Elucidating the role of *O*-GlcNAc glycosylation in neurobiology and neurodegeneration

Thesis by

Elizabeth Hwang Jensen

In Partial Fulfillment of the Requirements for

the degree of

Doctor of Philosophy

Biochemistry and Molecular Biophysics

Caltech

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2017

Defended November 13, 2017

© 2017

Elizabeth Hwang Jensen
ORCID: 0000-0002-6177-4304

for PSSETS, HwangLink, and the Jensensations

ACKNOWLEDGEMENTS

ABSTRACT

O-GlcNAc glycosylation is a dynamic, inducible post-translational modification (PTM) essential for neuronal homeostasis and found on proteins associated with neurodegenerative diseases such as α-synuclein, amyloid precursor protein, and tau. Intracellularly, O-GlcNAc modification is cycled by two enzymes in mammalian cells: O-GlcNAc transferase (OGT) appends O-GlcNAc to serine or threonine residues and O-GlcNAcase (OGA) removes O-GlcNAc. OGT modifies over 1000 different proteins, but the lack of a well-defined consensus sequence or substrate structural constraints has hampered efforts to predict sites *a priori*. Furthermore, the identification of O-GlcNAc modification sites has been obstructed by the difficulty of enriching and detecting O-GlcNAc using traditional biochemical methods. Here, we established and employed biological and chemical tools to illuminate the role of O-GlcNAc in neuronal function.

In Chapter 2, we sought to determine the role of O-GlcNAc in learning, memory, and neurodegeneration. Deletion of the OGT gene causes early postnatal lethality in mice, complicating efforts to study O-GlcNAc glycosylation in mature neuronal function and dysfunction. We demonstrated that the loss of OGT in the forebrain of adult mice (OGT cKO) leads to progressive neurodegeneration, including neuronal death, neuroinflammation, hyperphosphorylated tau, amyloidogenic Aβ-peptides, and memory deficits. In the hippocampus, we showed that OGT ablation lead to the upregulation of neuroinflammatory genes and the downregulation of cholesterol biosynthetic genes. Additionally, a gene network analysis (WGCNA), qPCR, and immunohistochemistry (IHC) revealed that loss of O-GlcNAc perturbed cell cycle progression in the hippocampal neurons. In the hippocampus, we identified increased neuroinflammatory gene transcription

in OGT cKO mice and both tau neurofibrillary tangle (NFT)-forming and amyloid-forming Alzheimer's disease (AD) mouse models. However, only OGT cKO and NFT-forming mice displayed decreased synaptic gene expression, suggesting that NFT formation and OGT cKO compromise hippocampal synaptic transcription. These studies indicate that *O*-GlcNAcylation regulates pathways vital for the maintenance of neuronal health and suggest that dysfunctional *O*-GlcNAc signaling may be an important contributor to neurodegenerative diseases.

In order to understand the critical *O*-GlcNAc-mediated neuronal functions that underlie OGT cKO dysfunction, we next developed and utilized novel biological and chemical tools in order to identify key OGT interactors and substrates in the brain in Chapter 3. Due to the lack of a well-defined OGT substrate sequence and structural constraints, OGT is believed to obtain its substrate specificity through its interactome where specific interactors target OGT to specific substrates. In order to identify these interactors, we used CRISPR/Cas9 to generate a novel mouse with a minimally tagged OGT in order to identify the endogenous OGT brain interactome using tandem affinity purification and MS methods. The preliminary OGT brain interactome consisted of previously identified OGT interactors and substrates as well as novel interactors. The identified OGT interactors were enriched for ribosomal and cytoskeletal proteins in addition to axonal, dendritic, and neuronal cell body proteins, implicating OGT as a pivotal mediator of neuronal structure and function.

In addition to the OGT interactome, we sought to uncover OGT's substrates or the *O*-GlcNAcome. We developed an improved approach to quantitatively label and enrich *O*-GlcNAcylated proteins for site identification. Chemoenzymatic labeling followed by Cu(I)-

catalyzed azide-alkyne cycloaddition (CuAAC) installed a new MS-compatible linker designed for facile purification and release of *O*-GlcNAcylated proteins for downstream MS analysis. We validated the approach by identifying several established *O*-GlcNAc sites on the proteins α-crystallin and OGT as well as discovering new, previously unreported sites on both proteins. Notably, these novel sites on OGT lie in key functional domains of OGT, underscoring how this site identification method can reveal important biological insights into protein activity and regulation.

Finally, in Chapters 4 and 5, we focus on the post-translational modification (PTM) code on a specific transcription factor (TF), CREB (cAMP response element binding protein). CREB regulates memory formation through its transcriptional control of neuronal metabolism, activity, differentiation, development, and survival. CREB phosphorylation at serine 133 has been previously shown to enhance CREB-mediated transcription while CREB glycosylation at serine 40 has been shown to decrease CREB-mediated transcription. However, the exact gene networks modulated by and potential interplay between CREB glycosylation and phosphorylation have not been explored. Through differential expression analysis with glycosylation-deficient (S40A) and phosphorylation-deficient (S133A) CREB mutants, we showed that CREB *O*-GlcNAcylation is important for neuronal activity and excitability while phosphorylation at serine 133 regulated the expression of genes involved in neuronal differentiation. Using WGCNA, we demonstrated that CREB *O*-GlcNAcylation at serine 40 and phosphorylation at serine 133 mediate mutually exclusive gene networks. The glycosylation-deficient mutant enhanced neuronal activity- and excitotoxicity-related gene networks while the phosphorylation-deficient mutant perturbed neuronal differentiation and amino and fatty acid metabolism-related

gene networks. Our work sheds light on the regulation of CREB through PTMs to modulate neuronal function and delineate the roles of *O*-GlcNAcylation and phosphorylation in modulating neuronal excitability and neuronal development and metabolism respectively. Altogether, these studies demonstrate that *O*-GlcNAc modification is a critical mediator of neuronal homeostasis and neurodegeneration.

PUBLISHED CONTENT AND CONTRIBUTIONS

Wang, AC, **Jensen EH**, Rexach JE, Vinters HV, and Hsieh-Wilson LC. "Loss of *O*-GlcNAc glycosylation in forebrain excitatory neurons induces neurodegeneration." *PNAS*. **2016**, *113*(52):15120-15125. doi: 10.1073/pnas.1606899113.

 E.H.J. participated in the execution of the experiments including the microarray, qPCR, IHC, and western blotting experiments, data analysis, and in the writing of the manuscript. The article, including figures, is reproduced in part within Chapter 2 with permission under the *PNAS* rights and permissions.

Griffin ME, **Jensen EH**, Mason DE, Jenkins CL, Stone SE, Peters EC, Hsieh-Wilson LC. "Comprehensive mapping of O-GlcNAc modification sites using a chemically cleavable tag." Mol. Biosys. 2016, 12 (6): 1756-1759. doi: 10.1039/c6mb00138f. Research article.

 E.H.J. participated in the execution of the experiments including the peptide and protein chemoenzymatic labeling and preparation for mass spectrometry and in writing of the manuscript. The article, including figures, is reproduced in part within Chapter 6 with permission under a Creative Commons Attribution 3.0 Unported License.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

NOMENCLATURE

A or Ala	alanine or adenosine
a.a.	amino acids
ACN	acetonitrile
AD	Alzheimer's disease
ALS	amyotrophic lateral sclerosis (a.k.a.Lou-Gehrig's disease or motor neuron disease)
AMPA	
Anti-anti	anti-mycotic, anti-microbial
B	biotin/biotinylated
BCA	bicinchoninic acid
BDNF	brain-derived neurotrophic factor
BEMAD	β-elimination Michael addition
BG	E11.5 basal ganglia
BM	bone marrow
Bp	base pair(s)
BSA	bovine serum albumin
C or Cys	cysteine or cytosine
CBP	CREB-binding protein
CID	collision-induced dissociation
CNs	cortical neurons (either E15.5 or E16.5)
Cre	Cre recombinase
CRE	consensus sequence for CREB binding (Full site: TGACGTCA, Half site: TGACG/CGTCA)
CREB	cAMP response element binding protein
CRISPR	clustered regularly interspaced short palindromic repeats
CuAAC	copper-catalyzed azide-alkyne cycloaddition
Cu(I)	copper (I), Cu^+
$CuSO_4$	copper sulfate
DAPI	4',6-diamidino-2-phenylindole
Dde	1-(4,4-dimethyl-2,6-dioxocyclohex-1-ylidene)ethyl
Ddv/ivDde	1-(4,4-dimethyl-2,6-dioxocyclohex-1-ylidene)isovaleryl
DE	differentially expressed
DMEM	Dulbecco's modified Eagle medium
DNA	deoxyribonucleic acid
dsDNA	double stranded DNA
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EGTA	ethylene glycol-bis(β-aminoethyl ether)tetraacetic acid
E#	embryonic day #

EOGT	EGF domain-specific *O*-linked *N*-acetylglucosamine transferase (extracellular)
ESC(s)	embryonic stem cell(s)
ETD	electron transfer dissociation
EtOH	ethanol
FBS	fetal bovine serum
FDR	false discovery rate
FLAG	protein tag with the sequence DYKDDDDK
FTD(P)	frontotemporal dementia (with Parkinsonism)
G or Gly	glycine or guanosine
Gal	galactose
GalN	galactosamine
GalNAc	*N*-acetylgalactosamine
GalNAz	*N*-azidoacetylgalactose
(Y289L) GalT	(Y289L) β-1,4-galactosyltransferase
gDNA	genomic DNA
GFP	green fluorescent protein
GlcNAc	*N*-acetylglucosamine
GlcNAz	*N*-azidoacetylglucosamine
HA	hemagglutinin tag (YPYDVPDYA)
HBSS	Hank's buffered saline solution
HCD	higher-energy collisional dissociation
HD	Huntington's disease
HDR	homology-directed repair/recombination
HEK293	human embryonic kidney cells
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HSV	replication-deficient herpes simplex virus
ICC	immunocytochemistry
IF	immunofluorescence
IHC	immunohistochemistry
Ile (or I)	isoleucine
IRDye800	infrared dye 800
Leu (or L)	leucine
IgG	immunoglobulin domain G
IgM	immunoglobulin domain M
©KO	(conditional) knockout
lncRNA	long-noncoding ribonucleic acid
LTM	long term memory
LTQ	linear trap quadrupole
LSM	laser scanning microscope
LV	lentivirus
LWAC	lectin weak affinity chromatography
MAPK	mitogen-activated protein kinase
MEF(s)	mouse embryonic fibroblast(s)
MeOH	methanol

Met (or M)	methionine
MS	mass spectrometry
NaAsc	sodium ascorbate
NFTs	neurofibrillary tangles
NHEJ	non-homologous end joining
NMR	nuclear magnetic resonance
NPCs	neural progenitor cells
nt	nucleotide
O-GlcNAc	*O*-linked *N*-acetylglucosamine
OGA	*O*-GlcNAcase
(s)OGT	(short isoform of) *O*-GlcNAc transferase
PAM	protospacer motif
PARP1	poly-ADP ribose polymerase 1
PBS(T)	phosphate buffered saline (Tween 20)
PC	photocleavable
PC12	pheochromocytoma 12 (cell line)
PCR	polymerase chain reaction
PD	Parkinson's disease
PEG	polyethylene glycol
Phe (or F)	phenylalanine
PI3K	phosphatidylinositol 3-kinase
P/S	penicillin/streptomycin
PSP	progressive supranuclear palsy
PTM	post-translational modification
PUGNAc	*O*-(2-acetamido-2-deoxy-D-glucopyranosylidene)amino-*N*-phenylcarbamate
qRT-PCR (qPCR)	quantitative reverse transcription- polymerase chain reaction
RapiGest	sodium 3-[(2-methyl-2-undecyl-1,3-dioxolan-4-yl)methoxy]-1-propanesulfonate
RNA	ribonucleic acid
mRNA	messenger ribonucleic acid
rRNA	ribosomal ribonucleic acid
RNAi	RNA interference
RT	room temperature
siRNA	short interfering RNA
SCX	strong cation exchange
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate protein acrylamide gel electrophoresis
S.E.M	standard error of the mean
Ser (or S)	serine
ssODN	single-stranded oligonucletide
STM	short term memory
TAE	40 mM Tris base, 20 mM acetic acid, 1 mM EDTA, pH 8.3
TBS(T)	Tris buffered saline (Tween 20)
TFA	trifluoroacetic acid

THPTA	tris(3-hydroxypropyltriazolylmethyl)amine
Thr (or T)	threonine
TNFα	tumor necrosis factor alpha
Trp (or W)	tryptophan
Tyr (or Y)	tyrosine
UDP	uridine diphosphate
(3' or 5')UTR	(3' or 5') untranslated region of RNA
UV	ultraviolet
(s)WGA	(succinylated) wheat germ agglutinin
Wnt	wingless-type MMTV integration site family member

C h a p t e r   1

The role of *O*-GlcNAc glycosylation in neurobiology and neurodegeneration

1.1 Introduction to glycobiology

The central dogma of biology begins with the DNA code, is followed by transcription to RNA, and ends with translation into proteins.[1] This well choreographed dance allows us to take the template-encoded blueprint inherited from our ancestors and execute complex, robust, and adaptive biological functions. In addition to the aforementioned major macromolecules, carbohydrates and lipids are the understudied brethren whose contributions to biology have been proven time and time again to be every bit as important as DNA, RNA, and proteins. Indeed, both glycans and lipids have been shown to be critical regulators of cellular fate and function and arbiters of disease.[2-4] In particular, despite the prevalence and importance of carbohydrates, their study has been less accessible than the central dogmatic macromolecules largely due to their heterogeneity and their non-template encoded nature. For these reasons, carbohydrates remain neglected and poorly understood in the context of biology.[5] Without the luxury of a blueprint or a "Rosetta Stone," researchers must fundamentally shift the way they approach the study of carbohydrates and develop and exploit novel chemical biological tools in order to decipher the language of the "glycocode" or "sugar code"− the concept that the specific glycan structure conveys biological information to cells.[5-7] Over the course of the next chapter, I will attempt to shed light on the intricacies and importance of a specific glycosylation modification, *O*-GlcNAc glycosylation.

1.2 Introduction to *O*-GlcNAc glycosylation

The post-translational modification (PTM) of serine or threonine residues with β-linked *N*-acetylglucosamine (*O*-GlcNAc) is cycled by two proteins in mammalian cells: *O*-GlcNAc transferase (OGT) and *O*-GlcNAcase (OGA), which catalyze the addition and

removal of *O*-GlcNAc respectively (Figure 1.1). *O*-GlcNAc glycosylation is a dynamic and inducible PTM found on over one thousand proteins and has been shown to critically regulate a variety of different cellular processes including transcription, epigenetics, translation, organelle biogenesis, cell cycle, autophagy, metabolism, protein turnover, apoptosis, and stress response.[8-11] Due to the importance of *O*-GlcNAc modification in most cellular processes, dysregulation of *O*-GlcNAcylation is associated with a variety of diseases including diabetes, cancer, and neurodegenerative diseases.[12] *O*-GlcNAc glycosylation has been observed across metazoans, insects, plants, fungi, and several bacteria.[12,13] While plants have two different genes that encode for OGTs, mammals have a single OGT-encoding gene for intracellular *O*-GlcNAc modification.[14,15] In 2011, a separate epidermal growth factor (EGF)-domain specific OGT (EOGT) protein was identified and shown to preferentially *O*-GlcNAcylate EGF-like repeats of extracellular and secreted proteins in the lumen of the endoplasmic reticulum (ER).[16,17] *O*-GlcNAc modification by EOGT has been shown to be important for cell-matrix interactions and notch signaling and its disruption leads to the congenital disorder, Adams-Oliver syndrome.[17-19]
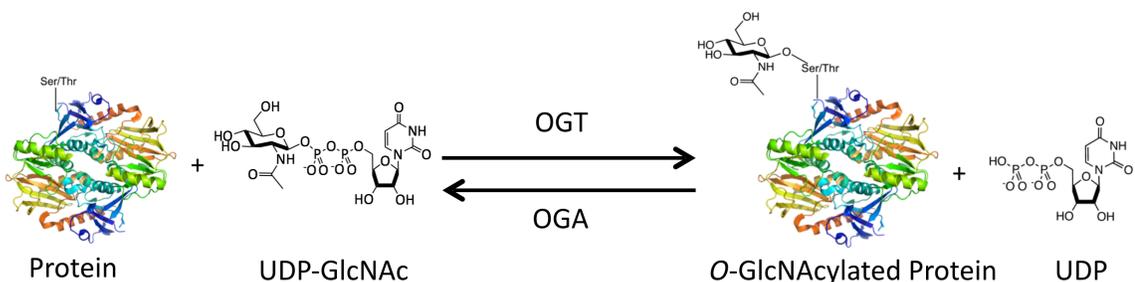
Figure 1.1 *O*-GlcNAc glycosylation. *O*-GlcNAc transferase (OGT) catalyzes the transfer of an *N*-acetylglucosamine (GlcNAc) from UDP-GlcNAc to a serine or threonine residue on a protein substrate (shown here is synapsin). *O*-GlcNAcase (OGA) catalyzes the removal of *O*-GlcNAc from protein substrates.

Despite the ubiquitous nature of this modification, *O*-GlcNAcylation was not discovered until the 1980s.[20] The delayed discovery of *O*-GlcNAc can be attributed to challenges inherent to the study of protein *O*-GlcNAc glycosylation; *O*-GlcNAc glycosylation (1) does not have a well-defined consensus sequence; (2) is chemically labile and vulnerable to hydrolysis during cell lysis; (3) adds a small and uncharged moiety to proteins precluding separation using electrophoretic gels; (4) like other PTMs, is substoichiometric on most proteins making MS detection difficult; and finally (5) cannot be separated from unglycosylated forms in reverse phase high performance liquid chromatography (HPLC).[21,22] Due to these difficulties, site detection has remained difficult, and novel techniques have been developed to detect and quantify endogenous *O*-GlcNAcylated sites and proteins, which we will discuss in section 1.5.

1.3 The enzymes that cycle *O*-GlcNAc: OGT and OGA

Notwithstanding the difficulties in determining *O*-GlcNAcylated proteins and sites, several properties of *O*-GlcNAc modification have been characterized based on the characterization of the enzymes that cycle *O*-GlcNAc. Human OGT is active as a homodimer or heterotrimer and is alternatively spliced to produce three isoforms: a 116-kDa nuclear/cytoplasmic isoform (ncOGT), a 103-kDa mitochondrial isoform (mOGT), and a 70-kDa short isoform (sOGT).[23,24] Recently, Conrad and colleagues showed alternative splicing further regulates OGT under high *O*-GlcNAc levels through the retention of an intron that leads to decreases in OGT mRNA levels.[25] The crystal structure of a truncated human OGT bound to UDP and a peptide substrate has recently been determined (Figure 1.2).[26] OGT has a C-terminus with a catalytic domain, a phosphatidylinositol (3,4,5)-triphosphate (PIP$_3$)-binding domain, and an N-terminal

protein-protein tetratricopeptide repeat (TPR) domain.[27] While most ncOGT is localized

in the cytoplasm and nucleus, ncOGT can be targeted to the plasma membrane through

its PIP$_3$-binding domain in response to insulin signaling.[28] OGT's TPR domains form a

superhelical structure similar to importin-α, which allows OGT to mediate protein-protein

interactions through asparagines along one face.[29] Ablation of these asparagines disrupts

protein-protein interactions and decreases *O*-GlcNAcylation globally, demonstrating that

the TPR domains are critical for its recruitment to substrates.[30]
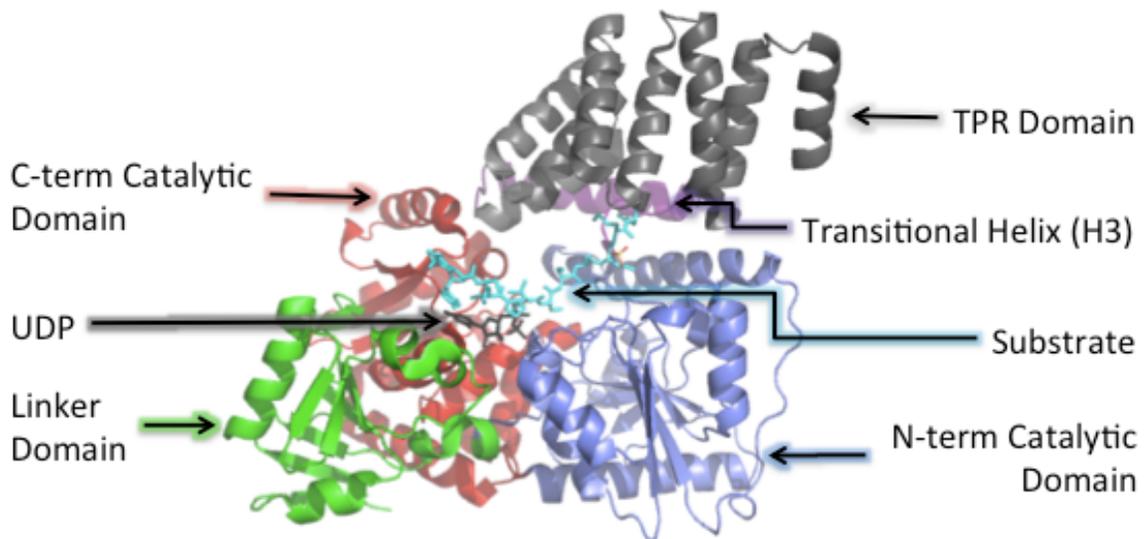
Figure 1.2. Structure of human OGT with UDP and peptide substrate. The crystal structure of hOGT displays 4.5 TPR (gray) although the number of TPR motifs differs for each isoform where ncOGT, mOGT, and sOGT, have 11.5, 9.5, and 2.5 TPR motifs respectively. The transitional helix (H3), peptide substrate, N-term catalytic domain, C-term catalytic domain, UDP, and linker domain are shown in purple, cyan, blue, red, black, and green respectively.[26]

A recent study showed that OGT can also *O*-glucosylate proteins with 25-fold less

efficiency than its *O*-GlcNAcylation activity, suggesting that OGT may also catalyze the

addition of *O*-Glc to proteins. However, more studies are required to definitively

demonstrate *in vivo* OGT *O*-glucosylation catalytic activity.[31] In addition to its

glycosyltransferase activity, OGT catalyzes the proteolytic cleavage and maturation of

host cell factor-1 (HCF-1) in its active site through a unique mechanism whereby an

HCF-1 glutamate is glycosylated followed by pyroglutamate formation and spontaneous backbone cleavage.[32-34] While this is the first demonstration of this cleavage mechanism, glutamate glycosylation-mediated amide hydrolysis may be a general cleavage mechanism employed by OGT or other glycosyltransferases.[33]

OGT's opposite, OGA, is a 103-kDa protein with an N-terminal glycoside hydrolase (GH) catalytic domain (residues 60-366), a nonfunctional C-terminal histone acetyltransferase (pseudoHAT) domain (residues 707-916), and a caspase-3 cleavage site in the stalk domain between the two domains (residues 367-706) (Figure 1.3).[35,36] Cleavage of OGA by caspase-3 at residue 413 occurs during apoptosis, producing two segments that remain tightly associated and maintain the same enzymatic activity.[37] Recently, three research groups independently solved the crystal structure of human OGA with the removal of the flexible linker region in the stalk domain (lacking residues ~400~535).[36,38,39] These structures revealed that OGA acts as a homodimer where the active site is composed of the catalytic domain of one monomer and the C-terminal helical bundle of another monomer.[39] There is also a shorter alternatively-spliced isoform of OGA (sOGA) that only contains the N-terminal O-GlcNAcase domain.[40] Full-length ncOGA is localized to the cytoplasm and nucleus while sOGA associates with the ER and lipid droplets.[41,42] Within the nucleus, OGA and OGT are often associated in complexes especially at transcription sites allowing for rapid O-GlcNAc cycling.[12] OGT and OGA together form a binary complex (a.k.a. the "O-GlcNAczyme"), which has been shown to be important for estrogen-dependent signaling, mammary development, and viral pathogenesis.[43,44] Importantly, OGT and OGA levels are carefully regulated transcriptionally and translationally in order to maintain proper O-GlcNAc levels.[25]
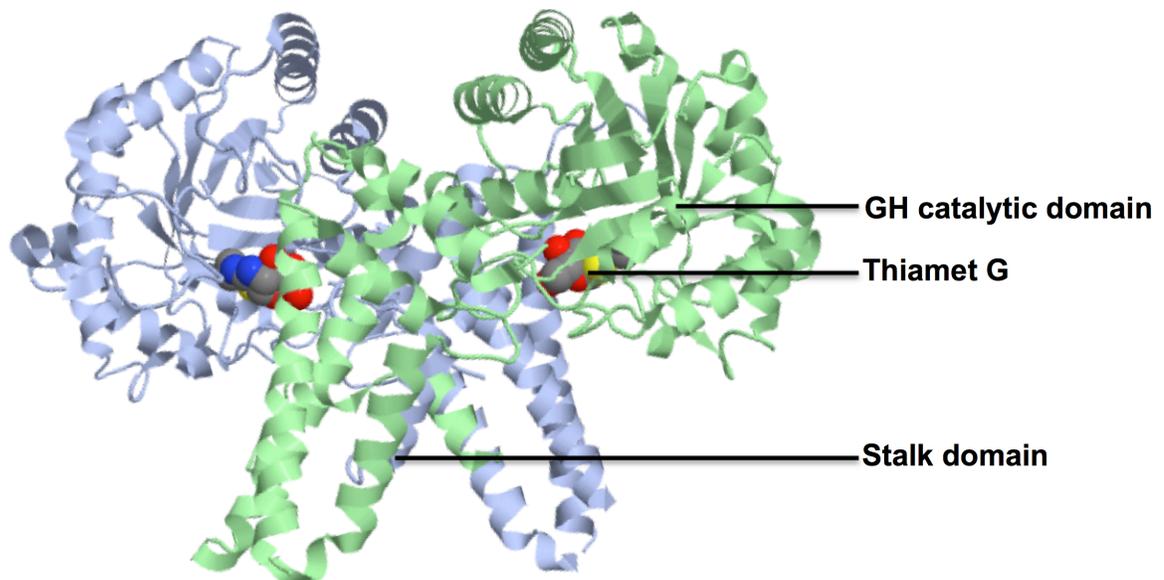
Figure 1.3 Structure of OGA homodimer. Shown here is the homodimer crystal structure of OGA including the N-terminal glycoside hydrolase (GH) catalytic domain and the separated stalk domain with the inhibitor Thiamet G (TMG). This figure was made using FirstGlance in Jmol using the crystal structure from Jiang and colleagues (PDB: 5UN9).[36,45]

1.4 Methods for the OGT interactome and *O*-GlcNAcome and the OGT substrate specificity hypothesis

The *O*-GlcNAc modification is often compared to phosphorylation due to the fact that it occurs dynamically on serine or threonine residues. However, unlike phosphorylation where there are over 500 of serine/threonine kinases and phosphatases each with their own substrate specificity, only OGT and OGA catalyze the addition and removal of *O*-GlcNAc on intracellular proteins. Several labs have attempted to determine the substrate specificity of OGT, but OGT appears to lack a well-defined consensus sequence although about half of substrates contain a PVS type motif.[12] A more recent study using a peptide library to screen substrates showed that the active site of OGT constrains the OGT-recognition sequence to [TS][PT][VT]**S/T**[RLV][ASY] (probability cutoff of 0.5, in bold is *O*-GlcNAcylated residue).[46] Based on known *O*-GlcNAc sites, several groups have developed computational methods to predict sites, but due to the

ambiguity of the recognition site, these computational methods fail to robustly predict sites, although methods may improve as more sites are identified.[47-50] Aside from the modification site sequence, studies have shown that many substrates have a disordered region C-terminal to the *O*-GlcNAcylation site while other studies have observed that substrates appear to have secondary structure surrounding the site such as coils.[30,51] However, studies have shown that there is no definite three-dimensional structural requirements for the substrate.[51] Additionally, OGT has been shown to *O*-GlcNAcylate proteins co-translationally.[52] Altogether, these studies show that OGT is an inherently promiscuous enzyme that lacks well-defined sequence or structural requirements for its substrates.

Given the high numbers of *O*-GlcNAc substrates and the lack of a consensus sequence or structural constraints, how then does OGT determine which substrates to modify? One prevailing theory in the *O*-GlcNAc field hypothesizes that OGT is recruited or targeted to specific substrates through specific interactors (Figure 1.4). If this hypothesis were true, abolishing the interaction between OGT and a specific interactor would result in reduction in the modification of specific substrates without interfering with other substrates' modification. Indeed, there may exist "hub" interactors, which are critical for recruitment of OGT to a large number of substrates. Several studies have supported this OGT interactor-substrate hypothesis by removing of portions of the OGT protein-binding TPR domains that are responsible for the specific interaction with certain interactors and demonstrating alterations in the *O*-GlcNAcylation of certain substrates. For example, ablation of the OGT TPR domain that interacts with TET2 results in a decrease in the *O*-GlcNAcylation serine 112 of histone 2B (H2B).[53] This TET2-

dependent *O*-GlcNAcylation of the H2B substrate occurs independently of TET2 catalytic activity suggesting that the TET2 interactor presence rather than activity is responsible for recruitment of OGT to this substrate.[53] The chromatin deacetylating mSin3a-HDAC complex interacts with TPR 1-6 region of OGT. Through ablation of OGT TPR 1-6, researchers demonstrated that the protein mSin3a recruits OGT to various substrates including RNA polymerase II and Sp1.[54] Overall, the TPR regions of OGT mediate its protein-protein interactions, which govern downstream substrate *O*-GlcNAcylation.
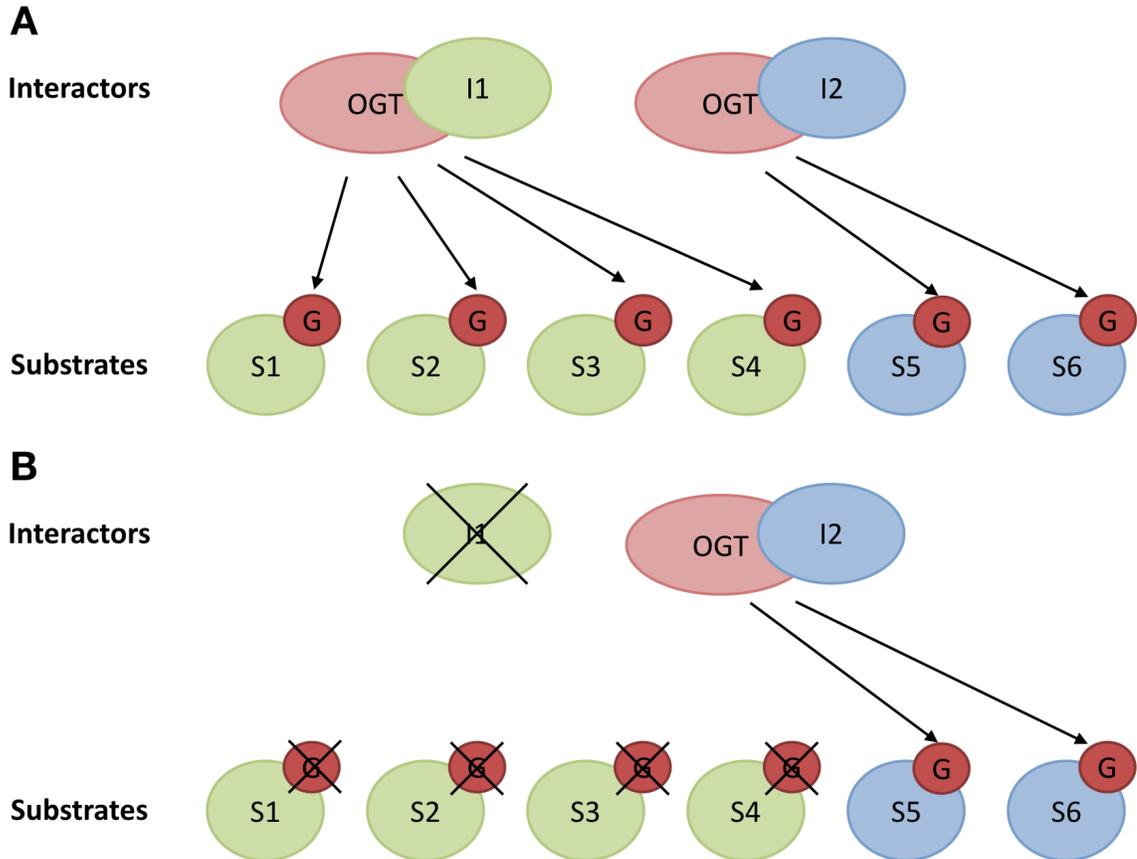
Figure 1.4 OGT substrate specificity hypothesis. The major hypothesis for OGT specificity involves OGT interacting with key "hub" interactors through its TPR domains. (A) These interactors then recruit OGT to certain substrates resulting in their *O*-GlcNAcylation. (B) If the interactor-OGT interaction is ablated, the targeted substrates will no longer be *O*-GlcNAcylated. I = interactor, S = substrate, G = *O*-GlcNAc.

Due to the importance of interactor-mediated recruitment of OGT to substrates, several attempts have been made to identify the global OGT interactome. The first such attempt was performed by Gerald Hart's lab, where a yeast two-hybrid study identified 25 different putative OGT interactors. They then verified that several of these putative interactors including MYPT1 and CARM1 interact *in vivo* with OGT and recruit OGT to substrates.[55] This provided a foundation for the next evolution in OGT interactome identification, which was performed by the Yang lab using a tandem affinity purification method. In this study, OGT-FLAG-HA was overexpressed in HEK293T cells, enriched using FLAG and then HA tag pull down, and then identified using MS. Using GFP-FLAG-HA as a control, this study found 853 putative OGT interactors and further verified that OGT interacts with HCF-1, which targets OGT to glycosylate peroxisome proliferator-activated receptor γ coactivator 1α (PGC-1α).[56] A final approach for the OGT interactome involved the use of a human proteome microarray, which identified 25 putative OGT-binding proteins.[57] These studies help to increase the total number of putative OGT interactors although more studies are required in order to validate these putative interactors and expand the interactome for different species and cell types.

As the lack of a conserved consensus sequence makes it difficult to predict OGT's targets *a priori*, MS methods are currently necessary to directly, efficiently, and globally determine sites of glycosylation. However, *O*-GlcNAc modifications are difficult to detect using traditional MS methods due to several properties of *O*-GlcNAc. First, *O*-GlcNAc and glycans generally are extremely labile during harsh mass spectrometric (MS) conditions precluding the use of higher-energy collisional induced dissociation (CID). Furthermore, *O*-GlcNAc is a large, uncharged moiety that leads to ion suppression

with the unglycosylated peptide and an inability to separate from unglycosylated forms in reverse phase high performance liquid chromatography (HPLC); this is exacerbated further by its substoichiometry on many proteins.[22] In order to circumvent these MS limitations, novel O-GlcNAc enrichment-LC/tandem MS methods have been developed to identify O-GlcNAc modified proteins and their modification sites.[58,59] The first method developed to successfully identify O-GlcNAc sites was β-elimination followed by Michael addition with dithiothreitol (DTT) (BEMAD) whereby the O-GlcNAc site is eliminated using DTT leaving behind dehydrated serine or threonine residues.[60] With BEMAD, external validation of the sites are required because false positives are possible with this method as any other eliminated serine or threonine modification will produced the same signature. Since the introduction of BEMAD, many other MS-based methods have been developed, including WGA lectin enrichment, mutant OGA enrichment, metabolic labeling, chemoenzymatic labeling, IsoTag, and native methods often coupled with lower energy electron-transfer dissociation (ETD) MS.[61-69] The advantages and disadvantages of many of these methods are discussed in detail in several reviews.[70,71]

In particular, our lab and others have utilized novel chemoenzymatic labeling methods in order to identify the O-GlcNAcome.[72] In this method, a mutant enzyme, GalT, specifically recognizes and appends an N-azidoacetylgalactosamine (GalNAz) moiety to O-GlcNAcylated proteins or peptides (Figure 1.5). Then, bioorthogonal "click chemistry" (copper (I)-catalyzed azide-alkyne cycloaddition, CuAAC) can be performed to append an alkynyl linker to the O-GlcNAc for subsequent O-GlcNAcome enrichment (Figure 1.5). Importantly, both the enzymatic and chemical steps are bioorthogonal and occur stoichiometrically.[73] While the enzyme and CuAAC conditions are consistent

11

across methods, many different linkers have been utilized, which commonly have: (1) a biotin (for streptavidin bead enrichment), (2) alkyne (for CuAAC), and (3) a cleavable linker portion that enables facile elution. Several different cleavable linkers have been used for *O*-GlcNAcome enrichment including a photocleavable linker and an acid-cleavable linker.[22,67] While these approaches have yielded several novel *O*-GlcNAcylation sites, these linkers can suffer from incomplete cleavage upon elution, hydrolysis of *O*-GlcNAc under harsh cleavage conditions, or off-target selectivity or reactivity.[74] Ideally, the linker will overcome these challenges and also confer MS advantages upon cleavage such as (1) a positive charge to enable facile MS identification like with the photocleavable linker, (2) a small size tag to avoid complex fragmentation, (3) isotopic labeling for facilitated MS identification such as in the IsoTag method, and (4) differential labeling (heavy/light) for multiplexing and labeling of multiple different samples.[22,67] In order to fully realize these goals, it is critical to synthesize and/or apply novel linkers toward *O*-GlcNAcome identification. Considerable variety is observed in the *O*-GlcNAcylation of proteins in different species, cell types, and conditions compelling researchers to determine the *O*-GlcNAcome in the specific conditions of interest. In Chapter 3, we will discuss the development of biological and chemical tools to identify the OGT interactome and *O*-GlcNAcome in the murine brain.
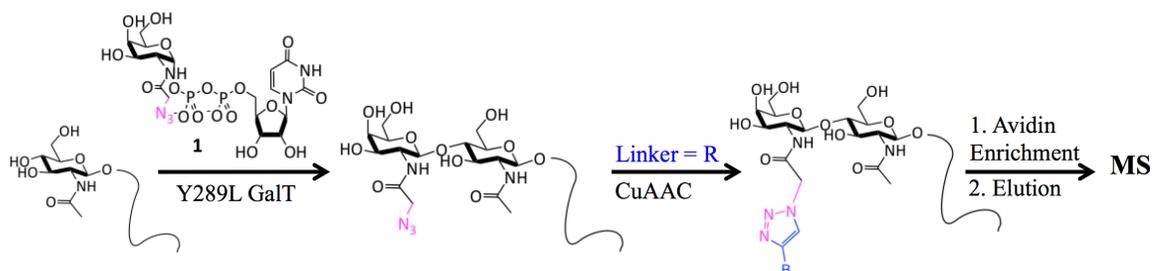
Figure 1.5 Overview of chemoenzymatic labeling for identification of the *O*-GlcNAcome. First, *O*-GlcNAcylated proteins or peptides are enzymatically labeled using Y289L GalT and **1** (UDP-GalNAz). After appendage of the azide handle, an alkynyl linker is added using Cu(I) azide-alkyne chemistry

(CuAAC), which appends a biotin. Finally, the *O*-GlcNAcylated peptides or proteins are enriched using streptavidin beads, mildly eluted in the case of a cleavable linker, and then prepared for mass spectrometric analysis.

1.5 *O*-GlcNAc crosstalk with other post-translational modifications

As we have touched upon before, the *O*-GlcNAc modification is akin to phosphorylation in that it is highly dynamic, inducible, and occurs on serine or threonine residues. In fact, *O*-GlcNAcylation and phosphorylation are often competing for the same residues where these two processes have a zen-like yin-yang relationship.[12] This relationship is borne out through three general mechanisms: (1) direct competition between phosphorylation and *O*-GlcNAcylation at the same site, (2) OGT association with phosphatases, (3) OGT crosstalk with kinases.

The first demonstration of direct *O*-GlcNAc and phosphate competition was found in the oncogenic transcription factor (TF) c-Myc. Phosphorylation of c-Myc at threonine 58 led to its proteosomal degradation thereby reducing c-Myc activity while *O*-GlcNAcylation at the same site increased its stability and activity.[75] Many more direct competition examples have been discovered since c-Myc with a recent proteomic study identifying a specific yin-yang phosphorylation/*O*-GlcNAcylation interplay motif that is enriched in the human proteome (40% of phosphorylated sites) supporting positive evolutionary selection.[76] Additionally, OGT and phosphatases form complexes that act on the same protein substrates at the same site; *id est*, the phosphatase dephosphorylates the site so that the associated OGT can add an *O*-GlcNAc moiety.[55,77,78] One example of this type of complex is seen in the OGT and PP1β/PP1γ (protein phosphatase 1) complex, which was shown to catalyze the dephosphorylation and subsequent *O*-GlcNAcylation of a peptide.[77] Finally, OGT has been shown to display considerable crosstalk with the

13

nutrient sensor adenosine-monophosphate-activated protein kinase (AMPK). AMPK-mediated phosphorylation of OGT disrupts OGT's interaction with chromatin, decreases H2B O-GlcNAcylation, and alters OGT's overall substrate specificity. In an interesting feedback mechanism, O-GlcNAcylation of AMPK increases AMPK's enzymatic activity.[79,80] Furthermore, in response to hyperglycemic conditions, the O-GlcNAc glycosylation of CaMKIIδ (calmodulin kinase IIδ) at serine 279 results in elevated CaMKIIδ activity exacerbating arrhythmogenesis in cardiomyocytes.[81] CaMKII is an important Ca^{2+}-dependent kinase that regulates cardiac and neuronal function, and its chronic activation leads to pathogenesis.[82,83] O-GlcNAcylation of CaMKIIδ links diabetic hyperglycemia to cardiac dysfunction and potentially neurodegenerative diseases. Similar to the case of AMPK, CaMKII phosphorylates and activates OGT, suggesting that this could be a general kinase/OGT feedback loop.[84] Overall, phosphorylation and O-GlcNAcylation are inextricably linked post-translational modifications that regulate both homeostatic cellular function and pathology.

Additionally, O-GlcNAcylation has been shown to have considerable crosstalk with ubiquitination and its associated ubiquitin-mediated proteosomal degradation.[10,76] First, O-GlcNAcylation of H2B at serine 112 facilitates its monoubiquitination at lysine 120 through enhancement of the interaction with the E3 ubiquitin ligase complex, BRE1A/1B. This monoubiquitination event does not affect H2B proteosomal degradation, but rather enhances transcriptional initiation and elongation.[85] Co-translational O-GlcNAcylation of the TF Sp1 spares it from ubiquitin-dependent proteosomal degradation, but also impairs its interaction with other transcription factors, thereby disrupting its transcriptional activity.[52,86] Similarly, the glycosylation of the

14

oncogenic TF p53 spares it from ubiquitin-dependent proteolysis.[87] However, unlike with Sp1, p53 glycosylation leads to enhanced TF activity leading to carcinogenesis.[87-89] The polycomb repressor complex (PRC) members, Bmi-1 (B lymphoma Mo-MLV insertion region 1 homolog) and EZH2 (enhancer of zeste homolog 2), are stabilized by *O*-GlcNAcylation leading to enhanced activity and alterations in oncogenesis.[90,91]

Furthermore, OGT is known to associate with and *O*-GlcNAcylate the deubiquitinase and tumor suppressor, BAP1 (BRCA1 associated protein 1), which regulates DNA replication, dsDNA break repair, and transcription.[92,93] *O*-GlcNAcylation of the TF PGC-1α encourages its association of BAP1, which prevents ubiquitin-mediated degradation of PGC-1α leading to PGC-1α-mediated gluconeogenesis.[56] *O*-GlcNAcylation of BAP1 itself results in BAP1 enhanced activity, which represses retinoic acid receptor (RAR) TF activity and stem cell differentiation.[92] BAP1 deubiquitinates and stabilizes OGT providing another feedback mechanism for OGT activity.[93] Indeed, BAP1 and OGT are often found together in large multiprotein complexes together and have been shown to target each other to different substrates.[94] *O*-GlcNAcylation has been found on dozens of other E1 ubiquitin ligases, E3 ubiquitin ligases, and deubiquitinases although the exact sites and biological consequences of *O*-GlcNAcylation have not yet been determined.[10] Through both specific *O*-GlcNAcylation events and the interaction between OGT and the enzymes that cycle phosphorylation and ubiquitination, *O*-GlcNAcylation displays considerable crosstalk between phosphorylation and ubiquitination.

1.6 Role of *O*-GlcNAc in cellular functions: development, survival, stress response, circadian rhythm, longevity, cell cycle, and protein turnover

In order to ascertain the role of *O*-GlcNAc in organismal development and survival, Marth and colleagues first attempted to knockout (KO) the X-linked *Ogt* gene in mice, but found that OGT KO is embryonic lethal.[95] When the Marth lab then created neuron-specific OGT conditional KO (nsOGT cKO) mice, they found that the nsOGT cKO mice were present at only 50% of the expected Mendelian frequencies. The nsOGT cKO led to postnatal death within 10 days due to a failure to develop proper locomotor activity. In addition, OGT knockout in T cells and fibroblasts resulted in apoptosis and growth arrest respectively.[96] Total OGA KO in mice is also embryonic lethal and displays impaired growth and postnatal death within 1 day due to delayed lung development.[97] Brain-specific KO of OGA in mice results in delayed brain differentiation and neurogenesis, obesity, and metabolic perturbations.[98] These OGT and OGA KO studies in mice revealed that *O*-GlcNAc is critical for organismal survival and development.

While most species like mice do not survive knocking out of OGT, an exception to this rule is *C. elegans*, which have led to its use as a model OGT KO organism. An *O*-GlcNAc ChIP-chip study (RL2) in wildtype (WT) *C. elegans* found that the *O*-GlcNAc resided upstream of genes involved in reproductive behavior, aging, axon/neuron differentiation, and glycolysis. When they compared the WT to *ogt-1* and *oga-1* mutants who expressed null/inactive OGT and OGA respectively, they found deregulation in genes involving aging, stress response, innate immunity, and metabolism.[99] A ChIP-chip study (WGA lectin) in *drosophila* showed that *O*-GlcNAc modified transcription factors associate with PREs (polycomb response elements), promoter elements for genes

16

responsible for homeotic maintenance. As in mice, knocking out the *drosophila* homologue for OGT, super sex combs (KO denoted *sxc*), is required for normal development as *sxc drosophila* die at the pharate adult stage and have highly altered homeotic states.[100] Indeed, many critical regulators of pluripotency and development such as Oct4 and Sox2 are regulated by *O*-GlcNAcylation supporting the demonstrated vital connection between *O*-GlcNAc modification and proper development.[101-105]

In addition to OGT and OGA KO creating developmental abnormalities, *O*-GlcNAc disruption undermines an organism's ability to respond to, recover from, and survive stress stimuli. Several studies have shown that dynamic *O*-GlcNAcylation is important for mounting a defense in cardiomyocytes against oxidative stress.[106-108] In both *ogt-1 C. elegans* (OGT KO) and cardiomyocyte-specific OGT KO mouse (cmOGT cKO) mouse studies, the OGT KO mutants were less tolerant to stress and had lower survival rates in response to stress.[99,109] The essential stress-related TFs, PGC-1α and PGC-1β, as well as their transcription target genes, were downregulated in cmOGT KO mice, which exacerbated heart failure.[99,109] OGT regulates stress through direct *O*-GlcNAcylation of stress-mediating TFs including PGC-1α, HCF-1, HIF-1α (hypoxia-inducible factor 1), and FoxO1/4 (forkhead box O1/4).[110-112]

In *C. elegans*, *daf-2* (IGF-1, insulin-like growth factor 1) mutants display enhanced dauer formation, which is a stress response where worms enter stasis and can survive harsh conditions. Dauer formation was further enhanced in a double *oga-1/daf-2* mutant compared to the *daf-2* single mutants, but reduced in *ogt-1/daf-2* mutants, suggesting that increasing *O*-GlcNAcylation facilitates the stress-induced survival response.[99] *O*-GlcNAcylation confers resistance to heat stress during development across many

different species from worms to zebrafish.[113] Across species, knocking out OGT leads to ER stress, apoptosis, and cell death through reduced Ca^{2+}- and Akt-based pro-growth signaling.[114-116] Consistent with a protective role for *O*-GlcNAc, elevation of *O*-GlcNAc levels confers resistance and protection to cardiomyocytes in response to ER stress.[117] In contrast, upregulation of OGT in hepatocellular carcinoma (HCC) increased palmitic acid levels, which in turn, led to elevated ER stress.[114] These studies show that a careful balance of *O*-GlcNAc is necessary for appropriate response to various stress stimuli and survival.

O-GlcNAc regulates stress response through many different mechanisms.[118,119] *O*-GlcNAc levels increase in response to physiological and heat stress, which protects cells from various stress stimuli.[115] In a global RNAi screen, OGT was found to be necessary for stress granule formation, a mechanism to conserve energy through decreasing protein synthesis.[120] Formed in response to various stress stimuli, stress granules are composed of P-bodies, mRNA processing enzymes, non-translating mRNAs, and translation initiation proteins.[121] Several key components of stress granules were found to be *O*-GlcNAcylated including RACK1 (receptor for activated C kinase 1), prohibitin-2, GAPDH (glyceraldehyde-3-phosphate dehydrogenase), and ribosomal proteins.[120] Another crucial stress response mechanism is autophagy; a process mediated by the formation of autophagosomes wherein cellular organelles and components are degraded through autophagosomal fusion with lysosomes.[122] *O*-GlcNAcylation has been shown to be necessary for autophagosome maturation through *O*-GlcNAcylation of SNAP-29 (synaptosomal-associated protein 29), a protein that controls the fusion of autophagosomes with lysosomes. Indeed, autophagy is critical for prevention of

aggregation of proteins in proteinopathies, like most neurodegenerative diseases, and is heavily dependent on OGT activity.[9] Across diverse cell types and organisms, O-GlcNAc has been shown to be a principle regulator of autophagy.[84,123-125] Thus, O-GlcNAcylation acts an important mediator of stress and autophagy, and its level must be carefully tuned to produce appropriate stress survival response.

Longevity and circadian rhythm are also regulated by O-GlcNAcylation. A study in C. elegans revealed that mutants lacking OGA had effects on overall longevity through its influence on insulin signaling.[99,126] In addition, O-GlcNAc modification of the C. elegans ortholog of TF Nrf2 (NF-E2-related factor 2) enhanced oxidative stress resistance and anti-aging effects of Nrf2-mediated transcription.[127] As mice age, the levels of O-GlcNAcylation gradually decrease along with ncOGT levels while the levels of sOGT increase in the brain.[128] This suggests that O-GlcNAc may play a different role in the aging process.

Aside from its role in aging and longevity, O-GlcNAc plays a prominent role in circadian rhythm.[129-131] In order to maintain proper circadian rhythm, the TF regulators BMAL1 and CLOCK transcribe the key regulators, period (Per1/2) and cryptochrome (Cry1/2), which in turn, repress BMAL1 and CLOCK.[132] Additionally, BMAL1 and CLOCK are phosphorylated by GSK3β, a critical regulator of circadian clocks. This leads to their proteosomal degradation and disrupted transcriptional activity.[133-135] These protein turnover and repressive feedback loop mechanisms generate oscillatory activity that gives rise to the 24-hour circadian rhythm within cells.

Altering O-GlcNAc levels changes the circadian rhythm in flies and mice; in fact, as the circadian period progresses, global protein O-GlcNAcylation changes as well.

19

OGT regulates circadian rhythm through the O-GlcNAcylation of PER2 (period 2), which competes directly with PER2 phosphorylation by casein kinase I (CKIδ) at S662. The O-GlcNAcylation of PER2 enhances its repressor activity thereby decreasing CLOCK/BMAL1-mediated transcription.[131] The effects of PER2 O-GlcNAcylation mirrors the effects of a S662G-PER2, which leads to reduced PER2 phosphorylation, increases PER2's transcriptional repression, and results in a familial advanced sleep phase disorder.[136,137] As the circadian cycle progresses, GSK3β phosphorylates OGT increasing OGT's activity.[131] Both BMAL1 and CLOCK are O-GlcNAcylated in response to high glucose leading to their stabilization likely through the recruitment of BAP1 for deubiquitination.[130] In addition, CLOCK coimmunoprecipitates with OGA at specific circadian times, suggesting that CLOCK is subject to temporally regulated O-GlcNAc removal.[131] The stabilization of PER2, BMAL1, and CLOCK through O-GlcNAc glycosylation in response to glucose produces a crucial link between the metabolic environment and circadian entrainment.

Cell cycle proteins are also subjected to extensive O-GlcNAc glycosylation (Figure 1.6).[138-141] Proper cell cycle advancement from G1 phase to S phase (DNA synthesis phase) to G2 phase (post-DNA replication) to M phase (mitosis) and back to G1 phase (or G0 phase) is required for control of appropriate cellular growth or senescence and prevention of the aberrant growth observed in cancer.[142] Manipulation of O-GlcNAc levels pharmacologically and genetically has been shown to perturb cytokinesis and mitosis through alterations of the expression of many different cell cycle regulators (Figure 1.6A).[97,138-140] O-GlcNAc levels on histones increase in G1 phase, decrease in S phase, and increase again in M phase, suggesting that O-GlcNAcylation mediates

chromatin remodeling for cell cycle progression.[143] Increasing OGT levels reduces the activity, mRNA, and protein levels of Cdk1 (cyclin-dependent kinase 1), an important regulator of the G1/S transition through M phase.[144]

The key cell cycle transcriptional regulators, E2F and Rb (retinoblastoma-like protein), are regulated by *O*-GlcNAc and regulate *O*-GlcNAcylation in turn. The G1/S transition is inhibited by E2F1 association with and sequestration by Rb (Rb1/2).[145] Phosphorylation of Rb by Cdk2/Cyclin E or Cdk4/6/Cyclin D complexes emancipates E2F1 from Rb permitting E2F1's unfettered pursuit of its cell cycle transcriptional programming (Figure 1.6B).[146] For the transition to S phase, E2F carries out its transcription program in complex with HCF-1, which, as we have discussed earlier, requires OGT for its proteolytic activation and is heavily glycosylated.[139,147] In addition, knockout of Rb1 or E2F1 leads to increased OGT and OGA expression.[148] Extensive *O*-GlcNAcylation of Rb1 has been observed in G1 phase, which decreases as cell cycle progresses, suggesting that *O*-GlcNAc may influence Rb1 activity in a cell-cycle dependent manner.[145] Interestingly, ablation of the serine 420 *O*-GlcNAcylation site of Rb2 impairs the activity of OGT supporting reciprocal regulatory roles for *O*-GlcNAc and Rb.[149]

While most cyclins and cyclin-dependent kinases (Cdk) have not yet been found to be *O*-GlcNAcylated, Cyclin K, Cdk12, Cdk13, Cyclin M1, and Cdk5 have been found to be *O*-GlcNAcylated.[150-152] Cyclin K and Cdk12/13 form a complex that phosphorylates RNA polymerase II, leading to enhanced transcriptional elongation (Figure 1.6B).[153] The activity of this cyclin complex is important for the transcription of DNA damage associated genes and therefore has been shown to be critical for DNA repair.[154] *O-*

GlcNAc, OGT, and OGA extensively regulate cell cycle-related transcription, chromatin stability, and DNA damage checkpoints that produce cell cycle arrest, which we will discuss in more detail in Chapter 4. The *O*-GlcNAc sites or peptides have been identified for Cyclin M1, Cdk5, Cdk12, Cdk13, but the sites on Cyclin K have not yet been determined.[150,151] Furthermore, the effects of *O*-GlcNAcylation on their cell cycle functions have not yet been explored. It would be interesting to elucidate the *O*-GlcNAcylation sites on Cyclin K, Cdk12, and Cdk13 and to determine the influence of *O*-GlcNAcylation on the complex's ability to mitigate DNA damage. In summary, *O*-GlcNAcylation is a central regulator of cytokinesis, but more information is needed to determine the role of specific *O*-GlcNAc modifications on cyclin and Cdk cell cycle regulators.
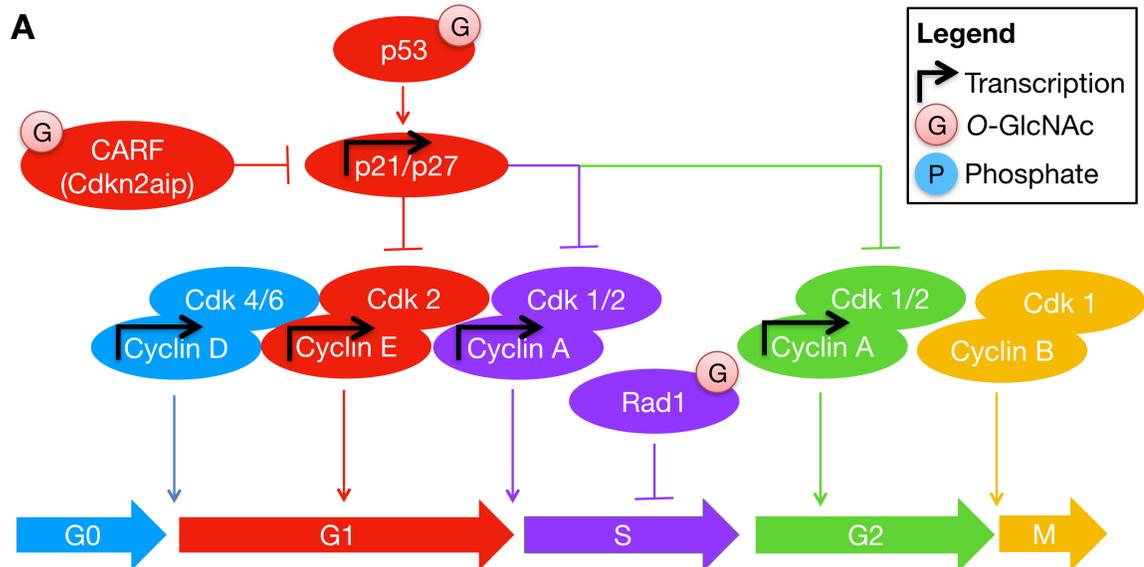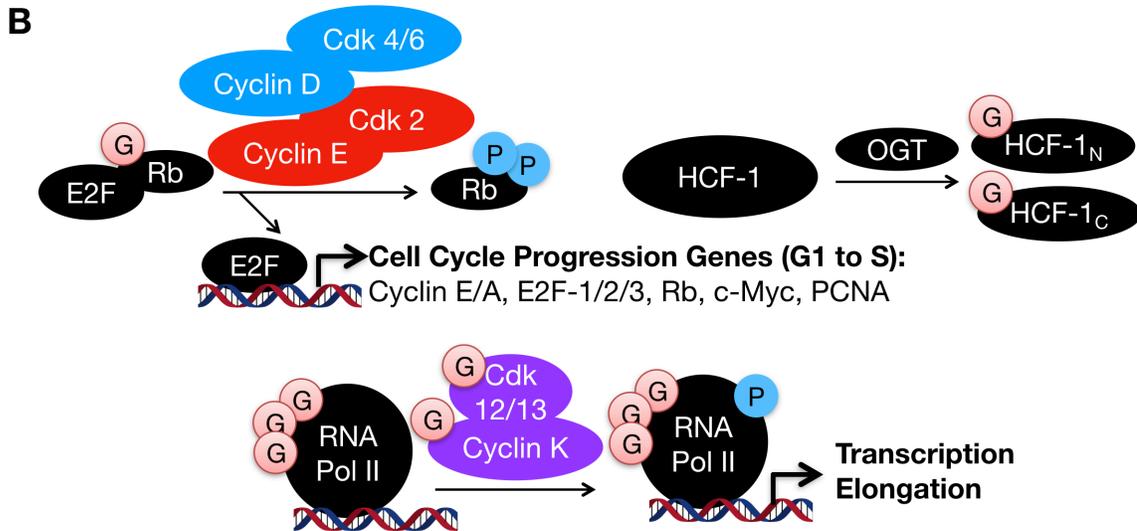
Figure 1.6 Cell cycle regulators are *O*-GlcNAcylated or regulated by *O*-GlcNAcylation. (A) Disrupting O-GlcNAc levels can lead to cell growth defects in a variety of organisms through alteration of the expression levels and *O*-GlcNAcylation of different cyclins, and cell cycle-related transcripts. (B) *O*-GlcNAc is involved in the transcriptional regulation of cell cycle. First, Rb is glycosylated in G1 phase, which is associated with association between Rb and E2F leading to inhibition of cell cycle transcription. Upon phosphorylation of Rb by the Cyclin D/Cdk4/6 or Cyclin E/Cdk2 complexes, E2F is freed from Rb sequestration and facilitates transcription of important cell cycle regulators. Additionally, OGT *O*-GlcNAcylates, proteolytically cleaves, and activates the critical cell cycle-regulating TF HCF-1. Finally, the Cdk12/13/Cyclin K complex mediates the phosphorylation of RNA polymerase II, which enhances transcription elongation. Cdk12, Cdk13, and Cyclin K were all found to be glycosylated, but the effect on their activity is not yet known.

O-GlcNAc regulates protein synthesis through its mediation of cell cycle, upstream growth-dependent signaling pathways, and the protein synthetic machinery. Protein synthesis is regulated in response to external stimuli such as growth factors and nutrients through intracellular signaling pathways such as the mitogen-activated pathway kinase (MAPK) or phosphatidyl inositide-3 kinase (PI3K) pathway.[155] After extracellular activation, a cascade of phosphorylation and dephosphorylation events are then integrated and converge upon the protein synthetic machinery to prepare for cellular growth.[156] Many of these growth-mediating kinases and phosphatases are *O*-GlcNAcylated (Figure 1.7).[21,157] Protein synthesis occurs in three major steps in eukaryotes: (1) initiation, the most heavily regulated step, which involves a large number of eukaryotic initiation factors (eIFs); (2) elongation, which is regulated by eukaryotic elongation factors (eEFs);

and (3) termination, which is controlled by eukaryotic release factors (eRFs).[158] *O*-GlcNAcylation also occurs on the majority of proteins in the ribosome and many of the proteins involved in translational initiation and elongation (Figure 1.7).[159] While *O*-GlcNAc heavily modifies these proteins, the exact consequences of these *O*-GlcNAc modifications have not yet been delineated.

In addition to regulating the ribosome and protein synthesis, protein turnover is regulated by *O*-GlcNAcylation. OGT cotranslationally modifies proteins such as Sp1 to stabilize them and prevent ubiquitin-mediated proteosomal degradation.[52] Indeed, the proteasome itself and proteosomal regulators are both known to be directly *O*-GlcNAcylated.[160,161] *O*-GlcNAcylation of the 26S and 19S proteasomes reduces proteosomal degradation.[160] The reduction of sOGA impairs proteosomal activity, leading to the accumulation of ubiquitinated proteins. When sOGA is localized to lipid droplets, proteosomal inhibition results in increases in the lipid droplet components, perilipin-2 and perilipin-3, which leads to enhanced lipid droplet formation.[41] Altogether, *O*-GlcNAcylation, OGT, and OGA must act in concert to regulate protein synthesis and degradation.
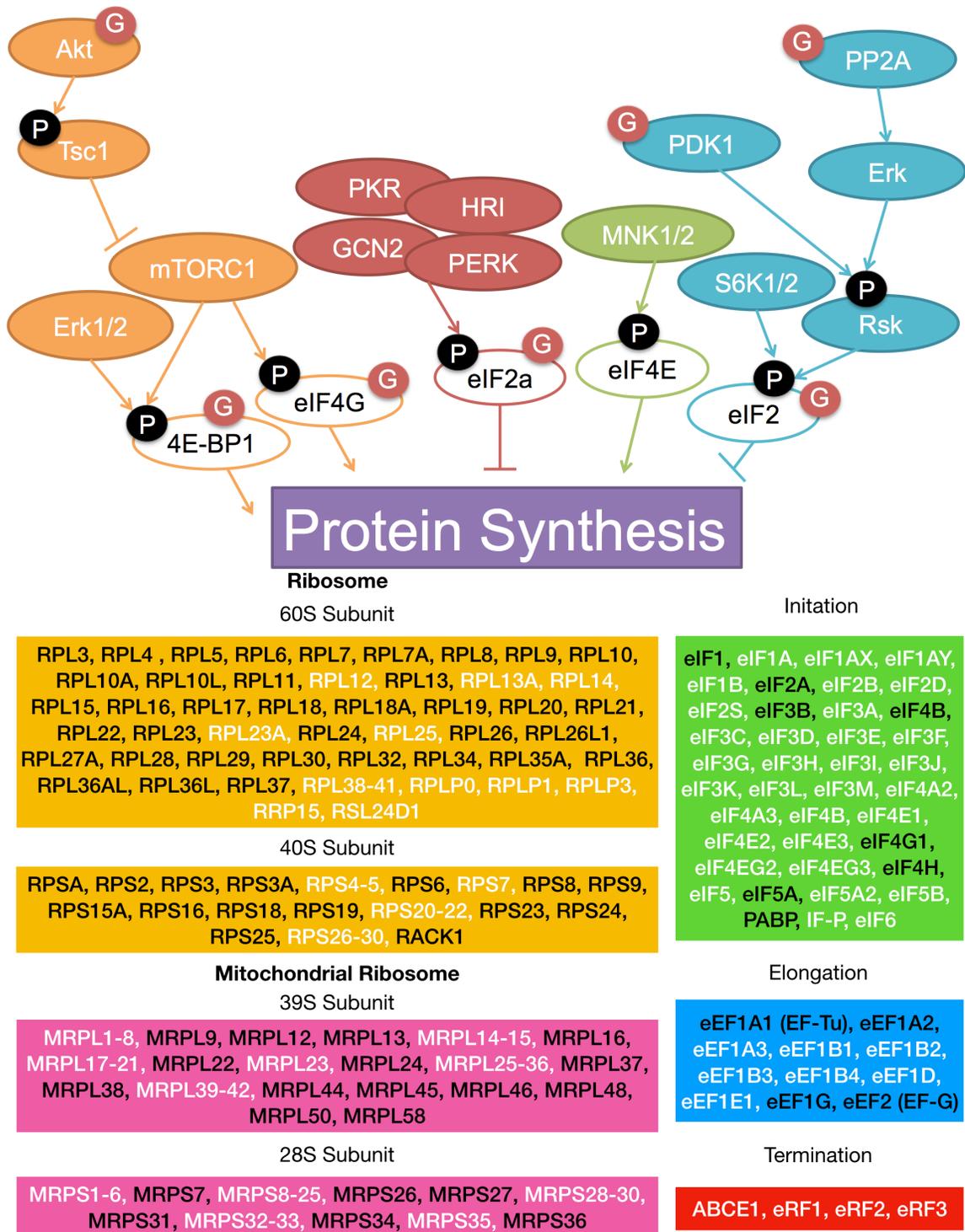
Figure 1.7 *O*-GlcNAcylation of protein synthesis upstream regulators and machinery. Shown here are the known glycosylated proteins for the upstream regulators of protein synthesis and the direct protein synthesis regulators including the ribosomal machinery. The critical phosphorylation events that modulate protein synthesis are shown as well as their ultimate effects on protein synthesis. The bottom image shows all the glycosylated proteins that are components of the ribosome and mitochondrial ribosome in addition to proteins involved in initiation, elongation, and termination of protein synthesis. The proteins written in

black are known to be glycosylated, while those written in white are not known to be glycosylated. G = GlcNAc, P = phosphate.

1.7 Role of *O*-GlcNAc in metabolic function and dysfunction

In eukaryotes, *O*-GlcNAcylation acts as a nutrient sensor due to its dependence on the abundance of UDP-GlcNAc, the end product of the hexosamine biosynthetic pathway (HBP) (Figure 1.8). The production of UDP-GlcNAc begins with the uptake of glucose, which can either enter glycolysis, pentose phosphate pathway (PPP), glycogen biosynthesis, or HBP. Specifically, the first three products in the HBP compete with these other pathways for substrates. The first committing and rate-limiting step of the HBP is catalyzed by glutamine:fructose amidotransferase (GFAT), which transfers an amine from a glutamine donor to fructose-6-phosphate in order to generate glucosamine-6-phosphate.[162] Along the HBP, the transformation of glucose to UDP-GlcNAc is reliant on the products of a variety of metabolic pathways, including amino acid, fatty acid, and nucleotide metabolism.[21] Researchers estimate 2-5% of glucose is shunted into the HBP to make UDP-GlcNAc levels the second highest metabolite population after ATP implicating *O*-GlcNAc glycosylation as an important PTM and nutrient sensor.[21] Indeed, *O*-GlcNAc, OGT, and OGA levels are all dependent upon the intracellular UDP-GlcNAc concentrations and therefore HBP flux.[25,162]
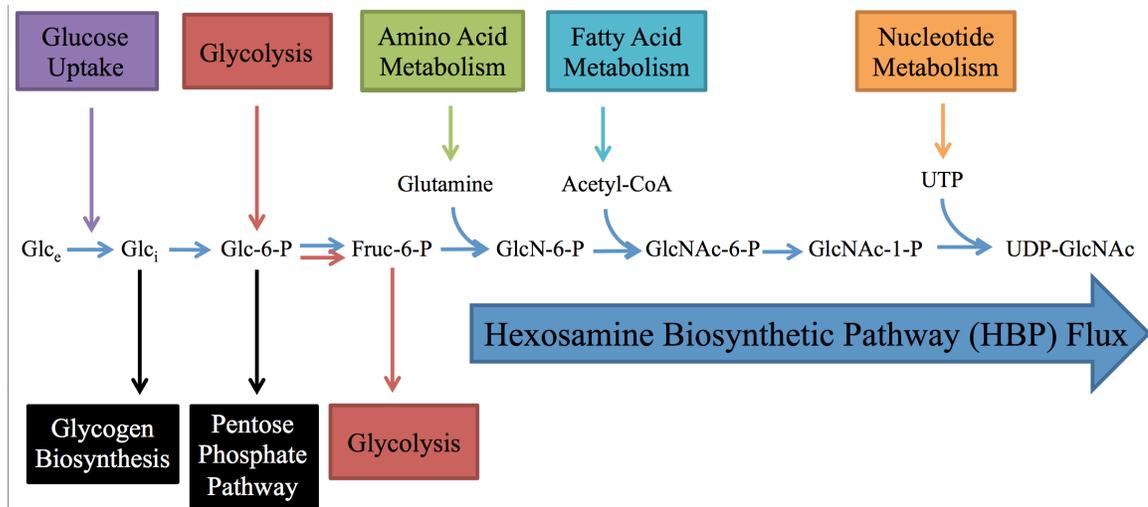
Figure 1.8 The hexosamine biosynthetic pathway (HBP) requires input from many different metabolic pathways. The HBP begins with the uptake of extracellular glucose ($Glc_e \rightarrow Glc_i$) followed by two glycolytic steps. The first 3 products are depleted when they enter the competing metabolic processes, glycogen biosynthesis, pentose phosphate pathway, and glycolysis. The first committed and rate-limiting step of HBP occurs when Fruc-6-P is converted to GlcN-6-P. The next HBP steps require metabolites that involved in amino acid, fatty acid, and nucleotide metabolism.

Given the dependence of O-GlcNAc on the cell's metabolic pathways, O-GlcNAc has been shown to be a key glucose sensor that is exquisitely sensitive to glucose and metabolite levels.[111,163-165] In addition, O-GlcNAc regulates glucose uptake, insulin signaling, glycogen synthesis, gluconeogenesis, lipogenesis, and mitochondrial function.[42,166] OGT has been shown to be recruited to the cellular membrane through direct interaction with PIP_3 downstream of insulin-induced PI3K signaling. Once localized to the membrane, OGT glycosylates the key insulin signaling regulators, IR-β (insulin receptor β), IRS1 (insulin receptor substrate 1), PDK1 (3-phosphoinositide-dependent protein kinase 1), PI3K, Akt (RAC-α serine/threonine-protein kinase), which changes their phosphorylation states and contributes to insulin resistance.[28] Interestingly, insulin signaling promotes the expression of OGT and the targeting of OGT to lipid rafts.[167]

The connection between O-GlcNAc and insulin signaling has been demonstrated to be present in and contribute to the chronic disease, diabetes.[162,168] Aberrant *O*-GlcNAc glycosylation has been observed in diabetic patients and diabetic mouse models.[81,108] In fact, several studies have proposed and demonstrated that *O*-GlcNAc has utility as a robust diagnostic biomarker for diabetic and, potentially more importantly, pre-diabetic metabolism.[169,170] Reduction of OGA in mice have been shown to cause obesity and insulin resistance implicating OGA in diabetic metabolic and transcriptional changes.[171] *O*-GlcNAcylation of CREB-regulated transcriptional coactivator 2 (CRTC2) reduces its cytoplasmic sequestration, increases its nuclear translocation, and enhances hepatic gluconeogenesis through elevated CRTC2-mediated transcription.[172] In addition to CRTC2, many metabolism-regulating TFs are glycosylated or regulated by *O*-GlcNAc, which affects their transcriptional activity including carbohydrate responsive element binding protein (ChREBP), nuclear receptor liver X receptor (LXR), sterol response element binding protein 1 (SREBP1), CRTC2, PGC-1α, and Sp1. Through these TFs, *O*-GlcNAc directly regulates insulin signaling, glycogen synthesis, gluconeogenesis, and lipogenesis in response to nutrient levels.[173-176]

O-GlcNAc has also been shown to regulate nutrient flux and mitochondrial activity. Mitochondria are the powerhouses of the cells, responsible for maximizing the ATP required for cellular activity while minimizing ROS (reactive oxygen species) generation. OGA deficiency modulates insulin and glucose sensitivity through the modulation of mitochondrial homeostasis.[177,178] Interestingly, GlcN (glucosamine), a precursor to UDP-GlcNAc (the product of GFAT), facilitated mitochondrial function, which led to lifespan extension in both *C. elegans* and mice.[179] This result led to the

hypothesis that increasing O-GlcNAc levels may be beneficial for mitochondrial function. However, increasing O-GlcNAc levels (TMG treatment) led to mitochondrial dysfunction and was associated with lower glycolytic flux and ROS generation while decreasing ncOGT and mOGT levels led to higher glycolysis and ROS generation in mitochondria.[180,181] The conflicting GlcN and TMG results can be explained by the ability for GlcN to affect multiple different metabolic pathways while TMG selectively and specifically inhibits OGA. OGT directly O-GlcNAcylates a variety of mitochondrial proteins involved in oxidative phosphorylation.[182] Interestingly, the mitochondrial OGT (mOGT) isoform has been shown to trigger apoptosis and potentially contribute to Alzheimer's disease (AD) pathology in diabetic patients.[23,183] However, the role of O-GlcNAc in mitochondrial function has been left unclear by two studies that showed (1) that ncOGT is sufficient for mitochondrial O-GlcNAcylation leaving the exact role of mOGT unclear and (2) high glucose can change mitochondrial function independent of O-GlcNAc.[184,185] While many studies suggest that O-GlcNAc tracks with and may regulate metabolic functions, establishment of either a causative or correlative link between O-GlcNAc and these functions require further study.

Cancer exploits perturbations in cell cycle, growth, and metabolism to circumvent constraints on growth and proliferation. Cancer has been shown to display altered metabolism including increased glucose import, aerobic glycolysis (a.k.a. the Warburg effect), glutamine addiction, reduced oxidative phosphorylation, and hyper-O-GlcNAcylation.[186] Increased O-GlcNAc, OGT, and OGA levels have been observed across many different cancer subtypes from including breast, lung, colon, and prostate cancer, suggesting that O-GlcNAc could be a useful malignancy diagnostic biomarker.[187]

Indeed, several studies have shown that *O*-GlcNAc plays a critical role in cancer cell biology, leading to many reviews on the link between *O*-GlcNAc and cancer.[186,188-190] As discussed in the section above, many TFs, transcriptional regulators, and epigenetic regulators that are known to regulate cancer biology are *O*-GlcNAcylated, which alter their oncogenic properties including p53, c-Myc, Bmi-1, EZH2, HIF-1, HCF-1, HDAC, mSin3a, TET, BAP1, Sp1, Rb, and FOXM1.[110,191-193]

O-GlcNAc achieves metabolic reprogramming of cancer cells through transcriptional and metabolic changes. The *O*-GlcNAc modification of two glycolytic pathway proteins, PFK1 (phosphofructokinase 1) and G6PD (glucose-6-phosphate dehydrogenase), reduces glycolytic flux, and therefore shunts carbon flux through the PPP.[194,195] Cancer cells rely more heavily on the PPP than glycolysis because the PPP reduces oxidative stress and capacitates rapid growth required for cellular proliferation.[196] Many studies have shown that *O*-GlcNAcylation and HBP flux are perturbed in and important for the cancer epithelial to mesenchymal transition (EMT), the initiative process involved in cancer metastasis.[197-199] Cancer metastasis, invasiveness, and angiogenesis are heavily influenced by *O*-GlcNAcylation.[191,192,200,201] Tumorigenesis, the promotion of tumor growth in cancers, is also enhanced by increasing *O*-GlcNAcylation.[202] Recently, researchers have also shown that cancer stem cell proliferation is also regulated by *O*-GlcNAcylation.[203] Cancer stem cells are the population of cells that with self-renewal and differentiation properties that are believed to be key to tumor initiation, progression, and metastasis.[204] In summary, *O*-GlcNAc has been demonstrated to be a key regulator of cancer biology and could be an important biomarker and therapeutic target.

1.8 Role of *O*-GlcNAc in neuronal function

We will now focus on the role of *O*-GlcNAcylation in neurobiology. Studies have shown that OGT and OGA are expressed at their highest levels in the hippocampus, localized to synapses, and modify thousands of proteins in neurons.[205] In fact, an estimated 40% of the neuronal and 19% of the synaptosomal proteomes appear to be *O*-GlcNAcylated with some proteins like bassoon modified with over a dozen *O*-GlcNAcylation sites.[150] In excitatory synapses, OGT is localized to the post-synaptic density where *O*-GlcNAcylation increases upon neuronal stimulation.[206] The addition of the OGA inhibitor, PUGNAc, enhances long term potentiation (LTP), short term potentiation (STP), and axonal branching while an inhibitor for OGT, alloxan, produces the opposite effects (Figure 1.9).[207,208] LTP is a mechanism that underlies learning and memory and becomes increasingly impaired in neurodegenerative diseases.[209] Increasing *in vivo* *O*-GlcNAc levels in synapses enhances the phosphorylation of synapsin I/II, which increases hippocampal synaptic plasticity and activity.[208] OGT KO leads to decreased expression of glutamate AMPA receptors (GluA2/3), reduction of the number of dendritic spines, and an increase in the proportion of immature dendritic spines.[206] On the other hand, another study in excitatory glutamatergic CA3-CA1 hippocampal neurons showed that increasing *O*-GlcNAcylation led to long term depression (LTD), disrupted LTP, and impaired memory formation.[210] Consistent with this observation, increasing *O*-GlcNAc levels was shown to be ameliorate the neuronal hyperexcitability in hippocampal slices and in mice, suggesting a neuroprotective role for *O*-GlcNAc.[211] OGA deficient mice (*Oga*$^{+/-}$) displayed elevated *O*-GlcNAc levels and LTP and LTD impairments through aberrant glutamate receptor phosphorylation.[212] One study showed that

pharmacologically increasing O-GlcNAc levels had no effect on basal synaptic transmission while a separate study showed that this facilitated basal synaptic transmission through increased trafficking of AMPA receptors.[208]

The inconsistency of these studies about the role of O-GlcNAc on neuronal function can be attributed to pharmacological inhibition vs. genetic ablation methods. In particular, some of the pharmacological agents used to inhibit the O-GlcNAc cycling enzymes were found to have off-target effects confounding potential results. Despite their contradictory results, these studies support an essential role of OGT, OGA, and O-GlcNAcylation in neuronal function and that the maintenance of homeostatic O-GlcNAc levels is necessary for proper neuronal homeostasis.
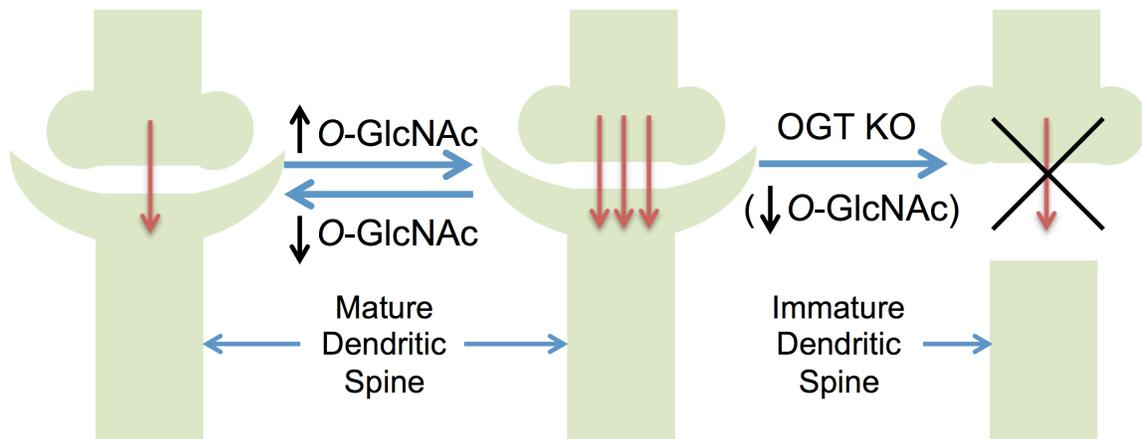
Figure 1.9 O-GlcNAc regulates neuronal activity. Increasing O-GlcNAc levels with an OGA inhibitor enhances LTP, STP, and axonal branching while decreasing O-GlcNAc levels with an OGT inhibitor has the reverse effects. Knocking out OGT results in fewer dendritic spines and a higher proportion of mature dendritic spines thereby interfering with neuronal signaling.

Through the examination of the protein- and site-specific effects of O-GlcNAcylation on neuronal proteins, the mechanism by which O-GlcNAc influences neuronal function has been further elucidated. Ablation of the O-GlcNAcylation site of synapsin I at T87 increases synapsin I synaptic localization, the synaptic vesicle localization to axons, as well as the reserve pool of synaptic vesicles, thereby enhancing

synaptic activity.[213] Similarly, *O*-GlcNAcylation of AMPA receptor GluA2 results in LTD and affects GluA2 trafficking to synapses.[209,210] Taken with the studies that globally change *O*-GlcNAc levels, this site-specific data supports a complex and essential role of *O*-GlcNAcylation in mediating neuronal activity and underscores the importance of pursuing and integrating systems level and reductionist approaches in parallel.

       O-GlcNAc plays a critical role in organismal metabolism through regulation of neuronal activity. In AgRP neurons in the arcuate nucleus of the hypothalamus, *O*-GlcNAc modification of the potassium channel, $K_v7.3$ (*Kcnq3*), increases neuronal excitability, which prevents the conversion of white to brown fat. OGT KO in AgRP neurons resulted in decreased hepatic gluconeogenesis, enhanced thermogenesis, and protection from obesity and insulin resistance.[214] Ablation of OGT in another region of the hypothalamus, the paraventricular nucleus, causes mice to overeat and become obese, demonstrating that specific ablation of OGT in different neurons of the same brain region can result in polar opposite metabolic effects.[215] In these ways, *O*-GlcNAcylation in specific neurons can affect an entire organism's metabolism.

       In addition, *O*-GlcNAc plays a vital role in cellular metabolism within neurons. The arborized nature of neurons necessitates proper trafficking and distribution of mitochondria to meet the localized energy demands of dendrites and axons.[216] The mitochondrial motor-adaptor protein, TRAK1 (trafficking kinesin binding protein 1), has been shown to stably interact with OGT, which can alter OGT's substrate specificity.[217] In conditions with high extracellular glucose, *O*-GlcNAcylation of TRAK1/Milton increases, which reduces the motility of migrating mitochondria in axons.[178] Through this mechanism, neurons can rapidly respond to the nutrient milieu. Our lab showed that a

33

nutrient sensor and important transcriptional regulator of neuronal activity, CREB (cAMP response element binding protein), is O-GlcNAc modified at S40 in our lab. This glycosylation event disrupts the association of CREB with the coactivator, CRTC, and thereby suppresses the transcription of neuronal activity genes in response to neuronal depolarization. Ablation of the S40 glycosylation site of CREB was shown to hasten the formation of memory in a fear conditioning paradigm.[218] The role of the O-GlcNAcylation of CREB will be explored in significantly greater depth in Chapters 4 and 5. Thus, O-GlcNAc regulates neuronal activity through its regulation of the neuronal activity gene transcription and neuronal mitochondrial trafficking in response to neuronal stimulation and glucose increase respectively.

Beyond the importance of O-GlcNAc in normal neuronal function, OGT, OGA, and O-GlcNAc dysregulation have been implicated in many neurological and psychiatric disorders such as schizophrenia, Adams-Oliver syndrome, X-linked intellectual disability, and stroke.[18,219-221] Given the link between stress and psychiatric and neurological diseases, future studies may reveal more connections between the stress-induced O-GlcNAcylation and these diseases. In the next section, we will explore the relationship between O-GlcNAcylation and neurodegenerative diseases.

1.9 Role of O-GlcNAc in neurodegenerative diseases

O-GlcNAcylation has also been shown to play a pivotal role in neurodegenerative diseases. Many neurodegenerative diseases are proteinopathies where particular proteins aberrantly aggregate due to hyperphosphorylation and/or differential protease-mediated cleavage. Specifically, these neurodegenerative diseases can be classified into amyloidoses (AD), tauopathies (AD and frontotemporal dementia (FTD)),

synucleinopathies (Parkinson's disease (PD)), TDP-43 (transactivation response-DNA binding protein-43) proteinopathies (amyotrophic lateral sclerosis (ALS) and frontotemporal lobar degeneration with tau-negative and ubiquitin-positive inclusions (FTLD-U)), and polyglutamine (polyQ) proteinopathies (Huntington's disease (HD) and spinocerebellar ataxia (SCA)).[222,223] Through *O*-GlcNAc's fundamental role in regulating autophagy, the major mechanism for removal of protein aggregates, *O*-GlcNAc affects the cellular response to these proteinopathies. In addition, *O*-GlcNAc has been implicated in specific neurodegenerative diseases based on genome-wide association studies (GWAS). The genomic locus for the *Oga* is associated with increased risk for late-onset Alzheimer's disease (AD).[224] In addition, *Oga* is subject to differential alternative splicing in AD.[225] The *Ogt* locus has been associated with X-linked mental retardation and a potential linkage to a rare form of dystonia-parkinsonism although a recent study suggested that the disease might arise from another resident gene on the locus.[226-228]

O-GlcNAc has been found on all of these aggregating proteins and has been shown to be important for increasing the solubility of these proteins.[229] Indeed, several studies have established a yin-yang relationship between *O*-GlcNAc and disease-associated hyperphosphorylation for many different proteinopathy-associated proteins including α-synuclein, neurofilaments, and TDP-43.[64] This led to the general hypothesis that hyperphosphorylation leads to protein aggregation while *O*-GlcNAcylation prevents aggregation and confers neuroprotection (Figure 1.9). In addition, *O*-GlcNAc modulates the protease-mediated cleavage of APP (amyloid precursor protein), α-synuclein, and TDP-43, which generates amyloidogenic species.[230] We will now discuss what is known

about the effect of *O*-GlcNAc on the etiology of neurodegenerative diseases and their biological mechanisms.
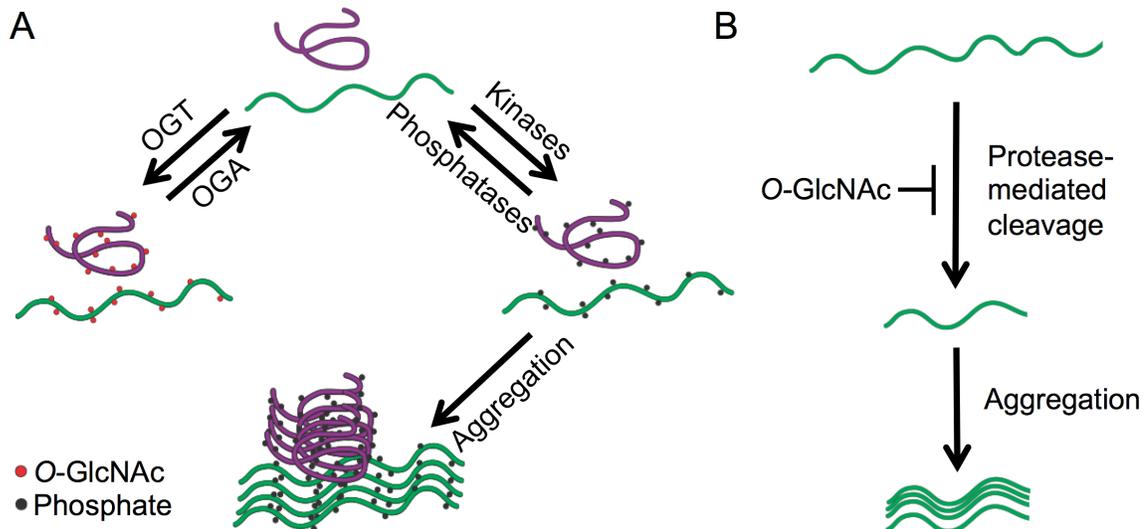
Figure 1.9 *O*-GlcNAc reduces protein aggregation in proteinopathies. (A) *O*-GlcNAcylation of proteins increases the solubility of proteins while hyperphosphorylation leads to aggregation and the formation of disease-related plaques, tangles, of inclusions. (B) *O*-GlcNAc also mediates the proteolytic cleavage of proteins in proteinopathies. Proteolytic cleavage of aggregating proteins often enhances their amyloidogenic capabilities and *O*-GlcNAcylation of these proteins can reduce their cleavage.

Alzheimer's disease (AD) is a debilitating disease that is characterized by progressive memory loss and is currently the largest unmet neurological need with no treatment to prevent, slow, or cure the disease.[231] *O*-GlcNAc levels and specific *O*-GlcNAcylation events are perturbed in AD patients.[232,233] The major hallmarks of AD are amyloid plaques and neurofibrillary tangles (NFTs), which are composed of β-amyloid peptide and hyperphosphorylated microtubule-associated protein tau respectively. Both disease-related proteins are known to be *O*-GlcNAcylated *in vivo*, and their *O*-GlcNAcylation is correlated with a decrease in protein aggregation.[234,235] The main familial Alzheimer's disease (FAD) mutations that lead to early-onset AD occur in proteins involved in β-amyloid (Aβ) processing, including mutations in APP, presenilin 1 (PS1), and presenilin 2 (PS2). Amyloid precursor protein is a transmembrane protein that

is subject to proteolytic processing by sheddases. APP can be cleaved either by α-secretase to generate non-aggregating species or β-secretase to generate the βCTF. After cleavage by β-secretase, γ-secretase cleaves βCTF to create Aβ40 and Aβ42, the more amyloidogenic species.[236] The *O*-GlcNAcylation of APP increases α-secretase non-amyloidogenic processing and decreases Aβ40 levels.[237] In addition, *O*-GlcNAc modification of APP at threonine 576 increases the trafficking rate to the membrane, decreases its endocytosis, and decreases Aβ peptide production.[238] Increasing *O*-GlcNAc levels using an OGA inhibitor, NButGT, reduced β-amyloid plaque formation, neuroinflammation, and memory impairment in a mutant mice that expresses FAD mutations in APP and PS1. The underlying mechanism of neuroprotection was via the direct *O*-GlcNAcylation of S708 of nicastrin, a component of γ-secretase, leading to a reduction in γ-secretase activity and Aβ generation.[239] These results led to the theory that *O*-GlcNAc modification plays a neuroprotective role in AD and generally in neurodegenerative diseases.

In addition to AD, FTD, FTD with parkinsonism (FTDP), and progressive supranuclear palsy (PSP) are tauopathies that are characterized by the hyperphosphorylation and then aberrant aggregation of microtubule-associated protein tau to form neurofibrillary tangles (NFTs).[240] OGT is known to modify and regulate glycogen synthase kinase-3b (GSK-3b) and Cdk5, kinases that abnormally hyperphosphorylates APP and tau.[151,241-243] In a mouse model of FTDP expressing amyloidogenic tau[P301L], the Vocadlo lab evinced that increasing *O*-GlcNAc levels with TMG raised global protein *O*-GlcNAcylation levels and reduced tau aggregation.[244] As the *O*-GlcNAcylation of tau at S400 increased, the solubility of tau increased, although

the hyperphosphorylation of tau was not affected.[245-247] This suggests that O-GlcNAcylation alone (without removal of hyperphosphorylation) may be sufficient for reducing the fibrillar tau. Another study showed that O-GlcNAc-S400 tau enhanced the solubility of tau relative to WT tau *in vitro*, although this was not compared to the purely phosphorylated species or pseudo-phosphorylated species where phosphorylation sites are replaced with phospho-mimetic residues.[248] Furthermore, increased O-GlcNAc levels led to the elevation of O-GlcNAc-S400 tau, reduction of tauopathy, decreased cerebrospinal fluid tau, and amelioration of the neurodegenerative phenotype in tau^{P301L} mice.[244,249] In another AD mouse model that expresses tau^{P301L} and APP$^{K670N/M671L}$ (APP^{Swe}, Tg2576), treatment with TMG reduces the amount of β-amyloid peptides and plaques and succors cognitive decline.[250] Across mouse models of neurodegeneration and especially of amyloidoses and tauopathies, studies have consistently shown that increasing O-GlcNAc levels ameliorates neurodegeneration supporting a neuroprotective role for O-GlcNAc.[249,251,252] In fact, several companies, one of which grew out of the Vocadlo group, are developing and exploring the utility and efficacy of O-GlcNAcase inhibitors as therapeutics for neurodegenerative diseases. A Swiss company, Asceneuron, now has a drug entering Phase I clinical trials for the treatment of the tauopathy PSP as of this year.[253]

In the synucleinopathy PD, α-synuclein aggregates to form Lewy bodies in dopaminergic neurons in substantia nigra pars compacta (SNpc), which causes neuronal death and progressive motor dysfunction.[254] Elevated O-GlcNAc levels were observed in the temporal cortex of PD patients. Enhancing O-GlcNAc levels through TMG inhibition of OGA was then shown to reduce autophagy activity, leading to α-synuclein aggregation

in PD, which suggests that globally increasing *O*-GlcNAc may exacerbate PD.[255] Knocking out the second most common causal genetic PD mutation (*LRRK2*) and its homolog *LRRK1* in mice reduced autophagic flux, recapitulated the specific SNpc dopaminergic neurodegeneration observed in PD, and intensified α-synuclein accumulation, suggesting that impaired autophagic flux could be a causal mechanism in PD generally or at least specifically in *LRRK2* mutation-induced PD.[256] *O*-GlcNAcylation of T72 and S87 on α-synuclein prevents α-synuclein hyperphosphorylation and aggregation and associated neurodegeneration.[257-259] In PD, α-synuclein is proteolytically cleaved by calpain to generate fragments found in Lewy bodies. α-synuclein *O*-GlcNAcylation on either T72 or S87 conferred resistance to the calpain-induced cleavage of α-synuclein.[230] Interestingly, increasing *O*-GlcNAc levels through GlcN or PUGNAc diminishes calpain activity and is cardioprotective in rats.[260] Similar to AD, the role of *O*-GlcNAc appears to be neuroprotective for PD through prevention of with aberrant hyperphosphorylation and proteolytic cleavage, which inhibits protein aggregation. In addition, up to half of patients with Parkinson's disease also have dementia with the major hallmarks of Alzheimer's disease, amyloid-β plaques and neurofibrillary tangles, suggesting a link between PD and AD.[261]

Amyotrophic lateral sclerosis (ALS) is another neurodegenerative disease that causes the progressive degeneration of motor neurons leading to paralysis.[262] ALS is characterized by the aggregation of TDP-43, NFs, FUS (FUS RNA-binding protein), and SOD1 (Cu/Zn superoxide dismutase 1) to form inclusion bodies. TDP-43, NF-H, NF-M, NF-L, FUS, and SOD1 are all known to be *O*-GlcNAcylated and at least some of their sites have been mapped.[150,263-266] Like with tau and α-synuclein in AD and PD

respectively, hyperphosphorylation of neurofilaments and TDP-43 occurs with aberrant proteolytic cleavage of TDP-43. TDP-43's *O*-GlcNAc sites overlap with known CK1 phosphorylation sites, suggesting that a yin-yang relationship may exist for *O*-GlcNAcylation and phosphorylation on TDP-43.[263] *O*-GlcNAc levels are decreased in the spinal cords of G93A-SOD1 mutant mice suggesting that *O*-GlcNAcylation is perturbed in ALS.[267] More research is required to explore the connection between ALS and *O*-GlcNAcylation.

Although *O*-GlcNAc modification has been shown to be important in neurodegenerative diseases, a causal link between *O*-GlcNAcylation and neurodegenerative diseases has not been established. This leads to the outstanding question: does *O*-GlcNAc dysregulation precede and produce neurodegeneration or is aberrant *O*-GlcNAcylation merely a side effect of neuronal dysfunction? What are the initial transcriptional or morphological changes that occur upon *O*-GlcNAc depletion, and does this depletion ultimately lead to a disease phenotype? In humans, these questions remain largely inaccessible, as diagnosis for AD, PD, and other idiopathic neurodegenerative diseases are typically made after neurodegeneration. Studies in different model organisms have been contradictory; studies in *C. elegans* show that increasing *O*-GlcNAc levels aggravates protein aggregation and neurodegeneration while studies in mice show that increasing *O*-GlcNAc produces the neuroprotective effects.[229,268,269] More incompatible evidence can be found in studies exploring the levels of *O*-GlcNAc in human samples; different studies have shown increased and decreased *O*-GlcNAc levels even in the same exact disease and brain region. For example, one study found increased levels of *O*-GlcNAc and decreased OGA protein levels in the

cerebral cortices of AD patients while another study found decreased levels of *O-*GlcNAc.[233,252] In light of these apparent antithetical studies, the question remains: is *O-*GlcNAc neuroprotective or neurodegenerative? Additionally, what is the role of *O-*GlcNAc in neuronal function and homeostasis? In order to answer these major questions, our lab has characterized a forebrain-specific OGT KO mouse, which will be the subject of Chapter 2.

1.10 References

1   Crick, F. Central Dogma of Molecular Biology. *Nature* (1970) **227**, 561-563.
2   Wymann, M.P. & Schneiter, R. Lipid signalling in disease. *Nat. Rev. Mol. Cell Biol.* (2008) **9**, 2:162-176.
3   Sasisekharan, R., Shriver, Z., Venkataraman, G. & Narayanasami, U. Roles of heparan-sulphate glycosaminoglycans in cancer. *Nat. Rev. Cancer* (2002) **2**, 7:521-528.
4   Shriver, Z., Raguram, S. & Sasisekharan, R. Glycomics: a pathway to a class of new and improved therapeutics. *Nat. Rev. Drug Discov.* (2004) **3**, 10:863-873.
5   Cocinero, E.J. & Çarçabal, P. in *Gas-Phase IR Spectroscopy and Structure of Biological Molecules* (eds Anouk M. Rijs & Jos Oomens) 299-333 (Springer International Publishing, 2015).
6   Gabius, H.J., Andre, S., Jimenez-Barbero, J., Romero, A. & Solis, D. From lectin structure to functional glycomics: principles of the sugar code. *Trends Biochem. Sci.* (2011) **36**, 6:298-313.
7   Marino, K., Bones, J., Kattla, J.J. & Rudd, P.M. A systematic approach to protein glycosylation analysis: a path through the maze. *Nat Chem Biol* (2010) **6**, 10:713-723.
8   Yang, X. & Qian, K. Protein O-GlcNAcylation: emerging mechanisms and functions. *Nat. Rev. Mol. Cell Biol.* (2017) **18**, 7:452-465.
9   Guo, B. *et al.* O-GlcNAc-modification of SNAP-29 regulates autophagosome maturation. *Nat. Cell Biol.* (2014) **16**, 12:1215-1226.
10   Ruan, H.B., Nie, Y. & Yang, X. Regulation of protein degradation by O-GlcNAcylation: crosstalk with ubiquitination. *Mol. Cell. Proteomics* (2013) **12**, 12:3489-3497.
11   Bond, M.R. & Hanover, J.A. A little sugar goes a long way: the cell biology of O-GlcNAc. *J. Cell Biol.* (2015) **208**, 7:869-880.
12   Hart, G.W., Slawson, C., Ramirez-Correa, G. & Lagerlof, O. Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease. *Annu. Rev. Biochem.* (2011) **80**, 1:825-858.
13   Selvan, N. *et al.* The Early Metazoan Trichoplax adhaerens Possesses a Functional O-GlcNAc System. *J. Biol. Chem.* (2015) **290**, 19:11969-11982.
14   Love, D.C. & Hanover, J.A. The hexosamine signaling pathway: deciphering the "O-GlcNAc code". *Sci. STKE* (2005) **2005**, 312:re13.
15   Xu, S.L. *et al.* Proteomic analysis reveals O-GlcNAc modification on proteins with key regulatory functions in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* (2017) **114**, 8:E1536-E1543.
16   Ogawa, M., Sawaguchi, S., Kamemura, K. & Okajima, T. Intracellular and extracellular O-linked N-acetylglucosamine in the nervous system. *Exp. Neurol.* (2015) **274**, Pt B:166-174.
17   Sakaidani, Y. *et al.* O-linked-N-acetylglucosamine on extracellular protein domains mediates epithelial cell-matrix interactions. *Nat Commun* (2011) **2**, 583.
18   Ogawa, M. *et al.* Impaired O-linked N-acetylglucosaminylation in the endoplasmic reticulum by mutated epidermal growth factor (EGF) domain-specific O-linked N-acetylglucosamine transferase found in Adams-Oliver syndrome. *J Biol Chem* (2015) **290**, 4:2137-2149.

19   Sawaguchi, S. *et al.* O-GlcNAc on NOTCH1 EGF repeats regulates ligand-induced Notch signaling and vascular development in mammals. *Elife* (2017) **6**, e24419.

20   Torres, C.R. & Hart, G.W. Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc. *J. Biol. Chem.* (1984) **259**, 5:3308-3317.

21   Hart, G.W., Housley, M.P. & Slawson, C. Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature* (2007) **446**, 7139:1017-1022.

22   Wang, Z. *et al.* Enrichment and site mapping of O-linked N-acetylglucosamine by a combination of chemical/enzymatic tagging, photochemical cleavage, and electron transfer dissociation mass spectrometry. *Mol. Cell. Proteomics* (2010) **9**, 1:153-160.

23   Shin, S.H., Love, D.C. & Hanover, J.A. Elevated O-GlcNAc-dependent signaling through inducible mOGT expression selectively triggers apoptosis. *Amino Acids* (2011) **40**, 3:885-893.

24   Janetzko, J. & Walker, S. The making of a sweet modification: structure and function of O-GlcNAc transferase. *J. Biol. Chem.* (2014) **289**, 50:34424-34432.

25   Park, S.K. *et al.* A conserved splicing silencer dynamically regulates *O*-GlcNAc transferase intron retention and *O*-GlcNAc homeostasis. *Cell Rep* (2017) **20**, 5:1088-1099.

26   Lazarus, M.B., Nam, Y., Jiang, J., Sliz, P. & Walker, S. Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature* (2011) **469**, 7331:564-567.

27   Kreppel, L.K., Blomberg, M.A. & Hart, G.W. Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and characterization of a unique O-GlcNAc transferase with multiple tetratricopeptide repeats. *J. Biol. Chem.* (1997) **272**, 14:9308-9315.

28   Yang, X. *et al.* Phosphoinositide signalling links O-GlcNAc transferase to insulin resistance. *Nature* (2008) **451**, 7181:964-969.

29   Jinek, M. *et al.* The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin alpha. *Nat. Struct. Mol. Biol.* (2004) **11**, 10:1001-1007.

30   Rafie, K. *et al.* Recognition of a glycosylation substrate by the O-GlcNAc transferase TPR repeats. *Open Biol* (2017) **7**, 6.

31   Shen, D.L. *et al.* Catalytic Promiscuity of O-GlcNAc Transferase Enables Unexpected Metabolic Engineering of Cytoplasmic Proteins with 2-Azido-2-deoxy-glucose. *ACS Chem Biol* (2017) **12**, 1:206-213.

32   Lazarus, M.B. *et al.* HCF-1 is cleaved in the active site of O-GlcNAc transferase. *Science* (2013) **342**, 6163:1235-1239.

33   Janetzko, J., Trauger, S.A., Lazarus, M.B. & Walker, S. How the glycosyltransferase OGT catalyzes amide bond cleavage. (2016) **12**, 899.

34   Kotzler, M.P. & Withers, S.G. Proteolytic Cleavage Driven by Glycosylation. *J Biol Chem* (2016) **291**, 1:429-434.

35   Gao, Y., Wells, L., Comer, F.I., Parker, G.J. & Hart, G.W. Dynamic O-glycosylation of nuclear and cytosolic proteins: cloning and characterization of a neutral, cytosolic beta-N-acetylglucosaminidase from human brain. *J. Biol. Chem.* (2001) **276**, 13:9838-9845.

36   Li, B., Li, H., Lu, L. & Jiang, J. Structures of human O-GlcNAcase and its complexes reveal a new substrate recognition mode. *Nat. Struct. Mol. Biol.* (2017) **24**, 4:362-369.

37   Butkinaree, C. *et al.* Characterization of beta-N-acetylglucosaminidase cleavage by caspase-3 during apoptosis. *J Biol Chem* (2008) **283**, 35:23557-23566.

38   Elsen, N.L. *et al.* Insights into activity and inhibition from the crystal structure of human O-GlcNAcase. *Nat. Chem. Biol.* (2017) **13**, 6:613-615.

39   Roth, C. *et al.* Structural and functional insight into human O-GlcNAcase. *Nat. Chem. Biol.* (2017) **13**, 6:610-612.

40   Hanover, J.A. Glycan-dependent signaling: O-linked N-acetylglucosamine. *FASEB J.* (2001) **15**, 11:1865-1876.

41   Keembiyehetty, C.N., Krzeslak, A., Love, D.C. & Hanover, J.A. A lipid-droplet-targeted O-GlcNAcase isoform is a key regulator of the proteasome. *Journal of Cell Science* (2011) **124**, 16:2851.

42   Ruan, H.B., Singh, J.P., Li, M.D., Wu, J. & Yang, X. Cracking the O-GlcNAc code in metabolism. *Trends Endocrinol. Metab.* (2013) **24**, 6:301-309.

43   Whisenhunt, T.R. *et al.* Disrupting the enzyme complex regulating O-GlcNAcylation blocks signaling and development. *Glycobiology* (2006) **16**, 6:551-563.

44   Groussaud, D. *et al.* Hijacking of the O-GlcNAcZYME complex by the HTLV-1 Tax oncoprotein facilitates viral transcription. *PLoS Pathog.* (2017) **13**, 7:e1006518.

45   FirstGlance in Jmol: a simple tool for macromolecular visualization, available on-line at firstglance.jmol.org v. 2.5.1 (free software, Creative Commons License, 2006).

46   Pathak, S. *et al.* The active site of O-GlcNAc transferase imposes constraints on substrate sequence. *Nat. Struct. Mol. Biol.* (2015) **22**, 9:744-750.

47   Kao, H.J. *et al.* A two-layered machine learning method to identify protein O-GlcNAcylation sites with O-GlcNAc transferase substrate motifs. *BMC Bioinformatics* (2015) **16 Suppl 18**, 18:S10.

48   Wang, J., Torii, M., Liu, H., Hart, G.W. & Hu, Z.Z. dbOGAP - an integrated bioinformatics resource for protein O-GlcNAcylation. *BMC Bioinformatics* (2011) **12**, 1:91.

49   Gupta, R. & Brunak, S. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac Symp Biocomput.* (2002) 310-322.

50   Jia, C.Z., Liu, T. & Wang, Z.P. O-GlcNAcPRED: a sensitive predictor to capture protein O-GlcNAcylation sites. *Mol. Biosyst.* (2013) **9**, 11:2909-2913.

51   Britto-Borges, T. & Barton, G.J. A study of the structural properties of sites modified by the O-linked 6-N-acetylglucosamine transferase. *PLoS One* (2017) **12**, 9:e0184405.

52   Zhu, Y. *et al.* O-GlcNAc occurs cotranslationally to stabilize nascent polypeptide chains. *Nat. Chem. Biol.* (2015) **11**, 5:319-325.

53   Chen, Q., Chen, Y., Bian, C., Fujiki, R. & Yu, X. TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature* (2013) **493**, 7433:561-564.

54   Yang, X., Zhang, F. & Kudlow, J.E. Recruitment of O-GlcNAc transferase to promoters by corepressor mSin3A: coupling protein O-GlcNAcylation to transcriptional repression. *Cell* (2002) **110**, 1:69-80.

55   Cheung, W.D., Sakabe, K., Housley, M.P., Dias, W.B. & Hart, G.W. O-linked beta-N-acetylglucosaminyltransferase substrate specificity is regulated by myosin phosphatase targeting and other interacting proteins. *J. Biol. Chem.* (2008) **283**, 49:33935-33941.

56   Ruan, H.B. *et al.* O-GlcNAc transferase/host cell factor C1 complex regulates gluconeogenesis by modulating PGC-1alpha stability. *Cell Metab.* (2012) **16**, 2:226-237.

57   Deng, R.P. *et al.* Global identification of O-GlcNAc transferase (OGT) interactors by a human proteome microarray and the construction of an OGT interactome. *Proteomics* (2014) **14**, 9:1020-1030.

58   Khidekel, N., Ficarro, S.B., Peters, E.C. & Hsieh-Wilson, L.C. Exploring the O-GlcNAc proteome: direct identification of O-GlcNAc-modified proteins from the brain. *Proc. Natl. Acad. Sci. U. S. A.* (2004) **101**, 36:13132-13137.

59   Vosseller, K. *et al.* O-linked N-acetylglucosamine proteomics of postsynaptic density preparations using lectin weak affinity chromatography and mass spectrometry. *Mol. Cell. Proteomics* (2006) **5**, 5:923-934.

60   Greis, K.D. *et al.* Selective detection and site-analysis of O-GlcNAc-modified glycopeptides by beta-elimination and tandem electrospray mass spectrometry. *Anal Biochem* (1996) **234**, 1:38-49.

61   Trinidad, J.C., Schoepfer, R., Burlingame, A.L. & Medzihradszky, K.F. N- and O-glycosylation in the murine synaptosome. *Mol. Cell. Proteomics* (2013) **12**, 12:3474-3488.

62   Hahne, H. *et al.* Proteome wide purification and identification of O-GlcNAc-modified proteins using click chemistry and mass spectrometry. *J. Proteome Res.* (2013) **12**, 2:927-936.

63   Vosseller, K. *et al.* Quantitative analysis of both protein expression and serine / threonine post-translational modifications through stable isotope labeling with dithiothreitol. *Proteomics* (2005) **5**, 2:388-398.

64   Kang, M.J. *et al.* Synapsin-1 and tau reciprocal O-GlcNAcylation and phosphorylation sites in mouse brain synaptosomes. *Exp. Mol. Med.* (2013) **45**, e29:e29.

65   Alfaro, J.F. *et al.* Tandem mass spectrometry identifies many mouse brain O-GlcNAcylated proteins including EGF domain-specific O-GlcNAc transferase targets. *Proc. Natl. Acad. Sci. U. S. A.* (2012) **109**, 19:7280-7285.

66   Zaro, B.W., Yang, Y.Y., Hang, H.C. & Pratt, M.R. Chemical reporters for fluorescent detection and identification of O-GlcNAc-modified proteins reveal glycosylation of the ubiquitin ligase NEDD4-1. *Proc. Natl. Acad. Sci. U. S. A.* (2011) **108**, 20:8146-8151.

67   Woo, C.M., Iavarone, A.T., Spiciarich, D.R., Palaniappan, K.K. & Bertozzi, C.R. Isotope-targeted glycoproteomics (IsoTaG): a mass-independent platform for intact N- and O-glycopeptide discovery and analysis. *Nat. Methods* (2015) **12**, 6:561-567.

68   Selvan, N. *et al.* A mutant O-GlcNAcase enriches Drosophila developmental regulators. *Nat Chem Biol* (2017) **13**, 8:882-887.

69   Leney, A.C., Rafie, K., van Aalten, D.M.F. & Heck, A.J.R. Direct Monitoring of Protein O-GlcNAcylation by High-Resolution Native Mass Spectrometry. *ACS Chem Biol* (2017) **12**, 8:2078-2084.

70   Rexach, J.E., Clark, P.M. & Hsieh-Wilson, L.C. Chemical approaches to understanding O-GlcNAc glycosylation in the brain. *Nat. Chem. Biol.* (2008) **4**, 2:97-106.

71   Banerjee, P.S., Hart, G.W. & Cho, J.W. Chemical approaches to study O-GlcNAcylation. *Chem. Soc. Rev.* (2013) **42**, 10:4345-4357.

72   Vocadlo, D.J., Hang, H.C., Kim, E.J., Hanover, J.A. & Bertozzi, C.R. A chemical approach for identifying O-GlcNAc-modified proteins in cells. *Proc Natl Acad Sci U S A* (2003) **100**, 16:9116-9121.

73   Rashidian, M., Dozier, J.K. & Distefano, M.D. Enzymatic labeling of proteins: techniques and approaches. *Bioconjug Chem* (2013) **24**, 8:1277-1294.

74   Szychowski, J. *et al.* Cleavable biotin probes for labeling of biomolecules via azide-alkyne cycloaddition. *J. Am. Chem. Soc.* (2010) **132**, 51:18351-18360.

75   Kamemura, K., Hayes, B.K., Comer, F.I. & Hart, G.W. Dynamic interplay between O-glycosylation and O-phosphorylation of nucleocytoplasmic proteins: alternative glycosylation/phosphorylation of THR-58, a known mutational hot spot of c-Myc in lymphomas, is regulated by mitogens. *J. Biol. Chem.* (2002) **277**, 21:19229-19235.

76   Leney, A.C., El Atmioui, D., Wu, W., Ovaa, H. & Heck, A.J.R. Elucidating crosstalk mechanisms between phosphorylation and O-GlcNAcylation. *Proc Natl Acad Sci U S A* (2017) **114**, 35:E7255-E7261.

77   Wells, L., Kreppel, L.K., Comer, F.I., Wadzinski, B.E. & Hart, G.W. O-GlcNAc transferase is in a functional complex with protein phosphatase 1 catalytic subunits. *J. Biol. Chem.* (2004) **279**, 37:38466-38470.

78   Dias, W.B., Cheung, W.D., Wang, Z. & Hart, G.W. Regulation of calcium/calmodulin-dependent kinase IV by O-GlcNAc modification. *J. Biol. Chem.* (2009) **284**, 32:21327-21337.

79   Xu, Q. *et al.* AMPK regulates histone H2B O-GlcNAcylation. *Nucleic Acids Res.* (2014) **42**, 9:5594-5604.

80   Bullen, J.W. *et al.* Cross-talk between two essential nutrient-sensitive enzymes: O-GlcNAc transferase (OGT) and AMP-activated protein kinase (AMPK). *J. Biol. Chem.* (2014) **289**, 15:10592-10606.

81   Erickson, J.R. *et al.* Diabetic hyperglycaemia activates CaMKII and arrhythmias by O-linked glycosylation. *Nature* (2013) **502**, 7471:372-376.

82   Anderson, M.E., Brown, J.H. & Bers, D.M. CaMKII in myocardial hypertrophy and heart failure. *J. Mol. Cell. Cardiol.* (2011) **51**, 4:468-473.

83   Ashpole, N.M. & Hudmon, A. Excitotoxic neuroprotection and vulnerability with CaMKII inhibition. *Mol. Cell. Neurosci.* (2011) **46**, 4:720-730.

84   Ruan, H.B. *et al.* Calcium-dependent O-GlcNAc signaling drives liver autophagy in adaptation to starvation. *Genes Dev* (2017) **31**, 16:1655-1665.

85   Fujiki, R. *et al.* GlcNAcylation of histone H2B facilitates its monoubiquitination. *Nature* (2011) **480**, 7378:557-560.

86   Yang, X. *et al.* O-linkage of N-acetylglucosamine to Sp1 activation domain inhibits its transcriptional capability. *Proc. Natl. Acad. Sci. U. S. A.* (2001) **98**, 12:6611-6616.

87   Yang, W.H. *et al.* Modification of p53 with O-linked N-acetylglucosamine regulates p53 activity and stability. *Nat. Cell Biol.* (2006) **8**, 10:1074-1083.

88   Shtraizent, N. *et al.* MPI depletion enhances O-GlcNAcylation of p53 and suppresses the Warburg effect. *Elife* (2017) **6**, e22477.

89   de Queiroz, R.M., Madan, R., Chien, J., Dias, W.B. & Slawson, C. Changes in O-Linked N-Acetylglucosamine (O-GlcNAc) Homeostasis Activate the p53 Pathway in Ovarian Cancer Cells. *J. Biol. Chem.* (2016) **291**, 36:18897-18914.

90   Li, Y. *et al.* O-GlcNAcylation modulates Bmi-1 protein stability and potential oncogenic function in prostate cancer. *Oncogene* (2017).

91   Chu, C.S. *et al.* O-GlcNAcylation regulates EZH2 protein stability and function. *Proc. Natl. Acad. Sci. U. S. A.* (2014) **111**, 4:1355-1360.

92   Moon, S., Lee, Y.K., Lee, S.W. & Um, S.J. Suppressive role of OGT-mediated O-GlcNAcylation of BAP1 in retinoic acid signaling. *Biochem Biophys Res Commun* (2017) **492**, 1:89-95.

93   Dey, A. *et al.* Loss of the tumor suppressor BAP1 causes myeloid transformation. *Science* (2012) **337**, 6101:1541-1546.

94   Yu, H. *et al.* The ubiquitin carboxyl hydrolase BAP1 forms a ternary complex with YY1 and HCF-1 and is a critical regulator of gene expression. *Mol Cell Biol* (2010) **30**, 21:5071-5085.

95   Shafi, R. *et al.* The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny. *Proc. Natl. Acad. Sci. U. S. A.* (2000) **97**, 11:5735-5739.

96   O'Donnell, N., Zachara, N.E., Hart, G.W. & Marth, J.D. Ogt-dependent X-chromosome-linked protein glycosylation is a requisite modification in somatic cell function and embryo viability. *Mol Cell Biol* (2004) **24**, 4:1680-1690.

97   Yang YR, S.M., Lee H, Jeon Y, Choi EJ, Jang HJ, Moon HY, Byun HY, Kim EK, Kim DH, Lee MN, Koh A, Ghim J, Choi JH, Lee-Kwon W, Kim KT, Ryu SH, Suh PG. O-GlcNAcase is essential for embryonic development and maintenance of genomic stability. *Aging Cell.* (2012) **11**, 3:439-448.

98   Olivier-Van Stichelen, S., Wang, P., Comly, M., Love, D.C. & Hanover, J.A. Nutrient-driven O-linked N-acetylglucosamine (O-GlcNAc) cycling impacts neurodevelopmental timing and metabolism. *J Biol Chem* (2017) **292**, 15:6076-6085.

99   Love, D.C. *et al.* Dynamic O-GlcNAc cycling at promoters of Caenorhabditis elegans genes regulating longevity, stress, and immunity. *Proc. Natl. Acad. Sci. U. S. A.* (2010) **107**, 16:7413-7418.

100   Gambetta, M.C., Oktaba, K. & Muller, J. Essential role of the glycosyltransferase sxc/Ogt in polycomb repression. *Science* (2009) **325**, 5936:93-96.

101   Constable, S., Lim, J.M., Vaidyanathan, K. & Wells, L. O-GlcNAc transferase regulates transcriptional activity of human Oct4. *Glycobiology* (2017) **27**, 10:927-937.

102   Myers, S.A. *et al.* SOX2 O-GlcNAcylation alters its protein-protein interactions and genomic occupancy to modulate gene expression in pluripotent cells. *Elife* (2016) **5**, e10647.

103   Jang, H. *et al.* O-GlcNAc regulates pluripotency and reprogramming by directly acting on core components of the pluripotency network. *Cell Stem Cell* (2012) **11**, 1:62-74.

104   Chen, Y. *et al.* NRAGE induces beta-catenin/Arm O-GlcNAcylation and negatively regulates Wnt signaling. *Biochem. Biophys. Res. Commun.* (2017) **487**, 2:433-437.

105   Koyama, T. & Kamemura, K. Global increase in O-linked N-acetylglucosamine modification promotes osteoblast differentiation. *Exp. Cell Res.* (2015) **338**, 2:194-202.

106   Chatham, J.C. & Marchase, R.B. The role of protein O-linked beta-N-acetylglucosamine in mediating cardiac stress responses. *Biochim. Biophys. Acta* (2010) **1800**, 2:57-66.

107   Ngoh, G.A., Watson, L.J., Facundo, H.T. & Jones, S.P. Augmented O-GlcNAc signaling attenuates oxidative stress and calcium overload in cardiomyocytes. *Amino Acids* (2011) **40**, 3:895-911.

108   Banerjee, P.S., Ma, J. & Hart, G.W. Diabetes-associated dysregulation of O-GlcNAcylation in rat cardiac mitochondria. *Proc. Natl. Acad. Sci. U. S. A.* (2015) **112**, 19:6050-6055.

109   Watson, L.J. *et al.* O-linked beta-N-acetylglucosamine transferase is indispensable in the failing heart. *Proc. Natl. Acad. Sci. U. S. A.* (2010) **107**, 41:17797-17802.

110   Ferrer, C.M. *et al.* O-GlcNAcylation regulates cancer metabolism and survival stress signaling via regulation of the HIF-1 pathway. *Mol. Cell* (2014) **54**, 5:820-831.

111   Housley, M.P. *et al.* A PGC-1alpha-O-GlcNAc transferase complex regulates FoxO transcription factor activity in response to glucose. *J. Biol. Chem.* (2009) **284**, 8:5148-5157.

112   Ho, S.R. *et al.* O-GlcNAcylation enhances FOXO4 transcriptional regulation in response to stress. *FEBS Lett* (2010) **584**, 1:49-54.

113   Radermacher, P.T. *et al.* O-GlcNAc reports ambient temperature and confers heat resistance on ectotherm development. *Proc. Natl. Acad. Sci. U. S. A.* (2014) **111**, 15:5592-5597.

114   Xu, W. *et al.* O-GlcNAc transferase promotes fatty liver-associated liver cancer through inducing palmitic acid and activating endoplasmic reticulum stress. *J. Hepatol.* (2017) **67**, 2:310-320.

115   Zachara, N.E. *et al.* Dynamic O-GlcNAc modification of nucleocytoplasmic proteins in response to stress. A survival response of mammalian cells. *J. Biol. Chem.* (2004) **279**, 29:30133-30142.

116   Alejandro, E.U. *et al.* Disruption of O-linked N-Acetylglucosamine Signaling Induces ER Stress and beta Cell Failure. *Cell Rep* (2015) **13**, 11:2527-2538.
117   Ngoh, G.A., Hamid, T., Prabhu, S.D. & Jones, S.P. O-GlcNAc signaling attenuates ER stress-induced cardiomyocyte death. *Am. J. Physiol. Heart Circ. Physiol.* (2009) **297**, 5:H1711-1719.
118   Groves, J.A., Lee, A., Yildirir, G. & Zachara, N.E. Dynamic O-GlcNAcylation and its roles in the cellular stress response and homeostasis. *Cell Stress Chaperones* (2013) **18**, 5:535-558.
119   Chatham, J.C. & Marchase, R.B. Protein O-GlcNAcylation: A critical regulator of the cellular response to stress. *Curr. Signal Transduct. Ther.* (2010) **5**, 1:49-59.
120   Ohn, T., Kedersha, N., Hickman, T., Tisdale, S. & Anderson, P. A functional RNAi screen links O-GlcNAc modification of ribosomal proteins to stress granule and processing body assembly. *Nat. Cell Biol.* (2008) **10**, 10:1224-1231.
121   Buchan, J.R. & Parker, R. Eukaryotic stress granules: the ins and outs of translation. *Mol Cell* (2009) **36**, 6:932-941.
122   He, C. & Klionsky, D.J. Regulation mechanisms and signaling pathways of autophagy. *Annu. Rev. Genet.* (2009) **43**, 1:67-93.
123   Wang, P. & Hanover, J.A. Nutrient-driven O-GlcNAc cycling influences autophagic flux and neurodegenerative proteotoxicity. *Autophagy* (2013) **9**, 4:604-606.
124   Marsh, S.A., Powell, P.C., Dell'italia, L.J. & Chatham, J.C. Cardiac O-GlcNAcylation blunts autophagic signaling in the diabetic heart. *Life Sci* (2013) **92**, 11:648-656.
125   Park, S. *et al.* O-GlcNAc modification is essential for the regulation of autophagy in Drosophila melanogaster. *Cell. Mol. Life Sci.* (2015) **72**, 16:3173-3183.
126   Rahman, M.M. *et al.* Intracellular protein glycosylation modulates insulin mediated lifespan in C.elegans. *Aging* (2010) **2**, 10:678-690.
127   Li, H. *et al.* O-GlcNAcylation of SKN-1 modulates the lifespan and oxidative stress resistance in Caenorhabditis elegans. *Sci Rep* (2017) **7**, 43601.
128   Liu, Y. *et al.* Developmental regulation of protein O-GlcNAcylation, O-GlcNAc transferase, and O-GlcNAcase in mammalian brain. *PLoS One* (2012) **7**, 8:e43724.
129   Kim, E.Y. *et al.* A role for O-GlcNAcylation in setting circadian clock speed. *Genes Dev.* (2012) **26**, 5:490-502.
130   Li, M.D. *et al.* O-GlcNAc signaling entrains the circadian clock by inhibiting BMAL1/CLOCK ubiquitination. *Cell Metab.* (2013) **17**, 2:303-310.
131   Kaasik, K. *et al.* Glucose sensor O-GlcNAcylation coordinates with phosphorylation to regulate circadian clock. *Cell Metab.* (2013) **17**, 2:291-302.
132   Menet, J.S., Pescatore, S. & Rosbash, M. CLOCK:BMAL1 is a pioneer-like transcription factor. *Genes Dev* (2014) **28**, 1:8-13.
133   Stojkovic, K., Wing, S.S. & Cermakian, N. A central role for ubiquitination within a circadian clock protein modification code. *Front Mol Neurosci* (2014) **7**, 69:69.
134   Sahar, S., Zocchi, L., Kinoshita, C., Borrelli, E. & Sassone-Corsi, P. Regulation of BMAL1 protein stability and circadian function by GSK3beta-mediated phosphorylation. *PLoS One* (2010) **5**, 1:e8561.
135   Yoshitane, H. *et al.* Roles of CLOCK phosphorylation in suppression of E-box-dependent transcription. *Mol Cell Biol* (2009) **29**, 13:3675-3686.
136   Toh, K.L. *et al.* An hPer2 phosphorylation site mutation in familial advanced sleep phase syndrome. *Science* (2001) **291**, 5506:1040-1043.
137   Xu, Y. *et al.* Modeling of a human circadian mutation yields insights into clock regulation by PER2. *Cell* (2007) **128**, 1:59-70.
138   Drougat, L. *et al.* Characterization of O-GlcNAc cycling and proteomic identification of differentially O-GlcNAcylated proteins during G1/S transition. *Biochim. Biophys. Acta* (2012) **1820**, 12:1839-1848.
139   Capotosti, F. *et al.* O-GlcNAc transferase catalyzes site-specific proteolysis of HCF-1. *Cell* (2011) **144**, 3:376-388.
140   Slawson, C. *et al.* Perturbations in O-linked beta-N-acetylglucosamine protein modification cause severe defects in mitotic progression and cytokinesis. *J. Biol. Chem.* (2005) **280**, 38:32944-32956.
141   Daou, S. *et al.* Crosstalk between O-GlcNAcylation and proteolytic cleavage regulates the host cell factor-1 maturation pathway. *Proc. Natl. Acad. Sci. U. S. A.* (2011) **108**, 7:2747-2752.

142   Buchakjian, M.R. & Kornbluth, S. The engine driving the ship: metabolic steering of cell proliferation and death. *Nat Rev Mol Cell Biol* (2010) **11**, 10:715-727.
143   Sakabe, K. & Hart, G.W. O-GlcNAc transferase regulates mitotic chromatin dynamics. *J. Biol. Chem.* (2010) **285**, 45:34460-34468.
144   Wang, S. *et al.* Extensive crosstalk between O-GlcNAcylation and phosphorylation regulates Akt signaling. *PLoS One* (2012) **7**, 5:e37427.
145   Wells, L., Slawson, C. & Hart, G.W. The E2F-1 associated retinoblastoma-susceptibility gene product is modified by O-GlcNAc. *Amino Acids* (2011) **40**, 3:877-883.
146   Giacinti, C. & Giordano, A. RB and cell cycle progression. *Oncogene* (2006) **25**, 38:5220-5227.
147   Zargar, Z. & Tyagi, S. Role of host cell factor-1 in cell cycle regulation. *Transcription* (2012) **3**, 4:187-192.
148   Muthusamy, S., Hong, K.U., Dassanayaka, S., Hamid, T. & Jones, S.P. E2F1 Transcription Factor Regulates O-linked N-acetylglucosamine (O-GlcNAc) Transferase and O-GlcNAcase Expression. *J. Biol. Chem.* (2015) **290**, 52:31013-31024.
149   Shi, J., Sharif, S., Ruijtenbeek, R. & Pieters, R.J. Activity Based High-Throughput Screening for Novel O-GlcNAc Transferase Substrates Using a Dynamic Peptide Microarray. *PLoS One* (2016) **11**, 3:e0151085.
150   Trinidad, J.C. *et al.* Global Identification and Characterization of Both O-GlcNAcylation and Phosphorylation at the Murine Synapse. *Molecular & Cellular Proteomics* (2012) **11**, 8:215-229.
151   Ning, X. *et al.* The O-GlcNAc Modification of CDK5 Involved in Neuronal Apoptosis Following In Vitro Intracerebral Hemorrhage. *Cell Mol Neurobiol* (2017) **37**, 3:527-536.
152   Teo, C.F. *et al.* Glycopeptide-specific monoclonal antibodies suggest new roles for O-GlcNAc. *Nat. Chem. Biol.* (2010) **6**, 5:338-343.
153   Kohoutek, J. & Blazek, D. Cyclin K goes with Cdk12 and Cdk13. *Cell Div* (2012) **7**, 1:12.
154   Blazek, D. *et al.* The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. *Genes Dev* (2011) **25**, 20:2158-2172.
155   Shimobayashi, M. & Hall, M.N. Making new contacts: the mTOR network in metabolism and signalling crosstalk. *Nat Rev Mol Cell Biol* (2014) **15**, 3:155-162.
156   Mendoza, M.C., Er, E.E. & Blenis, J. The Ras-ERK and PI3K-mTOR pathways: cross-talk and compensation. *Trends Biochem Sci* (2011) **36**, 6:320-328.
157   Butkinaree, C., Park, K. & Hart, G.W. O-linked beta-N-acetylglucosamine (O-GlcNAc): Extensive crosstalk with phosphorylation to regulate signaling and transcription in response to nutrients and stress. *Biochim. Biophys. Acta* (2010) **1800**, 2:96-106.
158   Jackson, R.J., Hellen, C.U. & Pestova, T.V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* (2010) **11**, 2:113-127.
159   Zeidan, Q., Wang, Z., De Maio, A. & Hart, G.W. O-GlcNAc cycling enzymes associate with the translational machinery and modify core ribosomal proteins. *Mol. Biol. Cell* (2010) **21**, 12:1922-1936.
160   Zhang, F. *et al.* O-GlcNAc modification is an endogenous inhibitor of the proteasome. *Cell* (2003) **115**, 6:715-725.
161   Sumegi, M., Hunyadi-Gulyas, E., Medzihradszky, K.F. & Udvardy, A. 26S proteasome subunits are O-linked N-acetylglucosamine-modified in Drosophila melanogaster. *Biochem Biophys Res Commun* (2003) **312**, 4:1284-1289.
162   Ma, J. & Hart, G.W. Protein O-GlcNAcylation in diabetes and diabetic complications. *Expert Rev Proteomics* (2013) **10**, 4:365-380.
163   Issad, T. & Kuo, M. O-GlcNAc modification of transcription factors, glucose sensing and glucotoxicity. *Trends Endocrinol. Metab.* (2008) **19**, 10:380-389.
164   Durning, S.P., Flanagan-Steet, H., Prasad, N. & Wells, L. O-Linked beta-N-acetylglucosamine (O-GlcNAc) Acts as a Glucose Sensor to Epigenetically Regulate the Insulin Gene in Pancreatic Beta Cells. *J. Biol. Chem.* (2016) **291**, 5:2107-2118.
165   Taylor, R.P. *et al.* Glucose deprivation stimulates O-GlcNAc modification of proteins through up-regulation of O-linked N-acetylglucosaminyltransferase. *J. Biol. Chem.* (2008) **283**, 10:6050-6057.
166   Tan, E.P. *et al.* Altering O-linked beta-N-acetylglucosamine cycling disrupts mitochondrial function. *J. Biol. Chem.* (2014) **289**, 21:14719-14730.
167   Perez-Cervera, Y. *et al.* Insulin signaling controls the expression of O-GlcNAc transferase and its interaction with lipid microdomains. *FASEB J* (2013) **27**, 9:3478-3486.

168   Hanover, J.A. *et al.* A Caenorhabditis elegans model of insulin resistance: altered macronutrient storage and dauer formation in an OGT-1 knockout. *Proc. Natl. Acad. Sci. U. S. A.* (2005) **102**, 32:11266-11271.

169   Springhorn, C., Matsha, T.E., Erasmus, R.T. & Essop, M.F. Exploring leukocyte O-GlcNAcylation as a novel diagnostic tool for the earlier detection of type 2 diabetes mellitus. *J Clin Endocrinol Metab* (2012) **97**, 12:4640-4649.

170   Myslicki, J.P., Shearer, J., Hittel, D.S., Hughey, C.C. & Belke, D.D. O-GlcNAc modification is associated with insulin sensitivity in the whole blood of healthy young adult males. *Diabetol Metab Syndr* (2014) **6**, 1:96.

171   Keembiyehetty, C. *et al.* Conditional knock-out reveals a requirement for O-linked N-Acetylglucosaminase (O-GlcNAcase) in metabolic homeostasis. *J. Biol. Chem.* (2015) **290**, 11:7097-7113.

172   Dentin, R., Hedrick, S., Xie, J., Yates, J., 3rd & Montminy, M. Hepatic glucose sensing via the CREB coactivator CRTC2. *Science* (2008) **319**, 5868:1402-1405.

173   Takemoto, T. *et al.* RBMX is a novel hepatic transcriptional regulator of SREBP-1c gene response to high-fructose diet. *FEBS Lett* (2007) **581**, 2:218-222.

174   Guinez, C. *et al.* O-GlcNAcylation increases ChREBP protein content and transcriptional activity in the liver. *Diabetes* (2011) **60**, 5:1399-1413.

175   Sodi, V.L. *et al.* Nutrient sensor O-GlcNAc transferase controls cancer lipid metabolism via SREBP-1 regulation. *Oncogene* (2017).

176   Ozcan, S., Andrali, S.S. & Cantrell, J.E. Modulation of transcription factor function by O-GlcNAc modification. *Biochim. Biophys. Acta* (2010) **1799**, 5-6:353-364.

177   Wang, X. *et al.* O-GlcNAcase deficiency suppresses skeletal myogenesis and insulin sensitivity in mice through the modulation of mitochondrial homeostasis. *Diabetologia* (2016) **59**, 6:1287-1296.

178   Pekkurnaz, G., Trinidad, J.C., Wang, X., Kong, D. & Schwarz, T.L. Glucose regulates mitochondrial motility via Milton modification by O-GlcNAc transferase. *Cell* (2014) **158**, 1:54-68.

179   Weimer, S. *et al.* D-Glucosamine supplementation extends life span of nematodes and of ageing mice. *Nat Commun* (2014) **5**, 3563.

180   Tan, E.P. *et al.* Sustained O-GlcNAcylation reprograms mitochondrial function to regulate energy metabolism. *J Biol Chem* (2017) **292**, 36:14940-14962.

181   Sacoman, J.L., Dagda, R.Y., Burnham-Marusich, A.R., Dagda, R.K. & Berninsone, P.M. Mitochondrial O-GlcNAc Transferase (mOGT) Regulates Mitochondrial Structure, Function, and Survival in HeLa Cells. *J Biol Chem* (2017) **292**, 11:4499-4518.

182   Ma, J. *et al.* O-GlcNAcomic Profiling Identifies Widespread O-Linked beta-N-Acetylglucosamine Modification (O-GlcNAcylation) in Oxidative Phosphorylation System Regulating Cardiac Mitochondrial Function. *J. Biol. Chem.* (2015) **290**, 49:29141-29153.

183   Lozano, L., Lara-Lemus, R., Zenteno, E. & Alvarado-Vasquez, N. The mitochondrial O-linked N-acetylglucosamine transferase (mOGT) in the diabetic patient could be the initial trigger to develop Alzheimer disease. *Exp Gerontol* (2014) **58**, Supplement C:198-202.

184   Trapannone, R., Mariappa, D., Ferenbach, A.T. & van Aalten, D.M.F. Nucleocytoplasmic human O-GlcNAc transferase is sufficient for O-GlcNAcylation of mitochondrial proteins. *Biochemical Journal* (2016) **473**, 12:1693-1702.

185   Dassanayaka, S. *et al.* High glucose induces mitochondrial dysfunction independently of protein O-GlcNAcylation. *Biochem J* (2015) **467**, 1:115-126.

186   Li, Z. & Yi, W. Regulation of cancer metabolism by O-GlcNAcylation. *Glycoconj. J.* (2014) **31**, 3:185-191.

187   Mi, W. *et al.* O-GlcNAcylation is a novel regulator of lung and colon cancer malignancy. *Biochim Biophys Acta* (2011) **1812**, 4:514-519.

188   Slawson, C. & Hart, G.W. O-GlcNAc signalling: implications for cancer cell biology. *Nat. Rev. Cancer* (2011) **11**, 9:678-684.

189   Jozwiak, P., Forma, E., Brys, M. & Krzeslak, A. O-GlcNAcylation and Metabolic Reprograming in Cancer. *Front. Endocrinol. (Lausanne)* (2014) **5**, 145.

190   Ferrer, C.M., Sodi, V.L. & Reginato, M.J. O-GlcNAcylation in Cancer Biology: Linking Metabolism and Signaling. *J Mol Biol* (2016) **428**, 16:3282-3294.

191   Ferrer, C.M. *et al.* O-GlcNAcylation regulates breast cancer metastasis via SIRT1 modulation of FOXM1 pathway. *Oncogene* (2017) **36**, 4:559-569.

192   Lynch, T.P. *et al.* Critical role of O-Linked beta-N-acetylglucosamine transferase in prostate cancer invasion, angiogenesis, and metastasis. *J. Biol. Chem.* (2012) **287**, 14:11070-11081.

193   Forma, E., Jozwiak, P., Brys, M. & Krzeslak, A. The potential role of O-GlcNAc modification in cancer epigenetics. *Cell Mol Biol Lett* (2014) **19**, 3:438-460.

194   Rao, X. *et al.* O-GlcNAcylation of G6PD promotes the pentose phosphate pathway and tumor growth. *Nat Commun* (2015) **6**, 8468.

195   Yi, W. *et al.* Phosphofructokinase 1 glycosylation regulates cell growth and metabolism. *Science* (2012) **337**, 6097:975-980.

196   Patra, K.C. & Hay, N. The pentose phosphate pathway and cancer. *Trends Biochem Sci* (2014) **39**, 8:347-354.

197   Lucena, M.C. *et al.* Epithelial Mesenchymal Transition Induces Aberrant Glycosylation through Hexosamine Biosynthetic Pathway Activation. *J Biol Chem* (2016) **291**, 25:12917-12929.

198   Taparra, K., Tran, P.T. & Zachara, N.E. Hijacking the Hexosamine Biosynthetic Pathway to Promote EMT-Mediated Neoplastic Phenotypes. *Front Oncol* (2016) **6**, 85:85.

199   Park, S.Y. *et al.* Snail1 is stabilized by O-GlcNAc modification in hyperglycaemic condition. *EMBO J* (2010) **29**, 22:3787-3796.

200   Huang, X. *et al.* O-GlcNAcylation of cofilin promotes breast cancer cell invasion. *J Biol Chem* (2013) **288**, 51:36418-36425.

201   Gu, Y. *et al.* GlcNAcylation plays an essential role in breast cancer metastasis. *Cancer Res* (2010) **70**, 15:6344-6351.

202   Qiao, Y.X. *et al.* High Glucose Stimulates Tumorigenesis in Hepatocellular Carcinoma Cells Through AGER-Dependent O-GlcNAcylation of c-Jun. *Diabetes* (2016) **65**, 3:619-632.

203   Guo, H. *et al.* O-Linked N-Acetylglucosamine (O-GlcNAc) Expression Levels Epigenetically Regulate Colon Cancer Tumorigenesis by Affecting the Cancer Stem Cell Compartment via Modulating Expression of Transcriptional Factor MYBL1. *J Biol Chem* (2017) **292**, 10:4123-4137.

204   Li, F., Tiede, B., Massague, J. & Kang, Y. Beyond tumorigenesis: cancer stem cells in metastasis. *Cell Res* (2007) **17**, 1:3-14.

205   Liu, K. *et al.* Accumulation of protein O-GlcNAc modification inhibits proteasomes in the brain and coincides with neuronal apoptosis in brain areas with high O-GlcNAc metabolism. *J. Neurochem.* (2004) **89**, 4:1044-1055.

206   Lagerlof, O., Hart, G.W. & Huganir, R.L. O-GlcNAc transferase regulates excitatory synapse maturity. *Proc. Natl. Acad. Sci. U. S. A.* (2017) **114**, 7:1684-1689.

207   Francisco, H. *et al.* O-GlcNAc post-translational modifications regulate the entry of neurons into an axon branching program. *Developmental Neurobiology* (2009) **69**, 2-3:162-173.

208   Tallent, M.K. *et al.* In vivo modulation of O-GlcNAc levels regulates hippocampal synaptic plasticity through interplay with phosphorylation. *J. Biol. Chem.* (2009) **284**, 1:174-181.

209   Kanno, T., Yaguchi, T., Nagata, T., Mukasa, T. & Nishizaki, T. Regulation of AMPA receptor trafficking by O-glycosylation. *Neurochem. Res.* (2010) **35**, 5:782-788.

210   Taylor, E.W. *et al.* O-GlcNAcylation of AMPA receptor GluA2 is associated with a novel form of long-term depression at hippocampal synapses. *J. Neurosci.* (2014) **34**, 1:10-21.

211   Stewart, L.T. *et al.* Acute Increases in Protein O-GlcNAcylation Dampen Epileptiform Activity in Hippocampus. *J Neurosci* (2017) **37**, 34:8207-8215.

212   Yang, Y.R. *et al.* Memory and synaptic plasticity are impaired by dysregulated hippocampal O-GlcNAcylation. *Sci Rep* (2017) **7**, 44921.

213   Skorobogatko, Y. *et al.* O-linked beta-N-acetylglucosamine (O-GlcNAc) site thr-87 regulates synapsin I localization to synapses and size of the reserve pool of synaptic vesicles. *J. Biol. Chem.* (2014) **289**, 6:3602-3612.

214   Ruan, H.B. *et al.* O-GlcNAc transferase enables AgRP neurons to suppress browning of white fat. *Cell* (2014) **159**, 2:306-317.

215   Lagerlof, O. *et al.* The nutrient sensor OGT in PVN neurons regulates feeding. *Science* (2016) **351**, 6279:1293-1296.

216   van Spronsen, M. *et al.* TRAK/Milton motor-adaptor proteins steer mitochondrial trafficking to axons and dendrites. *Neuron* (2013) **77**, 3:485-502.

217   Iyer, S.P. & Hart, G.W. Roles of the tetratricopeptide repeat domain in O-GlcNAc transferase targeting and protein substrate specificity. *J. Biol. Chem.* (2003) **278**, 27:24608-24616.

218   Rexach, J.E. *et al.* Dynamic O-GlcNAc Glycosylation Regulates CREB-Mediated Gene Expression and Memory Formation. *submitted* (2011).

219   Willems, A.P. *et al.* Mutations in N-acetylglucosamine (O-GlcNAc) transferase in patients with X-linked intellectual disability. *J Biol Chem* (2017) **292**, 30:12621-12631.

220   Vaidyanathan, K. *et al.* Identification and characterization of a missense mutation in the O-linked beta-N-acetylglucosamine (O-GlcNAc) transferase gene that segregates with X-linked intellectual disability. *J. Biol. Chem.* (2017) **292**, 21:8948-8963.

221   Ma, X., Li, H., He, Y. & Hao, J. The emerging link between O-GlcNAcylation and neurological disorders. *Cell Mol Life Sci* (2017) **74**, 20:3667-3686.

222   Dugger, B.N. & Dickson, D.W. Pathology of Neurodegenerative Diseases. *Cold Spring Harbor perspectives in biology* (2017) **9**, 7.

223   Hardy, J. & Gwinn-Hardy, K. Genetic classification of primary neurodegenerative disease. *Science* (1998) **282**, 5391:1075-1079.

224   Bertram, L. *et al.* Evidence for genetic linkage of Alzheimer's disease to chromosome 10q. *Science* (2000) **290**, 5500:2302-2303.

225   Twine, N.A., Janitz, K., Wilkins, M.R. & Janitz, M. Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease. *PLoS One* (2011) **6**, 1:e16266.

226   Wells, L. Mutations in O-GlcNAc Transferase Linked to X-linked Intellectual Disability. *The FASEB Journal* (2016) **30**, 1 Supplement:98.95.

227   Nemeth, A.H. *et al.* Refined linkage disequilibrium and physical mapping of the gene locus for X-linked dystonia-parkinsonism (DYT3). *Genomics* (1999) **60**, 3:320-329.

228   Ito, N. *et al.* Decreased N-TAF1 expression in X-linked dystonia-parkinsonism patient-specific neural stem cells. *Dis Model Mech* (2016) **9**, 4:451-462.

229   Gong, C.X., Liu, F. & Iqbal, K. O-GlcNAc cycling modulates neurodegeneration. *Proc. Natl. Acad. Sci. U. S. A.* (2012) **109**, 43:17319-17320.

230   Levine, P.M. *et al.* O-GlcNAc modification inhibits the calpain-mediated cleavage of alpha-synuclein. *Bioorg Med Chem* (2017) **25**, 18:4977-4982.

231   Citron, M. Alzheimer's disease: strategies for disease modification. *Nat Rev Drug Discov* (2010) **9**, 5:387-398.

232   Skorobogatko, Y.V. *et al.* Human Alzheimer's disease synaptic O-GlcNAc site mapping and iTRAQ expression proteomics with ion trap mass spectrometry. *Amino Acids* (2011) **40**, 3:765-779.

233   Liu, F. *et al.* Reduced O-GlcNAcylation links lower brain glucose metabolism and tau pathology in Alzheimer's disease. *Brain* (2009) **132**, Pt 7:1820-1832.

234   Liu, F., Iqbal, K., Grundke-Iqbal, I., Hart, G.W. & Gong, C.X. O-GlcNAcylation regulates phosphorylation of tau: a mechanism involved in Alzheimer's disease. *Proc. Natl. Acad. Sci. U. S. A.* (2004) **101**, 29:10804-10809.

235   Jacobsen, K.T. & Iverfeldt, K. O-GlcNAcylation increases non-amyloidogenic processing of the amyloid-β precursor protein (APP). *Biochemical and Biophysical Research Communications* (2011) **404**, 3:882-886.

236   Zhang, Y.W., Thompson, R., Zhang, H. & Xu, H. APP processing in Alzheimer's disease. *Mol Brain* (2011) **4**, 1:3.

237   Jacobsen, K.T. & Iverfeldt, K. O-GlcNAcylation increases non-amyloidogenic processing of the amyloid-beta precursor protein (APP). *Biochem. Biophys. Res. Commun.* (2011) **404**, 3:882-886.

238   Chun, Y.S., Kwon, O.H. & Chung, S. O-GlcNAcylation of amyloid-beta precursor protein at threonine 576 residue regulates trafficking and processing. *Biochem Biophys Res Commun* (2017) **490**, 2:486-491.

239   Kim, C. *et al.* O-linked beta-N-acetylglucosaminidase inhibitor attenuates beta-amyloid plaque and rescues memory impairment. *Neurobiol. Aging* (2013) **34**, 1:275-285.

240   Ballatore, C., Lee, V.M. & Trojanowski, J.Q. Tau-mediated neurodegeneration in Alzheimer's disease and related disorders. *Nat. Rev. Neurosci.* (2007) **8**, 9:663-672.

241   Nandi, A. *et al.* Global identification of O-GlcNAc-modified proteins. *Anal. Chem.* (2006) **78**, 2:452-458.

242   Iqbal, K. & Grundke-Iqbal, I. Metabolic/signal transduction hypothesis of Alzheimer's disease and other tauopathies. *Acta Neuropathol.* (2005) **109**, 1:25-31.

243   Cruz, J.C., Tseng, H.C., Goldman, J.A., Shih, H. & Tsai, L.H. Aberrant Cdk5 activation by p25 triggers pathological events leading to neurodegeneration and neurofibrillary tangles. *Neuron* (2003) **40**, 3:471-483.

244   Yuzwa, S.A. *et al.* Increasing O-GlcNAc slows neurodegeneration and stabilizes tau against aggregation. *Nat. Chem. Biol.* (2012) **8**, 4:393-399.

245   Yuzwa, S.A. *et al.* Mapping O-GlcNAc modification sites on tau and generation of a site-specific O-GlcNAc tau antibody. *Amino Acids* (2011) **40**, 3:857-868.

246   Smet-Nocca, C. *et al.* Identification of O-GlcNAc sites within peptides of the Tau protein and their impact on phosphorylation. *Mol. Biosyst.* (2011) **7**, 5:1420-1429.

247   Morris, M. *et al.* Tau post-translational modifications in wild-type and human amyloid precursor protein transgenic mice. *Nat. Neurosci.* (2015) **18**, 8:1183-1189.

248   Yuzwa, S.A., Cheung, A.H., Okon, M., McIntosh, L.P. & Vocadlo, D.J. O-GlcNAc modification of tau directly inhibits its aggregation without perturbing the conformational properties of tau monomers. *J. Mol. Biol.* (2014) **426**, 8:1736-1752.

249   Hastings, N.B. *et al.* Inhibition of O-GlcNAcase leads to elevation of O-GlcNAc tau and reduction of tauopathy and cerebrospinal fluid tau in rTg4510 mice. *Mol. Neurodegener.* (2017) **12**, 1:39.

250   Yuzwa, S.A. *et al.* Pharmacological inhibition of O-GlcNAcase (OGA) prevents cognitive decline and amyloid plaque formation in bigenic tau/APP mutant mice. *Mol. Neurodegener.* (2014) **9**, 1:42.

251   Graham, D.L. *et al.* Increased O-GlcNAcylation reduces pathological tau without affecting its normal phosphorylation in a mouse model of tauopathy. *Neuropharmacology* (2014) **79**, 307-313.

252   Forster, S. *et al.* Increased O-GlcNAc levels correlate with decreased O-GlcNAcase levels in Alzheimer disease brain. *Biochim. Biophys. Acta* (2014) **1842**, 9:1333-1339.

253   Beher, D. *Asceneuron Receives Approval for Phase I Study of Oral Tau Inhibitor*, <https://www.asceneuron.com/news/asceneuron-receives-approval-phase-i-study-oral-tau-inhibitor> (2017).

254   Dauer, W. & Przedborski, S. Parkinson's disease: mechanisms and models. *Neuron* (2003) **39**, 6:889-909.

255   Wani, W.Y. *et al.* O-GlcNAc regulation of autophagy and alpha-synuclein homeostasis; implications for Parkinson's disease. *Mol Brain* (2017) **10**, 1:32.

256   Giaime, E. *et al.* Age-Dependent Dopaminergic Neurodegeneration and Impairment of the Autophagy-Lysosomal Pathway in LRRK-Deficient Mice. *Neuron* (2017).

257   Marotta, N.P., Cherwien, C.A., Abeywardana, T. & Pratt, M.R. O-GlcNAc modification prevents peptide-dependent acceleration of alpha-synuclein aggregation. *ChemBioChem* (2012) **13**, 18:2665-2670.

258   Marotta, N.P. *et al.* O-GlcNAc modification blocks the aggregation and toxicity of the protein alpha-synuclein associated with Parkinson's disease. *Nat. Chem.* (2015) **7**, 11:913-920.

259   Lewis, Y.E. *et al.* O-GlcNAcylation of alpha-Synuclein at Serine 87 Reduces Aggregation without Affecting Membrane Binding. *ACS Chem Biol* (2017) **12**, 4:1020-1027.

260   Liu, J., Marchase, R.B. & Chatham, J.C. Increased O-GlcNAc levels during reperfusion lead to improved functional recovery and reduced calpain proteolysis. *Am J Physiol Heart Circ Physiol* (2007) **293**, 3:H1391-1399.

261   Irwin, D.J., Lee, V.M. & Trojanowski, J.Q. Parkinson's disease dementia: convergence of alpha-synuclein, tau and amyloid-beta pathologies. *Nat Rev Neurosci* (2013) **14**, 9:626-636.

262   Pasinelli, P. & Brown, R.H. Molecular biology of amyotrophic lateral sclerosis: insights from genetics. *Nat Rev Neurosci* (2006) **7**, 9:710-723.

263   Shan, X.Y. *O-linked N-acetylglucosamine protein modification in mouse models of neurodegenerative diseases* Ph.D. thesis, Simon Fraser University, (2011).

264   Ludemann, N. *et al.* O-glycosylation of the tail domain of neurofilament protein M in human neurons and in spinal cord tissue of a rat model of amyotrophic lateral sclerosis (ALS). *J. Biol. Chem.* (2005) **280**, 36:31648-31658.

265   Dong, D.L. *et al.* Glycosylation of mammalian neurofilaments. Localization of multiple O-linked N-acetylglucosamine moieties on neurofilament polypeptides L and M. *J. Biol. Chem.* (1993) **268**, 22:16679-16687.

266   Hornbeck, P.V. *et al.* PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res* (2015) **43**, Database issue:D512-520.

267   Shan, X., Vocadlo, D.J. & Krieger, C. Reduced protein O-glycosylation in the nervous system of the mutant SOD1 transgenic mouse model of amyotrophic lateral sclerosis. *Neurosci. Lett.* (2012) **516**, 2:296-301.
268   Hanover, J.A. & Wang, P. O-GlcNAc cycling shows neuroprotective potential in C. elegans models of neurodegenerative disease. *Worm* (2013) **2**, 4:e27043.
269   Wang, P. *et al.* O-GlcNAc cycling mutants modulate proteotoxicity in Caenorhabditis elegans models of human neurodegenerative diseases. *Proc. Natl. Acad. Sci. U. S. A.* (2012) **109**, 43:17669-17674.

Transcriptomic characterization of a forebrain-specific OGT cKO

Portions of this chapter are adapted from the published work:

2.1 Abstract

O-GlcNAc glycosylation is a dynamic, inducible post-translational modification (PTM) found on proteins associated with neurodegenerative diseases such as α-synuclein, amyloid precursor protein, and tau. Deletion of the *O*-GlcNAc transferase (*Ogt*) gene responsible for the modification causes early postnatal lethality in mice, complicating efforts to study *O*-GlcNAc glycosylation in mature neuronal function and dysfunction. Here, we report that forebrain-specific loss of OGT in adult mice leads to progressive neurodegeneration, including neuronal death, neuroinflammation, hyperphosphorylated tau, amyloidogenic Aβ-peptides, and memory deficits. In the OGT cKO hippocampus, we show the upregulation of neuroinflammatory genes and downregulation of cholesterol and lipid biosynthetic genes supporting a critical role of OGT in the regulation of lipid metabolism. Additionally, a gene network analysis (WGCNA) revealed that the OGT cKO mice showed perturbations in cell cycle, which were verified using immunohistochemistry (IHC). Another WGCNA comparing the OGT cKO with other FTDP and AD mouse models showed striking transcriptional similarities between the OGT cKO and neurofibrillary tangle (NFT)-forming FTDP/AD mouse models with increased correlation with a neuroinflammatory gene network and decreased correlation with a synaptic gene network. In contrast, amyloid plaque-forming AD mouse models showed no correlation with the synaptic gene network, suggesting that NFT formation specifically compromises synaptic transcription. These studies indicate that *O*-GlcNAcylation regulates pathways critical for the maintenance of neuronal health and suggest that dysfunctional *O*-GlcNAc signaling may be an important contributor to

neurodegenerative diseases.

2.2 General approach to generation of an OGT cKO mouse and validation

Previous studies have demonstrated that knocking out OGT in mice is embryonic lethal, and selectively knocking out neuronal OGT leads to motor dysfunction, brain development abnormalities, and eventual death.[1,2] This established OGT as a critical enzyme especially during development and in particular, neuronal development. In addition, *in situ* hybridization studies have shown that OGT and OGA are expressed at their highest levels in the hippocampus.[3] These previous studies encouraged our lab to cross a floxed OGT (OGTfl) mouse with mice with Cre downstream of a calcium/calmodulin kinase IIα (CaMKIIα) promoter (CaMKIIα-Cre, obtained from the Kennedy lab).[4] Wherever Cre integrase is expressed, it will selectively excise DNA segments that are flanked by LoxP sites, so that that excised DNA sequence is no longer transcribed.[5] The expression pattern for CaMKIIα in the murine brain is restricted primarily to the adult forebrain where it is expressed at high levels in the hippocampus and cortical neurons and at moderate levels in the caudate-putamen. Crucially, CaMKIIα is not expressed until after embryogenesis when its expression increases an order of magnitude from P1 to P21 and continues to gradually increase up to P90.[6] This design allows normal neural development to resume before *O*-GlcNAc modification is depleted in the adult. In this way, it mirrors the time course of the pathogenicity of adult-onset neurodegenerative disorders, which manifest well after neural development and considerable neural degeneration.[7] As OGT is an X-linked gene, only about half of males born from a CaMKIIα-Cre male x OGTfl female cross were forebrain-specific OGT knock-outs (fbOGT cKO).[2] The CaMKIIα-Cre transgene is expressed between postnatal

day 14-21 in excitatory neurons in the postnatal forebrain, including the cortex, hippocampus, caudate nucleus, thalamus, and hypothalamus.[8] Beginning at 1 month of age, the mice showed progressive depletion of *O*-GlcNAcylation in the hippocampus and cortex, but not in the cerebellum.

2.3 Overview of the phenotypes of the OGT cKO mouse: morphological and behavioral features

Andrew Wang, a former graduate student in the lab, characterized the OGT cKO mouse morphologically and behaviorally and showed the following OGT cKO features. Due to the embryonic and perinatal lethality phenotypes observed in total and neuronal-specific knockout, we sought to generate a knockout that would allow for normal neuronal development. Toward that end, the OGT cKO mouse was generated by crossing floxed *OGT* (OGTfl) mice with CaMKIIα-Cre transgenic mice.[2,8] By 6 months, few neurons were *O*-GlcNAc negative due to the loss of neurons lacking OGT and *O*-GlcNAc. Starting at 7 weeks of age, the OGT cKO began to show significant reductions in weight and brain size when compared to their wildtype littermates, which advanced to significant morphological changes in the cortex and hippocampus at 6 months of age (Figure 2.1A). At 2 months of age, progressive neuronal loss was observed in the hippocampus and cortex escalating to an 83% and 60% decrease in neuronal density in the dentate gyrus (DG) and CA1 respectively at 6 months of age (NeuN staining). In addition, the majority of DG region neurons and ~10% of cortical neurons displayed apoptotic neurodegeneration using Tdt dUTP Nick-Labeling (TUNEL) and Fluoro-Jade C (FJC) staining while their WT littermates displayed no TUNEL- or FJC-positive neurons (Figure 2.1B).

At 2 months of age, the OGT cKO mouse displayed significant increases in gliosis and neuroinflammation as measured by immunohistochemical (IHC) staining with the astrocyte marker, glial fibrillary acidic protein (GFAP), and the microglial marker, ionized calcium-binding adaptor molecule (Iba-1) (Figure 2.1B). Furthermore, the OGT cKO demonstrated increases in aberrantly and hyperphosphorylated tau and protein aggregation in the hippocampus (Thioflavine S staining) at both 2 and 6 months of age (Figure 2.1C). The ratio of the amyloidogenic 42-mer Aβ-peptide to the 40-mer Aβ-peptide increased 2.5-fold when compared to their wildtype littermates. These OGT cKO morphological changes are accompanied by long-term memory (LTM) impairments in amygdala-dependent cued and hippocampus-dependent contextual fear conditioning experiments at 4 months of age. Finally, Jessica Rexach, a former graduate student in the lab, showed that OGT protein levels were decreased by 1.6-fold in the brains of AD patients with the most severe memory decline (Braak VI) ($P < 0.0005$).
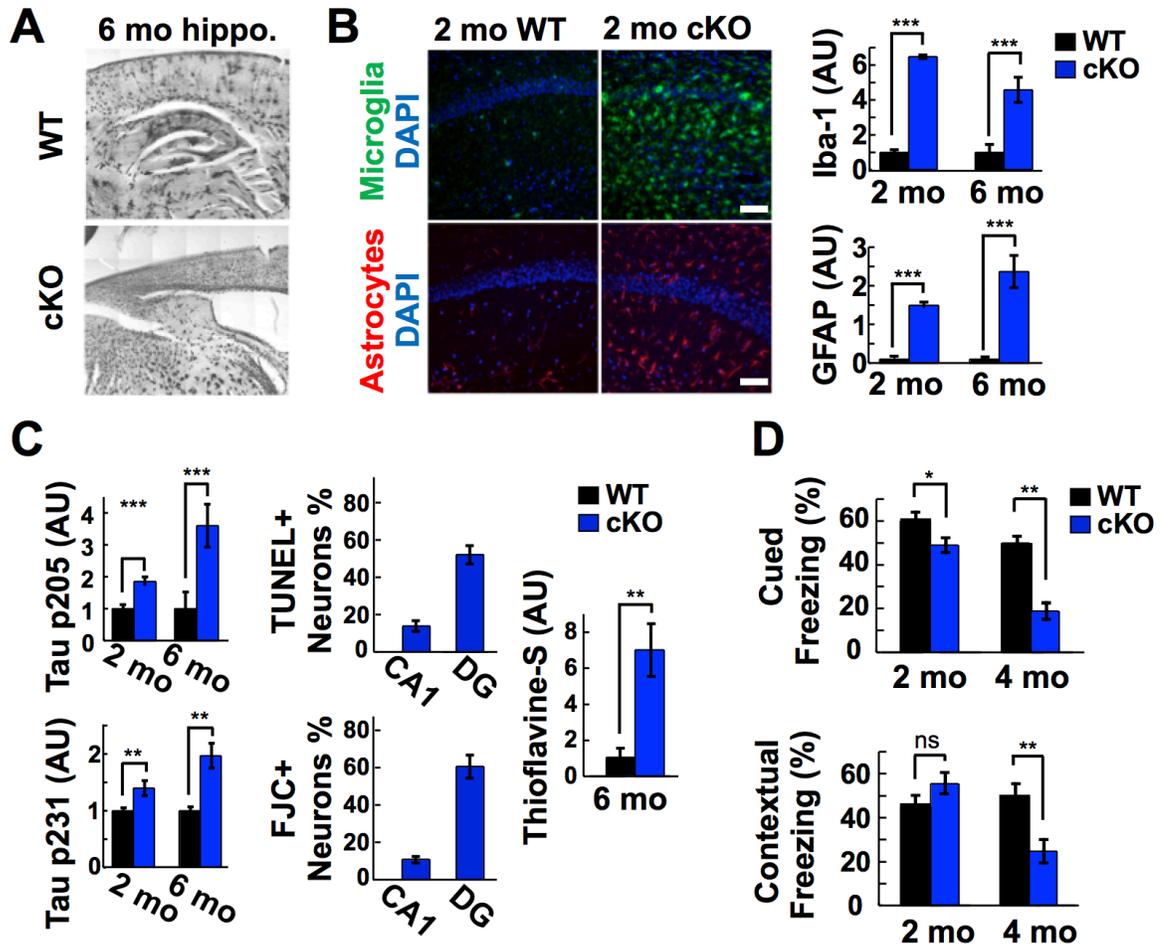
Figure 2.1 Summary of OGT cKO morphological and behavioral changes. (A) Hippocampal shrinkage, a decrease in the number of neurons, and disruption of the hippocampal neuronal network are revealed by golgi stains of the hippocampi of the WT and OGT cKO mice at 6 months of age. (B) Elevated expression of the neuroinflammatory markers Iba-1 (green) and GFAP (red) in the CA1 hippocampus of 2-month-old OGT cKO compared to WT mice. Total pixel intensity (green or red) from each field of view was quantified, and mean ± SEM is shown. $n = 3$. Scale bar = 50 μm. (C) Pathological phosphorylation of tau (pThr-205, pThr-231) increased in DG neurons of 2-month and 6-month-old OGT cKO mice based on IHC. Fluoro-Jade C and TUNEL staining identified degenerating and apoptotic neurons respectively in the DG and CA1 regions of the hippocampus of OGT cKO mice at 2 months of age. Thioflavine S-positive aggregates accumulated in 6-month-old OGT cKO mice. (D) Deficits in cued fear conditioning were observed at 2 months in OGT cKO mice. Deficits in both cued and contextual learning and memory were observed at 4 months. $n = 12$ per group. $*P < 0.05$, $**P < 0.005$, $***P < 0.0005$. Figure from Andrew Wang.

2.4 Few differentially-expressed genes in OGT cKO hippocampi at 3 weeks

To identify global transcriptional changes associated with the behavioral and histological phenotypes, we performed gene expression microarray analyses. At 3 weeks of age, only 10 differentially-expressed (DE) genes were identified in the hippocampus of

OGT cKO mice compared to WT mice, consistent with the minimal extent of OGT ablation and neuronal loss shown by histology (Table 2.1). Complement component 1, q subcomponent, C chain (*C1qc*) and lectin, galactoside-binding, soluble, 3 binding protein (*Lgals3bp*) are immune-response related genes that are upregulated in the OGT cKO mice (1.3-fold and 1.6-fold respectively). A 3.0-fold upregulation of the SLX4 structure-specific endonuclease subunit homolog (*Slx4*), a Holliday junction resolvase critical for homology-directed repair (HDR) in response to DNA interstrand crosslinking, replication, and telomeric extension.[9-11] This increase in SLX4 expression could be priming cells for genotoxic stress response or replication. The upregulation of *C1qc*, *Lgals3bp*, and potentially *Slx4* indicate incipient immune response in the 3-week OGT cKO.

Several genes that are involved in neuronal activity and growth are also DE at 3 weeks in the OGT cKO. Upregulated in OGT cKO (1.5-fold), abhydrolase domain containing 11 (ABHD11, *Abhd11*) is an α/β serine lipid hydrolase that is currently poorly characterized.[12,13] One study implicated ABHD11 in axonal pathfinding and astroglial organization in the forebrain during zebrafish embryogenesis.[14] Furthermore, Abhd11os, a long non-coding RNA (lncRNA), binds to the 3'UTR region of ABHD11 mRNA and upregulates ABHD11 expression has been shown to be neuroprotective in Huntington's disease suggesting that ABHD11 may play a role in HD.[15] Cadherin 6 (*Cdh6*), a protein critical for axonal guidance, is downregulated in the OGT cKO (-1.6-fold), suggesting that there may be impairments in axonal targeting in the OGT cKO mouse.[16] Proline-rich transmembrane protein 1 (SynDIG4, *Prrt1*) associates with extrasynaptic glutamate α-

amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors, but its function is not yet known.[17,18]

Finally, a metabolic and an mRNA turnover regulator are also upregulated in the OGT cKO at 3 weeks. The gene 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3 (PFKFB3, *Pfkfb3*) is upregulated 1.8-fold in the OGT cKO mice. PFKFB3 is a master regulator of glycolysis through its production of 2,6-biphosphate, a major allosteric agonist of 6-phosphofructokinase-1 (PFK1), the enzyme that catalyzes the rate-limiting step of glycolysis.[19] Under excitotoxic conditions, PFKFB3 is stabilized and shunts neurons to glycolysis exacerbating oxidative stress and leading to neurodegeneration.[19,20] Poly(A) binding protein, cytoplasmic 6 (*Pabpc6*), upregulated 1.6-fold in the OGT cKO, is required for mRNA translation initiation and mRNA degradation.[21] Finally, both *1810062O18Rik* and *4930597O21Rik* are RNA transcribed from the antisense strand of DNA, indicating that these are likely to be involved in RNA interference, but the actual functions or targets of these transcripts is currently unknown.[22] In summary, at 3 weeks of age, the OGT cKO shows the upregulation of a few immune response genes and the differential expression of neuronal activity, metabolism, and mRNA processing genes.

Table 2.1 Differentially-expressed genes in the OGT cKO at 3 weeks.

Genes	Description	FC	*P*-values
Pfkfb3	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3	1.8	1.4×10^{-19}
Prrt1	proline-rich transmembrane protein 1	1.6	1.8×10^{-11}
Cdh6	cadherin 6	-1.6	5.8×10^{-9}
Pabpc6	poly(A) binding protein, cytoplasmic 6	1.6	1.0×10^{-7}
Lgals3bp	lectin, galactoside-binding, soluble, 3 binding protein	1.6	4.9×10^{-7}
Abhd11	abhydrolase domain containing 11	1.5	1.9×10^{-5}

Slx4	SLX4 structure-specific endonuclease subunit homolog (*S. cerevisiae*)	3.0	2.6×10^{-5}
C1qc	complement component 1, q subcomponent, C chain	1.3	2.6×10^{-4}
1810062O18Rik	RIKEN cDNA 1810062O18 gene	1.3	5.7×10^{-4}
4930597O21rik	RIKEN cDNA 4930597O21 gene	-1.4	5.7×10^{-4}

Table 2.1 Listed above are the differentially-expressed genes in the OGT cKO mice compared to their WT littermates at 3 weeks of age. FC = fold-change where negative values denote decreases. Reported *P*-values are Bonferroni corrected (*P*-values < 0.001).

2.5 Upregulation of immune response and AD-related genes in the OGT cKO mouse at 2 months

At 2 months of age, we observed a dramatic upregulation of hundreds of genes in the hippocampus of OGT cKO mice (Appendix I). Among the most highly upregulated were glial proliferation and immune response genes such as glial fibrillary acidic protein (*Gfap*, 6.0-fold), complement component 1q (*C1qb*, 4.0-fold), and complement component 3 (*C3*, 2.1-fold) (Appendix I). In addition, both *Lgals3bp* (5.0-fold in 2 month) and *C1qc* (3.5-fold in 2 month) were upregulated in OGT cKO hippocampi at both 3 weeks and 2 months of age. Increased expression of these and other representative genes was confirmed independently by qRT-PCR (Table 2.2). Notably, many of the same genes are upregulated in established AD mouse models that express familial Alzheimer's disease (FAD) mutations in presenilin 1 (*Psen1*), presenilin 2 (*Psen2*), and/or amyloid precursor protein (*App*) that lead to early onset AD (Table 2.2, Figure 2.2).[23,24] In addition, we validated these upregulated genes using qPCR as well. Importantly, these AD mouse models show similar levels of neuroinflammatory gene expression at 12-18 months of age while our mice are only 2 months of age.

Table 2.2 Upregulated genes in the OGT cKO at 2 months and in AD mouse models

Genes	Description	OGT cKO/WT Microarray		OGT cKO/WT qRT-PCR	AD Mouse Model Microarray[1]			AD Mouse Model qRT-PCR[2]	
		FC	P-values[3]	FC[4]	A	B	C	B	SEM
Gfap	glial fibrillary acidic protein	6.0	1.0×10^{-37}	12	1.3	6.3	7.3	4.83	0.64
C1qb	complement component 1 q subcomponent beta polypeptide	4.0	1.0×10^{-37}	6.8	1.6	2.5	2.8	2.00	0.16
Tyrobp	TYRO protein tyrosine kinase binding protein	4.0	1.0×10^{-37}	6.1	2.0	3.6	4.8	2.54	0.06
Cd14	CD14 antigen	3.9	1.0×10^{-37}	5.4	1.4	1.8	2.3	-	-
C1qc	complement component 1, q subcomponent, C chain	3.5	1.0×10^{-37}	6.2	2.0	3.6	4.8	-	-
Gusb	glucuronidase beta	2.7	1.0×10^{-37}	3.6	1.4	2.3	2.4	-	-
B2m	beta-2 microglobulin	2.1	1.8×10^{-29}	4.4	-	-	-	2.03	0.09
C3	complement component 3	2.1	2.6×10^{-21}	8.1	-	-	-	-	-
Vim	vimentin	1.9	5.2×10^{-4}	3.1	-	-	-	1.81	0.14
Man2b1	mannosidase 2, alpha B1	1.7	6.4×10^{-34}	2.3	1.3	1.9	2.1	1.62	0.07

[1] Fold-expression changes of regulated genes in (A) 18-month old APP$^{NLh/NLh}$/PS-1$^{P264L/P264L}$ mice[24]; (B) 12-month-old Tg2576/PS-1$^{P264L/+}$ mice[24]; and (C) 12-month old Tg2576/PS-1$^{P264L/P264L}$ mice[24];

[2] Fold-expression changes of regulated genes in the 17-18-month old Tg2576/PS-1$^{P264L/P264L}$ mice[23]; SEM = standard error of the mean

[3] Microarray P-values were Bonferroni-corrected; FC = fold-change

[4] qRT-PCR P-values were calculated using a two-tailed Student's t-test (P-values < 0.05) ; FC = fold-change
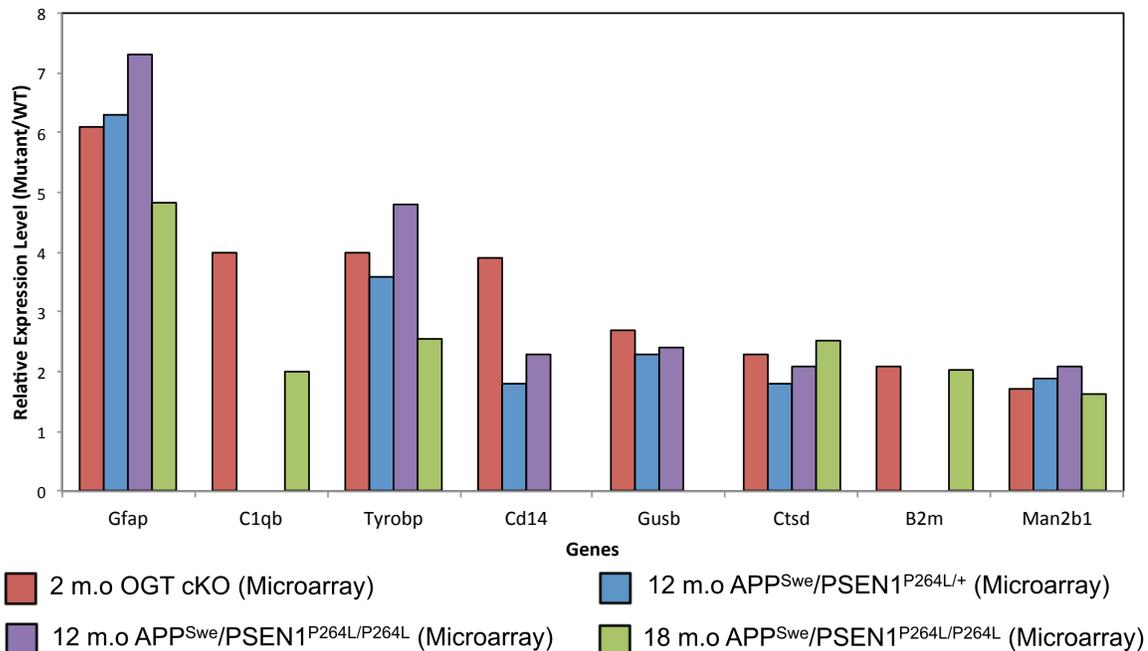
Figure 2.2 Barplot of upregulated genes in OGT cKO and AD mouse models. Here are the expression levels of representative upregulated genes from Table 2.2 (only microarray levels) in the 2 month old (m.o.) OGT cKO, 12 m.o. APP^{Swe}/PSEN1$^{P264L/+}$, 12 m.o. APP^{Swe}/PSEN1^{P264L}, and 18 m.o. APP^{Swe}/PSEN1^{P264L}. These upregulated genes are all involved in immune response.

Interestingly, there is also upregulation of genes that have been shown to affect the susceptibility for late-onset AD in humans, including the microglial genes: triggering receptor expressed on myeloid cells 2, *Trem2* (6.0-fold increase); inositol polyphosphate-5-phosphatase D, *Inpp5d* (2.3-fold); and phospholipase C, γ2, *Plcg2* (1.5-fold).[25-28] Using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) bioinformatics tool, we found the top 3 GO categories were immunity (FDR $< 2.0 \times 10^{-33}$), immune system process (FDR $< 1.4 \times 10^{-29}$), and innate immunity (FDR $< 1.6 \times 10^{-27}$) (Table 2.3). The upregulation of microglial and immune response genes indicates substantial neuroinflammation in the OGT cKO at 2 months of age, confirming the IHC data and revealing elevated GFAP and Iba-1 staining at this age.

Table 2.3 DAVID GO annotation of OGT cKO/WT upregulated genes at 2 months

GO term	#	Gene names	FE
Immunity	73	*Zc3hav1, Ly86, Tirap, Tlr2, Tlr3, Ly9, C1qc, Tlr7, Btk, B2m, Isg20, C1ra, Tmem173, Myd88, Oasl2, Oasl1, Mx2, Pik3cg, Gbp5, Bst2, Inppl1, Herc6, H2-Dmb1, Serping1, Cd84, C1qa, C1qb, Cd86, Lat2, Lrmp, H2-Aa, Eif2ak2, Tnfaip8l2, C3, Csf1, Ifitm3, Gsdmd, Unc93b1, Ifi30, Oas2, Cd74, Naip6, Tap2, Tap1, Naip5, Iigp1, Inpp5d, Dhx58, Csf1r, Havcr2, Tlr12, Irgm1, Cfb, Hck, Tlr13, Axl, Myo1g, Samhd1, H2-Ab1, Trim25, Fcgr1, Psmb8, Lgals9, Psmb9, Ddx58, Ifit2, Trim56, Ifit1, Irf5, Irf7, H2-Eb1, Clec7a, Cd14*	6.3
Immune system process	72	*Zc3hav1, Ly86, Tirap, Tlr2, Tlr3, Ly9, C1qc, Tlr7, Btk, B2m, Isg20, C1ra, Tmem173, Myd88, Oasl2, Oasl1, Mx2, Pik3cg, Gbp5, Bst2, Inppl1, Herc6, H2-Dmb1, Serping1, Cd84, C1qa, C1qb, Cd86, Lat2, Lrmp, H2-Aa, Eif2ak2, Tnfaip8l2, C3, Csf1, Ifitm3, Gsdmd, Unc93b1, Ifi30, Oas2, Cd74, Naip6, Tap2, Tap1, Naip5, Iigp1, Inpp5d, Dhx58, Csf1r, Havcr2, Tlr12, Irgm1, Cfb, Hck, Tlr13, Axl, Myo1g, Samhd1, H2-Ab1, Trim25, Fcgr1, Psmb8, Lgals9, Psmb9, Ddx58, Ifit2, Trim56, Ifit1, Irf5, Irf7, H2-Eb1, Cd14*	5.6
Innate immunity	53	*Zc3hav1, Ly86, Tlr2, Tirap, Tlr3, Ly9, C1qc, Tlr7, Isg20, Btk, C1ra, Tmem173, Myd88, Oasl2, Oasl1, Mx2, Bst2, Herc6, Serping1, C1qa, Cd84, C1qb, Eif2ak2, Tnfaip8l2, C3, Csf1, Ifitm3, Gsdmd, Unc93b1, Oas2, Naip6, Naip5, Iigp1, Dhx58, Csf1r, Tlr12, Havcr2, Irgm1, Cfb, Hck, Tlr13, Axl, Samhd1, Trim25, Fcgr1, Ddx58, Trim56, Ifit2, Ifit1, Irf5, Irf7, Clec7a, Cd14*	7.7
Glycoprotein	225	*Kcne1l, Scpep1, A2m, Adora3, Ltbp3, Atp1b2, Osmr, Fgfrl1, Ptgs1, Aqp4, Cd52, Cd53, Megf10, Cd48, C1ra, Cd44, Apod, Serpine2, Grin2c, Ch25h, Ggta1, F11r, Ptprf, Pdpn, F9, Serping1, Sirpa, C1qa, Sstr5, C1qb, Cd37, Npc2, Ccr5, Serpinf1, Hepacam, Pmp22, Wfdc3, Il1r2, Il1r1, Mfng, Ifi30, Fcgrt, Oas2, Cd72, Cd74, Slc11a1, Lgals3bp, Slc29a3, Cd68, Lamb2, Fgl2, Gcnt1, Spp1, St6gal1, Slamf9, Cfb, Gusb, Tgfbr2, Rhbdf1, Sepn1, Col16a1, Cd63, Fcgr1, Fcgr3, Notch1, Clec7a, Tmem119, Gpr84, Gpr160, Mpeg1, Ly86, Il4i1, Ly9, Cd151, Ednrb, Tnfrsf11a, Scrg1, P4ha3, Angptl, Sepp1,*	2.1

		Il13ra1, Lag3, Icam1, Gpr37ll, Bgn, Adam17, Fmod, C3, Csf1, Gpr65, Cxcl9, Tpcn1, Nagpa, Lect1, Ly6e, Il10ra, Thbs1, Thbs4, Pla2g15, Axl, H2-Ab1, P2ry13, Fcgr2b, Gria2, Abcc3, Slc15a3, Cd14, Slc44a2, Tspan4, Lgmn, Tlr2, Tlr3, Cspg5, Tlr7, Tapbp, Slc7a7, Olfml3, St3gal4, Csf3r, Lgi4, Asph, Il1a, Bst2, Plxnb2, H2-Dmb1, Hcst, Pnpla7, Gns, Cst7, Prcp, Reln, C3ar1, Ccl2, Abca9, Unc93b1, Lrig1, Itgb5, Itgb2, Abca1, Trf, Itgam, Timp1, Angptl6, P2ry6, Itgax, Ttyh2, P2ry1, Entpd4, Mfap3l, Entpd2, Csf1r, Havcr2, Tlr12, Mrc2, Tlr13, Tst, Lamp2, Gpr34, Itga6, Slc7a3, Cd274, H2-Eb1, Slc13a3, Scara3, Mertk, Igdcc4, Gm2a, Lrrc8a, Hexa, Hexb, Bcan, Cxadr, Sdc4, Sdc3, Glb1, Slc1a4, Ccrl2, Tor3a, Slc1a3, Hpse, Csf2rb, Itih3, Man2b1, Csf2ra, Slc43a3, Taar3, Lair1, Ctsz, Spint1, Ctss, Dnase2a, Cd84, Cd83, Ctsl, Serpina3n, Cd86, Grn, H2-Aa, Ctsd, Ctsb, Trem2, Pros1, Ctsh, Vim, Cd24a, Scarf2, Vcam1, Tnfrsf1a, Cd9, Tnfrsf1b, Glycam1, Smoc1, Gpnmb, Fgfbp1, Selplg, Gba, Tmc6, Ptprc, Hpn, Sun2, Sparc, Il6ra, Rgs20, Liph, Slc14a1	
Membrane	339	*Kcne1l, S100a6, Aif1, Atp1b2, Osmr, Ptgs1, Fgfrl1, Aqp4, Cd52, Cd53, Itsn1, Megf10, Rnf213, Btk, B2m, Cd48, Hmha1, Bak1, Cd44, Serpine2, Grin2c, Wnk4, Ch25h, Oasl1, Rapgef3, Ggta1, Pik3cg, Sgpl1, F11r, Gbp6, Gbp5, Ptprf, Pdpn, Ncf1, Ncf4, Pld4, Ptbp1, Ifi47, Gem, Lpcat2, Sirpa, Sstr5, Cd37, Aaas, Hepacam, Ccr5, Parp14, Cx3cr1, Tmem184b, Rab13, Pmp22, Eif2ak2, Il1r2, Mfng, Il1r1, Rtp4, Gnai2, Gsdmd, Fcgrt, Oas2, Cd72, Kcnj2, Sfxn5, Cd74, Slc11a1, Cd68, Slc29a3, Lgals3bp, Rac2, Cklf, Fcer1g, Gcnt1, Tyrobp, Blnk, St6gal1, Slamf9, Gusb, Rhbdf1, Tgfbr2, Nckap1l, Sepn1, Cd63, Fcgr1, Fcgr3, Atp13a4, Notch1, Gngt2, Rgs1, Parp9, Tmem119, Gpr84, Gfap, Slc20a1, Gpr160, Mpeg1, Fermt3, Kcnj10, Ly9, Cd151, Hvcn1, Ednrb, Tnfrsf11a, Pik3ap1, Fam129b, Il13ra1, Lag3, Icam1, Fmnl3, Inppl1, Clic1, Flnc, Arrdc3, Fcrls, Gpr37ll, Ripk1, Adam17, Lrmp, Susd3, Hsd3b7, Csf1, Gpr65, Hk2, Mfsd1, Apbb1ip, Tpcn1, Sft2d2, Glipr2, Nagpa, Lpxn, Lect1, Igtp, Dapp1, Ly6e, Klc1, Plin4, Il10ra, Rasa4, Irgm1, Pla2g15, Tbxas1, Irgm2, Plek, Baiap2l2, Rrbp1, Axl, Cyp4f14, Elavl1, H2-Ab1, Cyp4V3, Atp1a2, Gjb6, Capn2, Appl2, Itpr2, P2ry13, Rab31, Fcgr2b, Gria2, Plscr2, Abcc3, Cmtm7, Cmtm3, Slc15a3, Cd14, Tmem176b, Cmtm6, Tmem176a, Tspo, Gna15, Slc44a2, Tspan4, Tlr2, Tirap, Tlr3, Cspg5, Tlr7, Iqgap1, Slc7a7, Vcl, Tapbp, St3gal4, Csf3r, Asph, Scamp2, Myo6, Bst2, Plxnb2, Tor1aip1, H2-Dmb1, Pde4d, Psd2, Hcst, Pnpla7, Igsf6, Lat2, Plce1, Gbp3, Spata13, Gbp2, Mvp, Fxyd1, C3ar1, Gal3st4, Cav1, Abca9, Mcl1, Snx5, Ifitm3, Stk10, Phka1, Unc93b1, Itgb5, Lrig1, Itgb2, Abca1, Trf, Itgam, P2ry6, Dock1, Laptm5, Itgax, Ttyh2, Rasal3, P2ry1, Iigp1, Mfap3l, Entpd2, Csf1r, Hist1h4h, Arhgdib, Havcr2, Srebf1, Tlr12, Tlr13, Mrc2, Myo1g, Dock8, Reep3, Ddx58, Myo10, Lamp2, Gpr34, Slc7a3, Itga6, Ifi27l2a, H2-Eb1, Cd274, Plcg2, Slc13a3, Scara3, Mertk, Rhoj, Igdcc4, Tln1, Lrrc8a, Hexa, Hexb, Cxadr, Sdc4, Rhou, Syngr2, Sdc3, Ccrl2, Slc1a4, Cdc42, Tmem173, Slc1a3, Plekhb1, Hpse, Hmox1, Necap2, Csf2rb, Rhoc, Msn, Yap1, Csf2ra, Rhog, Ahnak, Slc43a3, Taar3, Lair1, Noxo1, Ms4a6c, Ms4a6d, Adipor2, Ctss, Slc9a3r1, Slc7a11, Vasp, Cd84, Stom, Cd83, Cd86, H2-Aa, Trem2, Cyth4, Cd24a, Scarf2, Vcam1, Cd9, Tnfrsf1a, Tnfrsf1b, Glycam1, Sh3glb1, Map3k1, Tap2, Shisa5, Tap1, Plcd4, Pik3r5, Inpp5d, Slc25a45, Appbp2, Gpnmb, Fgfbp1, Selplg, Inpp5a, Gba, Slc39a1, Snx20, Ehd4, Guca1a, Ptprc, Tmc6, Hpn, Gnao1, H2-M3, Hck, Sun2, Anxa5, Anxa3, Anxa2, Il6ra, Cyba, Rgs20, Itpripl2, Liph, Bmpr1b, Slc14a1*	1.5
Disulfide bond	183	*A2m, Adora3, Ltbp3, Atp1b2, Osmr, Lgmn, Fgfrl1, Ptgs1, Tirap, Tlr2, Tlr3, Cspg5, Megf10, C1qc, Tapbp, Cxcl10, Slc7a7, B2m, Cd48, C1ra, Olfml3, Isg15, Cd44, Apod, St3gal4, Csf3r, Asph, F11r, Bst2, Ptprf, Plxnb2, F9, H2-Dmb1, Serping1, Tcn2, Sirpa, Hcst, Igsf6, C1qa, Sstr5, C1qb, Npc2, Ccr5,*	2.0

64

		Hepacam, Cst7, Cx3cr1, Prcp, Reln, Wfdc3, Il1r2, C3ar1, Il1r1, Mfng, Ccl3, Ccl2, Gsdmd, Lrig1, Ifi30, Chchd2, Itgb5, Itgb2, Fcgrt, Abca1, Cd72, Ccl5, Ccl4, Trf, Itgam, Cd74, Timp1, Ccl6, Angptl6, Cd68, Lgals3bp, P2ry6, Lamb2, Itgax, P2ry1, Fcer1g, Fgl2, Entpd4, Mfap3l, Entpd2, Gcnt1, Csf1r, Tyrobp, Havcr2, St6gal1, Cfb, Slamf9, Tgfbr2, Mrc2, Fcgr1, Fcgr3, Notch1, Lamp2, Gpr34, Itga6, H2-Eb1, Cd274, Clec7a, Mertk, Igdcc4, Gm2a, Hexa, Ly86, Hexb, Bcan, Il4i1, Ly9, Cxadr, Ccrl2, Ednrb, Tnfrsf11a, Ang, Csf2rb, Angptl, Man2b1, Il13ra1, Lag3, Taar3, Icam1, Lair1, Ctsz, Spint1, Ctss, Clic1, Fcrls, Gpr37ll, Dnase2a, Slc7a11, Cd84, Ctsl, Cd83, Cd86, Bgn, Grn, Ctsd, H2-Aa, Adam17, Ctsb, Susd3, Trem2, Ctsh, Pros1, Fmod, Cxcl5, C3, Csf1, Ndp, Endou, Cxcl9, Gpr65, Scarf2, Vcam1, Cd9, Tnfrsf1a, Nagpa, Tnfrsf1b, Lect1, Ly6e, Il10ra, Smoc1, Thbs1, Selplg, Fgfbp1, Thbs4, Gba, Hpn, Pla2g15, H2-M3, Axl, H2-Ab1, Sparc, Il6ra, P2ry13, Ccl12, Fcgr2b, Gria2, Liph, Bmpr1b, Cd14, Igfbp5	
Extracellular exosome	174	*S100a4, Scpep1, S100a6, Tspo, A2m, Slc44a2, Ltbp3, Lgmn, Ptgs1, Cd53, Aldh1l2, C1qc, Iqgap1, Vcl, B2m, Cd48, C1ra, Cd44, Apod, St3gal4, Rapgef3, Ddah1, F11r, Gbp6, Myo6, Scamp2, Bst2, Ptprf, Plxnb2, Ptbp1, F9, Serping1, Tcn2, Sirpa, Gltp, C1qa, Gns, Arpc1b, C1qb, Lat2, Renbp, Cd37, Npc2, Serpinf1, Prcp, Rab13, Vsig4, Mvp, Gal3st4, Naglu, Gnai2, Stk10, Ifitm3, Apoc1, Itgb5, Itgb2, Tagln2, Trf, Itgam, Cd74, Timp1, Angptl6, Lgals3bp, Lamb2, Rac2, Rasal3, Fgl2, Entpd2, Hist1h4h, Arhgdib, Spp1, Havcr2, Ptpn6, St6gal1, S100a16, Cfb, Lgals1, Gusb, Ephx2, Myo1g, Fcgr4, S100a10, Nckap1l, Cd63, Psmb8, S100a13, Psmb9, Fcgr3, Tst, Lamp2, H2-Eb1, Plcg2, Cd274, Slc13a3, Rhoj, Tln1, Gm2a, Hexa, Fermt3, Hexb, Sdc4, Syngr2, Glb1, Slc1a4, Tor3a, Cdc42, Ang, Acot11, Tubb6, Rhoc, Angptl, Msn, Sepp1, Fam129b, Itih3, Cdk5rap2, Man2b1, Rhog, Ahnak, Lair1, Icam1, Ctsz, Padi2, Spint1, Clic1, Slc9a3r1, Was, Vasp, Dnase2a, Stom, Cd84, Mtmr11, Ctsl, Serpina3n, Cd86, Bgn, Grn, Ctsd, Ctsb, Ctsh, Pros1, Carhsp1, C3, Csf1, Vim, Vcam1, Aldh1a1, Glipr2, Sft2d2, Cd9, Sh3glb1, Shisa5, Thbs1, Inpp5a, Thbs4, Ehd4, Gba, Bcas1, Ptprc, Tmc6, Hpn, Pla2g15, Axl, Cotl1, Capn2, Anxa5, Appl2, Anxa4, Anxa3, Anxa2, Capg, Hpgd, Cmtm6, Cd14*	2.1
Cell surface	72	*Tln1, Lrrc8a, Tlr2, Tlr3, Cspg5, Ly9, Cd53, Sdc4, Cxadr, Sdc3, Slc1a4, Tnfrsf11a, Slc1a3, Cd44, Msn, Il1a, Icam1, Bst2, Plxnb2, Ctss, Slc7a11, Hcst, Cd86, Bgn, Ccr5, Adam17, Ctsb, Siglech, Cav1, Il1r1, Ifitm3, Ndp, Itgb5, Itgb2, Abca1, Cd24a, Itgam, Cd74, Trf, Vcam1, Slc11a1, Cd9, Tnfrsf1a, Itgax, P2ry1, Fcer1g, Pcsk6, Thbs1, Fgfbp1, Entpd2, Csf1r, Tyrobp, Havcr2, Ptprc, Hpn, Lgals1, Mrc2, Tgfbr2, Axl, Fcgr4, H2-Ab1, Sparc, Cd63, Anxa4, Anxa2, Il6ra, Notch1, Itga6, Fcgr2b, Gria2, Cd274, Cd14*	3.6
Phospho-protein	325	*S100a6, Adora3, Aif1, Pdlim4, Aqp4, Amotl1, Itsn1, Rnf213, Btk, Hmha1, C1ra, Fli1, Myd88, Cd44, Grin2c, Wnk4, Rapgef3, Ccna2, Ddah1, Pik3cg, Sgpl1, F11r, Ptprf, Suclg2, Ncf1, Ncf4, Ptbp1, F9, Gem, Sirpa, Aaas, Hepacam, Serpinf1, Ccr5, Parp12, Parp14, Cx3cr1, Tmem184b, Rab13, Eif2ak2, Il1r1, Gsdmd, Fcgrt, Cd72, Tagln2, Cd74, Tal1, Slc29a3, Lamb2, Fcer1g, Tyrobp, Blnk, Spp1, St6gal1, Ikzf1, Ptpn18, Lgals1, Rhbdf1, Tgfbr2, Fcgr1, S100a13, Tnni2, Notch1, Parp9, Tmod3, Clec7a, Tmem119, Sash3, Gpr84, Gfap, Zcchc24, Slc20a1, Fermt3, Fastk, Kcnj10, Ttc28, Ly9, Nfkb2, Hvcn1, Cdt1, Ednrb, Tnfrsf11a, Pbxip1, Tubb6, Epsti1, Pik3ap1, Sepp1, Fam129b, Zfp36, Batf3, Srpk3, Fmnl3, Hist1h1c, Inpp1l, Clic1, Flnc, Was, Gpr37ll, Dok1, Ripk1, Adam17, Lrmp, Map3k14, Phyhd1, Triobp, Carhsp1, Nmi, C3, Klc4, Mfsd1, Apbb1ip, Zfp36ll, Sft2d2, Plekhg2, Lpxn, Dapp1, Klc1, Plin4, Bcl3, Fyb, Bcas1, Irgm1, Plek, Baiap2l2, Rrbp1, Hcls1, Zfp703, Axl, Elavll1, Trim25, Atp1a2, Samd4, Itpr2, Stat2, Rab31, Fcgr2b, Gria2,*	1.5

		Rps6ka1, Plscr2, Capg, Rassf2, Abcc3, Sh3rf2, Slc15a3, Tmem176b, Cmtm6, Tmem176a, Slc44a2, Prc1, Zc3hav1, Tirap, Tlr3, Cspg5, Aldh1l2, Iqgap1, Slc7a7, Kank2, Vcl, Trim47, Eif4ebp1, Aspg, Rbms2, Asph, Dap, Il1a, Scamp2, Myo6, Socs3, Plxnb2, Dtx3l, Rela, Tor1aip1, Mlxipl, Pde4d, Psd2, Foxn3, Ddit3, Hcst, Pnpla7, Gns, Tns3, Arpc1b, Lat2, Renbp, Plce1, Hspb6, Timeless, Nfe2l2, Spata13, Mvp, Fxyd1, C3ar1, Cav1, Mcl1, Snx5, Ifitm3, Stk10, Phka1, Unc93b1, Rabgap1l, Itgb5, Itgb2, Abca1, Trf, Timp1, Spc25, Dock1, Laptm5, Gmip, Ttyh2, Rasal3, Mfap3l, Runx1, Csf1r, Hist1h4h, Arhgdib, Havcr2, Srebf1, Ptpn6, Phactr4, Mrc2, Ak3, Ephx2, Tead1, Myo1f, Dock8, Reep3, Vav1, Ddx58, Tst, Atf5, Trim56, Myo10, Itga6, Slc7a3, Plcg2, Mertk, Tprn, Kif23, Igdcc4, Tln1, Lrrc8a, Bcan, Zcwpw1, Arhgap18, Cxadr, Skap2, Rhou, Syngr2, Sdc3, Fubp1, Slc1a4, Phc3, Cdc42, Plcb3, Heatr5a, Tmem173, Slc1a3, Hmox1, Casp8, Acot11, Necap2, Csf2rb, Msn, Yap1, Cdk5rap2, Lrrfip1, Rhog, Lair1, Ms4a6d, Slc9a3r1, Slc7a11, Cdk2, Vasp, Arhgap25, Cd84, Arhgap30, Stom, Ssfa2, Gpsm3, Vim, Scarf2, Patl1, Aldh1a1, Rgs10, Tnfrsf1b, Glycam1, Sh3glb1, Tsc22d4, Map3k1, Pik3r5, Zfp521, Inpp5d, Gpnmb, Dcx, Selplg, Nfatc1, Rcsd1, Snx20, Ehd4, Inf2, Cebpa, Ptprc, Tmc6, Hck, Sun2, Samhd1, Samd14, Anxa5, Samsn1, Cotl1, Anxa4, Anxa3, Anxa2, Irf9, Cyba, Ifit1, Rgs20, Irf5, Dbp, Irf7, Itpripl2, Slc14a1, Igfbp5, Sh3bp2	
Inflammatory response	48	*C3ar1, Ccl3, Ccl2, Cxcl5, Aif1, C3, Csf1, Ly86, Ptgs1, Gsdmd, Tirap, Cxcl9, Tlr2, Tlr3, Nfkb2, Ccl5, Ccl4, Tlr7, Ccl6, Cxcl10, Ccrl2, Naip6, Tnfrsf1a, Slc11a1, Tnfrsf1b, Myd88, Naip5, Thbs1, Il1a, Csf1r, Spp1, Pik3cg, Havcr2, Tlr12, Gbp5, Ncf1, Hck, Rela, Tlr13, Axl, Ephx2, Ccl12, Cyba, Ccr5, Clec7a, Nfe2l2, Bmpr1b, Cd14*	4.1
Lysosome	37	*Gm2a, Lgmn, Hexa, Ifitm3, Hexb, Unc93b1, Ifi30, Il4i1, Tlr7, Tpcn1, Glb1, Slc29a3, Cd68, Laptm5, Hpse, Man2b1, Gba, Irgm1, Ctsz, Pla2g15, Hck, Gusb, H2-Dmb1, Ctss, Arrdc3, Cd63, Dnase2a, Pnpla7, Gns, Ctsl, Lamp2, Npc2, Prcp, Ctsd, Ctsb, Slc15a3, Ctsh*	5.2
Signal peptide	183	*Scpep1, A2m, Ltbp3, Osmr, Lgmn, Fgfrl1, Ptgs1, Tlr2, Tlr3, Cd52, Cspg5, Megf10, C1qc, Tlr7, Tapbp, Cxcl10, B2m, Cd48, C1ra, Olfml3, Cd44, Apod, Serpine2, Grin2c, Csf3r, Lgi4, Htra3, F11r, Ucma, Ptprf, Pdpn, F9, H2-Dmb1, Serping1, Tcn2, Sirpa, Hcst, Igsf6, C1qa, Gns, C1qb, Npc2, Serpinf1, Hepacam, Cst7, Prcp, Reln, Fxyd1, Il1r2, Il1r1, Ccl3, Ccl2, Apoc1, Lrig1, Itgb5, Loc68395, Itgb2, Fcgrt, Ccl5, Ccl4, Trf, Itgam, Timp1, Ccl6, Angptl6, Cd68, Lgals3bp, Lamb2, Itgax, Fcer1g, Fgl2, Mfap3l, Csf1r, Spp1, Tyrobp, Havcr2, Tlr12, Cfb, Slamf9, Gusb, Tgfbr2, Mrc2, Tlr13, Col16a1, Fcgr1, Fcgr3, Notch1, Lamp2, Itga6, H2-Eb1, Cd274, Mertk, Tmem119, Igdcc4, Gm2a, Mpeg1, Hexa, Ly86, Hexb, Bcan, Zcwpw1, Il4i1, Ly9, Cxadr, Sdc4, Glb1, Sdc3, Tor3a, Ednrb, Tnfrsf11a, Hpse, Ang, Scrg1, P4ha3, Csf2rb, Angpt1, Sepp1, Itih3, Man2b1, Il13ra1, Csf2ra, Lag3, Icam1, Lair1, Ctsz, Spint1, Ctss, Fcrls, Gpr37l1, Dnase2a, Cd84, Ctsl, Cd83, Cd86, Serpina3n, Bgn, Grn, Ctsd, H2-Aa, Adam17, Ctsb, Trem2, Ctsh, Pros1, Fmod, Cxcl5, C3, Csf1, Ndp, Cxcl9, Cd24a, Scarf2, Vcam1, Tnfrsf1a, Nagpa, Tnfrsf1b, Glycam1, Ly6e, Il10ra, Smoc1, Shisa5, Gpnmb, Thbs1, 2810459m11rik, Fgbp1, Selplg, Thbs4, Gba, Ptprc, Pla2g15, H2-M3, Axl, H2-Ab1, Sparc, Il6ra, Ccl12, Fcgr2b, Gria2, Itpripl2, Liph, Bmpr1b, Cd14, Igfbp5*	1.7
Disulfide bond	156	*A2m, Adora3, Ltbp3, Atp1b2, Osmr, Ptgs1, Fgfrl1, Tlr2, Tlr3, Cspg5, Megf10, C1qc, Tapbp, Cxcl10, B2m, Cd48, C1ra, Olfml3, Cd44, Apod, St3gal4, Csf3r, F11r, Ptprf, F9, Serping1, Tcn2, Sirpa, Igsf6, C1qa, Sstr5, C1qb, Npc2, Hepacam, Ccr5, Cst7, Cx3cr1, Reln, Il1r2, C3ar1, Il1r1, Ccl3, Ccl2, Lrig1, Itgb5, Loc68395, Fcgrt, Itgb2, Abca1, Ccl5, Cd72, Ccl4, Trf, Itgam, Cd74, Ccl6, Timp1, Angptl6, Lgals3bp, P2ry6, Cd68, Lamb2, Itgax, P2ry1, Fcer1g, Fgl2, Mfap3l, Gcnt1, Csf1r, Havcr2, St6gal1, Cfb, Slamf9,*	1.9

66

		Tgfbr2, Mrc2, Fcgr1, Fcgr3, Notch1, Lamp2, Gpr34, Itga6, H2-Eb1, Cd274, Clec7a, Mertk, Igdcc4, Gm2a, Hexa, Hexb, Bcan, Il4i1, Ly9, Cxadr, Ccrl2, Ednrb, Tnfrsf11a, Ang, Csf2rb, Angpt1, Man2b1, Il13ra1, Lag3, Taar3, Lair1, Icam1, Ctsz, Spint1, Ctss, Clic1, Gpr37ll, Dnase2a, Cd84, Ctsl, Cd83, Cd86, Bgn, Ctsd, H2-Aa, Adam17, Ctsb, Susd3, Trem2, Ctsh, Pros1, Fmod, Cxcl5, C3, Csf1, Ndp, Cxcl9, Gpr65, Scarf2, Vcam1, Tnfrsf1a, Nagpa, Tnfrsf1b, Lect1, Ly6e, Smoc1, Il10ra, Thbs1, Selplg, Fgfbp1, Thbs4, Gba, Hpn, Axl, H2-Ab1, Sparc, Il6ra, Ccl12, P2ry13, Fcgr2b, Liph, Cd14, Igfbp5	
Defense response to virus	32	*Apobec1, Zc3hav1, Ifitm3, Cxcl9, Unc93b1, Tlr3, Oas2, Tlr7, Cxcl10, Isg20, Tmem173, Itgax, Isg15, Oasl2, Ddx60, Oasl1, Mx2, Dhx58, Ptprc, Bst2, Trim34a, Samhd1, Trim25, Stat2, Ddx58, Ifit2, Trim56, Ifit1, Cd86, Irf5, Oas1c, Eif2ak2*	5.7
Osteoclast differentiation	30	*Il1r1, Csf1, Nfkb2, Btk, Tnfrsf1a, Tnfrsf11a, Pik3r5, Il1a, Csf1r, Blnk, Nfatc1, Tyrobp, Pik3cg, Ncf1, Socs3, Ncf4, Rela, Tgfbr2, Socs1, Fcgr4, Fcgr1, Sirpa, Fcgr3, Stat2, Irf9, Cyba, Fcgr2b, Plcg2, Trem2, Map3k14*	5.6

Table 2.3 contains the top 20 DAVID functional gene ontology annotations of the OGT cKO/WT upregulated differentially-expressed genes. Redundant GO categories were removed. FE = fold-enrichment; FDR <6.2x10[-11]

2.6 Synaptic genes, OGT, and OGA are not differentially-expressed in the OGT cKO mouse hippocampus at 2 months

At 2 months of age, genes that showed little alteration in the AD mouse model Tg2576/PS-1$^{P264L/P264L}$, such as the synaptic genes growth-associated protein 43 (*Gap43*), syntaxin (*Synt*), synaptophysin I (*Syp*), and synaptotagmin I (*Syt1*), were also not differentially expressed in OGT cKO hippocampi (Table 2.4) suggesting that there may not be synaptic impairments across all neurons at this point.[23] Importantly, we did not see *Ogt* or *Mgea5* (OGA) differentially expressed in the OGT cKO at 2 months despite the fact that *Ogt* is being knocked out progressively. This is likely due to the fact that our CaMKIIα promoter restricts expression of Cre recombinase to excitatory neurons in the forebrain while OGT is still present in other cell types such as inhibitory neurons or glia, which are proliferating in 2-month-old OGT cKO mice. Overall, the similarities in the transcriptional profiles between OGT cKO mice and AD mouse models suggest that they may share common underlying mechanisms of neurodegeneration.

Table 2.4 Non-differentially expressed genes in 2-month-old OGT cKO hippocampus

Genes	Description	OGT cKO/WT Microarray		OGT cKO/WT qRT-PCR		AD Mouse Model qRT-PCR[23]	
		FC	P-values	FC	P-values	FC	SEM
Synt	syntaxin	1.0	5.7×10^{-1}	-1.1	1.0	-1.1	0.07
Nr4a2	nuclear receptor subfamily 4, group A, member 2	1.2	1.3×10^{-1}	-1.1	0.89	-	-
Syt1	synaptotagmin I	-1.1	8.1×10^{-3}	-1.1	1.0	1.1	0.19
Mgea5	meningioma expressed antigen 5 (hyaluronidase) (OGA)	-1.2	1.8×10^{-2}	-1.1	1.0	-	-
Ogt	O-linked N-acetylglucosamine (GlcNAc) transferase 3	-1.3	3.7×10^{-3}	-1.1	1.0		
Syn1	synapsin I	1.1	7.8×10^{-2}	-1.1	1.0	1.0	0.07
Syp	synaptophysin I	-1.1	6.5×10^{-2}	1.0	1.0	1.0	0.05
Nr2b	glutamate receptor, ionotropic, NMDA2b (epsilon 2) (Grin2b)	1.2	2.8×10^{-4}	1.0	1.0	-1.2	0.05
Gap43	growth-associated protein 43	1.0	9.7×10^{-1}	1.4	1.0	1.0	0.01

Table 2.4 shows the genes that were not differentially expressed in 2-month-old OGT cKO hippocampi. Fold-expression changes of regulated genes in the 17-18-month old Tg2576/PS-1[P264L/P264L] mouse.[23] FC = fold-change where negative fold-change values denote decreases; SEM = standard error of the mean; microarray *P*-values are Bonferroni-corrected *P*-values; qRT-PCR *P*-values were calculated using a two-tailed Student's *t*-test.

2.7 Cholesterol and lipid biosynthesis genes are downregulated in OGT cKO hippocampi at 2 months

Based on the DAVID analysis, the downregulated genes in the OGT cKO mice at 2 months were enriched for genes involved in metabolism including cholesterol (FDR$<3.6 \times 10^{-7}$) and lipid biosynthesis (FDR$<1.3 \times 10^{-2}$) (Table 2.5, Appendix II). Several enzymes involved in the steroid biosynthetic pathway are downregulated in the OGT cKO mouse in addition to genes involved in glycolysis, the mevalonate pathway, and terpenoid biosynthesis (Table 2.5, Figures 2.3, 2.4). Cholesterol biosynthesis has been implicated in AD since one of the major late-onset AD mutations occurs in apolipoprotein E (*Apoe*), a critical regulator of cholesterol metabolism.[29,30] In addition, cholesterol turnover is required for proper LTM formation in the hippocampus.[31,32] Dysregulation of cholesterol has been shown to be a hallmark of several neurodegenerative diseases and psychiatric illnesses.[33-35] In addition to downregulation of

the enzymes responsible for cholesterol synthesis, we observed decreases in the expression of sterol regulatory element-binding protein 2 (SREBP2, *Srebf2,* -1.4-fold), the master TF regulator of cholesterol biosynthesis, and insulin induced gene 1 (*Insig1*, -1.5-fold), which prevents the proteolytic activation of the SREBPs.[35,36] OGT has been shown to be a critical regulator of a myriad of metabolic pathways including glycolysis, gluconeogenesis, glycogen synthesis, and lipogenesis.[37-40] In particular, OGT is known to *O*-GlcNAcylated nuclear liver X receptor (LXR), which leads to enhanced expression of SREBP-1c, the major TF regulator of lipogenesis.[37] Highlighting the interconnected nature of metabolic networks, LXR-mediated transcription of SREBP-1c requires a sterol ligand produced downstream of SREBP2 activity.[41,42] Therefore, SREBP2 downregulation will likely interfere with SREBP-1c-mediated lipogenesis. In these ways, the removal of OGT perturbs cholesterol and lipid biosynthesis by interfering with transcriptional regulators, biosynthetic enzymes, and post-transcriptional regulatory mechanisms. A previous study demonstrated that OGT knockout in *C. elegans* impaired insulin signaling, which resulted in a 3-fold reduction in lipid storage and triglyceride levels and a reduction in sterol esters accompanied by a 3-fold increase in trehalose and glycogen storage.[43] The connection between OGT and cholesterol biosynthesis contrasts a previous mammalian study that showed that inhibition of OGT failed to affect cholesterol synthesis, albeit on a significantly shorter time scale (30 min inhibitor treatment) in a different cell type (HepG2 liver carcinoma cells).[44] Overall, our study supports a critical role of OGT in cholesterol and lipid homeostasis.

Table 2.5 DAVID GO annotation of OGT cKO/WT downregulated genes at 2 months

GO term	#	Gene names	FE
Cholesterol biosynthetic process	9	*Tm7sf2, Cyp51, Mvd, Dhcr7, Insig1, Fdps, Idi1, Hsd17b7, Fdft1*	32

Steroid metabolism	11	*Tm7sf2, Cyp51, Sult4a1, Mvd, 0610007P14Rik, Dhcr7, Insig1, Fdps, Idi1, Fdft1, Srebf2*	17
Sterol biosynthesis	8	*Tm7sf2, Cyp51, Mvd, 0610007P14Rik, Dhcr7, Fdps, Idi1, Fdft1*	36
Biosynthesis of antibiotics	13	*Cyp51, Tm7sf2, Mvd, Fdps, Pfkp, Hk1, Pgam2, Acss2, Acat2, Gpi1, Fdft1, Idi1, Hsd17b7*	7.7
Lipid biosynthesis	10	*Tm7sf2, Cyp51, Mvd, 0610007P14Rik, Dhcr7, Fdps, Idi1, Hsd17b7, Fdft1, Fam57b*	7.3
Metabolic pathways	24	*Tm7sf2, Cyp51, Adss, Mvd, B3galt4, Pfkp, Fdps, Pgam2, Hk1, Acss2, Acat2, Gpi1, Acot5, Fdft1, Acot3, Tdo2, Dhcr7, Gls, B3galnt1, Inpp4b, Fuk, Idi1, Galnt13, Hsd17b7*	2.4

Table 2.5 contains the top DAVID functional gene ontology annotations of the OGT cKO/WT downregulated differentially-expressed genes at 2 months of age. Redundant and non-significant GO terms were removed from the list. Redundant GO terms were removed from the table. FE = fold-enrichment; FDR $<5.0 \times 10^{-2}$.
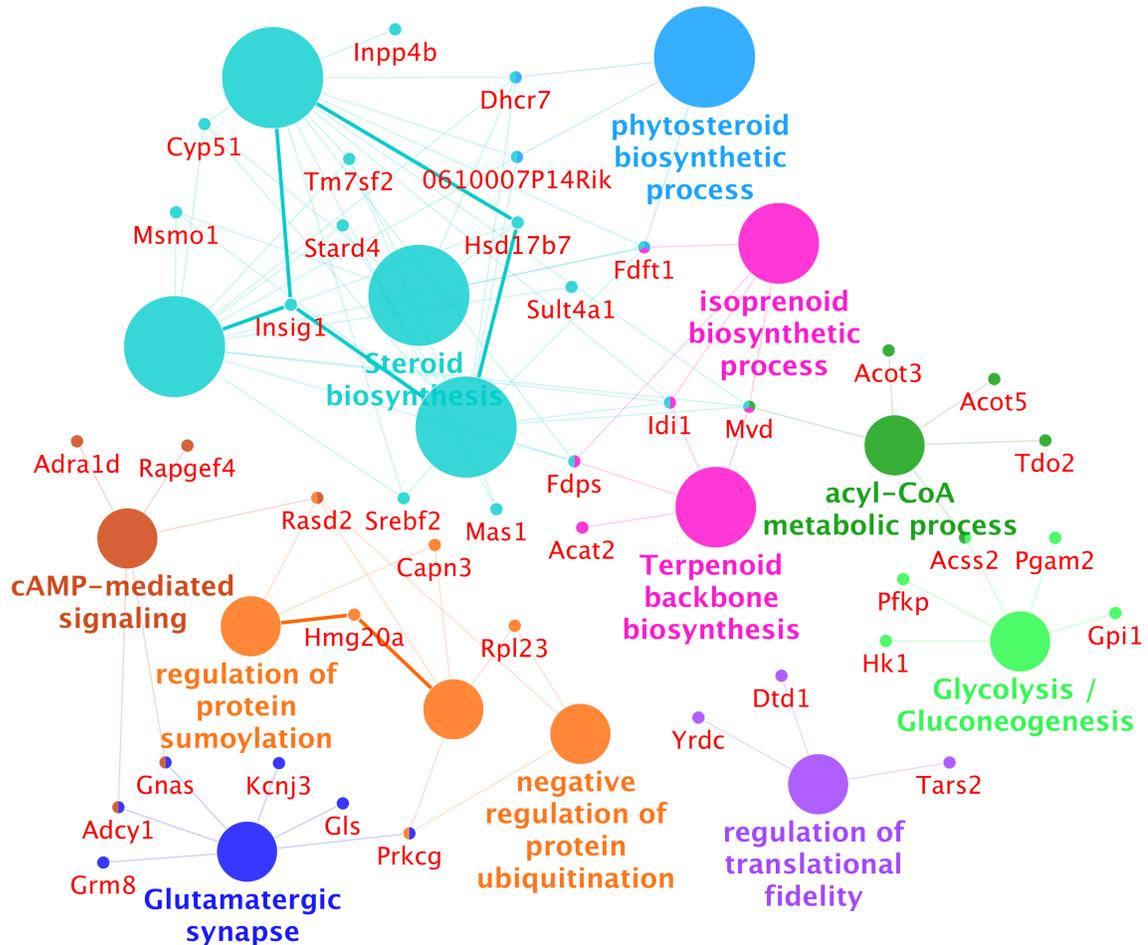
Figure 2.3 Cytoscape gene ontology annotations for OGT cKO 2 m.o. downregulated genes are enriched for metabolic genes. The categories include regulation of transcriptional fidelity, glycolysis/gluconeogenesis, glutamatergic synapse, acyl-CoA metabolic process, phytosteroid biosynthetic process, cAMP-mediated signaling, terpenoid backbone biosynthesis, isoprenoid biosynthetic process, regulation of protein sumoylation, and steroid biosynthesis. The genes that belong to and are shared by these gene ontology categories are also shown. The gene annotations were performed using the GO Biological Processes and KEGG Pathways database with the ClueGO extension in Cytoscape. The significance for all listed categories are Bonferroni step down corrected $P < 0.0042$.
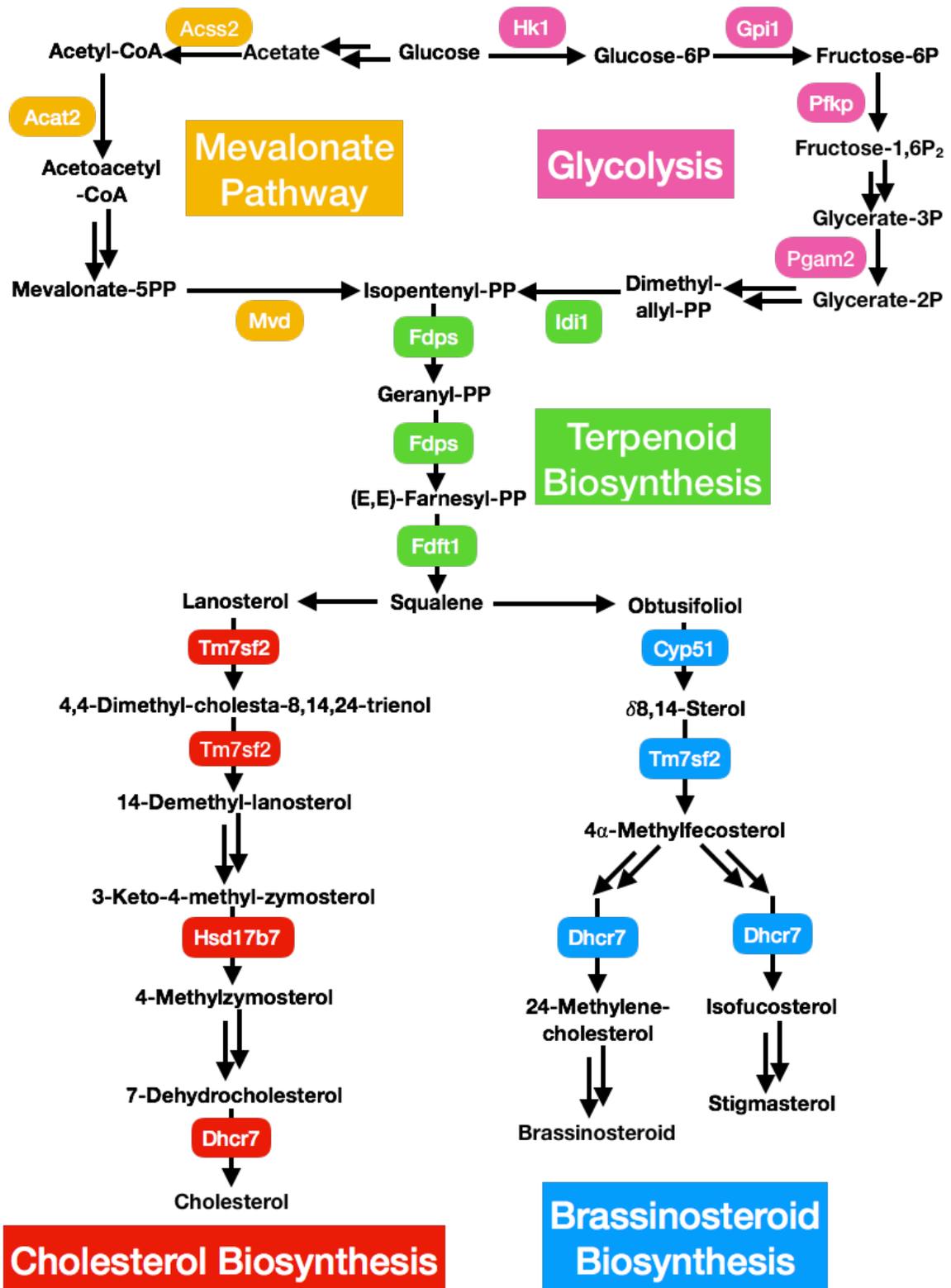
Figure 2.4 OGT cKO downregulated genes are involved in metabolism. Genes involved in glycolysis, mevalonate pathway, terpenoid biosynthesis, brassinosteroid biosynthesis, and cholesterol biosynthesis are

downregulated in the OGT cKO. All the genes listed are significantly downregulated (Bonferroni *P*-value < 0.001) and shown in the context of their substrates, products, and overall metabolic pathways.

Finally, glutamatergic synaptic genes including guanine nucleotide binding protein, α stimulating (*Gnas*, -1.5-fold); adenylate cyclase 1 (*Adcy1*, -1.3-fold); glutamate receptor, metabotropic 8 (*Grm8*, -1.5-fold); glutaminase (*Gls*, -1.5-fold); and the potassium channel (*Kcnj3*, -1.5-fold) were downregulated consistent with progressive knockout and degeneration of excitatory glutamatergic neurons. In summary, the genes downregulated in the OGT cKO at 2 months indicate impaired cholesterol and lipid biosynthesis and glutamatergic synaptic signaling.

2.8 OGT knockout is highly correlated with an immune response gene network

To determine whether entire networks of genes were deregulated in OGT cKO mice, a weighted gene coexpression network analysis (WGCNA) was performed to organize genes into biologically meaningful groups.[45,46] After performing WGCNA, we identified two gene modules that were significantly correlated with OGT deletion and not significantly correlated with age with cor = 0.71 ($P < 3.8 \times 10^{-28}$) and cor = 0.77 ($P < 1.2 \times 10^{-18}$) (Figure 2.5, Figure 2.6a). Using gene ontology analysis, we found that the first green module was enriched with genes involved in the immune response ($P < 3.4 \times 10^{-4}$) such as Bcl-2 homologous antagonist killer (*Bak1*, 1.4-fold increase) and Shc homology 2-domain containing phosphatidylinositol-3,4,5-trisphosphate 5-phosphatase 2 (*Inppl1*, 1.4-fold increase) (Figure 2.6B). Bak1, a central pro-apoptotic member of the Bcl-2 family, is required for mitochondrial permeabilization and release of cytochrome c into the cytosol in the early stages of apoptosis.[47] The inositol polyphosphate 5-phosphatase SHIP2 (*Inppl1*) regulates macrophage phagocytosis and links metabolic signaling to the immune response.[48] Interestingly, SHIP2 has recently been implicated in connecting Aβ

to tau pathology through dysregulated phosphoinositide metabolic signaling in the 3xTg AD mouse model.[49] Enhanced expression of apoptosis- and immune-related genes is consistent with the extensive apoptosis observed in OGT cKO (Figure 2.1), and the upregulation of immune response genes observed in other AD mouse models.[24]
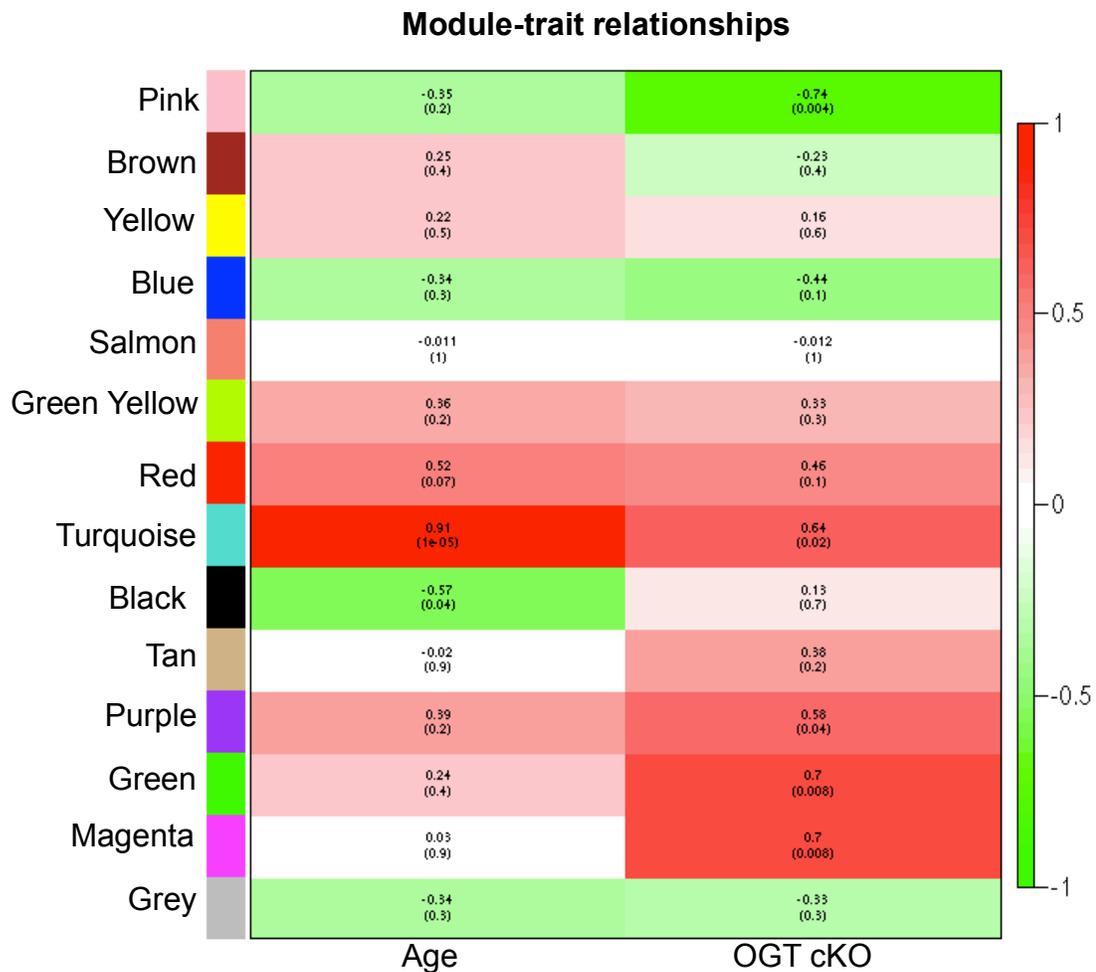
Figure 2.5 Heatmap indicating the module-trait relationships. Each of the 14 WGCNA modules can be correlated with certain traits including, in this case, mouse age and OGT cKO. Shown here is a heatmap indicating the strength of the correlation between each of the modules (given arbitrary colors) and mouse age and genotype (WT or OGT cKO).
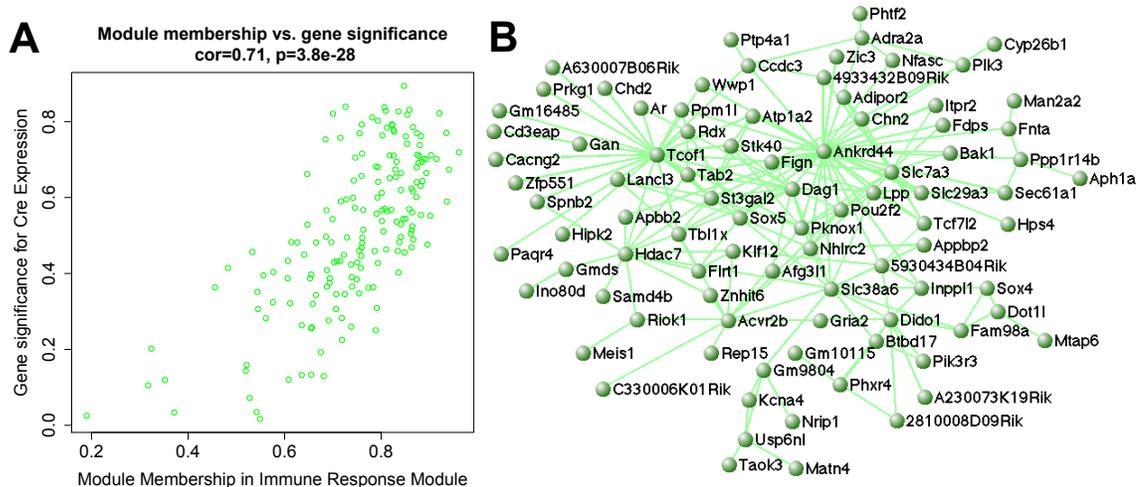
Figure 2.6 Green gene module is highly correlated with OGT cKO. (**A**) Cre expression in OGT cKO mice was significantly and highly correlated with membership in the green immune response module (cor = 0.71, P = 3.8 x 10^{-28}). (**B**) The green module gene network is enriched for immune response (gene ontology enrichment P < 3.75 x 10^{-5}). The gene network image was generated using VisANT (weight cutoff > 0.1).

Furthermore, the green module was enriched for genes involved in glial proliferation (Figure 2.7). Correlation of OGT cKO with a module enriched for glial differentiation and development is consistent with the extensive gliosis observed by IHC and confirmed by elevated expression of *Gfap* seen by both microarray and qPCR. Indeed, previous studies have shown that gliosis-related gene expression changes often dominate the transcriptional landscape in whole tissue global transcriptome studies.[50] Consistent with other studies looking at transcriptomic changes in AD hippocampi, we observed an upregulation of astrocytic genes such as *Gfap* and *Aqp4* and the microglial genes *Aif1* (Iba-1), *Cd68*, and *Emr1* with no changes in oligodendrocytic genes such as *Mbp, Sox10, Mog,* and *Mag* (Table 2.6).[51-53] Inconsistent with these AD human studies but consistent with AD mouse models, we did not yet observe downregulation of neuronal synaptic genes such as *Synt, Syn1, Syp,* and *Syt1* although there was a slight, non-significant decrease in the neuronal marker *Rbfox3* (NeuN, -1.2-fold, P < 6.4x10^{-5}, Bonferroni-corrected P = not significant) (Table 2.4).[23,51,52] This suggests that insufficient

neuronal loss has occurred in the OGT cKO at 2 months to fully recapitulate the downregulation of neuronal marker genes and/or that the OGT cKO mice show more transcriptional similarities to the AD mouse models than to human AD patients.
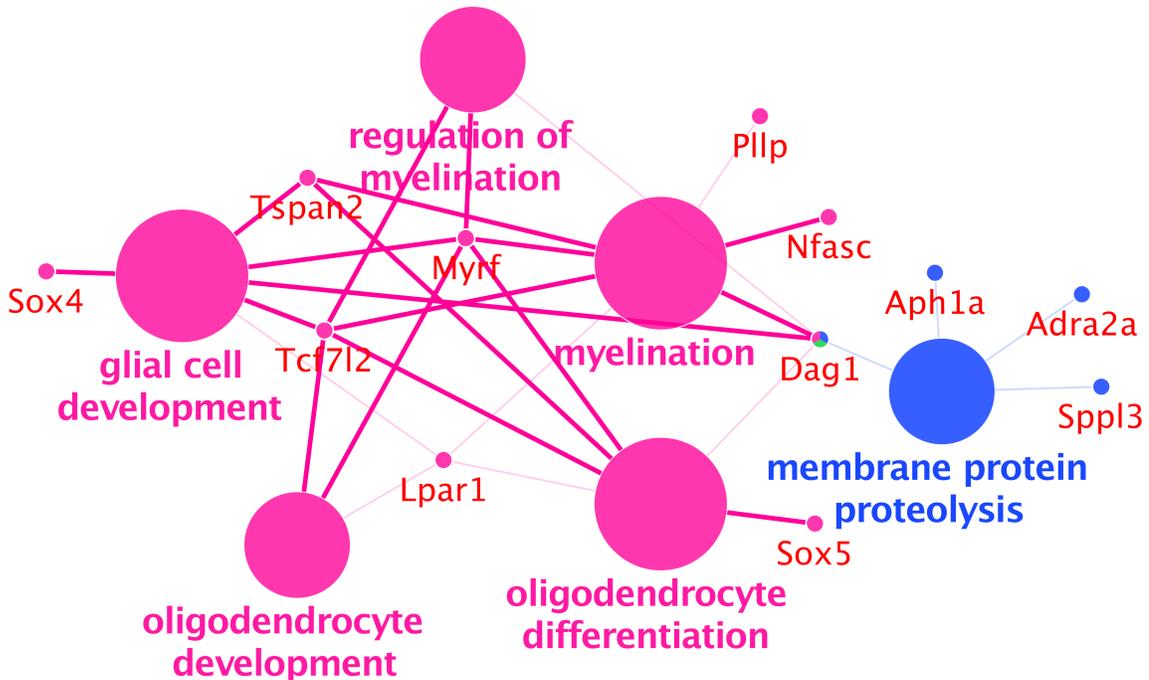
Figure 2.7 Cytoscape gene ontology annotations for the green module glial development-related genes. The categories include glial cell development and membrane protein proteolysis. The genes that belong to and are shared by these gene ontology categories are also shown. The gene annotations were created performed using the GO Biological Processes and Molecular Functions with the ClueGO extension in Cytoscape. The significance for all listed categories are Bonferroni step down corrected $P < 0.0039$.

Table 2.6 Cell type marker genes in the OGT cKO at 2 months

Genes	Description	Cell type	log$_2$(FC)	FC	*P*-values
Gfap	glial fibrillary acidic protein	Astrocyte	2.60	6.0	0
Aqp4	aquaporin 4	Astrocyte	0.86	1.8	1.3×10^{-7}
Aif1	allograft inflammatory factor 1	Microglia	1.41	2.7	0
Cd68	CD68 antigen	Microglia	1.80	3.5	0
Emr1	EGF-like module containing, mucin-like, hormone receptor-like sequence 1	Microglia	1.17	2.2	1.5×10^{-18}
Mbp	myelin basic protein	Oligodendrocyte	0.034	1.0	N.S.
Sox10	SRY-box containing gene 10	Oligodendrocyte	0.12	1.1	N.S.
Mog	myelin oligodendrocyte glycoprotein	Oligodendrocyte	-0.20	-1.1	N.S.
Mag	myelin-associated glycoprotein	Oligodendrocyte	0.28	1.2	N.S.

Table 2.6 contains the log$_2$(FC), FC, and *P*-values for different gene markers for astrocytes, microglia, and oligodendrocytes. OGT cKO/WT downregulated differentially-expressed genes at 2 months of age. FC = fold-change where negative fold-change values denote decreases; N.S. = not significant; *P*-values are Bonferroni corrected.

Furthermore, the green module was enriched for genes for proteins localized to the nucleus and involved in transcription ($P < 3.0 \times 10^{-3}$ and $P < 2.6 \times 10^{-2}$ respectively) suggesting that knocking out OGT is perturbing transcription and DNA-related processes. A few of the green module genes were DNA damage and repair genes, but these were not differentially expressed. This finding is consistent with the major role of OGT in transcriptional regulation and DNA damage response (see Chapter 4 for more on the role of *O*-GlcNAc in transcription and DNA damage response).[54] Finally, WGCNA allows one to identify "hub" genes with high connectivity within a given module (gene network). Knocking out these "hub" genes perturbs the entire gene network and affects traits that are highly correlated with its module.[55] Hub genes that have high connectivity in the green module include ankrin repeat domain 44 (*Ankrd44*, 40 interactors) and Treacher Collins Franceschetti syndrome 1, homolog (*Tcof1*, 28 interactors) (Figure 2.6B). *Ankrd44* is the gene for a regulatory component of the protein phosphatase 6 holoenzyme.[56] Protein phosphatase 6 (a.k.a. PP2a) regulates a myriad of functions, including regulating homology-directed repair of dsDNA breaks, mitotic spindle formation, E-cadherin maintenance at adherens junctions, ER to golgi transport, JNK pathway signaling, and NF-κB signaling.[57-62] *Tcof1* encodes a nucleolar protein that is important for (1) ribosomal RNA transcription, (2) maintenance of genomic integrity in response to DNA damage, (3) linking translation to DNA damage through roles (1) and (2), and (4) priming and protecting cells from oxidative stress during neuronal differentiation and development.[63-66] Overall, these hub genes could be key therapeutic targets for mediating the effects of *O*-GlcNAc loss on gliosis and neuroinflammation.

2.9 OGT knockout is highly correlated with a cell cycle arrest gene network

A second magenta module was enriched with genes involved in cell cycle arrest ($P < 3.8$ x 10^{-5}), including antigen KI-67 (*Mki67*), protein regulator of cytokinesis 1 (*Prc1*), and kinetochore protein Spc25 (*Spc25*) (Figures 2.5 and 2.8, Table 2.7). To explore further the connection between OGT ablation and cell cycle arrest, we probed OGT cKO mice for evidence of altered neuronal cell cycle progression. Although immunostaining of hippocampal neurons for proliferating cell nuclear antigen (PCNA) and bromodeoxyuridine (BrdU) revealed no appreciable differences (Figures 2.9 and 2.10), we observed a significant increase in the levels of cyclin A2 in the OGT cKO dentate gyrus compared to WT mice (Figure 2.11B). Cyclin A2 is an initiator of DNA replication during S-phase and is a well-established marker for cell cycle progression.[67] The Cyclin A2 positive neurons were found in *O*-GlcNAc-negative neurons throughout the dentate gyrus and the hilus and CA3 regions, but not in the CA1 and CA2 regions of the hippocampus (Figure 2.11A, C). This is consistent with the distribution of CaMKIIα expression where the highest expression in the mouse hippocampus is observed in the dentate gyrus followed by the CA3 region.[68] Importantly, previous studies have observed inappropriate cell cycle advancement in neurons of AD patients and AD mouse models and specifically the presence of Cyclin A2 in neurons in the hippocampus of AD mouse.[69,70] In human AD patient hippocampi and forebrains, post-mitotic neurons undergo S phase and DNA replication prior to neuronal loss.[71]
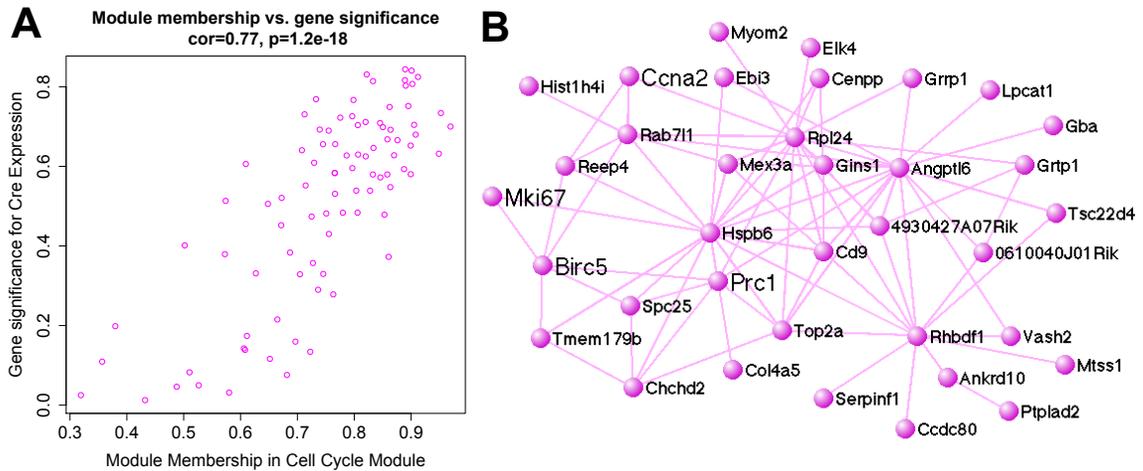
Figure 2.8 Magenta module is highly correlated with OGT cKO. (A) Cre expression in OGT cKO mice was significantly and highly correlated with membership in the magenta module (cor = 0.77, $P = 1.2 \times 10^{-18}$). (B) The magenta module was enriched for genes involved in cell cycle arrest (gene ontology enrichment $P < 3.4 \times 10^{-4}$). The magenta module gene network was generated in VisANT (weight cutoff > 0.1) and includes representative genes *Ccna2*, *Prc1*, and *Spc25*.

Table 2.7 DAVID GO annotation of magenta module genes

GO terms	#	Gene names	FE
Cell cycle	13	*Mitd1, Spc25, Mki67, Birc5, Cdca3, Ccna2, Cdkn1a, Gadd45a, Prc1, Reep3, Reep4, Ttc28, Txn14a*	6.0
Cell division	10	*Mitd1, Spc25, Birc5, Cdca3, Ccna2, Prc1, Reep3, Reep4, Ttc28, Txn14a*	7.7
Mitosis	8	*Spc25, Birc5, Cdca3, Ccna2, Reep3, Reep4, Ttc28, Txn14a*	8.9

Table 2.7 contains the top DAVID functional gene ontology annotations of the magenta module genes. Redundant and non-significant GO terms were removed from the list. FE = fold-enrichment; FDR < 5.0×10^{-2}.
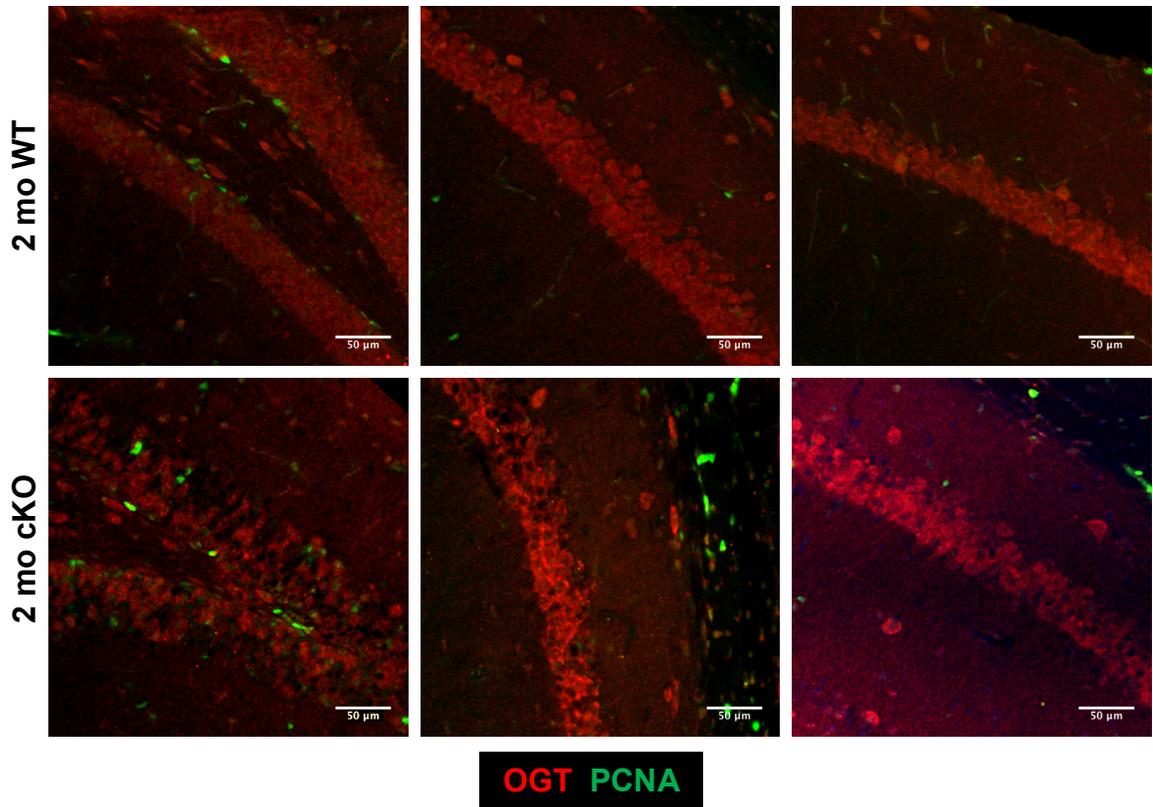
Figure 2.9 OGT cKO mice exhibit unchanged levels of PCNA-positive hippocampal neurons. No significant change in the number of PCNA-positive neurons was observed by immunostaining in the dentate gyrus or the CA1 regions of 2-month-old OGT cKO mice relative to WT mice. The few PCNA and BrdU-positive cells observed resided in the subgranular zone (SGZ) where adult neurogenesis occurs or outside of the hippocampus. Scale bar = 50 μm, n=4, each genotype.
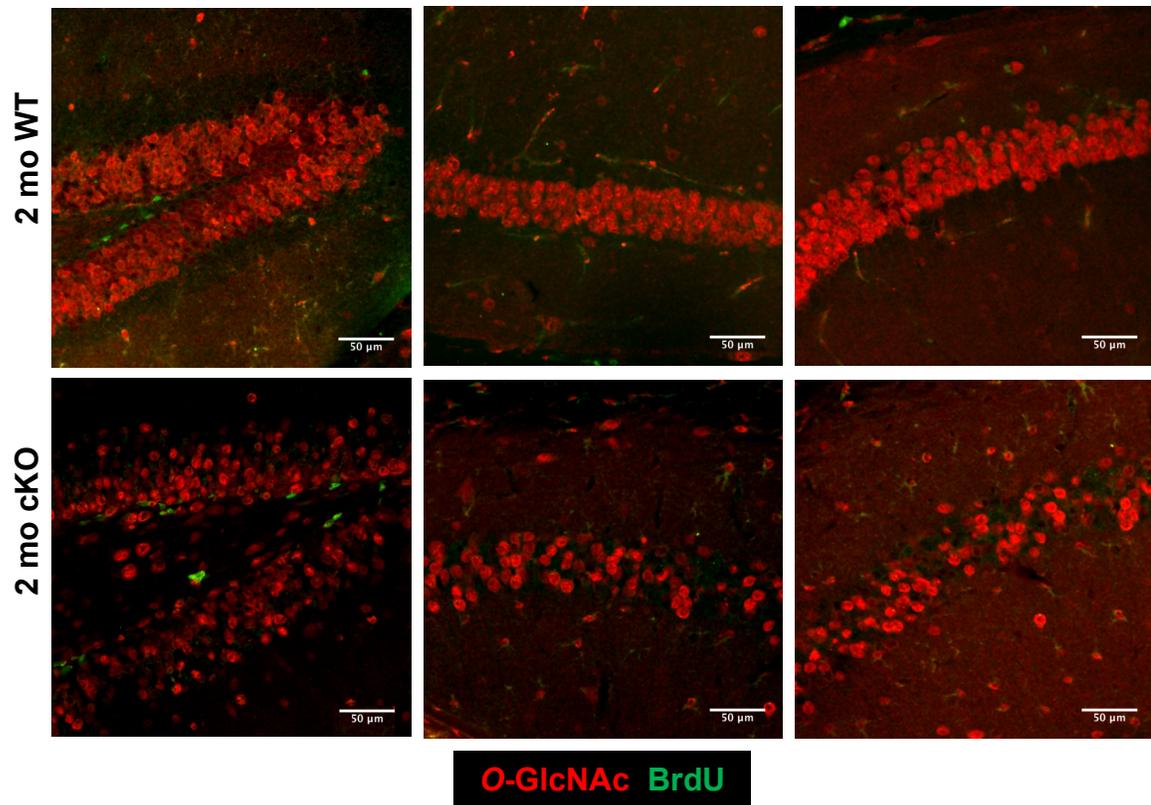
Figure 2.10 OGT cKO mice exhibit unchanged levels of BrdU-positive hippocampal neurons. No appreciable change in the number of BrdU-positive neurons was observed by immunostaining in the dentate gyrus or the CA1 regions of 2-month-old OGT cKO mice relative to WT mice. The few BrdU-positive cells observed resided in the subgranular zone (SGZ) where adult neurogenesis occurs. Scale bar = 50 μm, n=4, each genotype.
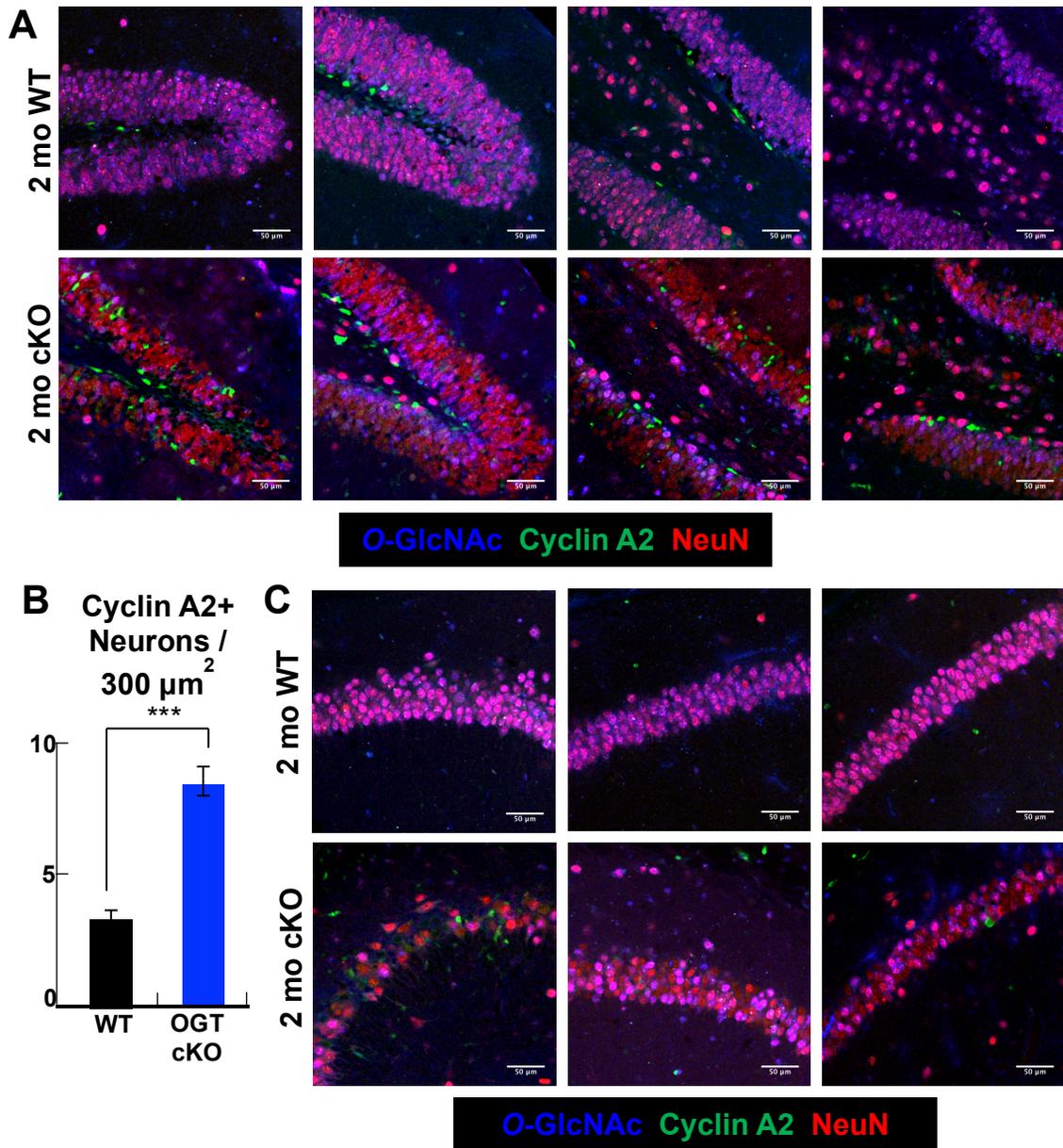
Figure 2.11 OGT cKO mice exhibit changes in levels of Cyclin A2-positive hippocampal neurons. (A) The number of Cyclin A2-positive neurons in the dentate gyrus, hilus, and CA3 regions of the hippocampus increased in the 2-month-old OGT cKO mice relative to WT mice. The Cyclin A2-positive cells observed in the WT hippocampus resided in the subgranular zone (SGZ) where adult neurogenesis occurs. (B) The number of Cyclin A2-positive neurons in the dentate gyrus significantly increased in the OGT cKO at 2 months of age. (C) The number of Cyclin A2-positive neurons was unaltered in the CA1 and CA2 regions of the OGT cKO mouse. ***$P < 0.0005$, n=4, each genotype. Scale bar = 50 μm.

We validated that several of the differentially-expressed cell cycle-related genes in the magenta gene network were significantly upregulated using qPCR (Figure 2.12). We verified that Cyclin A2 (*Ccna2*, 1.5-fold), Cyclin D3 (*Ccnd3*, 1.3-fold), and growth

arrest and DNA-damage-inducible protein α (*Gadd45a*, 1.6-fold) were upregulated in OGT cKO hippocampi at 2 months. Interestingly, several metabolic genes including *Pfkfb3* (2.2-fold) have been shown to regulate cell cycle directly. *Pfkfb3* upregulation leads to increased expression of *Ccnd3* and cell cycle progression toward G1 phase.[72,73] Importantly, increased PFKFB3 levels in neurons leads to apoptosis by shunting carbon flux through glycolysis rather than the pentose phosphate pathway (PPP), which ultimately exacerbates oxidative stress.[74] This shunting from glycolysis to PPP (a.k.a. the Warburg effect) in order to reduce oxidative stress is a strategy employed in cancer as well and is enhanced through the direct *O*-GlcNAc glycosylation of PFK1 under hypoxic conditions.[38] Increased expression of cell cycle-related genes has been observed in transgenic AD mice as well as in AD human patients.[75]
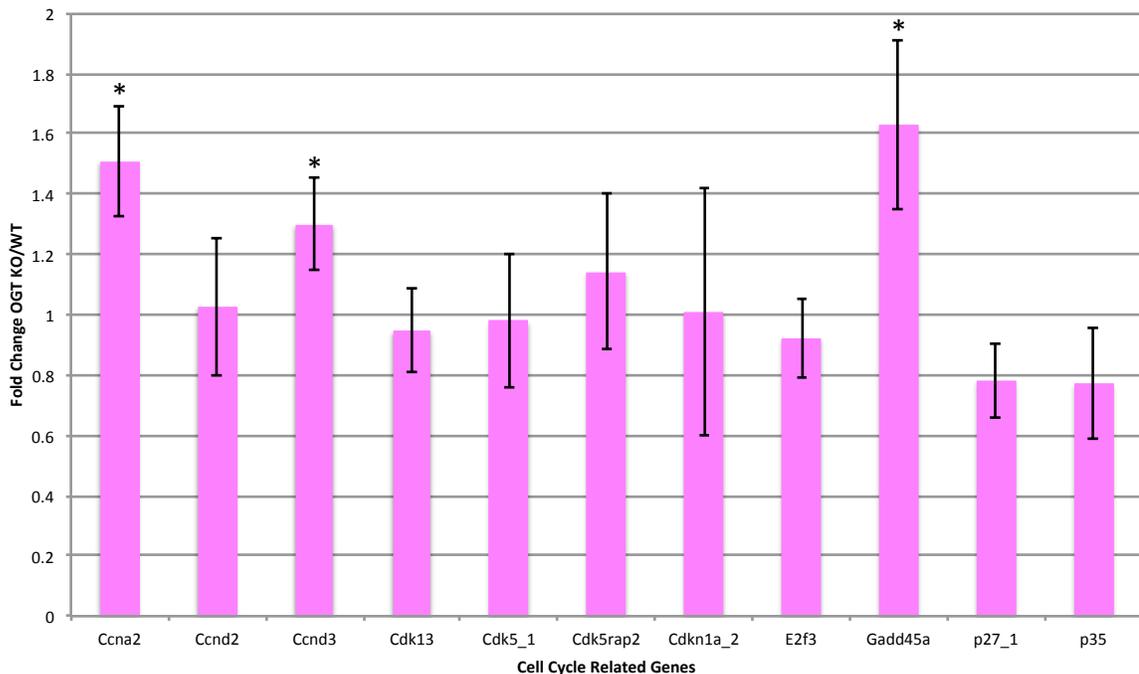
Figure 2.12 Cell cycle genes upregulated in OGT cKO mice at 2 months. We validated that cyclin A2 (*Ccna2*), cyclin D3 (*Ccnd3*), and growth arrest and DNA-damage-inducible protein (*Gadd45a*) were upregulated while other important cell cycle genes were not perturbed. *$P < 0.05$.

82

In addition, the magenta module's hub genes were heat shock protein family B (small) member 6 (*Hspb6*, HSPB6, 18 interactors), angiopoietin-like protein 6 (*Angptl6*, ANGPTL6, 17 interactors), and rhomboid 5 homolog 1 (*Rhbdf1*, 13 interactors), which were upregulated 2.0-fold, 1.4-fold, and 1.5-fold respectively in OGT cKO hippocampi (Figure 2.8B). A final magenta hub gene was ribosomal protein L24 (*Rpl24*, RPL24, 16 interactors), but this gene was not differentially expressed based on our microarray data. HSPB6 (a.k.a. Hsp20) has been shown to be involved in stress response, important for preventing aberrant aggregation of denatured or damaged proteins, and its association with pathological hallmarks in AD brains has been shown to be neuroprotective.[76-78] In adipose tissue, ANGPTL6 is known to be a critical mediator of energy metabolism: (1) knockout of *Angptl6* results in either embryonic lethality or morbid obesity and insulin resistance in mice and (2) inhibition of mitochondrial oxidative phosphorylation leads to enhanced *Angptl6* expression, which then increases the expression of fibroblast growth factor 21 (*Fgf21*).[79-81] Several studies have implicated *Rhbdf1*, which encodes a regulator of EGF/TGF-α signaling, in Alzheimer's patients.[82,83] The final hub gene, RPL24, is a critical component of the ribosome that mediates protein synthesis and plays a key role in cell cycle, growth, and survival.[84] Together, the hub genes, *Hspb6, Angptl6, Rhbdf1,* and *Rpl24*, are central mediators of cell cycle arrest gene expression in the OGT cKO mouse.

Finally, the protein levels of cyclin-dependent kinase 5 (Cdk5), a negative regulator of cell cycle progression, were significantly decreased in OGT cKO hippocampi when compared to WT.[85] Despite no change in *Cdk5* mRNA levels (Figure 2.12), IHC analysis also indicated a specific reduction in Cdk5 protein levels in *O*-GlcNAc-null hippocampal neurons (Figure 2.13), suggesting the potential for post-transcriptional

regulation of Cdk5 by *O*-GlcNAc. Cdk5 has been implicated as a critical regulator of synaptic plasticity, phosphorylation of tau, cell cycle progression, and DNA damage response in AD and other neurodegenerative diseases.[86-88] Taken together, these WGCNA, qPCR, microarray, IHC, and Western blot analyses provide evidence that there is enhanced cell cycle progression in the hippocampus of OGT cKO mice.
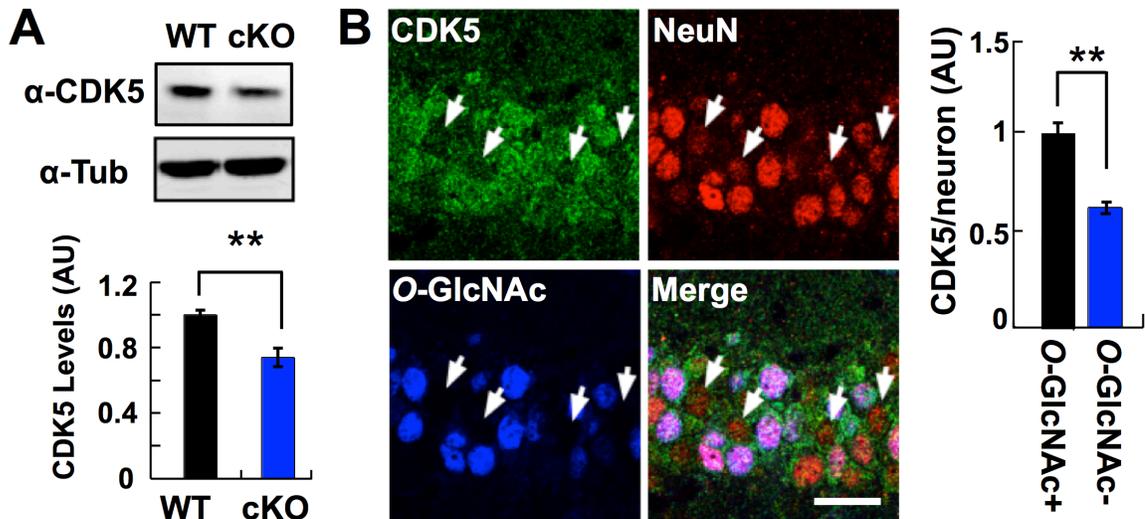
Figure 2.13 Cdk5 levels are depleted in OGT cKO neurons in the hippocampus. (*A*) A significant decrease in Cdk5 expression levels was detected by Western blotting in the hippocampus of 2-month-old OGT cKO mice. $n = 8$. *$P < 0.005$. (*B*) *O*-GlcNAc-negative neurons have decreased Cdk5 immunostaining (arrowheads) in the hippocampus at 2 months. Scale bar = 10 µm. Figure from Andrew Wang.

2.10 OGT cKO, amyloid-forming, and plaque-forming mice are highly correlated with a immune response gene network

We next wanted to compare our OGT cKO with other AD and neurodegenerative mouse models in a more global fashion to see if the OGT cKO fully recapitulates the transcriptional changes observed in these mouse models. In a recent study from Dr. Frances Edwards and colleagues, tangle-forming and amyloid-forming AD and FTDP17 (frontotemporal dementia with parkinsonism) mouse models were compared based on their disease pathology and gene expression changes (Figure 2.14).[89] The tangle-forming mice (MAPT^{P301L}) demonstrated increased hyperphosphorylated tau, leading to the

formation of neurofibrillary tangles (NFTs) in addition to increased immune response and decreased synaptic gene expression. On the other hand, the amyloid-forming mice had a variety of different genotypes: TAS10 (APP^{K670N,M671L}), TPM (PSEN1^{M146V}), HET-TASTPM (APP$^{Swe/+}$/PSEN1$^{M146V/+}$), HO-TASTPM (APP$^{Swe/Swe}$/PSEN1$^{M146V/M146V}$). Depending on the genotype and age, the amyloid-forming mice exhibited increased levels of Aβ peptides, accumulation of Aβ to form Aβ plaques, and elevated immune response gene expression.[89] We obtained the hippocampal samples from this data set, 113 samples altogether, and compared the gene expression traits to our OGT cKO mice using WGCNA (Figures 2.14, 2.15).
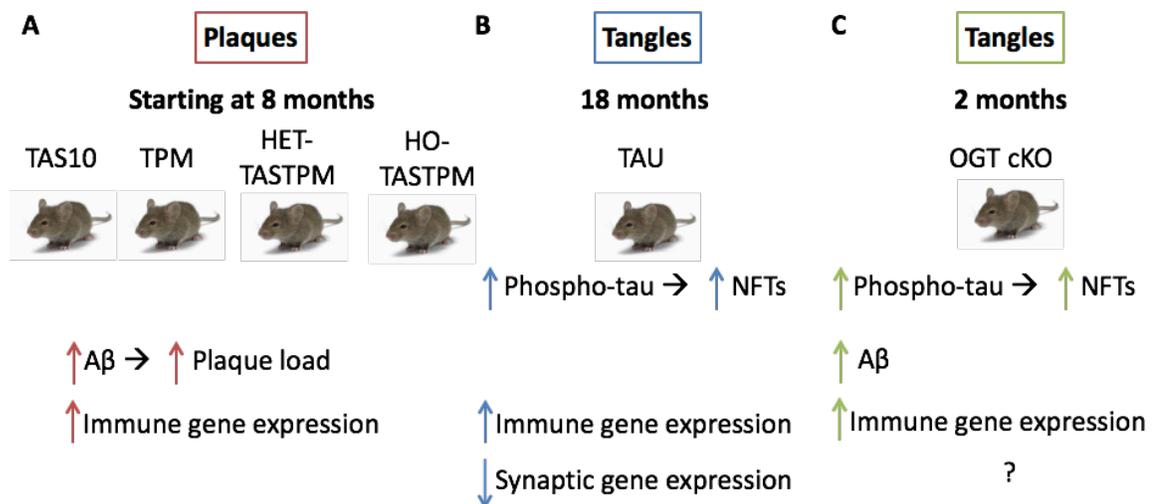
Figure 2.14 Comparison of the characteristics of the mice from the AD/FTDP mouse study and the OGT cKO mouse. (*A*) The plaque-forming mice are one of four different genotypes TAS10, TPM, HET-TASTPM, and HO-TASTPM. Depending on the genotype, these mice increase their Aβ peptide load, which leads to a 1:1 increase in plaque load (starting at 8 months) and an increase in immune gene expression. (*B*) The tangle-forming mice have the MAPT^{M301L} mutation, leading to increased hyperphosphorylated tau, NFTs (18 months), and immune gene expression and decreased synaptic gene expression. (C) The OGT cKO at 2 months of age displays increased hyperphosphorylated tau, NFTs, amyloidogenic Aβ peptides (not yet plaques), and immune gene expression.

85
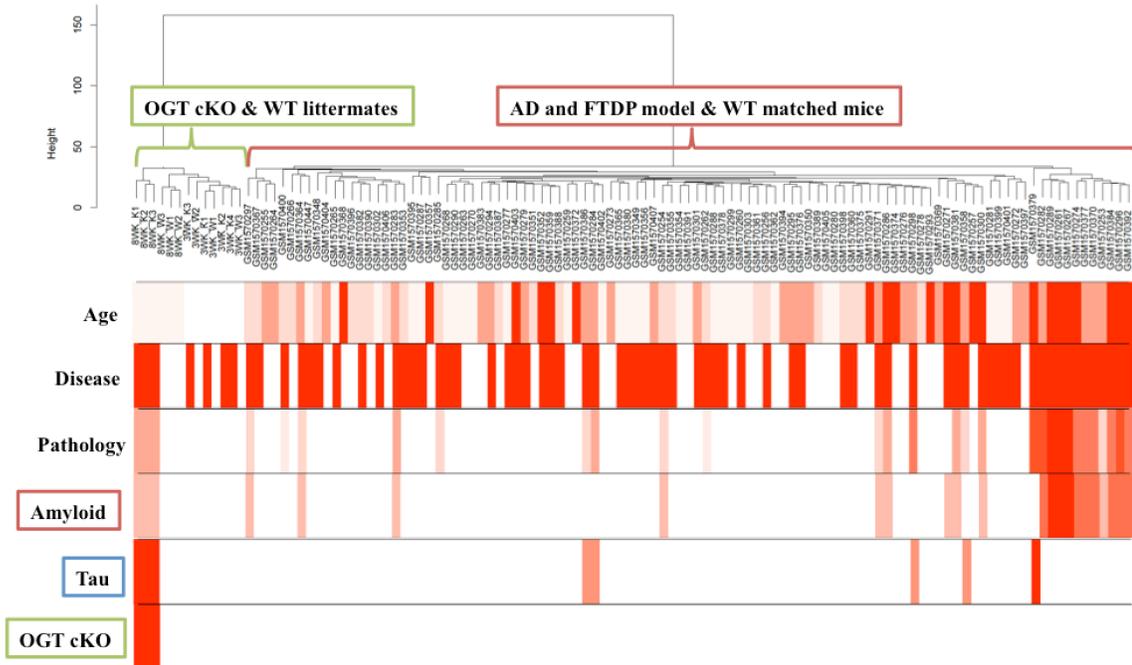
Figure 2.15 WGCNA dendrogram of hippocampal samples and traits. Shown here is the WGCNA sample dendrogram clustering similar samples and their annotation with the following traits: age (range was 3 weeks to 18 months), disease (presence or absence of disease), pathology (level of pathology associated with the presence of NFTs and plaques), amyloid (level of Aβ peptides and plaque formation), tau (level of hyperphosphorylated tau and NFT formation), OGT cKO (presence or absence of OGT knockout).

In a heatmap of the overall trait-module relationships, we observed that across most traits, OGT cKO and tau NFT-forming mice exhibit transcriptional similarities while the amyloid-plaque forming mice diverge from these two mice (Figure 2.16). The green module, enriched for immunity (FDR $< 6.2 \times 10^{-9}$), was positively correlated with all traits: age (cor = 0.41, $P < 2 \times 10^{-6}$), disease (cor = 0.29, $P < 9 \times 10^{-4}$), pathology (cor = 0.77, $P < 2 \times 10^{-25}$), OGT cKO (cor = 0.45, $P < 2 \times 10^{-7}$), tau tangle-forming (cor = 0.54, $P < 7 \times 10^{-11}$), amyloid plaque-forming (cor = 0.64, $P < 1 \times 10^{-15}$) (Figure 2.16). The green module was enriched for immunity (FDR $< 6.2 \times 10^{-9}$), both innate (FDR $< 1.0 \times 10^{-6}$) and adaptive (FDR $< 2.0 \times 10^{-4}$), NF-κB signaling pathway (FDR $< 1.0 \times 10^{-2}$), and lysosome (FDR $< 2.3 \times 10^{-2}$) (Table 2.8, Figure 2.17). This green module contains many of the upregulated immune-response and gliosis related genes in the OGT cKO mouse,

including *Gfap, Cd14, C3, Gusb, Tyrobp,* and *Man2b1*, which we validated with microarray and qPCR (Figure 2.18, Table 2.2).

The highest correlation with this module is with pathology supporting the idea that the more amyloid plaques and NFTs, the more immune response the mice exhibit. The next highest correlation is with the plaque formation followed by tangle formation, suggesting that amyloid plaque deposition is more highly correlated with immune response than tau NFTs. This is consistent with AD literature, which suggests that increasing levels of Aβ peptide lead directly to immune response, which in turn leads to hyperphosphorylation of tau and NFTs.[86] That being said, the causality vs. correlation relationship between immune response, Aβ pathology, and tau pathology in AD is still being adjudicated in the scientific courtroom. The OGT cKO had the next highest correlation with the green module suggesting that the immune response in the OGT cKO mice at 2 months was less pronounced than the amyloid plaque- and tau tangle-forming mice, which are generally older and have higher tangle and especially plaque formation than our 2-month-old OGT cKO mice. The high correlation of this immune response green module with pathology, amyloid plaque formation, NFT formation, and OGT cKO support previous findings showing extensive gliosis and immune response in these different mutant mice and show that amyloid plaque formation appears to correlate most highly with immune activation.
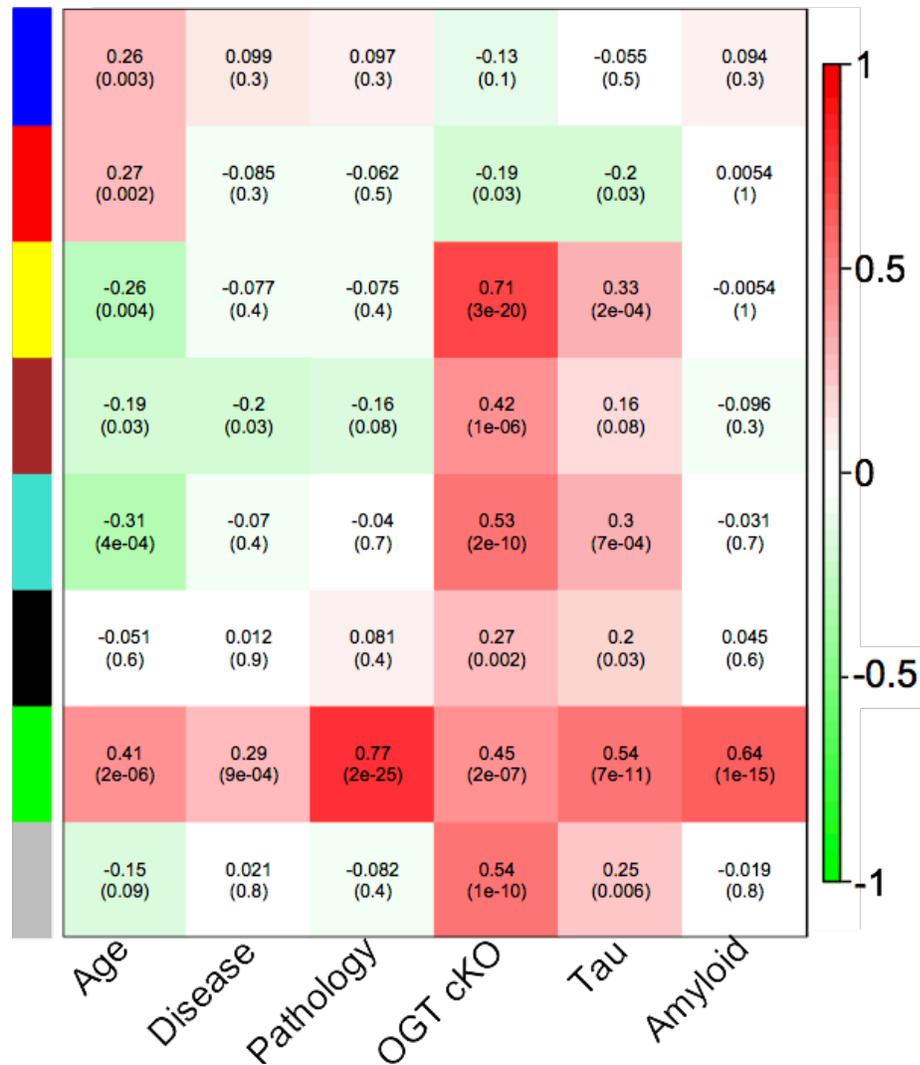
Figure 2.16 WGCNA trait and module correlations. The WGCNA trait and module correlations and *P*-values are shown here. Overall, there were 8 different modules that were generated from our WGCNA. Please note that the different modules are given arbitrary colors that have no correlation with the previous WGCNA analysis colors/modules.

Table 2.8 DAVID GO annotation of green module genes

GO term	#	Genes	FE
Immunity	23	*Havcr2, Tnfaip8l2, Itk, Ifitm2, C3, Ifitm3, Tlr13, Unc93b1, Tnfrsf17, Serping1, Oas2, Tlr7, Lgals9, Btk, Cd84, Ddx58, Prkd2, Cd86, Tap2, Inpp5d, Csk, Mx2, Cd14*	6.7
Immune system process	23	*Havcr2, Tnfaip8l2, Itk, Ifitm2, C3, Ifitm3, Tlr13, Unc93b1, Tnfrsf17, Serping1, Oas2, Tlr7, Lgals9, Btk, Cd84, Ddx58, Prkd2, Cd86, Tap2, Inpp5d, Csk, Mx2, Cd14*	6.0
Innate immune response	22	*Havcr2, Tnfaip8l2, Itk, Ifitm2, C3, Ifitm3, Tlr13, Unc93b1, Serping1, Oas2, Fes, Tlr7, Trim21, Btk, Cd84, Ddx58, Cyba, Fcer1g, Csk, Mx2, Cd14, Tyrobp*	5.5

Glycoprotein	66	*Kcne1l, Aebp1, Adora3, Osmr, Il4i1, Aqp4, Cd52, Ggt1, Kcnk13, Tlr7, Mmp2, Ednrb, Tor3a, Olfml3, Ch25h, Slc2a1, Lgi4, Asph, Man2b1, Ltb, Icam1, Cln3, Pnpla2, Serping1, Mcm3, Hcst, Cd84, Cd83, Ctsl, Cd37, Cd86, Serpina3n, Npc2, Bgn, Tnfsf13b, Adam17, Emp3, Pros1, Xdh, Abca9, C3, Slc39a12, Cd109, Unc93b1, Itgb5, Lrig1, Oas2, Sorcs3, Slc11a1, P2ry6, Lgals3bp, Smoc1, Pltp, Havcr2, Tmc6, Slamf9, Efemp2, Gusb, Rhbdf1, Tlr13, Frrs1, Tst, Fcgr2b, Liph, Slc15a3, Cd14*	2.0
Membrane	103	*Kcne1l, Tspo, Osmr, Aqp4, Ggt1, Cd52, Mmp2, Tlr7, Vcl, Btk, Ch25h, Slc2a1, Gng3, Asph, Dnajc3, Sgpl1, Scamp2, Tnfrsf17, Ifi47, Pnpla2, Sspn, Hcst, Igsf6, Cd37, Parp14, Emp3, Fgd2, Abca9, Ifitm2, Ifitm3, Unc93b1, Itgb5, Lrig1, Oas2, Llgl1, Sorcs3, Slc11a1, Lgals3bp, P2ry6, Laptm5, Vrk2, Fcer1g, Tyrobp, Havcr2, Slamf9, Acy3, Gusb, Rhbdf1, Fbxo2, Tlr13, Ephx1, Ddx58, Gngt2, Uaca, Slc25a10, Plcg2, Plau, Gfap, Ai467606, Kcnk13, Fes, Hvcn1, Ednrb, Ptges, Csk, Ltb, Rhog, Agpat2, Icam1, Cln3, Fmnl3, Ms4a6d, Mcm3, Cd84, Prkd2, Cd83, Cd86, Tnfsf13b, Prdx6, Adam17, Fhod1, Map3k11, Slc39a12, Hk2, Cd109, Atp5g3, Glipr2, Lpxn, Igtp, Tap2, Tgm2, Inpp5d, Ehd4, Tmc6, Slc10a3, Anxa3, Frrs1, Cyba, Rab32, Fcgr2b, Liph, Slc15a3, Cd14*	1.5
Extracellular exosome	54	*Gna14, Aebp1, Tspo, Mlph, Ggt1, Aldh1l2, Vcl, Tor3a, Slc2a1, Cdk5rap2, Csk, Man2b1, Dnajc3, Rhog, Idua, Icam1, Scamp2, Serping1, Arrdc1, Cd84, Ctsl, Renbp, Cd37, Cd86, Serpina3n, Npc2, Bgn, Prdx6, Pros1, C3, Ifitm3, Itgb5, Glipr2, Lgals3bp, Tgm2, Gsto2, Glo1, Ehd4, Havcr2, Tmc6, Acy3, Efemp2, Gusb, Fbxo2, Dbi, Anxa3, Akr1b8, Tst, Uaca, Capg, Plcg2, Hpgd, Cd14, Plau*	2.1
Adaptive immunity	10	*Cd84, Havcr2, Prkd2, Itk, Cd86, Tap2, Unc93b1, Tnfrsf17, Csk, Btk*	12
SH2 domain	10	*Itk, Socs3, Plcg2, Shc1, Inpp5d, Fes, Csk, Vav1, Sh3bp2, Btk*	9.4
Lysosome	13	*Cln3, Ifitm3, Gusb, Unc93b1, Il4i1, Tlr7, Ctsl, Npc2, Laptm5, Prdx6, Man2b1, Slc15a3, Idua*	6.1
Phosphoprotein	96	*Mocos, Gfap, Adora3, Ai467606, Pdlim4, Aqp4, Fes, Aldh1l2, Skap2, Hvcn1, Mmp2, Btk, Vcl, Ednrb, Plcb3, Oplah, Slc2a1, Casp8, Shc1, Cdk5rap2, Asph, Gng3, Csk, Dnajc3, Rhog, Sgpl1, Cln3, Fmnl3, Scamp2, Socs3, Ms4a6d, Anks1, Acads, Mlxipl, Pnpla2, Mcm3, Card10, Hcst, Cd84, Prkd2, Eml3, Renbp, Hspb6, Prdx6, Parp14, Adam17, Map3k14, Mcts1, Fhod1, Map3k11, Fgd2, Triobp, C3, Ifitm3, Mapkapk3, Unc93b1, Itgb5, Llgl1, Lpxn, Vrk2, Laptm5, Lyl1, Fcer1g, Glo1, Inpp5d, Tyrobp, Ehd4, Nr1h3, Havcr2, Fyb, Tmc6, Itk, Acy3, Rhbdf1, Fbxo2, Aff1, Samsn1, Vav1, Dbi, Anxa3, Tst, Ddx58, Ikbke, Cyba, Rab32, Rassf4, Uaca, Rps6ka1, Fcgr2b, Helb, Plcg2, Capg, Slc15a3, Plau, Sh3bp2, Igfbp5*	1.5
NF-κB signaling pathway	10	*Ddx58, Icam1, Tnfsf13b, Plcg2, Map3k14, Ltb, Ccl4, Plau, Cd14, Btk*	7.1
Disulfide bond	50	*Adora3, Osmr, Il4i1, Ggt1, Mmp2, Ednrb, Olfml3, Asph, Dnajc3, Man2b1, Idua, Icam1, Tnfrsf17, Serping1, Hcst, Igsf6, Cd84, Ctsl, Cd83, Cd86, Bgn, Npc2, Tnfsf13b, Prdx6, Adam17, Pros1, Xdh, Ccl3, C3, Cd109, Itgb5, Lrig1, Ccl4, P2ry6, Lgals3bp, Smoc1, Fcer1g, Glo1, Pltp, Tyrobp, Havcr2, Klk6, Slamf9, Efemp2, Igf1, Fcgr2b, Liph, Cd14, Plau, Igfbp5*	1.9
Lysosome	14	*Cln3, Gusb, Ifitm3, Unc93b1, Il4i1, Tlr7, Ctsl, Slc11a1, Npc2, Laptm5, Prdx6, Man2b1, Slc15a3, Idua*	4.4

Table 2.8 contains the significant DAVID functional gene ontology annotations of the green immune response module. Redundant and non-significant GO terms were removed from the list. FE = fold-enrichment; FDR $<5.0 \times 10^{-2}$.
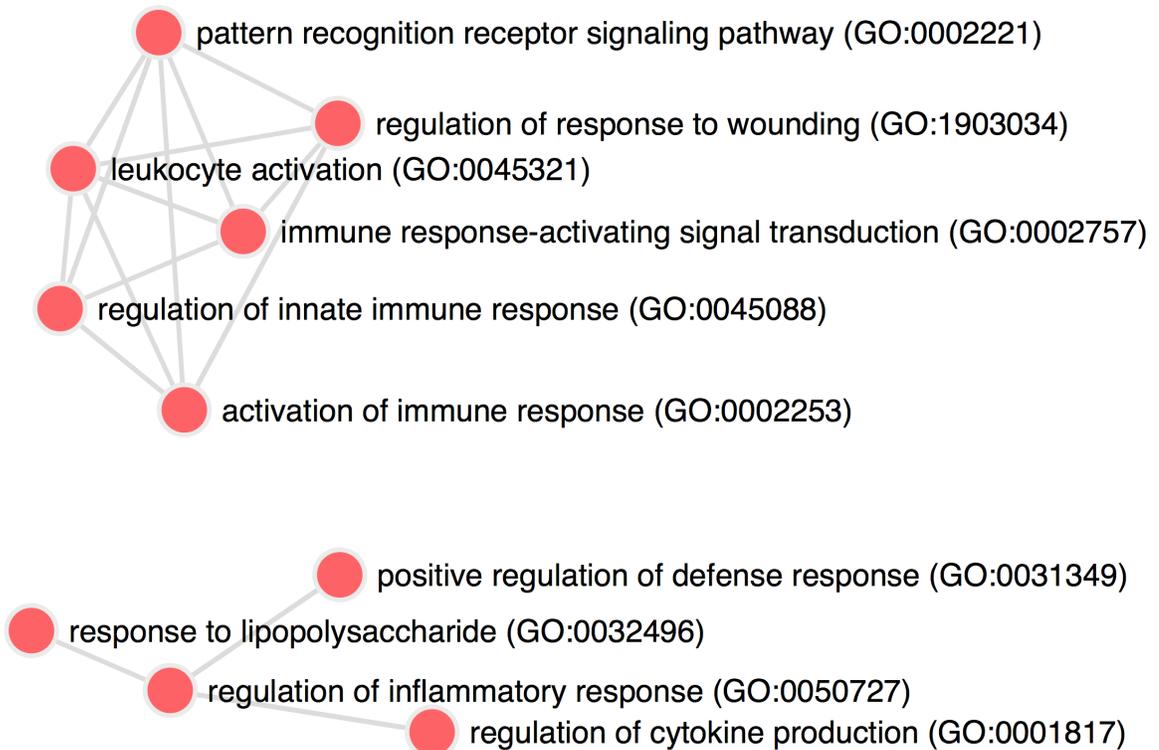
Figure 2.17 Top 10 gene ontology annotations for green immune response module. Shown above are the top 10 gene ontology annotations for the green module from the Enrichr database GO processes (adjusted $P < 6.4 \times 10^{-5}$).
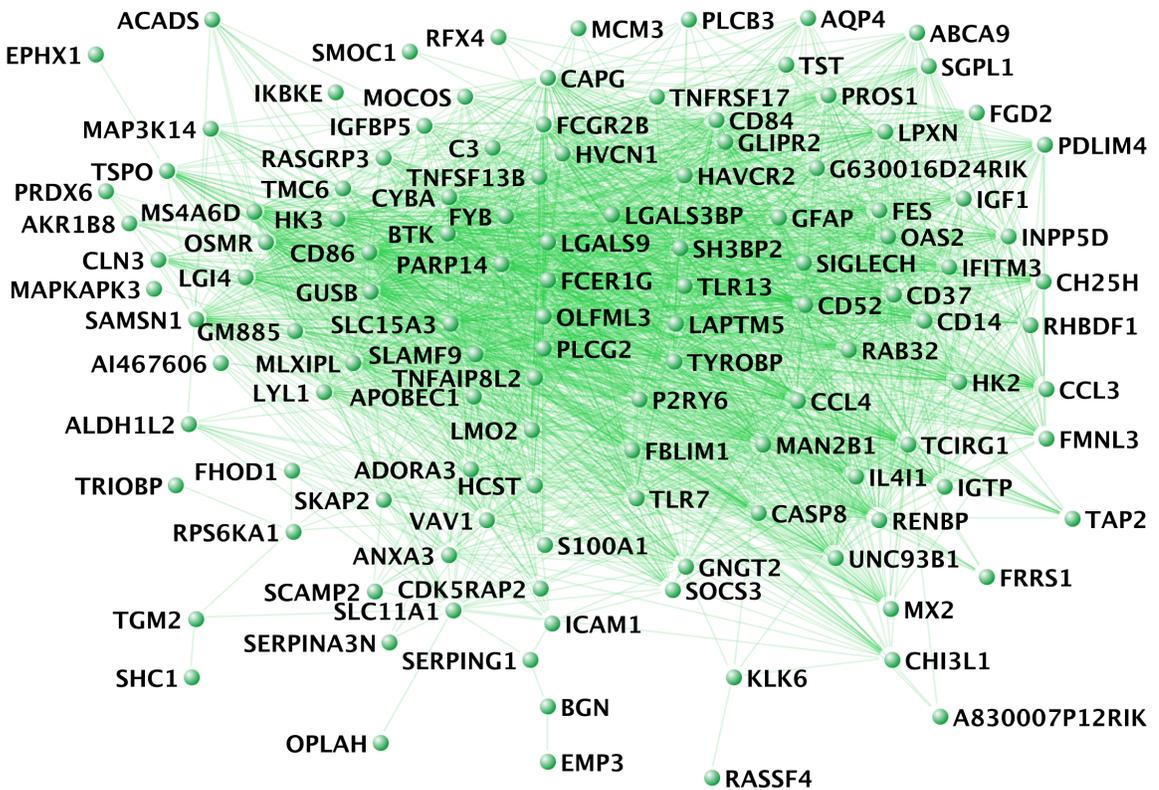
Figure 2.18 Green module is enriched for immune response genes. The green module was enriched for genes involved in immunity (DAVID GO enrichment FDR < 6.2×10^{-9}). The immune response gene network was generated in VisANT (weight cutoff > 0.1).

2.11 OGT cKO and plaque-forming mice but not amyloid forming mice are anti-correlated with a synaptic gene network

The red module was enriched for genes associated with the postsynaptic density (FDR < 6.4×10^{-9}) and the synapse (FDR < 0.029) was anti-correlated with the tau tangle-formation (cor = -0.19, $P < 0.03$) and OGT cKO (cor = -0.2, $P < 0.03$), but not correlated with pathology, disease, and amyloid plaque-forming mice (Figures 2.16 and 2.19, Table 2.9). In addition, the red module was enriched for genes associated with neuronal cell body (FDR < 2.8×10^{-4}), dendrite (FDR < 0.024), calcium transport (FDR < 0.038), growth cone (FDR < 0.040), and glutamatergic synaptic transmission (FDR < 0.043), suggesting that this module could be a more general neuronal and synaptic growth, activity, and plasticity module (Table 2.9). The anti-correlation with synaptic genes is consistent with (1) the Matarin and colleagues study that found that the tau tangle-forming mice are anti-correlated with a synaptic module and (2) our differential expression results in Section 2.6 revealing that the OGT cKO mouse displays downregulation of glutamatergic synaptic genes.[89] Both OGT cKO and tau are disrupting synaptic gene transcription while amyloid plaque-formation does not appear to disrupt it directly. This finding is consistent with previous studies demonstrating that synaptic loss is directly correlated with the presence of tau tangles in AD and removal of tau inoculates the hippocampus from stress-induced synaptic degeneration that occurs in AD.[90,91] Furthermore, a recent study showed that knockout of OGT in hippocampal neurons

results in fewer synapses and a higher proportion of immature synaptic spines, suggesting

that *O*-GlcNAc cycling is critical for the maintenance of synapses.[92]

Table 2.9 DAVID GO annotation of red synapse-related module

GO terms	#	Genes	FE
Phosphoprotein	81	Gprin1, Als2, Hmgn3, Pip5k1b, Sart3, Prkg1, Rgl1, Bzw2, Trim2, Diras2, Dab1, Trim9, Aak1, Trp63, Tlk1, Bcl7a, Kcnq2, Jph1, Brd9, Islr2, Tomm34, Nol6, Arhgef3, Cacng8, Baiap2, Actn1, Mbd4, Tle1, Cacng3, Rb1, Gal, Tmem74, Gtf2h2, Gtf2h1, Prkcb, Uhrf2, Rab11fip3, Ryr1, Zfpm1, Doc2b, Cpsf2, Dbn1, Unc13b, Ywhaz, Napa, Abi1, Ensa, Srf, Mlf2, Ephb2, Cdh8, Snn, Tnrc6c, Gp1bb, Syn2, Bcl11a, Pafah1b3, Rnf10, Camk2b, Nedd4l, Tcf4, Camk2a, Dtna, Trhde, Klf13, Grin1, Mapk10, Gas7, Zfp608, Atm, Itpr1, E130012a19Rik, Gria2, Rpl22, Bbc3, Hebp2, Trps1, Tsc2, Zfp532, Ube2e2, Plekha2	2.0
Postsynaptic density	16	Als2, Ywhaz, Cacng8, Baiap2, Grin1, Abi1, Mapk10, Itpr1, Dab1, Gria2, Syn2, Bcl11a, Camk2b, Camk2a, Dbn1, Lrp4	12
Alternative splicing	56	Als2, Fgf5, Hmgn3, Pip5k1b, Prkg1, Sart3, Pmvk, Dab1, Trim9, Aak1, Trp63, Rapgef4, Bcl7a, Kcnq2, Tomm34, Nol6, Nfkbiz, Arhgef3, Baiap2, Tle1, Mbnl2, Sez6l, Zcchc14, Prkcb, Rab11fip3, Uhrf2, Tlcd1, Dbn1, Unc13b, Camta2, Abi1, Ensa, Slco2a1, Ephb2, Mettl2, Dusp15, Bcl11a, Syn2, Rnf10, Nedd4l, Tcf4, Camk2a, Olfm1, Dtna, Grin1, Mapk10, Gas7, Zfp608, Itpr1, Rasl11b, Gria2, Tsc2, Parp6, Zfp532, Lrp4, Dusp7	2.2
Neuronal cell body	16	Als2, Baiap2, Grin1, Sez6l, Gal, Itpr1, Ephb2, Slc17a8, Dab1, Gria2, Tsc2, Rapgef4, Camk2a, Dbn1, Olfm1, Lrp4	5.3
Circadian entrainment	8	Gria2, Grin1, Ryr1, Camk2b, Prkg1, Camk2a, Itpr1, Prkcb	11
Dendrite	13	Als2, Slc17a8, Gria2, Trim9, Tsc2, Grin1, Trp63, Rapgef4, Camk2b, Camk2a, Lrp4, Itpr1, Ephb2	4.7
Synapse	11	Slc17a8, Gria2, Trim9, Cacng8, Syn2, Grin1, Abi1, Unc13b, Camk2a, Olfm1, Dtna	5.7
Calcium ion transport	8	Cacng8, Grin1, Ryr1, Cacng3, Camk2b, Camk2a, Itpr1, Prkcb	9.2
Growth cone	8	Gprin1, Als2, Gria2, Tsc2, Grin1, Abi1, Rapgef4, Dbn1	8.9
Synaptic transmission, glutamatergic	5	Cdh8, Als2, Grin1, Napa, Unc13b	28

Table 2.9 contains the significant DAVID functional gene ontology annotations of the red synapse-related module. Redundant and non-significant GO terms were removed from the list. FE = fold-enrichment. FDR $<5.0 \times 10^{-2}$.
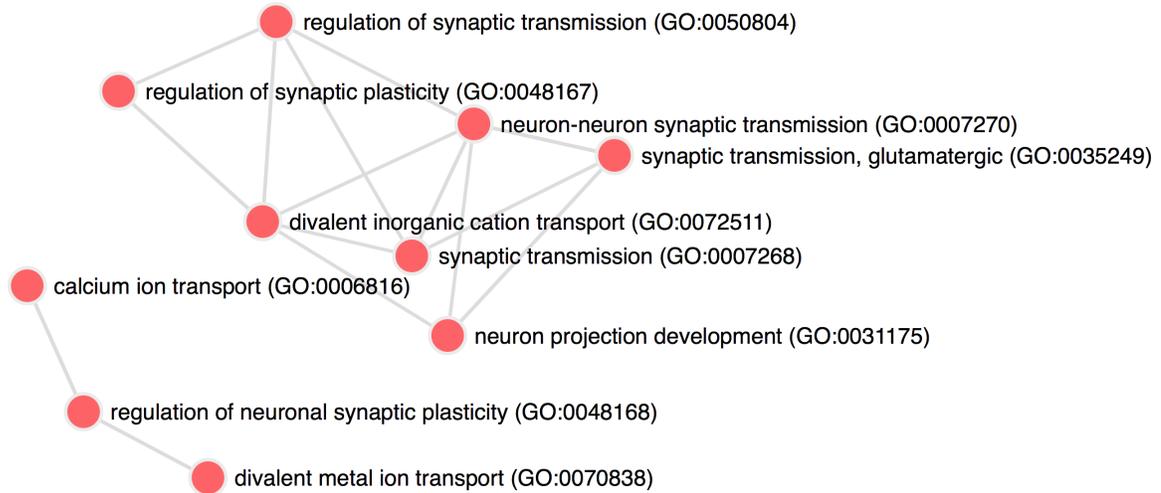
Figure 2.19 Top 10 gene ontology annotations for red synapse-related module. Shown above are the top 10 gene ontology annotations for the red module from the Enrichr database GO processes (adjusted *P* < 5.5×10^{-3}).
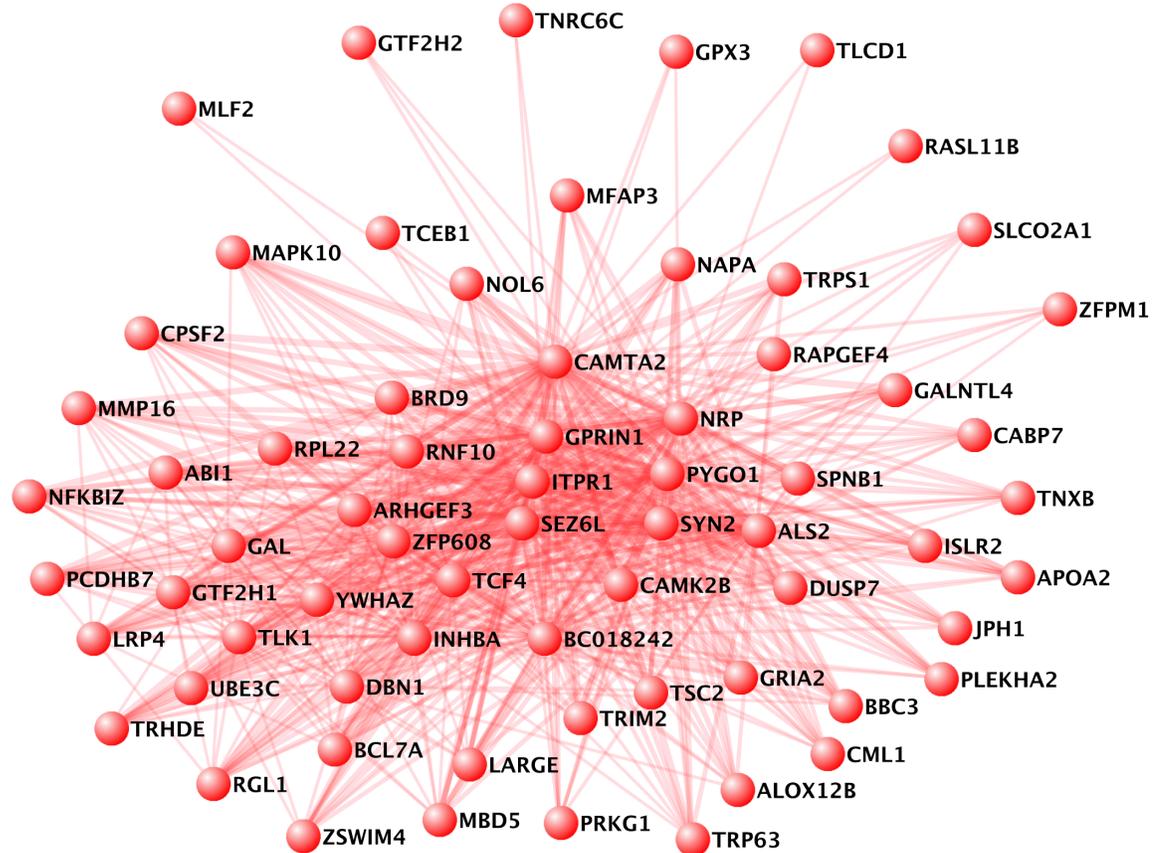
Figure 2.20 Red module is enriched for synaptic and neuronal activity genes. The red module was enriched for genes involved in postsynaptic density (FDR < 6.4×10^{-9}), synapse (FDR < 0.029), neuronal cell body (FDR < 2.8×10^{-4}), dendrite (FDR < 0.024), calcium transport (FDR < 0.038), growth cone (FDR < 0.040), and glutamatergic synapses (FDR < 0.043). The gene network was generated in VisANT (weight cutoff > 0.1).

2.12 OGT cKO mice are highly correlated with a nuclear gene network

93

The yellow module was enriched for genes associated with the nucleus (FDR < 3.5×10^{-9}), chromatin regulators (FDR < 6.0×10^{-4}), and transcription (FDR < 1.0×10^{-3}) was anti-correlated with the tau tangle-formation (cor = 0.33, P < 3.0×10^{-20}) and OGT cKO (cor = 0.71, P < 2.0×10^{-4}), but not correlated with pathology, disease, and amyloid plaque-forming mice (Figure 2.15, Table 2.10). OGT is known to plays a singular role in the nucleus through its interaction with several chromatin modifying complexes and modification of chromatin, nuclear pore proteins, and transcription factors (see Chapter 4). In particular, the yellow module is enriched for various post-translational modifications including phosphoprotein (FDR < 3.4×10^{-15}), acetylation (FDR < 3.3×10^{-6}), Ubl conjugation (FDR < 3.3×10^{-6}), and methylation (FDR < 1.8×10^{-3}), which, transcription by influencing transcription factor stability and the transcriptional accessibility of the chromatin state.[93] OGT and OGA mediate chromatin stability, nuclear pore stability, and DNA repair mechanisms through substantial crosstalk with these other PTMs (see Chapter 4).[54,94-96]

Finally, the yellow module is associated with alternative splicing (FDR < 2.6×10^{-6}) and splice variant (FDR < 7.1×10^{-3}). This module includes the neuron-specific RNA-binding protein (RNABP)-encoding genes, *Elavl3* and *Elavl4* ((embryonic lethal, abnormal vision, *Drosophila*)-like 3 and 4 (Hu antigens C and D)), which are responsible for mediating mRNA levels and splicing in order to control neuronal excitability and glutamate availability.[97] These neuronal alternative splicers are responsible for splicing APP itself, which leads to increased levels of Aβ peptides in AD.[98] OGT and OGA themselves are subject to alternative splicing, which is closely regulated by the *O*-GlcNAc levels and metabolic environment.[99] In particular, OGA (*Mgea5*) is associated

with late-onset AD and is alternatively spliced in AD.[100,101] While the role of *O*-GlcNAc in regulating alternative splicing has not yet been fully explored, *O*-GlcNAc has been found to modify at least 10 different splicing factors so far, but the roles of these modifications on splicing remains to be determined.[102,103] In addition to *Elavl3* and *Elavl4*, several other yellow module genes are members of and regulate the spliceosome including *Ddx41* (DEAD (Asp-Glu-Ala-Asp) box polypeptide 41), *Khsrp* (KH-type splicing regulatory protein), and *Aqr* (aquarius).[104-106]

Table 2.10 DAVID GO annotation of yellow nucleus-related module

GO terms	#	Genes	FE
Phosphoprotein	154	*Atp1b1, Xpo6, Plekhm1, Xpo4, Morf4l2, Bbx, Tssc4, Trim46, Cul2, Epc1, Top1, Cul5, Pacsin1, App, Sin3b, Ank2, Nt5c3, H2afy, Dnajc5, Frs2, Stag2, Rnf31, Scamp1, Slc12a6, Aldh6a1, Ptprk, Ctrb1, Mecp2, Spag1, Stim1, Dll1, Ythdf1, Stk4, Gcc2, Jup, Eif2ak1, Hif1a, Siah1a, Mapk4, Vamp2, Nek4, Glg1, Dcbld2, Palm, Hdlbp, Slc38a2, Napg, Pfkfb2, Azi2, Zfp451, Nipbl, Fbxo42, Ddx41, Bcor, Stk38l, Asf1a, Sdf4, Zc3h14, Ehmt1, Osbpl8, Syt11, Eef1a2, Smad2, Ubp1, E130308a19Rik, Pold3, Dnajb9, Hdac2, Atp2a2, Smarcc1, Prkar1a, Tmpo, Klf4, Gpbp1, Hp1bp3, Ctcf, Rsrc2, Pcbp2, Parg, Nckipsd, Twistnb, Nsf, Wdr33, Nfx1, Psmd9, Usp13, Golga2, Kcnma1, Lair1, Ssbp2, Sgk3, Nup88, Ralbp1, Snapc3, Adnp, Pfkp, Ldb3, Mpp5, Topors, Rad50, Jmy, Pja2, Kif1c, Eif4g2, Senp3, Myrip, Nab2, Khsrp, Carm1, Mdm1, Clock, Rev3l, Usp7, B230219D22rik, Pard3, Nup160, Tpm1, Aldh3a2, Hic1, Stk32c, Cdyl, Dgkb, Nr1d2, Rasgrp2, Kif4, Per1, Chd1, Tbc1d4, Gtf3c2, Kif21a, Nfatc2, Fen1, Rasa2, Cnnm3, Msh6, Pdk2, Foxl2, Dnm1l, Immt, Kif18a, Raf1, Nlgn2, Nlgn3, Elavl4, Ppp1r13l, Pwp2, Rps6ka4, Tfrc, Pecam1, Apbb2, Hist1h2ao, Tob1, Fez1, Cbs*	1.8
Nucleus	100	*Xpo6, Xpo4, Morf4l2, Bbx, Cbx7, Zfp92, Top1, Epc1, Sin3b, Aspa, Zfp90, Smarcd1, H2afy, Pdgfc, Scrt1, Stag2, Polk, Polh, Socs1, Hdac10, Mecp2, Dll1, Stk4, Aqr, Hif1a, Siah1a, Mapk4, Hdlbp, Zfp451, Nipbl, Ddx41, Bcor, Asf1a, Zc3h14, Klf7, Ehmt1, Osbpl8, Eef1a2, Smad2, Ubp1, Pold3, Hdac2, Smarcc1, Tmpo, Scmh1, Klf4, Hp1bp3, Gpbp1, Ctcf, Dph3, Pcbp2, Hey2, Parg, Nckipsd, Twistnb, Wdr33, Nfx1, Tbl1xr1, Ssbp2, Nup88, Snapc3, Adnp, Topors, Rad50, Jmy, Mxd4, Senp3, Smu1, Gnaq, Nab2, Khsrp, Carm1, Clock, Mdm1, Rev3l, Usp7, Zfp64, Nup160, Spo11, Hic1, Cdyl, Nr1d2, Kif4, Chd1, Per1, Nfatc2, Gtf3c2, Fen1, Msh6, Foxl2, L3mbtl3, Kif18a, Raf1, Ppp1r13l, Pwp2, Rps6ka4, Dusp8, Hist1h2ao, Cbs, Tob1*	2.0

Alternative splicing	96	*Slc8a3, Prr16, Xpo6, Gpbp1, Hp1bp3, Bbx, Tssc4, Trim46, Cul2, Epc1, Sin3b, App, Aasdhppt, Ank2, Chst10, Nt5c3, Pcbp2, Smarcd1, H2afy, Parg, Nfx1, Golga2, Rnf31, Kcnma1, Polk, Lair1, Ssbp2, Nrxn3, Nup88, Spag1, Mecp2, Pfkp, Ldb3, Rad50, Jmy, Man2a2, Pja2, Smu1, Eif4g2, Hif1a, Nab2, Lrp12, Nptn, Nek4, Eda, Carm1, Mdm1, C87436, Clock, Usp7, Palm, Pard3, Zfp64, Cabp2, Nup160, Spo11, Angel2, Tpm1, Azi2, Hic1, Nphp3, Nipbl, Cdyl, Dgkb, Pde1c, Rasgrp2, Tbc1d4, Slco3a1, Strbp, Bcor, Nfatc2, Kif21a, B4galt7, Sdf4, Stk38l, Zc3h14, Cnnm3, Dnm1l, Ehmt1, L3mbtl3, Immt, Raf1, Smad2, Elavl3, Elavl4, Ubp1, E130308a19Rik, Peo1, Atp2a2, Smarcc1, Pecam1, Mep1b, Tmpo, Scmh1, Apbb2, Cbs*	1.8
Acetylation	72	*Xpo6, Hp1bp3, Ube2g1, Tssc4, Ctcf, Cul2, Top1, Smarcd1, H2afy, Abcb10, Parg, Dnajc5, Nsf, Stag2, Wdr33, Scamp1, Aldh6a1, Tbl1xr1, Ssbp2, Ralbp1, Nup88, Mecp2, Adnp, Pfkp, Ythdf1, Stk4, Pkia, Gcc2, Rad50, Jup, Pja2, Smu1, Eif4g2, Hif1a, Aqr, Khsrp, Rab14, Vamp2, Usp7, Palm, Hdlbp, Pard3, Pfkfb2, Tpm1, Hic1, Nipbl, Pde1c, Nr1d2, Slc30a5, Tbc1d4, Mtmr9, Bcor, Kif21a, Stk38l, Fen1, Rasa2, Zc3h14, Msh6, Dnm1l, Osbpl8, Immt, Eef1a2, Smad2, Ppp1r13l, Pwp2, Pold3, Fem1c, Hdac2, Smarcc1, Prkar1a, Tmpo, Hist1h2ao*	2.0
Ubl conjugation	45	*Usp7, B230219D22rik, Slc38a2, Ctcf, Rsrc2, Hic1, Cul2, Top1, Zfp451, Cul5, App, Sin3b, H2afy, Per1, Pdgfc, Nfatc2, Frs2, Wdr33, Rnf31, Zc3h14, Foxl2, Ehmt1, Dnm1l, Polh, Syt11, Fzd1, Kif18a, Adnp, Dll1, Smad2, Fzd2, Topors, Drg2, Jmy, Pold3, Hdac2, Hif1a, Nab2, Smarcc1, Khsrp, Clock, Klf4, Hist1h2ao, Fez1, Cbs*	2.6
Isopeptide bond	32	*Usp7, B230219D22rik, Ctcf, Tpm1, Rsrc2, Hic1, Cul2, Top1, Zfp451, App, Cul5, H2afy, Wdr33, Zc3h14, Foxl2, Dnm1l, Ehmt1, Polh, Kif18a, Adnp, Dll1, Topors, Pold3, Hdac2, Hif1a, Atp2a2, Smarcc1, Khsrp, Clock, Klf4, Hist1h2ao, Cbs*	3.0
Cytoplasm	121	*Slc8a3, Xpo6, Plekhm1, Xpo4, Bbx, Cbx7, Top1, Cul5, Aspa, App, Sin3b, Pacsin1, Ank2, Nt5c3, Pdgfc, Frs2, Rnf31, Polh, Ctrb1, Socs1, Mecp2, Spag1, Hdac10, Stim1, Ythdf1, Pkia, Stk4, Gcc2, Jup, Pcf11, Eif2ak1, Hif1a, Mapk4, Siah1a, Nek4, Palm, Hdlbp, Pfkfb2, Wars2, Angel2, Azi2, Strbp, Stk38l, Zc3h14, Eef1a2, Syt11, Smad2, Ubp1, E130308a19Rik, Hdac2, Dnajb9, Atp2a2, Prkar1a, Klf4, Ube2g1, Dph3, Aasdhppt, Pigh, Ankrd12, Pcbp2, Hey2, Parg, Nsf, Nfx1, Psmd9, Golga2, Kcnma1, Ssbp2, Ldb3, Pfkp, Mpp5, Adnp, Jmy, Kif1c, Pja2, Eif4g2, Senp3, Smu1, Myrip, Pank2, Gzmn, Rab5a, Khsrp, Carm1, Mdm1, Clock, Usp7, Pard3, Cabp2, Upp2, Tpm1, Hic1, Numa1, Dgkb, Rasgrp2, Kif4, Per1, Tbc1d4, Chd1, Mtmr9, Kif21a, Nfatc2, Rasa2, Pdk2, Msh6, Dnm1l, Kif18a, Raf1, Nlgn2, Fzd2, Drg2, Elavl4, Ppp1r13l, Fem1c, Rps6ka4, Tfrc, Apbb2, Dusp8, Tob1, Fez1, Cbs*	1.5
Chromatin regulator	16	*Tbl1xr1, Ehmt1, L3mbtl3, Morf4l2, Hdac10, Ctcf, Cbx7, Epc1, Hdac2, Smarcc1, Smarcd1, Chd1, H2afy, Bcor, Carm1, Asf1a*	5.3
Protein binding	89	*Atp1b1, Gpbp1, Plekhm1, Ctcf, Cul2, Cul5, Aspa, Sin3b, App, Pacsin1, Ank2, Zfp90, Pcbp2, Hey2, Smarcd1, Abcb10, Dnajc5, Scrt1, Frs2, Nsf, Nfx1, Rnf31, Golga2, Scamp1, Slc12a6, Kcnma1, Polk, Tbl1xr1, Ssbp2, Ralbp1, Polg, Socs1, Mecp2, Ldb3, Adnp, Mpp5, Secisbp2, Stim1, Dll1, Topors, Gcc2, Mxd4, Jmy, Jup, Pja2, Senp3, Eif4g2, Myrip, Hif1a, Gnaq, Siah1a, Rab5a, Rab14, Vamp2, Carm1, Clock, Usp7, Pard3, Nr1d2, Per1, Mtmr9, Bcor, Ddx41, Nfatc2, Foxl2, Ehmt1, Osbpl8, L3mbtl3, Immt, Syt11, Raf1, Smad2, Fzd2, Ppp1r13l, Ppp1r9a, Dnajb9, Hdac2, Tfrc, Atp2a2, Smarcc1,*	1.6

		Pecam1, Prkar1a, Tmpo, Scmh1, Apbb2, Klf4, Tob1, Fez1, Bmpr1a	
Transcription	46	*Zfp64, Gpbp1, Morf4l2, Bbx, Ctcf, Cbx7, Hic1, Zfp92, Epc1, Zfp451, Sin3b, Cdyl, Nr1d2, Zfp90, Hey2, Chd1, Per1, Bcor, Gtf3c2, Twistnb, Nfatc2, Scrt1, Asf1a, Nfx1, Tbl1xr1, Foxl2, Klf7, Ssbp2, L3mbtl3, Snapc3, Hdac10, Adnp, Mecp2, Smad2, Ubp1, Ppp1r13l, Mxd4, Hif1a, Hdac2, Nab2, Smarcc1, Khsrp, Carm1, Scmh1, Clock, Klf4*	2.2
Methylation	30	*Palm, Gpbp1, Hic1, Zfp451, Cdyl, Smarcd1, Tbc1d4, H2afy, Strbp, Fen1, Wdr33, Mto1, Golga2, Ssbp2, Eef1a2, Mecp2, Adnp, Ldb3, Raf1, Elavl4, Ppp1r13l, Kif1c, Eif4g2, Smarcc1, Rab14, Khsrp, Tmpo, Carm1, Nek4, Hist1h2ao*	2.8
DNA-binding	41	*Zfp64, Gpbp1, Hp1bp3, Bbx, Ctcf, Spo11, Hic1, Zfp92, Top1, Zfp451, Nr1d2, Zfp90, Kif4, Hey2, Pcbp2, Chd1, H2afy, Strbp, Nfatc2, Scrt1, Nfx1, Polk, Msh6, Foxl2, Klf7, Ssbp2, Polh, Polg, Snapc3, Adnp, Mecp2, Smad2, Ubp1, Mxd4, Hif1a, Khsrp, Tmpo, Clock, Klf4, Hist1h2ao, Rev3l*	2.3
Covalent chromatin modification	16	*Tbl1xr1, Ehmt1, L3mbtl3, Morf4l2, Hdac10, Ctcf, Cbx7, Epc1, Hdac2, Smarcc1, Smarcd1, Chd1, H2afy, Bcor, Carm1, Asf1a*	4.6
Metal-binding	67	*Slc8a3, Slc6a2, Plekhm1, Ctcf, Dph3, Zfp92, Trim46, App, Aspa, Aasdhppt, Zfp90, Nt5c3, Yod1, Scrt1, Nsf, Nfx1, Usp13, Rnf31, Kcnma1, Polk, Polh, Nrxn3, Ldb3, Adnp, Pfkp, Stim1, Topors, Stk4, Rad50, Man2a2, Pja2, Myrip, Gnaq, Siah1a, Nek4, Aoc2, Rev3l, Zfp64, Cabp2, Spo11, Ppat, Zfp111, Hic1, Stk32c, Zfp451, Dgkb, Pde1c, Nr1d2, Rasgrp2, Ddx41, Sdf4, Stk38l, B4galt7, Fen1, Rasa2, Zc3h14, Klf7, Ehmt1, Syt11, Raf1, Smad2, Zfp709, Atp2a2, Mep1b, Klf4, Cbs, Bmpr1a*	1.7
Splice variant	91	*Prr16, Xpo6, Gpbp1, Hp1bp3, Bbx, Tssc4, Trim46, Cul2, Epc1, Sin3b, App, Aasdhppt, Ank2, Chst10, Nt5c3, Pcbp2, Smarcd1, H2afy, Parg, Nfx1, Rnf31, Slc12a6, Kcnma1, Polk, Lair1, Ssbp2, Nup88, Spag1, Mecp2, Pfkp, Ldb3, Rad50, Jmy, Pja2, Smu1, Eif4g2, Hif1a, Nab2, Lrp12, Nptn, Nek4, Eda, Carm1, Mdm1, C87436, Clock, Usp7, Palm, Pard3, Zfp64, Cabp2, Nup160, Spo11, Angel2, Tpm1, Azi2, Nphp3, Nipbl, Cdyl, Dgkb, Pde1c, Rasgrp2, Tbc1d4, Slco3a1, Strbp, Bcor, Nfatc2, Kif21a, B4galt7, Sdf4, Stk38l, Zc3h14, Cnnm3, Dnm1l, L3mbtl3, Immt, Raf1, Smad2, Elavl3, Elavl4, Ubp1, E130308a19Rik, Peo1, Atp2a2, Smarcc1, Pecam1, Mep1b, Tmpo, Scmh1, Apbb2, Cbs*	1.5
Repressor	20	*Tbl1xr1, Mecp2, Hdac10, Ctcf, Ppp1r13l, Cbx7, Hic1, Mxd4, Eif4g2, Sin3b, Cdyl, Hdac2, Nr1d2, Nab2, Zfp90, Hey2, Scrt1, Scmh1, Bcor, Nfx1*	3.3
Chromatin binding	20	*Msh6, Polg, Hp1bp3, Mecp2, Adnp, Ctcf, Smad2, Cbx7, Top1, Sin3b, Nipbl, Hdac2, Smarcc1, Smarcd1, Chd1, H2afy, Parg, Nfatc2, Asf1a, Stag2*	3.2
Ubiquitin protein ligase binding	15	*Usp7, Dnm1l, Syt11, Ube2g1, Smad2, Cul2, Cul5, Hif1a, Pcbp2, Prkar1a, Per1, Yod1, Cbs, Usp13, Rnf31*	3.9

97

Table 2.10 contains the significant DAVID functional gene ontology annotations of the yellow nuclear-related module. Redundant and non-significant GO terms were removed from the list. FE = fold-enrichment. FDR <5.0x10^{-2}.
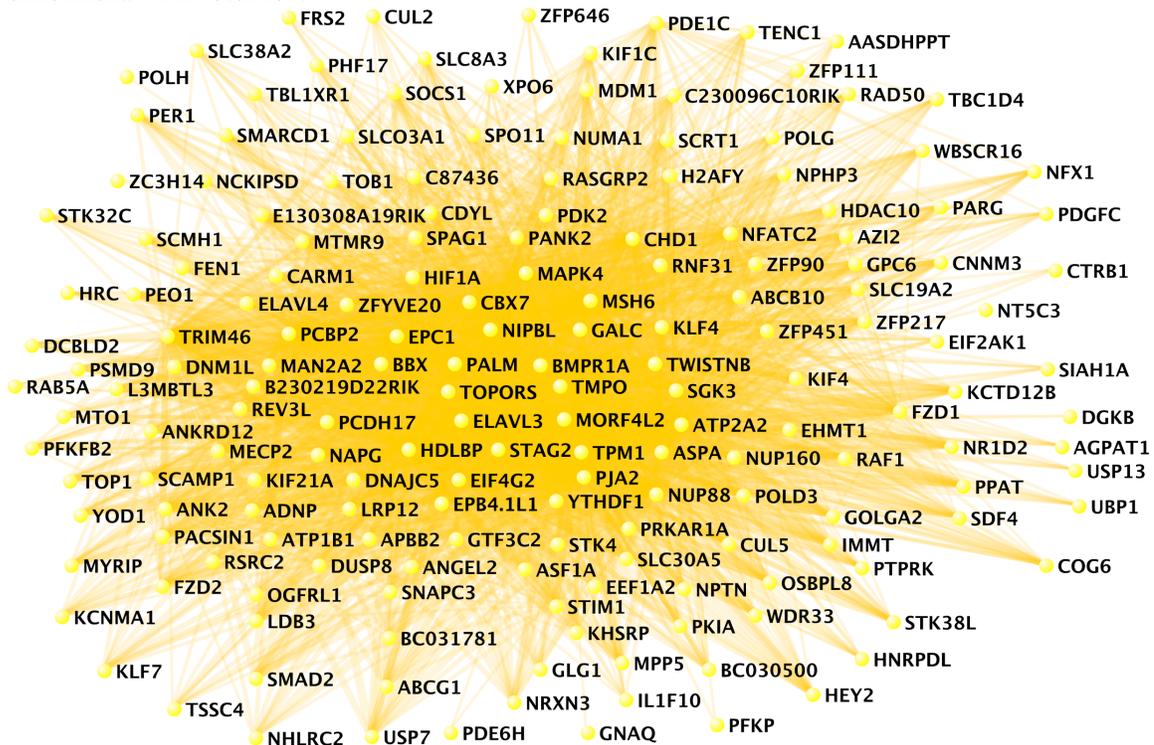
Figure 2.21 Yellow module is enriched for nuclear genes. The yellow module was enriched for genes involved in nucleus (FDR < 3.5x10^{-9}), chromatin regulators (FDR < 6.0x10^{-4}), transcription (FDR < 1.0x10^{-3}), and alternative splicing (FDR < 2.6x10^{-6}). In addition, the module is enriched for genes that are involved in PTMs including phosphorylation (FDR < 3.4x10^{-15}), acetylation (FDR < 3.3x10^{-6}), Ubl conjugation (FDR < 3.3x10^{-6}), and methylation (FDR < 1.8x10^{-3}). The gene network was generated in VisANT (weight cutoff > 0.1).

2.13 Discussion

In this study, we generated a forebrain-specific conditional *Ogt* knockout mouse to study the roles of *O*-GlcNAcylation in adult neurons. As the mice survived to adulthood, the effects of *Ogt* deletion on learning, memory, and adult neuronal function could be directly evaluated. Loss of OGT led to progressive neuronal death and other phenotypes associated with neurodegenerative diseases, including gliosis, activation of immune response and cell cycle genes, aberrant phosphorylation of tau, and amyloidogenic Aβ-42 peptides. OGT cKO mice also displayed behavioral deficits such as abnormal nesting behavior, and memory impairments, which are characteristics observed

in neurodegenerative mouse models of AD.[108] Notably, we found that human AD patients showed a significant decrease in OGT protein expression levels in the frontal cerebral cortex compared to control individuals, and this decrease correlated with progressive cognitive decline. Together, our studies reveal important neuroprotective roles for *O*-GlcNAcylation and suggest that alterations in *O*-GlcNAc signaling may contribute to neuronal pathology.

Both tau phosphorylation and APP processing were significantly altered in OGT cKO mice. Upon *Ogt* deletion, we observed enhanced tau aggregation and hyperphosphorylation at sites associated with neurofibrillary tangles. These findings further support previous studies suggesting a reciprocal relationship between *O*-GlcNAc and phosphorylation on tau. For example, increasing global *O*-GlcNAc levels reduced tau phosphorylation at Ser-199, Thr-212, Thr-217, Ser-262, Ser-396, and Ser-422 in mouse brain slices.[109] Moreover, inhibition of OGA elevated *O*-GlcNAc levels and decreased tau hyperphosphorylation in an AD mouse model.[110] Although we observed no significant change in the levels of tau *O*-GlcNAcylated at Ser-400, it is possible that *O*-GlcNAcylation of tau is differentially regulated at specific sites or that a small amount of residual OGT may be sufficient to maintain *O*-GlcNAc levels at this site.

With regard to APP, previous studies have suggested that higher *O*-GlcNAcylation levels can enhance non-amyloidogenic processing of APP by raising α-secretase activity in vitro and lowering γ-secretase activity.[111,112] In accordance with these observations, we found that loss of OGT enhanced the pathological processing of APP in vivo, increasing the ratio of the amyloidogenic 42-mer Aβ-peptide to the 40-mer. Notably, accumulation of hyperphosphorylated tau and Aβ-peptides is not observed in

many neurodegenerative mouse models, and both phenotypes are rarely observed together. For example, 5XFAD mice that express 5 different familial AD mutations in APP and presenilin 1 showed no accumulation of hyperphosphorylated tau, and mutant P301L tau transgenic mice displayed no increases of amyloid beta load, despite the presence of extensive neuronal death.[113,114] Thus, the defects in tau phosphorylation and APP processing observed in OGT cKO mice are not likely indirect effects caused by a global requirement for OGT in proper neuronal function and survival. Together, the findings suggest that the *O*-GlcNAc modification plays a central role in regulating both APP and tau, and dysfunctional *O*-GlcNAc signaling may contribute to improper APP processing and tau pathology.

We performed gene microarray analyses to obtain insights into the global, systems-level changes induced by loss of OGT. Our studies revealed that several cell cycle genes were upregulated in OGT cKO mice, including *Mki67*, *Prc1*, and *Spc25*. Consistent with an important role for OGT in cell cycle regulation, *O*-GlcNAcylation has previously been shown to modulate cell cycle progression by prolonging cyclin A and B expression and catalyzing proteolytic maturation of the critical cell cycle regulator, host cell factor-1 (HCF-1).[115,116] Interestingly, several genes involved in controlling oxidative stress were also significantly upregulated in OGT cKO mice, including heme oxygenase 1 (*Hmox1*), cyclooxygenase 1 (*Ptgs1*), and activity-dependent neuroprotective protein homeobox 2 (*ADNP2*). *O*-GlcNAcylation has previously been shown to be required for stress granule assembly in response to oxidative stress and to decrease oxidative stress in cardiomyocytes by reducing reactive oxygen species.[117,118] Thus, one important mechanism by which OGT may exert its neuroprotective effects is through the regulation

of critical cell cycle regulators and modulators of oxidative stress. The 'two-hit' hypothesis of AD postulates that both oxidative stress and mitotic dysregulation are necessary and sufficient to cause the neurodegeneration associated with AD.[119] Further supporting the gene expression analyses, we found that cyclin A2, a marker of cell cycle progression, was increased in OGT cKO mice, providing strong evidence for enhanced cell cycle progression in the hippocampi of these mice. Moreover, Cdk5 protein expression levels were significantly decreased upon loss of OGT. Cdk5 represses neuronal cell cycle advancement and is an important upstream regulator of oxidative stress through phosphorylation of peroxiredoxin I/II and p53.[119,120] Interestingly, a recent paper suggests that blocking *O*-GlcNAcylation of Cdk5 can lead to neuronal apoptosis by enhancing its association with the p53 pathway.[121] Several other studies have implicated aberrant cell cycle advancement in AD-associated neurodegeneration through deregulation of Cdk5 levels and activity.[120,122] Indeed, cyclin inhibitors that halt cell cycle progression have been shown to reduce the toxicity of Aβ peptides suggesting that they could be effective AD therapeutics.[87,123] Altogether, these results suggest that a major mode by which OGT ablation leads to neurodegeneration is through cell-cycle dysfunction. Future studies will focus on understanding the detailed mechanisms by which OGT affects Cdk5 function, the cell cycle and oxidative stress in neurons.

Lagerlof *et al.* reported that the tamoxifen-induced deletion of *Ogt* from CaMKIIα-positive neurons in adult mice leads to obesity from overeating.[124] This phenotype was due to reduced satiety caused by OGT ablation in the paraventricular nucleus (PVN) of the hypothalamus. Interestingly, no change in neuronal number in the hippocampus or PVN was noted upon quantification of DAPI+ cells, although the age of

the mice analyzed was not reported. The phenotypic differences between this and our model could be due to several explanations. First, there are likely differences in the timing of the OGT KO in the CaMKIIα-Cre versus the CaMKIIα-CreER^{T2} tamoxifen-inducible systems. In our study, loss of OGT and O-GlcNAcylation began at 4 weeks of age, and the neuronal loss and degenerative phenotypes were not significant until 8 weeks of age. In the Lagerlof study, tamoxifen injections were initiated at 6 weeks of age, and phenotypic responses were monitored up to 4 weeks later. Second, it is possible that the neurodegenerative phenotypes were not yet apparent during the window of their study as tamoxifen-induced CRE recombination can take several days. Furthermore, variability in the location and timing of CaMKIIα promoter-driven Cre recombination has been previously described and attributed to differences in the location of CaMKIIα-Cre transgene insertion within the genome.[125] Alternatively, excitatory forebrain neurons may be more vulnerable to OGT deletion at specific stages of maturation. Lastly, any effects on satiety in our model were likely obscured by the strong neurodegenerative phenotype. The phenotypic differences between the two OGT cKO models suggest that OGT plays specific roles in neuronal health and homeostasis during various stages, and they highlight the importance of the precise spatiotemporal coordination of O-GlcNAcylation.

The O-GlcNAc modification appears to play multiple, complex roles in neurodegeneration and, paradoxically, previous studies have suggested both neuroprotective and neurodegenerative effects upon decreasing O-GlcNAcylation.[110,126] We show that loss of O-GlcNAcylation in healthy neurons leads to progressive neurodegeneration *in vivo*, providing strong evidence that O-GlcNAc has neuroprotective functions in adult mammalian neurons. The OGT cKO and AD mouse models share

many of the same neurodegenerative, transcriptional, and behavioral phenotypes, suggesting common mechanisms of neurodegeneration. It is noteworthy that loss of OGT alone is sufficient to induce neurodegenerative pathologies and that this neurodegeneration occurs relatively rapidly. Induction of such phenotypes in other mouse models generally requires mutation or overexpression of multiple familial AD-related proteins such as tau, APP, and presenilin, and the pathological phenotypes take 9 to 12 months to progress.[108,127] The rapid course of neurodegeneration in OGT cKO mice underscores the critical importance of OGT in the maintenance of adult neuronal health and suggests that dysfunctional *O*-GlcNAc signaling may be an important contributor to neuronal pathology. Collectively, our studies provide a direct link between the ablation of *O*-GlcNAcylation and the induction of neurodegenerative phenotypes, suggesting that strategies to control *O*-GlcNAcylation and its neuroprotective effects may represent a novel approach for the treatment of neurodegenerative conditions.

2.14 Methods

2.14.1 Maintenance and breeding of OGT cKO mice

Mice were housed in groups whenever possible, and all animal procedures were performed in accordance with Institutional Animal Care and Use Committee guidelines of the California Institute of Technology. Only male mice were considered in this study due to random X inactivation of the *Ogt* gene in females, which would complicate phenotype analysis. OGT cKO mice were produced at the expected frequencies and were indistinguishable from WT littermates at birth.

As OGT cKO mice did not breed well, OGTfl and CaMKIIα-Cre mouse lines were maintained separately. By breeding homozygous OGTfl females (OGTfl-/-) with

heterozygous CaMKIIα-Cre males (CaMKIIα-Cre +/-), OGT cKO (OGTfl -/Y;CaMKIIα-Cre-/+) mice were produced, along with wild-type littermates (OGTfl -/Y; CaMKIIα-Cre+/+). The ROSA26L-eYFP mice were kindly provided by Dr. Ellen Rothenberg and bred with the CaMKIIα-Cre+/+ mice in order to determine the spatiotemporal expression and recombination of Cre recombinase.[128] As OGT cKO mice had an excessive grooming phenotype that resulted in skin lesions, the mice were culled at 6-7 months for humane reasons.

2.14.2 Behavioral studies

Fear conditioning was performed as described previously.[129] Briefly, mice were placed in a chamber (Med Associates) for the first time and allowed to explore for 120 s, after which an audible tone was played for 30 s. During the last 2 s of tone, a 0.7-mA foot shock was applied. This 150-s sequence was repeated, after which mice were returned to their home cages. After 24 h, mice were placed in the same chamber and monitored for 5 min to evaluate context-dependent memory. Mice were returned to their home cages for 1 h, and then placed into an altered context chamber (altered scent, shape, and lighting) for 5 min, with a tone playing for the last 180 s. All freezing quantification was done by Med Associates software, and freezing was normalized to pre-tone freezing % for cued tests. No significant difference in pre-tone freezing was observed between the two groups using an unpaired, two-tailed Student's *t*-test.

2.14.3 Antibodies

The following primary antibodies were used for immunohistochemistry (1:250 dilution) or Western (1:1000): OGT antibodies: TI-14 (Sigma, O6014), AL-25 and AL-28 (G.W. Hart, Johns Hopkins); Tubulin (Sigma, T9026); *O*-GlcNAc (RL2) (Pierce,

MA1-072); Tau antibodies: pThr231-AT180 (Pierce, MN1040), pSer202/Thr205-AT8 (Pierce, MN1020), pThr205 (Life Technologies, 44738G), pThr231 (Millipore, AB9668), tau-5 (Millipore, MAB361), pSer396 (Abcam, 32057); *O*-GlcNAc Ser400 (Anaspec, AS-55945, Western: 1:200); NeuN (Millipore, MAB377); NeuN (Abcam, ab134014, IHC: 1:100); GFP (Life Technologies, G10362); Iba-1 (WakoUSA, 019-19741); GFAP (Cell Signaling Technologies, 3670P); Cdk5 (Santa Cruz Biotechnology, SC-173); Cyclin A2 (Abcam, ab32386, IHC: 1:50); PCNA (Abcam, ab18197, IHC: 1:100); BrdU (Abcam, ab6326, IHC: 1:100). The following secondary antibodies were used for IHC (1:400 dilution): goat anti-mouse Alexa Fluor-405 (Life Technologies, A-31553), goat anti-mouse Alexa Fluor-488 (Life Technologies, A-11001), goat anti-mouse Alexa Fluor-546 (Life Technologies, A-11003), goat anti-rabbit Alexa Fluor-488 (Life Technologies, A-11008); goat anti-rabbit Alexa Fluor-405 (Life Technologies, A-31556); goat anti-rabbit Alexa Fluor-546 (Life Technologies, A-11035); goat anti-chicken Alexa Fluor-546 (Life Technologies, A-11040); goat anti-rat Alexa Fluor-488 (Life Technologies, A-11006). The following secondary antibodies for Western (1:10,000 dilution): goat anti-mouse Alexa Fluor-680, highly cross-adsorbed (Life Technologies, A21058); goat anti-mouse Alexa-Fluor-790 (Life Technologies, A11357); goat anti-rabbit Alexa Fluor-680 (Life Technologies, A21109); goat anti-rabbit Alexa Fluor-790 (Life Technologies, A11369).

2.14.4 Western blotting

Mouse cortical, hippocampal, and cerebellar samples extracted and then dounced in 2% sodium dodecyl sulfate (SDS) solution in water augmented with Roche cOmplete™, EDTA Protease Inhibitor Cocktail (Millipore Sigma, SKU 5056489001) and phosphatase inhibitors. Then, the protein amount was calculated using the Pierce

BCA (bicinchoninic acid) Protein Assay (Pierce, 23222 and 23224) as described by the manufacturer's instructions. Equal amounts of protein were diluted in water and 4X SDS-polyacrylamide gel electrophoresis (SDS-PAGE) loading buffer (200 mM Tris pH 6.8, 400 mM DTT, 8% SDS, 40% glycerol, 0.4% bromophenol blue) and heated to 95 °C for 10 min. The samples and ladder (Precision Plus Protein Dual Color Standards (Bio-Rad)) were separated using NuPAGE 4-12% Bis-Tris protein gels (NP0355BOX, ThermoFisher Scientific) at 180V at room temperature (RT) for 1 hour. Following resolution, the proteins were then transferred onto Immobilion-FL PVDF membrane (IPFL00010, EMD Millipore) at 250 mA at 4 °C for 1.5 hours. The membranes were blocked for 1 hour at RT in blocking buffer (either 5% BSA (bovine serum albumin) in TBST (50 mM Tris pH 7.4, 150 mM NaCl, 0.1% Tween) or with 5% nonfat milk in TBST according to the antibody manufacturer's specifications). Following the blocking step, the blots were incubated overnight at 4°C in the diluted primary antibody in the blocking buffer with gentle rocking. Blots were rinsed three times for 5 minutes with TBST, incubated with secondary antibodies at RT for 1.5 hours with gentle rocking, rinsed three times with TBST again, and then imaged using the Li-COR Odyssey® CLx Infrared Imaging System. The western blot intensities were quantified using Image Studio™ Lite Software (Li-COR) and significance in a pairwise comparison was determined using an unpaired, two-tailed Student's t-test.

2.14.5 Immunohistochemistry

For Nissl staining, slices were incubated in 100% EtOH followed by Histoclear (Electron Microscopy Sciences, 641101-01) for 2 min each. Slices were then sequentially rehydrated in 100%, 70%, and 50% aqueous EtOH for 2 min each. Slices were stained

with a 0.1% cresyl violet solution for 10 min, rinsed in distilled water, and differentiated in 90% EtOH with 1% acetic acid for 10 s. After a final wash in 100% EtOH, slices were cleared for 2 min in Histoclear and mounted with Vectashield.

For Fluoro-Jade C (Millipore, AG325) and Thioflavine S (Sigma, T1892) IHC staining, slices were incubated in 0.06% potassium permanganate solution for 10 min, washed in double-distilled H_2O, and transferred to a 0.0001% solution of Fluoro-Jade C in 0.1% acetic acid, or a 0.05% Thioflavine S solution in 50% EtOH for 10 min. Following a 5 min wash in H_2O, (or 50% EtOH for Thioflavine S), slices were mounted with Vectashield. For Cyclin A2 and PCNA staining, slices underwent antigen retrieval for one hour at 37 degrees in citrate buffer (0.01 M sodium citrate in 0.1% Triton X-100 in H_2O, pH 6.0) prior to blocking. For BrdU staining, slices were incubated in 2 N HCl for 20 minutes at 37 degrees followed by 10 min of neutralization in 0.1 M sodium borate buffer, pH 8.5 (two 5 min incubations). Next, the slices were rinsed three times with PBS and then blocked as normal.

Brain samples of frontal cortex from control ($n = 6$), Braak VI ($n = 8$), and Braak IV-V ($n = 4$) patients were obtained from the Alzheimer's Disease Research Center at the University of California, Los Angeles. Samples were from patients with a similar age, gender, and postmortem interval (PMI). Slices were pre-treated with 0.01 M citrate buffer (37 °C, pH 3.5) followed by treatment with Sudan black (0.1% w/v in 70% EtOH, Sigma, 199664)) and stained as described above. Four representative regions were imaged in each slice, and at least four slices were imaged per individual. Images were coded and discrete quantitative scoring of OGT protein signal (scale from 1-4) was carried out with researchers blind to sample ID.

2.14.6 BrdU Assay for Neurogenesis

The BrdU assay for neurogenesis was performed as previously described. Briefly, 2-month old male mice were injected with 100-200 μL of sterile 10 mg/ml BrdU (5-Bromo-2'-deoxyuridine, Sigma Aldrich, B5002-1g) in PBS buffer intraperitoneally in order to achieve a final concentration of 200 mg/kg BrdU. Twenty-four hours following the injection, the mice were transcardially perfused followed by dissection, fixation, and immunohistochemical staining as described previously.[130]

2.14.7 Aβ-Peptide ELISA

The Amyloid-beta ELISA Kits 1-40 and 1-42 (Covance, SIG-38954, 38956) were used to quantify Aβ-peptide levels following the manufacturer's protocol. Whole cortices were lysed in TBS (50 mM Tris-HCl, pH 7.5, 150 mM NaCl) containing 1% Triton-X100 and protease inhibitor cocktail (Roche, 11697498001), centrifuged at 100,000g, and the supernatant was assayed following the ELISA kit protocol.

2.14.8 RNA extraction, qRT-PCR, and Microarray Analysis

Hippocampal tissue was isolated from 3-week-old or 2-month-old mice, flash frozen, and stored at -80 °C in nuclease-free tubes until extraction. Total RNA was extracted using an RNeasy kit with RNase-free DNase treatment per the manufacturer's instructions (Qiagen, 79254). Following RNA extraction, the samples were sent to Phalanx Biotech Group (Belmont, CA) for microarray analysis. Briefly, a library was generated from the mRNA using the Ambion Amino Allyl MessageAmp II aRNA Amplification Kit (Life Technologies, AM1753) and applied to Mouse OneArrays v2 containing probes for over 23,000 genes (Phalanx Biotech). The microarray intensities were normalized using over 800 control probes throughout the array. In addition to

microarray analysis, the extracted mRNA was reverse transcribed using the iScript cDNA Synthesis Kit (Bio-rad, 1708891) for qRT-PCR. Following reverse transcription, the cDNA was used for qPCR analysis using Perfecta SYBR Green Fastmix with ROX (VWR, 101414-280) on an AB7300 Real Time System (Applied Biosystems). The qRT-PCR expression values were calculated using the comparative C_T method, and expression values were normalized to the geometric mean of 5 different reference genes as previously described.[131] The efficiencies of the qPCR primers were calculated using Python version 2.7.13 and were utilized in the calculation of the expression values as previously described.[131] A list of primers used in this study can be found in Appendix III. Statistical significance of qRT-PCR results was determined using an unpaired t-test. Reported *P* values for the microarray results from this study were calculated using a Bonferroni correction (all differentially-expressed genes have Bonferroni *P*-value < 0.001 unless otherwise noted).[132] All programs are available upon request.

2.14.9 WGCNA and Gene Ontology Analysis

Prior to performing the first WGCNA, the microarray data was pre-processed using limma in R version 3.3.2 and Python version 2.7.13. The AD and FTDP mouse model comparative analysis utilized hippocampal data from Matarin and colleagues downloaded from http://www.mouseac.org/ (GEO number GSE64398).[89] Then, the Matarin microarray data were aligned with our microarray data from the OGT cKO and the WT littermates using Python version 2.7.13. Following alignment, the Partek® Genomics Suite®, version 6.6 Copyright ©2017 was used to perform a quantile normalization across all of the microarray data sets.[133] Using the pre-processed microarray data, we performed WGCNA in R package version 3.0.0 on all microarray-

detected genes using the protocols previously described.[45,46] Following hierarchical clustering and module assignment, gene ontology enrichment analysis was performed using Bioconductor packages GO and the Database for Annotation, Visualization and Integrated Discovery (DAVID) as previously described.[134] The package VisANT version 5.0 (correlation cutoff of 0.1 (cor > 0.1)), Cytoscape version 3.5.1, ClueGO version 2.3.3, and CluePedia version 1.3.3 was used to visualize the WGCNA gene networks.[135-138] All programs are available upon request.

2.14.10 Statistical Analyses

Unless otherwise stated, all results are expressed as the mean ± standard error of the mean (SEM) and are representative of at least four experimental replicates. Statistical tests for significant deviation between samples were performed using an unpaired, two-tailed Student's *t*-test.

2.15 References

1   Shafi, R. *et al.* The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny. *Proceedings of the National Academy of Sciences* (2000) **97**, 11:5735-5739.
2   O'Donnell, N., Zachara, N.E., Hart, G.W. & Marth, J.D. Ogt-Dependent X-Chromosome-Linked Protein Glycosylation Is a Requisite Modification in Somatic Cell Function and Embryo Viability. *Molecular and Cellular Biology* (2004) **24**, 4:1680-1690.
3   Liu, K. *et al.* Accumulation of protein O-GlcNAc modification inhibits proteasomes in the brain and coincides with neuronal apoptosis in brain areas with high O-GlcNAc metabolism. *J Neurochem* (2004) **89**, 4:1044-1055.
4   Zhou, Y. *et al.* Interactions between the NR2B receptor and CaMKII modulate synaptic plasticity and spatial learning. *J Neurosci* (2007) **27**, 50:13843-13853.
5   Hayashi, S. & McMahon, A.P. Efficient recombination in diverse tissues by a tamoxifen-inducible form of Cre: a tool for temporally regulated gene activation/inactivation in the mouse. *Dev Biol* (2002) **244**, 2:305-318.
6   Dragatsis, I. & Zeitlin, S. CaMKIIalpha-Cre transgene expression and recombination patterns in the mouse brain. *Genesis* (2000) **26**, 2:133-135.
7   Kennedy, J.L., Farrer, L.A., Andreasen, N.C., Mayeux, R. & St George-Hyslop, P. The genetics of adult-onset neuropsychiatric disease: complexities and conundra? *Science* (2003) **302**, 5646:822-826.
8   Schweizer C, B.S., Bluethmann H, Mansuy IM, Fritschy JM, Mohler H, Lüscher B. The gamma 2 subunit of GABA(A) receptors is required for maintenance of receptors at mature synapses. *Mol Cell Neurosci.* (2003) **24**, 2:442-450.
9   Fekairi, S. *et al.* Human SLX4 is a Holliday junction resolvase subunit that binds multiple DNA repair/recombination endonucleases. *Cell* (2009) **138**, 1:78-89.

10   Garner, E., Kim, Y., Lach, F.P., Kottemann, M.C. & Smogorzewska, A. Human GEN1 and the SLX4-associated nucleases MUS81 and SLX1 are essential for the resolution of replication-induced Holliday junctions. *Cell Rep* (2013) **5**, 1:207-215.

11   Wilson, J.S. *et al.* Localization-dependent and -independent roles of SLX4 in regulating telomeres. *Cell Rep* (2013) **4**, 5:853-860.

12   Navia-Paldanius, D. *et al.* Chemoproteomic, biochemical and pharmacological approaches in the discovery of inhibitors targeting human alpha/beta-hydrolase domain containing 11 (ABHD11). *Eur J Pharm Sci* (2016) **93**, 253-263.

13   Arya, M., Srinivasan, M. & Rajasekharan, R. Human alpha beta hydrolase domain containing protein 11 and its yeast homolog are lipid hydrolases. *Biochem Biophys Res Commun* (2017) **487**, 4:875-880.

14   Barresi, M.J. *et al.* Essential genes for astroglial development and axon pathfinding during zebrafish embryogenesis. *Dev Dyn* (2010) **239**, 10:2603-2618.

15   Francelle, L. *et al.* Striatal long noncoding RNA Abhd11os is neuroprotective against an N-terminal fragment of mutant huntingtin in vivo. *Neurobiol Aging* (2015) **36**, 3:1601 e1607-1616.

16   Osterhout, J.A. *et al.* Cadherin-6 mediates axon-target matching in a non-image-forming visual circuit. *Neuron* (2011) **71**, 4:632-639.

17   Kirk, L.M. *et al.* Distribution of the SynDIG4/proline-rich transmembrane protein 1 in rat brain. *J Comp Neurol* (2016) **524**, 11:2266-2280.

18   Chen, N. *et al.* Interaction proteomics reveals brain region-specific AMPA receptor complexes. *J Proteome Res* (2014) **13**, 12:5695-5706.

19   Rodriguez-Rodriguez, P., Fernandez, E., Almeida, A. & Bolanos, J.P. Excitotoxic stimulus stabilizes PFKFB3 causing pentose-phosphate pathway to glycolysis switch and neurodegeneration. *Cell Death Differ* (2012) **19**, 10:1582-1589.

20   Rodriguez-Rodriguez, P., Almeida, A. & Bolanos, J.P. Brain energy metabolism in glutamate-receptor activation and excitotoxicity: role for APC/C-Cdh1 in the balance glycolysis/pentose phosphate pathway. *Neurochem Int* (2013) **62**, 5:750-756.

21   Schoenberg, D.R. & Maquat, L.E. Regulation of cytoplasmic mRNA decay. *Nat Rev Genet* (2012) **13**, 4:246-259.

22   Katayama, S. *et al.* Antisense transcription in the mammalian transcriptome. *Science* (2005) **309**, 5740:1564-1566.

23   Dickey, C.A. *et al.* Selectively reduced expression of synaptic plasticity-related genes in amyloid precursor protein + presenilin-1 transgenic mice. *J Neurosci* (2003) **23**, 12:5219-5226.

24   Wu, Z.L. *et al.* Comparative analysis of cortical gene expression in mouse models of Alzheimer's disease. *Neurobiol Aging* (2006) **27**, 3:377-386.

25   Ulrich, J.D., Ulland, T.K., Colonna, M. & Holtzman, D.M. Elucidating the Role of TREM2 in Alzheimer's Disease. *Neuron* (2017) **94**, 2:237-248.

26   Van Cauwenberghe, C., Van Broeckhoven, C. & Sleegers, K. The genetic landscape of Alzheimer disease: clinical implications and perspectives. *Genet Med* (2016) **18**, 5:421-430.

27   Sims, R. *et al.* Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. *Nat Genet* (2017) **49**, 9:1373-1384.

28   Ulland, T.K. *et al.* TREM2 Maintains Microglial Metabolic Fitness in Alzheimer's Disease. *Cell* (2017) **170**, 4:649-663 e613.

29   Mahley, R.W. Apolipoprotein E: cholesterol transport protein with expanding role in cell biology. *Science* (1988) **240**, 4852:622-630.

30   Puglielli, L., Tanzi, R.E. & Kovacs, D.M. Alzheimer's disease: the cholesterol connection. *Nat Neurosci* (2003) **6**, 4:345-351.

31   Kotti, T.J., Ramirez, D.M., Pfeiffer, B.E., Huber, K.M. & Russell, D.W. Brain cholesterol turnover required for geranylgeraniol production and learning in mice. *Proc Natl Acad Sci U S A* (2006) **103**, 10:3869-3874.

32   Kotti, T., Head, D.D., McKenna, C.E. & Russell, D.W. Biphasic requirement for geranylgeraniol in hippocampal long-term potentiation. *Proc Natl Acad Sci U S A* (2008) **105**, 32:11394-11399.

33   Luthi-Carter, R. *et al.* SIRT2 inhibition achieves neuroprotection by decreasing sterol biosynthesis. *Proc Natl Acad Sci U S A* (2010) **107**, 17:7927-7932.

34   Bryleva, E.Y. *et al.* ACAT1 gene ablation increases 24(S)-hydroxycholesterol content in the brain and ameliorates amyloid pathology in mice with AD. *Proc Natl Acad Sci U S A* (2010) **107**, 7:3081-3086.

35   Shimano, H. & Sato, R. SREBP-regulated lipid metabolism: convergent physiology - divergent pathophysiology. *Nat Rev Endocrinol* (2017) **advance online publication**.

36   Yang, T. *et al.* Crucial step in cholesterol homeostasis: sterols promote binding of SCAP to INSIG-1, a membrane protein that facilitates retention of SREBPs in ER. *Cell* (2002) **110**, 4:489-500.

37   Anthonisen, E.H. *et al.* Nuclear receptor liver X receptor is O-GlcNAc-modified in response to glucose. *J Biol Chem* (2010) **285**, 3:1607-1615.

38   Yi, W. *et al.* Phosphofructokinase 1 glycosylation regulates cell growth and metabolism. *Science* (2012) **337**, 6097:975-980.

39   Dentin, R., Hedrick, S., Xie, J., Yates, J., 3rd & Montminy, M. Hepatic glucose sensing via the CREB coactivator CRTC2. *Science* (2008) **319**, 5868:1402-1405.

40   Ruan, H.B., Singh, J.P., Li, M.D., Wu, J. & Yang, X. Cracking the O-GlcNAc code in metabolism. *Trends Endocrinol Metab* (2013) **24**, 6:301-309.

41   Rong, S. *et al.* Expression of SREBP-1c Requires SREBP-2-mediated Generation of a Sterol Ligand for LXR in Livers of Mice. *Elife* (2017) **6**, e25015.

42   Wong, J., Quinn, C.M. & Brown, A.J. SREBP-2 positively regulates transcription of the cholesterol efflux gene, ABCA1, by generating oxysterol ligands for LXR. *Biochem J* (2006) **400**, 3:485-491.

43   Hanover, J.A. *et al.* A Caenorhabditis elegans model of insulin resistance: altered macronutrient storage and dauer formation in an OGT-1 knockout. *Proc Natl Acad Sci U S A* (2005) **102**, 32:11266-11271.

44   Perez-Cervera, Y. *et al.* Insulin signaling controls the expression of O-GlcNAc transferase and its interaction with lipid microdomains. *FASEB J* (2013) **27**, 9:3478-3486.

45   Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* (2008) **9**, 559.

46   Langfelder, P. & Horvath, S. Fast R Functions for Robust Correlations and Hierarchical Clustering. *J Stat Softw* (2012) **46**, 11:17.

47   Tait, S.W. & Green, D.R. Mitochondria and cell death: outer membrane permeabilization and beyond. *Nat Rev Mol Cell Biol* (2010) **11**, 9:621-632.

48   Ooms, L.M. *et al.* The role of the inositol polyphosphate 5-phosphatases in cellular function and human disease. *Biochem J* (2009) **419**, 1:29-49.

49   Kam, T.I. *et al.* FcgammaRIIb-SHIP2 axis links Abeta to tau pathology by disrupting phosphoinositide metabolism in Alzheimer's disease model. *Elife* (2016) **5**, e18691.

50   Srinivasan, K. *et al.* Untangling the brain's neuroinflammatory and neurodegenerative transcriptional responses. *Nat Commun* (2016) **7**, 11295.

51   Hokama, M. *et al.* Altered expression of diabetes-related genes in Alzheimer's disease brains: the Hisayama study. *Cereb Cortex* (2014) **24**, 9:2476-2488.

52   Miller, J.A., Woltjer, R.L., Goodenbour, J.M., Horvath, S. & Geschwind, D.H. Genes and pathways underlying regional and cell type changes in Alzheimer's disease. *Genome Med* (2013) **5**, 5:48.

53   Wirz, K.T. *et al.* Cortical beta amyloid protein triggers an immune response, but no synaptic changes in the APPswe/PS1dE9 Alzheimer's disease mouse model. *Neurobiol Aging* (2013) **34**, 5:1328-1342.

54   Hart, G.W., Slawson, C., Ramirez-Correa, G. & Lagerlof, O. Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease. *Annu Rev Biochem* (2011) **80**, 1:825-858.

55   Langfelder, P., Mischel, P.S. & Horvath, S. When is hub gene selection better than standard meta-analysis? *PLoS One* (2013) **8**, 4:e61505.

56   Stefansson, B., Ohama, T., Daugherty, A.E. & Brautigan, D.L. Protein phosphatase 6 regulatory subunits composed of ankyrin repeat domains. *Biochemistry* (2008) **47**, 5:1442-1451.

57   Zhong, J. *et al.* Protein phosphatase PP6 is required for homology-directed repair of DNA double-strand breaks. *Cell Cycle* (2011) **10**, 9:1411-1419.

58   Stefansson, B. & Brautigan, D.L. Protein phosphatase 6 subunit with conserved Sit4-associated protein domain targets IkappaBepsilon. *J Biol Chem* (2006) **281**, 32:22624-22634.

59   Zeng, K., Bastos, R.N., Barr, F.A. & Gruneberg, U. Protein phosphatase 6 regulates mitotic spindle formation by controlling the T-loop phosphorylation state of Aurora A bound to its activator TPX2. *J Cell Biol* (2010) **191**, 7:1315-1332.

60   Ma, X. *et al.* PP6 Disruption Synergizes with Oncogenic Ras to Promote JNK-Dependent Tumor Growth and Invasion. *Cell Rep* (2017) **19**, 13:2657-2664.

61   Bhandari, D. *et al.* Sit4p/PP6 regulates ER-to-Golgi traffic by controlling the dephosphorylation of COPII coat subunits. *Mol Biol Cell* (2013) **24**, 17:2727-2738.

62   Ohama, T., Wang, L., Griner, E.M. & Brautigan, D.L. Protein Ser/Thr phosphatase-6 is required for maintenance of E-cadherin at adherens junctions. *BMC Cell Biol* (2013) **14**, 1:42.

63   Valdez, B.C., Henning, D., So, R.B., Dixon, J. & Dixon, M.J. The Treacher Collins syndrome (TCOF1) gene product is involved in ribosomal DNA gene transcription by interacting with upstream binding factor. *Proc Natl Acad Sci U S A* (2004) **101**, 29:10709-10714.

64   Ciccia, A. *et al.* Treacher Collins syndrome TCOF1 protein cooperates with NBS1 in the DNA damage response. *Proc Natl Acad Sci U S A* (2014) **111**, 52:18631-18636.

65   Larsen, D.H. *et al.* The NBS1-Treacle complex controls ribosomal RNA transcription in response to DNA damage. *Nat Cell Biol* (2014) **16**, 8:792-803.

66   Werner, A. *et al.* Cell-fate determination by ubiquitin-dependent regulation of translation. *Nature* (2015) **525**, 7570:523-527.

67   Gong, D. & Ferrell, J.E., Jr. The roles of cyclin A2, B1, and B2 in early and late mitotic events. *Mol Biol Cell* (2010) **21**, 18:3149-3161.

68   Wang, X., Zhang, C., Szabo, G. & Sun, Q.Q. Distribution of CaMKIIalpha expression in the brain in vivo, studied by CaMKIIalpha-GFP mice. *Brain Res* (2013) **1518**, 9-25.

69   Yang, Y., Mufson, E.J. & Herrup, K. Neuronal cell death is preceded by cell cycle events at all stages of Alzheimer's disease. *Journal of Neuroscience* (2003) **23**, 7:2557-2563.

70   Yang, Y., Varvel, N.H., Lamb, B.T. & Herrup, K. Ectopic cell cycle events link human Alzheimer's disease and amyloid precursor protein transgenic mouse models. *J Neurosci* (2006) **26**, 3:775-784.

71   Yang, Y., Geldmacher, D.S. & Herrup, K. DNA replication precedes neuronal cell death in Alzheimer's disease. *J Neurosci* (2001) **21**, 8:2661-2668.

72   Buchakjian, M.R. & Kornbluth, S. The engine driving the ship: metabolic steering of cell proliferation and death. *Nat Rev Mol Cell Biol* (2010) **11**, 10:715-727.

73   Yalcin, A. *et al.* Nuclear targeting of 6-phosphofructo-2-kinase (PFKFB3) increases proliferation via cyclin-dependent kinases. *J Biol Chem* (2009) **284**, 36:24223-24232.

74   Herrero-Mendez, A. *et al.* The bioenergetic and antioxidant status of neurons is controlled by continuous degradation of a key glycolytic enzyme by APC/C-Cdh1. *Nat Cell Biol* (2009) **11**, 6:747-752.

75   Esteras, N. *et al.* Altered cell cycle-related gene expression in brain and lymphocytes from a transgenic mouse model of Alzheimer's disease [amyloid precursor protein/presenilin 1 (PS1)]. *Eur J Neurosci* (2012) **36**, 5:2609-2618.

76   Weeks, S.D. *et al.* Molecular structure and dynamics of the dimeric human small heat shock protein HSPB6. *J Struct Biol* (2014) **185**, 3:342-354.

77   Wilhelmus, M.M. *et al.* Specific association of small heat shock proteins with the pathological hallmarks of Alzheimer's disease brains. *Neuropathol Appl Neurobiol* (2006) **32**, 2:119-130.

78   Cameron, R.T. *et al.* The phosphorylation of Hsp20 enhances its association with amyloid-beta to increase protection against neuronal cell death. *Mol Cell Neurosci* (2014) **61**, Supplement C:46-55.

79   Hato, T., Tabata, M. & Oike, Y. The role of angiopoietin-like proteins in angiogenesis and metabolism. *Trends Cardiovasc Med* (2008) **18**, 1:6-14.

80   Oike, Y. *et al.* Angiopoietin-related growth factor antagonizes obesity and insulin resistance. *Nat Med* (2005) **11**, 4:400-408.

81   Kang, S.G. *et al.* ANGPTL6 expression is coupled with mitochondrial OXPHOS function to regulate adipose FGF21. *J Endocrinol* (2017) **233**, 1:105-118.

82   Mukherjee, S. *et al.* Genome-Wide Analysis of Amyloid Beta (Ab) Peptide Obtained from Histelide Identifies Suggestive Hits in Rhbdf1, Grid1, and Ptprd Regions in the Adult Changes in Thought (Act) Study. *Alzheimer's & Dementia* (2014) **10**, 4:P317.

83   Raj, T. *et al.* Integrative analyses of splicing in the aging brain: role in susceptibility to Alzheimer's Disease. *bioRxiv* (2017).

84   Wilson-Edell, K.A. *et al.* RPL24: a potential therapeutic target whose depletion or acetylation inhibits polysome assembly and cancer cell growth. *Oncotarget* (2014) **5**, 13:5165-5176.

85   Zhang, J. *et al.* Cdk5 suppresses the neuronal cell cycle by disrupting the E2F1-DP1 complex. *J Neurosci* (2010) **30**, 15:5219-5228.

86   Heneka, M.T., Golenbock, D.T. & Latz, E. Innate immunity in Alzheimer's disease. *Nat Immunol* (2015) **16**, 3:229-236.

87   Folch, J. *et al.* Role of cell cycle re-entry in neurons: a common apoptotic mechanism of neuronal cell death. *Neurotox Res* (2012) **22**, 3:195-207.

88   Odajima, J. *et al.* Cyclin E constrains Cdk5 activity to regulate synaptic plasticity and memory formation. *Dev Cell* (2011) **21**, 4:655-668.

89   Matarin, M. *et al.* A genome-wide gene-expression analysis and database in transgenic mice during development of amyloid or tau pathology. *Cell Rep* (2015) **10**, 4:633-644.

90   Merino-Serrais, P. *et al.* The influence of phospho-tau on dendritic spines of cortical pyramidal neurons in patients with Alzheimer's disease. *Brain* (2013) **136**, Pt 6:1913-1928.

91   Lopes, S. *et al.* Tau protein is essential for stress-induced brain pathology. *Proc Natl Acad Sci U S A* (2016) **113**, 26:E3755-3763.

92   Lagerlof, O., Hart, G.W. & Huganir, R.L. O-GlcNAc transferase regulates excitatory synapse maturity. *Proc Natl Acad Sci U S A* (2017) **114**, 7:1684-1689.

93   Kouzarides, T. Chromatin modifications and their function. *Cell* (2007) **128**, 4:693-705.

94   Chen, Q. & Yu, X. OGT restrains the expansion of DNA damage signaling. *Nucleic Acids Res* (2016) **44**, 19:9266-9278.

95   Akan, I., Love, D.C., Harwood, K.R., Bond, M.R. & Hanover, J.A. Drosophila O-GlcNAcase Deletion Globally Perturbs Chromatin O-GlcNAcylation. *J Biol Chem* (2016) **291**, 19:9906-9919.

96   Lee, J.S. & Zhang, Z. *O*-linked N-acetylglucosamine transferase (OGT) interacts with the histone chaperone HIRA complex and regulates nucleosome assembly and cellular senescence. *Proc Natl Acad Sci U S A* (2016) **113**, 23:E3213-3220.

97   Ince-Dunn, G. *et al.* Neuronal Elav-like (Hu) proteins regulate RNA splicing and abundance to control glutamate levels and neuronal excitability. *Neuron* (2012) **75**, 6:1067-1080.

98   Fragkouli, A. *et al.* Neuronal ELAVL proteins utilize AUF-1 as a co-partner to induce neuron-specific alternative splicing of APP. (2017) **7**, 44507.

99   Park, S.K. *et al.* A conserved splicing silencer dynamically regulates *O*-GlcNAc transferase intron retention and *O*-GlcNAc homeostasis. *Cell Rep* (2017) **20**, 5:1088-1099.

100   Bertram, L. *et al.* Evidence for genetic linkage of Alzheimer's disease to chromosome 10q. *Science* (2000) **290**, 5500:2302-2303.

101   Twine, N.A., Janitz, K., Wilkins, M.R. & Janitz, M. Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease. *PLoS One* (2011) **6**, 1:e16266.

102   McKay, S.L. & Johnson, T.L. A bird's-eye view of post-translational modifications in the spliceosome and their roles in spliceosome dynamics. *Mol Biosyst* (2010) **6**, 11:2093-2102.

103   Shan, X.Y. *O-linked N-acetylglucosamine protein modification in mouse models of neurodegenerative diseases* Ph.D. thesis, Simon Fraser University, (2011).

104   De, I. *et al.* The RNA helicase Aquarius exhibits structural adaptations mediating its recruitment to spliceosomes. *Nat Struct Mol Biol* (2015) **22**, 2:138-144.

105   Zhu, X. *et al.* Spatiotemporal expression of KHSRP modulates Schwann cells and neuronal differentiation after sciatic nerve injury. *Int J Biochem Cell Biol* (2014) **48**, Supplement C:1-10.

106   Zhang, Z. *et al.* The helicase DDX41 senses intracellular DNA mediated by the adaptor STING in dendritic cells. *Nat Immunol* (2011) **12**, 10:959-965.

107   Su, C. & Schwarz, T.L. O-GlcNAc Transferase Is Essential for Sensory Neuron Survival and Maintenance. *J Neurosci* (2017) **37**, 8:2125-2136.

108   Oddo, S. *et al.* Triple-transgenic model of Alzheimer's disease with plaques and tangles: intracellular Aβ and synaptic dysfunction. *Neuron* (2003) **39**, 3:409-421.

109   Liu, F., Iqbal, K., Grundke-Iqbal, I., Hart, G.W. & Gong, C.X. O-GlcNAcylation regulates phosphorylation of tau: a mechanism involved in Alzheimer's disease. *Proc Natl Acad Sci U S A* (2004) **101**, 29:10804-10809.

110   Yuzwa, S.A. *et al.* Increasing O-GlcNAc slows neurodegeneration and stabilizes tau against aggregation. *Nat Chem Biol* (2012) **8**, 4:393-399.

111   Jacobsen, K.T. & Iverfeldt, K. O-GlcNAcylation increases non-amyloidogenic processing of the amyloid-beta precursor protein (APP). *Biochem Biophys Res Commun* (2011) **404**, 3:882-886.

112   Kim, C. *et al.* O-linked beta-N-acetylglucosaminidase inhibitor attenuates beta-amyloid plaque and rescues memory impairment. *Neurobiol Aging* (2013) **34**, 1:275-285.

113   Oakley, H. *et al.* Intraneuronal β-amyloid aggregates, neurodegeneration, and neuron loss in transgenic mice with five familial Alzheimer's disease mutations: potential factors in amyloid plaque formation. *Journal of Neuroscience* (2006) **26**, 40:10129-10140.

114   Götz, J., Chen, F., Van Dorpe, J. & Nitsch, R. Formation of neurofibrillary tangles in P301L tau transgenic mice induced by Aβ42 fibrils. *Science* (2001) **293**, 5534:1491-1495.

115   Slawson, C. *et al.* Perturbations in O-linked β-N-acetylglucosamine protein modification cause severe defects in mitotic progression and cytokinesis. *Journal of Biological Chemistry* (2005) **280**, 38:32944-32956.

116   Capotosti, F. *et al.* O-GlcNAc transferase catalyzes site-specific proteolysis of HCF-1. *Cell* (2011) **144**, 3:376-388.

117   Ngoh, G.A., Watson, L.J., Facundo, H.T. & Jones, S.P. Augmented O-GlcNAc signaling attenuates oxidative stress and calcium overload in cardiomyocytes. *Amino Acids* (2011) **40**, 3:895-911.

118   Ohn, T., Kedersha, N., Hickman, T., Tisdale, S. & Anderson, P. A functional RNAi screen links O-GlcNAc modification of ribosomal proteins to stress granule and processing body assembly. *Nature cell biology* (2008) **10**, 10:1224-1231.

119   Herrup, K. & Yang, Y. Cell cycle regulation in the postmitotic neuron: oxymoron or new biology? *Nat Rev Neurosci* (2007) **8**, 5:368-378.

120   Zhang, J. *et al.* Nuclear localization of Cdk5 is a key determinant in the postmitotic state of neurons. *Proceedings of the National Academy of Sciences* (2008) **105**, 25:8772-8777.

121   Ning, X. *et al.* The O-GlcNAc Modification of Cdk5 Involved in Neuronal Apoptosis Following In Vitro Intracerebral Hemorrhage. *Cellular and Molecular Neurobiology* (2016) 1-10.

122   Zhang, J., Li, H., Zhou, T., Zhou, J. & Herrup, K. Cdk5 levels oscillate during the neuronal cell cycle: Cdh1 ubiquitination triggers proteosome-dependent degradation during S-phase *Journal of Biological Chemistry* (2012) **287**, 31:25985-25994.

123   Xu, X. *et al.* Prevention of beta-amyloid induced toxicity in human iPS cell-derived neurons by inhibition of Cyclin-dependent kinases and associated cell cycle events. *Stem Cell Res* (2013) **10**, 2:213-227.

124   Lagerlöf, O. *et al.* The nutrient sensor OGT in PVN neurons regulates feeding. *Science* (2016) **351**, 6279:1293-1296.

125   Tsien, J.Z. *et al.* Subregion- and Cell Type-Restricted Gene Knockout in Mouse Brain. *Cell* (1996) **87**, 7:1317-1326.

126   Wang, P. *et al.* O-GlcNAc cycling mutants modulate proteotoxicity in Caenorhabditis elegans models of human neurodegenerative diseases. *Proc Natl Acad Sci U S A* (2012) **109**, 43:17669-17674.

127   Dickey, C.A. *et al.* Selectively reduced expression of synaptic plasticity-related genes in amyloid precursor protein+ presenilin-1 transgenic mice. *Journal of Neuroscience* (2003) **23**, 12:5219-5226.

128   Scripture-Adams, D.D. *et al.* GATA-3 dose-dependent checkpoints in early T cell commitment. *J Immunol* (2014) **193**, 7:3470-3491.

129   Haubensak, W. *et al.* Genetic dissection of an amygdala microcircuit that gates conditioned fear. *Nature* (2010) **468**, 7321:270-276.

130   Wojtowicz, J.M. & Kee, N. BrdU assay for neurogenesis in rodents. *Nat Protoc* (2006) **1**, 3:1399-1405.

131   Hellemans, J., Mortier, G., De Paepe, A., Speleman, F. & Vandesompele, J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol* (2007) **8**, 2:R19.

132   Dudoit, S., Yang, Y., Callow, M. & Speed, T. Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica Sinica* (2002) 111-139.

133   Smyth, G.K. in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* (eds Robert Gentleman *et al.*) 397-420 (Springer New York, 2005).

134   GO.db: A set of annotation maps describing the entire Gene Ontology v. Version 3.4.1 (2017).

135   Hu, Z. *et al.* VisANT 4.0: Integrative network platform to connect genes, drugs, diseases and therapies. *Nucleic Acids Res* (2013) **41**, Web Server issue:W225-231.

136   Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* (2003) **13**, 11:2498-2504.

137   Bindea, G. *et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* (2009) **25**, 8:1091-1093.

138   Bindea, G., Galon, J. & Mlecnik, B. CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. *Bioinformatics* (2013) **29**, 5:661-663.

Development of biological and chemical tools for discovery of the OGT interactome and *O*-GlcNAcome

3.1 Abstract

The intracellular post-translational modification of serine or threonine residues of proteins with a single N-acetylglucosamine monosaccharide (O-GlcNAcylation) is essential for neuronal homeostasis and a variety of cellular processes. This modification is cycled by only two enzymes in mammalian cells: O-GlcNAc transferase (OGT) and O-GlcNAcase (OGA). OGT contains several protein-binding TPR domains, which allow it to bind to specific interactors. A prevalent hypothesis in the field suggests that OGT is targeted OGT to specific substrates by certain key interactors. In order to identify these interactors, we used CRISPR/Cas9 to develop a novel mouse with a minimally tagged OGT in order to identify the OGT brain interactome using tandem affinity purification and proteomic methods. The preliminary OGT brain interactome showed agreement with previous OGT interactome studies and O-GlcNAcome studies although these studies were performed in different species and cell types. The identified OGT interactors are enriched for ribosomal and cytoskeletal proteins in addition to axonal, dendritic, and neuronal cell body proteins.

In addition to the OGT interactome, we sought to uncover OGT substrates in order to fully test the OGT interactor/substrate hypothesis. OGT modifies over 1000 different proteins, but the lack of a well-defined consensus sequence has hampered efforts to predict sites a priori. Furthermore, relatively few O-GlcNAc modification sites have been mapped due to the difficulty of enriching and detecting O-GlcNAcylated peptides from complex samples. Here, we describe an improved approach to quantitatively label and enrich O-GlcNAcylated proteins for site identification. Chemoenzymatic labeling followed by copper(I)-catalyzed azide-alkyne cycloaddition

(CuAAC) installs a new mass spectrometry (MS)-compatible linker designed for facile purification and release of O-GlcNAcylated proteins for downstream MS analysis. We validate the approach by unambiguously identifying several established O-GlcNAc sites on the proteins α-crystallin and O-GlcNAc transferase (OGT) as well as discovering new, previously unreported sites on both proteins. Notably, these novel sites on OGT lie in key functional domains of the protein, underscoring how this site identification method can reveal important biological insights into protein activity and regulation.

3.2 Overview of OGT interactome and *O*-GlcNAcome approach

The enzymes OGT and OGA cycle *O*-GlcNAc modification on over 1000 proteins in mammalian cells.[1] Although the recent crystal structures of OGT have augmented our understanding of the structural constraints imposed by the active site of OGT, OGT appears to lack a well-defined consensus sequence, which has hampered efforts to predict substrates and their sites.[2] Without a well-defined consensus sequence or site prediction, the choice of OGT substrate remains poorly understood. Several studies have demonstrated that key OGT interactors can effectively recruit OGT to specific substrates.[3-5] This has lead to the OGT interactor/substrate hypothesis whereby the protein-binding TPR (tetratricopeptide repeat) domain of OGT interacts with certain interactors, which in turn mediate the *O*-GlcNAcylation of specific target proteins.[6] While this hypothesis has been shown for specific substrate-interactor pairs, this hypothesis has not been tested in a global proteome-wide fashion. Toward that end, we have developed chemical and biological tools in order to identify the global OGT interactome and *O*-GlcNAcome.

In order to identify the OGT interactome, we sought to develop a biological system that would enable facile pull down of OGT to identify physiologically-relevant interactors (Figure 3.1). Through the utilization of Crispr/Cas9 genome editing methods, we installed two small FLAG and HA tags upon the C-terminus of OGT in mice for tandem affinity purification. There are multiple advantages of our approach: (1) tandem affinity purification methods have been shown to successfully enrich interactors with minimal off-target effects compared to single enrichment methods; (2) small tags are likely to minimally perturb interactors and OGT function; and (3) by labeling endogenous OGT instead of previously employed overexpression/knockdown strategies or yeast two-hybrid assays, we endeavor to maximize endogenous interactors and avoid false positive interactors.[7] It is important to note that tandem affinity purification will identify strong and stable interactions rather than transient interactions.

In addition to the OGT interactome, we also needed to identify OGT's substrates (the *O*-GlcNAcome). We employed chemoenzymatic-labeling methods in order to quantitatively label and then enrich *O*-GlcNAcylated proteins and peptides for site identification. Using this approach, a GalNAz group was appended to *O*-GlcNAcylated proteins using a mutant enzyme. Then, a novel biotin-Dde-alkyne linker was covalently appended to the GalNAz handle (Dde = 1-(4,4-dimethyl-2,6-dioxocyclohex-1-ylidene)ethyl). Now biotinylated *O*-GlcNAcylated proteins were enriched using streptavidin beads and eluted using mild hydrazine cleavage of the Dde group. The remaining amine moiety conferred a positive charge to peptides facilitating mass spectrometric identification. Through this method, we were able to identify known and novel sites on the *O*-GlcNAcylated proteins, α-crystallin and OGT itself. Finally, current

efforts are underway to expand the *O*-GlcNAcome approach to identify novel *O*-GlcNAc sites across the proteome in the brain (Figure 3.1). Once the *O*-GlcNAcome and OGT interactome have been identified, we can determine the important hub interactors as well as key substrates and interactors in different tissue-specific contexts. Importantly, with the OGT-FH mouse for OGT interactome enrichment, these tools and methods have broad applicability to different tissues, developmental time points, disease states (when OGT-FH mice are bred with other disease mouse models), or physiological relevant stimuli.
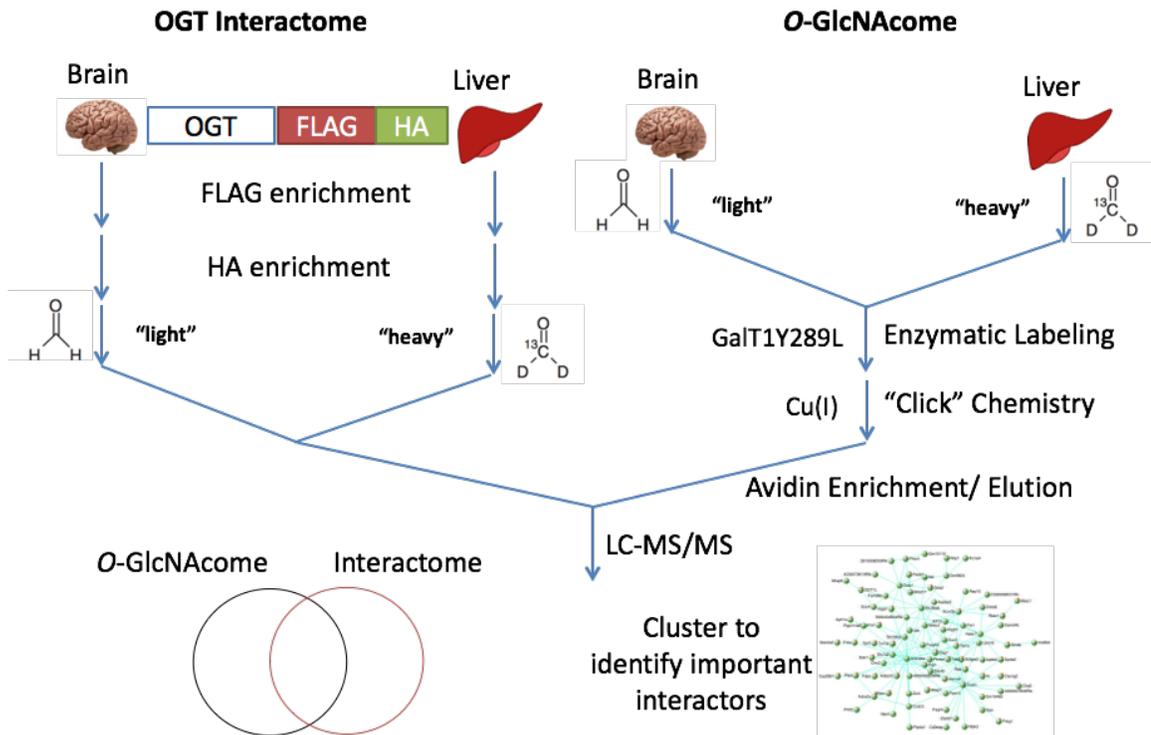
Figure 3.1 Overview of tissue-specific OGT interactome and *O*-GlcNAcome dual approach. Shown is the overall workflow that we have been developing in order to identify the OGT interactome and *O*-GlcNAcome (OGT substrates). Briefly, we tag endogenous OGT with the small FLAG and HA tags and perform tandem affinity purification followed by mass spectrometry (left side). In parallel, the *O*-GlcNAcome is obtained through differential sample labeling, chemoenzymatic labeling, avidin enrichment, and a mild elution prior to mass spectrometry (right side). After these dual analyses, the *O*-GlcNAcome and OGT interactome can be integrated to identify critical hub interactors and determine the differential regulation of *O*-GlcNAc across different tissues.

3.3 Development of biological tools for identifications of the OGT interactome

3.3.1 Validation of tandem affinity purification and C-terminal tagged OGT using OGT activity assay

Prior to generating a mouse with the FLAG-HA tag inserted into OGT, we first wanted to validate the tandem affinity purification and that the tags would not interfere with OGT activity. In order to check the first criterion, we used HEK293T cells that had been treated with lentivirus that inserted OGT-FLAG-HA (C-terminal tagged OGT) downstream of a doxycycline-inducible promoter. The workflow of the tandem affinity purification is shown in Figure 3.2. Briefly, we lysed cells or tissue expressing either OGT (wildtype or -dox) or OGT-FLAG-HA (OGT-FH or +dox) and enriched the OGT-FH using anti-FLAG beads. After washing and elution with a 3xFLAG tag peptide, we then enriched the OGT-FH with anti-HA beads, washed, and eluted using 3M NaSCN (Figure 3.3C). Prior to cell lysis, we also explored the usage of crosslinking techniques in order to obtain the transient OGT interactome (Figure 3.3A-B). While the crosslinking method appeared to pull down Nup62, a known interactor and substrate of OGT, we decided to forego the crosslinking in order to simplify the workflow and also focus on the strong, stable interactors.
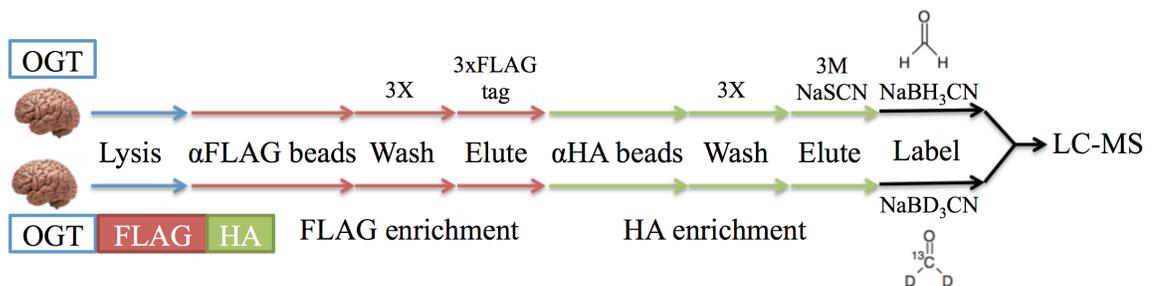
Figure 3.2 Workflow of tandem affinity purification for OGT interactome identification. In order to verify selective OGT-FH enrichment, we used HEK293T cells that would express OGT-FH in response to doxycycline treatment.
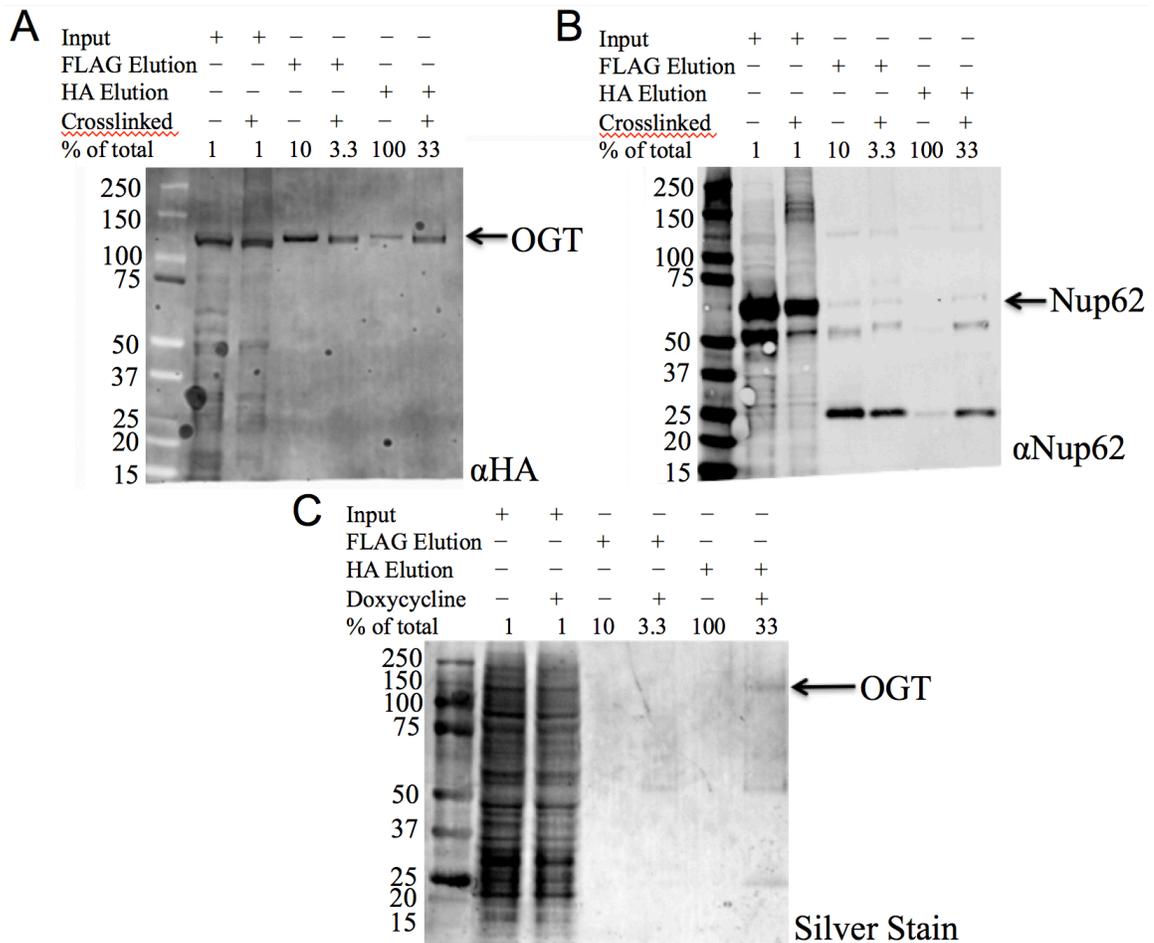
Figure 3.3 Validation of crosslinking conditions and doxycycline induction of OGT-FH expression. (A) We showed pull down of OGT itself with and without crosslinking using a HA antibody. (B) Nup62, a known substrate and interactor with OGT, is enriched in crosslinking conditions. (C) Using a silver stain, we show that we are able to pull down OGT using our tandem affinity purification methods only in the doxycycline-induced expression conditions but not without doxycycline.

In addition to validating the efficiency of tandem affinity purification, we sought to compare the activity of N-terminal tagged OGT to the C-terminal tagged OGT. The active site of OGT is closer to the C-terminus while the N-terminus of OGT contains the protein-protein interacting TPR domains. Toward that end, we transfected HEK293T cells with C-terminal and N-terminal tagged OGT, purified the OGT with a FLAG pull down, and then performed an activity assay. We observed no difference in the activity of the N-terminal and C-terminal tagged OGT as measured by an *in vitro* OGT activity assay (Figure 3.4). Importantly, the N-terminus of OGT undergoes differential splicing to

form mitochondrial (mOGT, 11.5 TPR domains) and short OGT (sOGT, 4.5 TPR domains). This would mean that an N-terminal tag would only tag the full-length isoform of OGT (13.5 TPR domains) while leaving the sOGT and mOGT untagged. We decided to move forward with tagging OGT on the C-terminus as this strategy would (1) tag all isoforms of OGT thereby ensuring the complete OGT interactome would be enriched and (2) not interfere with OGT activity.
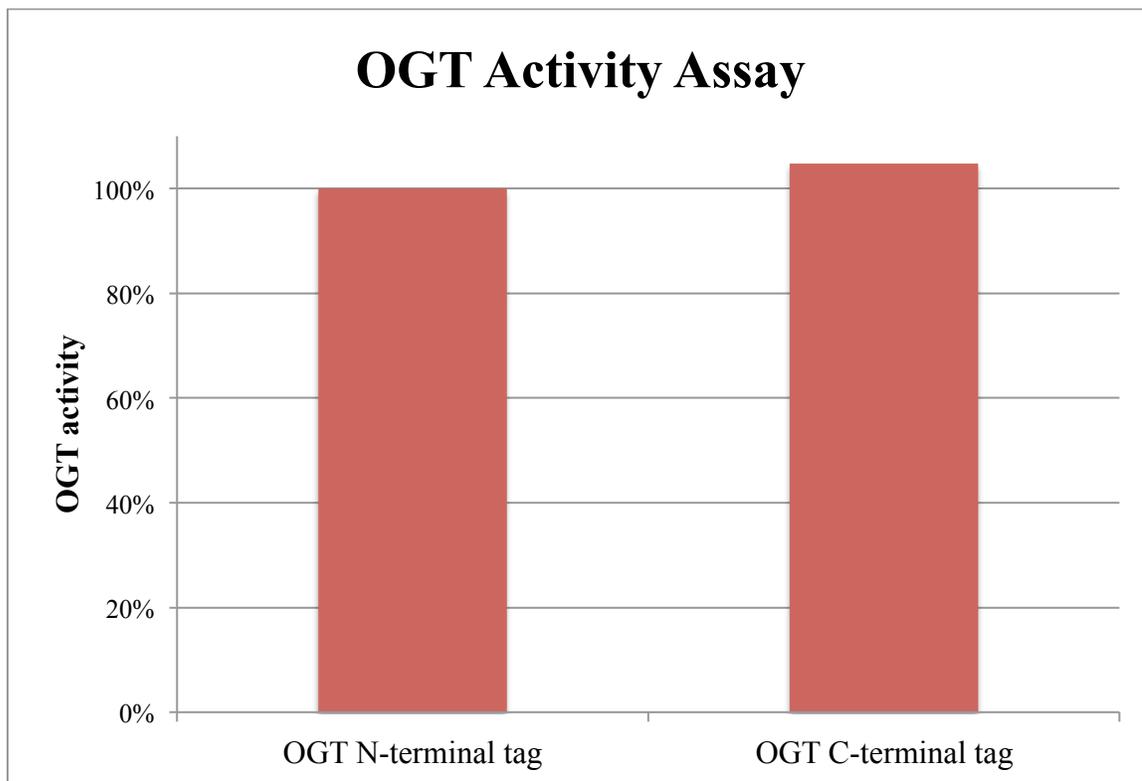
Figure 3.4 FLAG-HA C-terminal and N-terminal tagging of OGT does not affect OGT activity. Using the UDP-Glo™ Glycosyltransferase Assay from Promega, we showed that there was no discernible difference in the OGT activity. The average fluorescence was normalized to the OGT levels as measured by Western.

3.3.2 Validation of OGT targeting sgRNA for CRISPR/Cas9

Using the guide RNA program from the Feng Zhang, we obtained five sgRNA candidates that we then tested for cutting and homologous recombination efficiency.[8] Briefly, the guide RNA needs to have a requisite 3'-protospacer-adjacent motif (PAM) sequence (N(G or A)G) that allows the sgRNA and Crispr/Cas9 machinery to bind

efficiently to the cut site and ideally, good sequence homology for the target site while having low off-target cut site homology across the entire genome. We cloned the C-terminus of mouse OGT into the pCAG-EGxxFP construct that would produce EGFP upon dsDNA cleavage and successful homologous recombination (Figure 3.5). All five of the sgRNA sequences were efficient based on this assay, but we decided to use the highest rated (first) sgRNA as predicted by Feng Zhang lab's program that maximizes on-target cleavage while minimizing off-targets cleavage $[S_{guide} = 87\% = 100/[100 - \Sigma(i = 1 \rightarrow n_{mm})\ S_{hit}(h_i)]]$ (Figure 3.5).[8]
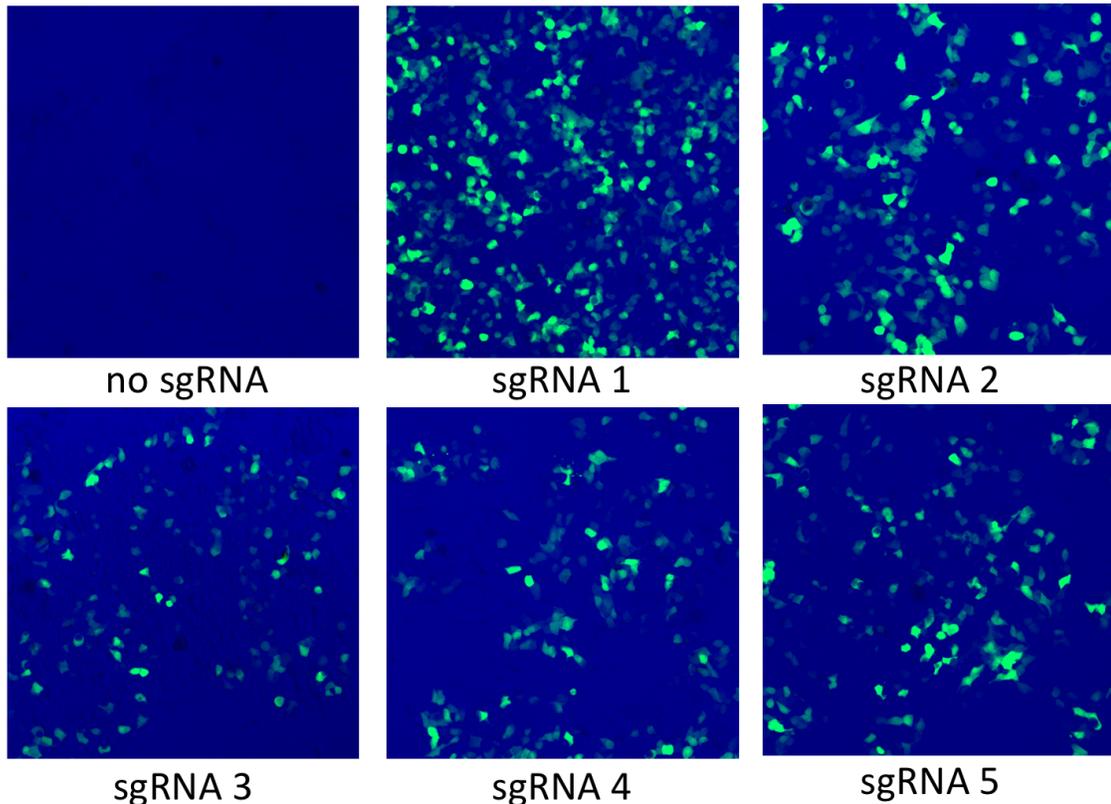
Figure 3.5 Screening of sgRNAs for efficiency. The top five sgRNA candidates were cloned into the pX330 plasmid while the C-terminus OGT target was cloned into the pCAG-EGxxFP construct. Then, the pX330-sgRNA and pCAG-EG_ogt_FP constructs were co-transfected into HEK293T cells. Upon CRISPR/Cas9/sgRNA mediated dsDNA cleavage and HDR, the cells fluoresced, indicating that all five sgRNA candidates were competent. Figure from Matt Griffin.

3.3.3 CRISPR/Cas9 to make novel OGT-FLAG-HA mice

After verifying the proper targeting of the sgRNA, we sought to design the ssODN used for homologous recombination to insert the short tags into the C-terminus of OGT. Based on previous studies exploring the efficacy of homologous recombination for small inserts following Crispr/Cas9 cleavage and our own needs, we sought to design the ssODN with the following criteria: (1) at least 60 bp of OGT homology arms on either side of the cut site, (2) inclusion of two novel cut sites for facilitation of PCR genotyping after incorporation, (3) codon optimization in order to ensure maximal translational efficiency, (4) removal of PAM sequence in the final construct (to prevent recutting by Crispr/Cas9), and (5) reintroduction of a stop codon to terminate the OGT sequence.[9] The final ssODN sequence along with its relevant features is shown in Figure 3.6.
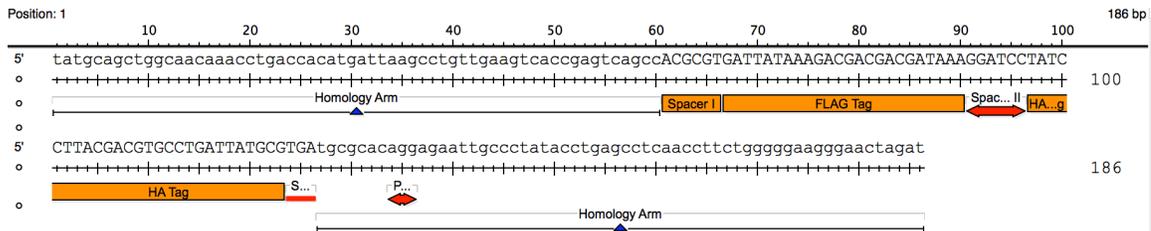
Figure 3.6 Sequence for ssODN for homologous recombination. The single-stranded DNA oligonucleotide was designed with the following components (from 5' to 3'): (1) 60 bases homologous to the C-terminus of OGT prior to the sgRNA cut site, (2) Spacer I (MluI cut site), (3) FLAG tag, (4) Spacer II (BamHI cut site), (5) HA tag, (6) stop codon, (7) 60 bases homologous to the region 3' to the original stop codon of OGT with the PAM sequence mutated out to prevent recutting by the CRISPR/Cas9/sgRNA system after proper HDR insertion.

In order to perform Crispr/Cas9 genome editing in mice, we followed the procedure previously published by Jaenisch and colleagues, where ssODN, Cas9 mRNA, and the sgRNA were directly injected into fertilized mouse C57B/L6 wildtype embryos.[9] Using pronuclear and cytoplasmic injections, only a single female mouse was heterozygous for the correct on-target sequence out of 10 total mice that were born although 4 died shortly after birth. This deviated significantly from the experimental efficiencies reported by Jaenisch and colleagues (30-50% efficiency).[9] Upon closer

examination, the Crispr/Cas9 targeting produced insertions and deletions in 5/10 mice (~5/12 chromosomes we were able to test) although only one had the appropriate insertion. This overall efficiency suggests that the Crispr/Cas9 targeting and cleavage had a high efficiency (~42%) while the homologous recombination efficiency was inefficient (<10%). Thus, the ssODN could have been optimized potentially with longer homology arms for example in order to maximize successful incorporation and reach the Jaenisch and colleagues reported efficiency (30-50%) for small insertion incorporation. After breeding this heterozygous female, we obtained the first group of homozygous mice (all males initially due to the OGT's location on the X chromosome). These homozygous mice exhibited no abnormalities in growth, behavior, or breeding and were indistinguishable from wildtype and heterozygous mice, consistent with our ultimate goal of integrating a minimally disruptive small tag. We verified the genotypes of the mice using on-target sequencing and on-target PCR (Figures 3.7 & 3.8). Importantly, the program provides the most likely off-target sites, which, in the case of sgRNA 1, was within a gene called *Rhox11*. We verified the proper sequence of *Rhox11* around the predicted sgRNA off-target site using sequencing (Figure 3.9). Finally, we verified expression of OGT-FLAG-HA protein in brain tissue of the adult homozygous and heterozygous mice using Western blot (Figure 3.10). In conclusion, we were able to generate a novel OGT-FLAG-HA (OGT-FH) mouse with the desired on-target incorporation and without off-target CRISPR/Cas9 mutation.
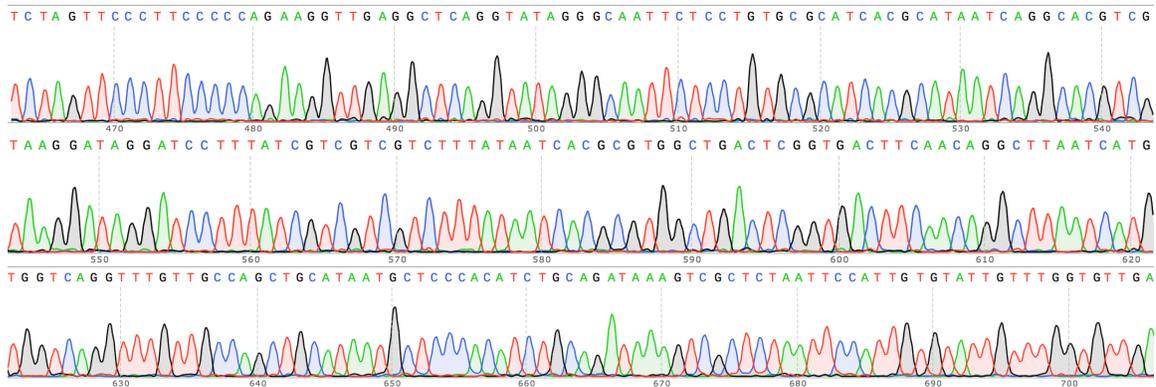
OGTFH7
GTCTGCAACACAGAACTACACAGATACACACAAAGCATGTTATCTAGTTCCCTTCCCCCA 480
ssODN ---atctagttccttcccca 19
                   ****

OGTFH7
GAAGGTTGAGGCTCAGGTATAGGGCAATTCTCCTGTGCGCATCACGCATAATCAGGCACG 540
ssODN gaaggttgaggctcaggtatagggcaattctcctgtgcgcaTCACGCATAATCAGGCACG 79
      ****

OGTFH7
TCGTAAGGATAGGATCCTTTATCGTCGTCGTCTTTATAATCACGCGTGGCTGACTCGGTG 600
ssODN TCGTAAGGATAGGATCCTTTATCGTCGTCGTCTTTATAATCACGCGTggctgactcggtg 139
      ****

OGTFH7
ACTTCAACAGGCTTAATCATGTGGTCAGGTTTGTTGCCAGCTGCATAATGCTCCCACATC 660
ssODN acttcaacaggcttaatcatgtggtcaggtttgttgccagctgcata------------- 186
      ****

Legend
3' C-terminal OGT homology arm `Stop codon` `HA tag` `BamHI cut site` `FLAG tag` `MluI cut site` 5' C-terminal OGT homology arm

Figure 3.7 Sequencing validation of tag insert. After one round of breeding, we obtained homozygous mice and were able to verify by sequencing the proper insertion of the tag. Shown here is the autoscaled sequencing results for the first homozygous mouse produced in SnapGene® Viewer 4.0.5. The homology alignment was performed using Clustal Omega version 1.2.4.
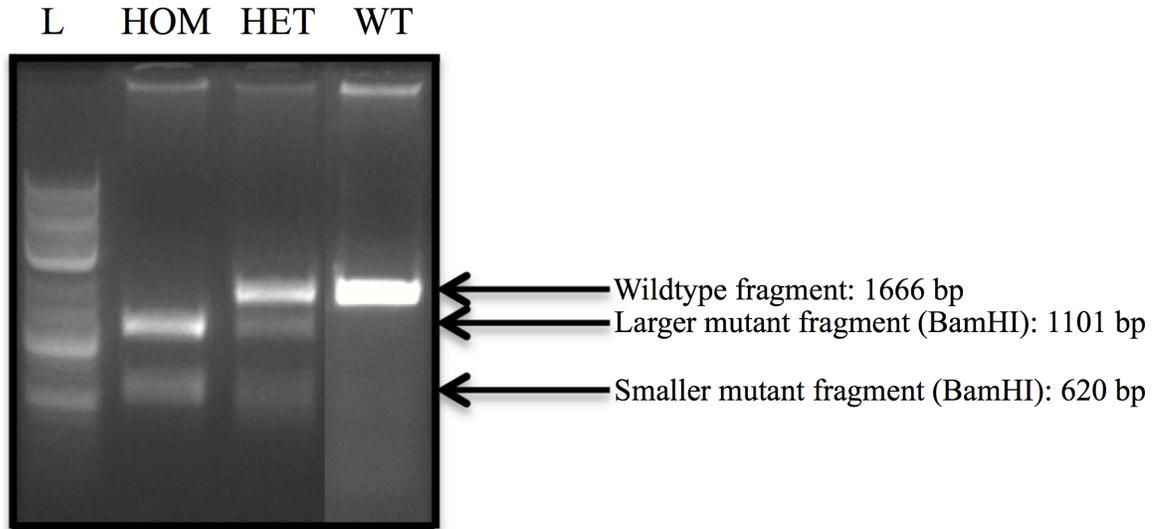
Figure 3.8 PCR and gel genotyping. We designed primers that amplified the regions surrounding the inserted tag. Then, the amplified DNA was cleaved using the BamHI restriction enzyme and then run on a 1% DNA agarose gel. Sample genotyping results are shown here where the fragment sizes are (1) 1666 bp for the wildtype, (2) 1101 bp for the larger mutant fragment, and (3) 620 bp for the smaller mutant fragment. As indicated, homozygous (HOM) mice have 2 distinct bands (2 and 3), heterozygous (HET) mice have three distinct bands (1, 2, and 3), and the wildtype (WT) mice have a single band (1).


```
off_F         actgattttctacgaacactgcttgcagttcccacagctgtccaggtgtaaatctgtatg         8100         2_F
ACTGATTTTCTACGAACACTGCTTGCAGTTCCCACAGCTGTCCAGGTGTAAATCTGTATG 571
     ***********************************************************                      off_F
ctttccgaggaatcctgaataacgactgctcactggtctcggccaccttactggtctcct         8220         2_F
CTTTCCGAGGAATCCTGAATAACGACTGCTCACTGGTCTCGGCCACCTTACTGGTCTCCT 631
     ***********************************************************
off_F ctggctcttctcgcccagtgtcctgagtgttaggcatgacacagctatgctcgcctatag 8280
2_F    CTGGCTCTTCTCGCCCAGTGTCCTGAGTGTTAGGCATGACACAGCTATGCTCGCCTATAG
751
          **********************************************************
```

Figure 3.9 Sequencing validation of most likely off-target site. We validated that the most likely off-target site had no discernible insertions or deletions using the original heterozygous OGT-FH we obtained. The most likely target site is highlighted in yellow. Shown here is the autoscaled sequencing results for the first homozygous mouse produced in SnapGene® Viewer 4.0.5. The homology alignment was performed using Clustal Omega version 1.2.4.
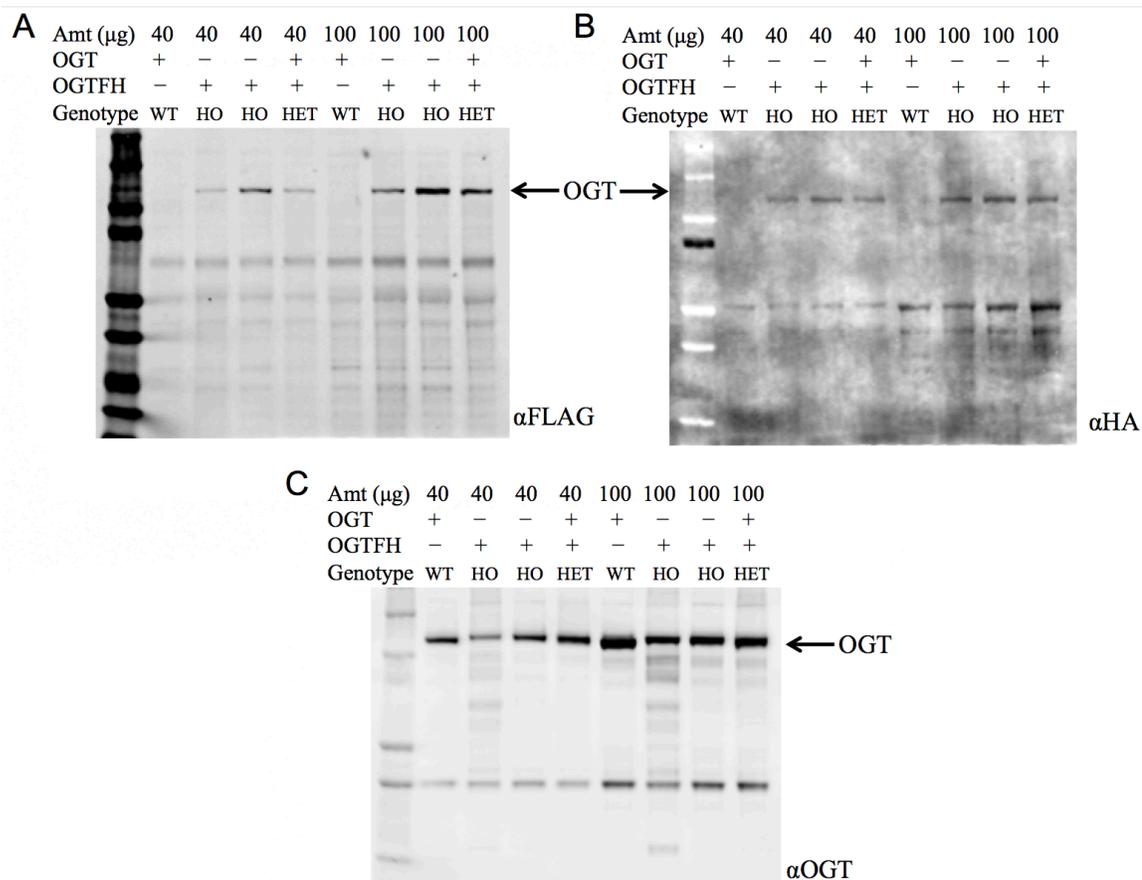
Figure 3.10 Western blotting verification of OGT-FLAG-HA protein expression in mice. We took 40 μg and 100 μg of adult (2 month old) mouse brain lysate of homozygous (HO), heterozygous (HET), and wildtype (WT) littermates. The homozygous and heterozygous, but not wildtype mice show expression of FLAG-tagged OGT (A) and HA-tagged OGT (B). (C) Homozygous, heterozygous, and wildtype mice showed expression of OGT.

3.3.4 OGT interactome preliminary results

After verification and validation of our OGT-FLAG-HA mouse, we began to optimize lysis conditions in order to ensure lysis of nuclear proteins while maintaining protein-protein interactions. We found that 0.5-1% (vol/vol) NP-40 and Triton X-100 were both capable of lysing brain tissue with douncing (Figure 3.11). After optimizing the tandem affinity pull down followed by differential labeling of the OGT and OGT-FH peptides, we were able to get a preliminary list of 76 proteins in the brain OGT interactome using a cutoff SILAC ratio of at least 5.0 (data from Yao Xiao). Shown below is the gene ontology enrichment for this initial list of 76 proteins (Table 3.1). The

proteins were enriched for phosphorylated (68), acetylated (32), methylated (21), and ubiquitylated (14) proteins consistent with the literature showing considerable crosstalk between *O*-GlcNAc glycosylation and other post-translational modifications.[1,10,11] Furthermore, the interactors were enriched for cytoskeletal proteins including those previously reported to be *O*-GlcNAc glycosylated such as bassoon (BSN); piccolo (PCLO); microtubule associated proteins 1B, 2, and 6 (MAP1B, MAP2, MAP6); α-internexin (INA); tubulin α1c, β2A, and β3 chains (TUBA1C, TUBB2A, TUBB3); and actin β (ACTB).[12] In a similar vein, we found several proteins involved in cytoskeletal dynamics that are known to be *O*-GlcNAc glycosylated: α-actinin 1 (ACTN1); spectrin α and β chains (SPTAN1, SPTBN1); myosin-10 (MYH10); unconventional myosin-Va (MYO5A); trafficking kinesin-binding proteins 1 and 2 (TRAK1, TRAK2); ankyrin-3 (ANK3); dynamin-1 (DNM1); and erythrocyte membrane protein band 4.1-like proteins 1 and 3 (EPB41L1, EPB41L3).[13-16] As discussed in Chapter 1, OGT has been shown to interact with and *O*-GlcNAcylate TRAK1/Milton, leading to changes in mitochondrial motility and OGT substrate *O*-GlcNAcylation.[17,18]
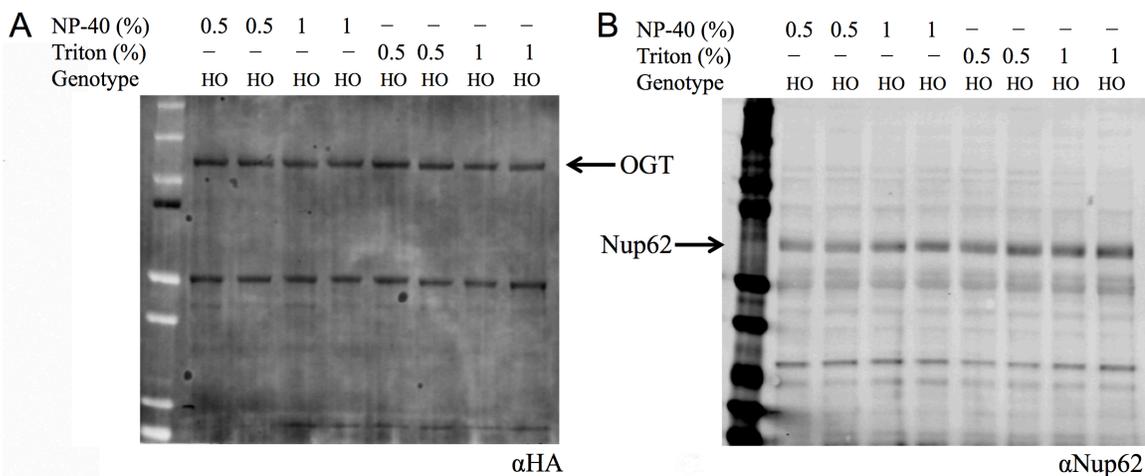
Figure 3.11 Lysis condition screen for brain tissue. We looked at the ability of different mild lysis conditions to lyse OGT and also the nuclear protein and known OGT interactor and substrate, Nup62.

Table 3.1 Preliminary OGT interactome from OGT-FH brain.

GO term	Proteins	FDR
Phosphoprotein	Atp1b1, Srcin1, Tubb2a, Bap1, Hsph1, Atp2b4, Ash2l, Ank3, Wdr77, Stk39, Agap2, Nsf, Tubb3, Jakmip1, Ina, Caskin1, Rbbp5, Arhgef2, Slc25a4, Pfkp, Wnk1, Bsn, Pfkm, Pclo, Mfn2, Mapk1, Ksr2, Atp5c1, Bin1, Carm1, Hcfc1, Klc2, Brsk1, Cxxc1, Camkv, Tpi1, Klc1, Ppp1r12a, Ppp3cb, Camk2b, Ppp3ca, Hspa5, Clasp2, Gapdh, Iqsec1, Tubb4a, Rtcb, Tubb4b, Plp1, Ncdn, Immt, Map1b, Csnk2b, Atp1a3, Atp1a2, Dpysl2, Ywhah, Map2, Hivep3, Pspc1, Ywhaq, Hivep2, Ahcyl1, Hivep1, Atp5a1, Map6, Hnrnph1, Dnm1	1.8×10^{-21}
Myelin sheath	Ina, Plp1, Atp1b1, Slc25a4, Slc25a5, Immt, Atp1a3, Dpysl2, Atp1a2, Slc25a12, Atp5c1, Atp5a1, Hspa5, Gapdh, Dnm1, Nsf, Tubb4a, Tubb4b	5.5×10^{-16}
Cytoplasm	Srcin1, Tubb2a, Hcfc1, Bap1, Brsk1, Klc2, Hsph1, Klc1, Ank3, Wdr77, Ppp1r12a, Camk2b, Stk39, Clasp2, Hspa5, Agap2, Gapdh, Iqsec1, Nsf, Tubb4a, Tubb3, Rtcb, Jakmip1, Tubb4b, Caskin1, Arhgef2, Ncdn, Map1b, Wnk1, Pfkp, Bsn, Dpysl2, Pfkm, Mapk1, Tet3, Ksr2, Ywhah, Map2, Hivep3, Pspc1, Ywhaq, Ahcyl1, Map6, Carm1, Bin1, Dnm1	6.9×10^{-12}
Protein binding	Atp1b1, Srcin1, Bap1, Wbp2, Hsph1, Ash2l, Ank3, Wdr77, Stk39, Agap2, Tubb3, Nsf, Ina, Rbbp5, Arhgef2, Slc25a4, Wnk1, Bsn, Pfkm, Pclo, Mfn2, Mapk1, Ksr2, Bin1, Carm1, Hcfc1, Klc2, Brsk1, Klc1, Ppp1r12a, Camk2b, Ppp3ca, Hspa5, Clasp2, Plp1, Ncdn, Immt, Map1b, Csnk2b, Dpysl2, Atp1a2, Tet3, Ywhah, Map2, Ywhaq, Pspc1, Ahcyl1, Atp5a1, Dnm1	4.8×10^{-11}
Microtubule	Arhgef2, Tubb2a, Map1b, Klc2, Dpysl2, Hsph1, Klc1, Map2, Map6, Clasp2, Bin1, Tubb3, Tubb4a, Dnm1, Tubb4b, Jakmip1	1.4×10^{-9}
Methylation	Caskin1, Srcin1, Ncdn, Slc25a5, Tubb2a, Map1b, Bsn, Hcfc1, Dpysl2, Brsk1, Tpi1, Tet3, Atp2b4, Ash2l, Pspc1, Atp5a1, Hspa5, Hnrnph1, Carm1, Gapdh, Dnm1	2.2×10^{-8}
Cytoskeleton	Arhgef2, Srcin1, Tubb2a, Map1b, Bsn, Klc2, Dpysl2, Brsk1, Mapk1, Klc1, Ank3, Map2, Camk2b, Clasp2, Map6, Gapdh, Dnm1, Tubb3, Tubb4a, Tubb4b, Jakmip1	1.6×10^{-7}
Acetylation	Tubb2a, Hcfc1, Cxxc1, Hsph1, Tpi1, Ppp3cb, Stk39, Hspa5, Ppp3ca, Gapdh, Nsf, Tubb4b, Ina, Arhgef2, Ncdn, Slc25a4, Immt, Slc25a5, Map1b, Pfkp, Csnk2b, Pfkm, Slc25a12, Mapk1, Ywhah, Ywhaq, Pspc1, Atp5c1, Ahcyl1, Atp5a1, Hnrnph1, Bin1	3.3×10^{-6}
Nucleotide-binding	Tubb2a, Pfkp, Wnk1, Atp1a3, Brsk1, Pfkm, Atp1a2, Mfn2, Mapk1, Hsph1, Atp2b4, Ksr2, Rhot1, Stk39, Camk2b, Atp5a1, Hspa5, Agap2, Dnm1, Nsf, Tubb3, Tubb4a, Rtcb, Tubb4b	4.9×10^{-6}
Coiled coil	Srcin1, Hcfc1, Bap1, Klc2, Cxxc1, Hsph1, Atp2b4, Trak2, Klc1, Ank3, Ppp1r12a, Clasp2, Hspa5, Iqsec1, Tubb3, Jakmip1, Tubb4b, Ina, Arhgef2, Immt, Map1b, Wnk1, Bsn, Atp1a2, Pclo, Mfn2, Tet3, Pspc1, Hivep3, Bin1	4.4×10^{-5}
Axon	Mapk1, Srcin1, Klc1, Ank3, Map1b, Atp1a3, Bsn, Hcfc1, Dpysl2, Bin1, Pclo, Tubb3	1.6×10^{-4}
cGMP-PKG signaling pathway	Mapk1, Atp1b1, Atp2b4, Slc25a4, Slc25a5, Atp1a3, Ppp3cb, Ppp1r12a, Ppp3ca, Atp1a2	1.9×10^{-4}
Microtubule binding	Arhgef2, Map2, Map1b, Map6, Clasp2, Dpysl2, Gapdh, Dnm1, Jakmip1	2.6×10^{-3}
Beta tubulin	Tubb2a, Tubb4a, Tubb3, Tubb4b	3.1×10^{-3}
Neuron projection	Atp2b4, Srcin1, Ncdn, Klc1, Ank3, Map2, Bsn, Klc2, Atp1a2, Dpysl2, Pclo	4.7×10^{-3}
Neuronal cell body	Arhgef2, Srcin1, Ncdn, Klc1, Map2, Map1b, Bsn, Hcfc1, Dpysl2, Pclo, Tubb4a, Tubb3	5.9×10^{-3}
Calmodulin-	Camkv, Atp2b4, Map2, Ppp3cb, Camk2b, Ppp3ca, Map6	7.4×10^{-3}

binding		
Set1c/ COMPASS complex	Rbbp5, Ash2l, Hcfc1, Cxxc1	1.0×10^{-2}
Protein kinase binding	Mapk1, Srcin1, Ppp1r12a, Wnk1, Stk39, Camk2b, Brsk1, Dpysl2, Agap2, Dnm1, Nsf	1.8×10^{-2}
Dendrite	Srcin1, Ncdn, Ank3, Map2, Map1b, Bsn, Hcfc1, Camk2b, Dpysl2, Pclo, Tubb3	1.8×10^{-2}
Substantia nigra development	Ina, Plp1, Ywhah, Ywhaq, Hspa5	2.5×10^{-2}
Microtubule-based process	Tubb2a, Map1b, Tubb4a, Tubb3, Tubb4b	3.4×10^{-2}
ATP-binding	Pfkp, Wnk1, Atp1a3, Atp1a2, Pfkm, Brsk1, Mapk1, Hsph1, Atp2b4, Ksr2, Camk2b, Stk39, Atp5a1, Hspa5, Nsf, Rtcb	3.4×10^{-2}
GTPase activity	Mfn2, Tubb2a, Rhot1, Agap2, Tubb4a, Dnm1, Tubb3, Tubb4b	3.7×10^{-2}
Postsynaptic density	Arhgef2, Srcin1, Map2, Map1b, Bsn, Camk2b, Pclo, Nsf	4.0×10^{-2}
Extracellular vesicle	Atp1b1, Tubb2a, Atp1a3, Atp1a2, Tubb4b	4.3×10^{-2}

Table 3.1 Shown here are the top GO categories for the significantly enriched OGT interactome proteins. FDR = false discovery rate. OGT-FH mouse brain MS data from Yao Xiao.

If we lowered the cutoff to an average SILAC ratio of 5.0, the list becomes 188 putative OGT interactors, which are heavily enriched for ribosomal proteins (*FDR* < 2.7×10^{-23}, 29 proteins, 17.4-fold enrichment). This is consistent with previous studies showing that the majority of ribosomal proteins appear to glycosylated in addition to translational regulators such as the elongation factors Tu translation elongation factor, mitochondrial (TUFM) and elongation factor 1-α1 (EEF1A1).[19,20] We also compared the brain OGT interactome to the OGT interactomes found in a tandem affinity purification experiment performed in HEK293T cells and a human protein microarray experiment (Figure 3.12). The top two gene ontology general categories for the overlapping proteins in the two tandem affinity purification experiments were ribosomal and cytoskeletal proteins suggesting that OGT interacts with these highly abundant proteins across cell types and species (Benjamini-corrected *P* < 1.8×10^{-19} and *P* < 1.3×10^{-6} respectively). In addition, we saw 75 of the 188 proteins in our OGT interactome overlapping with known *O*-GlcNAcylated proteins (Figure 3.13). The top annotation cluster 1 for those

overlapping 75 proteins includes dendritic, axonal, and neuronal cell body proteins (FDR $< 2.9 \times 10^{-5}$, 1.2×10^{-3}, 4.6×10^{-3} respectively) (Table 3.2). The next two annotation clusters were enriched for calmodulin binding and cytoskeletal proteins (FDR $< 1.7 \times 10^{-5}$ and 2.3×10^{-4}) (Table 3.2). In conclusion, our preliminary OGT brain interactome results reveal similar proteins as identified in other OGT interactome and *O*-GlcNAcome studies as well as some novel protein interactors. Experiments are ongoing within the lab to verify the interaction of OGT with these putative interactors and test to see if the *O*-GlcNAcome is influenced by the interaction of OGT with these interactors.
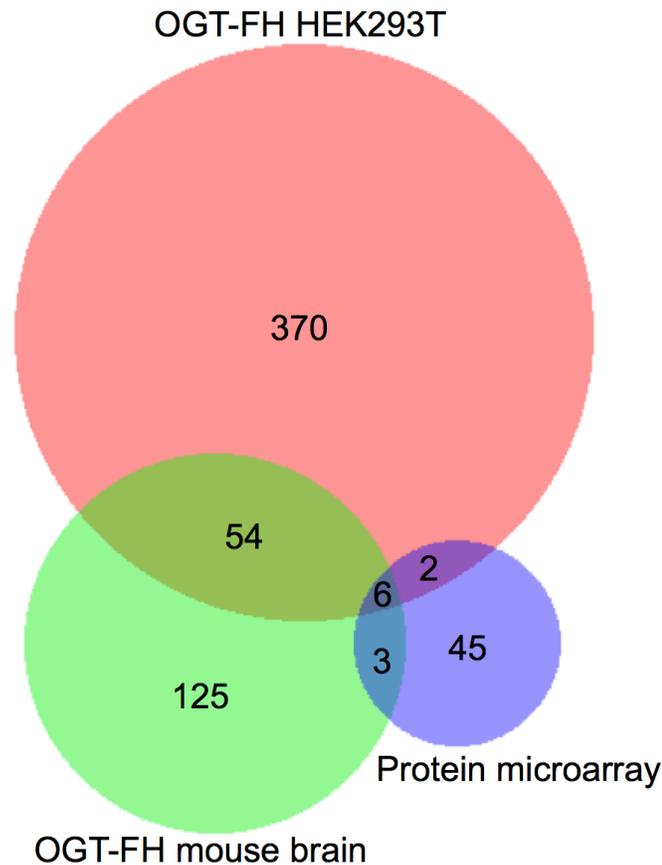
Figure 3.12 Comparison between OGT-FH mouse brain OGT interactome and other OGT interactome experiments. Shown here is a Venn Diagram displaying the number of common proteins in our OGT interactome study and other OGT interactome studies.[21,22] OGT-FH mouse brain MS data from Yao Xiao.
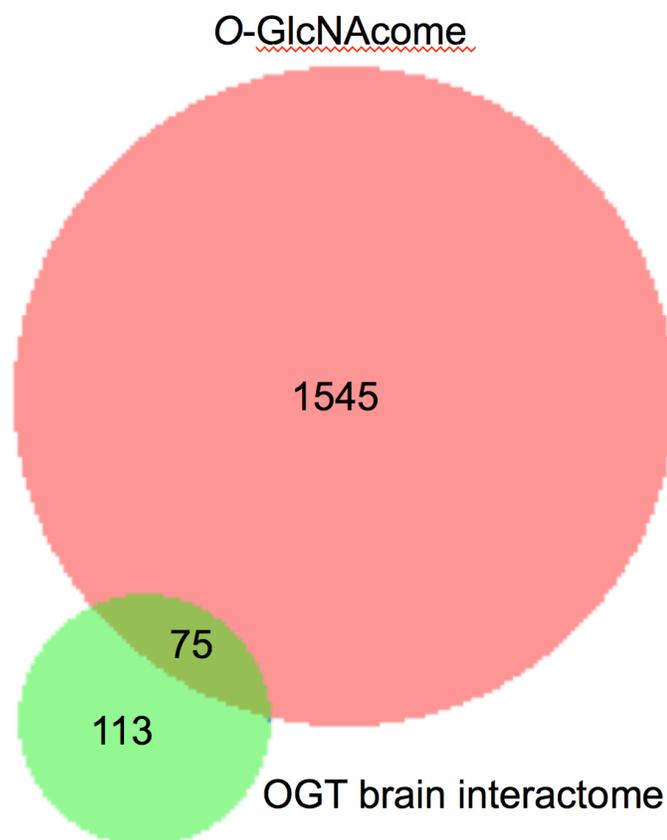
Figure 3.13. Comparison between *O*-GlcNAcome and the OGT brain interactome. Here is a Venn Diagram overview of the *O*-GlcNAcome and the OGT brain interactome. The *O*-GlcNAcome was obtained from several different species, cell types, and methods as previously described.[12,14-16,23,24] OGT-FH mouse brain MS data from Yao Xiao.

Table 3.2 Top 5 GO clusters for OGT brain interactome and *O*-GlcNAcome common proteins.

GO Term	#	FDR	Proteins	FE
Dendrite	14	2.89×10^{-5}	Myo5a, Ncdn, Map1b, Bsn, Hcfc1, Rps6, Pclo, Rps3, Gnb1, Ank3, Map2, Pabpc1, Camk2a, Tubb3	7.7
Axon	11	1.20×10^{-3}	Myo5a, Actb, Ank3, Map1b, Bsn, Hcfc1, Pclo, Camk2a, Ywhae, Tubb3, Myh10	8.0
Neuronal cell body	12	4.61×10^{-3}	Myo5a, Ncdn, Map2, Map1b, Bsn, Hcfc1, Ogt, Pclo, Camk2a, Tubb3, Mbp, Myh10	6.1
Calmodulin-binding	9	1.67×10^{-5}	Myo5a, Camkv, Map2, Sptbn1, Ppp3ca, Map6, Camk2a, Myh10, Sptan1	20
Cytoskeleton	17	2.27×10^{-4}	Myo5a, Ncdn, Map1b, Bsn, Hcfc1, Rps6, Pclo, Rps3, Gnb1, Ank3, Map2, Pabpc1, Camk2a, Tubb3	4.9
Microtubule	8	2.46×10^{-2}	Myo5a, Ncdn, Map1b, Bsn, Hcfc1, Rps6, Pclo, Rps3, Gnb1, Ank3, Map2, Pabpc1, Camk2a, Tubb5	9.4
Nucleotide-binding	23	1.51×10^{-5}	Myo5a, Ncdn, Map1b, Bsn, Hcfc1, Rps6, Pclo, Rps3, Gnb1, Ank3, Map2, Pabpc1, Camk2a, Tubb3	4.1
Postsynaptic density	11	2.06×10^{-5}	Actb, Epb41l3, Bcas1, Map2, Map1b, Bsn, Sptbn1, Pclo, Camk2a, Nsf, Dclk1	12

Table 3.2 shows the top gene ontology (GO) clusters for the overlapping proteins between the OGT brain interactome and the *O*-GlcNAcome. The clusters are indicated by color: red (cluster 1: enrichment 6.4), orange (cluster 2: enrichment 6.1), green (cluster 3: enrichment 5.0), blue (cluster 4: enrichment 4.4), purple (cluster 5: enrichment 3.3). Categories with FDR > 0.05 or that were redundant were removed from

the table. FE = Fold Enrichment; FDR = False Discovery Rate. OGT-FH mouse brain MS data from Yao Xiao.

3.4 Development of chemoenzymatic tools for the *O*-GlcNAcome

3.4.1 Overview of chemoenzymatic approach

Robust enrichment of O-GlcNAcylated proteins can be accomplished using a two-step chemoenzymatic approach.[25,26] First, the O-GlcNAc moiety is tagged with a non-natural azide group by treatment of cell lysates with UDP-GalNAz **1** and a mutant glycosyltransferase (Y289L GalT) that specifically recognizes terminal GlcNAc moieties.[27] Next, a biotin group is attached via copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC), which allows for affinity purification. Although a limited set of alkyne-biotin linkers are commercially available, many existing linkers are not ideal for mapping O-GlcNAc modification sites.[28] In particular, harsh conditions are usually required to disrupt the femtomolar biotin-streptavidin interaction, which may hydrolyze the labile O-GlcNAc moiety.[29] Additionally, many linkers contain a large spacer between the biotin group and the alkyne functionality, which appends a relatively large mass to the glycopeptide and can complicate its identification.[29] Therefore, a facile method to release the labeled peptides and proteins with minimal added mass would greatly facilitate downstream analysis.

Several cleavable linkers have been previously developed for enrichment of O-GlcNAcylated proteins.[30-33] However, each suffers from significant drawbacks for site identification. For example, a photocleavable linker was employed in conjunction with UDP-GalNAz **1** and Y289L GalT to sequence modified peptides from mouse brain lysate.[30,31] Importantly, the moiety retained after cleavage provided a positively-charged amine group, which increased the overall peptide charge and facilitated ionization by

electron-transfer dissociation (ETD), the most successful MS/MS method for O-GlcNAc peptide sequencing.[34,35] Unfortunately, cleavage of the linker was found to be incomplete.[30] In a recent report, a dibromine-containing, acid-cleavable linker was employed to identify various glycan modifications including O-GlcNAc.[32] However, cleavage of the linker revealed only a neutral hydroxyl group, and the halogenated glycopeptides demonstrated poor fragmentation efficiency using ETD. Therefore, we aimed to develop a tag that would be both quantitatively released and incorporate a positive charge upon cleavage to facilitate MS detection by ETD.

3.4.2 Validation of Dde cleavable linker

We aimed to develop a tag that would be both quantitatively released and incorporate a positive charge upon cleavage to facilitate MS detection by ETD. To achieve these dual goals, we chose to examine the 1-(4,4-dimethyl-2,6-dioxocyclohex-1-ylidene)ethyl (Dde) functional group (Figure 3.14A). The Dde moiety has been used extensively as a protecting group for lysine in peptide synthesis, demonstrating its compatibility with biomolecules.[36] The group is stable to both acid and base and can be quantitatively removed by hydrazine.[37] However, it was reported that the Dde group is incompatible with sodium dodecyl sulfate (SDS) and amine-containing buffers, common additives to protein labeling protocols.[38]

We first investigated the labeling of a model O-GlcNAcylated peptide with commercially available alkyne-Dde-biotin **2** followed by cleavage of the linker using liquid chromatography (LC)-MS (Figure 3.14B and Figure 3.15). Commercially available peptide TAPT(gS)TIAPG (Figure 3.15A), where gS is the O-GlcNAcylated residue, was incubated with 100 ng/µL Y289L GalT and 1 mM UDP-GalNAz of **1** in 10 mM HEPES

pH 7.9, 5.5 mM MnCl2 overnight at 4 °C. LC-MS analysis revealed quantitative conversion to the desired GalNAz-labeled product (Figure 3.15B). Next, the azide-containing peptide was reacted with 100 μM of **2** in 10 mM sodium phosphate pH 7.6 containing 2 mM sodium ascorbate (NaAsc), 100 μM THPTA, and 1 mM CuSO4. After 1 h, stoichiometric biotinylation of the peptide was observed (Figure 3.15C). Treatment with 2% aqueous hydrazine for 1 h at RT resulted in quantitative cleavage of the linker to afford a minimal, positively-charged aminomethyltriazolyl group (Figures 3.14B and 3.15D). To test whether the linker would be stable under stringent wash conditions, we incubated the labeled peptide with 1% RapiGest, a MS-compatible analogue of SDS, or 6 M urea for 1 h at RT (Figure 3.16). In both cases, the linker remained intact, highlighting the compatibility of the linker with rigorous washing steps.
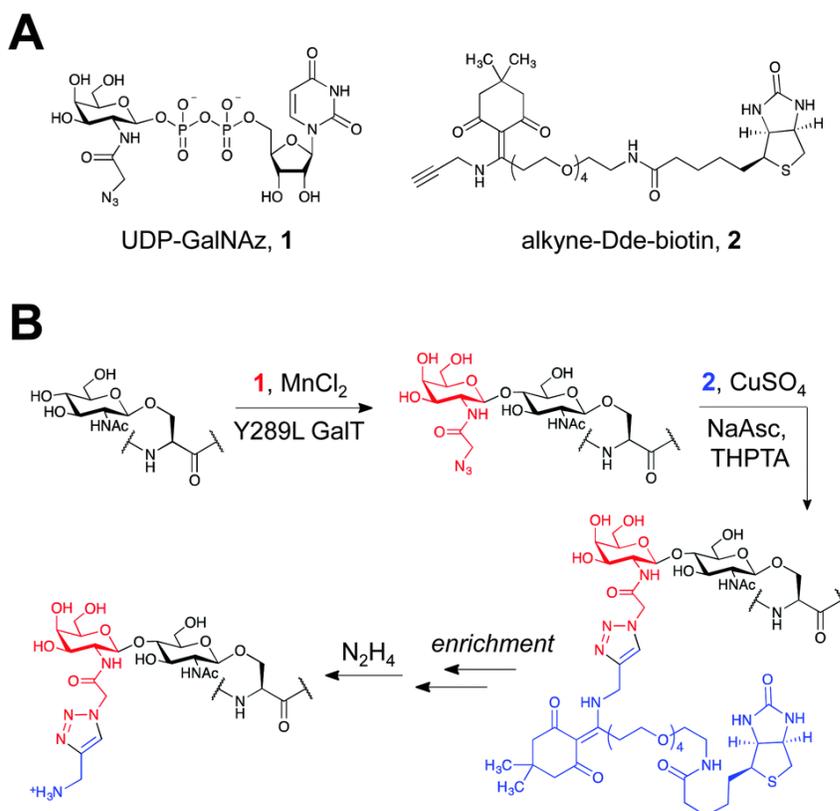
Figure 3.14 Overview of chemicals and workflow of chemoenzymatic linker labeling. (A) Chemicals used in the labeling protocol. (B) Schematic of *O*-GlcNAc protein enrichment and elution using two-step chemoenzymatic/CuAAC labeling protocol.
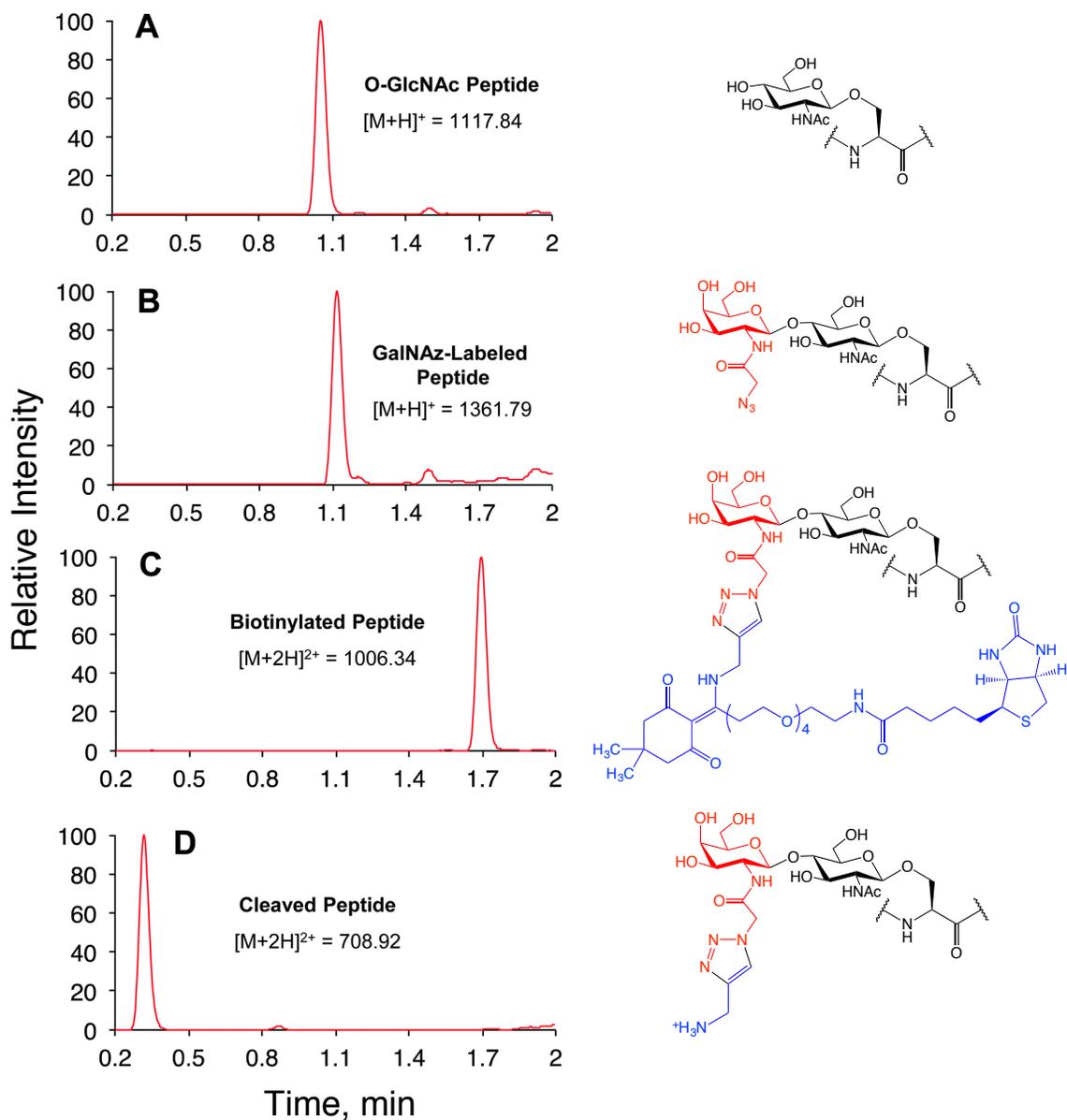
Figure 3.15 Labeling and cleavage reactions proceed quantitatively. Reverse phase LC-MS analysis of O-GlcNAc peptide labeling reactions at (A) time 0, (B) 16 h after addition of 1 and Y289L GalT, (C) 1 h after CuAAC with 2, and (D) 1 h after cleavage with 2% aqueous hydrazine. See methods for experimental details. (A) and (B) show base peak chromatograms. (C) and (D) show extracted ion chromatograms of the starting material and product within ±1 m/z of calculated values.
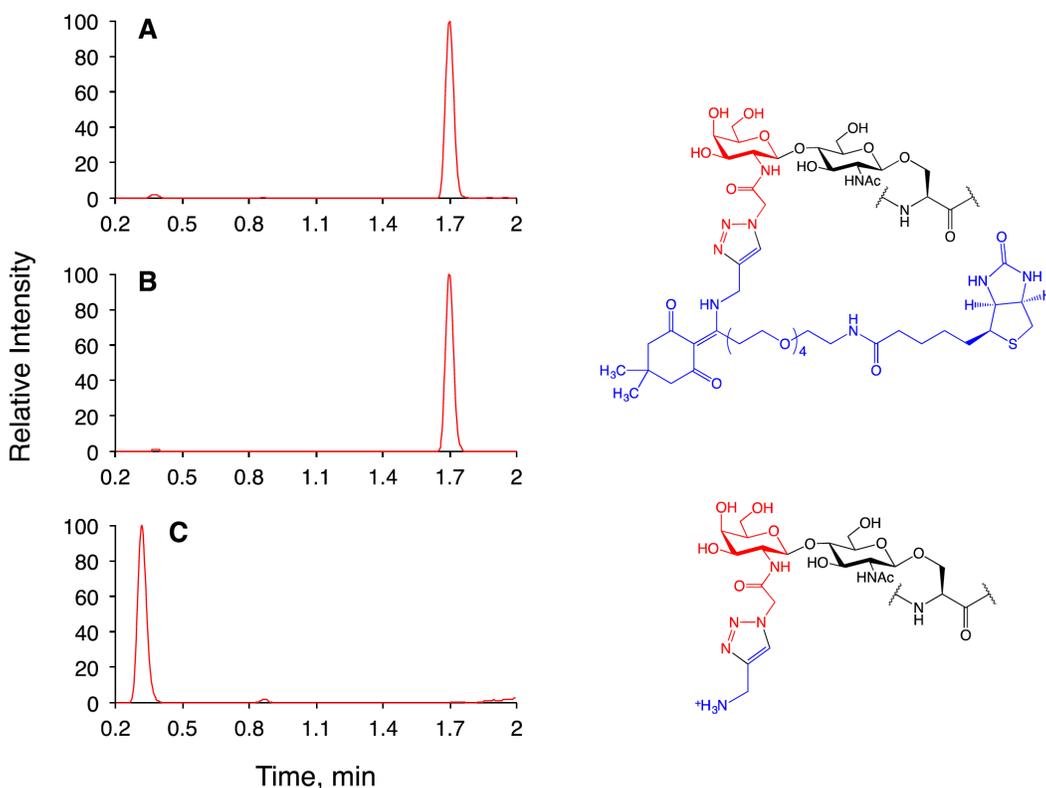
Figure 3.16 Labeled peptide is stable to wash conditions. Reverse-phase LC-MS analysis of alkyne-Dde-biotin-labeled peptide after 1 h incubation with (A) 6 M urea, (B) 1% RapiGest, or (C) 2% hydrazine. All graphs show extracted ion chromatograms of the starting material and possible product within ± 1 *m/z*.

3.4.3 Comparison with a photocleavable linker

We then tested the performance of our linker in comparison to the previously described photocleavable linker (alkyne-PC-biotin).[30] Briefly, 500 µg HEK-293T cell lysate was subjected to chemoenzymatic labeling with **1** using Y289L GalT as described above. The azide-labeled protein was split into two fractions of 200 µg and reacted with **2** or alkyne-PC-biotin by CuAAC. A sample of each biotinylated mixture was reserved, and the remaining material was subjected to cleavage by 2% hydrazine monohydrate or UV irradiation by 365 nm. Samples were resolved by SDS-PAGE and probed for biotin using fluorescently tagged streptavidin (Figure 3.17). Notably, lysate labeled by **2** showed a stronger biotin signal than the photocleavable linker, suggesting higher labeling efficiency. Furthermore, although both linkers cleaved well, slightly higher residual

signal was observed for the photocleavable linker compared to **2**, implying that the Dde moiety was also released more efficiently than the photocleavable group.
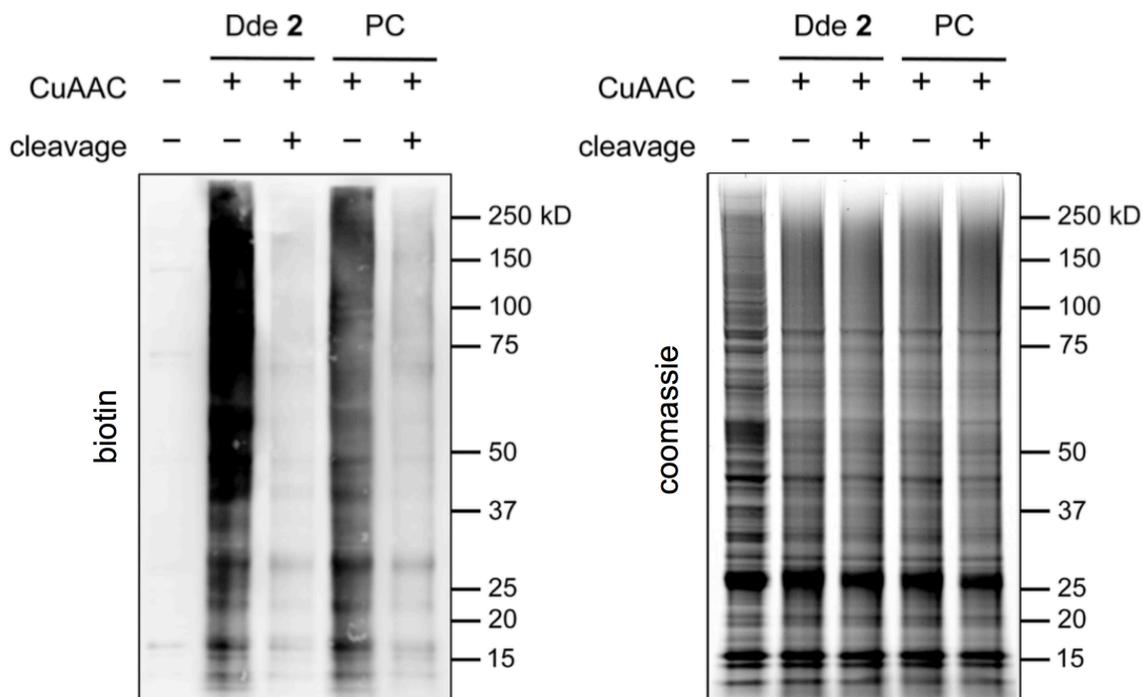
Figure 3.17 Dde linker 2 outperforms photocleavable linker. Protein (20 μg) labeled with alkyne-Dde-linker **2** shows both higher biotin signal after labeling (lane 2 vs. lane 4) and lower residual signal after cleavage (lane 3 vs. lane 5) compared to alkyne-PC-linker (PC). Coomassie staining on the right confirms equal protein loading in each lane.

3.4.4 Validation using known *O*-GlcNAcylated proteins- α-crystallin and *O*-GlcNAc transferase

We next evaluated the potential of the approach to pull down known *O*-GlcNAcylated proteins and identify sites of modification. The well-characterized *O*-GlcNAcylated protein α-crystallin was selected to assess the sensitivity of the method because it has a relatively low glycosylation stoichiometry (< 10%).[39] Short-form OGT (sOGT) has multiple sites of *O*-GlcNAcylation and was thus used to determine whether comprehensive site mapping could be achieved.[35,40,41] To test the robustness of our method in a complex mixture, each protein was added to 200 μg of adult mouse cortical lysate and subjected to chemoenzymatic labeling and CuAAC using **2**. The labeled

proteins were applied to high-capacity Neutravidin resin and washed five times each with 0.5 mL of 1% SDS, 6 M urea, and phosphate buffered saline (PBS). The resin was then rotated end-over-end for 1 h with 2% aqueous hydrazine to cleave the O-GlcNAcylated proteins from the resin. Eluted samples were acetone precipitated, re-dissolved in denaturing buffer and subjected to reduction, alkylation, and proteolytic digestion. Digested peptides were separated by nanoLC-MS and analyzed on an LTQ-Velos by a combination of collision-induced dissociation (CID) and ETD-MS.

A large number of O-GlcNAcylation sites were identified on α-crystallin and sOGT (Table 3.3). First, the known O-GlcNAc site on α-crystallin A (Ser-162) was readily recognized despite the low abundance of the O-GlcNAc modification at this site.[42] Importantly, we observed both known and novel sites on sOGT.[35,43] For example, we identified the previously reported Thr-662 site, which is found in the catalytic domain of sOGT.[16] The new linker design also as a neighboring modification site at Ser-664, which to our knowledge has not been reported.[31] Furthermore, we mapped a novel site of glycosylation at Ser-580, which is found in the catalytic domain of sOGT. The new linker design also revealed a number of new O-GlcNAcylation sites within the N-terminal tetratricopeptide repeat-containing (TPR) domain (Ser-10, Thr-12, Ser-20, and Thr-23) of sOGT, and we observed a doubly modified peptide at both Thr-12 and Ser-20. As the TPR domains of OGT are thought to mediate protein-protein interactions, such modifications could play an integral role in OGT regulation and may provide a mechanism to selectively modulate its activity toward specific substrates.[5,44,45]

Table 3.3 O-GlcNAc sites identified on alpha-crystallin and OGT

Protein	Peptide Sequence	Site(s)	Mascot ion score	Mascot delta ion score	Method
α-crystallin A	AIPVSREEKPSSAPSS	Ser-162	24.9	23.5	ETD

sOGT	ISPTFADAYSNMGNTLK	Ser-10*/ Thr-12*	46.5	-	ETD
sOGT	ISPTFADAYSNMGNTLK	Ser-20*	21.6	13.6	ETD
sOGT	ISPTFADAYSNMGNTLK	Ser-10*/ Thr-12*/ Ser-20*	38.4	-	ETD
sOGT	EMQDVQGALQCYTR	Thr-38	41.8	35.0	ETD
sOGT	AIQINPAFADAHSNLASIHKDS GNIPEAIASYR	Ser-52*	53.5	41.0	CID
sOGT	AIQINPAFADAHSNLASIHK	Ser-56*	56.8	15.7	CID
sOGT	LYLQMWEHYAAGNKPDHMIK PVEVTESA	Thr-662	33.1	8.0	CID

Table 3.3 *O*-GlcNAc sites identified following labeling with **2**, Neutravidin affinity purification, and hydrazine mediated elution. Sites within the peptide sequence are denoted in red. Novel site identifications are marked by an asterisk.

3.5 Discussion

We describe an improved method to facilitate the comprehensive mapping of *O*-GlcNAc modification sites. Chemoenzymatic attachment of an azide-containing monosaccharide onto *O*-GlcNAc sugars provides a stoichiometric, bioorthogonal handle, which is further functionalized with an alkyne-Dde-biotin linker to isolate and enrich *O*-GlcNAcylated proteins. The cleavable Dde linker provides numerous benefits over other reported structures. First, the linker is commercially available and inexpensive. Second, it is stable to rigorous, denaturing wash conditions and can be quantitatively cleaved under mild chemical conditions. Finally, the cleaved moiety that remains on the modified peptide minimally changes the peptide mass and generates an additional positive charge, which facilitates peptide sequencing by ETD. Together, in combination with the commercially available chemoenzymatic labeling kit, the method provides an accessible and practical system for the broader community.[26] Using this approach, we identified both established and new modification sites on α-crystallin and sOGT, including potentially novel sites for OGT regulation. Our results showcase the method's capacity to comprehensively profile protein *O*-GlcNAcylation. The stoichiometric nature of the chemoenzymatic labeling, CuAAC reaction, and elution steps

provides an ideal platform for future quantitative MS analyzes to profile global *O*-GlcNAcylation and will enable the discovery of novel functional roles for *O*-GlcNAcylation in diverse biological contexts. Our lab is currently developing next generation linkers that have improved stability and differential labelling capabilities in order to obviate the dimethyl-labelling step. These next-generation linkers are currently being synthesized in the lab and will be used to identify novel *O*-GlcNAc sites as well as perform quantitative site-specific *O*-GlcNAcylation changes that accompany neuronal activation.

Furthermore, we describe the generation of a novel OGT-FLAG-HA mouse that facilitates tandem affinity purification of OGT in order to identify the OGT interactome. Importantly, the integration of the minimal tag into the mouse allows for pull down of endogenous OGT while avoiding the false positives and other potential pitfalls that could befall overexpression-based tandem affinity purification. Using our novel OGT-FH mouse, we were able to obtain the brain-specific OGT interactome, which is consistent with previous OGT interactome studies. In addition, many of the proteins within the brain-specific OGT interactome are also known to be OGT substrates. The next steps are to validate the novel OGT interactors and then to test the OGT interactor/substrate hypothesis using interactor knockdown or ideally OGT-interactor specific interaction disruption followed by our newly optimized *O*-GlcNAcome identification methods. In this way, we can directly test the OGT interactor/substrate hypothesis and identify the key "hub" interactors that mediate substrate targeting in the brain. In summary, we have developed robust OGT interactome and *O*-GlcNAcome tools that facilitate OGT

interactor and substrate identification toward the goal of understanding the key role of *O*-GlcNAc in neuronal function.

3.6 Methods

3.6.1 Reagents and materials for OGT interactome

Fetal bovine serum (FBS); ViraPower™ Lentiviral Packaging Mix, Lipofectamine™ 3000; Na pyruvate (100X, Gibco); Penicillin/Streptavidin (P/S, 100X, Gibco); Nonessential amino acids (100X, Gibco); HEPES (100X, Gibco); DMEM (Dulbecco's Modified Eagle Media) High Glucose and GlutaMAX (Gibco); FastDigest MluI; NuPAGE™ Novex™ 4-12% Bis-Tris protein gels (1.0 mm, 10-well); Triton®-X-100, Electrophoresis Grade; ElectroMax™ DH5α-E™ Competent Cells; Pierce™ BCA Protein Assay Kit; Bovine Serum Albumin, Fraction V; MEGAshortscript™ T7 Transcription Kit; MEGAclear™ Transcription Clean-up Kit; UltraPure™ DNase/RNase-Free Distilled Water; FastDigest SalI; FastDigest MluI; Pierce™ Trypsin Protease, MS Grade; DTSSP (3,3'-dithiobis(sulfosuccinimidyl propionate)); and BCA assay reagents were obtained from ThermoFisher Scientific (Waltham, MA). We purchased BD Bacto™ Yeast Extract, Tryptone, and Agar from BD Biosciences (San Jose, CA). Amicon Ultra-15 Centrifugal Filter Units with Ultracel-10 membrane (10 kDa); Amicon Ultra-15 Centrifugal Filter Units with Ultracel-30 membrane (30 kDa); phosphate buffered saline powder, pH 7.4; DL-Dithiothreitol (DTT); 3x FLAG peptide; ANTI-FLAG M2® affinity gel; α-Tubulin mouse mAb; Anti-*O*-GlcNAc transferase (DM-17) rabbit antibody; α-Tubulin mouse mAb (T9026); iodoacetamide, BioUltra; Ammonia Solution, 25%; Monoclonal ANTI-FLAG® M2 mouse antibody; Triethylammonium Bicarbonate (TEAB) Buffer, Volatile Buffer for HPLC; Sodium

cyanoborohydride (NaBH$_3$CN) were purchased from MilliporeSigma (Burlington, MA).

HA-Tag (C29F4) Rabbit mAb was obtained from Cell Signaling (Danvers, MA). Q5®

Site-Directed Mutagenesis Kit, Q5® Hot Start High-Fidelity 2X Master Mix, BamHI-

HF®, Quick Ligation® Kit were obtained from New England BioLabs, Inc. (Ipswich,

MA). GenePulser/MicroPulser Cuvettes (0.1 cm gap sterile electroporation cuvettes) and

Silver Stain Plus were purchased from Bio-Rad (Hercules, CA). The pX330-U6-

Chimeric_BB-CBh-hSpCas9 construct was obtained from Addgene (Cambridge, MA).

DNeasy Blood & Tissue Kits were purchased from Qiagen (Hilden, Germany). The

CMV6-Entry construct was obtained from Origene (Rockville, MD). Polyethylenimine,

linear transfection grade (PEI, MW 25,000) was purchased from Polysciences, Inc.

(Warrington, PA). Casein kinase II (CKII) peptide was synthesized and purchased from

GenScript (Piscataway, NJ). cOmplete^{TM} protease inhibitor cocktail without EDTA (PIC-

EDTA) was purchased from Roche Diagnostics Corp. (Indianapolis, IN).

3.6.2 Tandem affinity purification protocol for OGT pull down and lentiviral production

The conditions for tandem affinity purification were optimized using HEK293T

cells transfected with lentivirus expressing rat OGT-FLAG-HA and FLAG-HA-OGT. In

order to generate lentivirus, 10 cm plates were coated with 0.1% wt/vol. gelatin A for 2

hours followed by rinsing twice with sterile water and air drying. Then, HEK293FT cells

were plated in COMPLETE media (10% FBS, Na pyruvate, P/S, nonessential amino

acids, and HEPES in DMEM High Glucose). At 60% confluence, the media was

removed, and then the cells were plated with serum-free DMEM High Glucose with

GlutaMAX transfected with ViraPower™ Lentiviral Packaging Mix (pLP1, pLP2, and

146

pLP/VSVG), the Tet-OGT-FH or Tet-FH-OGT and Lipofectamine™ 3000 according to

the manufacturer's instructions. After 6 hours, the serum-free media was replaced with

COMPLETE media and then the media was collected and then replaced every 12 hours

thereafter for 72 hours total. The collected media was then filtered using a 0.45 μm sterile

filter or spun 2000 x *g* for 5 min in order to remove or pellet cellular debris. Then, the

media was concentrated using 50-ml pre-washed Amicon-15 30 kDa centrifugal

concentrators. HEK293T cells were transduced with the lentivirus in order to create

HEK293T cells that would express OGT-FH with the addition of doxycycline (Tet-OGT-

FH HEK293T cells).

Tet-OGT-FH HEK293T cells cultured with or without 1 μg/ml doxycycline or

OGT-FH mouse brains or WT mouse brains were then lysed in 1% Triton X-100 in TBS

with PIC-EDTA followed by enrichment using ANTI-FLAG M2® M2 affinity gel

agarose beads with end-over-end rotation overnight at 4°C. After this, the FLAG beads

were washed three times with TBS (Tris Buffered Saline with Triton X-100). Following

washing, the OGT-FH was eluted using 150 ng/μL 3xFLAG peptide in TBS (19 mM

Tris, pH 7.4, 137 mM NaCl, 2.7 mM KCl) with 1% Triton X-100. Then, the eluted

proteins were added to HA beads and enriched with end-over-end rotation overnight at

4°C. The beads were washed with TBS three times followed by elution with 3M NaSCN.

Proteins in the final tandem affinity purification elution were separated on NuPAGE™

Novex™ 4-12% Bis-Tris protein gels in MOPS buffer (50 mM MOPS, 50 mM Tris Base,

0.1% SDS, 1 mM EDTA, pH 7.7). Following protein gel electrophoresis, the proteins

were either then prepared for mass spectrometry (see below), silver stain, or transferred

onto an Immobilon-FL PVDF membrane in transfer buffer (25 mM Tris, 190 mM

glycine, 20% methanol) for western blot. For silver stain, the gel was stained using the Silver Stain Plus kit according to the manufacturer's instructions. For western blot, the membrane was blocked for 1 h at RT with 5% bovine serum albumin (BSA) in 1x TBST (TBS with 0.1% Tween 20). The blot was then incubated for 1 h at RT or overnight at 4°C with the indicated primary antibodies (1:1,000 dilution): HA-Tag (C29F4) Rabbit mAb, ANTI-FLAG M2® mouse mAb, α-Tubulin mouse mAb, and OGT rabbit Ab. After primary incubation, the membrane was rinsed three times with 1x TBST and then incubated with the appropriate Alexa Fluor® 680 or 800 conjugated secondary antibody (1:5,000) in 5% BSA in TBST for 1 h at RT, washed three times with TBST for 5 min, and imaged with an Odyssey scanner.

For crosslinking, the cells were placed on ice, washed once with PBS/Ca^{2+}/Mg^{2+} buffer (0.1 mM $CaCl_2$, 1 mM $MgCl_2$ in PBS), then incubated for 2 hours in either 1 mM DTSSP (crosslinking solution) in PBS/Ca^{2+}/Mg^{2+} or PBS/Ca^{2+}/Mg^{2+} (non-crosslinking solution), and quenched in 20 mM Tris, pH 7.4 in PBS/Ca^{2+}/Mg^{2+} buffer for 15 min. Following crosslinking, the cells were lysed in 0.5% Triton X-100 in Buffer A (10 mM HEPES, 150 mM NaCl, 1 mM EGTA, 0.1 mM $MgCl_2$, pH 7.4) with PIC-EDTA for 30 min with rocking. Following lysis, the cells were scraped from the plate and centrifuged at 15,000 x g for 15 min. All crosslinking and lysis conditions were performed on ice or at 4°C with ice-cold buffers. Then, the supernatant was analyzed for protein concentration using the BCA assay according to the manufacturer's instructions. Afterwards, the tandem affinity purification proceeded with 10 mg of each sample in parallel as described above.

3.6.3 Mass spectrometry for OGT interactome

Following tandem affinity purification, the final HA pull down eluent was added to two prewetted 10K MWCO Amicon filter units and washed 3 times in TEAB buffer (100 mM triethylammonium bicarbonate). Then the concentrate was spun down to 50 μL of liquid and then 10 mM DTT in 100 mM TEAB was incubated with gentle rotation for 60 min at 37°C. After reduction, the solution was washed twice with 100 mM TEAB, concentrated to 50 μL, and then alkylated with 50 mM 2-iodoacetamide (IAA) in 100 mM TEAB in the dark with rotation for 40 min at RT. The solution was washed twice with 100 mM TEAB, concentrated to 20 μL, and then digested overnight at 37°C with trypsin in 1 mM $CaCl_2$, 100 mM TEAB with a final protease to protein ratio of 1:20 (wt/wt). The peptides were washed twice with 100 mM TEAB and then combined for a final volume of 100 μL. Then, 4 μL of either 4% (vol/vol) CH_2O or CD_2O (for light or intermediate labelling respectively) was added to the samples, mixed, and then spun down as previously described.[46] Afterwards, 4 μL of 0.6M $NaBH_3CN$ was added to the samples and then rotated for 1 hour at RT. The solution was quenched by adding 16 μL of 1% (vol/vol) ammonia solution and then acidified with 8 μL formic acid. The sample was desalted using HPLC and then resuspended in 10 μL of 0.1% formic acid. At this point, the light- and intermediate-labelled samples were mixed and analyzed by LC-MS/MS at the Caltech Proteomics Exploration Laboratory. About 250 ng of digested peptides were analyzed with an EASY-nLC 1000 and Thermo Orbitrap Fusion™ Tribrid™ Mass Spectrometer (ThermoFisher Scientific) as previously described.[47]

3.6.4 Activity assay to check activity of OGT tags

N-terminal tagged (FLAG-HA-OGT) and C-terminal tagged (OGT-FLAG-HA) were cloned into the CMV6-Entry construct using Q5® Site-Directed Mutagenesis Kit

according to the manufacturer's instructions (New England BioLabs). HEK293T cells were cultured on a 10 cm plate in DMEM, high glucose, GlutaMAX, with 10% FBS and were transfected with these two constructs with a 3:1 PEI:DNA ratio (wt:wt). Twenty-four hours following transfection, the media was removed from cells prior to lysis and scraping in ice-cold lysis buffer (50 mM Tris/HCl pH 7.4, 250 mM mannitol, 50 mM NaF, 1 mM EDTA, 1 mM EGTA, 1 mM DTT, 1 mM Na_2VO_4, 1 % Triton X-100, and 1X protease inhibitors). The cell lysate was dounced 10 times with a 2 ml douncer on ice followed by centrifugation at 20,000 x g for 10 min at 4°C. Next, 500 μL of soluble lysate was added to 40 μL of ANTI-FLAG M2® M2 affinity gel agarose beads, washed 1X with 0.2% triton in TBS, washed 3X in TBS, and then resuspended in 80 μL of 1X OGT buffer (250 mM Tris, 125 mM $MgCl_2$, 0.6 mg/ml BSA, 10 mM DTT) and then divided evenly into four wells. Then, we added a mixture of 1 mM UDP-GlcNAc with or without 100 μM CKII peptide (KKKYPGGSTPVSSANMM) for 1 hour at RT with gentle mixing. After 1 hour, we added the UDP-Glo^{TM} Solution in order to convert all free UDP to ATP followed by a luciferase/luciferin reaction for visualization. All conditions were performed in duplicate or triplicate per plate and the C-terminal and N-terminal tagged OGT luminescence with peptide was normalized to the luminescence lacking peptide. Following completion of the UDP-Glo^{TM} Glycosyltransferase assay, the beads were collected from the wells and then boiled in 1X loading buffer. The boiled elution was then resolved using protein gel electrophoresis and then western blotting as described above. The membrane was blotted using anti-HA and anti-FLAG in order to quantify the levels of OGT that were used in the assay. After quantification, the normalized

luminescence levels were normalized to the average blot intensities of HA and FLAG

tagged OGT as determined by Western blotting.

3.6.5 Design and screen of CRISPR/Cas9 sgRNA

We used the Dr. Feng Zhang's lab's sgRNA design website to generate potential

guide RNAs (http://crispr.mit.edu/guides). The top five sgRNA sequences (with scores)

were (1) 5'-CCTGAATAAAGACTGCGCAC-3' (87%); (2) 5'-

GCTGACTCGGTGACTTCAAC-3' (84%); (3) 5'-CCTGTGCGCAGTCTTTATTC-3'

(77%); (4) 5'-CTTCAACAGGCTTAATCAT-3' (75%); and (5) 5'-

ACAGGCTTAATCATGTGGTC-3' (74%). These sgRNA candidates were cloned into

the pX330 plasmid (U6-sgRNA-Chimeric_BB-CBh-hSpCas9), which was a gift from

Feng Zhang (Addgene plasmid #42230).

We cloned the genomic region neighboring the C-terminal end of mouse OGT

into the pCAG-EGxxP plasmid in order to generate the pCAG-EG_ogt_FP plasmid

(Figure 3.18). Briefly, we extracted gDNA from mouse tails using a DNeasy kit and

amplified the C-terminal region of OGT with Q5® Hot Start High-Fidelity DNA

polymerase using the following primers: OGT_F: 5'-

CAACCACTGAGGATCCTGCCCACAGACAAGTTTGA-3' and OGT_R: 5'-

TATCGAATTCGTCGACTCAATCCACGGCATCACAA-3' [98°C for 30 sec; [98°C

for 10 sec; 62°C for 20 sec; 72°C for 30 sec (34 cycles)], and then 72°C for 2 min] to

produce the following amplicon: chrX:101681950+101682750

(https://tinyurl.com/yajf2aqt). Then, we double digested the amplicon and the pCAG-

EGxxP using SalI and BamHI restriction enzymes for 1 hour at 37°C followed by PCR

clean-up. The Quick Ligation® Kit was used to ligate the cleaned up amplicon and cut

pCAG-EGxxP plasmid as per the manufacturer's instructions. The newly ligated plasmid was transformed into DH5α electrocompetent cells using the Bio-Rad MicroPulser™ Electroporation Apparatus at the 18 kV/cm (Ec1) setting using 0.1 cm gap electroporation cuvettes. The transformed E. coli was grown overnight at 37°C in 1 mg/mL ampicillin in sterile lysogeny broth (LB) (12 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) with shaking and then streaked on 100 µL/mL ampicillin sterile agar (10 g/L agar in LB) plates.   The final plasmid, pCAG-EG_ogt_FP is shown in Figure 3.18B. Using HEK293T cells grown in DMEM, high glucose, GlutaMAX, with 10% FBS, we screened the top 5 guide RNA candidates for cleavage and homologous repair efficiency by co-transfecting the pCAG-EG_ogt_FP construct the pX330-sgRNA with Lipofectamine™ 3000 according to the manufacturer's instructions. Fluorescent images were taken using the Zeiss 700 LSM confocal microscope.
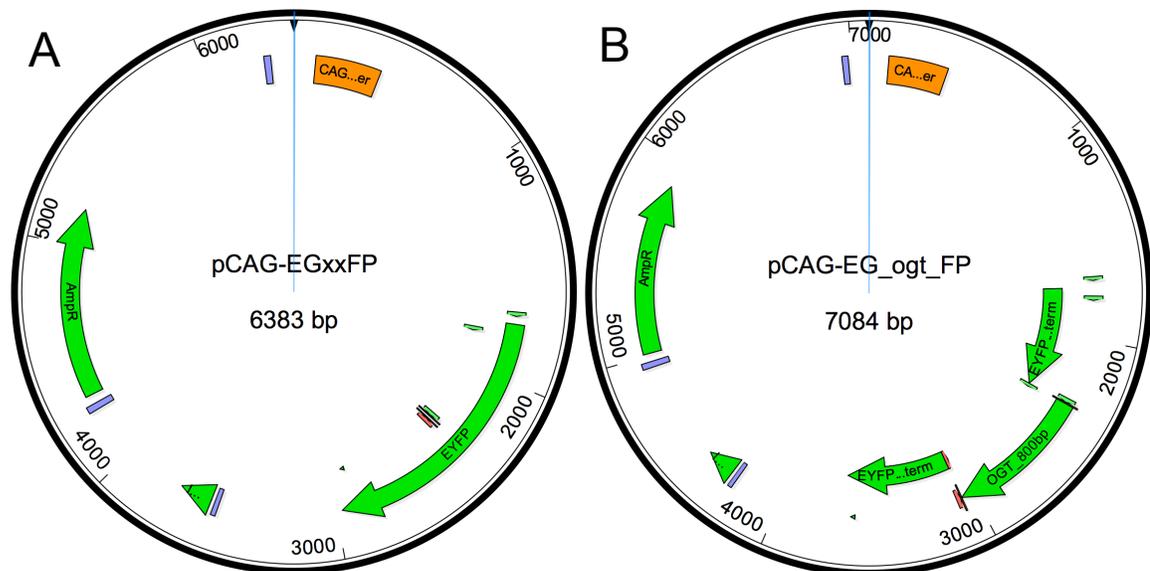
Figure 3.18 Plasmid map for pCAG-EGxxFP for sgRNA screening. (A) Shown to the left is the original pCAG-EGxxFP plasmid. (B) Shown to the right is the pCAG-EG_ogt_FP plasmid with the C-terminus of OGT inserted into the plasmid.

3.6.6 Generation of OGT-FLAG-HA mice using CRISPR/Cas9 and genotyping

We synthesized the sgRNA by performing PCR with Q5® Hot Start High-

Fidelity DNA polymerase on the pX330-sgRNA1 construct (10 ng) using the following

primers T7-sgRNA_F: 5'-ttaatacgactcactataggCCTGAATAAAGACTGCGCAC-3'; and

T7-sgRNA-R: 5'-AAAAGCACCGACTCGGTGCC-3' [98°C for 30 sec; [98°C for 10

sec; 60°C for 20 sec; 72°C for 30 sec (34 cycles)], and then 72°C for 2 min]. We then gel

purified the T7-sgRNA PCR product using the QIAQuick PCR purification kit and *in

vitro* transcribed the T7-sgRNA PCR product with the MEGAshortscript[TM] T7

Transcription Kit followed by MEGAclear[TM] Transcription Clean-up Kit according to the

manufacturer's instructions and Yang *et al.* protocol.[9] Some of the sgRNA was diluted to

500 ng/µL in RNase-free water and tested for quality using a 2% (wt/vol.) agarose gel in

TAE buffer. OGT-FLAG-HA mouse was generated using C57/BL6 wildtype zygotic

pronuclear and cytoplasmic injections as previously described.[9,48,49] For pronuclear

injections, 2.5 ng/µL sgRNA, 10 ng/µL ssODN, and 5 ng/µL Cas9 mRNA in TE buffer

(10 mM Tris-HCl, 0.1 mM Na$_2$EDTA, pH 8.0) [final concentrations: 0.36 ng/µL sgRNA,

1.43 ng/µL ssODN, and 0.71 ng/µL Cas9 mRNA] were injected into the nucleus of

mouse zygotes. For the cytoplasmic injections, 50 ng/µL sgRNA, 100 ng/µL ssODN, and

100 ng/µL Cas9 mRNA in TE buffer [final concentrations per embryo: 7.14 ng/µL

sgRNA, 14.3 ng/µL ssODN, and 14.3 ng/µL Cas9 mRNA] were injected into the

cytoplasm of mouse zygotes. Following injection, the zygotes were grown to the two-cell

stage and implanted into pseudopregnant foster mothers (up to 30 two-cell zygotes).[9]

Approximately 19.5 days after implantation, the pups were delivered, and 3 weeks after

birth the pups were tailed and separated by gender.

Genotyping of the OGT-FLAG-HA mice originally involved direct sequencing

and DNA agarose gel genotyping. The DNA samples were isolated from tail tips using

the DNeasy Blood & Tissue Kit. Then, PCR amplification was performed on 50 ng of template genomic DNA using the Q5® Hot Start High-Fidelity DNA polymerase and OGT-FH_geno_F1/OGT-FH_geno_R1 primer pair [98°C for 30 sec; [98°C for 10 sec; 65°C for 20 sec; 72°C for 30 sec (34 cycles)], and then 72°C for 2 min] producing the following amplicon: chrX:101681261+101682926 (https://tinyurl.com/ya2zt875). After amplification, the amplicon was either submitted for sequencing at Laragen, Inc. using the sequencing primer OGT-FH_Seq_R or incubated with BamHI-HF® enzyme for two hours at 37°C. After amplicon cleavage, the DNA was separated using agarose electrophoresis (1.5% agarose gel in TAE buffer) to distinguish wildtype (WT, 1666 bp band), heterozygotes (HET, 1068 bp + 658 bp bands), and homozygotes (HOM, 1666 bp + 1068 bp + 658 bp bands). The primers used were as follows: OGT-FH_geno_F1: 5'-CCATCTCACCAGCCCAATAC-3', OGT-FH_geno_R1: 5'-ACTGACAGTGCCAAGCATTA-3'; and OGT-FH_Seq_R: 5'-ACTGATATAGGCTCATGTGGTTT-3'.

For off-target sequence validation, we performed PCR on 50 ng of the DNeasy-derived gDNA using the Q5® Hot Start High-Fidelity DNA polymerase (New England BioLabs, Inc.) and the off_F/off_R primer pair [98°C for 30 sec; [98°C for 10 sec; 67°C for 20 sec; 72°C for 30 sec (34 cycles)], and then 72°C for 2 min] producing the following amplicon near the *Rhox11* gene: chrX:38076598+38085139 (https://tinyurl.com/ybpvwll2). Then, the amplicon and the off_F primer were submitted for sequencing at Laragen, Inc. The primers used were as follows: off_F: 5'-CTTCGTGGGTTTAAGGCCGA-3' and off_R: 5'-CTCACACAGGTTTGTGAGTTGAAG-3'.

3.6.7 Reagents and materials for *O*-GlcNAcome

All chemicals and reagents were of analytical grade, obtained from Millipore Sigma, and used without further purification unless specified. The NuPAGE[TM] Novex[TM] 4-12% Bis-Tris protein gels (1.0 mm, 10-well), Imperial[TM] protein stain, *O*-GlcNAcylated peptide TAPT(gS)TIAPG, high capacity Neutravidin agarose resin, Alexa Fluor® 680 conjugated streptavidin, Alexa Fluor® 680 Goat Anti-Rabbit IgG (H+L), Alexa Fluor® 790 Goat Anti-Rabbit IgG (H+L), Alexa Fluor® 680 Goat Anti-Mouse IgG (H+L), Alexa Fluor® 790 Goat Anti-Mouse IgG (H+L), Accela LC, PAL autosampler, highly cross-adsorbed, spin columns, and C18 desalting tips were purchased from ThermoFisher Scientific (Waltham, MA). All protein concentrations were measured using the BCA assay (ThermoFisher Scientific). Immobilon-FL PVDF membrane (EMD Millipore, 0.45 μM), gelatin from porcine skin (Type A, lyophilized powder, γ-irradiated, BioXtra, suitable for cell culture), α-crystallin from bovine eye lens were obtained from Millipore Sigma (St. Louis, MO). The CORTECS UPLC C18+ column (2.1 x 50 mm) was received from Waters Corp. (Milford, MA). Thiamet G was obtained from Tocris Biosciences (Avonmouth, Bristol, UK). cOmplete[TM] protease inhibitor cocktail without EDTA (PIC-EDTA) was purchased from Roche Diagnostics Corp. (Indianapolis, IN). Baculovirus preparation and protein expression of short-form OGT (sOGT) in *Spodoptera frugiperda* (Sf9) cells was performed as previously described.[50] Cerebral cortices were obtained from adult C57BL/6 mice bred in the Caltech Animal Facility according to NIH guidelines. RapiGest and UDP-GalNAz **1** were synthesized as referenced.[51,52] Y289L GalT was expressed and purified as described previously.[53] Tris(3-hydroxypropyltriazolylmethyl)amine (THPTA), alkyne-Dde-biotin **2**, and the

photocleavable alkyne-biotin (alkyne-PC-biotin) were purchased from Click Chemistry

Tools (Scottsdale, AZ). The Odyssey scanner was purchased from LI-COR Biosciences.

3.6.8 *O*-GlcNAcylated peptide labeling

The labeling protocol was adapted from a previously reported method.[54] The

peptide TAPT(gS)TIAPG (20 μM final) was dissolved in a 200 μL solution of 10 mM

HEPES pH 7.9, 5.5 mM MnCl$_2$, 1 mM UDP-GalNAz **1**, and 100 ng/μL Y289L GalT and

rotated end-over-end overnight at 4°C. Prior to enzyme addition, an aliquot was removed

as an initial time point for LC-MS analysis. The reaction was acidified to 0.1% TFA,

desalted using a C18 tip, and an aliquot was saved for analysis. The labelled peptide (10

μM final) was diluted into a 400 μL solution of 10 mM sodium phosphate pH 7.6, 100

μM alkyne-Dde-biotin **2**, 2 mM sodium ascorbate, and 100 μM THPTA. CuSO$_4$ was

added (1 mM final), and the reaction was incubated while rotating end-over-end at RT for

1 h. After removing a sample, the reaction was acidified and desalted again. The peptide

(10 μM) was then split into fractions of 50 μL containing 25 mM sodium phosphate pH

7.6 and either 1% RapiGest, 6 M urea, or 2% hydrazine monohydrate and incubated for 1

h at RT. Samples were acidified, desalted, and subjected to LC-MS analysis.

3.6.9 LC-MS analysis of *O*-GlcNAc peptide labeling

Liquid chromatography and mass spectrometry (LC-MS) were performed using

an LTQ linear ion trap mass spectrometer combined with an Accela LC and PAL

autosampler. Approximately 10 pmol peptide from each sample was injected onto a

CORTECS UPLC C18+ column (2.1 x 50 mm). Flow rate was set at 0.4 mL/min. Solvent

A (ddH$_2$O, 1% formic acid) and Solvent B (acetonitrile, 1% formic acid) were used to

create a gradient. The gradient consisted of 0-0.2 min, 5% B; 0.2-3.5 min 5-65% B, 3.5-

4.0 min 65% B with injection into the MS starting at 0.2 min to avoid salt contamination.

All peptide products were found to elute during the linear gradient between 0.2 and 2.0

min. For the biotinylated and cleaved peptide, alkyne reagent **2** was not sufficiently

removed by the C18 tips. Therefore, the reaction was monitored using an extracted ion

chromatogram by extracting all ions within ± 1 *m/z* of the calculated masses.

3.6.10 Chemoenzymatic labeling using Dde and photocleavable linkers

Labeling with Y289L GalT and UDP-GalNAz **1** was conducted as previously

described.[26] Briefly, 500 µg HEK-293T cell lysate in 1% SDS, 1x PBS pH 7.4 (10 mM

Na_2HPO_4, 1.8 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl), 10 µM Thiamet-G, and 1x

PIC-EDTA was diluted to a protein concentration of 1 mg/mL using 1% SDS, 1x PBS pH

7.4. For the α-crystallin and sOGT labeling, 200 µg cortical lysate was spiked with 20 µg

of α-crystallin and 5 µg of sOGT. Cortical lysate was obtained from adult 2 month old

mice and dounced 10 times in 2% SDS, 1x PBS, 10 µM Thiamet-G, and 1x PIC-EDTA

and then diluted to a protein concentration of 1 mg/mL using 1% SDS, 1x PBS pH 7.4,

10 µM Thiamet-G, and 1x PIC-EDTA. Protein was precipitated by adding three volumes

of methanol, one volume of chloroform, and two volumes of ddH_2O with vortex mixing

after each addition and then pelleted at the aqueous-organic interface by centrifuging at

21,000 x *g* for 5 min. The top, aqueous layer was removed, and one volume of methanol

was added with mixing. The protein was pelleted again, and all liquid was removed. After

the pellets were air-dried, samples were redissolved at 5 mg/mL (100 µL) in 1% SDS, 20

mM HEPES pH 7.9 by sonication. To each sample, the following were added in the given

order: 10 µL of 50x PIC-EDTA, 112.5 µL of ddH_2O, 200 µL of 2.5x labeling buffer (50

mM HEPES pH 7.9, 125 mM NaCl, 5% NP-40), and 27.5 µL of 100 mM $MnCl_2$.

Samples were briefly mixed by pipetting and placed on ice. Next, 25 μL of 0.5 mM UDP-GalNAz **1** was added followed by pipetting to mix. Finally, 25 μL of 2 mg/mL Y289L GalT was added, and samples were rotated end-over-end for 16 h at 4 °C. Proteins were then precipitated again, and pellets were air dried. Pellets were next dissolved at 4 mg/mL (125 μL) in 1% SDS, 20 mM HEPES pH 7.9. An aliquot of the GalT-labeled sample was removed (25 μL), and the remaining sample was split in two (50 μL each) to be labeled with either alkyne-Dde-biotin **2** or alkyne-PC-biotin. All manipulations with the photocleavable linker were performed in the dark.

To each sample, the following were added in the given order with mixing: 4 μL of 50x PIC-EDTA, 78 μL of ddH$_2$O, and 10 μL of 20x PBS pH 7.4. Next, the CuAAC reagents were added with vortex mixing after each addition: 4 μL of 5 mM alkyne-dde-biotin **2** or alkyne-PC-biotin (stock in DMSO), 4 μL of 100 mM sodium ascorbate (freshly prepared), 10 μL of 2 mM THPTA (stock in 4:1 tBuOH/DMSO), and 4 μL of 50 mM CuSO$_4$ (freshly prepared). Samples were rotated end-over-end for 1 h at RT, and the reaction was halted by the addition of 25 μL EDTA pH 8.0. Samples were precipitated, and the pellet was washed once with 1 mL MeOH to remove residual, unreacted linker. The pellet was then air-dried and redissolved at 4 mg/mL (50 μL) in 1% SDS, 20 mM HEPES pH 7.9. An aliquot of each sample was reserved (25 μL), and the remaining sample was cleaved. For the alkyne-Dde-biotin sample, the mixture was diluted to 1 mg/mL with 2% hydrazine monohydrate, and the sample was rotated end-over-end for 1 h at RT. For the alkyne-PC-biotin sample, the protein was diluted to 1 mg/mL with ddH$_2$O, and the liquid was irradiated from the open top of the tube (2 cm distance) with 365 nm UV light (UVGL-25 handheld UV lamp, 1.5 mW/cm^2) for 1 h at RT with mixing

every 10 min. Both samples were then precipitated by addition of four volumes of -20 °C

acetone and storage at -20 °C for 1 h. Samples were air-dried and then redissolved at 4

mg/mL (25 µL) in 1% SDS, 20 mM HEPES, pH 7.9.

3.6.11 Coomassie staining and western blotting

Aliquots corresponding to 20 µg protein (5 µL) for each sample was resolved by

SDS-PAGE as follows. Samples were added to 10 µL ddH$_2$O and 5 µL 4x SDS-PAGE

loading buffer (200 mM Tris pH 6.8, 400 mM DTT, 8% SDS, 0.4% bromophenol blue,

40% glycerol) and were then used directly without boiling to avoid cleaving the linkers.

The mixtures were loaded in duplicate on NuPAGE^{TM} Novex^{TM} 4-12% Bis-Tris protein

gels and separated using protein gel electrophoresis in MOPS buffer. One duplicate was

then stained with Imperial^{TM} protein stain according to the manufacturer's specifications

and imaged by an Odyssey scanner. The other set of samples was transferred onto an

Immobilon-FL PVDF membrane in transfer buffer and blocked for 1 h at RT with 5%

bovine serum albumin (BSA) in 1x TBST. The blot was then incubated with 1:20,000

Alexa Fluor® 680 conjugated streptavidin in 5% BSA/TBST for 1 h at RT, washed three

times with TBST for 5 min, and imaged with an Odyssey scanner.

3.6.12 Enrichment and elution of labeled proteins

Labeled samples were diluted to 1 mL using 1x PBS pH 7.4, 1x PIC-EDTA. For

each sample, 25 µL (settled volume) of high-capacity Neutravidin agarose was washed

twice with 500 µL of 1x PBS in spin columns, and samples were added to the washed

beads. Mixtures were rotated end-over-end for 1 h at RT. Lysate was removed by

centrifugation at 2,000 x g for 30 s. The beads were washed with 1% SDS (5 x 0.5 mL), 6

M urea (5 x 0.5 mL), and 1x PBS pH 7.4 (5 x 0.5 mL). Beads were then resuspended in

50 μL of 2% hydrazine monohydrate in ddH$_2$O and rotated end-over-end for 1 h at RT. The elution volume was removed, and beads were washed with 50 μL of 1x PBS pH 7.4. The wash volume was combined with the elution volume, and samples were flash-frozen and stored at -80 °C prior to analysis.

3.6.13 *O*-GlcNAcome sample processing for MS analysis

Samples were thawed and precipitated by addition of four volumes of -20 °C acetone. Samples were stored at -20 °C for 1 h and centrifuged at 21,000 x *g* for 5 min. Pellets were redissolved in 20 μL of 8 M urea, 100 mM Tris pH 8.0, 10 mM DTT and incubated at 60 °C with shaking for 20 min. Cysteine residues were blocked by addition of 25 mM iodoacetamide for 45 min. Samples were diluted four-fold with 100 mM Tris pH 8.0. Samples were split in two and digested with 0.01 mg/mL trypsin or chymotrypsin for 4-16 h at 37 °C. A portion of the trypsin digests was further digested with 7 μg/mL AspN for 6 h at 37 °C. Digests were acidified to a final concentration of 0.5% formic acid, 0.05% TFA.

3.6.14 LC separation and MS analysis

The digests were analyzed by nanoLC/MS on the LTQ-Velos with a 0 to 30% B in 120 min gradient with top 5 MS/MS (A: ddH$_2$O, 0.1% formic acid; B: acetonitrile, 0.1% formic acid). Samples were desalted on a 360 x 100 μm Kasil fritted pre-column (2 cm Monitor C18) prior to separation on a 360 x 75 μm (10 cm BEH130 C18, 1.7 μm) analytical column/tip. Full scan MS was acquired at 60,000 resolution followed by top 5 tandem MS in the linear ion trap alternating between ETD and CID modes of the same precursor. The ETD reaction time was 100 ms with supplemental activation. RAW files were converted to MGF files for Mascot searching using Proteome Discoverer with CID

and ETD spectra extracted to separate MFG files. Data was searched against a custom database with fixed modifications of carbamidomethyl (C) and variable mods of oxidation (M) and a custom modification for the tagged *O*-GlcNAc. The custom modification was defined as addition of C(19) H(30) N(6) O(10) to Ser or Thr (net addition of 502.202341 Da). For CID, a scoring neutral loss of C(19) H(30) N(6) O(10) was included, but this was omitted for ETD. Enzyme specificity was trypsin (KR), chymotrypsin (FLYW), or trypsin-AspN_ND (cleave C-term KR and N-term ND). Mass tolerances were 25 ppm and 0.8 Da for precursor and fragments ions, respectively. The instrument type was chosen as either ESI-TRAP or ETD-TRAP. Search results were combined in Scaffold 4.4 Proteome Software, filtered for 80% peptide confidence and modifications manually evaluated.

3.7 References

1   Hart, G.W., Slawson, C., Ramirez-Correa, G. & Lagerlof, O. Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease. *Annu Rev Biochem* (2011) **80**, 1:825-858.
2   Pathak, S. *et al.* The active site of O-GlcNAc transferase imposes constraints on substrate sequence. *Nat Struct Mol Biol* (2015) **22**, 9:744-750.
3   Yang, X., Zhang, F. & Kudlow, J.E. Recruitment of O-GlcNAc transferase to promoters by corepressor mSin3A: coupling protein O-GlcNAcylation to transcriptional repression. *Cell* (2002) **110**, 1:69-80.
4   Cheung, W.D., Sakabe, K., Housley, M.P., Dias, W.B. & Hart, G.W. O-linked beta-N-acetylglucosaminyltransferase substrate specificity is regulated by myosin phosphatase targeting and other interacting proteins. *J Biol Chem* (2008) **283**, 49:33935-33941.
5   Chen, Q., Chen, Y., Bian, C., Fujiki, R. & Yu, X. TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature* (2013) **493**, 7433:561-564.
6   Rafie, K. *et al.* Recognition of a glycosylation substrate by the O-GlcNAc transferase TPR repeats. *Open Biol* (2017) **7**, 6.
7   Dalvai, M. *et al.* A Scalable Genome-Editing-Based Approach for Mapping Multiprotein Complexes in Human Cells. *Cell Rep* (2015) **13**, 3:621-633.
8   Hsu, P.D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* (2013) **31**, 9:827-832.
9   Yang, H., Wang, H. & Jaenisch, R. Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. *Nat Protoc* (2014) **9**, 8:1956-1968.
10   Leney, A.C., El Atmioui, D., Wu, W., Ovaa, H. & Heck, A.J.R. Elucidating crosstalk mechanisms between phosphorylation and O-GlcNAcylation. *Proc Natl Acad Sci U S A* (2017) **114**, 35:E7255-E7261.
11   Ruan, H.B., Nie, Y. & Yang, X. Regulation of protein degradation by O-GlcNAcylation: crosstalk with ubiquitination. *Mol Cell Proteomics* (2013) **12**, 12:3489-3497.

12   Wang, J., Torii, M., Liu, H., Hart, G.W. & Hu, Z.Z. dbOGAP - an integrated bioinformatics resource for protein O-GlcNAcylation. *BMC Bioinformatics* (2011) **12**, 1:91.

13   Nandi, A. *et al.* Global identification of O-GlcNAc-modified proteins. *Anal Chem* (2006) **78**, 2:452-458.

14   Clark, P.M. *et al.* Direct in-gel fluorescence detection and cellular imaging of O-GlcNAc-modified proteins. *J Am Chem Soc* (2008) **130**, 35:11576-11577.

15   Skorobogatko, Y.V. *et al.* Human Alzheimer's disease synaptic O-GlcNAc site mapping and iTRAQ expression proteomics with ion trap mass spectrometry. *Amino Acids* (2011) **40**, 3:765-779.

16   Alfaro, J.F. *et al.* Tandem mass spectrometry identifies many mouse brain O-GlcNAcylated proteins including EGF domain-specific O-GlcNAc transferase targets. *Proc Natl Acad Sci U S A* (2012) **109**, 19:7280-7285.

17   Iyer, S.P. & Hart, G.W. Roles of the tetratricopeptide repeat domain in O-GlcNAc transferase targeting and protein substrate specificity. *J Biol Chem* (2003) **278**, 27:24608-24616.

18   Pekkurnaz, G., Trinidad, J.C., Wang, X., Kong, D. & Schwarz, T.L. Glucose regulates mitochondrial motility via Milton modification by O-GlcNAc transferase. *Cell* (2014) **158**, 1:54-68.

19   Ohn, T., Kedersha, N., Hickman, T., Tisdale, S. & Anderson, P. A functional RNAi screen links O-GlcNAc modification of ribosomal proteins to stress granule and processing body assembly. *Nat Cell Biol* (2008) **10**, 10:1224-1231.

20   Zeidan, Q., Wang, Z., De Maio, A. & Hart, G.W. O-GlcNAc cycling enzymes associate with the translational machinery and modify core ribosomal proteins. *Mol Biol Cell* (2010) **21**, 12:1922-1936.

21   Deng, R.P. *et al.* Global identification of O-GlcNAc transferase (OGT) interactors by a human proteome microarray and the construction of an OGT interactome. *Proteomics* (2014) **14**, 9:1020-1030.

22   Ruan, H.B. *et al.* O-GlcNAc transferase/host cell factor C1 complex regulates gluconeogenesis by modulating PGC-1alpha stability. *Cell Metab* (2012) **16**, 2:226-237.

23   Woo, C.M., Iavarone, A.T., Spiciarich, D.R., Palaniappan, K.K. & Bertozzi, C.R. Isotope-targeted glycoproteomics (IsoTaG): a mass-independent platform for intact N- and O-glycopeptide discovery and analysis. *Nat Methods* (2015) **12**, 6:561-567.

24   Ma, J. *et al.* O-GlcNAcomic Profiling Identifies Widespread O-Linked beta-N-Acetylglucosamine Modification (O-GlcNAcylation) in Oxidative Phosphorylation System Regulating Cardiac Mitochondrial Function. *J Biol Chem* (2015) **290**, 49:29141-29153.

25   Khidekel, N. *et al.* A Chemoenzymatic Approach toward the Rapid and Sensitive Detection of O-GlcNAc Posttranslational Modifications. *J. Am. Chem. Soc.* (2003) **125**, 52:16162-16163.

26   Clark, P.M. *et al.* Direct in-gel fluorescence detection and cellular imaging of O-GlcNAc-modified proteins. *J. Am. Chem. Soc.* (2008) **130**, 35:11576-11577.

27   Ramakrishnan, B. & Qasba, P.K. Structure-based design of beta 1,4-galactosyltransferase I (beta 4Gal-T1) with equally efficient N-acetylgalactosaminyltransferase activity: point mutation broadens beta 4Gal-T1 donor specificity. *J Biol Chem* (2002) **277**, 23:20833-20839.

28   McKay, C.S. & Finn, M.G. Click chemistry in complex mixtures: bioorthogonal bioconjugation. *Chem Biol* (2014) **21**, 9:1075-1101.

29   Szychowski, J. *et al.* Cleavable biotin probes for labeling of biomolecules via azide-alkyne cycloaddition. *J Am Chem Soc* (2010) **132**, 51:18351-18360.

30   Wang, Z. *et al.* Enrichment and Site Mapping of O-Linked N-Acetylglucosamine by a Combination of Chemical/Enzymatic Tagging, Photochemical Cleavage, and Electron Transfer Dissociation Mass Spectrometry. *Mol. Cell Proteomics* (2010) **9**, 1:153-160.

31   Alfaro, J.F. *et al.* Tandem mass spectrometry identifies many mouse brain O-GlcNAcylated proteins including EGF domain-specific O-GlcNAc transferase targets. *Proc. Natl. Acad. Sci. USA* (2012) **109**, 19:7280-7285.

32   Woo, C.M., Iavarone, A.T., Spiciarich, D.R., Palaniappan, K.K. & Bertozzi, C.R. Isotope-targeted glycoproteomics (IsoTaG): a mass-independent platform for intact N- and O-glycopeptide discovery and analysis. *Nat. Methods* (2015) **12**, 6:561-569.

33   Zaro, B.W., Yang, Y.-Y., Hang, H.C. & Pratt, M.R. Chemical reporters for fluorescent detection and identification of O-GlcNAc-modified proteins reveal glycosylation of the ubiquitin ligase NEDD4-1. *Proc. Natl. Acad. Sci. U.S.A.* (2011) **108**, 20:8146-8151.

34   Myers, S.A., Daou, S., Affar, E.B. & Burlingame, A.L. Electron transfer dissociation (ETD): The mass spectrometric breakthrough essential for O-GlcNAc protein site assignments—a study of the O-GlcNAcylated protein …. *Proteomics* (2013).

35   Khidekel, N. *et al.* Probing the dynamics of O-GlcNAc glycosylation in the brain using quantitative proteomics. *Nature Chemical Biology* (2007) **3**, 6:339-348.

36   Bycroft, B.W., Chan, W.C., Chhabra, S.R. & Hone, N.D. A Novel Lysine-Protecting Procedure for Continuous-Flow Solid-Phase Synthesis of Branched Peptides. *J. Chem. Soc. Chem. Comm.* (1993) 9:778-779.

37   Chhabra, S.R., Parekh, H., Khan, A.N., Bycroft, B.W. & Kellam, B. A Dde-based carboxy linker for solid-phase synthesis. *Tet. Lett.* (2001) **42**, 2189-2192.

38   Yang, Y. & Verhelst, S.H.L. Cleavable trifunctional biotin reagents for protein labelling, capture and release. *Chem. Comm.* (2013) **49**, 47:5366-5368.

39   Chalkley, R.J. & Burlingame, A.L. Identification of GlcNAcylation sites of peptides and alpha-crystallin using Q-TOF mass spectrometry. *Journal of the American Society for Mass Spectrometry* (2001) **12**, 10:1106-1113.

40   Hanover, J.A. *et al.* Mitochondrial and nucleocytoplasmic isoforms of O-linked GlcNAc transferase encoded by a single mammalian gene. *Arch Biochem Biophys* (2003) **409**, 2:287-297.

41   Lazarus, B.D., Love, D.C. & Hanover, J.A. Recombinant O-GlcNAc transferase isoforms: identification of O-GlcNAcase, yes tyrosine kinase, and tau as isoform-specific substrates. *Glycobiology* (2006) **16**, 5:415-421.

42   Roquemore, E.P. *et al.* Vertebrate lens alpha-crystallins are modified by O-linked N-acetylglucosamine. *J Biol Chem* (1992) **267**, 1:555-563.

43   Tai, H.C., Khidekel, N., Ficarro, S.B., Peters, E.C. & Hsieh-Wilson, L.C. Parallel identification of O-GlcNAc-modified proteins from cell lysates. *J Am Chem Soc* (2004) **126**, 34:10500-10501.

44   Zeytuni, N. & Zarivach, R. Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module. *Structure* (2012) **20**, 3:397-405.

45   Iyer, S.P. & Hart, G.W. Roles of the tetratricopeptide repeat domain in O-GlcNAc transferase targeting and protein substrate specificity. *J. Biol. Chem.* (2003) **278**, 27:24608-24616.

46   Boersema, P.J., Raijmakers, R., Lemeer, S., Mohammed, S. & Heck, A.J.R. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat. Protocols* (2009) **4**, 4:484-494.

47   Sung, M.-K., Reitsma, J.M., Sweredoski, M.J., Hess, S. & Deshaies, R.J. Ribosomal proteins produced in excess are degraded by the ubiquitin–proteasome system. *Mol Biol Cell* (2016) **27**, 17:2642-2652.

48   Behrenger, R., Gertsenstein, M., Nagy, K.V. & Nagy, A. *Manipulating the Mouse Embryo: A Laboratory Manual*. Fourth edn, 5, 6, and 7 (Cold Spring Harbor Laboratory Press, 2014).

49   Yang, H. *et al.* One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell* (2013) **154**, 6:1370-1379.

50   Tai, H.-C., Khidekel, N., Ficarro, S.B., Peters, E.C. & Hsieh-Wilson, L.C. Parallel Identification of *O*-GlcNAc-Modified Proteins from Cell Lysates. *J. Am. Chem. Soc.* (2004) **126**, 10500-10501.

51   Lee, P.J.J. & Compton, B.J. Destructible surfactants and uses thereof. (2007) US patent 7,229,539.

52   Hang, H.C., Yu, C., Pratt, M.R. & Bertozzi, C.R. Probing glycosyltransferase activities with the Staudinger ligation. *J. Am. Chem. Soc.* (2004) **126**, 6-7.

53   Clark, P.M., Rexach, J.E. & Hsieh-Wilson, L.C. in *Curr Prot Chem Biol*   (John Wiley & Sons, Inc., 2009).

54   Khidekel, N. *et al.* A chemoenzymatic approach toward the rapid and sensitive detection of O-GlcNAc posttranslational modifications. *J. Am. Chem. Soc.* (2003) **125**, 52:16162-16163.

Chapter 4

The roles of *O*-GlcNAc and CREB in the transcription of key neuronal gene networks

4.1 The histone and PTM codes

Transcription must integrate a variety of extracellular signals in order to faithfully execute distinct biological outcomes. A critical mechanism conferring biological fate specificity is through the post-translational modification (PTM) of transcriptional regulators such as histones. The 'histone code hypothesis' involves acetylation, methylation, phosphorylation, glycosylation, and many other modifications of histones at over a hundred sites.[1] Combinations of these modifications are integrated to modulate DNA accessibility and recruit transcriptional factors to control biological function.[1-5] In neurons, distinct epigenetic codes dictate everything from neurodevelopment and neuronal activity to neurodegenerative diseases.[6-8] Similarly, transcription factors (TFs) are modified by a variety of PTMs, which confer a PTM 'code'.[9]

In order to impart different neuronal outcomes, two potential mechanisms could be in play: either (1) certain subsets of modifications confer distinct biological consequences or (2) each individual modification contributes to activity through additive effects to influence net activity at particular sites.[4] While the individual contributions of each of these PTMs have been explored in isolation, their combinatorial integration to elicit specific transcriptional destinies are still poorly understood.[2]

4.2 The role of *O*-GlcNAc in the epigenetic code

The "epigenetic code" refers to the integration of the histone code and DNA methylation code in order to govern transcription and biological fates.[10] OGT has been shown to regulate chromatin remodeling (Figure 4.1). Chromatin, where genomic DNA envelopes histone octamers, can be made more (euchromatin) or less (heterochromatin) accessible to transcription through the dynamic PTM regulation of the histone code.[11]

165

Generally, acetylation of histone lysines is a dynamic modification that increases the accessibility of chromatin for transcriptional activation.[12] Methylation often at the same histone lysine residues is a longer-lived modification that is associated with heterochromatin and gene silencing with some exceptions (notably, H3K4 methylation is activating).[13] DNA methylation on cytosine residues at CpG dinucleotides is an important epigenetic regulation that impairs the binding of many TFs to DNA and recruits proteins including histone methyltransferases.[14,15] The net result of DNA methylation is to confer an epigenetic "memory" whereby critical genes are silenced during various cellular processes and in particular, during development.[16]

In 2009, it was discovered that super sex combs, an essential component of the polycomb repressive complex 2 (PRC2), is actually *O*-GlcNAc transferase in drosophila.[17] PRC2 mediates the silencing of chromatic regions through the methylation of histone H3 on lysine 27 and is necessary for the proper *O*-GlcNAcylation of proteins.[18,19] In addition, OGT *O*-GlcNAcylates and stabilizes EZH2 (enhancer of zeste 2 polycomb repressive complex 2 subunit), a component of the PRC2 complex.[20] A major component of polycomb repressive complex 1 (PRC1), RING1B or E3 ubiquitin ligase RING1, preferentially associates with neuronal genes or cell cycle genes when it is *O*-GlcNAc glycosylated or non-glycosylated respectively.[21] OGT further influences PRC1 activity by directly *O*-GlcNAcylating and protecting the PRC1 oncogene Bmi-1 (BMI1 proto-oncogene, polycomb ring finger) from proteosomal degradation.[22] In these ways, *O*-GlcNAc can alter the transcriptional regimes regulated by PRC1 influencing cellular pluripotency, differentiation, and proliferation.

In addition, OGT is also a major component of the SET1/COMPASS histone H3K4 methylation complex, whose activity is important for elevating transcription.[23] Part of the SET1/COMPASS (histone-lysine *N*-methyltransferase Set1/complex proteins associated with Set1) complex, the heavily *O*-GlcNAc glycosylated host-cell factor 1 (HCF1) is activated upon proteolytic cleavage, which is catalyzed by OGT itself.[24] Another component of the SET1/COMPASS complex, MLL5 (histone-lysine N-methyltransferase 2E (mixed lineage leukemia 5)), is glycosylated at S435 and T440, which prevents its proteosomal degradation.[25] Additionally, OGT interacts with and mediates the activity of the mSin3A/HDAC1 (SIN3 transcriptional regulator family member A/histone deacetylase 1) complex, which mediates transcriptional repression by deacetylating histones.[26] *O*-GlcNAc glycosylation of HDAC1 regulates its enzymatic activity; indeed, ablation of the *O*-GlcNAc sites on HDAC1 leads to a slight reduction in HDAC1 activity.[27] CARM1 (coactivator-associated arginine methyltransferase 1) and OGT interact to regulate mitosis through the methylation of histone 3 R2, R17, and R38.[28,29] Finally, studies have shown that OGT forms a complex with NSL3 (nonspecific lethal protein 3), a component of the histone acetyltransferase complex that glycosylates histone 4 at K5, K8, and K16.[30] The glycosylation of NSL3 stabilizes the protein and increases the methylation of histone 4. The interdependence of OGT and these transcriptional complexes results in a complex interplay that define the cellular fate.

OGT further regulates transcriptional activity by directly modifying histones.[31-39] Initially, Hart and colleagues identified *O*-GlcNAc glycosylation sites on Thr101 of H2A, Ser36 of H2B, Ser47 of H4, and H3.[40] However, some studies have found that detection of *O*-GlcNAc on certain site is sensitive to specific growth or stimulation conditions (like

cell cycle state), enrichment, mass spectrometry, and western blotting conditions.[33,36,39] Furthermore, the conservation of site-specific histone *O*-GlcNAcylation can vary by across species with one study demonstrating that Ser40 *O*-GlcNAcylation on H2A evolved with viviparity.[36]

Despite the difficulty and variability in detection of histone *O*-GlcNAcylation, researchers have begun to dissect the functions of specific histone glycosylation events. For example, one study found that histone H2B *O*-GlcNAcylation at Ser112 facilitates the ubiquitination of histone H2B at Lys120, which both lead to active transcription.[39] The phosphorylation of OGT by AMPK (5'adenosine monophosphate-activated protein kinase) downstream of glucose stimulation leads to impaired chromatin association and a reduction in *O*-GlcNAc and monoubiquitination of H2B.[31] *O*-GlcNAc glycosylation of H2A at threonine 101 perturbs the dimerization of H2A and H2B resulting in euchromatin and enhanced transcription.[41] *O*-GlcNAc glycosylation of H3S10 is dependent on cell cycle and directly precludes with phosphorylation at the same site.[35] Furthermore, glycosylation on the same histone at T32 prevents the cell cycle-dependent phosphorylation events at S10 and S28 (and T32 through direct competition).[42] *O*-GlcNAcylation of H2AX S139 occurs in response to DNA damage, competes directly with S139 phosphorylation, and most importantly, demarcates the DNA damage region preventing its suffusion.[43] Moreover, OGT plays a critical role in cellular senescence through the modification of HIRA (histone cell cycle regulator) and Ubn1 (ubinuclein 1), consituents of the HIRA chaperone complex necessary for installation of histone variant 3.3.[44] *O*-GlcNAcylation of HIRA and Ubn1 is necessary for the appropriate formation and timing of heterochromatic, senescence-related nucleosome assembly.[44] Either

through direct modification or through the targeting of its interacting partners, OGT dynamically determines the structure of chromatin and ultimately the transcriptional availability of genes.

In addition to OGT's aforementioned role in chromatin structure, deletion of *O*-GlcNAcase globally perturbs chromatin structure, chromatin *O*-GlcNAcylation, and gene expression.[45,46] Interestingly, OGA also contains a pseudo-histone acetyltransferase (HAT) domain that contains high homology with other HATs, but lacks the requisite acetyl-coenzyme A (acetyl-CoA) binding residues.[47] The function of the pseudo-HAT domain of OGA remains to be determined especially since it appears to be unnecessary for the proper *O*-GlcNAcase enzymatic activity.[47,48] Moreover, *O*-GlcNAcylation has been shown to mediate the selectivity of the nuclear pore and be necessary in order maintain the nuclear pore integrity through its modification and stabilization of nucleoporins, the major structural elements that comprise the nuclear pore.[49] Together, OGA and OGT act in concert to regulate global gene transcription, chromatin structure, and nuclear pore stability and selectivity.

In addition, OGT has been shown to be a major regulator of DNA methylation as well. As discussed earlier, methylation of DNA at 5-methylcytosine (5mC) in CpG islands or gene bodies leads to gene silencing.[50] The Tet (ten-eleven translocation methylcytosine dioxygenase) family catalyzes the conversion of 5mC to 5-hydroxymethylcytosine (5hmC) and then to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC), which are required for base excision repair-mediated removal of 5-methylcytosine.[50,51] The Tet protein family is made up of Tet1, Tet2, and Tet3, which have been shown to be involved in neuronal differentiation, activity, memory

formation, and neurodegeneration.[52-56] OGT and the Tet proteins interact and enhance each others activity across the genome.[23,57-62] In particular, the *O*-GlcNAcylation of Tet1 prevents the proteosomal degradation of Tet1.[62] In addition, the interaction of OGT with Tet3 increases its stability and DNA association.[63] Through regulation of Tet activity and localization, OGT regulates transcriptional activity.
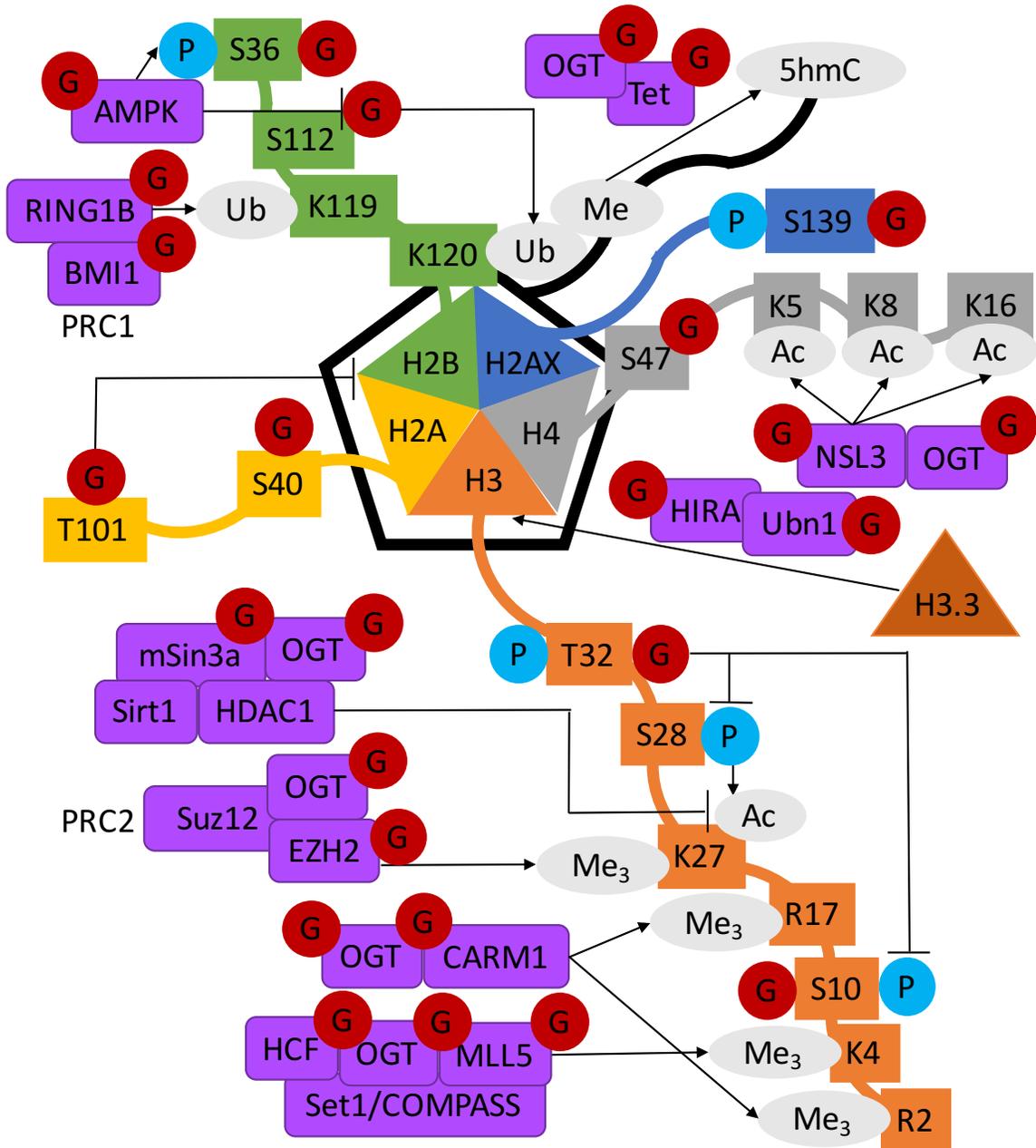
Figure 4.1 The histone and epigenetic codes are heavily regulated by OGT. **H2A** (in yellow) has been shown to be glycosylated at S40, which is associated with vivaparity. Glycosylation at H2A T101 disrupts the association of H2A and H2B leading to euchromatin and transcriptional activation. **H2B** (in green) is glycosylated at S36, which directly competes with AMPK-catalyzed phosphorylation at the same site. AMPK, which itself is activated through glycosylation, phosphorylates OGT, which reduces the glycosylation of H2B at another site, S112. H2B glycosylation at S112 enhances ubiquitination at K120, which is associated with transcriptional activation. **H2AX** (in blue) is glycosylated at S139 in response to DNA damage, which abrogates phosphorylation at the same site and delimits the DNA damage remodeling boundaries. **H4** (in grey) is glycosylated at S47, but its function is currently unknown. **H3** (in orange) is glycosylated at T32 and T10, which directly competes with cell-cycle dependent phosphorylation at the same sites. In addition, H3 T32 glycosylation indirectly reduces the phosphorylation at S28.

     In addition to directly modifying histones, OGT interacts and influences the activity of several histone and DNA modifying complexes. OGT interacts with the polycomb repressor complex 2 (PRC2) and modifies and stabilizes the EZH2 protein, which directly methylates H3K27 leading to gene silencing. In addition, OGT glycosylates, interacts with, and stabilizes HCF and MLL5, which are members of the Set1/COMPASS H3K4 methylation complex. Methylation via the Set1/COMPASS complex at H3K4 leads to enhanced transcriptional activity. The H2B K119 residue is ubiquitinated by the RING1B protein, the active enzyme in the PRC1. This ubiquitination event is associated with gene silencing. The *O*-GlcNAc glycosylation of RING1B causes RING1B to associate with neuronal genes while unglycosylated RING1B is associated with cell cycle genes. Another component of the PRC1, BMI1, is glycosylated and stabilized by OGT. OGT also enhances the activity of NSL3, the histone acetyltransferase responsible for acetylating histone 4 and therefore increasing chromatin accessibility. *O*-GlcNAcylation of the histone chaperone complex proteins, HIRA and Ubn1, is necessary for the proper installation of the senescence-related histone 3.3 variant. OGT directly glycosylates and stabilizes Tet proteins, which catalyze the removal of DNA methylation. The removal of CpG methylation leads to transcriptional activation.

Abbreviations and legend: G (red) = *O*-GlcNAc; P (turquoise) = phosphate; Me/Me3 (pale grey) = methyl/trimethyl group; 5hmC (pale grey) = 5'-hydroxy methylcytosine; Ac (pale grey) = acetyl group; Ub (pale grey) = ubiquitin; OGT = *O*-GlcNAc transferase; H# = histone #; PRC = polycomb repressive complex; AMPK = 5'adenosine monophosphate-activated protein kinase; Tet = ten-eleven translocation methylcytosine dioxygenase; RING1B = E3 ubiquitin-protein ligase RING1; BMI1 = BMI1 Proto-Oncogene, Polycomb Ring Finger; CARM1 = coactivator-associated arginine methyltransferase 1; Set1 = histone-lysine N-methyltransferase Set1; COMPASS = complex proteins associated with Set1; MLL5 = histone-lysine N-methyltransferase 2E (mixed lineage leukemia 5); HCF = host cell factor; EZH2 = enhancer of zeste 2 polycomb repressive complex 2 subunit; Suz12 = SUZ12 polycomb repressive complex 2 subunit; HIRA = histone cell cycle regulator; Ubn1 = ubinuclein 1; mSin3a = histonedeacetylase complex subunit Sin3a; Sirt1 = sirtuin 1; HDAC1 = histone deacetylase 1; NSL3 = nonspecific lethal protein 3

4.3 The role of *O*-GlcNAc in the PTM code

     Aside from influencing chromatin structure, *O*-GlcNAcylation plays an essential and specific role in transcriptional regulation.[64,65] The major component of the transcriptional machinery, RNA polymerase II (pol II), is heavily modified by O-GlcNAc, which is associated with pre-initiation complexes while loss of O-GlcNAcylation on pol II is linked to elongation.[66,67] *O*-GlcNAcase (OGA), the enzyme responsible for removing *O*-GlcNAc, is an pol II elongation factor itself.[68] In these ways,

O-GlcNAcylation and the enzymes that cycle it are essential regulators of basal transcription.

OGT and OGA dynamically cycle *O*-GlcNAc glycosylation on almost all transcription factors in eukaryotes.[34] *O*-GlcNAc glycosylation and phosphorylation have been shown to coordinate cellular function and transcriptional activity in tandem.[69] As discussed in Chapter 1, *O*-GlcNAc glycosylation and phosphorylation compete for the same serine and threonine residues leading to a "yin-yang" competitive relationship.[69] A classic example of this "yin-yang" relationship occurs on the TF c-myc (MYC proto-oncogene, bHLH (basic helix-loop-helix) TF); specifically, c-myc is rapidly degraded upon phosphorylation at Thr58, but the *O*-GlcNAc glycosylation at the same site increases the stability of c-myc.[70,71] In addition to competing for the same residues, a recent paper showed that phosphorylation at neighboring residues can directly prevent *O*-GlcNAcylation especially on the conserved (pSp/T)P(V/A/T)(gS/gT).[72] Together, *O*-GlcNAc and phosphorylation are integrated into a larger PTM Switchboard that controls the activity and specificity of transcription factors.[73]

Moreover, the *O*-GlcNAc modification of NF-κB (nuclear factor kappa-light-chain enhancer of activated B cells) RelA/p65 protein at Thr352 leads to disruption of its interaction with the protein IκBα, which normally sequesters NF-κB from the nucleus and prevents its transcriptional activation.[74] *O*-GlcNAc modification of RelA at another site, Thr305, is required for p300 to acetylate and fully activate RelA.[75] Thus, *O*-GlcNAcylation of RelA at Thr352 and Thr305 is necessary for the full activation of NF-κB transcription. Furthermore, our lab along with the Baltimore lab showed that *O*-GlcNAcylation at Ser350 of another NF-κB protein, c-Rel, enhances the expression of

toll-like receptor-dependent cytokine genes, but had no impact on TNFα-regulated gene expression.[76] Therefore, *O*-GlcNAc regulates both (1) global NF- κB activity by increasing its nuclear availability and activating acetylation and (2) stimulation-specific NF-κB gene networks important for immune and stress response. Global gene expression studies are required to determine effects of glycosylation on all NF-κB-regulated processes including cell survival and synaptic plasticity.[77,78] In summary, *O*-GlcNAc and its cycling enzymes cooperate to encode complex histone and PTM codes that are decoded to alter TF global and gene-specific activity.

4.4 CREB is a key regulator of critical gene networks in neurons

The TF, CREB, or cyclic adenosine monophosphate (cAMP) response element (CRE)-binding protein is a key regulator of a variety of different gene networks in neurons. CREB homo- and heterodimers bind to the full CRE consensus sequence (TGACGTCA) with high affinity and to half CRE sites (TGACG/CGTCA) with lower affinity.[79] CpG methylation of the CRE sites disrupt CREB binding to its CRE site.[80] CREB is an intrinsically disordered protein that contains an N-terminal glutamine rich domain (Q1, residues 1-86), a short α peptide (alternatively spliced in α/δ isoforms, 87-101), a kinase-inducible domain (KID, residues 102-165), another glutamine rich domain (Q2, residues 166-282), and a C-terminal DNA-binding and dimerization domain (basic region/leucine zipper domain (bZIP), residues 283-341) (Figure 4.2).[81-84] While full length CREB is intrinsically disordered, researchers have obtained a crystal structure of the bZIP domain in complex with DNA and an NMR structure of the KID domain interacting with a coactivator, CBP (CREB binding protein).[81,85] A canonical nuclear localization sequence (NLS) resides in the bZIP domain.[86] The Q2 domain constitutively

interacts with the TATA-binding protein-associated factor, $TAF_{II}130$, and is required for transcriptional activation of CREB.[87,88] Indeed, certain isoforms of the CREB homologue, cAMP response element modulator (CREM), lacks the Q2 domain and therefore acts as a transcriptional repressor at CRE sites.[87] In addition to this basic structure of CREB, there are several different isoforms expressed including α (341 residues.), β (287 residues), and δ (327 residues, lacks the α peptide) isoforms.[89]
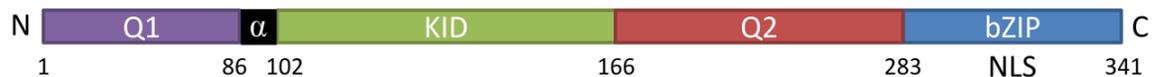
Figure 4.2 Structure of CREB. CREB contains 4 major domains: an N-terminal glutamine-rich Q1 domain (residues 1-86), kinase-inducible domain (KID) (residues 102-165), another glutamine-rich Q2 domain (residues 166-282), and a C-terminal bZIP domain (residues 283-341). The nuclear localization sequence (NLS) resides, DNA-binding, and dimerization capacity resides in the bZIP domain.

CREB has been shown to be important in memory formation from aplysia to drosophila to humans.[90] Nobel laureate Eric Kandel was the first to show that cAMP plays a role in learning and memory in aplysia.[91] Eventually, cAMP was shown to activate of protein kinase A (PKA), which in turn phosphorylated and activated the TF CREB at serine 133.[92,93] Researchers later discovered that many other stimuli and signaling pathways converge to activate and phosphorylate CREB at serine 133 including growth factors, neuronal membrane depolarization, calcium signaling, hormones, and stress (Figure 4.3).[94] Upon activation, CREB mediates the protein-synthesis independent transcription of immediate early genes (IEGs) such as *Fos* (c-Fos), *Jun* (c-Jun), and early growth response gene 1 (*Egr1*) within minutes of activation.[84] These IEGs are themselves TFs that lead to a downstream cascade of transcriptional changes that are (1) important for learning and memory and (2) last for hours after CREB's initial activation.[94] In addition to the cascade transcriptional changes promulgated by IEG expression, CREB changes the expression of synaptic growth-related genes underlying intrinsic neuronal

excitability including brain-derived neurotrophic factor (*Bdnf*), glutamate ionotropic receptor AMPA type subunit 1 (*Gria1*), and neuronal nitric oxide synthase (*Nos1*).[79] The CREB-mediated expression of these neuronal growth, activity, and excitability genes strengthens both short term memory (STM) and long term memory (LTM) formation.[95]
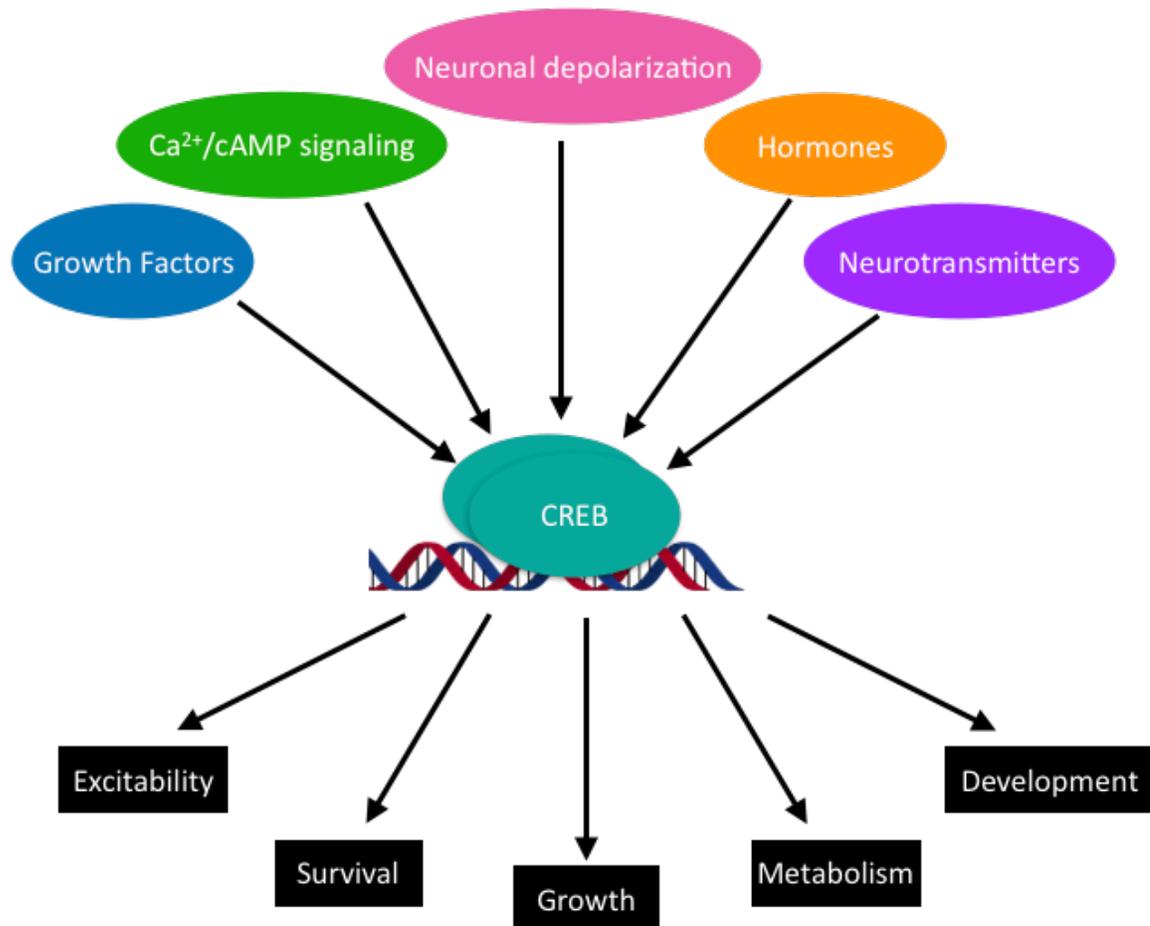
Figure 4.3 Overview of the stimuli that activate CREB and gene networks regulated by CREB activity. CREB is activated through phosphorylation at serine 133 and other mechanisms in response to growth factors, calcium, depolarization, cAMP, and neurotransmitters. CREB leads to the transcription of genes critical to neuronal excitability, survival, growth, metabolism, and development.

CREB's role in memory formation has been demonstrated repeatedly across species leading to its designation as "the memory gene".[94] Although CREB has been shown to enhance STM, CREB has mostly been shown to be important for LTM formation, a process that requires protein synthesis. Increasing expression of CREB can

lead to enhanced STM and LTM through its ability to increase neuronal excitability.[96] In neurons, either elevated CREB levels or higher intrinsic excitability preceding training leads to the preferential recruitment of these neurons into the memory trace and strengthening of the memory recall.[97-100] In accordance with these studies, the expression of the constitutively active CREB (caCREB) mutants, Y134F-CREB, DIEDML-CREB, and VP16-CREB, enhances LTM formation.[98,100-107] Reducing CREB activity through CREB knockout or the expression of a dominant-negative CREB mutant that lacks the DNA-binding domain (dnCREB, A-CREB, or K-CREB) results in severe LTM and synaptic plasticity impairments.[108-110] Integrating the results from experiments with higher and lower CREB activity lead to the proposal that there was a CREB dosage-dependent memory effect where higher CREB activity resulted in facilitated CREB activity.[111]

In addition to its role in learning and memory, CREB has been shown to have a key role in neuronal development, growth, and neurodegeneration. Günther Schütz's lab developed the first CREB[-/-] total knockout mouse, which could not be assessed for memory deficits due to its perinatal lethality resulting from a lack of sufficient pulmonary surfactant.[112] The CREB[-/-] mice were born at a reduced frequency (only 15% instead of the expected Mendelian 25% from breeding heterozygotes) and displayed reduced birth weight (70% of wildtype littermates) suggesting a developmental disadvantage.[112] Overexpressing S133A-CREB caused impaired synaptogenesis and neurogenesis demonstrating that CREB phosphorylation at serine 133 was important for neuronal and synaptic growth and development.[113] S133A-CREB mutant mice showed reduced allelic

frequencies (only 11% instead of the expected 25%) as in the CREB[-/-] mice, supporting a pivotal role of phosphorylation at serine 133 in early mouse development.[114]

Ginty and colleagues further explored the role of CREB in neuronal development by showing that CREB[-/-] mice displayed defects in axonal growth and increased neuronal apoptosis due to impairments in nerve growth factor (NGF) signaling.[115] Downstream of NGF and other growth factors, CREB regulates the expression of genes critical for neuronal growth and survival such as the anti-apoptotic protein, B-cell lymphoma 2 (*Bcl-2*), the Bcl-2 family member, *Mcl-1,* and neuronal growth factors such as insulin-like growth-factor 1 (IGF-1, *Igf1*), leptin (*Lep*), pituitary adenylate cyclase-activating peptide (PACAP, *Pacap*), *Bdnf*.[116-122] CREB also mediates the expression of the neuronal growth factor receptors including tyrosine receptor kinase B (TrkB, *Ntrk2*), the receptor for BDNF/NT-3 family of growth factors.[123] Through studies knocking out CREB, expressing dnCREB, or expressing an unphosphorylatable CREB mutant, S133A-CREB, researchers have observed widespread neuronal apoptosis through the inability to respond to and express pro-survival and growth neurotrophins including nerve growth factor (NGF), BDNF, IGF-1, and Bcl-2.[124,125] Increasing levels of wild-type (WT) CREB increases pro-survival genes (a.k.a. activity-regulated inhibitor of death (AID) genes) including *Atg3*, *Btg2, Gadd45β*, and *Gadd4γ* in mouse hippocampal neurons.[126] Through the mediation of the expression of these AID and pro-growth genes, CREB is able to direct neuronal development and protect neurons from apoptosis. Finally, CREB regulates a suite of metabolic genes in neurons including the gluconeogenic genes pyruvate carboxylase (*Pc*), phosphoenolpyruvate carboxykinase 1 (*Pepck1*), nuclear receptor subfamily 4 group A1 (*Nr4a1*), and glucose-6-phosphatase (*G6pc*).[79,127] CREB

is found within the mitochondria and mediates the major Complex I components NADH:ubiquinone oxidoreducatase subunits 2, 4, and 5 (*Nd2, Nd4, Nd5*).[128] Loss of CREB results in perturbations in the genes responsible for maintaining cholesterol homeostasis in response to neuronal activity.[129] While CREB has been shown to be important for neuronal development, survival, activity, and metabolism and mediating specific genes involved in these processes, the vast majority of CREB-mediated genes that are required for synaptic plasticity and memory formation are unknown.

Several studies have begun to delve into the global CREB-mediated gene networks underlying these processes. Genome-wide ChIP studies have shown that CREB binds to thousands of binding sites across the genome.[130-132] However, CREB-mediated transcription typically requires the recruitment of coactivators, which often means that only a small subset of genes bound by CREB are transcribed.[87] Also, despite similar genomic CREB occupancy across cell types, CREB-mediated, stimuli-specific gene expression is cell-type specific and restricted often to very few genes. For example, exposure to cAMP induces the expression of genes involved in cellular growth and survival in beta islet cells, fasting glucose and lipid metabolism in hepatocytes, and neuronal growth and activity in neurons.[131,133] A few early studies attempted to explore the transcriptome-wide, CREB-mediated memory and plasticity-associated genes and have identified a few genes including 4 differentially-expressed (DE) genes from mouse hippocampal overexpression of a constitutively active CREB and 41 DE genes from rat hippocampi following LTP induction or fear conditioning training.[102,134] A 2011 study showed that overexpressing the caCREB, VP16-CREB, for 3 or 6DIV resulted in 10% of the entire transcriptome differentially expressed in WT mouse hippocampal neurons.[133]

Another study identified 757 CREB-dependent, LTM-associated DE genes by comparing CREB knockout, CREB rescue, and CREB overexpression from WT with and without LTM training in *C. elegans.*[135] In this study, Murphy and colleagues showed that CREB basally mediated the transcription of genes involved in cellular growth, development, and metabolism without altering either longevity or memory genes. Upon LTM training, CREB altered expression of genes involved in neuronal activity, synaptic plasticity, calcium signaling, ion channels, and neurotransmitter signaling. Importantly, the genes that were mediated by CREB transcription under basal conditions were distinct and generally mutually exclusive from those regulated by LTM (only 12 genes were shared between the 463 LTM CREB-dependent genes and 281 basal CREB-dependent genes).[135] While this paper has provided great insight into CREB-mediated transcriptional networks, it remains to be determined whether or not these networks will be conserved in other species. Given the stark contrast in phenotypes in the CREB knockout between the mouse and worm (CREB KO in *C. elegans* results in no lifespan impairments while CREB KO in mice is perinatal lethal), the role of CREB will likely diverge in many ways the mouse and worm. These studies begin to probe the CREB-regulon and demonstrated that CREB regulates the expression of many different gene subsets, but much remains to be determined about the activity-dependent global gene networks regulated by CREB.

4.5 The CREB family of transcription factors: CREB, ATF1 and CREM

CREB cooperates with a suite of coactivators in order to regulate its transcriptional targets. We mentioned in the previous section that CREB can form heterodimers and bind to CRE sites with other TFs. These partners include two highly

179

homologous TFs, ATF1 or activating transcription factor 1 (65% homology) and CREM or cAMP response element modulator.[136] ATF1 and CREM share high sequence homology within the bZIP domain (91% for ATF1, 75% or 95% for CREM depending on alternative splicing) and therefore, are able to heterodimerize stably.[137,138] The affinity of dimers for the CRE site is highest for CREB/CREB homodimers followed by CREB/ATF1 heterodimers and then ATF1/ATF1 homodimers.[139,140] ATF1 acts as an activator while CREM, depending on its alternative splicing, can act as an activator or repressor.[141] In particular, CREM α, β, γ, and ϵ isoforms act as repressors since they lack the Q2 domain required for recruitment of $TAF_{II}130$ while CREMτ acts as an activator.[87,141,142] CREM activity both as a repressor and as an activator has been shown to be critical for neurogenesis, learning, and memory.[143,144] Both ATF1 and CREM activity is modulated by cAMP as they both have a PKA consensus site, but their ability to activate or repress transcription varies by cell type and expression levels.[141,145-147] Importantly, both CREM and ATF1 in concert with CREB has been shown to be critical in neuronal development since $CREB^{-/-}/ATF1^{-/-}$ and brain-specific $CREB^{-/-}/CREM^{-/-}$ double mutants were embryonic lethal.[148,149]

Despite ATF1, CREB, and CREM sharing considerable homology with other bZIP family members (such as the activator protein 1 (AP-1) family including c-Jun and c-Fos), ATF1, CREB, and CREM cannot heterodimerize with them, leading to a distinct classification of ATF1, CREB, and CREM into the "CREB family" of TFs.[137] The AP-1 family can heterodimerize with other AP-1 members, bind to the AP-1 site on DNA, and be activated through phosphorylation by c-Jun N-terminal kinase (JNK).[150] Unlike the rest of the CREB family of TFs, the homologous ATF2, ATF3, and ATF4 TFs can

heterodimerize with the AP-1 family, are activated by phosphorylation by JNK, but remain understudied, so it is still unclear if they act as activators or repressors. While little is known about the roles of ATF2-4, they have been shown to role similar to CREM, as a dominant negative regulator of CREB in the emotional response to stress and amphetamines.[151,152] In addition, ATF2-4 has been shown to heterodimerize with c-Jun and change the affinity of the heterodimer from AP-1 sites to CRE sites.[153] Furthermore, a more recent study demonstrated that ATF4 and c-Jun are important mediators of long-term facilitation, a process akin to LTP in aplysia.[154] We discussed earlier how CREB activity leads to the expression of the AP-1 family members, c-Jun and c-Fos, suggesting that extensive crosstalk occurs between the CREB and AP-1 TF families. In summary, much remains unknown about the individual contributions of CREB and its family members to learning and memory as well as any potential crosstalk within and without the CREB family.

4.6 CREB coactivators: CBP/p300 and CRTCs

Aside from the recruitment of RNA polymerase II, $TAF_{II}130$, important for basal transcription discussed in section 4.4, CREB interacts with and requires the activity of several key transcriptional coactivators in order to fully activate its gene networks. Phosphorylation of CREB at S133 strengthens the binding of the coactivators CBP and p300.[155] In addition, studies have shown that transcriptional activation is dictated by the interaction between the CREB KID and the KIX domain of CREB-binding protein.[156] The mechanisms by which CBP and p300 enzymatic activity enhances CREB-mediated transcription are two-fold: CBP and p300 acetylate (1) histones leading to increased chromatin accessibility, and (2) CREB directly at K91, K94, and K136, which increases

CREB trans-activation.[157,158] It is important to note that CBP and p300 also bind to a variety of other transcription factors and the pair can mediate distinct genes and cellular fates.[159] Finally, a ChIP-Seq experiment revealed that the genomic occupancy of CBP in neurons highly dependent on neuronal stimulation.[160] Indeed, both CBP and p300 have been shown to be important for LTM, and CBP has been shown to also be necessary for STM as well.[161-164] While there is some debate over whether or not CREB can recruit RNA polymerase II independent of phosphorylation and CBP association, some studies have demonstrated that CBP reinforces the interaction with RNA polymerase II through direct association with RNA helicase A.[165-168] The debate still continues over whether CBP/p300 is required for basal CREB transcription or solely for activity-induced CREB transcription.

CRTCs (cAMP-regulated transcriptional coactivator, formerly known as TORCs) are a family of coactivators that have proven to be critical regulators of CREB functions. CRTCs bind to the bZIP domain of CREB where arginine 314 has been shown to be critical for mediating that interaction.[169] Prior to stimulation, CRTCs is phosphorylated at S171, S275, and S307 by AMPK, which leads to its association with 14-3-3 proteins and cytoplasmic sequestration.[170] Upon cAMP and calcium stimulation, calcineurin dephosphorylates CRTCs, leading to their translocation to the nucleus where they can interact with CREB and mediate transcription.[171] The three members of the CRTC family, CRTC1, CRTC2, and CRTC3, all share an N-terminal CREB-binding domain (CBD), a central regulatory domain, an alternative splicing domain, and a transactivation domain.[79] All three CRTCs have been shown to mediate the CREB regulome through binding of the CRTC CBD and CREB bZIP domain to the CRE site (with a 2:2:1 of

stoichiometry), which facilitates the complex's association with $TAF_{II}130$.[169,172] The CRTCs have been shown to be nutrient sensors and master regulators of metabolic processes including gluconeogenesis, lipogenesis, mitochondrial activity, and stress.[173,174] In particular, CRTC1 was shown to reduce the expression of lipogenic genes including fatty acid synthase (*Fasn*) and nuclear sterol regulatory element binding protein 1c (*Srebp1c*).[173] CRTC2 has been shown to maintain glucose homeostasis through the facilitation of gluconeogenesis in response to fasting in the liver.[175] In addition, CRTCs has been shown to be an important regulator of lifespan and circadian rhythm.[176,177] The CRTC mediation of lifespan has been shown to be regulated through mitochondrial metabolism and catecholamine signaling.[178,179] Aside from the role of CRTCs as metabolic and lifespan regulators, CRTC1 and CRTC2 have been shown to be calcium and cAMP coincidence detectors, which, in the case of CRTC1, leads to the facilitation of synaptic plasticity and LTM.[171,180,181] Following learning, CRTC1 is translocated from synapses to the nucleus and modulates the strength of memory formation through regulation of CREB-mediated transcription of fibroblast growth factor 1b (*Fgf1b*), an important mediator of hippocampal-dependent associative memory.[182] Importantly, several studies have demonstrated that CRTC1 translocation is required for associative and context-dependent LTM.[182-184]

Due to the demonstrated roles of both CBP/p300 and CRTCs in memory formation, it has been proposed that the full activation of CREB for memory formation involves two phosphorylation-related events: (1) phosphorylation of CREB at serine 133 leading to the recruitment CBP/p300 to CRE sites and (2) dephosphorylation of the CRTCs, leading to their translocation into the nucleus where they can mediate

transcription.[79] While most studies have focused on either CBP/p300 or CRTCs, a few studies have begun to dissect the interplay and differential contribution of these two sets of coactivators on CREB transcription. One study found that the presence of CRTC2 could rescue CREB-mediated transcription of specific genes in CBP/p300 null cells.[185] A study performed in drosophila by the Saitoe lab showed that spaced training memory formation required the activity of CBP while fasting-dependent learning required CRTC2 activity suggesting that CBP and CRTC2 regulate distinct memory paradigms.[181] In a follow-up to this study, the Saitoe lab later showed that LTM formation requires the shifting from CBP-dependent to CRTC-dependent CREB transcription.[186] Another study showed that associative learning was enhanced by the recruitment of both CBP and CRTC to CREB on CRE sites, but CRTC alone is recruited in a graded fashion proportional to the amount of training.[182] Altogether, these studies suggest that both CBP and CRTCs are important for memory formation, but further studies are necessary to tease apart the individual contributions of each, their potential interdependence, and the memory-related transcriptional networks mediated by both.[84]

4.7 The role of CREB and its coactivators in neurodegeneration

Consistent with CREB's critical role in neuronal activity and homeostasis, CREB and its coactivators have been shown to be deregulated in neurodegenerative diseases.[187] As mentioned in section 4.4.1, knocking out CREB (and CREM) or expressing a dominant negative form of CREB in the brain leads to apoptosis and neurodegeneration.[149,188] In AD mouse models and post-mortem human AD brains, lower *Creb1* and *Bdnf* expression and/or CREB and BDNF protein levels were observed.[189] Furthermore, treatment of rat hippocampal neurons with Aβ peptides decreased CREB

activity.[189] On the other hand, increasing CREB activity in CA1 region rescues spatial memory deficits due to AD and normal aging as well.[190,191] While expression of slightly more active CREB (Y134F/DIEDML-CREB) leads to enhanced LTM and neuronal growth, elevated CREB activity through the overexpression of the constitutively active VP16-CREB can lead to excitotoxicity-induced neurodegeneration.[102,106,192] In addition, higher CREB transcription was associated with Huntington's disease (HD) in a mouse model although CREB has also been shown to play a neuroprotective role in mHtt-expressing mice.[193,194] Contrary to this study, other studies have suggested that elevated CREB transcription is neuroprotective and diminished CREB activity is neurodegenerative in HD model mice.[195-197] Furthermore, CREB activity has been shown to be impaired in several different human diseases including Coffin-Lowry syndrome and Rubenstein-Taybi syndrome (RSTS).[137] Altogether, these studies show that fine tuning CREB activity is necessary for maintaining neuronal homeostasis and preventing neurodegeneration (Figure 4.4).
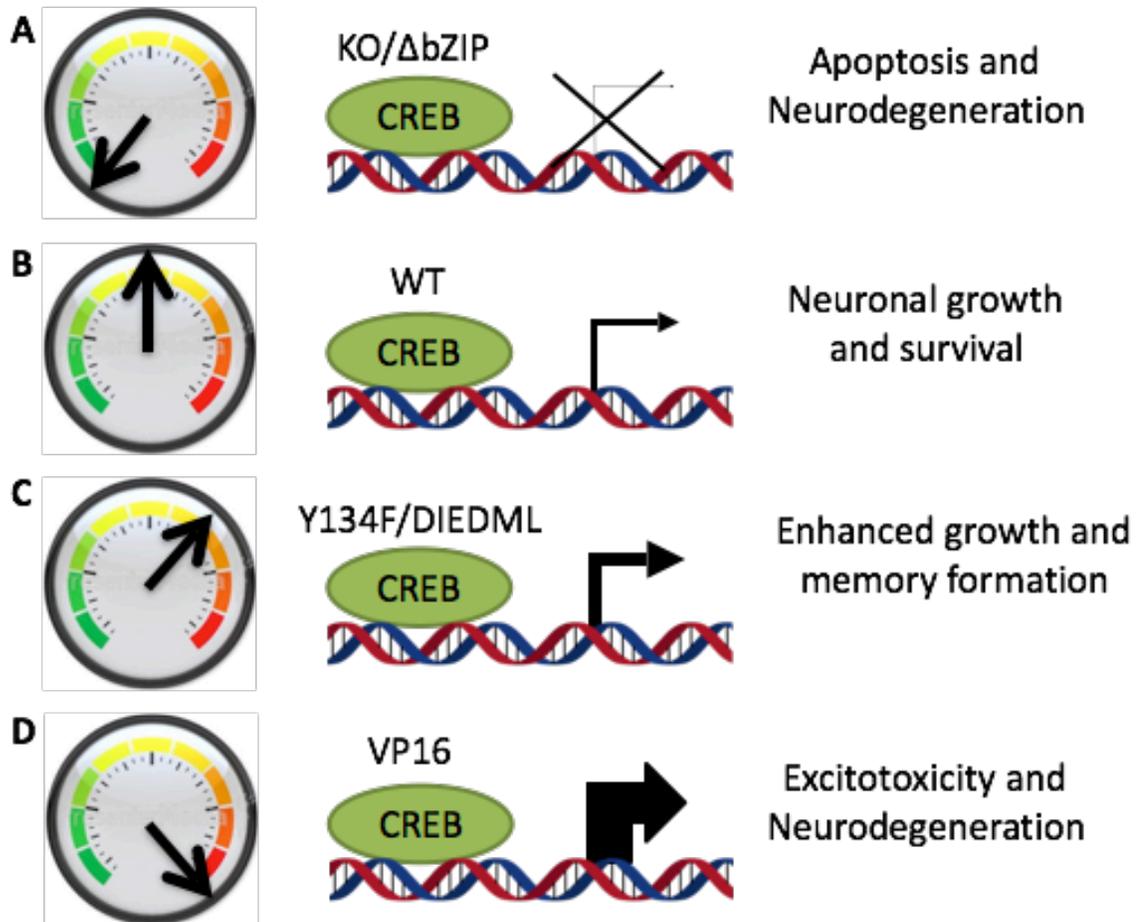
Figure 4.4 Fine tuning CREB activity is critical for neuronal growth and survival. Shown here is a summary of the different CREB mutants and their relative activity levels. (A) The lowest activity of CREB occurs in CREB knockout or the expression of the dominant negative mutants that lack the DNA-binding domain (dnCREB→ K-CREB or A-CREB) and leads to apoptosis-induced neurodegeneration. (B) WT CREB is able to regulate the transcription of neuronal growth and survival genes leading to thriving neurons. (C) The expression of the slightly more active Y134F/DIEDML mutants yields enhanced growth and memory formation. (D) Ramping up CREB activity through expression of the constitutively-active VP16-CREB mutant results in excitotoxicity-induced neurodegeneration.

In addition to CREB itself, CREB's coactivators CBP and CRTC have been linked to neurodegeneration. Loss of CBP or p300 has been shown to underlie Rubenstein-Taybi Syndrome (RSTS), a debilitating disease characterized by physical abnormalities and mental retardation due to histone acetylation and CREB transcriptional perturbations.[198-201] Supportive of a neuroprotective role for CBP, reduced CBP activity has been associated with HD while increased CBP function ameliorates learning and

memory deficits in a mouse model of AD.[202-204] CRTC1 has been shown to have a critical role in AD Aβ peptides disrupted calcium signaling leading to decreased dephosphorylation of CRTC1 leading to impaired CRTC- and CREB-mediated transcription in AD mouse models.[205] Saura and colleagues revealed that CRTC1-dependent transcription is impaired in AD mouse models during the early stages of AD mouse model cognitive decline, which is observed in the human hippocampus along with lower CRTC1 protein levels.[206] In two different AD mouse models (*PS* cDKO and APP$_{Swe,Ind}$), the authors showed a reversal in memory deficits upon overexpression of CRTC1 due to the recovery of CRTC1-mediated transcription.[183,206] In summary, CREB and its coactivators have proven to be critical mediators of neurodegenerative diseases underscoring the importance of precise regulation of CREB activity.

4.8 The CREB PTM code

The correct activation of CREB is critical for proper neuronal function: too little CREB activity either by expressing dominant negative CREB or knocking out CREB leads to apoptosis and neurodegeneration through a dearth of CREB-mediated pro-survival and growth genes; too much CREB activity from overexpression of the constitutively active VP16-CREB results in excitotoxicity and neurodegeneration through heightened expression of neuronal activity and excitation genes. Thus, it is critical to fine tune the activity of CREB in order to ensure proper neuronal activity and prevent neurodegeneration.

In order to modulate its activation, CREB is regulated by a variety of post-translational modifications (PTMs) including phosphorylation, acetylation, sumoylation, ubiquitylation, and glycosylation.[79] Many studies have focused on CREB's major

activity-induced phosphorylation site, S133, which resides in the kinase-inducible domain (KID). In neurons, this phosphorylation event occurs downstream of signaling in response to cAMP, calcium, hormones, growth factors, membrane depolarization, and stress.[94] Phosphorylation at S133 by protein kinase A (PKA), mitogen- and stress-activated kinase 1/2 (MSK1/2), ribosomal S6 kinase (RSK), protein kinase B (AKT), MAP kinase-activated protein kinase 2 (MAPKAP K2), and Ca^{2+}-calmodulin kinase IV (CaMKIV) recruits the coactivators, CBP/p300, which then acetylate CREB at K91, K94, and K136 and increase CREB trans-activation.[157] Several studies have shown that dephosphorylation by protein phosphatases 1 and 2A (PP1, PP2A) at S133 decreases CREB transcriptional activity.[207,208] Furthermore, overexpression of S133A-CREB results in neurodegeneration, some LTP deficits, and impaired LTM.[95,209-211] In stark contrast to these findings using S133A overexpression, S133A-CREB knock-in results in normal LTP in the hippocampus and amygdala and LTM, suggesting that phosphorylation at serine 133 may not be critical for learning and memory.[114,212] While S133A-CREB knock-in did not appear to affect learning and memory, Blendy and co-workers suggested a role of S133A-CREB in development.[114]

Despite many studies showing the importance of serine 133 phosphorylation, its transcriptional consequences are still poorly understood.[94] Several genes have been shown to be mediated by CREB phosphorylation, but CREB phosphorylation itself and the genes it regulates are stimuli- and cell type-specific.[131,213] Robust CREB phosphorylation occurs in pancreatic β cells upon exposure to cAMP, but not upon depolarization.[213] Furthermore, elevation of cAMP leads to increased expression of (1) anti-apoptotic and pro-growth genes in islet cells, (2) glucose and lipid metabolism genes

in hepatocytes, and (3) regulation of cell death and transcription factor genes in HEK293T cells.[131] In hippocampal neurons, tetracycline-inducible knock-in of S133A-CREB leads to a decrease in the levels of glutamate receptors (specifically AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid) and NMDA (*N*-Methyl-D-aspartic acid) subunits), which shifts synaptic plasticity from LTP to LTD.[95,209] In other cell types, heme oxygenase-1 (*Hmox1*), amphiregulin (*Areg*), interleukin-6 (*Il6*), and nuclear receptor subfamily 4 group A members 1 and 3 (*Nr4a1, Nr4a3*) were downregulated when S133A-CREB was expressed while *Crem* and cyclin A2 (*Ccna2*) were upregulated.[214-216]

In brain microarray and RNA-Seq studies, the transcriptional effects of S133A-CREB have been further expanded. Silva and colleagues reported only a single gene, 14-3-3ζ (*Ywhah*), that is downregulated upon overexpression of S133A-CREB following fear conditioning.[211] In the nucleus accumbens, an area important for addiction, the overexpression of S133A-CREB for 8 weeks resulted in upregulation of 4 genes and downregulation of 20 genes when compared to control WT mice without CREB overexpression.[217] The genes downregulated upon S133A-CREB overexpression included genes important for neuronal growth and development including *Bdnf*, T-box brain protein 1 (*Tbox1*), and cholecystokinin (*Cck*) in addition to genes for synaptic proteins such as synaptophysin (*Syp*) and synaptotagmin XVII (*Syt17*).[217] Antithetically, a more recent RNA-Seq study found that no genes were differentially expressed in a S133A-CREB knock-in mouse.[114] By integrating the results from the behavioral and transcriptional studies, the effects of ablating CREB phosphorylation at S133 appear to differ depending on the S133A-CREB expression method *videlicet* S133A

overexpression shows behavioral and transcriptional alterations while S133A knock-in expressed at endogenous CREB levels appears to have little or no memory or gene expression consequences. These studies reveal two major findings: (1) the level of CREB dictates CREB-mediated transcription and (2) the exact transcriptional effects of CREB phosphorylation at serine 133 in the KID remain unclear.

Also within the CREB KID, ataxia telangiectasia-mutated (ATM) kinase phosphorylates S111 followed by casein kinases 2 (CK2) phosphorylation at S108 and then by casein kinases 2 (CK1) phosphorylation at S114 and S117 in response to genotoxic stress.[218,219] Only after the sequential phosphorylation of all of these sites can ATM phosphorylate S121, which only occurs upon significant DNA damage.[219] CREB phosphorylation at S108, S111, S114, S117, and S121 (called the "ATM/CK cluster") has been shown to regulate the transcription of genes involved in metabolic homeostasis in response to cAMP and genotoxic damage (Figure 4.5).[220] In addition, phosphorylation at serine 142 by CaMKIIα or CK2 in response to Ca^{2+} signaling plays a critical role in circadian entrainment and inflammatory nociception.[221-223] Phosphorylation of the ATM/CK cluster or S142 inhibit the interaction with CBP/p300 and decrease CREB-mediated transcription.[219,220,224] Interestingly, phosphorylation of S142 *and* S143 in addition to S133 leads to full activation of Ca^{2+}- and CREB-mediated transcription in neurons despite disrupting the interaction with CBP.[221] Likewise, phosphorylation of S108, S111, and S114 in addition to S133 is required for full activation of cAMP- and CREB-mediated transcription in MEFs by mediating CREB-DNA binding, which in turn mediates its association with the co-activator CRTC2.[220] Furthermore, phosphorylation of CREB at S270 and S271 by cyclin-dependent kinase 1 (CDK1) in response to genotoxic

stress minimizes the binding of CREB to DNA thereby disrupting with downstream CREB-mediated transcription.[224] Aside from CREB phosphorylation, exposure to TNFα leads to (1) ubiquitylation and subsequent proteosomal degradation of CREB and (2) K285 and K304 SUMOylation of CREB, which spares CREB from proteolysis.[225] Altogether, these PTMs, both individually and combinatorially, determine CREB activity in response to a variety of stimuli.
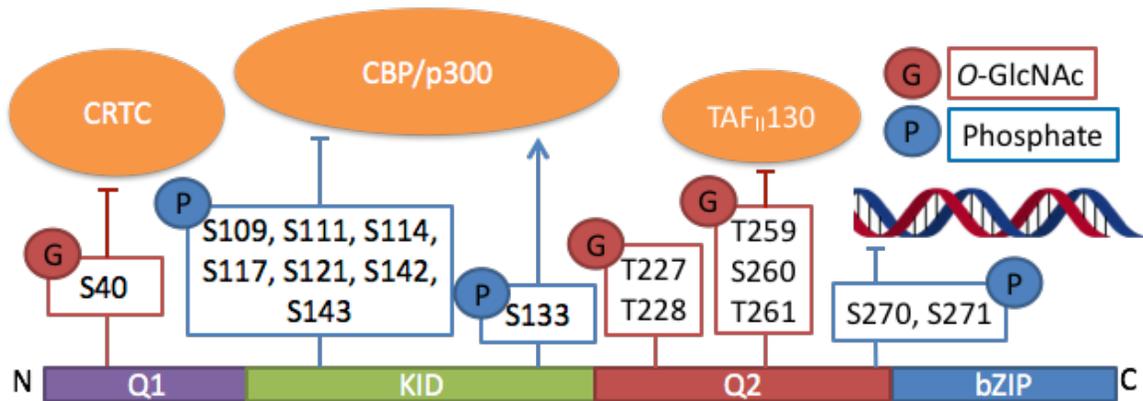
Figure 4.5 Overview of CREB phosphorylation and *O*-GlcNAc glycosylation sites. Shown here are all of the known CREB *O*-GlcNAc glycosylation and phosphorylation sites. In the *N*-terminal Q1 domain, CREB is glycosylated at serine 40, which disrupts the interaction with CRTC. Within the KID domain, phosphorylation at S133 enhances the binding of CBP/p300 to CREB while phosphorylation at S109, S111, S114, S117, S121, S142, and S142 interfere with this interaction. Glycosylation at T259, S260, T261 leads to impaired recruitment of the basal transcription factor, TAF$_{II}$130, to the Q2 domain of CREB. Finally, phosphorylation at S270 and S271 perturbs CREB binding to DNA.

Through mass spectrometric analysis, our lab identified several potential *O*-GlcNAc glycosylation sites where CREB including S40, T227/T228, and T259/S260/T261 (Figure 4.5).[226] Our lab went on to show the glycosylation of CREB at T259-T261 in the Q2 domain interferes with the association of CREB with TAF$_{II}$130.[227] The major glycosylation site on CREB, S40, is glycosylated in response to neuronal depolarization and occludes CREB's interaction with CRTC1 and CRTC2 thereby decreasing CREB-dependent transcription of certain key neuronal growth and activity genes including *Bdnf, c-fos, Wnt2,* and *Arc*.[226] Using an electrophoretic mobility shift

assay (EMSA), our lab showed that these transcriptional changes are not due to differences in the DNA binding ability of S40 glycosylated CREB. Interestingly, CREB glycosylation at S40 occurs preferentially on CREB phosphorylated at S133 suggesting interplay between CREB glycosylation and phosphorylation. Moreover, ablation of glycosylation in a S40A-CREB mutant leads to enhanced neuronal outgrowth and accelerated memory formation when compared to wild-type (WT) CREB in a fear conditioning paradigm.[226] In summary, glycosylation at S40 was shown to act as a brake to repress the expression of specific CREB-mediated genes involved in neuronal growth and activity. Therefore, S133 phosphorylation and S40 *O*-GlcNAcylation have been shown to have opposing effects on CREB transcription. However, the exact contribution of each of these PTMs to (1) overall CREB activity and (2) the CREB-mediated transcriptional networks globally have not yet been studied.

4.9 How are CREB phosphorylation and glycosylation integrated in order to confer biological outcomes?

Given the gaps in our understanding of the PTM code for CREB, several outstanding questions remain unanswered. While much is known about the modifications that adorn CREB, we still do not know how to decipher the complex PTM code that regulates CREB. Specifically, does phosphorylation at S133 globally increase and glycosylation at S40 globally decrease CREB transcription across all gene networks or are there specific gene networks affected by CREB PTMs? In addition, is there substantial interplay or overlap between the gene networks regulated by phosphorylation or glycosylation of CREB? Furthermore, are multiple PTMs on CREB integrated in order to determine cellular fate as seen in the multiple genotoxic-induced and Ca^{2+}-signaling

phosphorylation events discussed earlier? If indeed PTMs are integrated, then studying individual PTMs in isolation may provide an incomplete picture of the effects of TF PTMs. In our next chapter, we will leverage transcriptome-wide studies and targeted CREB mutants to begin to answer some of these questions toward the ultimate goal of determining the PTM code for CREB.

4.10 References

1   Su, Z. & Denu, J.M. Reading the Combinatorial Histone Language. *ACS Chem Biol* (2016) **11**, 3:564-574.
2   Ernst, J. & Kellis, M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol* (2010) **28**, 8:817-825.
3   Berger, S.L. Histone modifications in transcriptional regulation. *Curr Opin Genet Dev* (2002) **12**, 2:142-148.
4   Strahl, B.D. & Allis, C.D. The language of covalent histone modifications. *Nature* (2000) **403**, 6765:41-45.
5   Berger, S.L. The complex language of chromatin regulation during transcription. *Nature* (2007) **447**, 7143:407-412.
6   Hon, G.C. *et al.* Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat. Genet.* (2013) **45**, 10:1198-U1340.
7   Gapp, K., Woldemichael, B.T., Bohacek, J. & Mansuy, I.M. Epigenetic regulation in neurodevelopment and neurodegenerative diseases. *Neuroscience* (2014) **264**, 99-111.
8   Luo, C. *et al.* Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* (2017) **357**, 6351:600.
9   Khidekel, N. & Hsieh-Wilson, L.C. A 'molecular switchboard'--covalent modifications to proteins and their impact on transcription. *Org Biomol Chem* (2004) **2**, 1:1-7.
10   Day, J.J. & Sweatt, J.D. Epigenetic mechanisms in cognition. *Neuron* (2011) **70**, 5:813-829.
11   Narlikar, G.J., Sundaramoorthy, R. & Owen-Hughes, T. Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell* (2013) **154**, 3:490-503.
12   Verdin, E. & Ott, M. 50 years of protein acetylation: from gene regulation to epigenetics, metabolism and beyond. *Nat Rev Mol Cell Biol* (2015) **16**, 4:258-264.
13   Greer, E.L. & Shi, Y. Histone methylation: a dynamic mark in health, disease and inheritance. *Nat Rev Genet* (2012) **13**, 5:343-357.
14   Cedar, H. & Bergman, Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet* (2009) **10**, 5:295-304.
15   Suzuki, M.M. & Bird, A. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* (2008) **9**, 6:465-476.
16   Kohli, R.M. & Zhang, Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* (2013) **502**, 7472:472-479.
17   Sinclair, D.A. *et al.* Drosophila O-GlcNAc transferase (OGT) is encoded by the Polycomb group (PcG) gene, super sex combs (sxc). *Proc Natl Acad Sci U S A* (2009) **106**, 32:13427-13432.
18   Gambetta, M.C., Oktaba, K. & Muller, J. Essential role of the glycosyltransferase sxc/Ogt in polycomb repression. *Science* (2009) **325**, 5936:93-96.
19   Myers, S.A., Panning, B. & Burlingame, A.L. Polycomb repressive complex 2 is necessary for the normal site-specific O-GlcNAc distribution in mouse embryonic stem cells. *Proc Natl Acad Sci U S A* (2011) **108**, 23:9490-9495.

20   Chu, C.S. *et al.* O-GlcNAcylation regulates EZH2 protein stability and function. *Proc Natl Acad Sci U S A* (2014) **111**, 4:1355-1360.

21   Maury, J.J. *et al.* RING1B O-GlcNAcylation regulates gene targeting of polycomb repressive complex 1 in human embryonic stem cells. *Stem Cell Res* (2015) **15**, 1:182-189.

22   Li, Y. *et al.* O-GlcNAcylation modulates Bmi-1 protein stability and potential oncogenic function in prostate cancer. *Oncogene* (2017).

23   Deplus, R. *et al.* TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *EMBO J* (2013) **32**, 5:645-655.

24   Lazarus, M.B. *et al.* HCF-1 is cleaved in the active site of O-GlcNAc transferase. *Science* (2013) **342**, 6163:1235-1239.

25   Ding, X. *et al.* Mixed Lineage Leukemia 5 (MLL5) Protein Stability Is Cooperatively Regulated by O-GlcNac Transferase (OGT) and Ubiquitin Specific Protease 7 (USP7). *PLoS One* (2015) **10**, 12:e0145023.

26   Yang, X., Zhang, F. & Kudlow, J.E. Recruitment of O-GlcNAc transferase to promoters by corepressor mSin3A: coupling protein O-GlcNAcylation to transcriptional repression. *Cell* (2002) **110**, 1:69-80.

27   Zhu, G. *et al.* O-GlcNAcylation of histone deacetylases 1 in hepatocellular carcinoma promotes cancer progression. *Glycobiology* (2016) **26**, 8:820-833.

28   Sakabe, K. & Hart, G.W. O-GlcNAc transferase regulates mitotic chromatin dynamics. *J Biol Chem* (2010) **285**, 45:34460-34468.

29   Charoensuksai, P., Kuhn, P., Wang, L., Sherer, N. & Xu, W. O-GlcNAcylation of co-activator-associated arginine methyltransferase 1 regulates its protein substrate specificity. *Biochem J* (2015) **466**, 3:587-599.

30   Wu, D. *et al.* O-Linked N-acetylglucosamine transferase 1 regulates global histone H4 acetylation via stabilization of the nonspecific lethal protein NSL3. *J Biol Chem* (2017) **292**, 24:10014-10025.

31   Xu, Q. *et al.* AMPK regulates histone H2B O-GlcNAcylation. *Nucleic Acids Res* (2014) **42**, 9:5594-5604.

32   Gambetta, M.C. & Muller, J. A critical perspective of the diverse roles of O-GlcNAc transferase in chromatin. *Chromosoma* (2015) **124**, 4:429-442.

33   Gagnon, J. *et al.* Undetectable histone O-GlcNAcylation in mammalian cells. *Epigenetics* (2015) **10**, 8:677-691.

34   Hardiville, S. & Hart, G.W. Nutrient regulation of signaling, transcription, and cell physiology by O-GlcNAcylation. *Cell Metab* (2014) **20**, 2:208-213.

35   Zhang, S., Roche, K., Nasheuer, H.P. & Lowndes, N.F. Modification of histones by sugar beta-N-acetylglucosamine (GlcNAc) occurs on multiple residues, including histone H3 serine 10, and is cell cycle-regulated. *J Biol Chem* (2011) **286**, 43:37483-37495.

36   Hirosawa, M. *et al.* Novel O-GlcNAcylation on Ser(40) of canonical H2A isoforms specific to viviparity. *Sci Rep* (2016) **6**, 1:31785.

37   Dehennaut, V., Leprince, D. & Lefebvre, T. O-GlcNAcylation, an Epigenetic Mark. Focus on the Histone Code, TET Family Proteins, and Polycomb Group Proteins. *Front Endocrinol (Lausanne)* (2014) **5**, 155.

38   Lewis, B.A. & Hanover, J.A. O-GlcNAc and the epigenetic regulation of gene expression. *J Biol Chem* (2014) **289**, 50:34440-34448.

39   Fujiki, R. *et al.* GlcNAcylation of histone H2B facilitates its monoubiquitination. *Nature* (2011) **480**, 7378:557-560.

40   Sakabe, K., Wang, Z. & Hart, G.W. Beta-N-acetylglucosamine (O-GlcNAc) is part of the histone code. *Proc Natl Acad Sci U S A* (2010) **107**, 46:19915-19920.

41   Lercher, L. *et al.* Generation of a synthetic GlcNAcylated nucleosome reveals regulation of stability by H2A-Thr101 GlcNAcylation. *Nat Commun* (2015) **6**, 7978.

42   Fong, J.J. *et al.* beta-N-Acetylglucosamine (O-GlcNAc) is a novel regulator of mitosis-specific phosphorylations on histone H3. *J Biol Chem* (2012) **287**, 15:12195-12203.

43   Chen, Q. & Yu, X. OGT restrains the expansion of DNA damage signaling. *Nucleic Acids Res* (2016) **44**, 19:9266-9278.

194

44   Lee, J.S. & Zhang, Z. *O*-linked N-acetylglucosamine transferase (OGT) interacts with the histone chaperone HIRA complex and regulates nucleosome assembly and cellular senescence. *Proc Natl Acad Sci U S A* (2016) **113**, 23:E3213-3220.

45   Akan, I., Love, D.C., Harwood, K.R., Bond, M.R. & Hanover, J.A. Drosophila O-GlcNAcase Deletion Globally Perturbs Chromatin O-GlcNAcylation. *J Biol Chem* (2016) **291**, 19:9906-9919.

46   Keembiyehetty, C. Disruption of O-GlcNAc cycling by deletion of O-GlcNAcase (Oga/Mgea5) changed gene expression pattern in mouse embryonic fibroblast (MEF) cells. *Genom Data* (2015) **5**, 30-33.

47   Li, B., Li, H., Lu, L. & Jiang, J. Structures of human O-GlcNAcase and its complexes reveal a new substrate recognition mode. *Nat Struct Mol Biol* (2017) **24**, 4:362-369.

48   Elsen, N.L. *et al.* Insights into activity and inhibition from the crystal structure of human O-GlcNAcase. *Nat Chem Biol* (2017) **13**, 6:613-615.

49   Zhu, Y. *et al.* Post-translational O-GlcNAcylation is essential for nuclear pore integrity and maintenance of the pore selectivity filter. *J Mol Cell Biol* (2016) **8**, 1:2-16.

50   Hahn, M.A. *et al.* Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis. *Cell Rep* (2013) **3**, 2:291-300.

51   Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* (2011) **333**, 6047:1300-1303.

52   Rudenko, A. *et al.* Tet1 is critical for neuronal activity-regulated gene expression and memory extinction. *Neuron* (2013) **79**, 6:1109-1122.

53   Zhang, R.R. *et al.* Tet1 regulates adult hippocampal neurogenesis and cognition. *Cell Stem Cell* (2013) **13**, 2:237-245.

54   Santiago, M., Antunes, C., Guedes, M., Sousa, N. & Marques, C.J. TET enzymes and DNA hydroxymethylation in neural development and function - how critical are they? *Genomics* (2014) **104**, 5:334-340.

55   Jin, S.G. *et al.* Tet3 Reads 5-Carboxylcytosine through Its CXXC Domain and Is a Potential Guardian against Neurodegeneration. *Cell Rep* (2016) **14**, 3:493-505.

56   Wu, X. & Zhang, Y. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.* (2017) **advance online publication**.

57   Zhang, Q. *et al.* Differential regulation of the ten-eleven translocation (TET) family of dioxygenases by O-linked beta-N-acetylglucosamine transferase (OGT). *J Biol Chem* (2014) **289**, 9:5986-5996.

58   Bauer, C. *et al.* Phosphorylation of TET proteins is regulated via O-GlcNAcylation by the O-linked N-acetylglucosamine transferase (OGT). *J Biol Chem* (2015) **290**, 8:4801-4812.

59   Chen, Q., Chen, Y., Bian, C., Fujiki, R. & Yu, X. TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature* (2013) **493**, 7433:561-564.

60   Vella, P. *et al.* Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. *Mol Cell* (2013) **49**, 4:645-656.

61   Balasubramani, A. & Rao, A. O-GlcNAcylation and 5-methylcytosine oxidation: an unexpected association between OGT and TETs. *Mol Cell* (2013) **49**, 4:618-619.

62   Shi, F.T. *et al.* Ten-eleven translocation 1 (Tet1) is regulated by O-linked N-acetylglucosamine transferase (Ogt) for target gene repression in mouse embryonic stem cells. *J Biol Chem* (2013) **288**, 29:20776-20784.

63   Ito, R. *et al.* TET3-OGT interaction increases the stability and the presence of OGT in chromatin. *Genes Cells* (2014) **19**, 1:52-65.

64   Wu, D., Cai, Y. & Jin, J. Potential coordination role between O-GlcNAcylation and epigenetics. *Protein Cell* (2017).

65   Hanover, J.A., Krause, M.W. & Love, D.C. Bittersweet memories: linking metabolism to epigenetics through O-GlcNAcylation. *Nat Rev Mol Cell Biol* (2012) **13**, 5:312-321.

66   Ranuncolo, S.M., Ghosh, S., Hanover, J.A., Hart, G.W. & Lewis, B.A. Evidence of the involvement of O-GlcNAc-modified human RNA polymerase II CTD in transcription in vitro and in vivo. *J Biol Chem* (2012) **287**, 28:23549-23561.

67   Lu, L. *et al.* Distributive O-GlcNAcylation on the Highly Repetitive C-Terminal Domain of RNA Polymerase II. *Biochemistry* (2016) **55**, 7:1149-1158.

68   Resto, M. *et al.* O-GlcNAcase Is an RNA Polymerase II Elongation Factor Coupled to Pausing Factors SPT5 and TIF1beta. *J Biol Chem* (2016) **291**, 43:22703-22713.

69   Hart, G.W., Slawson, C., Ramirez-Correa, G. & Lagerlof, O. Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease. *Annu Rev Biochem* (2011) **80**, 1:825-858.

70   Kamemura, K., Hayes, B.K., Comer, F.I. & Hart, G.W. Dynamic interplay between O-glycosylation and O-phosphorylation of nucleocytoplasmic proteins: alternative glycosylation/phosphorylation of THR-58, a known mutational hot spot of c-Myc in lymphomas, is regulated by mitogens. *J Biol Chem* (2002) **277**, 21:19229-19235.

71   Swamy, M. *et al.* Glucose and glutamine fuel protein O-GlcNAcylation to control T cell self-renewal and malignancy. *Nat Immunol* (2016) **17**, 6:712-720.

72   Leney, A.C., El Atmioui, D., Wu, W., Ovaa, H. & Heck, A.J.R. Elucidating crosstalk mechanisms between phosphorylation and O-GlcNAcylation. *Proc. Natl. Acad. Sci. U. S. A.* (2017).

73   Lothrop, A.P., Torres, M.P. & Fuchs, S.M. Deciphering post-translational modification codes. *FEBS Lett* (2013) **587**, 8:1247-1257.

74   Yang, W.H. *et al.* NFkappaB activation is associated with its O-GlcNAcylation state under hyperglycemic conditions. *Proc Natl Acad Sci U S A* (2008) **105**, 45:17345-17350.

75   Allison, D.F. *et al.* Modification of RelA by O-linked N-acetylglucosamine links glucose metabolism to NF-kappaB acetylation and transcription. *Proc Natl Acad Sci U S A* (2012) **109**, 42:16888-16893.

76   Ramakrishnan, P. *et al.* Activation of the transcriptional function of the NF-kappaB protein c-Rel by O-GlcNAc glycosylation. *Sci Signal* (2013) **6**, 290:ra75.

77   Napetschnig, J. & Wu, H. Molecular basis of NF-kappaB signaling. *Annu Rev Biophys* (2013) **42**, 1:443-468.

78   Meffert, M.K., Chang, J.M., Wiltgen, B.J., Fanselow, M.S. & Baltimore, D. NF-kappa B functions in synaptic signaling and behavior. *Nat Neurosci* (2003) **6**, 10:1072-1078.

79   Altarejos, J.Y. & Montminy, M. CREB and the CRTC co-activators: sensors for hormonal and metabolic signals. *Nat Rev Mol Cell Biol* (2011) **12**, 3:141-151.

80   Yin, Y. *et al.* Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* (2017) **356**, 6337.

81   Schumacher, M.A., Goodman, R.H. & Brennan, R.G. The crystal structure of a CREB bZIP-SSCRE complex reveals the basis for selective dimerization and divalent cation-enhanced DNA binding. *J. Biol. Chem.* (2000) **275**, 35242-35247.

82   Richards, J.P., Bachinger, H.P., Goodman, R.H. & Brennan, R.G. Analysis of the structural properties of cAMP-responsive element-binding protein (CREB) and phosphorylated CREB. *J Biol Chem* (1996) **271**, 23:13716-13723.

83   Quinn, P.G. Distinct activation domains within cAMP response element-binding protein (CREB) mediate basal and cAMP-stimulated transcription. *Journal of Biological Chemistry* (1993) **268**, 23:16999-17009.

84   Mayr, B. & Montminy, M. Transcriptional regulation by the phosphorylation-dependent factor CREB. *Nat Rev Mol Cell Biol* (2001) **2**, 8:599-609.

85   Radhakrishnan, I. *et al.* Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. *Cell* (1997) **91**, 6:741-752.

86   Waeber, G. & Habener, J.F. Nuclear translocation and DNA recognition signals colocalized within the bZIP domain of cyclic adenosine 3',5'-monophosphate response element-binding protein CREB. *Mol Endocrinol* (1991) **5**, 10:1431-1438.

87   Ferreri, K., Gill, G. & Montminy, M. The cAMP-regulated transcription factor CREB interacts with a component of the TFIID complex. *Proc Natl Acad Sci U S A* (1994) **91**, 4:1210-1213.

88   Saluja, D., Vassallo, M.F. & Tanese, N. Distinct subdomains of human TAFII130 are required for interactions with glutamine-rich transcriptional activators. *Mol Cell Biol* (1998) **18**, 10:5734-5743.

89   Blendy, J.A., Kaestner, K.H., Schmid, W., Gass, P. & Schutz, G. Targeting of the CREB gene leads to up-regulation of a novel CREB mRNA isoform. *EMBO J* (1996) **15**, 5:1098-1106.

90   Carlezon, W.A., Jr., Duman, R.S. & Nestler, E.J. The many faces of CREB. *Trends Neurosci* (2005) **28**, 8:436-445.

91   Brunelli, M., Castellucci, V. & Kandel, E.R. Synaptic facilitation and behavioral sensitization in Aplysia: possible role of serotonin and cyclic AMP. *Science* (1976) **194**, 4270:1178-1181.

92   Dash, P.K., Hochner, B. & Kandel, E.R. Injection of the cAMP-responsive element into the nucleus of Aplysia sensory neurons blocks long-term facilitation. *Nature* (1990) **345**, 6277:718-721.

93   Montminy, M.R. & Bilezikjian, L.M. Binding of a nuclear protein to the cyclic-AMP response element of the somatostatin gene. *Nature* (1987) **328**, 6126:175-178.

94   Benito, E. & Barco, A. CREB's control of intrinsic and synaptic plasticity: implications for CREB-dependent memory models. *Trends Neurosci* (2010) **33**, 5:230-240.

95   Middei, S. *et al.* CREB is necessary for synaptic maintenance and learning-induced changes of the AMPA receptor GluA1 subunit. *Hippocampus* (2013) **23**, 6:488-499.

96   Yiu, A.P. *et al.* Neurons are recruited to a memory trace based on relative neuronal excitability immediately before training. *Neuron* (2014) **83**, 3:722-735.

97   Kim, J., Kwon, J.-T., Kim, H.-S., Josselyn, S.A. & Han, J.-H. Memory recall and modifications by activating neurons with elevated CREB. *Nat. Neurosci.* (2014) **17**, 1:65-72.

98   Dong, Y. *et al.* CREB modulates excitability of nucleus accumbens neurons. *Nat Neurosci* (2006) **9**, 4:475-477.

99   Han, J.H. *et al.* Neuronal competition and selection during memory formation. *Science* (2007) **316**, 5823:457-460.

100   Zhou, Y. *et al.* CREB regulates excitability and the allocation of memory to subsets of neurons in the amygdala. *Nat Neurosci* (2009) **12**, 11:1438-1443.

101   Barco, A., Alarcon, J.M. & Kandel, E.R. Expression of constitutively active CREB protein facilitates the late phase of long-term potentiation by enhancing synaptic capture. *Cell* (2002) **108**, 5:689-703.

102   Barco, A. *et al.* Gene expression profiling of facilitated L-LTP in VP16-CREB mice reveals that BDNF is critical for the maintenance of LTP and its synaptic capture. *Neuron* (2005) **48**, 1:123-137.

103   Huang, Y.H. *et al.* CREB modulates the functional output of nucleus accumbens neurons: a critical role of N-methyl-D-aspartate glutamate receptor (NMDAR) receptors. *J Biol Chem* (2008) **283**, 5:2751-2760.

104   Han, M.H. *et al.* Role of cAMP response element-binding protein in the rat locus ceruleus: regulation of neuronal activity and opiate withdrawal behaviors. *J Neurosci* (2006) **26**, 17:4624-4629.

105   Lopez de Armentia, M. *et al.* cAMP response element-binding protein-mediated gene expression increases the intrinsic excitability of CA1 pyramidal neurons. *J Neurosci* (2007) **27**, 50:13909-13918.

106   Viosca, J., Lopez de Armentia, M., Jancic, D. & Barco, A. Enhanced CREB-dependent gene expression increases the excitability of neurons in the basal amygdala and primes the consolidation of contextual and cued fear memory. *Learn Mem* (2009) **16**, 3:193-197.

107   Serita, T., Fukushima, H. & Kida, S. Constitutive activation of CREB in mice enhances temporal association learning and increases hippocampal CA1 neuronal spine density and complexity. *Sci Rep* (2017) **7**, 42528.

108   Bourtchuladze, R. *et al.* Deficient long-term memory in mice with a targeted mutation of the cAMP-responsive element-binding protein. *Cell* (1994) **79**, 1:59-68.

109   Pittenger, C. *et al.* Reversible inhibition of CREB/ATF transcription factors in region CA1 of the dorsal hippocampus disrupts hippocampus-dependent spatial memory. *Neuron* (2002) **34**, 3:447-462.

110   Pittenger, C. *et al.* Impaired bidirectional synaptic plasticity and procedural memory formation in striatum-specific cAMP response element-binding protein-deficient mice. *J Neurosci* (2006) **26**, 10:2808-2813.

111   Gass, P. *et al.* Deficits in memory tasks of mice with CREB mutations depend on gene dosage. *Learn Mem* (1998) **5**, 4-5:274-288.

112   Rudolph, D. *et al.* Impaired fetal T cell development and perinatal lethality in mice lacking the cAMP response element binding protein. *Proc Natl Acad Sci U S A* (1998) **95**, 8:4481-4486.

113   Zhang, J. *et al.* CREB-mediated synaptogenesis and neurogenesis is crucial for the role of 5-HT1a receptors in modulating anxiety behaviors. *Sci Rep* (2016) **6**, 29551.

114   Briand, L.A., Lee, B.G., Lelay, J., Kaestner, K.H. & Blendy, J.A. Serine 133 phosphorylation is not required for hippocampal CREB-mediated transcription and behavior. *Learn Mem* (2015) **22**, 2:109-115.

115   Lonze, B.E., Riccio, A., Cohen, S. & Ginty, D.D. Apoptosis, axonal growth defects, and degeneration of peripheral neurons in mice lacking CREB. *Neuron* (2002) **34**, 3:371-385.

116   Sakamoto, K., Karelina, K. & Obrietan, K. CREB: a multifaceted regulator of neuronal plasticity and protection. *J Neurochem* (2011) **116**, 1:1-9.

117   Wilson, B.E., Mochon, E. & Boxer, L.M. Induction of bcl-2 expression by phosphorylated CREB proteins during B-cell activation and rescue from apoptosis. *Mol Cell Biol* (1996) **16**, 10:5546-5556.

118   Fukuchi, M., Tabuchi, A. & Tsuda, M. Transcriptional regulation of neuronal genes and its effect on neural functions: cumulative mRNA expression of PACAP and BDNF genes controlled by calcium and cAMP signals in neurons. *J Pharmacol Sci* (2005) **98**, 3:212-218.

119   Wang, J.M. *et al.* The antiapoptotic gene mcl-1 is up-regulated by the phosphatidylinositol 3-Kinase/Akt signaling pathway through a transcription factor complex containing CREB. *Mol. Cell. Biol.* (1999) **19**, 9:6195-6206.

120   Lambert, H.W., Weiss, E.R. & Lauder, J.M. Activation of 5-HT receptors that stimulate the adenylyl cyclase pathway positively regulates IGF-I in cultured craniofacial mesenchymal cells. *Developmental Neuroscience* (2001) **23**, 1:70-77.

121   Maymó, J.L. *et al.* Regulation of placental leptin expression by cyclic adenosine 5'-monophosphate involves cross talk between protein kinase A and mitogen-activated protein kinase signaling pathways. *Endocrinology* (2010) **151**, 8:3738-3751.

122   Tabuchi, A., Sakaya, H., Kisukeda, T., Fushiki, H. & Tsuda, M. Involvement of an upstream stimulatory factor as well as cAMP-responsive element-binding protein in the activation of brain-derived neurotrophic factor gene promoter I. *J Biol Chem* (2002) **277**, 39:35920-35931.

123   Kingsbury, T.J., Murray, P.D., Bambrick, L.L. & Krueger, B.K. Ca(2+)-dependent regulation of TrkB expression in neurons. *J Biol Chem* (2003) **278**, 42:40744-40748.

124   Riccio, A., Ahn, S., Davenport, C.M., Blendy, J.A. & Ginty, D.D. Mediation by a CREB family transcription factor of NGF-dependent survival of sympathetic neurons. *Science* (1999) **286**, 5448:2358-2361.

125   Bonni, A. *et al.* Cell survival promoted by the Ras-MAPK signaling pathway by transcription-dependent and -independent mechanisms. *Science* (1999) **286**, 5443:1358-1362.

126   Tan, Y.W., Zhang, S.J., Hoffmann, T. & Bading, H. Increasing levels of wild-type CREB up-regulates several activity-regulated inhibitor of death (AID) genes and promotes neuronal survival. *BMC Neurosci* (2012) **13**, 1:48.

127   St-Pierre, J. *et al.* Suppression of reactive oxygen species and neurodegeneration by the PGC-1 transcriptional coactivators. *Cell* (2006) **127**, 2:397-408.

128   Lee, J. *et al.* Mitochondrial cyclic AMP response element-binding protein (CREB) mediates mitochondrial gene expression and neuronal survival. *J Biol Chem* (2005) **280**, 49:40398-40401.

129   Lemberger, T., Parkitna, J.R., Chai, M., Schütz, G. & Engblom, D. CREB has a context-dependent role in activity-regulated transcription and maintains neuronal cholesterol homeostasis. *The FASEB Journal* (2008) **22**, 8:2872-2879.

130   Impey, S. *et al.* Defining the CREB regulon: a genome-wide analysis of transcription factor regulatory regions. *Cell* (2004) **119**, 7:1041-1054.

131   Zhang, X. *et al.* Genome-wide analysis of cAMP-response element binding protein occupancy, phosphorylation, and target gene activation in human tissues. *Proc Natl Acad Sci U S A* (2005) **102**, 12:4459-4464.

132   Euskirchen, G. *et al.* CREB binds to multiple loci on human chromosome 22. *Mol Cell Biol* (2004) **24**, 9:3804-3814.

133   Benito, E., Valor, L.M., Jimenez-Minchan, M., Huber, W. & Barco, A. cAMP response element-binding protein is a primary hub of activity-driven neuronal gene expression. *J Neurosci* (2011) **31**, 50:18237.

134   Ploski, J.E., Park, K.W., Ping, J., Monsey, M.S. & Schafe, G.E. Identification of plasticity-associated genes regulated by Pavlovian fear conditioning in the lateral amygdala. *J Neurochem* (2010) **112**, 3:636-650.

135   Lakhina, V. *et al.* Genome-wide functional analysis of CREB/long-term memory-dependent transcription reveals distinct basal and memory gene expression programs. *Neuron* (2015) **85**, 2:330-345.

136   De Cesare, D., Fimia, G.M. & Sassone-Corsi, P. Signaling routes to CREM and CREB: plasticity in transcriptional activation. *Trends Biochem Sci* (1999) **24**, 7:281-285.

137   Shaywitz, A.J. & Greenberg, M.E. CREB: a stimulus-induced transcription factor activated by a diverse array of extracellular signals. *Annu Rev Biochem* (1999) **68**, 1:821-861.

138   Hai, T.W., Liu, F., Coukos, W.J. & Green, M.R. Transcription factor ATF cDNA clones: an extensive family of leucine zipper proteins able to selectively form DNA-binding heterodimers. *Genes Dev* (1989) **3**, 12B:2083-2090.

139   Kobayashi, M. & Kawakami, K. ATF-CREB heterodimer is involved in constitutive expression of the housekeeping Na$^+$-ATPase α subunit gene. *Nucleic Acids Res.* (1995) **23**, 15:2848-2855.

140   Hurst, H.C., Totty, N.F. & Jones, N.C. Identification and functional characterisation of the cellular activating transcription factor 43 (ATF-43) protein. *Nucleic Acids Res* (1991) **19**, 17:4601-4609.

141   Foulkes, N.S., Mellstrom, B., Benusiglio, E. & Sassone-Corsi, P. Developmental switch of CREM function during spermatogenesis: from antagonist to activator. *Nature* (1992) **355**, 6355:80-84.

142   Brindle, P., Linke, S. & Montminy, M. Analysis of a PK-A dependent activator in CREB reveals a new role for the CREM family of repressors. *Nature* (1993) **364**, 821-824.

143   Borlikova, G. & Endo, S. Inducible cAMP early repressor (ICER) and brain functions. *Mol Neurobiol* (2009) **40**, 1:73-86.

144   Gundersen, B.B. *et al.* Increased hippocampal neurogenesis and accelerated response to antidepressants in mice with specific deletion of CREB in the hippocampus: role of cAMP response-element modulator tau. *J Neurosci* (2013) **33**, 34:13673-13685.

145   Kobayashi, M., Shimomura, A., Hagiwara, M. & Kawakami, K. Phosphorylation of ATF-1 enhances its DNA binding and transcription of the Na,K-ATPase alpha 1 subunit gene promoter. *Nucleic Acids Res* (1997) **25**, 4:877-882.

146   Loriaux, M.M., Brennan, R.G. & Goodman, R.H. Modulatory function of CREB-CREM alpha heterodimers depends upon CREM alpha phosphorylation. *J Biol Chem* (1994) **269**, 46:28839-28843.

147   Hurst, H.C., Masson, N., Jones, N.C. & Lee, K.A. The cellular transcription factor CREB corresponds to activating transcription factor 47 (ATF-47) and forms complexes with a group of polypeptides related to ATF-43. *Mol Cell Biol* (1990) **10**, 12:6192-6203.

148   Bleckmann, S.C. *et al.* Activating transcription factor 1 and CREB are important for cell survival during early mouse development. *Mol. Cell. Biol.* (2002) **22**, 6:1919-1925.

149   Mantamadiotis, T. *et al.* Disruption of CREB function in brain leads to neurodegeneration. *Nat Genet* (2002) **31**, 1:47-54.

150   Shaulian, E. & Karin, M. AP-1 as a regulator of cell life and death. *Nat. Cell Biol.* (2002) **4**, 5:E131-E136.

151   Green, T.A. *et al.* Induction of activating transcription factors (ATFs) ATF2, ATF3, and ATF4 in the nucleus accumbens and their regulation of emotional behavior. *J. Neurosci.* (2008) **28**, 9:2025-2032.

152   Green, T.A. *et al.* Induction of inducible cAMP early repressor expression in nucleus accumbens by stress or amphetamine increases behavioral responses to emotional stimuli. *J Neurosci* (2006) **26**, 32:8235-8242.

153   Hai, T. & Curran, T. Cross-family dimerization of transcription factors Fos/Jun and ATF/CREB alters DNA binding specificity. *Proc Natl Acad Sci U S A* (1991) **88**, 9:3720-3724.

154   Hu, J.Y., Levine, A., Sung, Y.J. & Schacher, S. cJun and CREB2 in the Postsynaptic Neuron Contribute to Persistent Long-Term Facilitation at a Behaviorally Relevant Synapse. *J Neurosci* (2015) **35**, 1:386.

155   Kwok, R.P. *et al.* Nuclear protein CBP is a coactivator for the transcription factor CREB. *Nature* (1994) **370**, 6486:223-226.

156   Shaywitz, A.J., Dove, S.L., Kornhauser, J.M., Hochschild, A. & Greenberg, M.E. Magnitude of the CREB-Dependent Transcriptional Response Is Determined by the Strength of the Interaction between the Kinase-Inducible Domain of CREB and the KIX Domain of CREB-Binding Protein. *Mol Cell Biol* (2000) **20**, 24:9409-9422.

157   Lu, Q., Hutchins, A.E., Doyle, C.M., Lundblad, J.R. & Kwok, R.P. Acetylation of cAMP-responsive element-binding protein (CREB) by CREB-binding protein enhances CREB-dependent transcription. *J Biol Chem* (2003) **278**, 18:15727-15734.

158   Ogryzko, V.V., Schiltz, R.L., Russanova, V., Howard, B.H. & Nakatani, Y. The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell* (1996) **87**, 5:953-959.

159   Ramos, Y.F. *et al.* Genome-wide assessment of differential roles for p300 and CBP in transcription regulation. *Nucleic Acids Res* (2010) **38**, 16:5396-5408.

160   Kim, T.K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* (2010) **465**, 7295:182-187.

161   Wood, M.A. *et al.* Transgenic mice expressing a truncated form of CREB-binding protein (CBP) exhibit deficits in hippocampal synaptic plasticity and memory storage. *Learn Mem* (2005) **12**, 2:111-119.

162   Oliveira, A.M., Wood, M.A., McDonough, C.B. & Abel, T. Transgenic mice expressing an inhibitory truncated form of p300 exhibit long-term memory deficits. *Learn Mem* (2007) **14**, 9:564-572.

163   Chen, G., Zou, X., Watanabe, H., van Deursen, J.M. & Shen, J. CREB binding protein is required for both short-term and long-term memory formation. *J Neurosci* (2010) **30**, 39:13066.

164   Barrett, R.M. *et al.* Hippocampal focal knockout of CBP affects specific histone modifications, long-term potentiation, and long-term memory. *Neuropsychopharmacology* (2011) **36**, 8:1545-1556.

165   Felinski, E.A. & Quinn, P.G. The CREB constitutive activation domain interacts with TATA-binding protein-associated factor 110 (TAF110) through specific hydrophobic residues in one of the three subdomains required for both activation and TAF110 binding. *J. Biol. Chem.* (1999) **274**, 17:11672-11678.

166   Felinski, E.A., Kim, J., Lu, J. & Quinn, P.G. Recruitment of an RNA polymerase II complex is mediated by the constitutive activation domain in CREB, independently of CREB phosphorylation. *Mol Cell Biol* (2001) **21**, 4:1001-1010.

167   Nakajima, T., Uchida, C., Anderson, S.F., Parvin, J.D. & Montminy, M. Analysis of a cAMP-responsive activator reveals a two-component mechanism for transcriptional induction via signal-dependent factors. *Genes Dev* (1997) **11**, 6:738-747.

168   Nakajima, T. *et al.* RNA helicase A mediates association of CBP with RNA polymerase II. *Cell* (1997) **90**, 6:1107-1112.

169   Conkright, M.D. *et al.* TORCs: transducers of regulated CREB activity. *Mol Cell* (2003) **12**, 2:413-423.

170   Ch'ng, T.H. *et al.* Activity-dependent transport of the transcriptional coactivator CRTC1 from synapse to nucleus. *Cell* (2012) **150**, 1:207-221.

171   Screaton, R.A. *et al.* The CREB coactivator TORC2 functions as a calcium- and cAMP-sensitive coincidence detector. *Cell* (2004) **119**, 1:61-74.

172   Luo, Q. *et al.* Mechanism of CREB recognition and coactivation by the CREB-regulated transcriptional coactivator CRTC2. *Proc Natl Acad Sci U S A* (2012) **109**, 51:20865-20870.

173   Kim, H. The transcription cofactor CRTC1 protects from aberrant hepatic lipid accumulation. *Sci Rep* (2016) **6**, 37280.

174   Wang, B. *et al.* The insulin-regulated CREB coactivator TORC promotes stress resistance in Drosophila. *Cell Metab* (2008) **7**, 5:434-444.

175   Koo, S.H. *et al.* The CREB coactivator TORC2 is a key regulator of fasting glucose metabolism. *Nature* (2005) **437**, 7062:1109-1111.

176   Mair, W. *et al.* Lifespan extension induced by AMPK and calcineurin is mediated by CRTC-1 and CREB. *Nature* (2011) **470**, 7334:404-408.

177   Jagannath, A. *et al.* The CRTC1-SIK1 pathway regulates entrainment of the circadian clock. *Cell* (2013) **154**, 5:1100-1111.

178   Burkewitz, K. *et al.* Neuronal CRTC-1 governs systemic mitochondrial metabolism and lifespan via a catecholamine signal. *Cell* (2015) **160**, 5:842-855.

179   Song, Y. *et al.* CRTC3 links catecholamine signalling to energy balance. *Nature* (2010) **468**, 7326:933-939.

180   Kovacs, K.A. *et al.* TORC1 is a calcium- and cAMP-sensitive coincidence detector involved in hippocampal long-term synaptic plasticity. *Proc Natl Acad Sci U S A* (2007) **104**, 11:4700-4705.

181   Hirano, Y. *et al.* Fasting launches CRTC to facilitate long-term memory formation in Drosophila. *Science* (2013) **339**, 6118:443-446.

182   Uchida, S. *et al.* CRTC1 Nuclear Translocation Following Learning Modulates Memory Strength via Exchange of Chromatin Remodeling Complexes on the Fgf1 Gene. *Cell Rep* (2017) **18**, 2:352-366.

183   Parra-Damas, A. *et al.* CRTC1 Function During Memory Encoding Is Disrupted in Neurodegeneration. *Biol Psychiatry* (2017) **81**, 2:111-123.

184   Nonaka, M. *et al.* Region-specific activation of CRTC1-CREB signaling mediates long-term fear memory. *Neuron* (2014) **84**, 1:92-106.

185   Kasper, L.H. *et al.* CBP/p300 double null cells reveal effect of coactivator level and diversity on CREB transactivation. *EMBO J* (2010) **29**, 21:3660-3672.

186   Hirano, Y. *et al.* Shifting transcriptional machinery is required for long-term memory maintenance and modification in Drosophila mushroom bodies. *Nature Communications* (2016) **7**, 13471.

187   C.A., S. & Valero, J. The role of CREB signaling in Alzheimer's disease and other cognitive disorders. *Rev. Neurosci.* (2011) **22**, 2:153.

188   Jancic, D., Lopez de Armentia, M., Valor, L.M., Olivares, R. & Barco, A. Inhibition of cAMP response element-binding protein reduces neuronal excitability and plasticity, and triggers neurodegeneration. *Cereb Cortex* (2009) **19**, 11:2535-2547.

189   Pugazhenthi, S., Wang, M., Pham, S., Sze, C.I. & Eckman, C.B. Downregulation of CREB expression in Alzheimer's brain and in Abeta-treated rat hippocampal neurons. *Mol Neurodegener* (2011) **6**, 1:60.

190   Yiu, A.P., Rashid, A.J. & Josselyn, S.A. Increasing CREB function in the CA1 region of dorsal hippocampus rescues the spatial memory deficits in a mouse model of Alzheimer's disease. *Neuropsychopharmacology* (2011) **36**, 11:2169-2186.

191   Yu, X.W., Curlik, D.M., Oh, M.M., Yin, J.C. & Disterhoft, J.F. CREB overexpression in dorsal CA1 ameliorates long-term memory deficits in aged rats. *Elife* (2017) **6**, e19358.

192   Viosca, J. *et al.* Chronic enhancement of CREB activity in the hippocampus interferes with the retrieval of spatial information. *Learn Mem* (2009) **16**, 3:198-209.

193   Obrietan, K. & Hoyt, K.R. CRE-mediated transcription is increased in Huntington's disease transgenic mice. *J Neurosci* (2004) **24**, 4:791-796.

194   Jeong, H. *et al.* Sirt1 mediates neuroprotection from mutant huntingtin by activation of the TORC1 and CREB transcriptional pathway. *Nat Med* (2011) **18**, 1:159-165.

195   Choi, Y.S. *et al.* CREB is a key regulator of striatal vulnerability in chemical and genetic models of Huntington's disease. *Neurobiol Dis* (2009) **36**, 2:259-268.

196   Giampa, C. *et al.* Striatal modulation of cAMP-response-element-binding protein (CREB) after excitotoxic lesions: implications with neuronal vulnerability in Huntington's disease. *Eur J Neurosci* (2006) **23**, 1:11-20.

197   Milnerwood, A.J. *et al.* Early increase in extrasynaptic NMDA receptor signaling and expression contributes to phenotype onset in Huntington's disease mice. *Neuron* (2010) **65**, 2:178-190.

198   Petrif, F. *et al.* Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP. *Nature* (1995) **376**, 6538:348-351.

199   Roelfsema, J.H. *et al.* Genetic heterogeneity in Rubinstein-Taybi syndrome: mutations in both the CBP and EP300 genes cause disease. *Am J Hum Genet* (2005) **76**, 4:572-580.

200   Park, E. *et al.* Epigenetic mechanisms of Rubinstein-Taybi syndrome. *Neuromolecular Med* (2014) **16**, 1:16-24.

201   Alarcon, J.M. *et al.* Chromatin acetylation, memory, and LTP are impaired in CBP+/- mice: a model for the cognitive deficit in Rubinstein-Taybi syndrome and its amelioration. *Neuron* (2004) **42**, 6:947-959.

202   Caccamo, A., Maldonado, M.A., Bokov, A.F., Majumder, S. & Oddo, S. CBP gene transfer increases BDNF levels and ameliorates learning and memory deficits in a mouse model of Alzheimer's disease. *Proc Natl Acad Sci U S A* (2010) **107**, 52:22687-22692.

203   Giralt, A. *et al.* Long-term memory deficits in Huntington's disease are associated with reduced CBP histone acetylase activity. *Hum Mol Genet* (2012) **21**, 6:1203-1216.

204   Nucifora, F.C., Jr. *et al.* Interference by huntingtin and atrophin-1 with cbp-mediated transcription leading to cellular toxicity. *Science* (2001) **291**, 5512:2423-2428.

205   España, J. *et al.* β-Amyloid Disrupts Activity-Dependent Gene Transcription Required for Memory through the CREB Coactivator CRTC1. *J Neurosci* (2010) **30**, 28:9402.

206   Parra-Damas, A. *et al.* Crtc1 activates a transcriptional program deregulated at early Alzheimer's disease-related stages. *J Neurosci* (2014) **34**, 17:5776-5787.

207   Hagiwara, M. *et al.* Transcriptional attenuation following cAMP induction requires PP-1-mediated dephosphorylation of CREB. *Cell* (1992) **70**, 1:105-113.

208   Wadzinski, B.E. *et al.* Nuclear protein phosphatase 2A dephosphorylates protein kinase A-phosphorylated CREB and regulates CREB transcriptional stimulation. *Mol Cell Biol* (1993) **13**, 5:2822-2834.

209   Middei, S. *et al.* CREB selectively controls learning-induced structural remodeling of neurons. *Learn Mem* (2012) **19**, 8:330-336.

210   Ao, H., Ko, S.W. & Zhuo, M. CREB activity maintains the survival of cingulate cortical pyramidal neurons in the adult mouse brain. *Mol Pain* (2006) **2**, 15.

211   Kida, S. *et al.* CREB required for the stability of new and reactivated fear memories. *Nat Neurosci* (2002) **5**, 4:348-355.

212   Rammes, G. *et al.* Synaptic plasticity in the basolateral amygdala in transgenic mice expressing dominant-negative cAMP response element-binding protein (CREB) in forebrain. *Eur J Neurosci* (2000) **12**, 7:2534-2546.

213   Blanchet, E. *et al.* Feedback inhibition of CREB signaling promotes beta cell dysfunction in insulin resistance. *Cell Rep* (2015) **10**, 7:1149-1157.

214   Kronke, G. *et al.* Oxidized phospholipids induce expression of human heme oxygenase-1 involving activation of cAMP-responsive element-binding protein. *J Biol Chem* (2003) **278**, 51:51006-51014.

215   Naqvi, S., Martin, K.J. & Arthur, J.S. CREB phosphorylation at Ser133 regulates transcription via distinct mechanisms downstream of cAMP and MAPK signalling. *Biochem J* (2014) **458**, 3:469-479.

216   Kamiya, K. *et al.* Phosphorylation of the cyclic AMP response element binding protein mediates transforming growth factor beta-induced downregulation of cyclin A in vascular smooth muscle cells. *Mol Cell Biol* (2007) **27**, 9:3489-3498.

217   McClung, C.A. & Nestler, E.J. Regulation of gene expression and cocaine reward by CREB and DeltaFosB. *Nat Neurosci* (2003) **6**, 11:1208-1215.

218   Shi, Y. *et al.* Direct regulation of CREB transcriptional activity by ATM in response to genotoxic stress. *Proc Natl Acad Sci U S A* (2004) **101**, 16:5898-5903.

219   Shanware, N.P., Trinh, A.T., Williams, L.M. & Tibbetts, R.S. Coregulated ataxia telangiectasia-mutated and casein kinase sites modulate cAMP-response element-binding protein-coactivator interactions in response to DNA damage. *J Biol Chem* (2007) **282**, 9:6283-6291.

220   Kim, S.H. *et al.* Tunable regulation of CREB DNA binding activity couples genotoxic stress response and metabolism. *Nucleic Acids Res* (2016) **44**, 20:9667-9680.

221   Kornhauser, J.M. *et al.* CREB transcriptional activity in neurons is regulated by multiple, calcium-specific phosphorylation events. *Neuron* (2002) **34**, 2:221-233.

222   Gau, D. *et al.* Phosphorylation of CREB Ser142 regulates light-induced phase shifts of the circadian clock. *Neuron* (2002) **34**, 2:245-253.

223   Niederberger, E. *et al.* The impact of CREB and its phosphorylation at Ser142 on inflammatory nociception. *Biochem Biophys Res Commun* (2007) **362**, 1:75-80.

224   Trinh, A.T., Kim, S.H., Chang, H.Y., Mastrocola, A.S. & Tibbetts, R.S. Cyclin-dependent kinase 1-dependent phosphorylation of cAMP response element-binding protein decreases chromatin occupancy. *J Biol Chem* (2013) **288**, 33:23765-23775.

225   Comerford, K.M. *et al.* Small ubiquitin-related modifier-1 modification mediates resolution of CREB-dependent responses to hypoxia. *Proceedings of the National Academy of Sciences of the United States of America* (2003) **100**, 3:986-991.

226   Rexach, J.E. *et al.* Dynamic O-GlcNAc modification regulates CREB-mediated gene expression and memory formation. *Nat Chem Biol* (2012) **8**, 3:253-261.

227   Lamarre-Vincent, N. & Hsieh-Wilson, L.C. Dynamic glycosylation of the transcription factor CREB: a potential role in gene regulation. *J Am Chem Soc* (2003) **125**, 22:6612-6613.

Global analysis of the interplay between CREB *O*-GlcNAc glycosylation and phosphorylation

Portions of this chapter will be published as:

5.1 Abstract

The post-translational modification (PTM) "code" refers to the control of transcription factor (TF) activity through post-translational modifications (PTMs).[1] CREB (cAMP response element binding protein) is a TF that is controlled by a variety of PTMs to regulate neuronal metabolism, activity, differentiation, development, and survival.[2] CREB phosphorylation at serine 133 has been shown to enhance CREB-mediated transcription while CREB glycosylation at serine 40 has been shown to decrease CREB-mediated transcription.[3] The exact gene networks modulated by and potential interplay between CREB glycosylation and phosphorylation have not been explored. Through differential expression analysis with glycosylation-deficient (S40A) and phosphorylation-deficient (S133A) CREB mutants, we show that CREB O-GlcNAcylation is important for neuronal activity and excitability while phosphorylation at serine 133 regulates the expression of genes involved in neuronal differentiation. Furthermore, many of the S40A and S133A differentially-expressed genes were directly bound by (1) CREB and its co-activators, CREB-binding protein and p300, (2) activating histone modifications, (3) OGT and O-GlcNAc and (4) Tet1, an critical regulator of neuronal activity and differentiation. Finally, we found that CREB O-GlcNAcylation regulates activity- and excitotoxicity-related gene networks while CREB phosphorylation regulates neuronal differentiation and amino and fatty acid metabolism-related gene networks. This study demonstrates that CREB O-GlcNAcylation at serine 40 and phosphorylation mediate different gene networks. Together, O-GlcNAc and phosphorylation impart a TF code,

which CREB must integrate and decode to modulate neuronal activity, differentiation, and metabolism.

4.2 General approach and validation

Despite an understanding of the importance of CREB phosphorylation and glycosylation in mediating particular CREB-regulated genes, the relative importance of each of CREB's PTMs and their potential interplay in negotiating neuronal metabolism, survival, development, and excitability are not well understood. What are the global gene networks regulated by phosphorylation and glycosylation? Are the gene networks distinct or do the PTMs globally affect CREB-mediated transcription across all gene networks? To identify the global transcriptional changes regulated by specific CREB PTMs, we sought to find a system that would allow us to express different CREB modification-defective mutants with minimal interference from endogenous WT CREB.

While the total Creb1 knockout is postnatal lethal, we obtained the Creb1$^{\alpha\delta}$ knockout mice, which lack the Creb1 α and δ isoforms, but express 5% of the β isoform of CREB.[4] We cultured E15-16 Creb1$^{\alpha\delta-/-}$ cortical neurons and used replication-defective herpes simplex virus (HSV) expressing various CREB mutants and GFP (Figure 5.1). Expression of GFP, WT CREB (CREB), a glycosylation-deficient mutant (S40A-CREB), a phosphorylation-deficient mutant (S133A-CREB), and a glycosylation- and phosphorylation-deficient mutant (S40A-S133A-CREB) began at 2 hours with increasing levels up to 12 hours after transduction as observed by immunohistochemical (IHC) staining (Figure 5.2). We chose two time points: 4 hours, when CREB is beginning to be expressed, and 8 hours, when CREB is fully expressed (Figure 5.3). Our previous study showed that neuronal depolarization induced CREB glycosylation, so we KCl

depolarized neurons for 2 hours followed by RNA extraction and RNA-Seq analysis as previously described.[5]
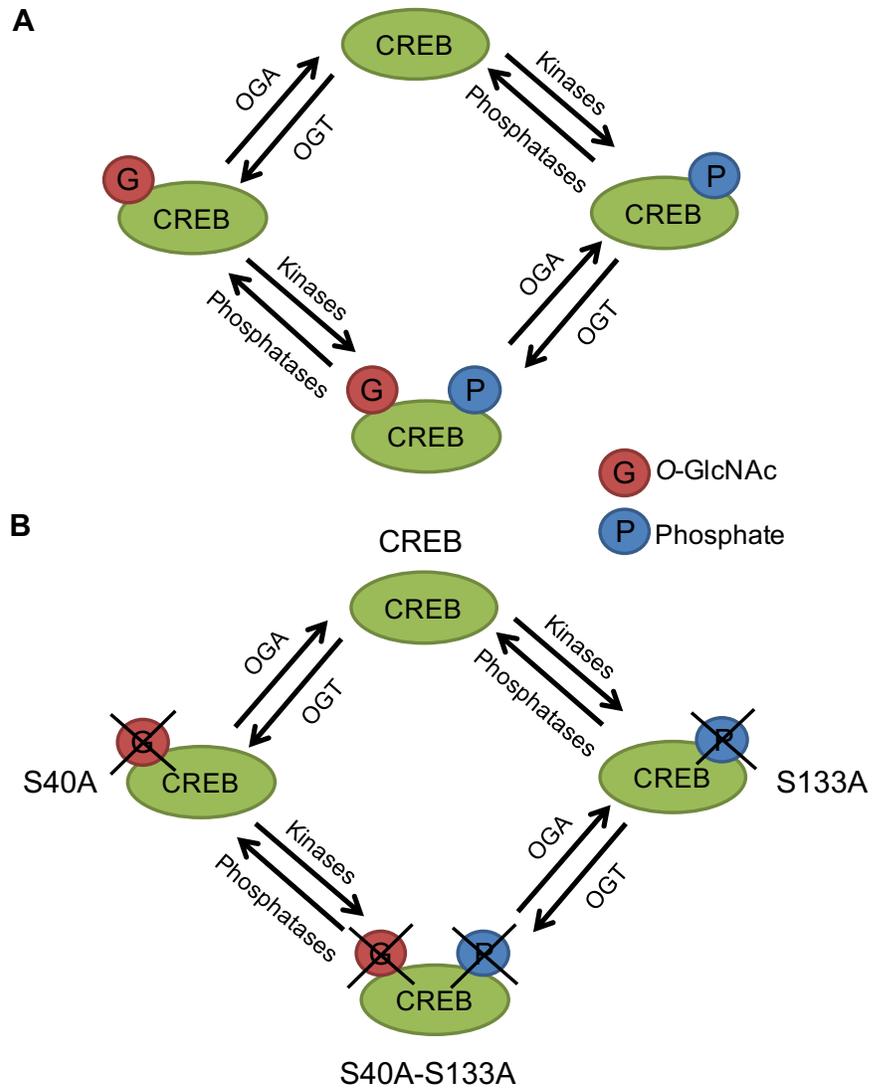
Figure 5.1 Overview of CREB mutants. (A) CREB can be glycosylated at S40 and/or phosphorylated S133. (B) Generation of glycosylation- or phosphorylation- deficient mutants allow us to tease apart contribution of these post-translational modifications to CREB transcription. In particular, we create a non-glycosylatable mutant, S40A-CREB, a non-phosphorylatable mutant, S133-CREB, and a S40A-S133A-CREB mutant that can neither be glycosylated nor phosphorylated.
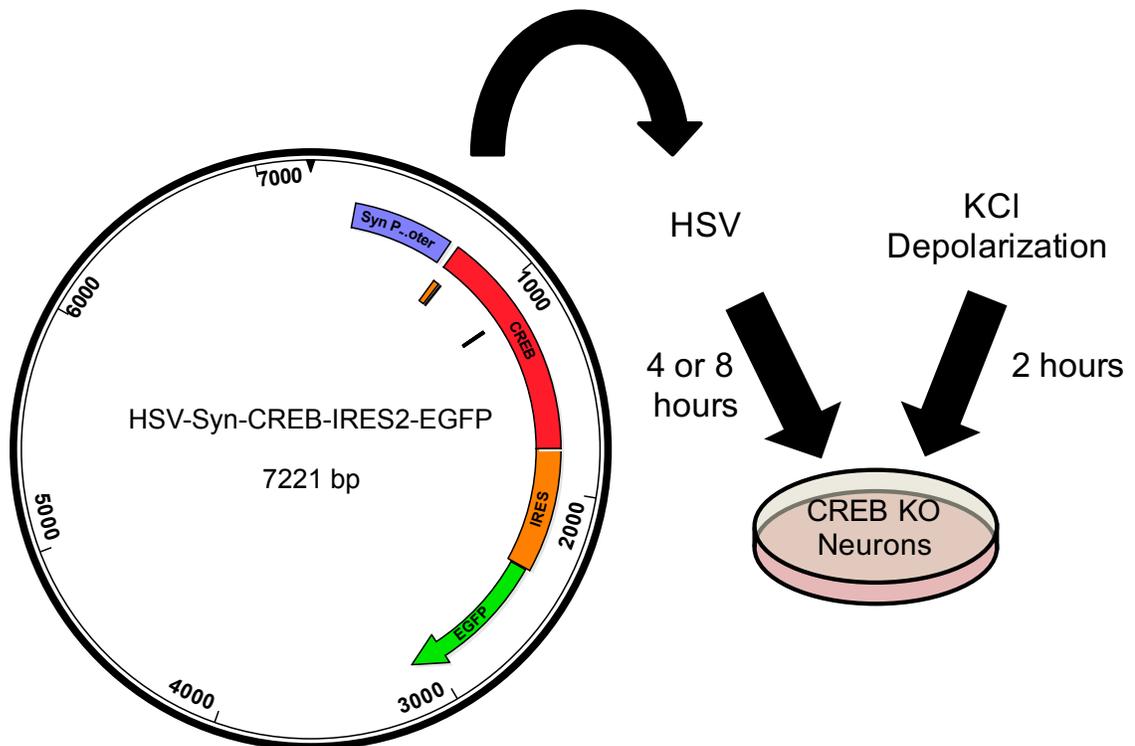
Fig. 5.2. Schematic of experimental overview. Shown here is the overall schematic of the expression of the various CREB mutants using HSV on the CREB KO cortical neuron background. Based on ICC, we chose 2 times points to explore the early and late stage changes. At 4 hours, CREB is beginning to be expressed (incipient changes). At 8 hours, CREB has had several hours to express and sufficient time has elapsed for significant translational changes to occur. We also depolarized neurons with KCl for a total of 2 hours and then we extracted RNA and submitted to RNA-Seq.
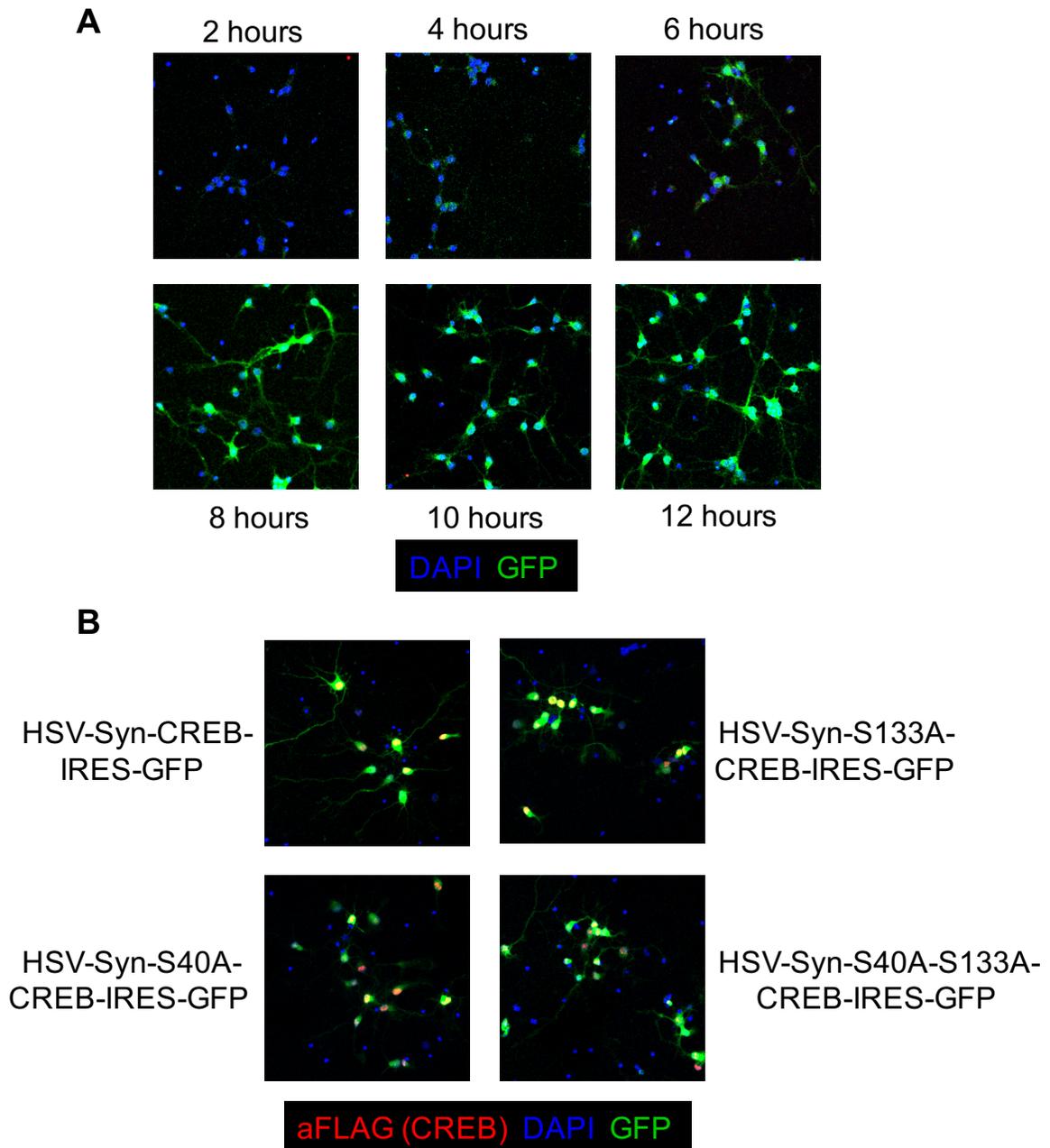
Figure 5.3 HSV transduces neurons rapidly. (A) E16.5 mouse cortical neurons were transfected at 5DIV with HSV-Syn-IRES-GFP for 2, 4, 6, 8, 10, and 12 hours. Expression of GFP is observed starting at 2 hours and increases at 4 hours post-transduction. By 8 hours, the neuronal processes have GFP throughout. (B) E16.5 cortical neurons (5DIV) were transfected with the following HSV for 8 hours: HSV-Syn-CREB-IRES-GFP, HSV-Syn-S40A-CREB-IRES-GFP, HSV-Syn-S133A-CREB-IRES-GFP, HSV-Syn-S40A-S133A-CREB-IRES-GFP. We observed the robust nuclear staining of FLAG (CREB) at 8 hours with all CREB-expressing HSV.

We also verified that the CREB expression levels were consistent using qPCR and RNA-Seq. Indeed, the S40A/WT levels of *Creb1* expression at 4 hours were found to be the same (1.00-fold, q-value = 1.00). There was a 19-fold increase in CREB expression in both the 4-hour WT and S40A conditions when compared to the GFP condition on CREB KO background (q-value = 0.004). Similarly, we verified using qPCR the CREB levels across the 8 hour WT, S40A, S133A, and S40A-S133A within the replicates (Figure 5.4).
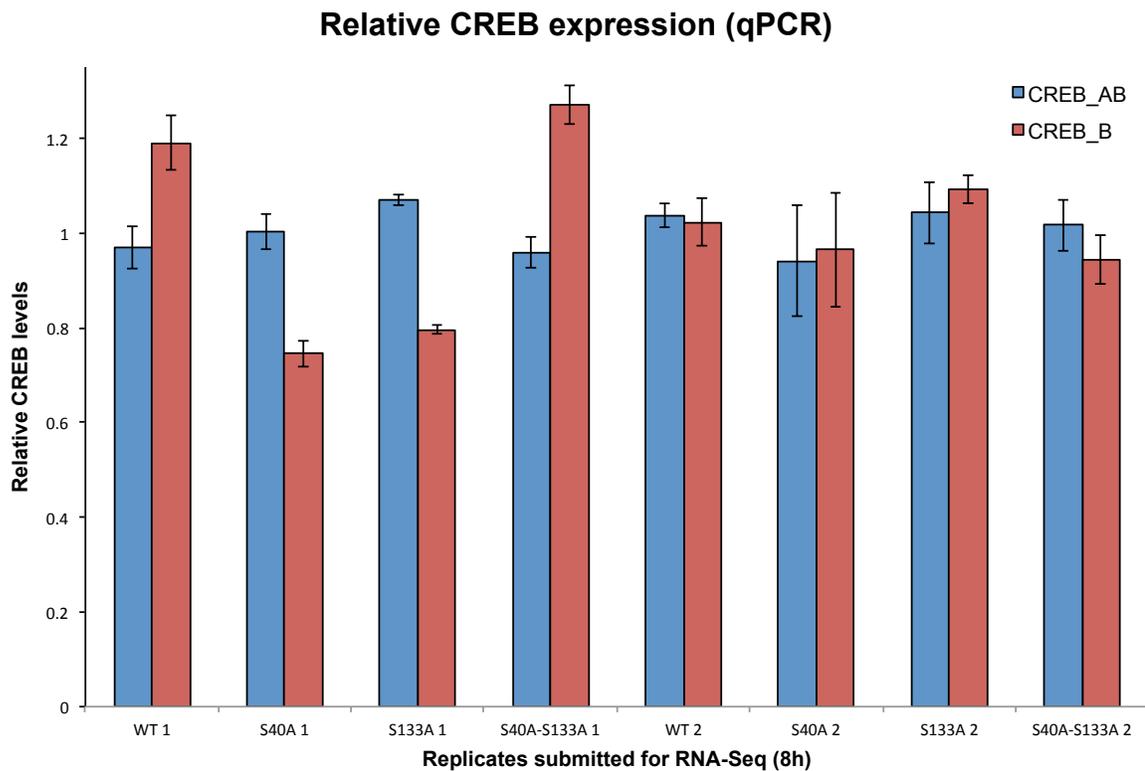
Figure 5.4 Relative CREB expression (qPCR). We monitored the CREB expression levels for the first and second replicates in 8 hours. Shown here are the Creb1 α and β isoform levels (CREB_AB) and the Creb1 β isoform levels (CREB_B). Each replicate group is normalized by the within group average CREB_AB or CREB_B expression levels. The error bars represent the relative standard error of the mean.

5.3 Neuronal polarization and axonogenesis genes are upregulated in the S40A CREB condition at 4 hours

RNA-Seq analysis was first performed with GFP, CREB, and S40A-CREB expression for 4 hours in order to identify the incipient gene expression changes (2

replicates each) (Table 5.1). Differentially-expressed (DE) genes were identified using the Cuffdiff pipeline on the Galaxy interface, and significance is reported as q-values (FDR).[6,7] At 4 hours, we observed upregulation of genes involved in neuronal growth and polarization in the glycosylation-deficient S40A-CREB mutant when compared to the WT CREB condition (FDR <0.05). Specifically, kinesin family member 1C (*Kif1c*, 17-fold), rotatin (*Rttn*, 36-fold), and brain-specific serine/threonine-protein kinase 2 (*Brsk2*, 4.3-fold) are involved in neuronal polarization and axonogenesis (Figure 5.5A).[8] The other two genes were myelin regulatory factor (*Myrf*, 3.5-fold) and CCR4-NOT transcription complex, subunit 3 (*Cnot3*, 4.8-fold), which are critical (1) for differentiation and pluripotency and (2) for synaptic rewiring and mRNA regulation respectively.[9,10]

Table 5.1 S40A/WT differentially-expressed genes at 4 hours

Gene names	Gene Description	S40A/WT log2(FC)	q-value
Actr5	ARP5 Actin Related Protein 5 Homolog	6.6	0.004
Rttn	Rotatin	5.2	0.004
Kif1c	Kinesin Family Member 1C	4.1	0.004
Cnot3	CCR4-NOT Transcription Complex Subunit 3	2.2	0.004
Brsk2	BR Serine/Threonine Kinase 2	2.1	0.004
Myrf	Myelin Regulatory Factor	1.8	0.05
Pdia4	Protein Disulfide Isomerase Family A Member 4	-1.2	0.012
Irak1	Interleukin 1 Receptor Associated Kinase 1	-3.5	0.004
2010111I01Rik	Aminopeptidase O	-6.6	0.004

Table 5.1 shows the differentially-expressed genes at 4 hours including the gene names, descriptions, $\log_2$(FC) where "FC" refers to fold-change, and q-values for the S40A/WT comparison at 4 hours. The upregulated genes are highlighted in pink while the downregulated genes are highlighted in green. q-values (FDR) < 0.05.
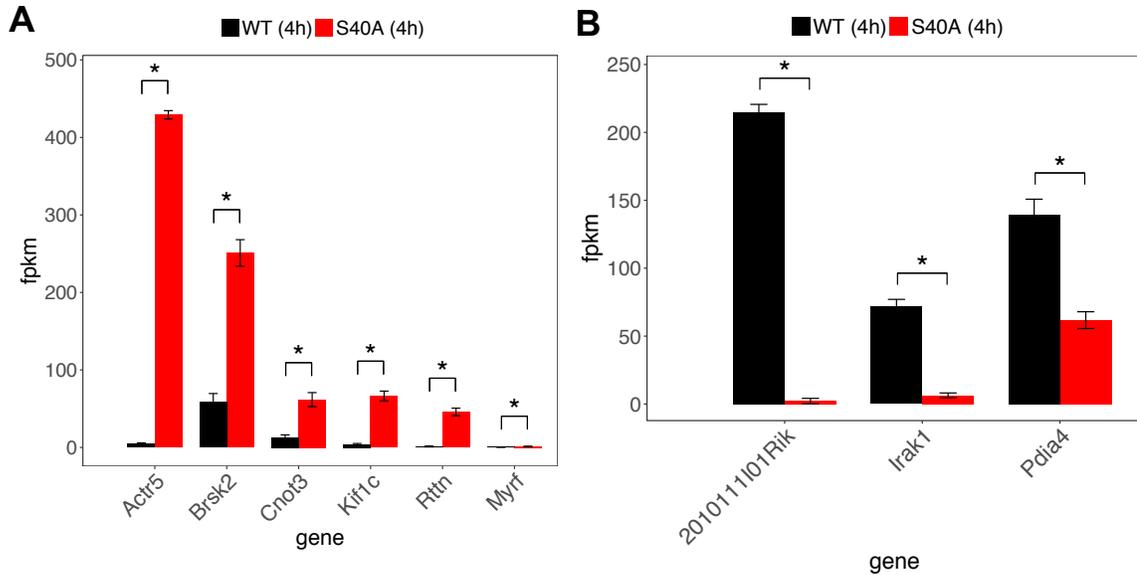
Figure 5.5 Differentially-expressed genes in S40A/WT at 4 hours. (A) The upregulated genes in the S40A condition over the WT CREB condition are involved in neuronal polarization, cytoskeletal rearrangement, and axonogenesis. (B) The downregulated genes in the S40A condition over the WT CREB condition are stress- and immune response-related genes. *FDR < 0.1.

Given the overrepresentation of genes important for neuronal growth and polarization amongst the upregulated genes, we expressed S40A and WT CREB on DIV1 E16.5 CREB KO cortical neurons. After 2 days of HSV expression, we observed enhanced neurite outgrowth in the S40A-expressing neurons consistent with the neurite outgrowth observed in our previous paper (Fig 5.6).[3] Our current and previous results show that ablation of the glycosylation site at serine 40 results in enhanced neuronal growth.

GFP

WT

S40A

Tau1 Map2 GFP

Figure 5.6 S40A-CREB produces enhanced neurite outgrowth. We transduced DIV1 E16.5 CREB KO cortical neurons with HSV-Synapsin-IRES-GFP, HSV-Synapsin-CREB-IRES-GFP, or HSV-Synapsin-S40A-CREB-IRES-GFP (green is GFP). After 2 days of expression, the neurons were fixed and then stained for Tau1 (red) and Map2 (blue). Tau1 is a marker for axons while Map2 stains dendrites and neuronal cell bodies. We observed enhanced neurite outgrowth in the S40A-CREB condition only. Scale bars represent 400

The S40A/WT downregulated genes at 4 hours were broadly involved in stress

and immune response, including protein disulfide isomerase family A member 4 (*Pdia4*,

212

2.2-fold), interleukin 1 receptor associated kinase 1 (*Irak1*, 11-fold), and aminopeptidase O (*2010111I01Rik*, 98-fold) (Figure 5.5B).[11,12] Irak1 is a serine/threonine kinase, which increases the stability of interleukin-1 and is important for immune response.[12] Pdia4 is upregulated in response to endoplasmic reticulum stress and can protect cells from oxidative stress.[11] Aminopeptidase O (ApO) has not been fully characterized, but ApO is believed to be important for angiogenesis.[13] In summary, four hours of expression yielded changes in few genes, but overall, the S40A mutant increased expression of neuronal growth genes and decreased expression of stress and immune response genes.

5.4 Neuronal excitability genes are upregulated in the S40A CREB condition at 8 hours

We next explored the longer-term CREB-induced changes in gene expression. After 8 hours of expression, we identified 87 upregulated genes when we compared S40A-CREB to WT CREB (FDR < 0.1) (Tables 5.2, 5.3). The upregulated genes were enriched for genes encoding synaptic proteins ($P = 2.4 \times 10^{-4}$), calcium signaling proteins ($P = 3.6 \times 10^{-3}$), and voltage-gated channels ($P = 8.4 \times 10^{-4}$) (Table 5.2). Upregulated calcium signaling genes included Ca^{2+}/calmodulin kinase II (*Camk2a*, 1.6-fold) and calcium voltage-gated channel subunit alpha1 I (*Cacna1i*, 1.6-fold) (Figure 5.7A). Calcium signaling is very important for modulating neuronal excitability and LTP.[14] The expression of cAMP signaling genes was also increased in the S40A/WT comparison, including protein kinase A (PKA, *Prkacb*, 1.4-fold), *Camk2a*, and brain-derived neurotrophic factor (*Bdnf*, 1.6-fold), a canonical CREB target important for neuronal growth, LTP, and long-term depression (LTD) (Figure 5.7B).[15]

Table 5.2 DAVID GO annotation of S40A/WT upregulated genes at 8 hours

213

GO term	#	Gene names	FE
Dendrite	14	*Kcnh1, Ache, Kcnc4, Cplx1, Ddn, Shh, Crhr1, Kcnj4, Bdnf, Lynx1, Chrm1, Negr1, Camk2a, Synpo*	7.0
Neuronal cell body	13	*Kcnh1, Ache, Kcnc4, Cplx1, Ptprn, Shh, Crhr1, Kcnj4, Bdnf, Crh, Nrsn1, Camk2a, Negr1*	5.9
Synapse	12	*Kcnh1, Kcnj4, Ache, Cplx1, Syndig1, Chrm1, Grin2d, Psd3, Lgi3, Ptprn, Camk2a, Synpo*	5.8
Voltage-gated channel	7	*Kcnh1, Kcnj4, Kcnc4, Kcns2, Kcnj9, Cacna1i, Kcnh3*	14
Glycoprotein	31	*Kcnh1, Kcnc4, Ache, Slc6a1, Spock3, Epha10, Shh, Col26a1, Bdnf, Rspo1, Creg2, Grin2d, Lgi3, Cntnap1, Etl4, Loxl2, Negr1, Synpo, Phyhip, Cckbr, Cdhr1, Ai593442, Ptprn, Crhr1, Lynx1, Slc6a7, Chrm1, Cemip, Wif1, Car4, Kcnh3*	2.1
Circadian entrainment	6	*Kcnj9, Grin2d, Cacna1i, Prkacb, Gng4, Camk2a*	14
Potassium transport	6	*Kcnh1, Kcnj4, Kcnc4, Kcns2, Kcnj9, Kcnh3*	15
Calcium signaling pathway	7	*Cckbr, Chrm1, Grin2d, Cacna1i, Prkacb, Itpka, Camk2a*	9.1
Cholinergic synapse	6	*Kcnj4, Ache, Chrm1, Prkacb, Gng4, Camk2a*	12
Axon	9	*Kcnh1, Bdnf, Ache, Kcnc4, Slc6a1, Cntnap1, Camk2a, Shh, Synpo*	5.9
Cell junction	11	*Kcnj4, Ache, Syndig1, Chrm1, Grin2d, Psd3, Lgi3, 9430020k01rik, Ptprn, Camk2a, Synpo*	4.5
Postsynaptic membrane	7	*Kcnj4, Ache, Syndig1, Chrm1, Grin2d, Psd3, Synpo*	7.7
Membrane	44	*Kcnh1, Ache, Kcnc4, Syndig1, Slc6a1, Ildr2, Epha10, Tmem151b, Tmem151a, Shh, Tmcc2, Kcns2, Pacsin3, Grin2d, Rasl10b, Cntnap1, Prkacb, Loxl2, Gng4, Negr1, Camk2a, Synpo, Ngef, Cckbr, Cdhr1, Ai593442, Psd3, Ptprn, Ddn, Crhr1, Kcnj4, Lynx1, Slc6a7, Kcnj9, Mtfp1, Chrm1, Vsnl1, Cemip, Nrsn1, Car4, Cend1, Parp1, Fam163b, Kcnh3*	1.5
Lipoprotein	11	*Ache, Lynx1, Cckbr, Ncald, Vsnl1, Rasl10b, Prkacb, Car4, Gng4, Negr1, Shh*	3.7
Secreted	16	*Ache, Spock3, Fam24a, Shh, Bdnf, Col26a1, Lynx1, Rspo1, Creg2, Cemip, Crh, Cartpt, Wif1, Lgi3, Loxl2, Scg2*	2.5
Phosphoprotein	43	*Kcnh1, Gda, Kcnc4, Syndig1, Tcap, Slc6a1, Ankrd34c, Cnot3, Ildr2, 9430020k01rik, Itpka, Tmcc2, Rasal1, Fbxw7, Pacsin3, Inpp5j, Grin2d, Osbpl1a, Apba3, Etl4, Cntnap1, Prkacb, Camk2a, Negr1, Scg2, Synpo, Ngef, Map1a, Esrrg, Psd3, Ptprn, Ddn, Rcan2, Crhr1, Dact2, Slc6a7, Nab2, Chrm1, Zbtb4, Cartpt, Parp1, Cend1, Fam163b*	1.5
Ion transport	9	*Kcnh1, Kcnj4, Kcnc4, Kcns2, Kcnj9, Grin2d, Cacna1i, Atp6v1g2, Kcnh3*	3.8
Postsynaptic density	6	*Syndig1, Chrm1, Map1a, Psd3, Camk2a, Synpo*	6.1

Table 5.2 contains the DAVID functional gene ontology annotations of the S40A-CREB/ CREB upregulated differentially-expressed genes. Benjamini-corrected p-values are all less than 0.05. FE = fold enrichment.

Table 5.3 List of differentially-expressed S40A/WT genes at 8 hours

Gene names	Gene description	S40A/WT log2(FC)	q-value
Loxl2	lysyl oxidase-like 2(Loxl2)	7.2	0.01
Susd5	sushi domain containing 5(Susd5)	2.4	0.03
Rcan3	regulator of calcineurin 3(Rcan3)	2.2	0.01
Cartpt	CART prepropeptide(Cartpt)	2.1	0.01

Car4	carbonic anhydrase 4(Car4)	2.0	0.01
Wif1	Wnt inhibitory factor 1(Wif1)	1.8	0.01
Rspo1	R-spondin 1(Rspo1)	1.6	0.03
Ankrd34c	ankyrin repeat domain 34C(Ankrd34c)	1.6	0.03
Cckbr	cholecystokinin B receptor(Cckbr)	1.4	0.01
Itpka	inositol 1,4,5-trisphosphate 3-kinase A(Itpka)	1.2	0.01
Apba3	amyloid beta (A4) precursor protein-binding, family A, member 3(Apba3)	1.2	0.01
Parp1	poly (ADP-ribose) polymerase family, member 1(Parp1)	1.2	0.01
Tcap	titin-cap(Tcap)	1.1	0.02
Crh	corticotropin releasing hormone(Crh)	1.1	0.05
Creg2	cellular repressor of E1A-stimulated genes 2(Creg2)	1.0	0.01
Pacsin3	protein kinase C and casein kinase substrate in neurons 3(Pacsin3)	1.0	0.01
Mybpc1	myosin binding protein C, slow-type(Mybpc1)	1.0	0.02
Lynx1	Ly6/neurotoxin 1(Lynx1)	0.85	0.01
Cdhr1	cadherin-related family member 1(Cdhr1)	0.85	0.03
Kcns2	K+ voltage-gated channel, subfamily S, 2(Kcns2)	0.85	0.02
Lgi3	leucine-rich repeat LGI family, member 3(Lgi3)	0.84	0.02
Nab2	Ngfi-A binding protein 2(Nab2)	0.81	0.04
Nrsn1	neurensin 1(Nrsn1)	0.77	0.01
Cntnap1	contactin associated protein-like 1(Cntnap1)	0.74	0.03
Rasal1	RAS protein activator like 1 (GAP1 like)(Rasal1)	0.74	0.05
Mtfp1	mitochondrial fission process 1(Mtfp1)	0.74	0.04
Gng4	guanine nucleotide binding protein (G protein), gamma 4(Gng4)	0.73	0.01
Cemip	cell migration inducing protein, hyaluronan binding(Cemip)	0.73	0.01
Slc6a7	solute carrier family 6 (neurotransmitter transporter, L-proline), member 7(Slc6a7)	0.73	0.01
Inpp5j	inositol polyphosphate 5-phosphatase J(Inpp5j)	0.73	0.02
Cnot3	CCR4-NOT transcription complex, subunit 3(Cnot3)	0.72	0.02
Dact2	dishevelled-binding antagonist of beta-catenin 2(Dact2)	0.71	0.04
Bdnf	brain derived neurotrophic factor(Bdnf)	0.68	0.01
Kcnh3	potassium voltage-gated channel, subfamily H (eag-related), member 3(Kcnh3)	0.68	0.06
Kcnj9	potassium inwardly-rectifying channel, subfamily J, member 9(Kcnj9)	0.67	0.01
Camk2a	calcium/calmodulin-dependent protein kinase II alpha(Camk2a)	0.66	0.01
Esrrg	estrogen-related receptor gamma(Esrrg)	0.66	0.09
Rcan2	regulator of calcineurin 2(Rcan2)	0.65	0.01
Ddn	dendrin(Ddn)	0.65	0.01
Kcnc4	potassium voltage gated channel, Shaw-related subfamily, member 4(Kcnc4)	0.65	0.03
Cacna1i	calcium channel, voltage-dependent, alpha 1I subunit(Cacna1i)	0.65	0.01

Spock3	sparc/osteonectin, cwcv and kazal-like domains proteoglycan 3(Spock3)	0.65	0.01
Etl4	enhancer trap locus 4(Etl4)	0.65	0.02
9430020K01 Rik	RIKEN cDNA 9430020K01 gene(9430020K01Rik)	0.64	0.01
Fam163b	family with sequence similarity 163, member B(Fam163b)	0.64	0.03
Vsnl1	visinin-like 1(Vsnl1)	0.63	0.01
Shh	sonic hedgehog(Shh)	0.63	0.03
Col26a1	collagen, type XXVI, alpha 1(Col26a1)	0.62	0.06
Synpo	synaptopodin(Synpo)	0.62	0.03
Rassf3	Ras association (RalGDS/AF-6) domain family member 3(Rassf3)	0.61	0.05
Ncald	neurocalcin delta(Ncald)	0.60	0.01
Phyhip	phytanoyl-CoA hydroxylase interacting protein(Phyhip)	0.59	0.01
Ngef	neuronal guanine nucleotide exchange factor(Ngef)	0.59	0.01
Cend1	cell cycle exit and neuronal differentiation 1(Cend1)	0.59	0.01
Syndig1	synapse differentiation inducing 1(Syndig1)	0.58	0.03
Tmem151a	transmembrane protein 151A(Tmem151a)	0.58	0.01
Ildr2	immunoglobulin-like domain containing receptor 2(Ildr2)	0.58	0.06
Epha10	Eph receptor A10(Epha10)	0.57	0.06
Nrip3	nuclear receptor interacting protein 3(Nrip3)	0.57	0.02
Tmcc2	transmembrane and coiled-coil domains 2(Tmcc2)	0.57	0.02
Kcnh1	potassium voltage-gated channel, subfamily H (eag-related), member 1(Kcnh1)	0.56	0.01
Grin2d	glutamate receptor, ionotropic, NMDA2D (epsilon 4)(Grin2d)	0.56	0.02
A830018L16 Rik	RIKEN cDNA A830018L16 gene(A830018L16Rik)	0.55	0.10
Me3	malic enzyme 3, NADP(+)-dependent, mitochondrial(Me3)	0.55	0.10
Fam81a	family with sequence similarity 81, member A(Fam81a)	0.55	0.03
Slc6a1	solute carrier family 6 (neurotransmitter transporter, GABA), member 1(Slc6a1)	0.54	0.04
Crhr1	corticotropin releasing hormone receptor 1(Crhr1)	0.54	0.09
Atp6v1g2	ATPase, H+ transporting, lysosomal V1 subunit G2(Atp6v1g2)	0.54	0.02
Kcnj4	potassium inwardly-rectifying channel, subfamily J, member 4(Kcnj4)	0.54	0.05
Osbpl1a	oxysterol binding protein-like 1A(Osbpl1a)	0.54	0.01
Map1a	microtubule-associated protein 1 A(Map1a)	0.54	0.03
AI593442	expressed sequence AI593442(AI593442)	0.53	0.03
Chrm1	cholinergic receptor, muscarinic 1, CNS(Chrm1)	0.52	0.08
Scg2	secretogranin II(Scg2)	0.52	0.06
1700020I14Rik	RIKEN cDNA 1700020I14 gene(1700020I14Rik)	0.51	0.07
Zbtb4	zinc finger and BTB domain containing 4(Zbtb4)	0.51	0.09
D3Bwg0562e	phospholipid phosphatase related 4(Plppr4)	0.51	0.05

Ache	acetylcholinesterase(Ache)	0.50	0.09
Cplx1	complexin 1(Cplx1)	0.49	0.09
Ptprn	protein tyrosine phosphatase, receptor type, N(Ptprn)	0.49	0.06
Rasl10b	RAS-like, family 10, member B(Rasl10b)	0.49	0.06
Prkacb	protein kinase, cAMP dependent, catalytic, beta(Prkacb)	0.48	0.09
Fbxw7	F-box and WD-40 domain protein 7(Fbxw7)	0.48	0.08
Gda	guanine deaminase(Gda)	0.48	0.10
Psd3	pleckstrin and Sec7 domain containing 3(Psd3)	0.48	0.08
Tmem151b	transmembrane protein 151B(Tmem151b)	0.47	0.10
Negr1	neuronal growth regulator 1(Negr1)	0.46	0.08
Grik3	glutamate receptor, ionotropic, kainate 3(Grik3)	-0.48	0.09
Islr2	immunoglobulin superfamily containing leucine-rich repeat 2(Islr2)	-0.51	0.08
Draxin	dorsal inhibitory axon guidance protein(Draxin)	-0.51	0.07
Shb	src homology 2 domain-containing transforming protein B(Shb)	-0.51	0.10
Dkk3	dickkopf WNT signaling pathway inhibitor 3(Dkk3)	-0.52	0.06
D8Ertd82e	DNA segment, Chr 8, ERATO Doi 82, expressed(D8Ertd82e)	-0.53	0.09
Bcl2l11	BCL2-like 11 (apoptosis facilitator)(Bcl2l11)	-0.53	0.06
Ezh2	enhancer of zeste 2 polycomb repressive complex 2 subunit(Ezh2)	-0.54	0.09
Nefm	neurofilament, medium polypeptide(Nefm)	-0.54	0.04
Sox11	SRY (sex determining region Y)-box 11(Sox11)	-0.55	0.04
Tle4	transducin-like enhancer of split 4(Tle4)	-0.57	0.02
Fst	follistatin(Fst)	-0.57	0.10
Epha2	Eph receptor A2(Epha2)	-0.58	0.04
Gpc2	glypican 2 (cerebroglycan)(Gpc2)	-0.62	0.01
Ebf3	early B cell factor 3(Ebf3)	-0.64	0.08
Igf2	insulin-like growth factor 2(Igf2)	-0.66	0.08
Nckap5	NCK-associated protein 5(Nckap5)	-0.70	0.01
Trp53i11	transformation related protein 53 inducible protein 11(Trp53i11)	-0.70	0.02
Carhsp1	calcium regulated heat stable protein 1(Carhsp1)	-0.72	0.01
Nxph3	neurexophilin 3(Nxph3)	-0.74	0.03
Arhgap28	Rho GTPase activating protein 28(Arhgap28)	-0.75	0.04
Id1	inhibitor of DNA binding 1(Id1)	-0.75	0.03
Tshz2	teashirt zinc finger family member 2(Tshz2)	-0.76	0.01
Fn1	fibronectin 1(Fn1)	-0.77	0.01
Foxp2	forkhead box P2(Foxp2)	-0.78	0.08
Plin2	perilipin 2(Plin2)	-0.79	0.02
Kif26a	kinesin family member 26A(Kif26a)	-0.81	0.01
Adam8	a disintegrin and metallopeptidase domain 8(Adam8)	-0.81	0.05
Zfhx3	zinc finger homeobox 3(Zfhx3)	-0.83	0.03

Id3	inhibitor of DNA binding 3(Id3)	-0.84	0.01
Hmox1	heme oxygenase 1(Hmox1)	-0.87	0.01
Bcas1	breast carcinoma amplified sequence 1(Bcas1)	-0.89	0.09
Tcf7l2	transcription factor 7 like 2, T cell specific, HMG box(Tcf7l2)	-0.92	0.02
Txnip	thioredoxin interacting protein(Txnip)	-0.95	0.01
Myrf	myelin regulatory factor(Myrf)	-1.0	0.08
Nts	neurotensin(Nts)	-1.1	0.01
Plxnb3	plexin B3(Plxnb3)	-1.1	0.09
Plp1	proteolipid protein (myelin) 1(Plp1)	-1.1	0.05
Igfbpl1	insulin-like growth factor binding protein-like 1(Igfbpl1)	-1.2	0.01
Cldn11	claudin 11(Cldn11)	-1.3	0.02
Nckap1l	NCK associated protein 1 like(Nckap1l)	-1.5	0.06
Laptm5	lysosomal-associated protein transmembrane 5(Laptm5)	-1.6	0.01
Mag	myelin-associated glycoprotein(Mag)	-1.6	0.04
H19	H19, imprinted maternally expressed transcript(H19)	-1.6	0.01
Lgals3	lectin, galactose binding, soluble 3(Lgals3)	-1.7	0.01
Cyba	cytochrome b-245, alpha polypeptide(Cyba)	-1.7	0.02
Rsad2	radical S-adenosyl methionine domain containing 2(Rsad2)	-1.7	0.08
Pdia4	protein disulfide isomerase associated 4(Pdia4)	-1.8	0.01
Pou2f2	POU domain, class 2, transcription factor 2(Pou2f2)	-1.8	0.01
Enpp6	ectonucleotide pyrophosphatase/phosphodiesterase 6(Enpp6)	-1.9	0.01
Slc11a1	solute carrier family 11 (proton-coupled divalent metal ion transporters), member 1(Slc11a1)	-2.0	0.05
C3ar1	complement component 3a receptor 1(C3ar1)	-2.1	0.01
Ctss	cathepsin S(Ctss)	-2.1	0.01
E130102H24 Rik,Mir101a	microRNA 101a(Mir101a)	-2.1	0.04
Cybb	cytochrome b-245, beta polypeptide(Cybb)	-2.5	0.02
Itgb2	integrin beta 2(Itgb2)	-2.5	0.01
Cd36	CD36 antigen(Cd36)	-2.5	0.01
Itgam	integrin alpha M(Itgam)	-2.5	0.02
Gpnmb	glycoprotein (transmembrane) nmb(Gpnmb)	-2.6	0.01
Ttr	transthyretin(Ttr)	-2.9	0.01
Spp1	secreted phosphoprotein 1(Spp1)	-2.9	0.01
C1qb	complement component 1, q subcomponent, beta polypeptide(C1qb)	-3.0	0.03
Lyz2	lysozyme 2(Lyz2)	-3.0	0.01
2010111I01Ri k	RIKEN cDNA 2010111I01 gene(2010111I01Rik)	-3.2	0.01
C1qa	complement component 1, q subcomponent, alpha polypeptide(C1qa)	-3.3	0.03
Tyrobp	TYRO protein tyrosine kinase binding protein(Tyrobp)	-3.4	0.01
C1qc	complement component 1, q subcomponent, C chain(C1qc)	-3.7	0.02

Kif1c	kinesin family member 1C(Kif1c)	-3.8	0.01
Mmp12	matrix metallopeptidase 12(Mmp12)	-4.6	0.01
Tlcd1	TLC domain containing 1(Tlcd1)	-5.7	0.01

Table 5.3 shows the gene names, descriptions, $\log_2$(FC) where "FC" refers to fold-change, and q-values for the differentially-expressed genes in the S40A/WT comparison at 8 hours. The upregulated genes are highlighted in pink while the downregulated genes are highlighted in green. q-values (FDR) < 0.1.
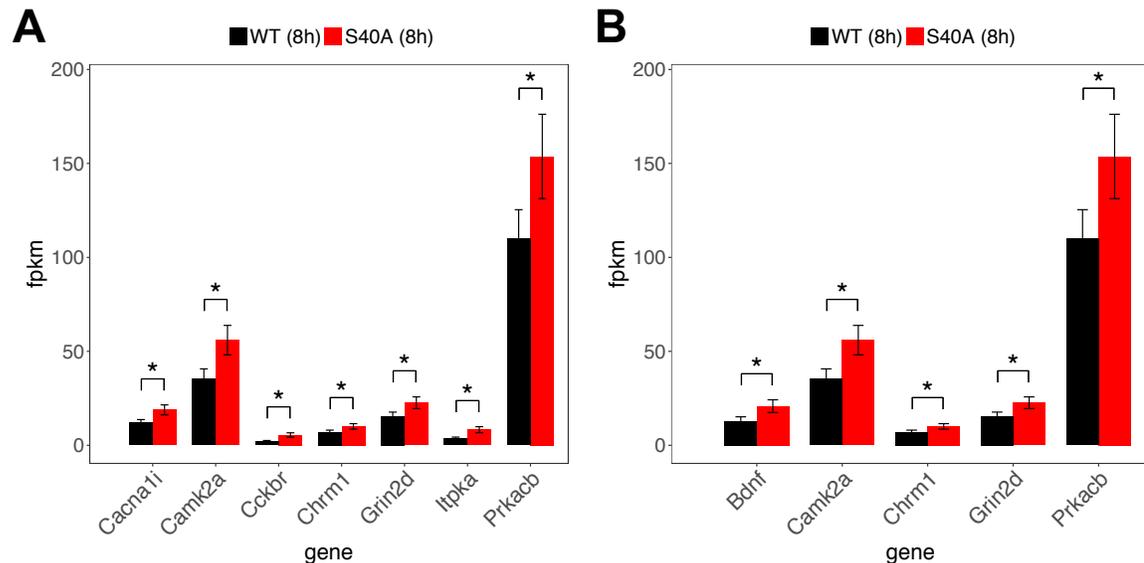
Figure 5.7 Expression levels of upregulated genes in S40A/WT involved in calcium and cAMP signaling pathways at 8 hours. (A) Calcium signaling pathway genes and (B) cAMP signaling pathway genes are upregulated in the S40A condition over the WT CREB condition. *FDR < 0.1.

LTP-related genes were also upregulated in the S40A condition, including glutamate ionotropic receptor NMDA type subunit 2D (*Grin2D*, 1.5-fold) in addition to *Bdnf* and *Prkacb*. Voltage-gated channels such as several potassium (*Kcnc4, Kcnh1, Kcnh3, Kcnj4, Kcnj9, Kcns2,* 1.5-1.8-fold) and calcium (*Cacna1i*, 1.6-fold) channels were enriched in the S40A condition (Figure 5.7A). Increased levels of voltage-gated channels lower the barrier for neuronal depolarization, thereby facilitating LTP and synaptic plasticity.[14] Finally, the S40A mutant increased the expression of genes that encode for proteins found in the dendrite ($P = 1.0 \times 10^{-5}$), axon ($P = 3.9 \times 10^{-3}$), neuronal cell body ($P = 9.7 \times 10^{-5}$), and synapse ($P = 2.4 \times 10^{-4}$), suggesting that S40A leads to enhanced neuronal growth and activity (Figures 5.8B-E, 5.9, Table 5.2). Overall, these

results suggest that CREB glycosylation at serine 40 plays a critical role in calibrating
homeostatic neuronal excitability and LTP.
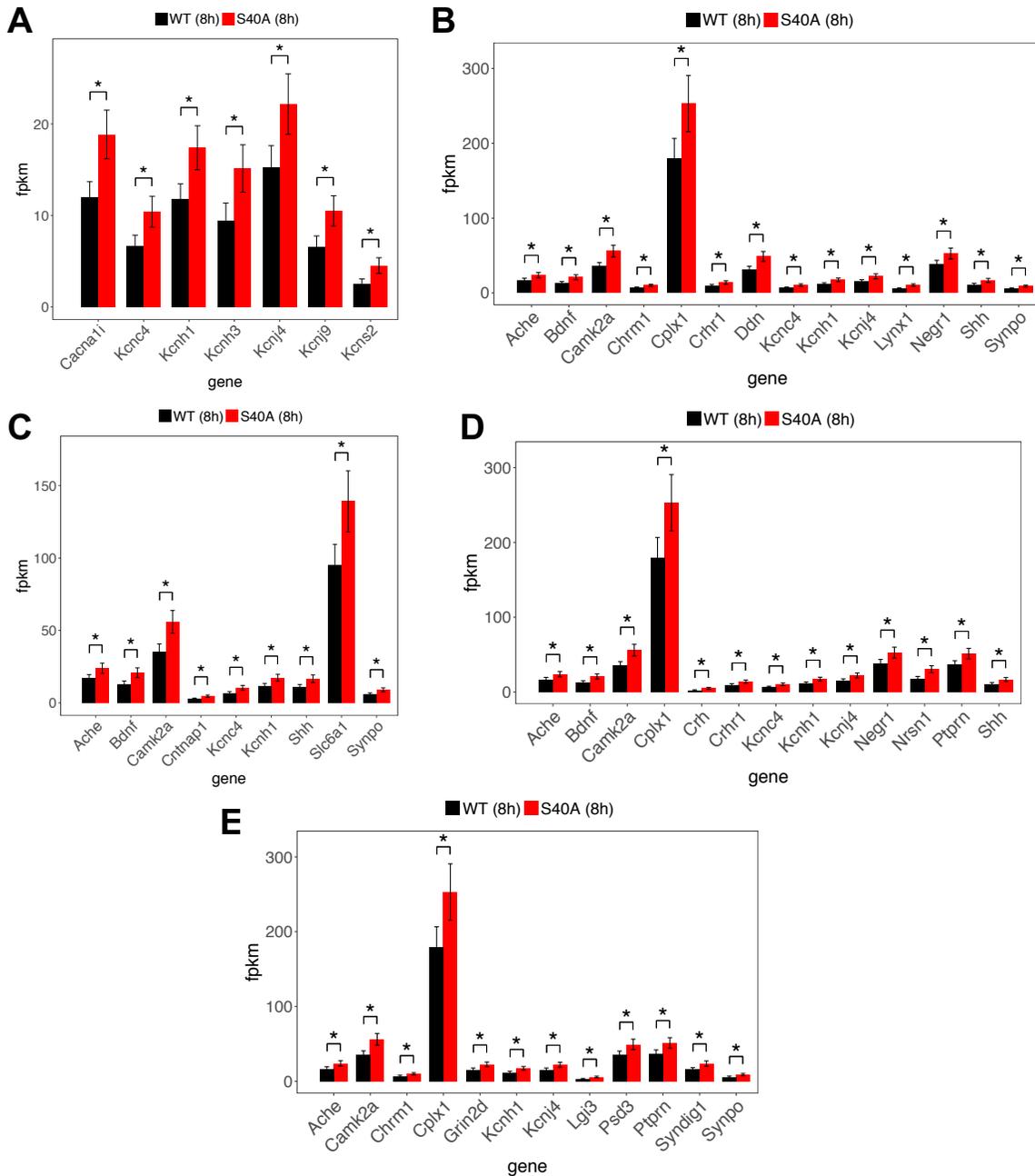
Figure 5.8 Expression levels for neuronal activity upregulated genes in S40A/WT at 8 hours. (A)
Voltage-gated channels, (B) dendritic genes, (C) axonal genes, (D) neuronal cell body, and (E) synaptic
genes are upregulated in S40A/WT after 8 hours of HSV treatment. *FDR < 0.1.
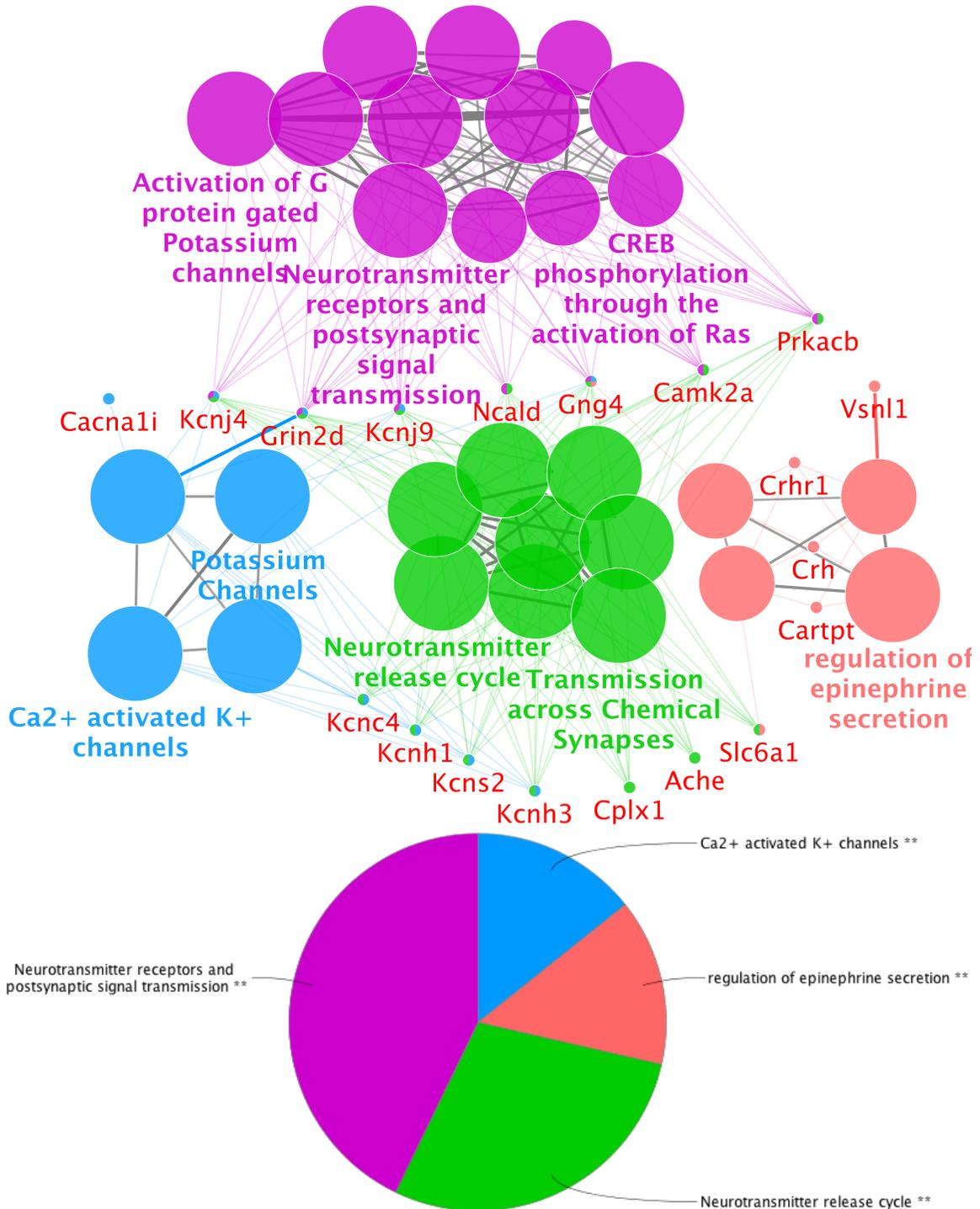
Figure 5.9 Cytoscape gene ontology annotations for the S40A/WT upregulated genes at 8 hours. Shown here are the top gene ontology categories for (A) the upregulated genes and (B) the downregulated genes in the S40A/WT comparison. The genes that belong to and are shared by these gene ontology categories are also shown. The gene annotations were created performed using the Reactome Pathways database for the downregulated genes and GO Molecular Functions, GO Biological Process, and Reactome Pathways databases for the upregulated genes with the ClueGO extension in Cytoscape. The significance for all listed categories are Bonferroni step down corrected $P < 0.0029$.

221

Our paper corroborates our previous study showing that S40A-CREB primes neurons for memory formation.[3] In particular, our previous study showed that important memory-related genes such as *Bdnf* showed increased expression with S40A-CREB expression when compared to CREB expression. This study also found that *Bdnf* expression was enhanced in the S40A-CREB condition in addition to neuronal genes from the dendrites to the axons to the neuronal cell body. Indeed, our current results support the findings in our previous paper where we demonstrated that injecting S40A-CREB-expressing HSV into the lateral amygdala lead to enhanced memory consequences at 2h, which leveled off at 24 hours.[3] Increased *Bdnf* expression was upregulated in other studies expressing other more active CREB mutants, VP16-CREB, Y134F-CREB, and DIEDML-CREB.[16,17]

5.5 Innate immune response and phagosome genes are downregulated in the glycosylation-deficient mutant at 8 hours

Next, we found that the 70 downregulated genes (Table 5.3) in the S40A/WT comparison were involved in innate-immune response ($P = 4.8\text{x}10^{-2}$) and the phagosome ($P = 2.3\text{x}10^{-2}$) at 8 hours (Table 5.4). S40A displayed a reduction in innate immune response-related genes, including complement component 1 subunits (*C1qa*, *C1qb*, *C1qc*, 7.8-13-fold); lectin, galactose-binding, soluble 3 (galectin-3, *Lgals3*, 3.2-fold); and radical *S*-adenosyl methionine domain containing 2 (*Rsad2*, 3.3-fold) (Figure 5.10A).[18] Immune activation is critical for synaptic pruning especially during early stages of neuronal development.[18] Therefore, impaired immune response at this stage could lead to inappropriate synaptic connections and pruning. Finally, phagosomal genes were downregulated in the S40A-CREB condition, including cytochrome b-245 light chain (*Cyba*, 3.2-fold) and cytochrome b-245 heavy chain (*Cybb*, 5.2-fold), which are the major

components of the phagocytic oxidase responsible for generating superoxide (Figure 5.10B).[19] Altogether, the upregulated genes in the S40A over WT CREB condition were involved in neuronal activation and enhancement of neuronal excitability while the downregulated genes were related to immune response and phagocytosis (Figure 5.11).

Table 5.4 DAVID functional annotation of S40A/WT downregulated genes

GO term	#	Gene names	FE
Glycoprotein	31	*Nxph3, Enpp6, C3ar1, Igfbpl1, Grik3, Fst, Ezh2, Itgb2, Pdia4, Itgam, Slc11a1, Ttr, Gpc2, Adam8, Gpnmb, Nefm, Spp1, Fn1, Islr2, Mag, Draxin, Plxnb3, Ctss, Mmp12, Epha2, C1qa, Dkk3, C1qb, Cybb, Cd36, Myrf*	2.7
Signal	32	*Nxph3, Enpp6, Igfbpl1, Grik3, Fst, Itgb2, Pdia4, C1qc, Itgam, Ttr, Gpc2, Adam8, Gpnmb, Spp1, Tyrobp, Fn1, Islr2, Mag, Lyz2, Lgals3, Draxin, Plxnb3, Igf2, Ctss, Mmp12, Epha2, Bcl2l11, C1qa, Dkk3, C1qb, Tlcd1, Nts*	2.4
Staphylococcus aureus infection	6	*C1qa, C3ar1, C1qb, Itgb2, C1qc, Itgam*	26
Secreted	17	*Nxph3, Lyz2, Lgals3, Igfbpl1, Draxin, Fst, Igf2, C1qc, Mmp12, C1qa, Dkk3, C1qb, Ttr, Gpc2, Nts, Spp1, Fn1*	3.4
Cell surface	12	*Slc11a1, Cd36, Lgals3, Plxnb3, Itgb2, Ctss, Pdia4, Adam8, Itgam, Epha2, Tyrobp, Islr2*	5.4
Extracellular region	18	*Nxph3, Enpp6, Lyz2, Lgals3, Igfbpl1, Draxin, Fst, Igf2, C1qc, Mmp12, C1qa, Dkk3, C1qb, Ttr, Gpc2, Nts, Spp1, Fn1*	2.9
Pertussis	5	*C1qa, C1qb, Itgb2, C1qc, Itgam*	15
Phagosome	6	*Cyba, Cybb, Cd36, Itgb2, Ctss, Itgam*	7.4
Repressor	8	*Tshz2, Id1, Ezh2, Tle4, Id3, Zfhx3, Tcf7l2, Foxp2*	5.0
Leukocyte transendothelial migration	5	*Cyba, Cybb, Itgb2, Cldn11, Itgam*	8.9
Respiratory burst	3	*Slc11a1, Cyba, Cybb*	105
Integrin-mediated signaling pathway	5	*Plp1, Itgb2, Adam8, Itgam, Tyrobp*	15
Cell adhesion	9	*Mag, Cd36, Itgb2, Cldn11, Gpnmb, Itgam, Epha2, Fn1, Spp1*	5.2
Neutrophil chemotaxis	5	*Lgals3, Nckap1l, Itgb2, Itgam, Spp1*	19
Negative regulation of transcription from RNA polymerase II promoter	11	*Txnip, Cd36, Id1, Sox11, Fst, Ezh2, Tle4, Id3, Zfhx3, Tcf7l2, Foxp2*	4.2
Innate immune response	8	*C1qa, C1qb, Cyba, Cybb, Lgals3, Rsad2, C1qc, Tyrobp*	5.6
Complement pathway	3	*C1qa, C1qb, C1qc*	36

Table 5.4 contains the DAVID functional gene ontology annotations of the S40A/WT downregulated differentially-expressed genes at 8 hours. Benjamini-corrected p-values are all less than 0.05. FE=fold enrichment.
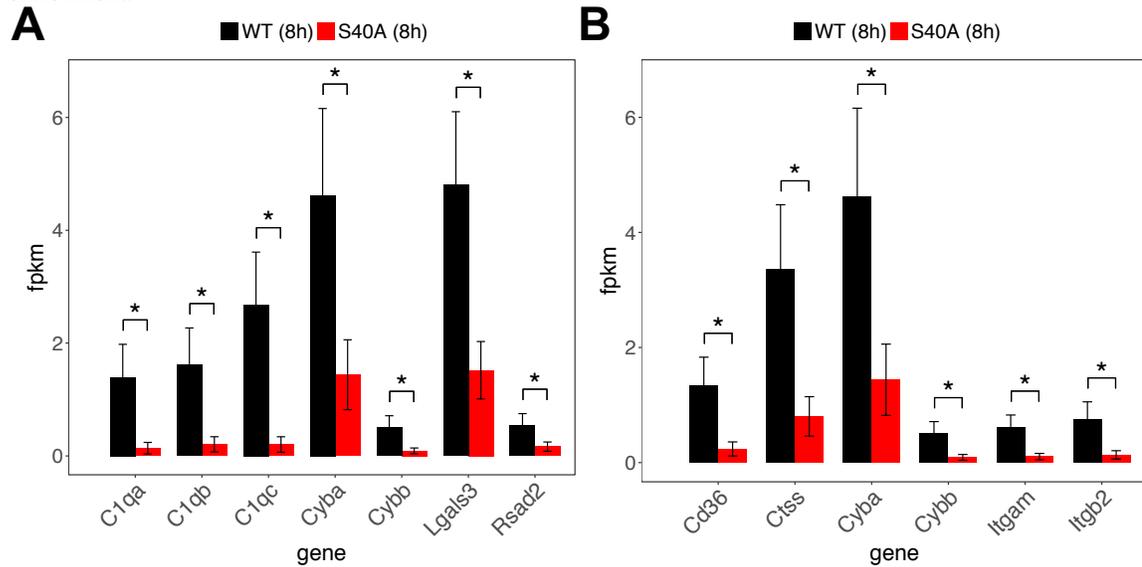
Figure 5.10 Downregulated genes in S40A/WT are involved in innate immune response and phagosome at 8 hours. (A) Innate immune response and (B) phagosome-related genes are downregulated in the S40A condition over the WT CREB condition. *FDR < 0.1.
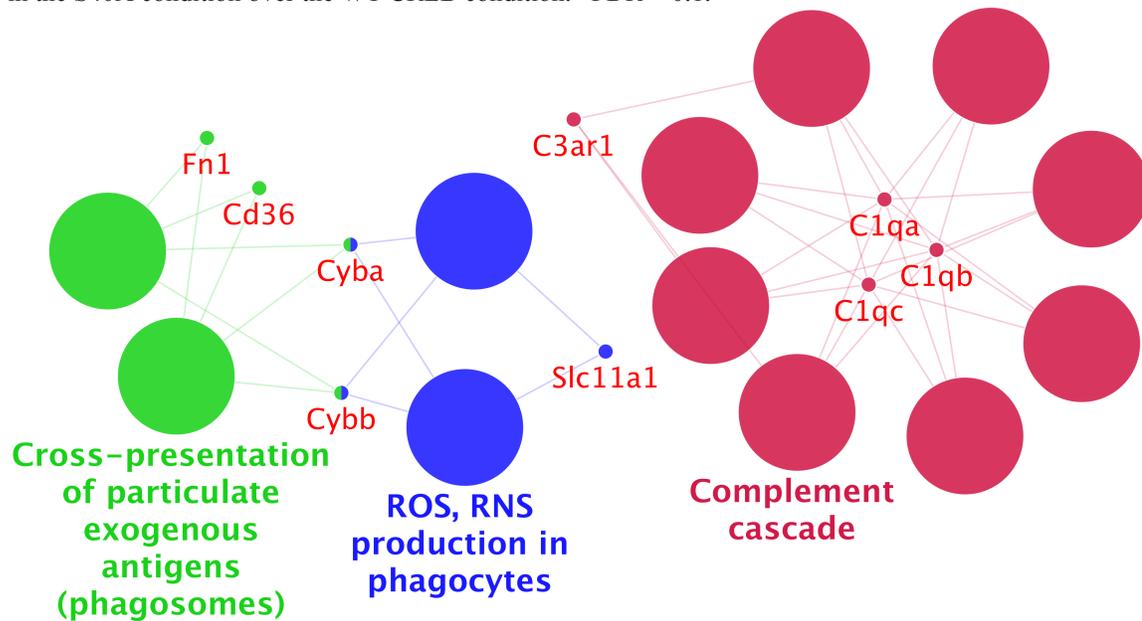
Figure 5.11 Gene ontology annotations for the downregulated genes in the S40A/WT comparison at 8 hours. Shown here are the top gene ontology categories for the downregulated genes in the S40A/WT comparison. The top categories include complement cascade innate immune response as well as phagosomal related categories. The genes that belong to and are shared by these gene ontology categories are also shown. The gene annotations were created performed using the Reactome Pathways database with the ClueGO extension in Cytoscape. The significance for all listed categories are Bonferroni step down corrected $P < 0.00046$.

5.6 Loss of CREB phosphorylation at serine 133 affects nervous system development at 8 hours

After exploring the glycosylation-related changes in transcription, we delved into the phosphorylation-dependent CREB gene expression changes. At 8 hours, abrogation of serine 133 phosphorylation resulted in increased expression of 17 genes (Table 5.5), which were enriched for genes involved in nervous system development ($P = 9.0x10^{-8}$), including wingless-type MMTV integration site family, member 7B (*Wnt7b*, 1.8-fold), netrin G2 (*Nntg2*, 2.2-fold), plexin D1 (*Plxnd1*, 1.7-fold), huntingtin-associated protein 1 (*Hap1*, 2.8-fold), dachshund homolog 2 (*Dach2*, 1.7-fold), and oligodendrocyte TFs 1 and 2 (*Olig1*, *Olig2*, 1.7- and 2.0-fold respectively) (Figures 5.12A, 5.13A, Table 5.6). In addition, the S133A/WT comparison displayed decreased expression of 40 genes (Table 5.5), which were enriched for genes found in the extracellular matrix ($P = 2.0x10^{-4}$) and important for cellular differentiation ($P = 4.0x10^{-2}$) (Figure 5.12B-C, Table 5.7). The DE extracellular matrix genes included aggrecan (*Acan*, 3.2-fold), collagen, type XIX, 1 (*Col19a1*, 1.9-fold), nephronectin (*Npnt* 1.8-fold), and *Lgals3* (3.9-fold) (Figure 5.12B, Figure 5.13B). Among the 18 downregulated genes were genes associated with differentiation into glia and other cell types, such as semaphorin 3C (*Sema3c*, 1.7-fold), eyes absent homolog 4 (*Eya4*, 2.1-fold), alanyl (membrane) aminopeptidase N (*Anpep*, 4.1-fold), and nephronectin (*Npnt*, 1.8-fold) (Figure 5.12C, Figure 5.13B). Because our primary cells were obtained from E16.5 cortices, our cell population was comprised of neural progenitor cells (NPCs) and neurons. We observe that loss of phosphorylation at serine 133 resulted in loss of pluripotency in this mixed cellular population and progression toward neuronal development.

225
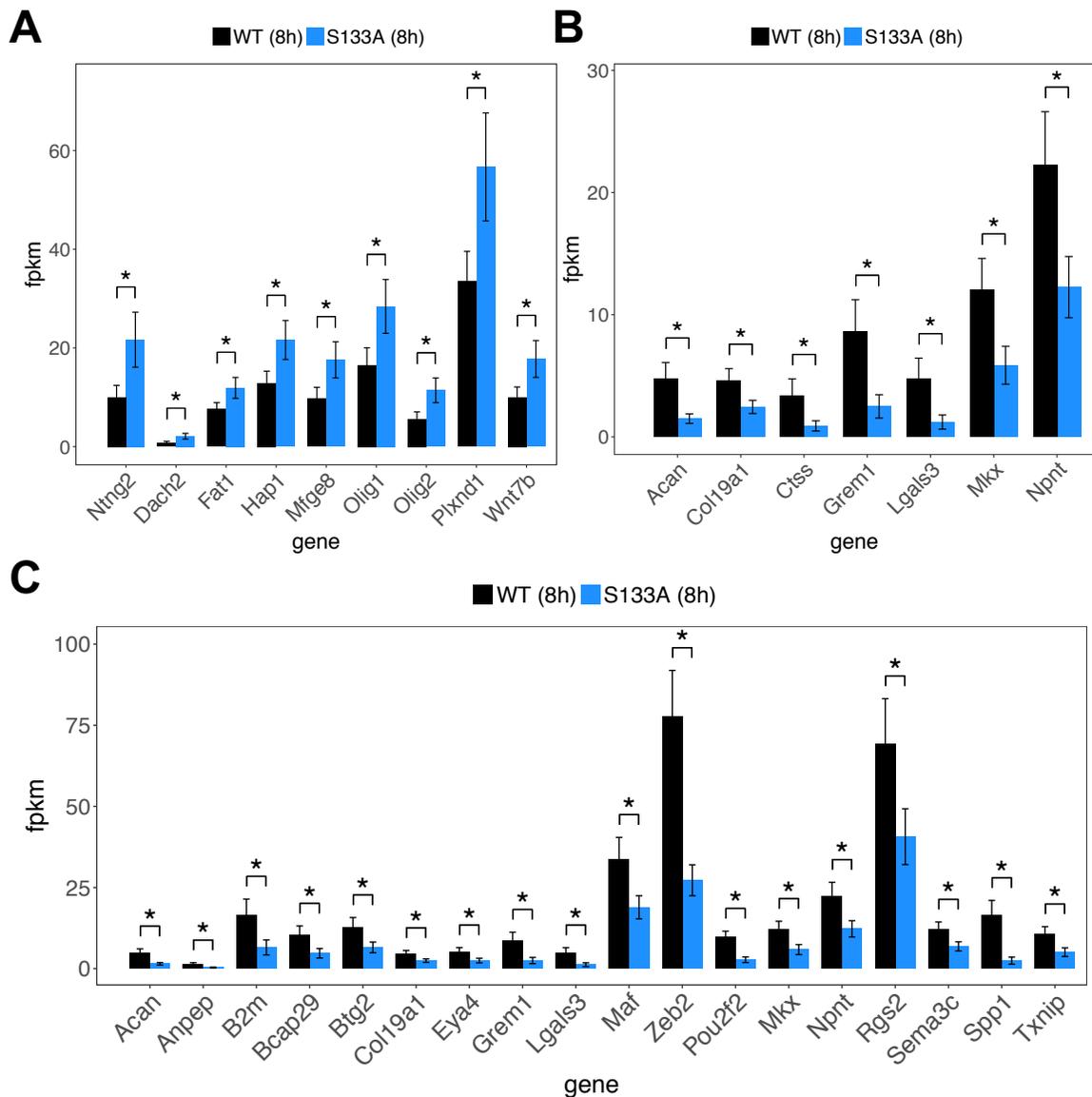
Figure 5.12 Differentially-expressed genes in S133A/WT at 8 hours are involved in neuronal differentiation and development. (A) Genes involved in the nervous system development are upregulated in the S133A-CREB condition over the CREB condition. (B) Extracellular matrix organization genes and (C) cell differentiation genes are downregulated in S133A-CREB condition compared to the CREB condition. *FDR < 0.1.

Table 5.5 List of differentially-expressed S133A/WT genes at 8 hours

Gene names	Gene description	S133A/WT log2(FC)	q-value
Col7a1	collagen, type VII, alpha 1(Col7a1)	6.4	0.004
Pdgfra	platelet derived growth factor receptor, alpha polypeptide(Pdgfra)	5.4	0.004
Rttn	rotatin(Rttn)	4.3	0.004
Rcan3	regulator of calcineurin 3(Rcan3)	3.2	0.004
Dach2	dachshund 2 (Drosophila)(Dach2)	1.5	0.07

Fn1	fibronectin 1(Fn1)	1.1	0.004
Ntng2	netrin G2(Ntng2)	1.1	0.01
Olig2	oligodendrocyte transcription factor 2(Olig2)	1.0	0.05
Aldh1l1	aldehyde dehydrogenase 1 family, member L1(Aldh1l1)	1.0	0.07
Mfge8	milk fat globule-EGF factor 8 protein(Mfge8)	0.86	0.10
Cebpb	CCAAT/enhancer binding protein (C/EBP), beta(Cebpb)	0.85	0.05
Wnt7b	wingless-type MMTV integration site family, member 7B(Wnt7b)	0.83	0.06
Olig1	oligodendrocyte transcription factor 1(Olig1)	0.78	0.08
Plxnd1	plexin D1(Plxnd1)	0.76	0.05
Hap1	huntingtin-associated protein 1(Hap1)	0.75	0.05
Fxyd6	FXYD domain-containing ion transport regulator 6(Fxyd6)	0.70	0.07
Fat1	FAT atypical cadherin 1(Fat1)	0.66	0.09
Nufip2	nuclear fragile X mental retardation protein interacting protein 2(Nufip2)	-0.69	0.09
Rgs2	regulator of G-protein signaling 2(Rgs2)	-0.77	0.09
Sema3c	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3C(Sema3c)	-0.81	0.07
Me3	malic enzyme 3, NADP(+)-dependent, mitochondrial(Me3)	-0.82	0.10
Maf	avian musculoaponeurotic fibrosarcoma oncogene homolog(Maf)	-0.83	0.04
Npnt	nephronectin(Npnt)	-0.86	0.02
Igfbpl1	insulin-like growth factor binding protein-like 1(Igfbpl1)	-0.88	0.05
Col19a1	collagen, type XIX, alpha 1(Col19a1)	-0.91	0.04
Btg2	B cell translocation gene 2, anti-proliferative(Btg2)	-1.0	0.08
Gabrd	gamma-aminobutyric acid (GABA) A receptor, subunit delta(Gabrd)	-1.0	0.03
Mkx	mohawk homeobox(Mkx)	-1.0	0.03
Eya4	EYA transcriptional coactivator and phosphatase 4(Eya4)	-1.1	0.07
Txnip	thioredoxin interacting protein(Txnip)	-1.1	0.03
Nts	neurotensin(Nts)	-1.1	0.08
Dlx6os1	distal-less homeobox 6, opposite strand 1(Dlx6os1)	-1.1	0.04
Hmgcll1	3-hydroxymethyl-3-methylglutaryl-Coenzyme A lyase-like 1(Hmgcll1)	-1.1	0.05
Bcap29	B cell receptor associated protein 29(Bcap29)	-1.1	0.08
Lpl	lipoprotein lipase(Lpl)	-1.2	0.004
Zfp612	zinc finger protein 612(Zfp612)	-1.2	0.08
Prss35	protease, serine 35(Prss35)	-1.3	0.05
B2m	beta-2 microglobulin(B2m)	-1.3	0.05
Plin2	perilipin 2(Plin2)	-1.5	0.010
Zeb2	zinc finger E-box binding homeobox 2(Zeb2)	-1.5	0.004
Acan	aggrecan(Acan)	-1.7	0.004
H19	H19, imprinted maternally expressed transcript(H19)	-1.8	0.02

Grem1	gremlin 1, DAN family BMP antagonist(Grem1)	-1.8	0.02
Pou2f2	POU domain, class 2, transcription factor 2(Pou2f2)	-1.8	0.004
Ctss	cathepsin S(Ctss)	-1.9	0.07
Pdia4	protein disulfide isomerase associated 4(Pdia4)	-1.9	0.004
C1qc	complement component 1, q subcomponent, C chain(C1qc)	-2.0	0.09
Lgals3	lectin, galactose binding, soluble 3(Lgals3)	-2.0	0.06
Lyz2	lysozyme 2(Lyz2)	-2.0	0.02
Anpep	alanyl (membrane) aminopeptidase(Anpep)	-2.0	0.03
Tyrobp	TYRO protein tyrosine kinase binding protein(Tyrobp)	-2.5	0.10
Spp1	secreted phosphoprotein 1(Spp1)	-2.7	0.004
Gpnmb	glycoprotein (transmembrane) nmb(Gpnmb)	-2.8	0.004
Mirlet7b	microRNA let7b(Mirlet7b)	-3.1	0.02
Mmp12	matrix metallopeptidase 12(Mmp12)	-3.4	0.01
Cd36	CD36 antigen(Cd36)	-3.6	0.02
2010111I01Rik	RIKEN cDNA 2010111I01 gene(2010111I01Rik)	-3.9	0.004
Tlcd1	TLC domain containing 1(Tlcd1)	-5.1	0.004
Ttr	transthyretin(Ttr)	-5.5	0.08
Xist	inactive X specific transcripts(Xist)	-6.2	0.004

Table 5.5 shows the gene names, descriptions, $\log_2$(FC) where "FC" refers to fold-change, and q-values for the differentially-expressed genes in the S133A/WT comparison at 8 hours. The upregulated and downregulated genes are highlighted in pink and green respectively. q-values (FDR) < 0.1.

Table 5.6 PANTHER gene ontology classifications for S133A/WT upregulated genes

Gene ontology	Gene names	FE	P-value
Nervous system development	*Mfge8, Wnt7b, Fat1, Olig2, Ntng2, Hap1, Olig1, Plxnd1, Dach2*	18	9.1×10^{-8}
Ectoderm development	*Fat1, Olig2, Hap1, Olig1, Dach2*	17	2.2×10^{-3}
System development	*Fn1, Rttn, Pdgfra, Mfge8, Wnt7b, Cebpb, Fat1, Olig2, Ntng2, Hap1, Olig1, Plxnd1, Dach2*	12	2.5×10^{-7}
Developmental process	*Fn1, Rttn, Pdgfra, Mfge8, Wnt7b, Cebpb, Fat1, Olig2, Ntng2, Hap1, Olig1, Plxnd1, Dach2*	6.5	9.8×10^{-5}

Table 5.6 displays the major GO slim biological processes PANTHER gene ontology categories that are enriched in the S133A/WT upregulated genes at 8 hours. P-values are Bonferroni corrected. FE= fold enrichment.

Table 5.7 PANTHER gene ontology classifications for S133A/WT downregulated genes

Gene ontology	#	Gene names	FE	P-value
Extracellular matrix organization	7	*Ctss, Col19a1, Lgals3, Acan, Mkx, Grem1, Npnt*	23	2.0×10^{-4}
Extracellular structure organization	7	*Ctss, Col19a1, Lgals3, Acan, Mkx, Grem1, Npnt*	23	2.1×10^{-4}
Cell differentiation	18	*Spp1, Btg2, Col19a1, Lgals3, Zeb2, Bcap29, B2m, Acan, Eya4, Mkx, Rgs2, Txnip, Sema3c, Grem1, Anpep, Pou2f2, Maf, Npnt*	3.0	4.4×10^{-2}

Table 5.7 displays the major GO biological processes PANTHER gene ontology categories that are enriched in the S133A/WT downregulated genes at 8 hours. P-values are Bonferroni corrected. FE= fold enrichment.
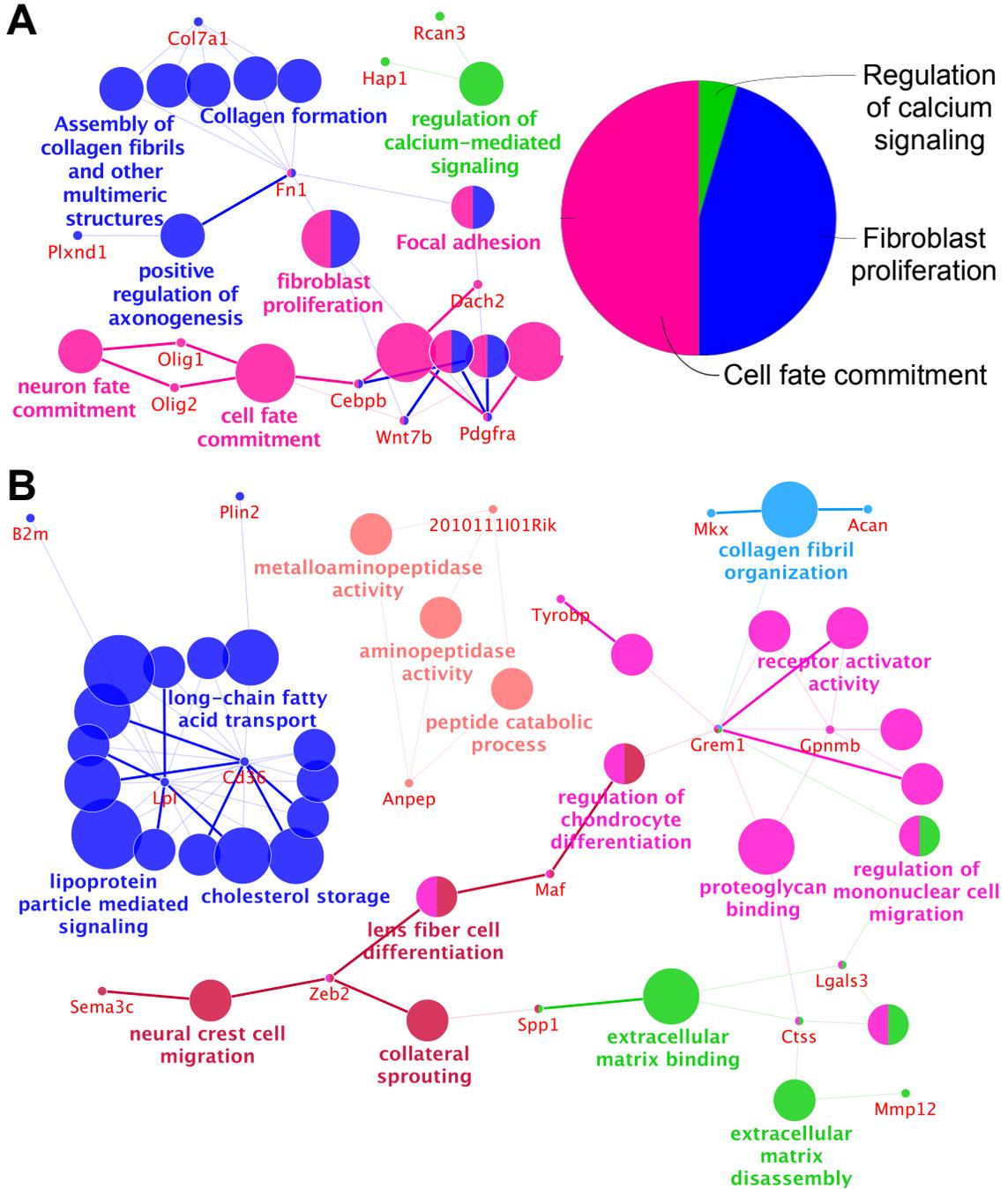
Figure 5.13 Cytoscape gene ontology annotations for the differentially-expressed genes in the S133A/WT comparison at 8 hours. (A) The upregulated genes are enriched for genes in involved in cell fate commitment and nervous and other system development. (B) The downregulated genes were enriched for lipid localization, extracellular matrix organization, neuronal development, and differentiation. The genes that belong to and are shared by these gene ontology categories are also shown. The gene annotations were created performed using the GO Biological Processes and Molecular Functions databases with the

5.7 The S40A-S133A double mutant affects nervous system development and the regulation of lipid localization

We next explored the DE genes from the glycosylation- and phosphorylation-deficient mutant (the S40A-S133A/WT comparison) (Table 5.8). In particular, genes related to nervous system development ($P = 4.7 \times 10^{-5}$) were upregulated in the S40A-S133A DE genes, including fatty acid-binding protein, brain *(Fabp7,* 2.1-fold), *Olig1* (1.8-fold), *Nntg2* (2.0-fold), *Hap1* (1.7-fold), *Wnt7b* (1.8-fold), and *Dach2* (3.3-fold) (Figure 5.14A). In addition to their upregulation in the double mutant, *Olig1*, *Nntg2, Hap1, Wnt7b,* and *Dach1* were also upregulated in S133A, indicating that ablation of phosphorylation at serine 133 influences nervous system development independent of the CREB glycosylation state. The downregulated genes in the double mutant were enriched for genes involved in the positive regulation of lipid localization ($P = 3.0 \times 10^{-2}$). Downregulated lipid localization genes included long-chain-fatty-acid-CoA ligase 5 (*Acsl5*, 2.1-fold), lipoprotein lipase (*Lpl*, 2.2-fold), and platelet glycoprotein IV (*Cd36*, 8.0-fold) (Figure 5.14B).

Table 5.8 List of differentially-expressed S40A-S133A/WT genes at 8 hours

Gene names	Gene description	S40A-S133A/ WT log2(FC)	q-value
Loxl2	lysyl oxidase-like 2(Loxl2)	7.4	0.004
Actr5	ARP5 actin-related protein 5(Actr5)	6.8	0.004
Parp1	poly (ADP-ribose) polymerase family, member 1(Parp1)	3.2	0.004
Dach2	dachshund 2 (Drosophila)(Dach2)	1.7	0.03
Tnc	tenascin C(Tnc)	1.5	0.004
Mfge8	milk fat globule-EGF factor 8 protein(Mfge8)	1.3	0.004
Fabp7	fatty acid binding protein 7, brain(Fabp7)	1.1	0.004
Apba3	amyloid beta (A4) precursor protein-binding, family A, member 3(Apba3)	1.1	0.05

Ntng2	netrin G2(Ntng2)	1.0	0.04
Fn1	fibronectin 1(Fn1)	1.0	0.004
Aldoc	aldolase C, fructose-bisphosphate(Aldoc)	1.0	0.04
Aldh1l1	aldehyde dehydrogenase 1 family, member L1(Aldh1l1)	1.0	0.06
Olig1	oligodendrocyte transcription factor 1(Olig1)	0.85	0.04
Kcnc4	potassium voltage gated channel, Shaw-related subfamily, member 4(Kcnc4)	0.83	0.09
Wnt7b	wingless-type MMTV integration site family, member 7B(Wnt7b)	0.82	0.06
Hap1	huntingtin-associated protein 1(Hap1)	0.78	0.04
Txnip	thioredoxin interacting protein(Txnip)	-0.90	0.09
Hmgcll1	3-hydroxymethyl-3-methylglutaryl-Coenzyme A lyase-like 1(Hmgcll1)	-1.0	0.09
Acsl5	acyl-CoA synthetase long-chain family member 5(Acsl5)	-1.0	0.07
Akt1s1	AKT1 substrate 1 (proline-rich)(Akt1s1)	-1.0	0.004
Zfp612	zinc finger protein 612(Zfp612)	-1.1	0.10
Lpl	lipoprotein lipase(Lpl)	-1.1	0.004
Btg2	B cell translocation gene 2, anti-proliferative(Btg2)	-1.2	0.01
Mbp	myelin basic protein(Mbp)	-1.3	0.02
Nts	neurotensin(Nts)	-1.3	0.02
Prss35	protease, serine 35(Prss35)	-1.3	0.04
Zeb2	zinc finger E-box binding homeobox 2(Zeb2)	-1.5	0.004
Plin2	perilipin 2(Plin2)	-1.5	0.03
Grem1	gremlin 1, DAN family BMP antagonist(Grem1)	-1.6	0.05
Plxnb3	plexin B3(Plxnb3)	-1.7	0.06
Acan	aggrecan(Acan)	-1.7	0.004
Anpep	alanyl (membrane) aminopeptidase(Anpep)	-1.9	0.04
C1qc	complement component 1, q subcomponent, C chain(C1qc)	-1.9	0.08
H19	H19, imprinted maternally expressed transcript(H19)	-2.0	0.004
Lyz2	lysozyme 2(Lyz2)	-2.0	0.01
Lgals3	lectin, galactose binding, soluble 3(Lgals3)	-2.0	0.04
Gpnmb	glycoprotein (transmembrane) nmb(Gpnmb)	-2.5	0.004
Spp1	secreted phosphoprotein 1(Spp1)	-2.5	0.004
Cd36	CD36 antigen(Cd36)	-3.0	0.01
Itgb2	integrin beta 2(Itgb2)	-3.2	0.09
Mir344g	microRNA 344g(Mir344g)	-3.2	0.004
Enpp6	ectonucleotide pyrophosphatase/phosphodiesterase 6(Enpp6)	-3.4	0.04
Kif1c	kinesin family member 1C(Kif1c)	-3.8	0.004
2010111I01Rik	RIKEN cDNA 2010111I01 gene(2010111I01Rik)	-3.8	0.004
Mmp12	matrix metallopeptidase 12(Mmp12)	-4.1	0.007

A330023F24Rik	RIKEN cDNA A330023F24 gene(A330023F24Rik)	-5.3	0.004
Xist	inactive X specific transcripts(Xist)	-5.7	0.004

Table 5.8 Shown here are the gene names, descriptions, $\log_2$(FC) where "FC" refers to fold-change, and q-values for the differentially-expressed genes in the S40A-S133A/WT comparison at 8 hours. The upregulated genes are highlighted in pink while the downregulated genes are highlighted in green. q-values (FDR) < 0.1.
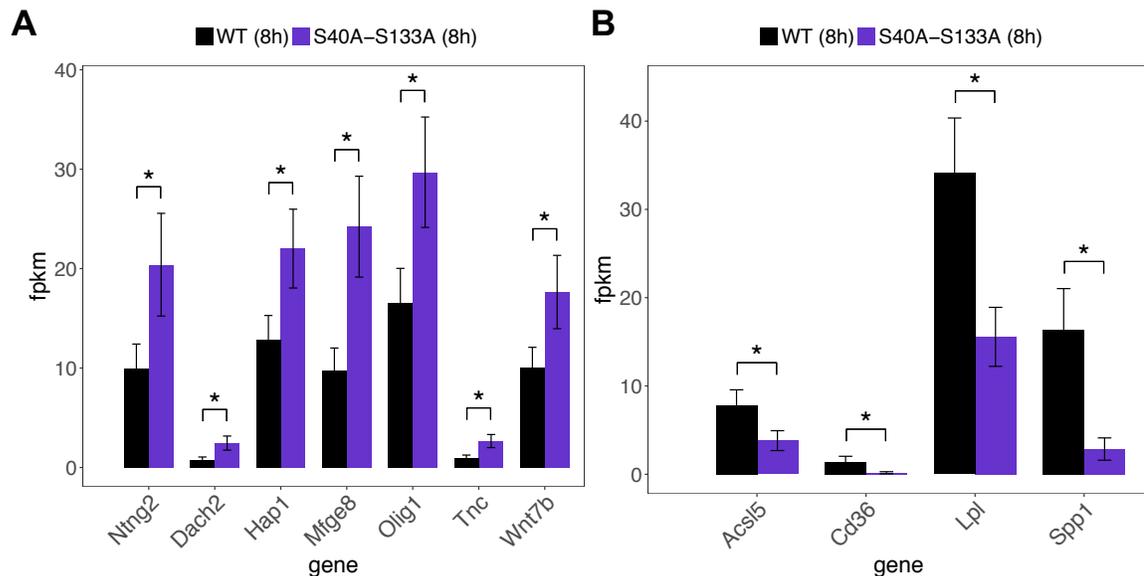
Figure 5.14 Differentially-expressed genes in S40A-S133A/WT at 8 hours are enriched for nervous system development and lipid localization genes. Genes involved in the (A) nervous system development and (B) positive regulation of lipid localization are upregulated in the S40A-S133A-CREB condition over the CREB condition. *FDR < 0.1.

5.8 CREB glycosylation and phosphorylation regulate different gene networks with the double mutant similar to the phosphorylation-deficient mutant

In addition to the three pairwise comparisons discussed earlier, there were 28 possible pairwise comparisons, which are represented in venn diagrams showing the number of common genes between the downregulated and upregulated genes for each pairwise comparison (Figures 5.15, 5.16). Overall, the largest differences could be observed when comparing 4 hours and 8 hours of expression, which is consistent with downstream second "wave" of transcription and late-phase LTP (L-LTP), the type of LTP that requires CREB-mediated transcription followed by translation.[2,20]
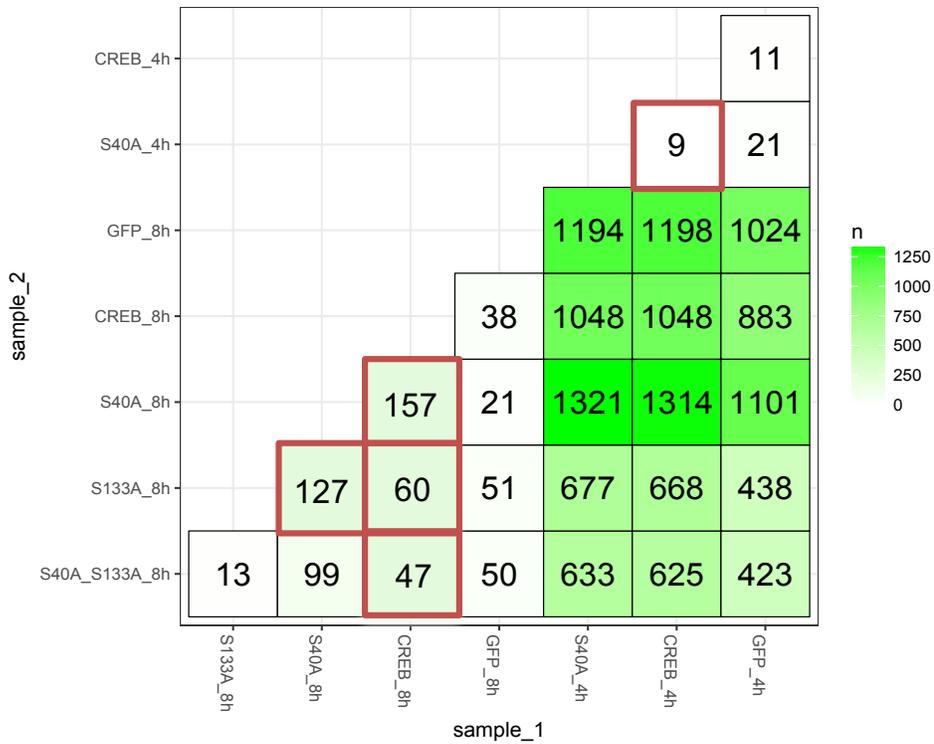
Figure 5.15 Differentially-expressed genes from the pairwise comparisons between different HSV treatment conditions. Following RNA-Seq and differential expression analysis using Cuffdiff, we obtained 28 different possible pairwise comparisons including the 4 hour and 8 hour time points. In the body of the paper, we focus on 5 of these comparisons: CREB_4h vs. S40A_4h, CREB_8h vs. S40A_8h, CREB_8h vs. S133A_8h, CREB_8h vs. S40A_S133A_8h, and S40A_8h vs. S133A_8h. These conditions that are explored are boxed in red. The q-value (FDR) cut-off used is 0.1 unless otherwise noted.
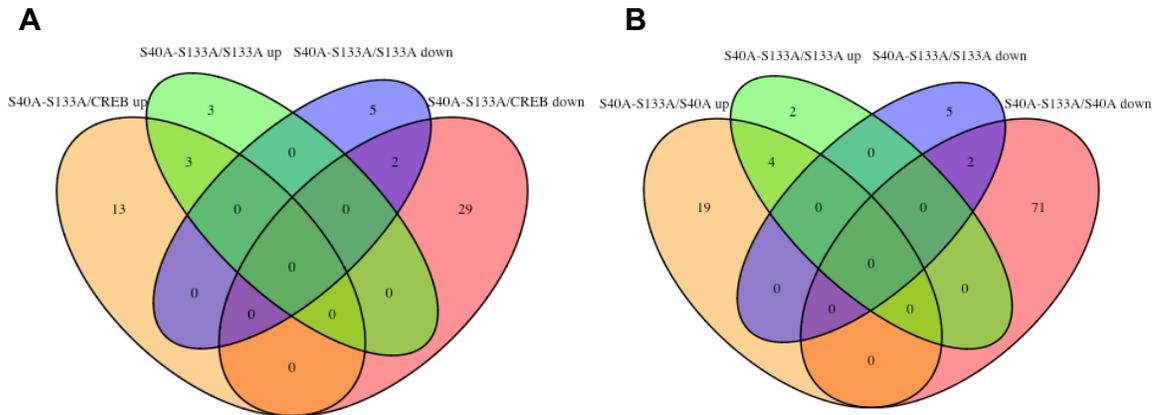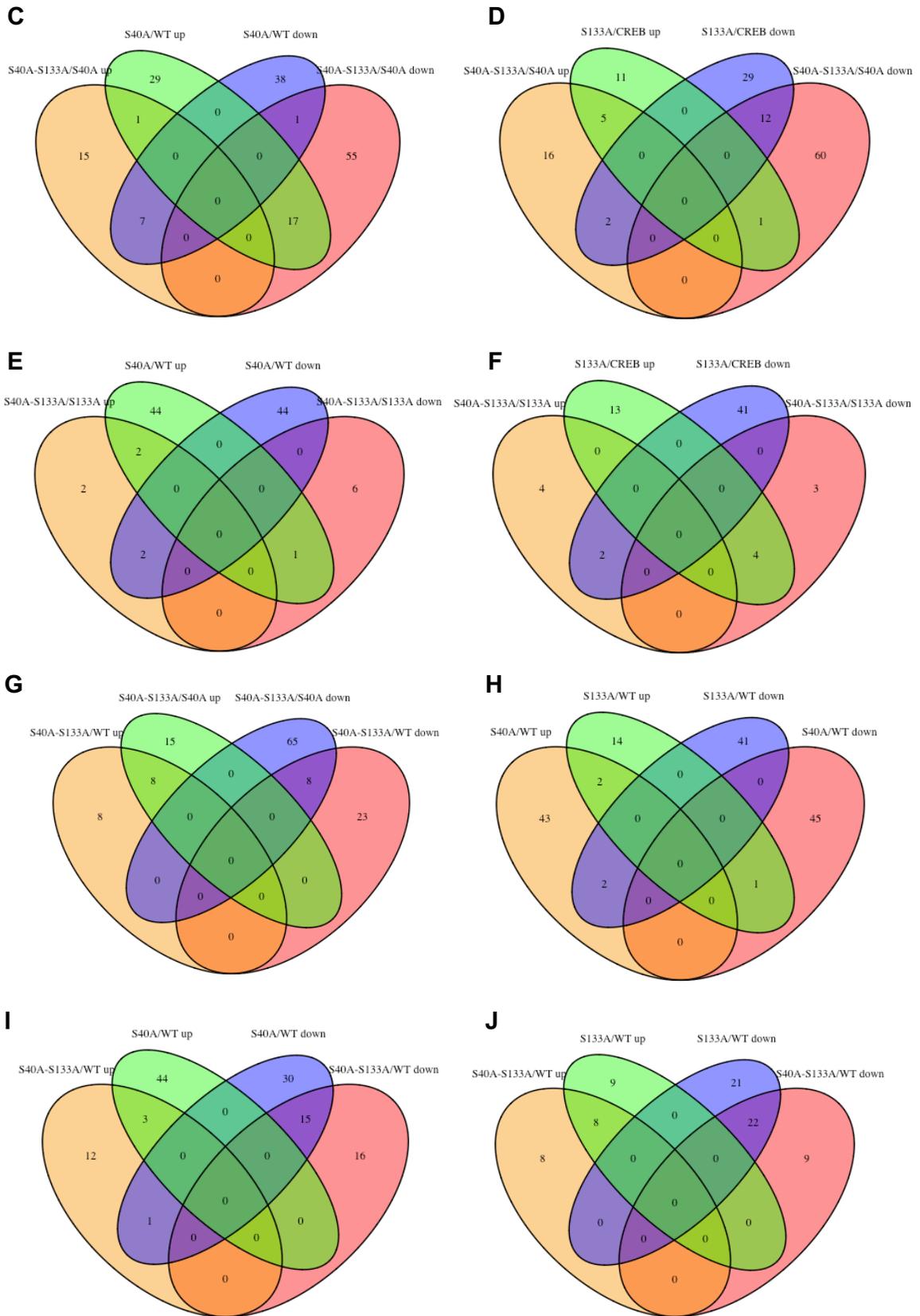
Figure 5.16 Venn diagrams showing pairwise CREB mutant comparisons at 8 hours. Shown here are the extra pairwise comparisons for the different CREB mutants separating upregulated and downregulated genes. (A) S40A-S133A/WT vs. S40A-S133A/S133A, (B) S40A-S133A/S40A vs. S40A-S133A/S133A, (C) S40A-S133A/S40A vs. S40A/WT, (D) S40A-S133A/S40A vs. S133A/WT, (E) S40A-S133A/S133A vs. S40A/WT, (F) S40A-S133A/S133A vs. S133A/WT, (G) S40A-S133A/WT vs. S40A-S133A/S40A, (H) S40A/WT vs. S133A/WT, (I) S40A-S133A/WT vs. S40A/WT, (J) S40A-S133A/WT vs. S133A/WT.

We also compared the DE genes in the S40A/WT, S133A/WT, and S40A-S133A/WT comparisons (Figure 5.17). Interestingly, there was no overlap between upregulated genes of the three mutant conditions, but there was some overlap between the downregulated genes. The downregulated genes that were common to all three pairwise comparisons were enriched for secreted proteins ($P = 2.3 \times 10^{-2}$). The genes that were only downregulated in the S40A/WT comparison were enriched for immune response ($P = 3.2 \times 10^{-3}$) genes. Neuronal excitability genes were exclusively upregulated in the S40A/WT condition and not in the S40A-S133A/WT and S133A/WT conditions. The pairwise comparisons between the S40A and S133A conditions show that S133A-CREB expression leads to enhanced expression of differentiation-related genes while S40A-CREB increases expression of calcium signaling, ion channel activity, and synaptic plasticity (Figures 5.9, 5.13). Finally, the S40A-S133A condition showed increased expression of nervous system development genes ($P = 4.7 \times 10^{-5}$) similar to the S133A condition and decreased expression of genes involved in lipid localization ($P = 3.1 \times 10^{-2}$) (Figures 5.13, 5.14). After comparing the double mutant with the single mutants, it appears that the double mutant displays more similarities with the S133A condition while the S40A condition alone enhances neuronal excitability. The DE genes suggest that glycosylation at serine 40 modulates distinct gene networks from phosphorylation at serine 133.

235
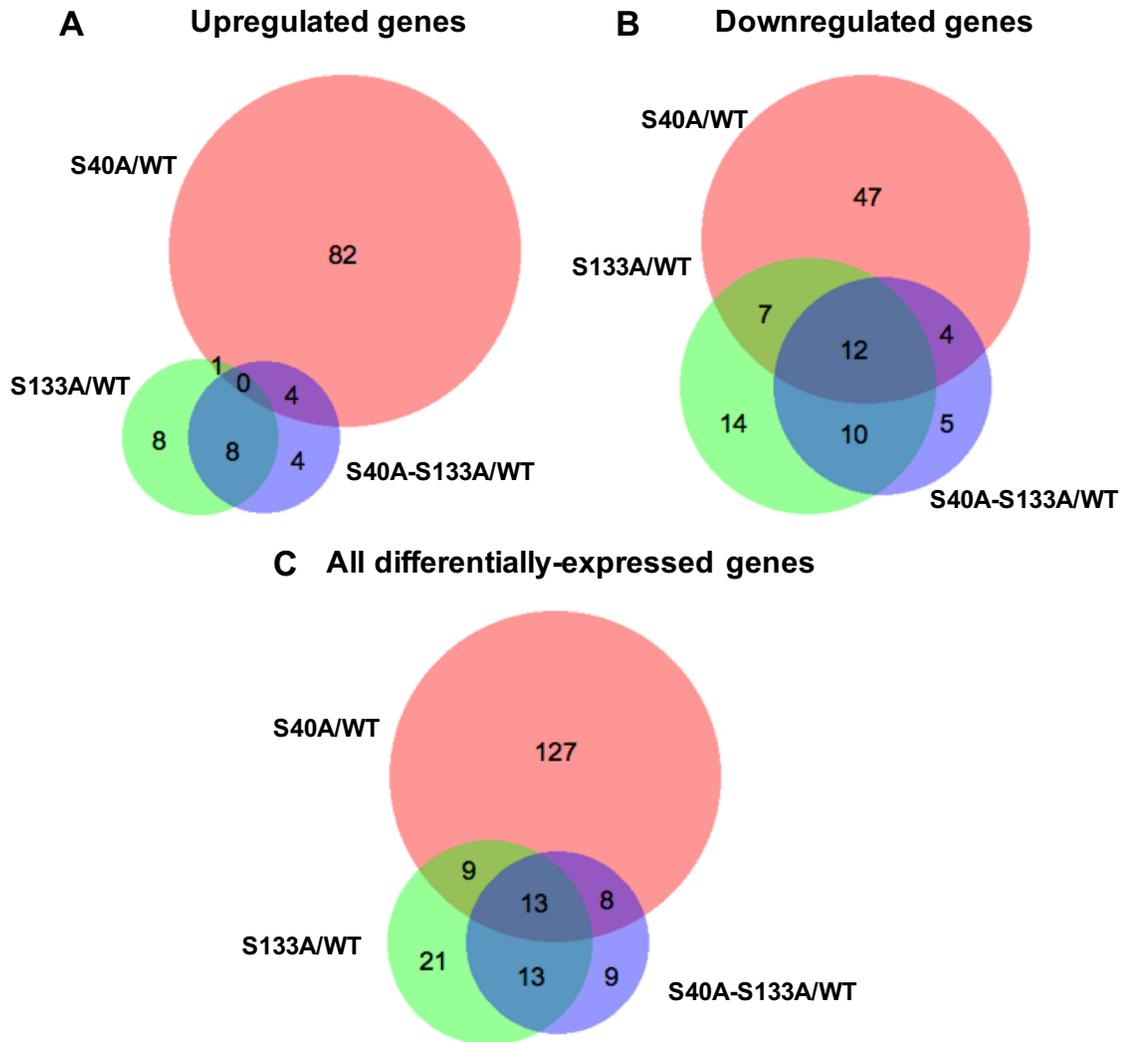
Figure 5.17 Venn diagrams showing overlap DE genes between various CREB mutants at 8 hours. These venn diagrams summarize the overlap between the DE genes in the following three conditions: S40A/WT CREB, S133A/WT CREB, and S40A-S133A/WT CREB. The venn diagrams show the overlap for (A) the upregulated DE genes, (B) the downregulated DE genes, and (C) all the DE genes. These venn diagrams show that there are some common genes downregulated by phosphorylation at S133 and glycosylation at S40, but no overlap between the upregulated genes. The Venn diagrams were generated using the BioVenn website.

5.9 CREB and its co-activators bind directly to DE gene promoters

We investigated whether CREB directly or indirectly regulated the DE genes identified from the S40A/WT, S133A/WT, and S40A-S133A/WT comparisons. CREB binds to full CRE sites (TGACGTCA) with high affinity or half CRE sites (TGACG/CGTCA) with slightly lower affinity.[21] Importantly, we found that 77-95% of

all the DE genes contained half CRE sites, suggesting that they are likely direct CREB targets (Table 5.9, Figure 5.18). Previous studies found no difference between the DNA binding capabilities of WT CREB and the S40A and S133A mutants, so we explored the WT CREB bound to these DE gene promoters in previous CREB ChIP-Seq studies.[3,22,23]

When we compared all our DE genes (S40A/WT, S133A/WT, S40A-S133A/WT) to CREB ChIP-Seq studies, we observed widely variable CREB ChIP-Seq enrichment on our genes across studies (0-48%). The CREB ChIP-Seq study performed in E16.5 mouse cortical neurons using the same conditions as ours (Kim *et al.* data set) yielded very low CREB ChIP-Seq binding to our DE gene promoters (0-6%).[5] A CREB ChIP-Seq study in mouse liver showed that 13-24% of DE genes were occupied by CREB while a CREB ChIP-Seq study using rat hippocampi had significantly higher overlap of 29-48%.[22,24] In the rat hippocampal CREB ChIP-Seq study, Lesiak and colleagues reanalyzed the mouse cortical study's CREB ChIP-Seq data (Kim *et al*. data set) using the same methods that they used for their ChIP-Seq analysis.[24] With this reanalyzed Kim *et al*. CREB ChIP-Seq peak data set (with FDR < 0.001), we found that about 8-17% of the DE genes were bound by the CREB. This coverage is closer to that observed in the CREB ChIP-Seq study from the mouse liver although the specific genes only show ~40% concordance for our DE genes when you compare the CREB occupancy from the E16.5 cortical neurons and mouse liver CREB ChIP-Seq studies (Figure 5.19). It is also important to note that we could have also used the rat hippocampal data set to compare our mouse data set as well since this has been done previously.[25] However, we decided to use the reanalyzed Kim data set at FDR < 0.001 because (1) we endeavored to utilize the ChIP-Seq data sets that were most similar to our conditions in terms of cell type and treatment and (2) the

FDR was sufficiently low to provide statistical confidence (FDR < 0.001). Thus, of all the DE genes, only 8-17% were occupied by CREB suggesting that many of the DE genes are downstream indirect targets of CREB potentially from the second "wave" of transcription following CREB activation (Table 5.10).[5,24]

Table 5.9 Full and half CRE sites on DE genes at 8 hours

Comparison	Up/Down	Full CRE	Half CRE	% CRE
S133A/WT	Up	1	20	80%
S133A/WT	Down	3	38	78%
S40A/WT	Up	1	37	77%
S40A/WT	Down	3	39	87%
S40A-S133A/WT	Up	1	20	95%
S40A-S133A/WT	Down	2	29	94%

Table 5.9 We used cruzdb to search 5000 bases upstream and 500 bases downstream of the transcription start sites in order to identify the total number of full and half CRE sites present for the DE genes. The reported %CRE gives the percent of genes that have either full or half CRE sites present. The number of full CRE sites includes the number of half CRE sites.
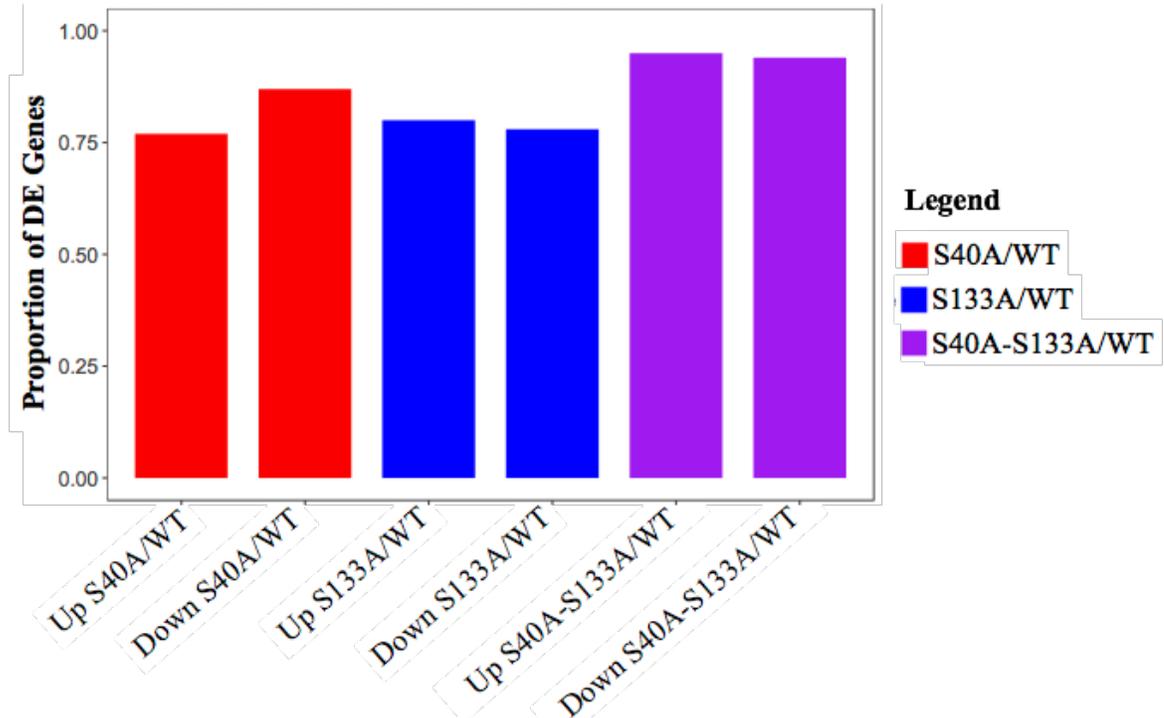
Figure 5.18 Barplot of half CRE sites on DE genes at 8 hours. The proportion of DE genes containing half CRE sites is shown in barplot format.

Table 5.10 CREB-regulated DE genes at 8 hours

	S133A/WT	S40A/WT	S40A-S133A/WT

Condition	% Up	% Down	% Up	% Down	% Up	% Down	Cell type
CBP_KCl[5]	52%	61%	60%	42%	57%	55%	CNs
CBP_noKCl[5]	4%	8%	2%	4%	10%	3%	CNs
CBP_KCl_dependent[5]	48%	53%	58%	38%	48%	52%	CNs
CREB[5,24]	8%	10%	17%	16%	14%	16%	CNs
CREB[22]	24%	18%	13%	13%	14%	19%	Liver
p300[26]	4%	10%	6%	13%	19%	3%	FB/MB
CRTC2[27]	8%	12%	8%	7%	10%	6%	3T3-L1
Up S133A/WT[28]	4%	4%	0%	0%	5%	5%	NA
Down S133A/WT[28]	0%	0%	2%	2%	0%	0%	NA
Up VP16-CREB/VP16[29]	8%	14%	42%	9%	10%	19%	HIPN
Down VP16-CREB/VP16[29]	12%	8%	35%	11%	5%	10%	HIPN
Up VP16-CREB/WT[16]	0%	8%	2%	11%	5%	10%	HIP
DownVP16-CREB/WT[16]	0%	0%	0%	0%	0%	0%	HIP

Table 5.10 shows an overview of the percent of differentially-expressed genes that are regulated by CREB and its co-activators. We compared our DE genes to ChIP-Seq studies (shown in blue) and microarray experiments (shown in purple). CN = E16.5 cortical neurons (DIV7 with KCl or without KCl), Liver = liver from mice fasted and then re-fed (males, age 8-12 weeks) FB/MB = E11.5 forebrain and midbrain, 3T3-L1 = 3T3-L1 mouse adipocytes treated with 10 μM forskolin, NA = overexpressing S133A or WT in nucleus accumbens (age 8 weeks), HIPN = overexpressing VP16-CREB or VP16 for 6 days in hippocampal neuronal cultures (10DIV), HIP = overexpressing VP16 or WT in hippocampus (age 8 weeks)

A   S133A/WT
B   S40A/WT
C   S40A-S133A/WT
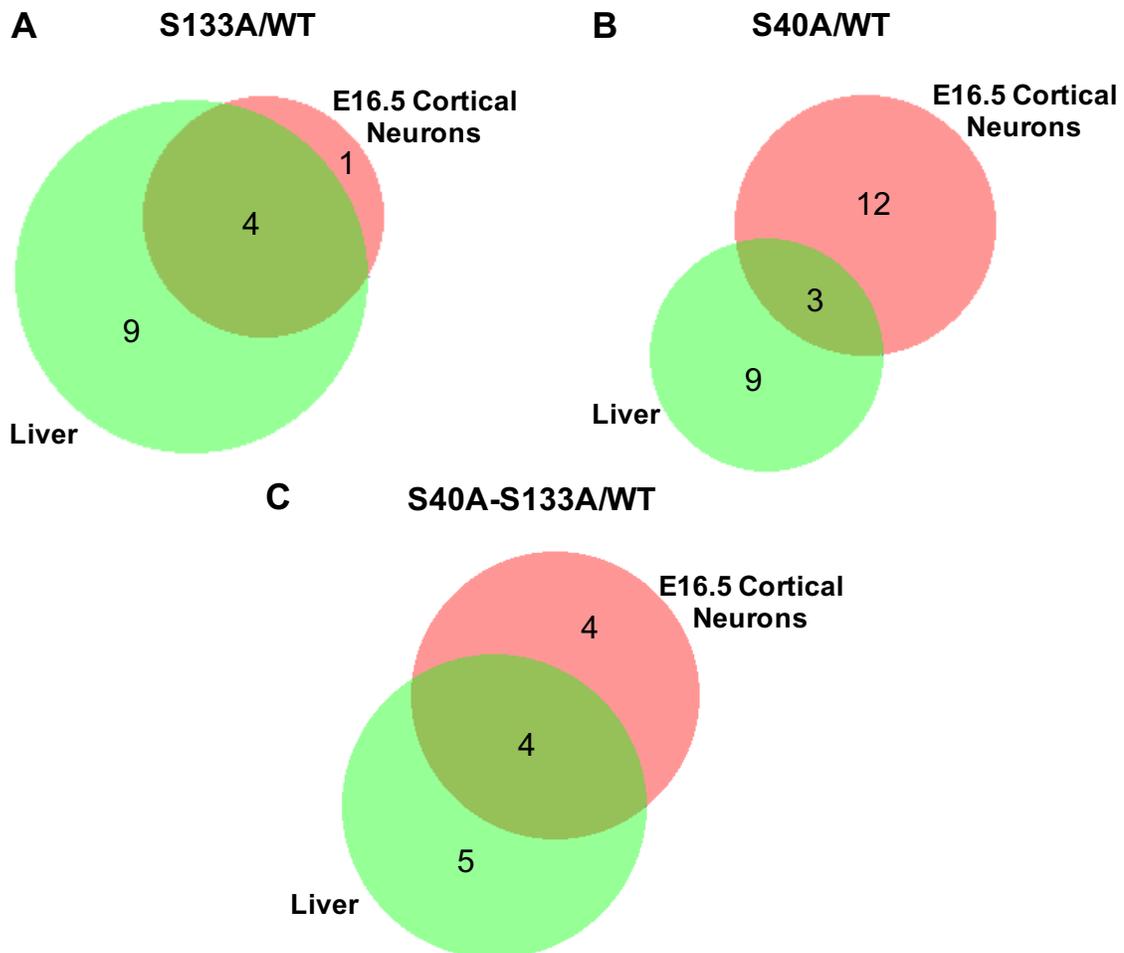
Figure 5.19 Overlap between CREB ChIP-Seq studies. Shown here are the overlapping DE genes that are occupied by CREB in E16.5 cortical neurons and 8-12 week livers (fasted and re-fed).

We next looked at whether other CREB-related proteins occupied the DE gene promoters.[5] Half of the DE gene promoters were occupied by the CREB coactivator CBP, but only under depolarization (not basal) conditions, indicating an activity-inducible CBP binding for these promoters (Table 5.10). The related CREB coactivator, p300, was found on fewer (3-19%) DE gene promoters in E11.5 brains.[26] For both p300 and CBP, CREB binding is mediated and encouraged by CREB phosphorylation at serine 133. However, we did not observe an overrepresentation of DE gene promoters occupied by p300 or CBP in the S133A/WT or S40A-S133A/WT conditions when compared to the S40A/WT condition. Given that CBP/p300 are coactivators for many TFs other than CREB and that the timing of our studies allows for secondary indirect CREB effects, the change in CBP or p300 binding cannot be solely attributed to direct CREB binding changes.[30] Finally, CRTC2, the major CREB glycosylation-dependent coactivator, was bound to 6-12% of DE gene promoters in 3T3-L1 mouse adipocytes.[27] We did not find an enrichment CRTC2 on the promoters of the DE genes in the S40A or S40A-S133A conditions when compared to the S133A condition likely for the same aforementioned reasons that we did not see differential binding of CBP and p300 in S133A mutant conditions. In summary, CREB and its coactivators were present on many of the S40A/WT, S133A/WT, and S40A-S133A/WT DE gene promoters.

5.10 Our study shows neuronal activity genes are upregulated by both VP16-CREB and S40A-CREB and minimal overlap between S133A and other studies exploring S133A-CREB gene changes

When we compared our DE genes to microarray and RNA-Seq studies expressing S133A mutant CREB, we found very little accordance except for a few key genes. First, McClung and colleagues showed that S133A-CREB overexpression results in opposing effects on transcription when compared to WT CREB in the nucleus accumbens of mice with a total of 24 differentially-expressed genes between the S133A-CREB and WT CREB conditions.[28] We first compared our DE genes to the 24 DE genes from a microarray study where McClung and colleagues overexpressed WT and S133A-CREB in the nucleus accumbens (NA).[28] Only one common gene was upregulated in both our and their S133A/WT comparisons: huntington-associated protein 1 (*Hap1*) (note: *Hap1* was not statistically significantly upregulated in the NA S133A/WT study).[28] Hap1 is associated with neuronal differentiation, signaling, and morphogenesis.[31] There was also one neuronal growth gene, *Bdnf*, shared between the downregulated S133A/WT NA genes and the S40A/WT upregulated genes.[28] Finally, *Fabp7* (fatty acid brain protein 7), a gene important for neuronal differentiation, was shared between the downregulated S133A/WT NA comparison and the upregulated S40A-S133A/WT comparison. This is supportive of the role of S133A in regulating expression of neuronal differentiation genes.

Overall, we did not observe significant overlap between the McClung and colleagues microarray study; potential reasons for the differences could be (1) that the study involved overexpression of S133A-CREB and WT CREB in the nucleus accumbens over course of 8 weeks while our study used a Creb1$^{\alpha\delta}$ background and explored considerably shorter time periods (4 and 8 hours) in a E16.5 cortical neuronal

population and (2) the differing sensitivity in detection for microarray studies when compared to RNA-Seq experiments, especially with low abundance transcripts.[32]

Another RNA-Seq study by Briand and coworkers showed that a phosphorylation-deficient S133A mutant mouse had no DE genes in the hippocampus when compared with WT mice.[23] In this study, breeding of the heterozygous S133A-CREB mice resulted in lower numbers than the expected Mendelian frequencies (only 11% homozygotes), leading the authors to suggest that while S133 phosphorylation may not affect transcription and memory-related behavior, it may affect development.[23] Prior studies have shown that CREB plays a vital role in neuronal development and differentiation.[2,33,34] Our results corroborate the paramount role of phosphorylation at S133 in arbitrating neuronal development and differentiation.

Next, we compared our DE genes to another microarray study that explored the effects of constitutively active VP16-CREB overexpression in the hippocampus of mice. The VP16-CREB mutant heightened neuronal activity, but lead to eventual excitotoxicity.[16] VP16-CREB-expressing mice displayed increased the expression of immune response genes and impaired spatial memory retrieval despite exhibiting enhanced LTP.[2,16,35] We found that *Bdnf* was upregulated in our S40A/WT and their VP16-CREB/WT comparisons, which is supportive of more active CREB leading to neuronal growth gene expression.[16] We found that the immune response related genes, *Spp1* and *C1qa,* were both upregulated in the VP16-CREB/WT comparison and downregulated all of the comparisons S133A/WT, S40A/WT, and S40A-S133A/WT. This suggests that the glycosylation and phosphorylation mutants may have impaired transcription of immune response genes while the VP16-CREB mouse might have

enhanced expression likely due to the presence of excitotoxic effects in the VP16-CREBexpressing mouse. Altogether, the comparison with the VP16-CREB mouse shows that loss of glycosylation at serine 40 leads to similar enhancements in neuronal activity-related genes such as *Bdnf* and dissimilar decreases in immune response-related gene expression like *Spp1* and *C1qa*. Importantly, the VP16-CREB overexpression analysis only yielded 4 differentially-expressed genes total across all time points explored (1-5 weeks expression), and the only gene discussed here that was differentially-expressed in the VP16-CREB mice was *Bdnf*.

Finally, another study conducted by Benito and colleagues explored the effects of overexpressing constitutively active (ca) VP16-TF fusions (CREB, EGR1, FOS, SRF) in regulating transcription in cultured mouse hippocampal neurons. Briefly, wildtype or hippocampal neurons were treated with (1) lentivirus expressing GFP and caCREB, caEGR1, caFOS, caSRF, or VP16 (control) for 6 days or (2) various stimuli including forskolin (fors), bicuculline (bic), and BDNF.[29] The DE genes for each of these conditions and our mutant CREB conditions are provided in Table 5.11 (percentage of total DE genes are reported in Table 5.10). Through a comparison of the caCREB/VP16 and S40A/WT conditions, we sought to determine which DE genes are either expressed (1) similarly in response to heightened CREB activity (DE genes shared by both conditions) (type 1) or (2) in opposite directions in the caCREB and S40A-CREB conditions (DE genes that show contrasting expression in the two conditions) (type 2). Critically, the VP16-CREB fusion facilitates the association with the coactivator, CBP, whereas S40A-CREB facilitates the association with the coactivator, CRTC, and potentially CBP as well (because S40A-CREB can be phosphorylated at S133).

Therefore, any differences between VP16-CREB- and S40A-CREB-mediated expression may reflect the different genes that are regulated by CRTC specifically.

First, we looked at the overlap between the S40A/WT and caCREB/VP16 conditions in order to see which genes were differentially expressed in response to elevated CREB activity (type 1 genes). The upregulated genes in the caCREB/VP16 and the S40A/WT conditions were enriched for genes involved in chemical synaptic transmission ($P < 6.3$x10^{-3}) including *Bdnf*, *Cartpt* (cocaine- and amphetamine-related transcript protein), *Ache* (acetylcholinesterase), *Crh* (corticotropin releasing hormone), *Lynx1* (Ly6/neurotoxin 1), and *Cplx1* (complexin 1). This implicates enhanced CREB activity in elevating the expression of synaptic activity genes, which is consistent with the well-established role of CREB in mediating synaptic and neuronal activity.[20] Interestingly, many of these upregulated genes are also upregulated in response to forskolin, suggesting that increased cAMP signaling can induce the expression of some of the same neuronal activity genes that are induced in response to caCREB and S40A-CREB. The genes that were downregulated in the S40A/WT and caCREB/VP16 conditions were involved in neuronal axonal growth and motility (*Nefm*, *Islr2*, and *Gpc2*) and oxidative stress response (*Carhsp1* and *Txnip*). Altogether, these common DE genes likely reflect the genes regulated by CREB and both of the coactivators, CBP and CRTC.

We next explored the DE genes that displayed opposing expression in the S40A/WT and caCREB/VP16 comparisons (type 2 genes). The S40A/WT condition showed elevated expression of ion and solute transport genes ($P < 7.5$x10^{-3}) such as *Slc6a1*, *Cacna1i*, *Kcns2*, *Slc6a7*, and *Kcnh3*, while the caCREB/VP16 condition revealed downregulation of these genes. Specifically, GAT1 (*Slc6a1*) is the most prevalent brain-

specific GABA (gamma-aminobutyric acid) transporter that removes GABA from synapses thereby enhancing excitatory signaling.[36] These results indicate that S40A-CREB specifically strengthens whereas caCREB diminishes neuronal excitability through ion channel expression. Furthermore, genes involved in the regulation of cell adhesion ($P < 6.1 \times 10^{-3}$) and cytokine production ($P < 3.6 \times 10^{-2}$) were downregulated in S40A/WT, but upregulated in caCREB/VP16 conditions. This suggests that immune response and cell adhesion gene expression was downregulated in the S40A-CREB condition and upregulated in the caCREB/VP16 condition. Altogether, this comparative analysis suggested that enhanced CREB activity (through either S40A-CREB or caCREB expression) facilitates the expression of genes involved in neuronal activity. Unlike caCREB, ablation of CREB glycosylation specifically upregulated genes that enhance neuronal excitability and downregulated genes involved in immune response and cell adhesion likely through specific enhancement of CRTC binding capabilities.

In order to check to see if these DE genes were mediated by CBP or CRTC, we explored the ChIP-Seq data sets for the genes. Overall, CBP could be found on the promoters of 15 of the similarly expressed (type 1) genes and 13 of the differently expressed (type 2) genes in mouse cortical neuronal cultures.[5] In contrast, CRTC1 was solely found on 8 of the differently expressed (type 2) genes and none of the similarly expressed (type 1) genes in drosophila mushroom bodies in response to learning.[37] This supports the hypothesis that the differently expressed genes between the S40A-CREB and VP-CREB reflect genes that are regulated by CRTC rather than CBP. However, ChIP-Seq in mouse neurons rather than drosophila would more firmly establish this connection.

245

Table 5.11 Differentially-expressed genes in the Benito study and our study

Gene	caCREB	caFOS	caEGR1	caSRF	bic	fors	BDNF	S133A	S40A	S40A-S133A
Spock3	2.86								1.55	
Ptprn	2.49					2.16			1.41	
Bcap29	2.35							-2.19		
Scg2	2.31	1.8			1.49	2.44			1.43	
Crh	2.26								2.14	
Kcnc4	2.26								1.57	1.78
Gpnmb	2.18							-6.92	-6.10	-5.68
Ache	2.15								1.42	
Bdnf	2.12	2.08			1.66	1.75			1.58	
Gng4	2.03		-2.18			1.3			1.66	
Fn1	2.02					1.73		2.19	-1.71	2.04
Lynx1	1.93								1.80	
Cntnap1	1.9								1.67	
Btg2	1.9				6.01			-1.94		-2.34
Gda	1.81								1.39	
Nrsn1	1.78								1.69	
Prss35	1.69							-2.50		-2.51
Rcan2	1.54								1.56	
Car4	1.5								3.87	
Kcnh1	1.5	-1.4							1.47	
Acan	1.49							-3.17		-3.28
Nckap1l	1.49								-2.93	
Ngef	1.45								1.51	
Zfp612	1.43							-2.30		-2.12
Cplx1	1.43								1.41	
Acsl5	1.41									-2.05
Cebpb	1.4							1.80		
Eya4	1.39							-2.08		
Cartpt	1.39					1.77			4.43	
Shb*	1.36	1.68							-1.43	
Zbtb4	1.34								1.43	
Shh	1.27								1.54	
Parp1	-1.29					1.39			2.22	9.10
Osbpl1a*	-1.31								1.46	
Kcns2*	-1.32								1.79	
2310003H01Rik	-1.34							1.41		
Cacna1i	-1.35								1.57	
Bai2	-1.42							1.16		
Fxyd6	-1.45		-1.47					1.62		
Crhr1*	-1.46	1.63							1.46	
Fam81a	-1.47								1.46	
D3Bwg0562e	-1.47								1.42	
Negr1*	-1.47								1.38	
Nefm	-1.48								-1.46	
Ildr2	-1.59	1.43							1.49	

Creg2	-1.68					-1.26			2.02	
*Itpka**	-1.72					-1.25			2.35	
Zeb2	-1.75						-2.85			-2.78
Slc6a7	-1.75								1.65	
Gpc2	-1.75								-1.54	
Phyhip	-1.76								1.51	
*Kcnh3**	-1.81								1.61	
A830018L16Rik	-1.83								1.46	
Txnip	-1.91					-2.48		-2.08	-1.93	-1.87
Islr2	-1.91								-1.42	
Sema3c	-1.94		-2.69					-1.75		
Carhsp1	-2.02								-1.64	
Camk2a	-2.04								1.58	
*Slc6a1**	-2.14								1.46	
Lpl	-2.73	-2.52		-2.53			-1.72	-2.25		-2.19
Ntng2							-1.86		2.17	2.04
Mfge8						-2.36			1.81	2.49
Plxnd1						1.21	1.69			
Rgs2				2.59			-1.70			
Npnt		-1.66				1.2	-1.82			
Cckbr		1.55							2.69	
Rasal1			1.98	1.56					1.67	
Ncald						-1.16			1.51	
Nrip3							1.82		1.48	
AI593442						-1.49			1.45	
Grik3				1.83					-1.39	
Bcl2l11					1.25				-1.45	
Sox11		-1.82			1.27				-1.46	
Fst		-1.36							-1.49	
Tnc						1.69				2.75
Aldoc						-2.04				2.03
Akt1s1						1.28				-2.06
Itgb2		1.41							-5.66	-9.13
Mmp12				1.57				-10.5	-24.8	-16.9
Nab2			5.53	1.58					1.75	

Table 5.11 indicates the fold changes for various conditions in the Benito and colleagues study and for our CREB mutants in our study. The results are color-coded based on whether the genes are upregulated (red) or downregulated (green). The hippocampal neurons were treated with (1) lentivirus expressing GFP and constitutively active (ca) caCREB, caEGR1, caFOS, and caSRF for 6 days (normalized to lentivirus expressing the VP16 domain and GFP) or (2) various stimulation conditions including forskolin (fors), bicuculline (bic), and BDNF (normalized to vehicle). In our conditions, embryonic cortical neurons were treated for 8 hours with S40A, S133A, and S40A-S133A mutant CREB (normalized to WT CREB). *genes that are bound by CRTC1 in response to learning in drosophila (Hirano and colleagues, 2016).

5.11 DE gene promoters are occupied by activating histone modifications

After identifying several CREB-related activators were present across the DE gene promoters, we assessed the overall chromatin accessibility across all of the S40A/WT, S133A/WT, and S40A-S133A/WT DE gene promoters at 8 hours (both upregulated and downregulated). We found that the promoters of these DE genes were mostly associated with activating histone modifications such as H3K4me$_1$, H3K4me$_3$, and H3K27Ac and less associated with the repressive histone modification H3K27me$_3$ in E13.5 basal ganglia (BG).[38] For the activating histone modifications, we found H3K4me$_1$, H3K4me$_3$, and H3K27Ac on 42-57%, 37-44%, and 15-38% of the DE gene promoters respectively (Table 5.12, Figure 5.20).[38] In contrast, the repressive histone modification H3K27me$_3$ occupied 6-20% of the DE gene promoters across all comparisons.[38] Few of our DE genes have bivalent promoters, which are modified by both the activating H3K4me$_3$ and repressive H3K27me$_3$ modifications (0-6%). Bivalent promoters occur during differentiation and indicate that the gene is poised to change to either an activating or repressive state.[39] Under our conditions, most of our DE gene promoters contain univalent modifications and are therefore committed to either repressive or activating cell fates. Upon the DE gene promoters, we also observed the presence of activating gene-body associated H3K79me$_2$ and H3K36me$_3$ histone modifications, which are associated with cell cycle (highest at G2 phase) and actively transcribed exons respectively.[40,41] Based on the predominance of activating histone modifications, both the upregulated and downregulated DE genes are generally found in euchromatic and more transcriptionally active regions of the genome.

Table 5.12 Histone code for DE gene promoters at 8 hours

	S133A/WT		S40A/WT		S40A-S133A/WT		
Condition	**% Up**	**% Down**	**% Up**	**% Down**	**% Up**	**% Down**	**Cell type**

H3K4me$_3$[42]	64%	63%	83%	56%	71%	58%	ESCs
H3K4me$_3$[43]	60%	45%	31%	27%	57%	23%	dNPCs
H3K27me$_3$[43]	0%	8%	15%	2%	5%	6%	dNPCs
Both H3K4me$_3$ and H3K27me$_3$[43]	0%	4%	2%	0%	5%	6%	dNPCs
H3K4Me$_1$[5]	48%	57%	56%	42%	57%	52%	CNs
Increasing H3K27Ac with KCl[44]	16%	6%	6%	0%	10%	3%	CNs
Decreasing H3K27Ac with KCl[44]	4%	10%	6%	7%	10%	10%	CNs
Constant H3K27Ac with KCl[44]	8%	4%	8%	7%	19%	6%	CNs
Total H3K27Ac[44]	28%	20%	21%	13%	38%	19%	CNs
H2B-S112-O-GlcNAc[45]	44%	20%	33%	42%	52%	29%	ESCs
H3K4me3[46]	32%	45%	56%	56%	48%	48%	BM
H3K27me$_3$ gain with neuronal differentiation[47]	8%	0%	10%	2%	0%	0%	CNs/NPCs
H3K27me$_3$ loss with neuronal differentiation[47]	12%	8%	8%	4%	10%	6%	CNs/NPCs
H3K27me$_3$[38]	12%	16%	15%	20%	14%	19%	BG
H3K4me$_1$[38]	48%	35%	38%	33%	33%	39%	BG
H3K27Ac[38]	36%	37%	40%	40%	33%	35%	BG
H3K4me$_3$[38]	44%	37%	40%	42%	38%	39%	BG
Tet2, OGT, H2B-S112-O-GlcNAc[45]	16%	4%	17%	18%	19%	6%	ESCs
Tet2, O-GlcNAc, H3K4me$_3$[46]	0%	8%	0%	2%	5%	6%	BM
H3K79me$_2$[48]	36%	18%	4%	20%	24%	16%	ESCs
H3K36me$_3$[48]	20%	14%	2%	16%	14%	13%	ESCs

Table 5.12 shows an overview of the percent of the promoters of the differentially-expressed genes that are regulated by various histone modifications identified in ChIP-Seq experiments. We identified which of our DE gene promoters were occupied by activating (pink), repressive (green), and bivalent (blue) histone modifications. ESCs = v6.5 embryonic stem cells (E3.5), dNPCs = v6.5 embryonic stem cells differentiated into neural progenitor cells, CN = E16.5 cortical neurons (DIV7 with KCl or without KCl), BM = bone marrow, CNs/NPCs = FACS sorted E15.5 neural progenitor cells and neurons, BG = E13.5 basal ganglia
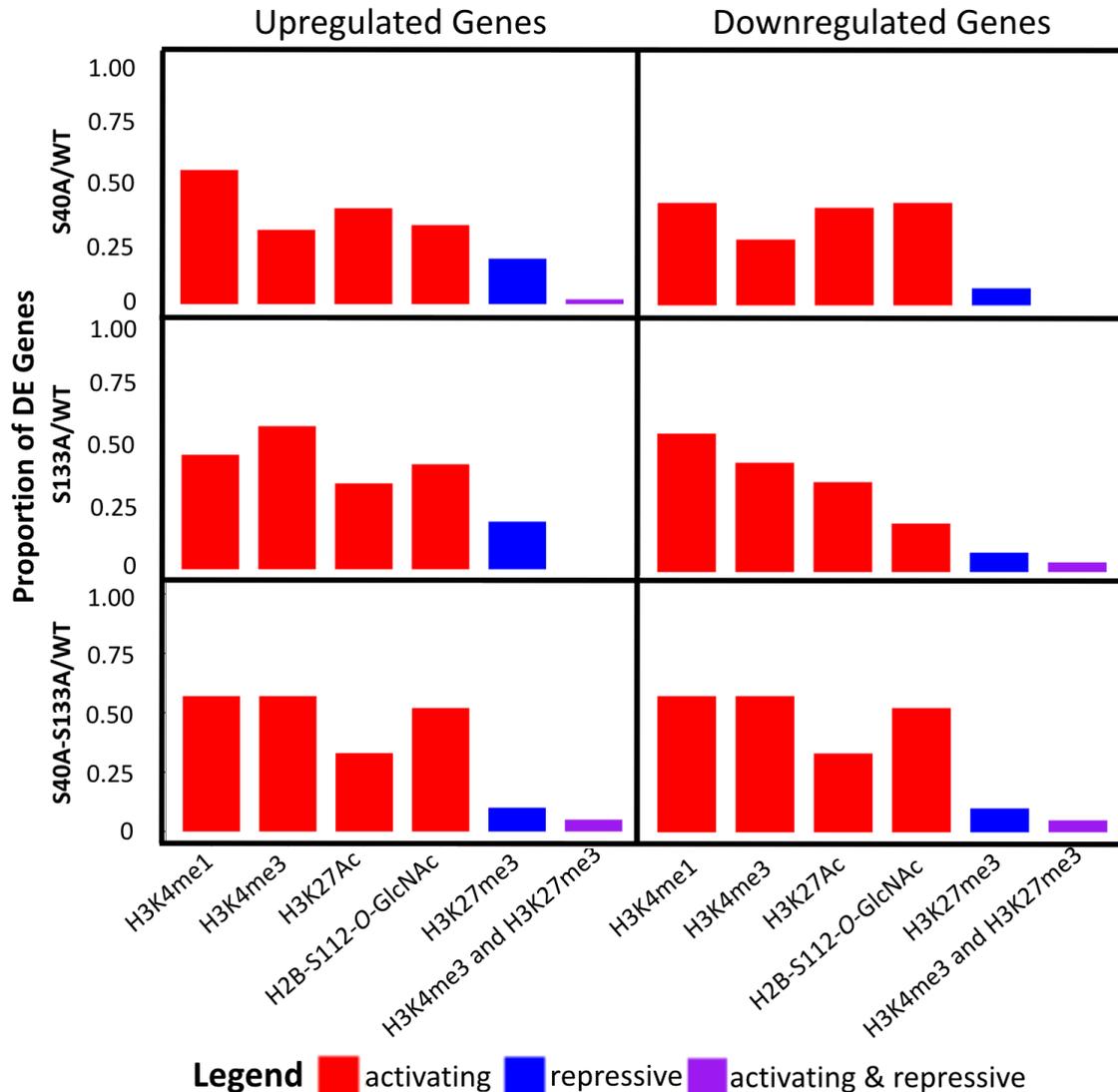
Figure 5.20 Histone modifications associated with the promoters of S40A/WT, S133A/WT, and S40A-S133A/WT DE genes. Shown here are the proportion of the promoters of DE genes occupied by either activating, repressive, or activating and repressive histone modifications including H3K4me$_1$[5], H3K4me$_3$[43], H3K27Ac[38], H2B-S112-*O*-GlcNAc[45], H3K27me$_3$[47], and bivalent modified genes (occupied by both H2K4me$_3$ and H3K27me$_3$)[43]. Overall, we observed that activating (red) more so than repressive (blue) histone modifications are associated with the 8-hour DE gene promoters. Few DE genes were occupied by bivalent, poised histone modifications (purple).

5.12 OGT, *O*-GlcNAc, and OGT-associated proteins and DNA modifications regulate the S40A/WT and S133A/WT DE genes at 8 hours

We then focused on the glycosylation-deficient mutant DE genes and found that the S40A/WT DE genes at 8 hours were regulated by *O*-GlcNAc and OGT directly. In a ChIP-seq study in mouse embryonic stem cells (ESCs), we found many of these gene

promoters were bound by OGT (65% of upregulated (up); 38% of downregulated (down) genes) as well as a variety of other TFs known to associate with OGT including the mSin3a corepressor (40% up; 44% down) and polycomb repressor complex 2 (PRC2) proteins EZH2 (46% up; 9% down) and Suz12 (31% up; 7% down) (Table 5.13).[42,45,46,49] The function and activity of the polycomb repressor complex 2 has been shown to be necessary for regulating *O*-GlcNAcylation throughout cells.[50,51] Furthermore, *O*-GlcNAcylation of RING1B, a major subunit of PRC1, increases the targeting of RING1B to neuronal genes and decreases the association with cell cycle genes in ESCs.[52] We observed 6-24% of DE gene promoters are occupied by RING1B in ESCs (Table 5.13). OGT also *O*-GlcNAc modifies EZH2 and the pluripotency regulating TFs, Sox2 and Oct4, and mediates their function.[53-55] In mouse bone marrow cells, *O*-GlcNAc (13% of upregulated and 16% of downregulated genes) was found on these S40A/WT DE gene promoters as well (Table 5.13).[46]

Table 5.13 OGT-related protein bound to DE gene promoters at 8 hours

	S133A/WT		S40A/WT		S40A-S133A/WT		
Condition	% Up	% Down	% Up	% Down	% Up	% Down	Cell type
O-GlcNAc[46]	12%	16%	13%	16%	19%	16%	BM
OGT[45]	32%	24%	33%	33%	38%	35%	ESCs
OGT[42]	72%	41%	65%	38%	48%	32%	ESCs
Tet1[42]	72%	71%	98%	76%	86%	71%	ESCs
Tet2[45]	32%	47%	46%	42%	29%	39%	ESCs
Tet2[46]	4%	20%	0%	11%	14%	16%	BM
Tet3[56]	4%	0%	2%	0%	5%	3%	CNs/NPCs
5hmC coverage in NPC[47]	32%	18%	21%	24%	29%	13%	CNs
5hmC coverage in neurons[47]	44%	14%	27%	22%	38%	16%	NPCs
5hmC loss with neuronal differentiation[47]	4%	4%	4%	2%	5%	0%	CNs/NPCs
5hmC gain with neuronal	20%	4%	10%	9%	19%	6%	CNs/NPCs

differentiation[47]							
5hmC in either neurons or NPCs[47]	48%	20%	38%	29%	38%	16%	CNs/NPCs
mSin3a[42]	48%	33%	75%	44%	57%	35%	ESCs
mSin3a[57]	36%	37%	40%	44%	43%	39%	pre-iPSCs
Ezh2[49]	32%	31%	46%	9%	24%	16%	ESCs
Suz12[48]	24%	22%	31%	9%	14%	10%	ESCs
Suz12[49]	24%	24%	31%	7%	14%	13%	ESCs
Ring1B[49]	24%	14%	17%	7%	14%	6%	ESCs
Oct4[48]	16%	16%	15%	27%	10%	16%	ESCs
Sox2[48]	16%	14%	15%	24%	19%	13%	ESCs
Nanog[48]	24%	8%	10%	16%	19%	10%	ESCs

Table 5.13 shows is an overview of the percent of differentially-expressed gene promoters that are bound by different OGT-related proteins and modifications and OGT and *O*-GlcNAc itself. We found the percentage of our DE gene promoters bound or modified by OGT and *O*-GlcNAc (green), Tet proteins or their products (orange), known OGT interactors including polycomb repressor complex 2 proteins (pink), polycomb repressor 1 complex proteins (purple), and pluripotency-related proteins (blue). All results come from ChIP-Seq experiments except for the 5hmC results, which are derived from hMeDIP experiments. BM = bone marrow, ESCs = v6.5 embryonic stem cells (E3.5), CNs/NPCs = FACS sorted E15.5 neural progenitor cells and neurons, pre-iPSCs = neural stem cell (NSC)-derived pre-induced pluripotent stem cells

In particular, we discovered that the promoters of the upregulated S40A/WT genes were overwhelmingly occupied by the OGT-associated Tet family of proteins [Tet1 (98%) in ESCs; Tet2 (46%) in ESCs; Tet3 (2%) in CNs/NPCs (E15.5 cortical neurons and neural progenitor cells)] (Table 5.13, Figure 5.21).[45,47,56] The Tet (ten-eleven translocation methylcytosine dioxygenase) family catalyzes the conversion of 5-methylcytosine (5mC) to 5-hydroxy-methylcytosine (5hmC), which are the first steps in the removal of 5-methylcytosine, a modification associated with gene silencing.[47] The 5hmC modification and Tet activity have been shown to be important for regulating pluripotency, neuronal activity, and gene activity.[45,46] When compared to a hMeDIP experiment that measured the occupancy of 5hmC, we found that the S40A/WT upregulated genes were 5hmC modified (38%) in CNs/NPCs (Table 5.13).[47] The presence Tet1 and 5hmC at gene promoters is associated with neuronal activity and

higher transcription.[58] The co-regulation of our S40A/WT upregulated genes with Tet1 and 5hmC underscores the impact of CREB glycosylation in regulating genes important for neuronal activity and growth.
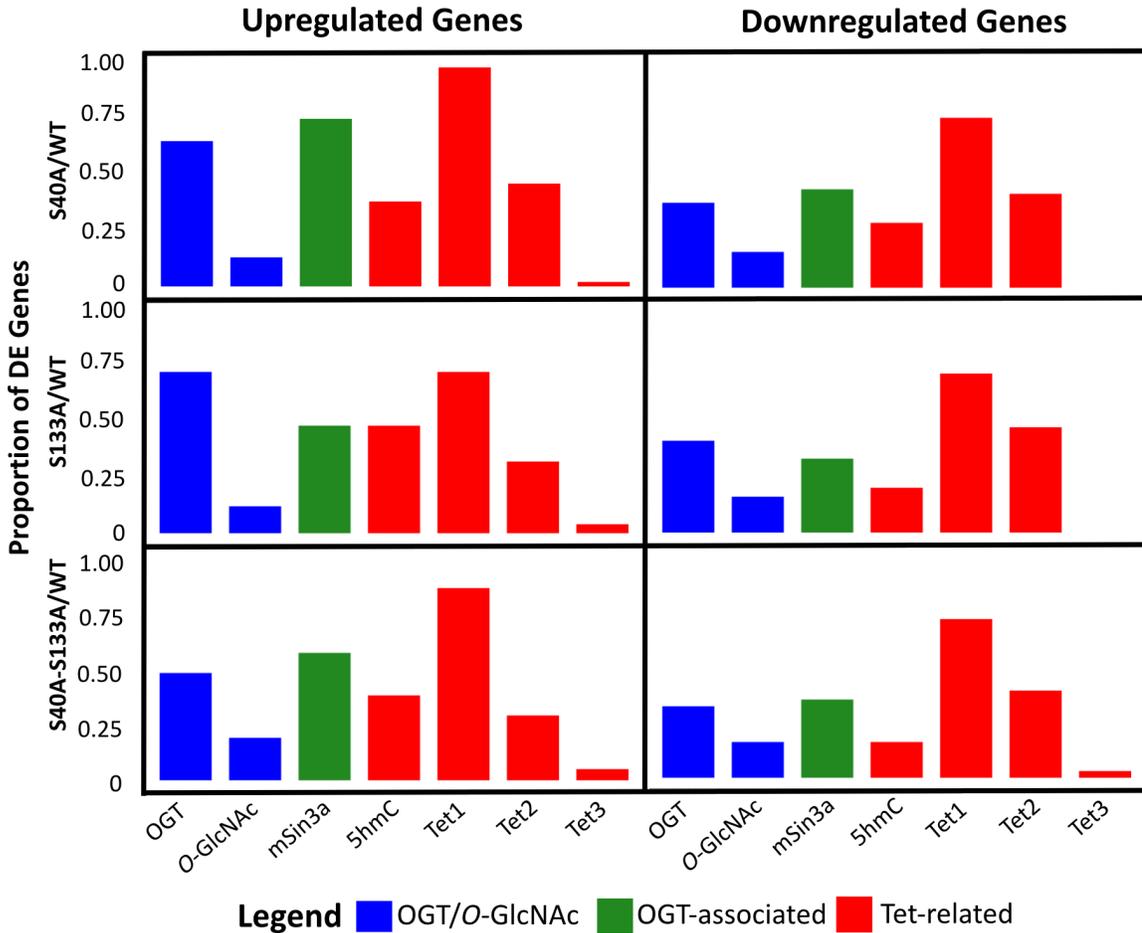
Figure 5.21 OGT-related proteins and modifications association with the S40A/WT, S133A/WT, and S40A-S133A/WT DE genes. Shown here are the proportion of genes occupied by OGT-related proteins and modifications including OGT[42], *O*-GlcNAc[46], mSin3a[42], 5hmC[47], Tet1[42], Tet2[45], and Tet3[56]. Overall, we observed that glycosylation-mediated genes are enriched for association with these different proteins and DNA modifications.

Previous studies have shown 5hmC and Tet1 association along active genes associated with neuronal differentiation.[47] Indeed, we see a particular increase in 5hmC with neuronal differentiation on the upregulated S133A/WT gene promoters: 20% of these promoters were associated with 5hmC gain with neuronal differentiation and only

4% of these promoters were associated 5hmC loss with differentiation (Table 5.12).[47] In addition, the TF Nanog was bound to the promoters of the upregulated S133A/WT genes (24%) more than the promoters of the downregulated S133A/WT genes (6%) indicating that these genes are important for pluripotency and self-renewal.[57] This supports the role of phosphorylation of CREB at serine 133 in adjudicating neuronal differentiation.

5.13 S133A and S40A-S133A is associated with neuronal differentiation and energy metabolism

To complement the pairwise comparisons of specific genes moderated by S133 phosphorylation and S40 glycosylation ablation, we next determined the gene networks through a weighted gene network analysis (WGCNA). WGCNA extracts biological meaningful gene networks from gene expression data using an unsupervised clustering approach based on gene coexpression (Figure 5.22).[59] The S133A and S40A-S133A conditions were anti-correlated (cor_{S133A} = $cor_{S40A-S133A}$ = -0.32, P = 4.1×10^{-7}) with the cyan module enriched for genes involved in the regulation of NPC proliferation (P = 3.6×10^{-2}) and integration of energy metabolism (P = 2.0×10^{-2}) (Figures 5.23, 5.24). The differentiation and metabolism module was neither correlated with WT CREB (cor = 0.032, P = 0.62) nor S40A-CREB (cor = 0.024, P = 0.71) (Figure 5.24). The anti-correlation of the phosphorylation-deficient mutants with a module enriched for NPC proliferation genes supports the differential expression analysis revealing an important role of CREB phosphorylation at serine 133 in neuronal differentiation. In addition, the anti-correlation of both the single and double phosphorylation with the cyan module show the similarity in the genes and gene networks regulated by the S133A and S40A-S133A mutants. Finally, we expand upon our DE gene analysis by discovering a role in

CREB phosphorylation in the regulation of several metabolic pathways: fatty acid metabolism ($P = 2.0 \times 10^{-2}$), amino acid metabolism ($P = 2.0 \times 10^{-2}$), insulin secretion ($P = 2.0 \times 10^{-2}$), and mitochondrial targeting ($P = 2.1 \times 10^{-2}$). CREB governs the metabolism of glucose, mitochondria, insulin, lipids, and fatty acids.[60] Furthermore, CREB phosphorylation at serine 133 increased the transcription of gluconeogenic genes.[61] Our study demonstrates a key role of CREB serine 133 phosphorylation in controlling metabolic and differentiation gene expression.
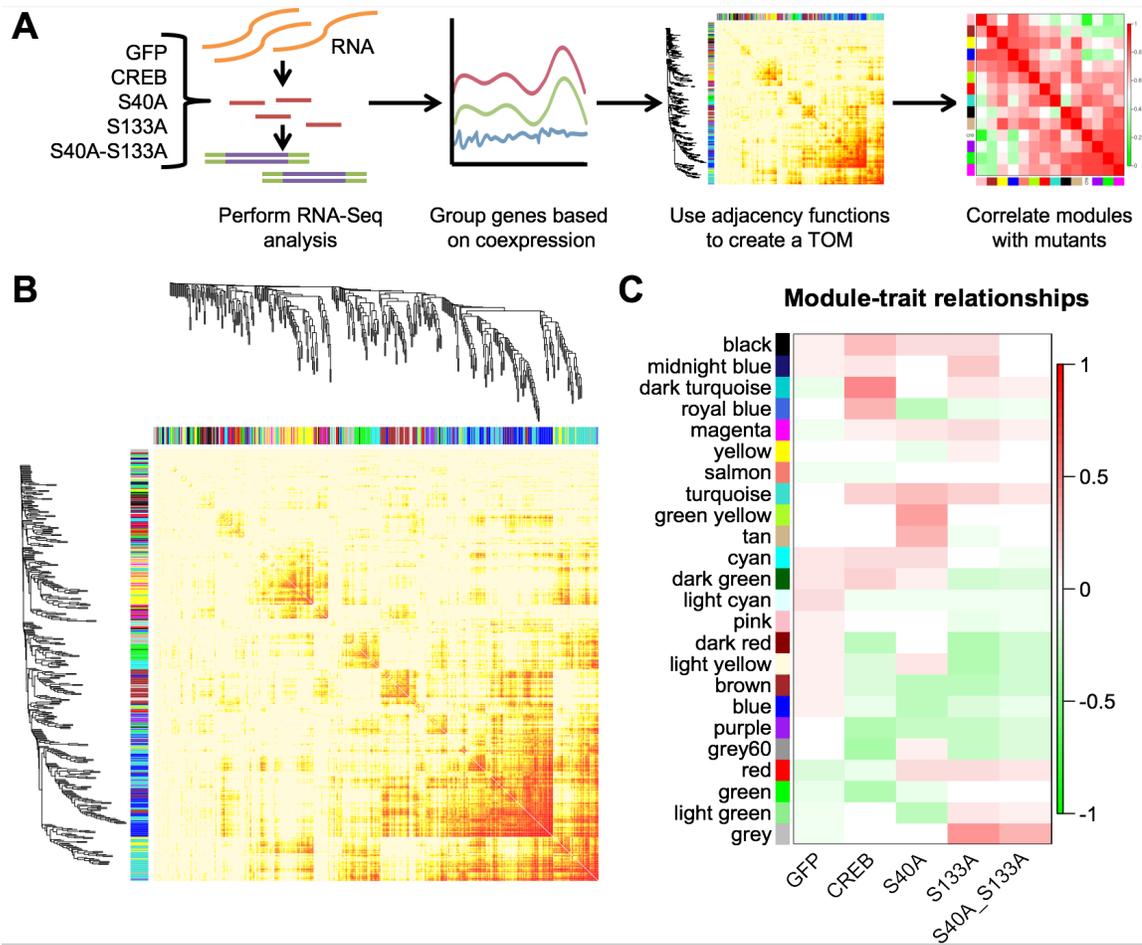
Figure 5.22 WGCNA overview. (A) Here is an overview of the procedure for WGCNA. Briefly, after RNA-Seq processing of various CREB-expressing neurons, the genes are grouped by co-expression. Then, using a weighted adjacency function, genes are clustered into biologically meaningful gene networks. Finally, networks are correlated with the different CREB mutants. (B) We generated gene dendrogram and network heatmap using our RNA-Seq data. Using a power of 10, the genes were divided into 24 different gene networks. (C) This heatmap indicates the module-trait relationships− specifically, the gene networks that are correlated with the different CREB mutants. The colors are arbitrarily assigned.
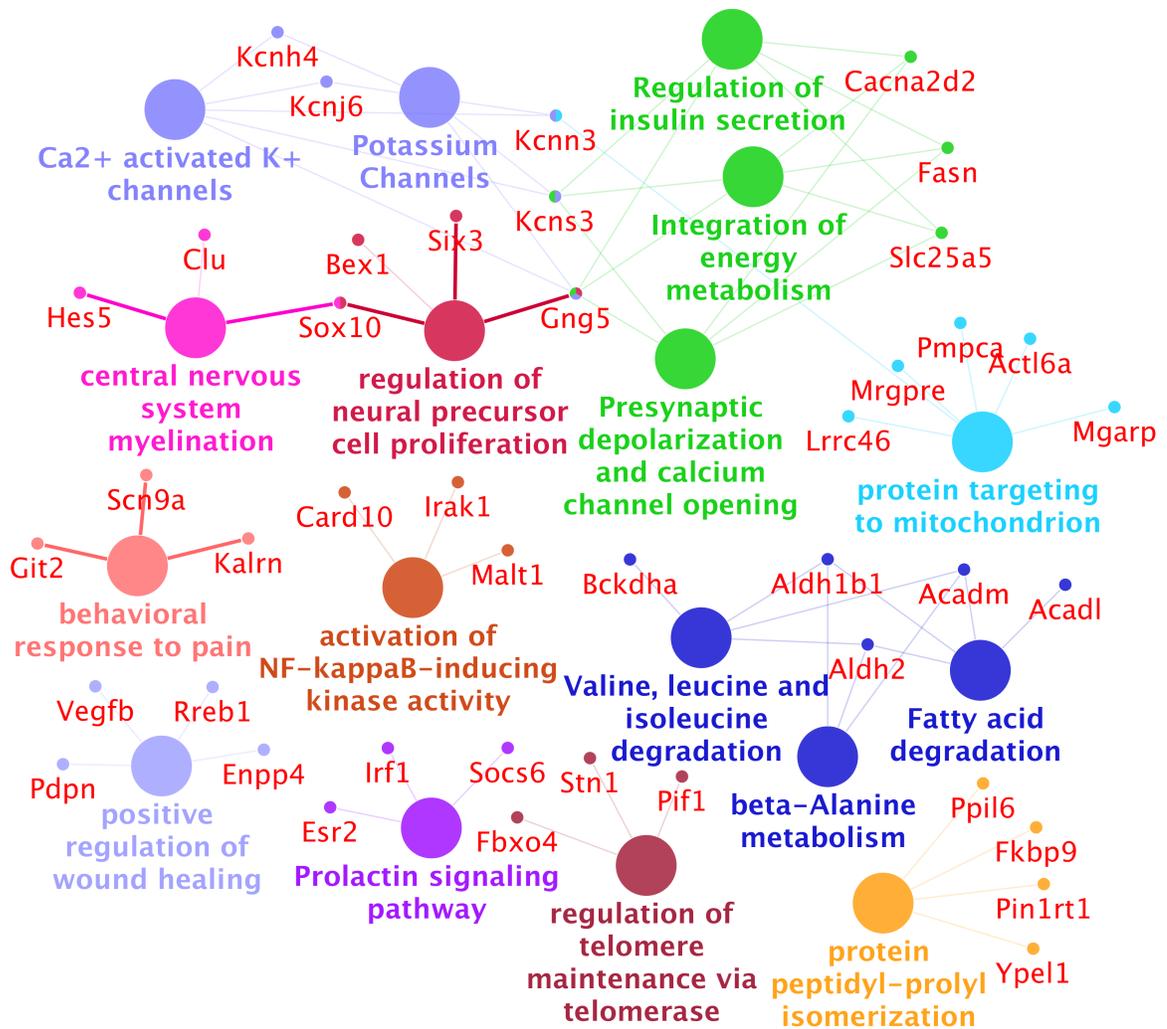
Figure 5.23 Gene ontology annotations for the S133A and S40A-S133A associated NPC proliferation-and metabolism-related module. Shown here are the top gene ontology categories, which include neural precursor cell proliferation ($P < 0.036$) and integration of energy metabolism ($P < 0.02$). The genes that belong to and are shared by these gene ontology categories are also shown. The module was annotated using the GO Biological Process, KEGG Pathways, and Reactome Pathways through the ClueGO extension in Cytoscape. The significance for all listed categories are Bonferroni step down corrected $P < 0.05$.

257
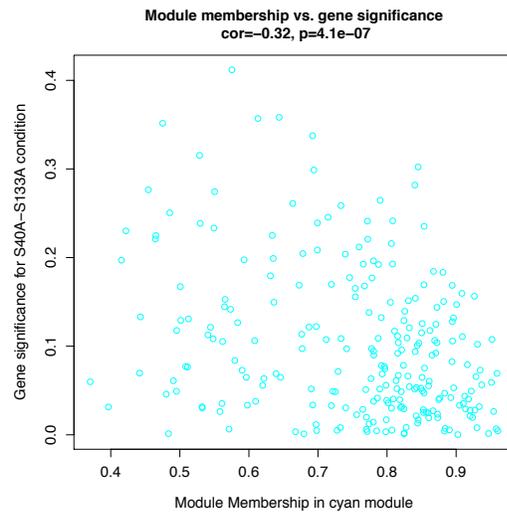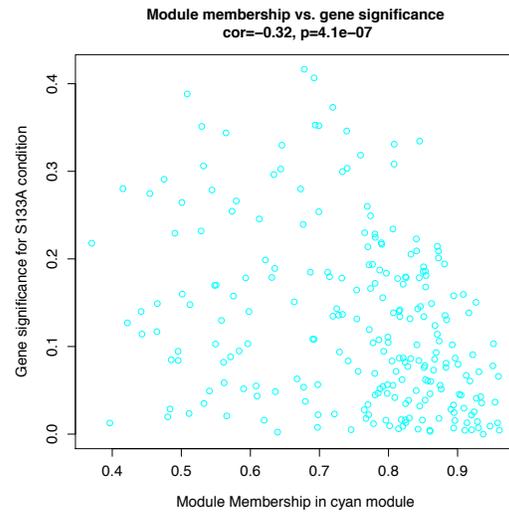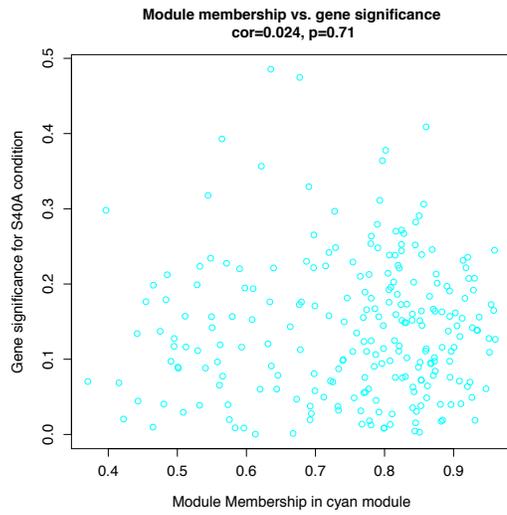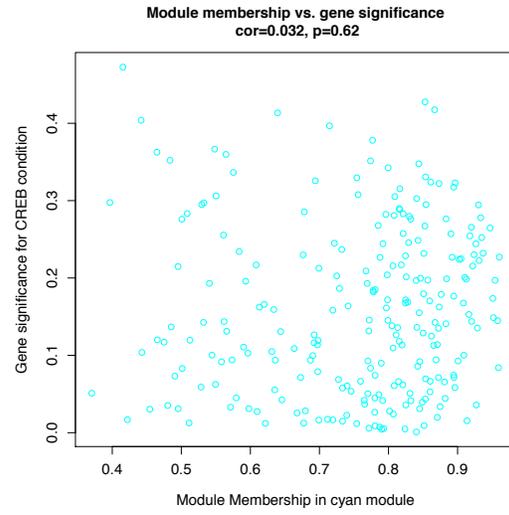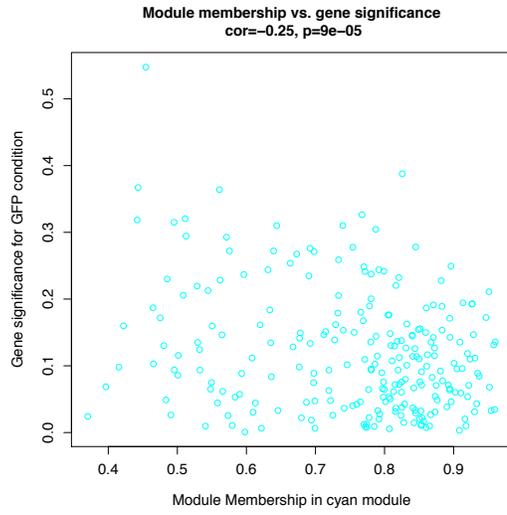
Figure 5.24 A differentiation- and metabolism-related cyan module is anti-correlated with GFP, S133A and S40A-S133A, and not correlated with WT and S40A. Module membership for the cyan module is anti-correlated with GFP (cor = -0.25, P = 9.0x10^{-5}), S133A-CREB (cor = -0.32, P = 4.1x10^{-7}), and S40A-S133A-CREB (cor = -0.32, P = 4.1x10^{-7}), but not correlated with WT CREB (cor = 0.032, P = 0.62) and S40A-CREB (cor = 0.024, P = 0.71).

In addition, genes with the highest and strongest intramodular connections (called "hub genes") likely play a critical role in regulating their modules.[62] The top five hub genes for the cyan module are solute carrier family 29 member 1 and family 25 member 37 (*Slc29a1, Slc25a37*), helicase like transcription factor (*Hltf*, HLTF), caspase recruitment domain family member 10 (*Card10*, CARD10), and arylsulfatase B (*N*-acetylgalactosamine-4-sulfatase, *Arsb,* Arsb) (Figure 5.25, Table 5.14). *Slc25a37* (mitoferrin 1) and *Slc29a1* both encode transporters that localize to the mitochondrial membrane (and plasma membrane in the case of *Slc29a1*) and regulate the uptake of heme and nucleosides respectively, which in turn dictates heme and nucleotide biosynthesis and metabolism.[63,64] HLTF is a translocase that is critical for coordinating and mediating DNA replication fork progression and DNA damage response.[65-68] CARD10 is a caspase recruitment domain that participates in apoptosis and mediates NF-κB signaling.[69] Arylsulfatase B catalyzes the removal of 4-sulfate groups from chondroitin-4-sulfate and dermatan sulfate and has been shown to play a critical role in neurite outgrowth and aerobic metabolism.[70,71] Together, *Slc25a37, Slc29a1, Hltf, Card10,* and *Arsb* are the hub genes that could be critical mediators of NPC proliferation and metabolic energy processes.
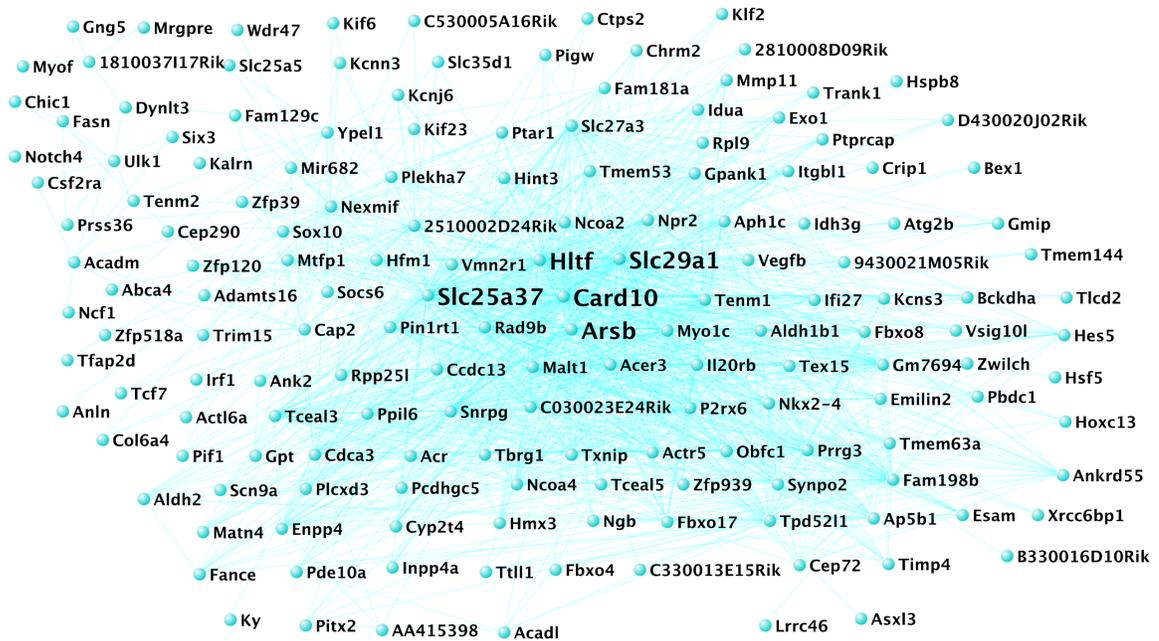
Figure 5.25 The cyan module is enriched for NPC proliferation- and metabolism-related genes. The cyan module gene network is enriched for genes involved in NPC proliferation ($P = 3.6\text{x}10^{-2}$) and integration of energy metabolism ($P = 2.0\text{x}10^{-2}$). The gene network image was generated using VisANT (edge weight cutoff > 0.1). The top five hub genes *Slc25a37, Slc29a1, Hltf, Card10*, and *Arsb* are enlarged.

Table 5.14 Cyan module hub gene connectivity

Genes	Description	Interactors	weight	k_{Within}	k_{Out}
Slc25a37	solute carrier family 25, member 37	111	0.134	55	450
Hltf	helicase-like transcription factor	109	0.141	44	661
Card10	caspase recruitment domain family, member 10	98	0.136	52	438
Arsb	arylsulfatase B	89	0.128	51	416
Slc29a1	solute carrier family 29 (nucleoside transporters), member 1	89	0.133	44	537
Slc27a3	solute carrier family 27 (fatty acid transporter), member 3	65	0.127	36	608
Fam198b	family with sequence similarity 198, member B	58	0.130	31	624
Malt1	mucosa associated lymphoid tissue lymphoma translocation gene 1	53	0.122	48	332
Tpd52l1	tumor protein D52-like 1	52	0.117	35	502
Prrg3	proline rich Gla (G-carboxyglutamic acid) 3 (transmembrane)	50	0.126	38	426
P2rx6	purinergic receptor P2X, ligand-gated ion channel, 6	49	0.123	51	259
Tceal3	transcription elongation factor A (SII)-like 3	49	0.119	46	351
2510002D24Rik	RIKEN cDNA 2510002 D24 gene	49	0.115	42	429
Rad9b	RAD9 homolog B	48	0.116	45	381
Aph1c	anterior pharynx defective 1c homolog	46	0.119	44	382

259

Gm7694	predicted gene 7694	44	0.119	47	353
Actr5	ARP5 actin-related protein 5 homolog	41	0.131	51	195
Ncoa4	nuclear receptor coactivator 4	40	0.121	46	286
Myo1c	myosin IC	39	0.120	50	292
C030023E24Rik	RIKEN cDNA C030023E24 gene	39	0.118	39	395
Synpo2	synaptopodin 2	35	0.129	28	87
Acer3	alkaline ceramidase 3	35	0.114	35	462
Nkx2-4	NK2 transcription factor related, locus 4	35	0.119	33	466
Txnip	thioredoxin interacting protein	32	0.129	46	195
Cdca3	cell division cycle associated 3	31	0.118	48	219

Table 5.14 shows the top 25 hub genes in the cyan module including their gene names, descriptions, the number of interactors within the module, the average connection weight, the connectivity within the module (k_{Within}), and the connectivity outside of the network (k_{Out}). The number of interactors only includes interactions with a connectivity strength > 0.1.

5.14 S40A is associated with gene networks involved in neuronal activity and excitotoxicity

Consistent with the differential expression analysis, the S40A-CREB condition was correlated with a synaptic activity module (cor = 0.47, P = 2.3x10^{-19}) that was enriched for genes associated with neurotransmitter release genes (P = 2.7x10^{-3}) and genes involved in transmission across chemical synapses (P = 2.7x10^{-3}) (Figures 5.26, 5.27). Interestingly, the synaptic activity module was anti-correlated with all other conditions including S133A-CREB (cor = -0.44, P = 4.8x10^{-17}) (Figure 5.27). This is consistent with a previous study in our lab that found that ablation of CREB glycosylation at serine 40 lead to increased transcription of neuronal activity genes.[3]
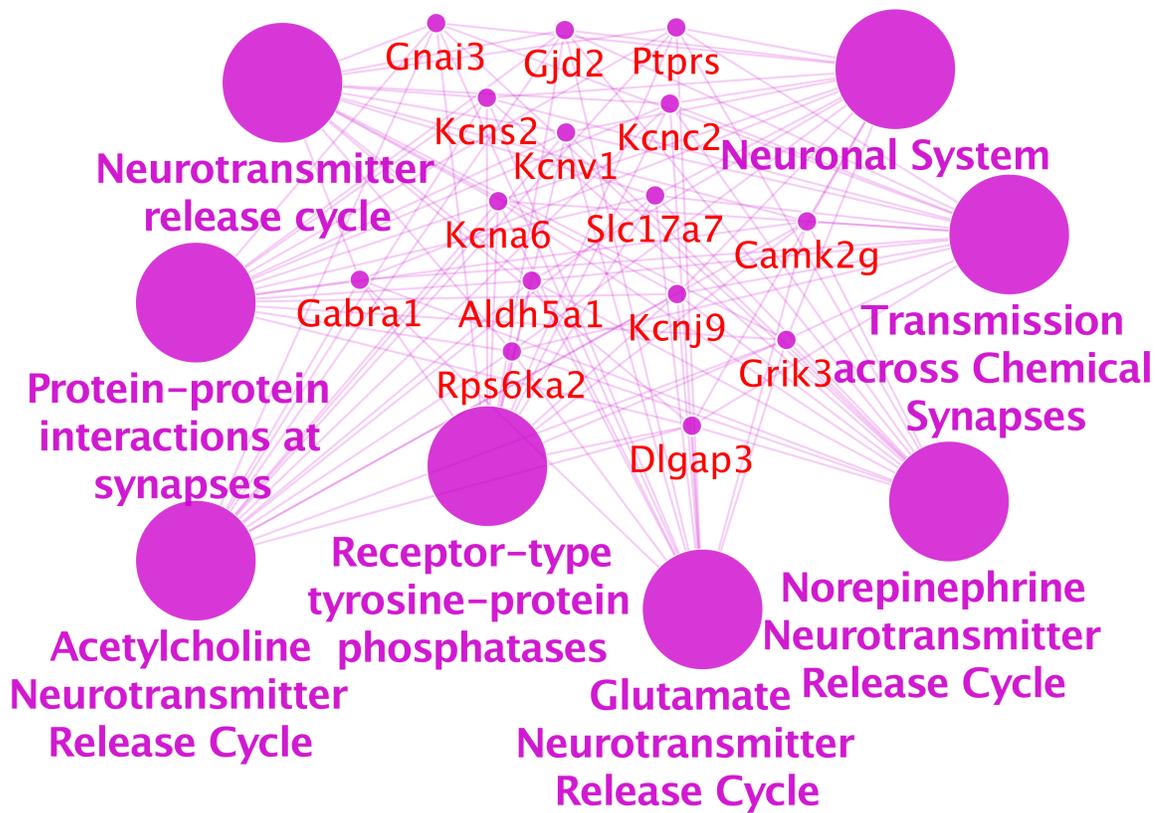
Figure 5.26 Gene ontology annotations for S40A-associated neuronal activity-related module. Shown here are the top gene ontology categories, which include neurotransmitter release cycle as well as synaptic activity. The neuronal activity module is enriched for genes involved in transcriptional activation ($P = 3.0\text{x}10^{-2}$) and the synapse ($P = 2.4\text{x}10^{-2}$). The genes that belong to and are shared by these gene ontology categories are also shown. The module was annotated using the Reactome pathways through the ClueGO extension in Cytoscape. The significance for all listed categories are Bonferroni corrected $P < 0.00027$.
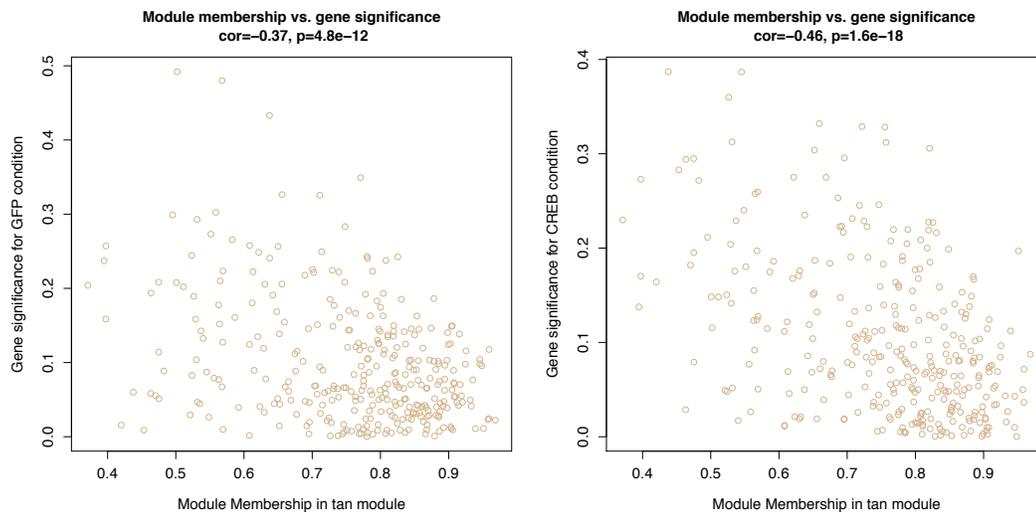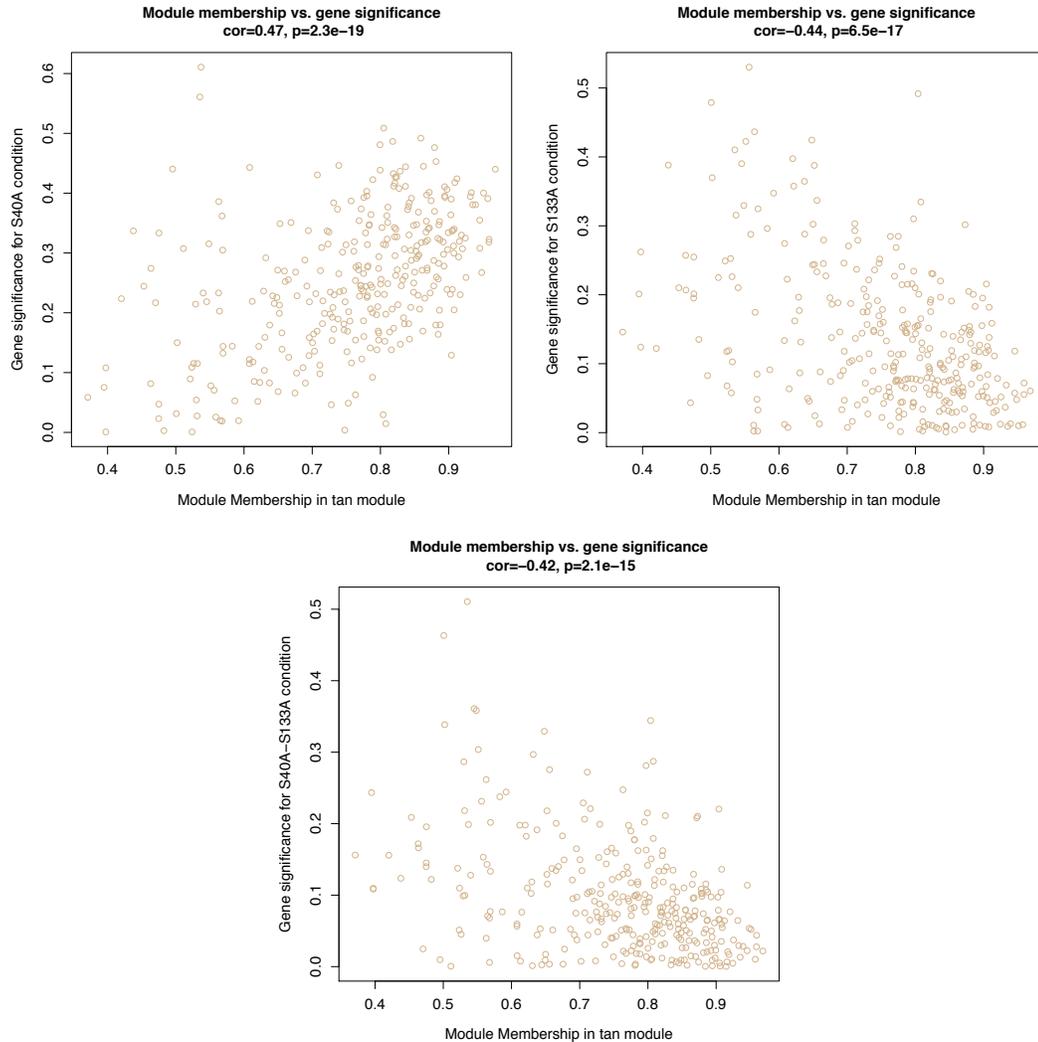
Figure 5.27 Synaptic activity-related module is correlated with S40A and anti-correlated with all other conditions. Module membership for the tan module is anti-correlated with GFP (cor = -0.37, P = 4.8×10^{-12}), WT CREB (cor = -0.46, P = 1.6×10^{-18}), S133A-CREB (cor = -0.44, P = 6.5×10^{-17}), S40A-S133A-CREB (cor = -0.42, P = 2.1×10^{-15}) and correlated with S40A-CREB (cor = 0.47, P = 2.3×10^{-19}).

The hub genes for the tan module included G protein-activated inward rectifier potassium channel 3 (Girk3, *Kcnj9*), pre-B-cell leukemia transcription factor (PBX1, *Pbx1*), SRY-box containing gene 11 (SOX11, *Sox11*), RNA binding motif protein X chromosome (RBMX, *Rbmx*), and oxysterol binding protein 2 (OSBP2, *Osbp2*) (Figure 5.28, Table 5.15). Girk3 is an inhibitory potassium channel that is important for mediating the excitability of neurons and has been shown to regulate dopaminergic

addiction pathways.[72,73] PBX1 is a transcription factor that has been shown to be important for neuronal patterning throughout the cortex, and disruption of PBX1 signaling has been implicated in Parkinson's disease.[74,75] SOX11 is another transcription factor responsible for neuronal development and differentiation.[76] RBMX is an RNA-binding protein that mediates alternative splicing and transcription and has been shown to necessary during neural development.[77-79] OSBP2 links sterol levels to sphingolipid metabolism and can regulate mitogenic signaling.[80] Together, these hub genes are poised to be critical regulators of the synaptic activity-related tan module.

Figure 5.28 The tan module is enriched for synaptic activity-related genes. The tan module gene network is enriched for neurotransmitter release genes ($P = 2.7\text{x}10^{-3}$) and genes involved in transmission across chemical synapses ($P = 2.7\text{x}10^{-3}$). The gene network image was generated using VisANT (weight cutoff > 0.1). The top five hub genes *Kcnj9, Pbx1, Sox11, Rbmx,* and *Osbp2* are enlarged.

Table 5.15 Tan module hub gene connectivity

Genes	Description	Interactors	weight	k_{Within}	k_{Out}
Kcnj9	potassium inwardly-rectifying channel, subfamily J, member 9	182	0.164	64	1135
Pbx1	preB-cell leukemia transcription factor 1	180	0.149	70	824
Sox11	SRY-box containing gene 11	174	0.166	59	1166
Rbmx	RNA binding motif protein, X chromosome	168	0.165	56	1171
Osbp2	oxysterol binding protein 2	164	0.147	53	1010

Dnajc27	DnaJ (Hsp40) homolog, subfamily C, member 27	146	0.131	67	740
Rfxank	regulatory factor X-associated ankyrin-containing protein	144	0.130	74	683
Mul1	mitochondrial ubiquitin ligase activator of NF-κB 1	129	0.146	51	982
Cds1	CDP-diacylglycerol synthase 1	126	0.129	67	688
Lsm11	U7 snRNP-specific Sm-like protein LSM11	124	0.137	49	967
Lphn2	latrophilin 2	119	0.135	57	711
Snph	syntaphilin	118	0.160	43	1237
Gpbp1	GC-rich promoter binding protein 1	116	0.153	41	1094
2900026A02Rik	RIKEN cDNA 2900026A02 gene	114	0.129	59	775
Smarcc1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily c, member 1	112	0.128	66	574
Ift27	intraflagellar transport 27 homolog	108	0.136	43	1007
Ssx2ip	synovial sarcoma, X break point 2 interacting protein	108	0.132	44	966
Tes	testis derived transcript	108	0.128	56	652
Lamb2	laminin, beta 2	107	0.132	55	736
Dgka	diacylglycerol kinase, alpha	105	0.124	68	558
Recql5	RecQ protein-like 5	98	0.135	48	890
Ly6g5b	lymphocyte antigen 6 complex, locus G5B	92	0.119	56	599
Carnmt1	carnosine *N*-methyltransferase 1	89	0.125	61	648
Slc17a7	solute carrier family 17 (sodium-dependent inorganic phosphate cotransporter), member7	89	0.128	47	818
Gm996	predicted gene 996	83	0.127	53	768

Table 5.15 shows the top 25 hub genes in the tan module including their gene names, descriptions, the number of interactors within the module, the average connection weight, the connectivity within the module (k_{Within}), and the connectivity outside of the network (k_{Out}). The number of interactors only includes interactions with a connectivity strength > 0.1.

Intriguingly, the S40A-CREB condition was correlated with a neuronal excitotoxicity module (cor = 0.46, P = 1.8x10^{-19}) that was enriched for genes involved in neurodegeneration (Alzheimer's (P = 6.5x10^{-8}), Parkinson's (P = 1.2x10^{-7}) and Huntington's diseases (P = 1.4x10^{-7})), mitochondrial activity (P = 1.2x10^{-7}), and in ion channels (P = 1.0x10^{-4}) (Figures 5.29 and 5.30).[81] This green yellow module was anti-correlated with all other conditions (cor = -0.27 to -0.32, P < 9.8x10^{-7}) (Figure 5.30). Our results suggest that glycosylation at serine 40 acts as a deterrent against aberrant neuronal

activity and excitotoxicity. The major hub genes for this module regulate synaptic activity [cholinergic receptor muscarinic 3 (*Chrm3*), calcium voltage-gated channel auxiliary subunit γ 2 (*Cacng2*), and γ-aminobutyric acid type A receptor δ subunit (*Gabrd*)] and cell growth and survival [nephronectin (*Npnt*) and inhibitor of DNA binding, HLH protein (*Id1*)] (Figure 5.31, Table 5.16). The WGCNA results support and extend the DE gene data by uncovering the association of S40A with synaptic activity and excitotoxicity modules.

Figure 5.29 Gene ontology annotations for the neuronal activity- and excitotoxicity-related module correlated with S40A. Shown here are the top gene ontology categories, which include oxidative phosphorylation and neurodegenerative diseases in addition to long-term potentiation and calcium

signaling. The green yellow module is enriched for genes involved in neurodegeneration (Alzheimer's ($P = 6.5 \times 10^{-8}$), Parkinson's ($P = 1.2 \times 10^{-7}$) and Huntington's diseases ($P = 1.4 \times 10^{-7}$)), mitochondrial activity ($P = 1.2 \times 10^{-7}$), and ion channels ($P = 1.0 \times 10^{-4}$). The genes that belong to and are shared by these gene ontology categories are listed. This excitotoxicity module was annotated using the KEGG pathways through the ClueGO extension in Cytoscape. The significance for all listed categories are Bonferroni step down corrected $P < 0.002$.

Module membership vs. gene significance
cor=−0.32, p=1.2e−09

(y-axis: Gene significance for S40A–S133A condition; x-axis: Module Membership in greenyellow module)
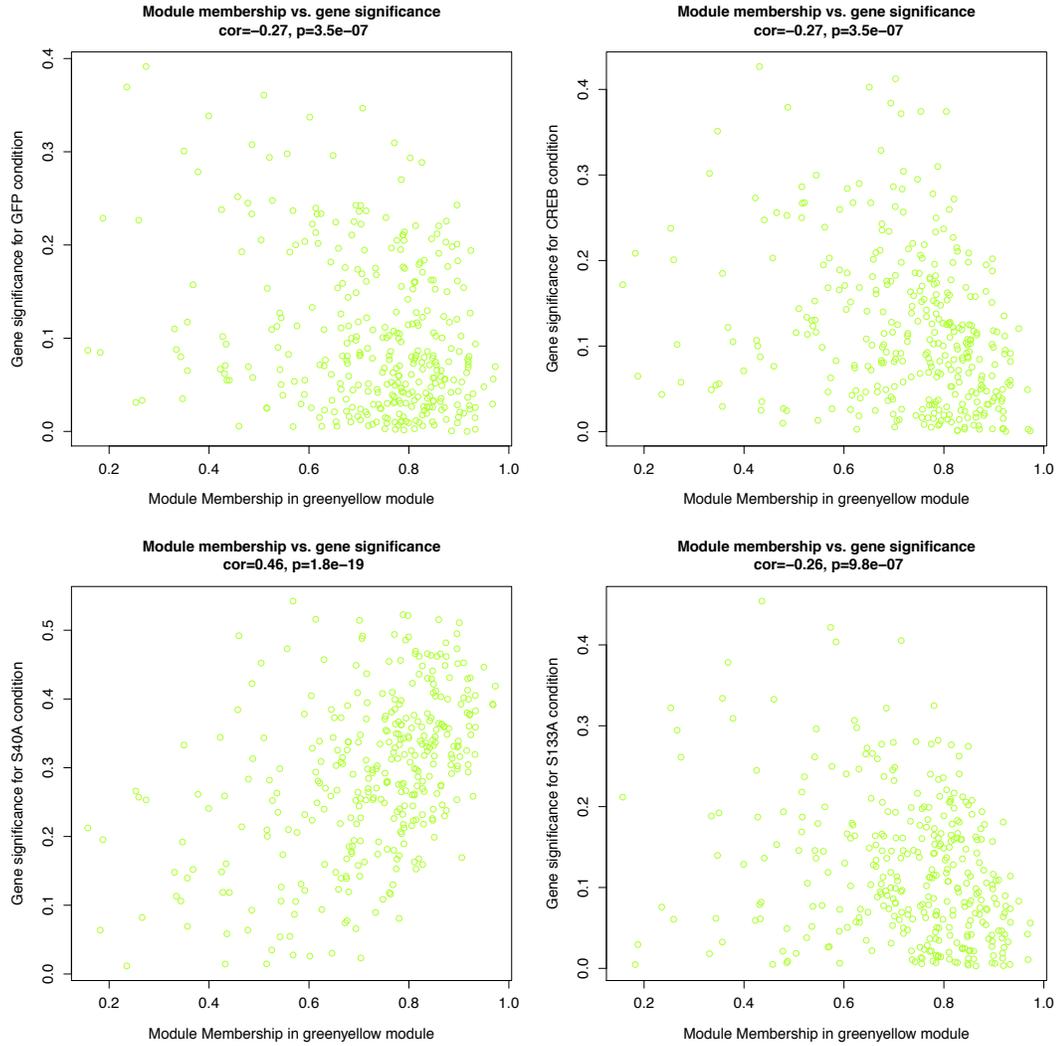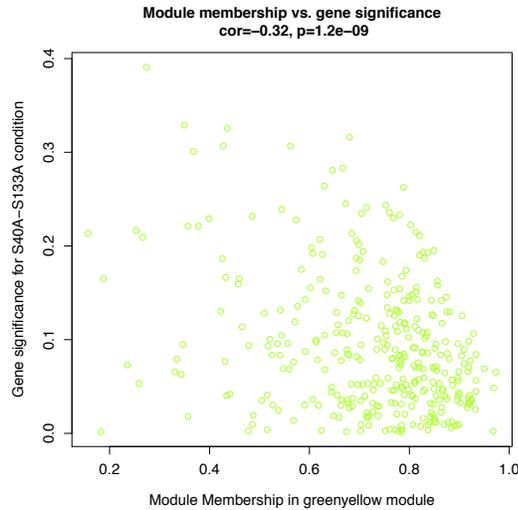
Figure 5.30 Green yellow module is positively correlated with S40A and anti-correlated with all other conditions. Module membership for the green yellow module is anti-correlated with GFP (cor = -0.27, P = 3.5×10^{-7}), WT CREB (cor = -0.27, P = 3.5×10^{-7}), S133A-CREB (cor = -0.26, P = 9.8×10^{-7}), S40A-S133A-CREB (cor = -0.32, P = 1.2×10^{-9}) and correlated with S40A-CREB (cor = 0.46, P = 1.8×10^{-19}).
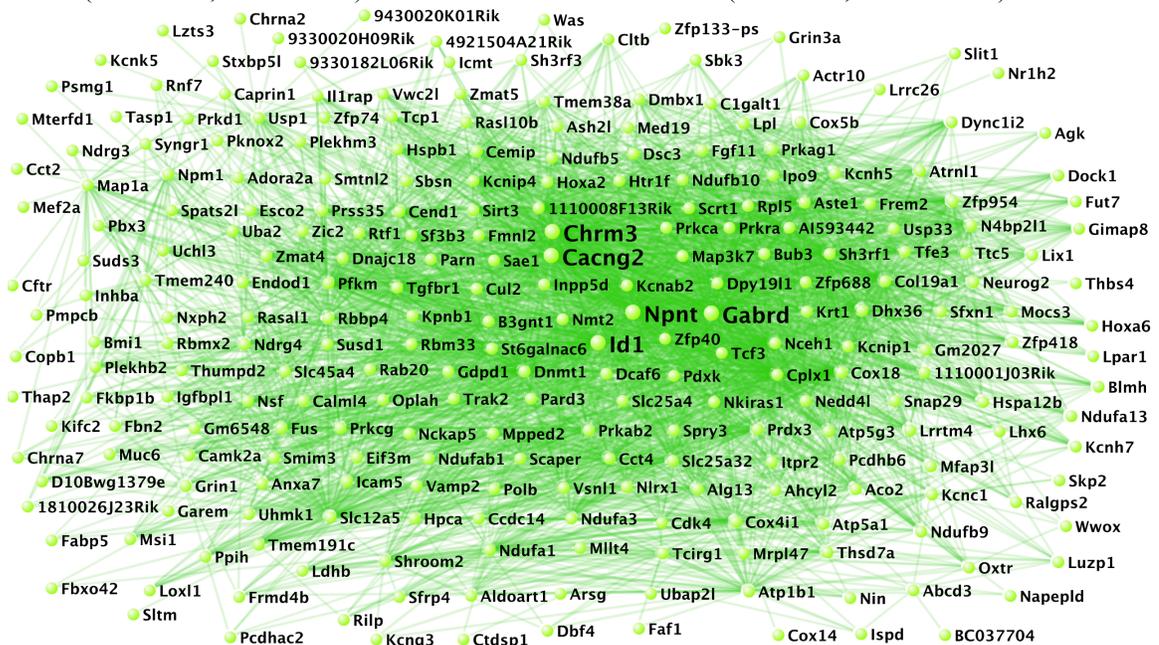
Figure 5.31 The green yellow module is enriched for neuronal activity and excitotoxicity genes. The tan module gene network is enriched for genes involved in neurodegeneration (Alzheimer's (P = 6.5×10^{-8}), Parkinson's (P = 1.2×10^{-7}) and Huntington's diseases (P = 1.4×10^{-7})), mitochondrial activity (P = 1.2×10^{-7}), and in ion channels (P = 1.0×10^{-4}). The gene network image was generated using VisANT (weight cutoff > 0.1). The top five hub genes *Chrm3, Cacng2, Gabrd, Npnt,* and *Id1* are enlarged.

Table 5.16 Green yellow module hub gene connectivity

Genes	Description	Interactors	weight	k_{Within}	k_{Out}
Cacng2	calcium channel, voltage-dependent, γ subunit 2	150	0.143	51	1050
Npnt	nephronectin	147	0.138	65	834

Gabrd	γ-aminobutyric acid (GABA) A receptor, subunit δ	146	0.149	51	1069
Id1	inhibitor of DNA binding 1	146	0.144	43	1104
Chrm3	cholinergic receptor, muscarinic 3, cardiac	143	0.136	46	1063
Kcnab2	potassium voltage-gated channel, shaker-related subfamily, β member 2	138	0.127	64	747
Cplx1	complexin 1	130	0.127	73	660
Cox18	COX18 cytochrome c oxidase assembly homolog	128	0.135	35	1089
Dnmt1	DNA methyltransferase (cytosine-5) 1	124	0.124	59	703
Dpy19l1	dpy-19-like 1	124	0.132	56	852
Pfkm	phosphofructokinase, muscle	121	0.126	69	707
Tgfbr1	transforming growth factor, β receptor I	121	0.126	60	762
Aste1	asteroid homolog 1	103	0.140	49	858
Slc12a5	solute carrier family 12, member 5	102	0.130	47	784
Dhx36	DEAH(Asp-Glu-Ala-His) box polypeptide 36	100	0.127	42	916
Col19a1	collagen, type XIX, α1	98	0.128	59	720
Nceh1	aryl acetamide deacetylase-like 1	98	0.124	56	698
Sh3rf1	SH3 domain containing ring finger 1	98	0.140	40	976
Rbbp4	retinoblastoma binding protein 4	96	0.129	43	877
1110008F13Rik	RIKEN cDNA 1110008F13 gene	94	0.134	33	1081
AI593442	expressed sequence AI593442	89	0.129	62	642
Bub3	budding uninhibited by benzimidazoles 3 homolog	82	0.124	45	769
St6galnac6	ST6 (α-*N*-acetyl-neuraminyl-2,3-β-galactosyl-1,3)-*N*-acetylgalactosaminide α -2,6-sialyltransferase 6	80	0.117	43	807
Cox4i1	cytochrome c oxidase subunit IV isoform 1	74	0.126	46	759
Cend1	cell cycle exit and neuronal differentiation 1	72	0.119	55	677

Table 5.16 shows the top 25 hub genes in the green yellow module including their gene names, descriptions, the number of interactors within the module, the average connection weight, the connectivity within the module (k_{Within}), and the connectivity outside of the network (k_{Out}). The number of interactors only includes interactions with a connectivity strength > 0.1.

5.15 Discussion

Overall, our results show that regulation of CREB through glycosylation and phosphorylation is critical for negotiating distinct transcriptional repertoires. In particular, we show that removal of glycosylation at S40 of CREB leads to the expression of neuronal excitability genes while abrogation of S133 phosphorylation affects the transcription of genes involved in neuronal development and differentiation. Minimal

overlap between the S40A and S133A conditions especially in the upregulated genes reveals that CREB glycosylation and phosphorylation mediate different sets of genes. Furthermore, we found that some of the genes differentially regulated by CREB phosphorylation and glycosylation were (1) known targets of CREB and its coactivators, CBP and p300 and (2) associated with euchromatic regions of the genome. Furthermore, OGT, *O*-GlcNAc, mSin3a, and Tet1 bind directly to the genes upregulated in the glycosylation-deficient CREB. Finally, we found that phosphorylation of CREB is important for mediating neuronal differentiation and several metabolic pathways while glycosylation of CREB is a central regulator of synaptic activity- and excitotoxicity-related gene networks. Together, CREB *O*-GlcNAcylation at serine 40 and phosphorylation at serine 133 are critical regulators of CREB transcription and mediators of neuronal activity and homeostasis.

Previous studies had implicated CREB phosphorylation at S133 as enhancing CREB transcription of specific genes through the enhancement of CREB's interaction with CBP and p300.[2,82] More recent studies have begun to explore the role of S133 phosphorylation on CREB-mediated transcription globally. In a microarray study, McClung and colleagues showed that S133A-CREB overexpression results in opposing effects on transcription when compared to WT CREB in the nucleus accumbens of mice with a total of 24 differentially-expressed genes between the S133A-CREB and WT CREB conditions.[28] We did not observe overlap between the microarray study; potential reasons for the differences could be (1) that the study involved overexpression of S133A-CREB and WT CREB in the nucleus accumbens over course of 8 weeks while our study used a Creb1$^{\alpha\delta}$ background and explored considerably shorter time periods (4 and 8

hours) in a E16.5 cortical neuronal population and (2) the differing sensitivity in detection for microarray studies when compared to RNA-Seq experiments, especially with low abundance transcripts.[32]

Another RNA-Seq study by Briand and coworkers showed that a phosphorylation-deficient S133A mutant mouse had no DE genes in the hippocampus when compared with WT mice.[23] In this study, breeding of the heterozygous S133A-CREB mice resulted in lower numbers than the expected Mendelian frequencies (only 11% homozygotes), leading the authors to suggest that while S133 phosphorylation may not affect transcription and memory-related behavior, it may affect development.[23] Prior studies have shown that CREB plays a vital role in neuronal development and differentiation.[2,33,34] Moreover, we found that phosphorylation of CREB modulates key metabolic pathways including fatty acid, mitochondrial, and amino acid metabolism as well as insulin secretion. CREB governs the metabolism of glucose, mitochondria, insulin, lipids, and fatty acids.[60] Furthermore, CREB phosphorylation at serine 133 increased the transcription of gluconeogenic genes.[61] Our study demonstrates a key role of CREB serine 133 phosphorylation in controlling metabolic gene expression and arbitrating neuronal development and differentiation.

Our work demonstrates that glycosylation of CREB, rather than just acting as a brake to overall CREB-mediated transcription, specifically mitigates the expression of neuronal activity and excitability genes. Removal of the glycosylation at serine 40 upregulated the expression of neuronal activity genes important for every part of the neuron from the dendrite to the axon. This is consistent with a previous study in our lab that found that ablation of CREB glycosylation at serine 40 led to increased transcription

of neuronal activity genes, enhanced neurite outgrowth, and accelerated memory formation when compared to WT CREB.[3] Elevated CREB activity through the expression of more active VP16, Y134F, and DIEDML CREB mutants led to heightened transcription of neuronal excitability genes.[16,17] We show that glycosylation of CREB at serine 40 is critical for fine-tuning the transcription of neuronal activity and excitability genes.

In addition, our S40A-CREB displayed key transcriptional similarities and differences with the constitutively active CREB mutant, VP16-CREB, which enhances the association of CREB with its coactivators, CBP and p300; similar to VP16-CREB, S40A-CREB upregulated genes associated with synaptic activity, but unlike VP16-CREB, S40A-CREB also increased neuronal excitation gene expression (Table 5.11). Interestingly, these neuronal excitability genes were known targets of CRTC1 in drosophila in response to learning (Table 5.11).[37] In contrast, the similarly regulated genes in S40A-CREB and VP16-CREB were not known targets of CRTC1, but instead were targets of CBP. This suggests that CREB glycosylation at serine 40 specifically affects genes directly regulated by CRTC1, consistent with our previous study showing that S40A-CREB enhanced binding to CRTC.[3] Indeed, recent studies have shown that CRTC is more critical for long-term memory formation and neuronal activity, whereas CBP may be important for the expression of genes early on in memory formation.[37,83-85]

Finally, our WGCNA showed that ablation of CREB serine 40 glycosylation correlated with excitotoxicity gene expression. Our results suggest that glycosylation at serine 40 acts as a significant deterrent against aberrant neuronal activity and excitotoxicity. This excitotoxicity association was also seen in a constitutively active

CREB mutant, VP16-CREB; mice expressing VP16-CREB across the brain displayed increased expression of immune response genes and impaired spatial memory retrieval despite exhibiting enhanced LTP.[2,16,35] Indeed, neurons with higher CREB activation appear to be more prone to excitotoxic effects.[86] Consistent with this finding, researchers discovered elevated CRE-mediated transcription in Huntington's disease model mice suggesting that elevated CREB transcription may lead to an excitotoxicity-induced neurodegenerative phenotype.[87] In contrast, Alzheimer's disease, Rubenstein-Taybi syndrome, and Coffin-Lowry syndrome are associated with impaired CREB-mediated transcription.[88]

These studies highlight the necessity of a judicious PTM rheostat to avoid neuronal apoptosis through CREB under-activation or excitotoxicity-induced neurodegeneration through CREB over-activation. Through a global analysis of the gene networks regulated by phosphorylation, glycosylation, and both, we have begun to decipher the PTM transcription code of CREB that ordains neuronal fate. Control of CREB activity through O-GlcNAcylation at serine 40 attenuates neuronal activity and excitotoxicity while phosphorylation at serine 133 moderates neuronal development. Together, CREB O-GlcNAcylation at serine 40 and phosphorylation at serine 133 are critical for the maintenance of neuronal homeostasis and could therefore be therapeutic targets for neurodegenerative diseases.

5.16 Methods

5.16.1 Breeding and genotyping Creb1$^{\alpha_\delta}$ mice

All animal procedures were performed in accordance with Institutional Animal Care and Use Committee (IACUC) guidelines of the California Institute of Technology.

CREB (α and δ) knock-out mice heterozygous mice (129S2/SvPasCrl background and C57BL/6 background) were group-housed with ad libitum access to food and water. Mice were bred and genotyped as previously described.[89,90] The mice were maintained with outbreeding with their respective genetic backgrounds- either 129/SVJ mice or C57BL/6 mice.[91,92] The heterozygotic 129S2/SvPasCrl CREB+/- (129.Creb1$^{\alpha_\delta+/-}$) mice were mated with the heterozygotic C57BL/6 CREB+/- (B6.Creb1$^{\alpha_\delta+/-}$) mice in order to yield homozygous, heterozygous, and wild-type embryos (Figure 5.32). We encountered some pregnancy detection and breeding issues that were rectified by breeding B6.Creb1$^{\alpha_\delta+/-}$ females with 129.Creb1$^{\alpha_\delta+/-}$ males and extra maternal support (lower rack placement and cage rearrangement to allow mothers to descend from upper play area easier) respectively.

♂   X   ♀

B6.Creb1$^{\alpha\delta+/+}$ B6.Creb1$^{\alpha\delta+/-}$
WT   HET

B6.Creb1$^{\alpha\delta+/+}$ X   B6.Creb1$^{\alpha\delta+/-}$
WT ~50%   HET ~50%

♂   X   ♀

129.Creb1$^{\alpha\delta+/-}$ 129.Creb1$^{\alpha\delta+/-}$
WT   HET

129.Creb1$^{\alpha\delta+/-}$ X   129.Creb1$^{\alpha\delta+/+}$
HET ~50%   WT ~50%

♀   X   ♂

129.B6.Creb1$^{\alpha\delta+/+}$ 129.B6.Creb1$^{\alpha\delta+/-}$ 129.B6.Creb1$^{\alpha\delta-/-}$
WT ~35%   HET ~50%   HOMO ~15%

273
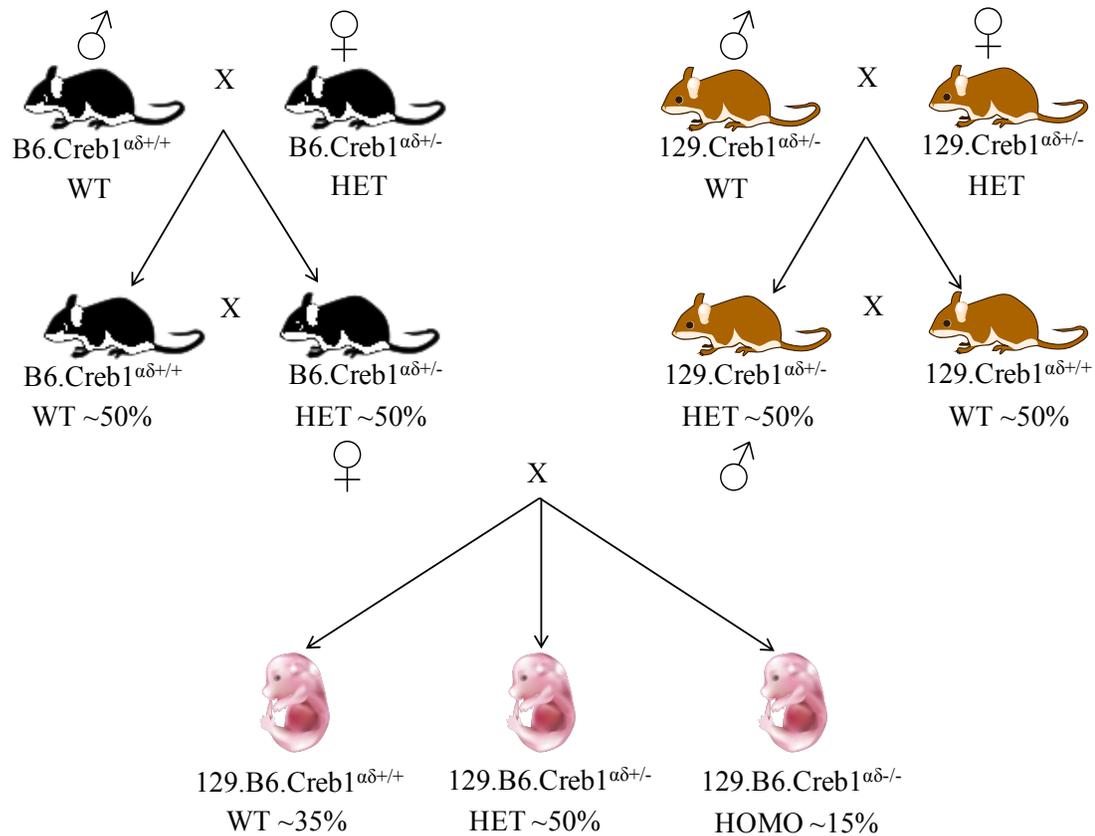
Figure 5.32 Breeding scheme for Creb1^{α_s} homozygous knockout mice. Shown here is the overall schematic for breeding the C57BL/6 Creb1^{α_s} heterozygous mice (denoted $\text{B6.Creb1}^{\alpha_s+/-}$) and 129S2/SvPasCrl Creb1^{α_s} (denoted $\text{129.Creb1}^{\alpha_s+/-}$) in order to obtain $\text{129.B6.Creb1}^{\alpha_s-/-}$ embryos. Mouse image from ChemDraw Professional 16.0 and mouse embryo image from the DataBase Center for Life Sciences (DBCLS).

Only homozygous neuronal cultures were used for subsequent experiments. Genotyping was performed by PCR as described by previous studies.[91] The DNA samples were isolated from tail tips using the standard procedure from the DNeasy Blood & Tissue Kits (Qiagen, 69504). Then, PCR amplification was performed on 50 ng of template genomic DNA using the Q5 Hot Start High-Fidelity DNA polymerase (New England BioLabs, Inc.) [98°C for 30 sec; [98°C for 10 sec; 51.8°C for 20 sec; 72°C for 30 sec (34 cycles)], and then 72°C for 2 min] to identify the wild-type (150 bp band), heterozygous (150 + 350 bp band), and homozygous (350 bp) using DNA agarose electrophoresis (1.5% agarose gel in TAE buffer). The primers used were as follows: oIMR3081 (Mutant): 5'-TGATGGATACTTTCTCGGCA-3', oIMR3082 (Common): 5'-ATGTATTTTTATACCTGGGC-3'; and oIMR3083 (Wild type): 5'-TATTGTAGGTAACTAAATGA-3'. We used Laragen's genotyping services to corroborate our genotyping results.

5.16.2 Creb1^{α_s} E16.5 cortical dissections

Mouse cortical neurons were prepared as described previously [5]. Briefly, we dissected E16.5 neurons by separating embryos and then using 1X TrypLE (ThermoFisher, 12605028) to trypsinize individual cortices. Following 20 minutes of trypsinization, 1X Defined Trypsin Inhibitor (ThermoFisher, R007100) was added to the neurons to quench trypsinization. Next, DMEM, high glucose, GlutaMAX Supplement (ThermoFisher, 10566024) with 10% fetal bovine serum, heat-inactivated

(ThermoFisher, 10082147) was added to the mixture followed by trituration using a flame-tipped glass pipette with washing in clean 1X HBSS, no calcium, no magnesium, phenol red (ThermoFisher, 14170161). Then, neurons were plated at approximately $2x10^6$ cells in 6 cm dishes. The dishes were coated for at least 2 hours in poly-D-lysine hydrobromide MW 70,000-150,000 (ThermoFisher, P6407-5MG) in 0.2 μm filter-sterilized PBS (10 mM PO_4, 137 mM NaCl, 2.7 mM KCl), pH 7.4 at 37°C and 5% CO_2 and then washed twice with sterile-filtered water and allowed to dry. The neurons were cultured in 5 ml of neuronal medium (Neurobasal Medium (ThermoFisher, 21103049) supplemented with GS21 Neural Supplement (Amsbio, GSM-3100), Penicillin-Streptomycin (ThermoFisher, 15070063), and GlutaMAX Supplement (ThermoFisher, 35050061)). Two hours after plating all the media was removed and replaced with warmed media. After 4 days *in vitro* (DIV), half of the media was replaced with fresh media. Following 6DIV, half of the media was replaced with fresh media supplemented with a final concentration of 1 μM tetrodotoxin (TTX) (Tocris Biosciences, 1078) and 100 μM D-(-)-2-amino-5-phosphonopentanoic acid (D-AP5) (Tocris Biosciences, 0106-1mg). At 12-16 hours following silencing, we added HSV for 4 hours or 8 hours and then depolarized neurons for 2 hours using 55 mM KCl. For the axonal and dendritic growth assays, neurons were plated and at 1DIV, HSV was added to the neurons. At 3DIV, the neurons were fixed and then prepared for ICC as described below.

5.16.3 Herpes simplex virus (HSV) transduction and immunocytochemistry (ICC)

We cloned FLAG-tagged rat Creb1 gene into the Gateway pENTR 1A entry dual selection vector (ThermoFisher, A10462). Then, we used the Q5 Site-Directed Mutagenesis Kit (New England BioLabs, E0554S) in order to mutate serine 40 and/or

serine 133 to alanines. The MIT Viral Gene Transfer Core (Dr. R. Neve) then recombined the various CREB mutants into the Gateway HSV-Syn -IRES-GFP construct (Syn refers to the synapsin promoter) and packaged this in high titer herpes simplex virus (HSV). Viruses were diluted in sterile 25 mM HEPES, pH 7.3 (ThermoFisher, 1530080) and stored at $-80^{\circ}C$ until addition to cells. The viral titer was tested initially in primary cortical neuronal cultures using immunocytochemistry (ICC) until ~100% transduction was achieved as measured by GFP readout. For the RNA-Seq experiments, the viral levels were titered to ensure that the amount of Creb1 expression was controlled across samples through qPCR (see qPCR section in Chapter 2.14.8).

For ICC experiments, 15mm coverslips were coated as described above with poly-D-lysine hydrobromide in 12-well plates. Then, E16.5 cortical neurons were seeded at 100,000 neurons/well in 1 ml of neuronal medium. On 5DIV (or 3DIV for the neuronal growth assays), the neurons were treated with HSV-Syn-IRES-GFP (GFP), HSV-Syn-CREB-IRES-GFP (WT), HSV-Syn-S40A-CREB-IRES-GFP (S40A), HSV-Syn-S133A-CREB-IRES-GFP (S133A), or HSV-Syn-S40A-S133A-CREB-IRES-GFP (S40A-S133A) for the indicated amount of time (2, 4, 6, 8, 10, 12, or 48 hours). After HSV treatment, the media was removed and the cells were rinsed once with ice cold PBS and then fixed with 4% paraformaldehyde for 20 min. Following fixation, the coverslips were rinsed 3X with PBS then permeabilized and blocked with 0.3% Triton X-100 in 10% normal goat serum (NGS) (ThermoFisher, 16310064) in PBS for 30 min. Next, the cells were incubated for 2 hours at room temperature in the following primary antibodies: 1:200 anti-M2 FLAG (Sigma, F1804-1MG), 1:200 anti-Tau-1 (Millipore, clone PC1C6, MAB3420), and 1:1,000 anti-microtubule-associated protein 2 (MAP2) (Millipore,

AB5622) in dilution buffer (2% NGS in PBS). The coverslips were rinsed 3X with PBS and incubated in the following secondary antibodies: 1:1,000 Goat anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 546, ThermoFisher, A-11030 or Goat anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 405, ThermoFisher, A-31556 in PBS for 2 hours at room temperature and rinsed 3X in PBS. Finally, the coverslips were mounted on slides in Vectashield mounting medium with DAPI (Vector Labs, H-1200) or without DAPI (Vector Labs, H-1000). The slides were visualized using a Zeiss LSM700 confocal microscope with 20X magnification with and without 5x5 tiling.

5.16.4 RNA extraction, qPCR, and RNA-Seq

The neuronal cultures were lysed and total RNA was extracted using an RNeasy Plus Mini Kit per the manufacturer's instructions (Qiagen, 74134). Following RNA extraction, qPCR analysis was performed as described in Chapter 2 with primers listed in Appendix III. For RNA-Seq, the total RNA samples were processed by Igor Antoshechkin and Vijaya Kumar at the Millard and Muriel Jacobs Genomics Laboratory (Caltech) using the mRNA-Seq Sample Preparation Kit following the manufacturer's instructions (Illumina). Briefly, the samples were enriched for poly-A mRNA using Sera-Mag magnetic Oligo(dT) beads and then subjected to divalent cation-catalyzed fragmentation to an average of ~300-350 bp. The enriched mRNA was converted to double-stranded cDNA through reverse transcription using random primers and SuperScript™ II Reverse Transcriptase (ThermoFisher) followed by degradation of mRNA and synthesis of the second strand of cDNA using DNA polymerase I. Then, cDNA purification, end repair, adaptor ligation, DNA purification, and several rounds of

PCR amplification were performed sequentially. Finally, the amplified cDNA was run on a short read the Illumina HiSeq2500. Sample quality and library validation were assessed using the Agilent 2100 Bioanalyzer and Nanodrop ND-1000 spectrophotometer (ThermoFisher).

Data quality processing and analysis were performed on the usegalaxy.org interface.[7] First, we took the raw fastq files and preprocessed by FastQC and low quality data was trimmed using FASTQ Trimmer. Following quality control preprocessing, the data was aligned to the mouse genome (*mm10*) using the TopHat and Cufflinks with bias correction.[93] The TopHat mapping results are displayed in Table 5.17.

Table 5.17 Summary of TopHat alignments to mm10 genome using Galaxy

Condition	Rep.	Input	Mapped	% input	MA	% input	>20 MA
GFP_4h	1	34195901	26456732	77.4%	3255290	12.3%	38
GFP_4h	2	27460980	24016164	87.5%	2959035	12.3%	2106
WT_4h	1	33535091	27341653	81.5%	3417555	12.5%	22
WT_4h	2	26656082	23080495	86.6%	3032739	13.1%	50
S40A_4h	1	36484023	29348405	80.4%	3582952	12.2%	36
S40A_4h	2	26409455	22548830	85.4%	2876216	12.8%	38
GFP_8h	1	42732370	34177292	80.0%	4494319	13.2%	69
GFP_8h	2	47061738	35357771	75.1%	4301622	12.2%	87
WT_8h	1	36321945	29029453	79.9%	3648703	12.6%	42
WT_8h	2	29035772	20981169	72.3%	2629444	12.5%	42
S40A_8h	1	44277390	34987651	79.0%	4275386	12.2%	77
S40A_8h	2	37784130	27882309	73.8%	3478354	12.5%	54
GFP_8h	1	33562997	23993641	71.5%	2949809	12.3%	40
GFP_8h	2	31777703	21733281	68.4%	2933432	13.5%	38
WT_8h	1	34856413	27226805	78.1%	3217959	11.8%	19
WT_8h	2	36621109	21798551	59.5%	2941624	13.5%	35
S40A_8h	1	35949409	30959749	86.1%	3668572	11.8%	45
S40A_8h	2	33798531	24387908	72.2%	3317849	13.6%	31
S133A_8h	1	34109018	29721320	87.1%	3514517	11.8%	34
S133A_8h	2	36072611	27299133	75.7%	3541779	13.0%	38
S40A-S133A_8h	1	33474368	28320199	84.6%	3300468	11.7%	22
S40A-S133A_8h	2	30656808	22663665	73.9%	2961173	13.1%	24

Table 5.17 shows an overview of the alignments of all the different RNA-Seq samples using the TopHat program in Galaxy. The first column denotes the HSV treatment (GFP, WT, S40A, S133A, S40A-S133A) along with the length of time (4 hours or 8 hours of expression). The second column shows the replicate number. The third and fourth columns show the number of input reads and the number of reads mapped reads after TopHat mapping to the mm10 genome respectively. The % of input represents the percent of the input reads that were mapped uniquely by TopHat. The next Multiple Alignments (MA) and % of input columns represent the total number of reads and percentage of reads that did not map uniquely to the mm10 mouse genome. The final column shows the number of non-unique alignments that had over 20 multiple alignments.

Next, the transcript assemblies were merged using Cuffmerge. The differentially expressed genes were determined using Cuffdiff with a cutoff of $\alpha = 0.1$. The differential expression results were then indexed and visualized using CummeRbund (version 2.16.0) and R package version 3.3.2. Venn diagrams of differentially expressed genes were generated using the VennDiagram R program (version 1.6.17) or BioVenn.[94,95] Reported false discovery rates for the RNA-Seq from this study were calculated q-values.[96] Gene ontology analysis is described in the "WGCNA and Gene Ontology Analysis" section.

5.16.5 ChIP-Seq, RNA-Seq, and microarray comparative analysis

Identifying the frequency of CRE sites in promoters

We used cruzdb in Python (version 2.7) in order to find the number of half and full CRE sites (TGACG/CGTCA or TGACGTCA) present 5000 bases upstream and 500 bases downstream from the transcription start site of all the differentially-expressed genes of interest.[97] The regions were mapped onto the *mm9* reference genome.

Identifying ChIP-Seq peaks

For several ChIP-Seq data sets, a list of chromosomes and start and stop sites for peaks were provided in Excel format instead of a list of nearest genes. In order to annotate the peaks with the nearest gene, we used cruzdb in Python (version 2.7) to identify the nearest gene such that the peak is upstream of the gene of interest transcription start site in the correct orientation of the gene.[97] We used this procedure to annotate peaks for the

following ChIP-Seq data sets: CREB;[5] p300;[26] H3K4me3, mSin3a, TetC, TetN, OGT;[42] H3K4me3, *O*-GlcNAc, Tet2.[46]

For the histone modification data set from Sandberg and colleagues, the peak data was obtained from GEO database as ChIP vs. Input peaks with detection $p < 0.00001$ (GSM2281997-GSM2282000).[38] The data was converted to the BED format and then aligned to the *mm9* mouse RefSeq gene coordinates using Galaxy.[7] This procedure was used for the H3K27me3, H3K4me1, H3K27Ac, and H3K4me3 ChIP-Seq data from Sandberg and colleagues.[38] Finally, for the CREB ChIP-Seq data set originally generated by Kim and colleagues, we obtained the reanalyzed peak data set from Lesiak and coworkers with FDR < 0.001.[5,24] Then, we used cruzdb in Python to annotate the new peaks as described above.

Microarray annotation and other RNA-Seq data set comparison

Several microarray data sets reported the data sets with microarray probe IDs. These data sets were annotated in R using the correct Affymetrix array annotation programs. For the McClung and Nestler *Nature Neuroscience* (2003) paper and Barco, *et al Neuron* (2005), we used the mgu74av2.db (version 3.2.3) and mgu74a.db (version 3.2.3) respectively.[16,28,98,99] For the Qiu, *et al eLife* (2016) RNA-Seq data set comparison, we used a cutoff of alpha=0.1 for differentially-expressed genes (KCl depolarized/ not depolarized) from the mouse DIV4 or DIV10 data sets.[100]

For the following data sets, we used the published annotated data without any pre-processing:

• H3K4me1, CBP, Npas4, Pol2, and SRF ChIP-Seq; RNA-Seq (KCl/no KCl)[5]

• 5hmC and H3K27me3 ChIP-Seq; RNA-Seq (NPC/neuron)[47]

280

- CREB ChIP-Seq[22]

- H3K27Ac ChIP-Seq[44]

- Tet2, OGT, and H2B-S112-*O*-GlcNAc ChIP-Seq[45]

- Tet3 ChIP-Seq[56]

- H3K4me3, H3K27me3 ChIP-Seq[43]

Program code for analysis is available upon request.

5.16.6 WGCNA and gene ontology analysis

Prior to performing WGCNA, we took the TopHat output from Galaxy and processed the data using HTSeq from the samtools program in order to determine the total counts for all the different samples.[101] This entailed sorting and indexing the BAM files and finally creating SAM files in command line for downstream analysis.[102] Next, the sorted SAM files were processed by HTSeq's counting program in order to get the final count for all the genes. Finally, we performed a data normalization using the variance stabilization transformation from DESeq2 (version 1.14.1) in R.[102-104]

Using the preprocessed RNA-Seq data, we performed WGCNA (version 1.51) in R on all detected genes using the protocols previously described.[59,105] Following hierarchical clustering and module assignment, gene ontology enrichment analysis was performed using the Bioconductor R packages AnnotationDbi (version 1.36.2), GO.db (version 3.4.0), and org.Mm.eg.db (version 3.4.0).[106-108] In addition, the Database for Visualization and Integrated Discovery (DAVID), PANTHER, and Enrichr as previously described.[109-111] The package Cytoscape version 3.5.1 was used to visualize the differentially-expressed gene ontology annotations using ClueGO version 2.3.3 and CluePedia version 1.3.3.[112-114] The package VisANT version 5.0 (weight cutoff of 0.1

(cor > 0.1)), Cytoscape, ClueGO, and CluePedia was used to visualize the WGCNA gene networks.[112-115] The top "hub" genes were identified as the genes with the most intramodular interactors with an edge weight cutoff of 0.1. The total number of interactors is listed in the Tables 5.13-5.15 above (all hub genes had at least 30 interactors).

5.17 References

1. 1 Khidekel, N. & Hsieh-Wilson, L.C. A 'molecular switchboard'--covalent modifications to proteins and their impact on transcription. *Org. Biomol. Chem.* (2004) **2**, 1:1-7.
2. 2 Sakamoto, K., Karelina, K. & Obrietan, K. CREB: a multifaceted regulator of neuronal plasticity and protection. *J. Neurochem.* (2011) **116**, 1:1-9.
3. 3 Rexach, J.E. *et al.* Dynamic O-GlcNAc modification regulates CREB-mediated gene expression and memory formation. *Nat. Chem. Biol.* (2012) **8**, 3:253-261.
4. 4 Blendy, J.A., Kaestner, K.H., Schmid, W., Gass, P. & Schutz, G. Targeting of the CREB gene leads to up-regulation of a novel CREB mRNA isoform. *EMBO J.* (1996) **15**, 5:1098-1106.
5. 5 Kim, T.K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* (2010) **465**, 7295:182-187.
6. 6 Trapnell, C. *et al.* Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotech* (2013) **31**, 1:46-53.
7. 7 Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* (2016) **44**, W1:W3-W10.
8. 8 Theisen, U., Straube, E. & Straube, A. Directional persistence of migrating cells requires Kif1C-mediated stabilization of trailing adhesions. *Dev. Cell* (2012) **23**, 6:1153-1166.
9. 9 Meng, J. *et al.* Myrf ER-Bound Transcription Factors Drive C. elegans Synaptic Plasticity via Cleavage-Dependent Nuclear Translocation. *Dev. Cell* (2017) **41**, 2:180-194 e187.
10. 10 Zheng, X. *et al.* CNOT3-Dependent mRNA Deadenylation Safeguards the Pluripotent State. *Stem Cell Reports* (2016) **7**, 5:897-910.
11. 11 Kuo, T.F. *et al.* Protein disulfide isomerase a4 acts as a novel regulator of cancer growth through the procaspase pathway. *Oncogene* (2017).
12. 12 Jensen, L.E. & Whitehead, A.S. IRAK1b, a novel alternative splice variant of interleukin-1 receptor-associated kinase (IRAK), mediates interleukin-1 signaling and has prolonged stability. *J. Biol. Chem.* (2001) **276**, 31:29037-29044.
13. 13 Diaz-Perales, A. *et al.* Identification of human aminopeptidase O, a novel metalloprotease with structural similarity to aminopeptidase B and leukotriene A4 hydrolase. *J. Biol. Chem.* (2005) **280**, 14:14310-14317.
14. 14 Simms, B.A. & Zamponi, G.W. Neuronal voltage-gated calcium channels: structure, function, and dysfunction. *Neuron* (2014) **82**, 1:24-45.
15. 15 Flavell, S.W. & Greenberg, M.E. Signaling mechanisms linking neuronal activity to gene expression and plasticity of the nervous system. *Annu. Rev. Neurosci.* (2008) **31**, 1:563-590.
16. 16 Barco, A. *et al.* Gene expression profiling of facilitated L-LTP in VP16-CREB mice reveals that BDNF is critical for the maintenance of LTP and its synaptic capture. *Neuron* (2005) **48**, 1:123-137.
17. 17 Suzuki, A. *et al.* Upregulation of CREB-mediated transcription enhances both short- and long-term memory. *J. Neurosci.* (2011) **31**, 24:8786-8802.
18. 18 Stephan, A.H., Barres, B.A. & Stevens, B. The complement system: an unexpected role in synaptic pruning during development and disease. *Annu. Rev. Neurosci.* (2012) **35**, 369-389.

19. 19 Cachat, J., Deffert, C., Hugues, S. & Krause, K.H. Phagocyte NADPH oxidase and specific immunity. *Clin. Sci. (Lond.)* (2015) **128**, 10:635-648.

20. 20 Benito, E. & Barco, A. CREB's control of intrinsic and synaptic plasticity: implications for CREB-dependent memory models. *Trends Neurosci.* (2010) **33**, 5:230-240.

21. 21 Richards, J.P., Bachinger, H.P., Goodman, R.H. & Brennan, R.G. Analysis of the structural properties of cAMP-responsive element-binding protein (CREB) and phosphorylated CREB. *J. Biol. Chem.* (1996) **271**, 23:13716-13723.

22. 22 Everett, L.J. *et al.* Integrative genomic analysis of CREB defines a critical role for transcription factor networks in mediating the fed/fasted switch in liver. *BMC Genomics* (2013) **14**, 1:337.

23. 23 Briand, L.A., Lee, B.G., Lelay, J., Kaestner, K.H. & Blendy, J.A. Serine 133 phosphorylation is not required for hippocampal CREB-mediated transcription and behavior. *Learn. Mem.* (2015) **22**, 2:109-115.

24. 24 Lesiak, A. *et al.* A genome-wide screen of CREB occupancy identifies the RhoA inhibitors Par6C and Rnd3 as regulators of BDNF-induced synaptogenesis. *PLoS One* (2013) **8**, 6:e64658.

25. 25 Hansen, K.F., Sakamoto, K., Pelz, C., Impey, S. & Obrietan, K. Profiling status epilepticus-induced changes in hippocampal RNA expression using high-throughput RNA sequencing. *Sci. Rep.* (2014) **4**, 6930.

26. 26 Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* (2009) **457**, 7231:854-858.

27. 27 Hallenborg, P. *et al.* MDM2 facilitates adipocyte differentiation through CRTC-mediated activation of STAT3. *Cell Death Dis.* (2016) **7**, 6:e2289.

28. 28 McClung, C.A. & Nestler, E.J. Regulation of gene expression and cocaine reward by CREB and DeltaFosB. *Nat. Neurosci.* (2003) **6**, 11:1208-1215.

29. 29 Benito, E., Valor, L.M., Jimenez-Minchan, M., Huber, W. & Barco, A. cAMP response element-binding protein is a primary hub of activity-driven neuronal gene expression. *J. Neurosci.* (2011) **31**, 50:18237.

30. 30 Ogryzko, V.V., Schiltz, R.L., Russanova, V., Howard, B.H. & Nakatani, Y. The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell* (1996) **87**, 5:953-959.

31. 31 Mejia, L.A. *et al.* A novel Hap1-Tsc1 interaction regulates neuronal mTORC1 signaling and morphogenesis in the brain. *J. Neurosci.* (2013) **33**, 46:18015-18021.

32. 32 Zhao, S., Fung-Leung, W.P., Bittner, A., Ngo, K. & Liu, X. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One* (2014) **9**, 1:e78644.

33. 33 Zhang, J. *et al.* CREB-mediated synaptogenesis and neurogenesis is crucial for the role of 5-HT1a receptors in modulating anxiety behaviors. *Sci. Rep.* (2016) **6**, 29551.

34. 34 Mantamadiotis, T. *et al.* Disruption of CREB function in brain leads to neurodegeneration. *Nat. Genet.* (2002) **31**, 1:47-54.

35. 35 Viosca, J. *et al.* Chronic enhancement of CREB activity in the hippocampus interferes with the retrieval of spatial information. *Learn. Mem.* (2009) **16**, 3:198-209.

36. 36 Pramod, A.B., Foster, J., Carvelli, L. & Henry, L.K. SLC6 transporters: structure, function, regulation, disease association and therapeutics. *Mol. Aspects Med.* (2013) **34**, 2-3:197-219.

37. 37 Hirano, Y. *et al.* Shifting transcriptional machinery is required for long-term memory maintenance and modification in Drosophila mushroom bodies. *Nature Communications* (2016) **7**, 13471.

38. 38 Sandberg, M. *et al.* Transcriptional Networks Controlled by NKX2-1 in the Development of Forebrain GABAergic Neurons. *Neuron* (2016) **91**, 6:1260-1275.

39. 39 Bernhart, S.H. *et al.* Changes of bivalent chromatin coincide with increased expression of developmental genes in cancer. *Sci. Rep.* (2016) **6**, 37393.

40. 40 Kolasinska-Zwierz, P. *et al.* Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat. Genet.* (2009) **41**, 3:376-381.

41. 41 Nguyen, A.T. & Zhang, Y. The diverse functions of Dot1 and H3K79 methylation. *Genes Dev.* (2011) **25**, 13:1345-1358.

42. 42 Vella, P. *et al.* Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. *Mol. Cell* (2013) **49**, 4:645-656.

43. 43 Mikkelsen, T.S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* (2007) **448**, 7153:553-560.

44. 44 Malik, A.N. *et al.* Genome-wide identification and characterization of functional neuronal activity-dependent enhancers. *Nat. Neurosci.* (2014) **17**, 10:1330-1339.
45. 45 Chen, Q., Chen, Y., Bian, C., Fujiki, R. & Yu, X. TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature* (2013) **493**, 7433:561-564.
46. 46 Deplus, R. *et al.* TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *EMBO J.* (2013) **32**, 5:645-655.
47. 47 Hahn, M.A. *et al.* Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis. *Cell Rep* (2013) **3**, 2:291-300.
48. 48 Marson, A. *et al.* Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* (2008) **134**, 3:521-533.
49. 49 Ku, M. *et al.* Genome-wide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet.* (2008) **4**, 10:e1000242.
50. 50 Myers, S.A., Panning, B. & Burlingame, A.L. Polycomb repressive complex 2 is necessary for the normal site-specific O-GlcNAc distribution in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* (2011) **108**, 23:9490-9495.
51. 51 Gambetta, M.C., Oktaba, K. & Muller, J. Essential role of the glycosyltransferase sxc/Ogt in polycomb repression. *Science* (2009) **325**, 5936:93-96.
52. 52 Maury, J.J. *et al.* RING1B O-GlcNAcylation regulates gene targeting of polycomb repressive complex 1 in human embryonic stem cells. *Stem Cell Res* (2015) **15**, 1:182-189.
53. 53 Jang, H. *et al.* O-GlcNAc regulates pluripotency and reprogramming by directly acting on core components of the pluripotency network. *Cell Stem Cell* (2012) **11**, 1:62-74.
54. 54 Myers, S.A. *et al.* SOX2 O-GlcNAcylation alters its protein-protein interactions and genomic occupancy to modulate gene expression in pluripotent cells. *Elife* (2016) **5**, e10647.
55. 55 Chu, C.S. *et al.* O-GlcNAcylation regulates EZH2 protein stability and function. *Proc. Natl. Acad. Sci. U. S. A.* (2014) **111**, 4:1355-1360.
56. 56 Jin, S.G. *et al.* Tet3 Reads 5-Carboxylcytosine through Its CXXC Domain and Is a Potential Guardian against Neurodegeneration. *Cell Rep* (2016) **14**, 3:493-505.
57. 57 Saunders, A. *et al.* The SIN3A/HDAC Corepressor Complex Functionally Cooperates with NANOG to Promote Pluripotency. *Cell Rep* (2017) **18**, 7:1713-1726.
58. 58 Rudenko, A. *et al.* Tet1 is critical for neuronal activity-regulated gene expression and memory extinction. *Neuron* (2013) **79**, 6:1109-1122.
59. 59 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* (2008) **9**, 559.
60. 60 Altarejos, J.Y. & Montminy, M. CREB and the CRTC co-activators: sensors for hormonal and metabolic signals. *Nat. Rev. Mol. Cell Biol.* (2011) **12**, 3:141-151.
61. 61 Herzig, S. *et al.* CREB regulates hepatic gluconeogenesis through the coactivator PGC-1. *Nature* (2001) **413**, 6852:179-183.
62. 62 Langfelder, P., Mischel, P.S. & Horvath, S. When is hub gene selection better than standard meta-analysis? *PLoS One* (2013) **8**, 4:e61505.
63. 63 Rouault, T.A. Iron metabolism in the CNS: implications for neurodegenerative diseases. *Nat Rev Neurosci* (2013) **14**, 8:551-564.
64. 64 Choi, D.S. *et al.* The type 1 equilibrative nucleoside transporter regulates ethanol intoxication and preference. *Nat Neurosci* (2004) **7**, 8:855-861.
65. 65 Lin, J.R., Zeman, M.K., Chen, J.Y., Yee, M.C. & Cimprich, K.A. SHPRH and HLTF act in a damage-specific manner to coordinate different forms of postreplication repair and prevent mutagenesis. *Mol Cell* (2011) **42**, 2:237-249.
66. 66 Motegi, A. *et al.* Polyubiquitination of proliferating cell nuclear antigen by HLTF and SHPRH prevents genomic instability from stalled replication forks. *Proc Natl Acad Sci U S A* (2008) **105**, 34:12411-12416.
67. 67 Unk, I. *et al.* Human HLTF functions as a ubiquitin ligase for proliferating cell nuclear antigen polyubiquitination. *Proceedings of the National Academy of Sciences of the United States of America* (2008) **105**, 10:3768-3773.
68. 68 Kile, A.C. *et al.* HLTF's Ancient HIRAN Domain Binds 3' DNA Ends to Drive Replication Fork Reversal. *Mol Cell* (2015) **58**, 6:1090-1100.

69. 69 Wang, L. *et al.* Card10 is a novel caspase recruitment domain/membrane-associated guanylate kinase family member that interacts with BCL10 and activates NF-kappa B. *J Biol Chem* (2001) **276**, 24:21405-21409.

70. 70 Zhang, X. *et al.* Arylsulfatase B modulates neurite outgrowth via astrocyte chondroitin-4-sulfate: dysregulation by ethanol. *Glia* (2014) **62**, 2:259-271.

71. 71 Bhattacharyya, S., Feferman, L. & Tobacman, J.K. Restriction of Aerobic Metabolism by Acquired or Innate Arylsulfatase B Deficiency: A New Approach to the Warburg Effect. *Sci Rep* (2016) **6**, 32885.

72. 72 Lujan, R., Marron Fernandez de Velasco, E., Aguado, C. & Wickman, K. New insights into the therapeutic potential of Girk channels. *Trends Neurosci* (2014) **37**, 1:20-29.

73. 73 Munoz, M.B. *et al.* A role for the GIRK3 subunit in methamphetamine-induced attenuation of GABAB receptor-activated GIRK currents in VTA dopamine neurons. *J Neurosci* (2016) **36**, 11:3106-3114.

74. 74 Golonzhka, O. *et al.* Pbx regulates patterning of the cerebral cortex in progenitors and postmitotic neurons. *Neuron* (2015) **88**, 6:1192-1207.

75. 75 Villaescusa, J.C. *et al.* A PBX1 transcriptional network controls dopaminergic neuron development and is impaired in Parkinson's disease. *The EMBO Journal* (2016) **35**, 18:1963.

76. 76 Muralidharan, B. *et al.* LHX2 Interacts with the NuRD Complex and Regulates Cortical Neuron Subtype Determinants Fezf2 and Sox11. *J Neurosci* (2017) **37**, 1:194-203.

77. 77 Yasushi, O. *et al.* SAFB1, an RBMX-binding protein, is a newly identified regulator of hepatic SREBP-1c gene. *BMB Rep.* (2009) **42**, 4:232-237.

78. 78 Tsend-Ayush, E. *et al.* RBMX gene is essential for brain development in zebrafish. *Dev Dyn* (2005) **234**, 3:682-688.

79. 79 Shashi, V. *et al.* The RBMX gene as a candidate for the Shashi X-linked intellectual disability syndrome. *Clin Genet* (2015) **88**, 4:386-390.

80. 80 Olkkonen, V.M. & Li, S. Oxysterol-binding proteins: sterol and phosphoinositide sensors coordinating transport, signaling and metabolism. *Prog Lipid Res* (2013) **52**, 4:529-538.

81. 81 Prentice, H., Modi, J.P. & Wu, J.Y. Mechanisms of Neuronal Protection against Excitotoxicity, Endoplasmic Reticulum Stress, and Mitochondrial Dysfunction in Stroke and Neurodegenerative Diseases. *Oxid. Med. Cell. Longev.* (2015) **2015**, 964518.

82. 82 Shaywitz, A.J., Dove, S.L., Kornhauser, J.M., Hochschild, A. & Greenberg, M.E. Magnitude of the CREB-dependent transcriptional response is determined by the strength of the interaction between the kinase-inducible domain of CREB and the KIX domain of CREB-binding protein. *Mol. Cell. Biol.* (2000) **20**, 24:9409-9422.

83. 83 Hirano, Y. *et al.* Fasting launches CRTC to facilitate long-term memory formation in Drosophila. *Science* (2013) **339**, 6118:443-446.

84. 84 Uchida, S. *et al.* CRTC1 Nuclear Translocation Following Learning Modulates Memory Strength via Exchange of Chromatin Remodeling Complexes on the Fgf1 Gene. *Cell Rep* (2017) **18**, 2:352-366.

85. 85 Nonaka, M. *et al.* Region-specific activation of CRTC1-CREB signaling mediates long-term fear memory. *Neuron* (2014) **84**, 1:92-106.

86. 86 Lopez de Armentia, M. *et al.* cAMP response element-binding protein-mediated gene expression increases the intrinsic excitability of CA1 pyramidal neurons. *J. Neurosci.* (2007) **27**, 50:13909-13918.

87. 87 Obrietan, K. & Hoyt, K.R. CRE-mediated transcription is increased in Huntington's disease transgenic mice. *J. Neurosci.* (2004) **24**, 4:791-796.

88. 88 Saura, C. & Valero, J. in *Rev. Neurosci.* Vol. 22   153 (2011).

89. 89 Briand, L.A. & Blendy, J.A. Not all stress is equal: CREB is not necessary for restraint stress reinstatement of cocaine-conditioned reward. *Behav. Brain Res.* (2013) **246**, 63-68.

90. 90 Pandey, S.C., Roy, A., Zhang, H. & Xu, T. Partial deletion of the cAMP response element-binding protein gene promotes alcohol-drinking behaviors. *J. Neurosci.* (2004) **24**, 21:5022-5030.

91. 91 Bourtchuladze, R. *et al.* Deficient long-term memory in mice with a targeted mutation of the cAMP-responsive element-binding protein. *Cell* (1994) **79**, 1:59-68.

92. 92 Hummler, E. *et al.* Targeted mutation of the CREB gene: compensation within the CREB/ATF family of transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* (1994) **91**, 12:5647-5651.

93.   93 Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* (2012) **7**, 3:562-578.
94.   94 Chen, H. & Boutros, P.C. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* (2011) **12**, 1:35.
95.   95 Hulsen, T., de Vlieg, J. & Alkema, W. BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics* (2008) **9**, 488.
96.   96 Storey, J.D., Taylor, J.E. & Siegmund, D. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society Series B-Statistical Methodology* (2004) **66**, 1:187-205.
97.   97 Pedersen, B.S., Yang, I.V. & De, S. CruzDB: software for annotation of genomic intervals with UCSC genome-browser database. *Bioinformatics* (2013) **29**, 23:3003-3006.
98.   98 mgu74av2.db: Affymetrix Murine Genome U74v2 annotation data (chip mgu74av2) v. Version 3.2.3 (2016).
99.   99 mgu74a.db: Affymetrix Murine Genome U74v2 annotation data (chip mgu74a) v. Version 3.2.3 (2016).
100. 100   Qiu, J. *et al.* Evidence for evolutionary divergence of activity-dependent gene expression in developing neurons. *Elife* (2016) **5**, e20337.
101. 101   Anders, S., Pyl, P.T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* (2015) **31**, 2:166-169.
102. 102   Anders, S. *et al.* Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat. Protoc.* (2013) **8**, 9:1765-1786.
103. 103   McCarthy, D.J., Chen, Y. & Smyth, G.K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* (2012) **40**, 10:4288-4297.
104. 104   Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* (2014) **15**, 12:550.
105. 105   Langfelder, P. & Horvath, S. Fast R Functions for Robust Correlations and Hierarchical Clustering. *J Stat Softw* (2012) **46**, 11:17.
106. 106   GO.db: A set of annotation maps describing the entire Gene Ontology v. Version 3.4.1 (2017).
107. 107   AnnotationDbi: Annotation Database Interface. v. 1.38.1. (2017).
108. 108   org.Mm.eg.db: Genome wide annotation for Mouse v. Version 3.4.1 (2017).
109. 109   Mi, H., Muruganujan, A., Casagrande, J.T. & Thomas, P.D. Large-scale gene function analysis with the PANTHER classification system. *Nat. Protoc.* (2013) **8**, 8:1551-1566.
110. 110   Huang da, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* (2009) **4**, 1:44-57.
111. 111   Kuleshov, M.V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* (2016) **44**, W1:W90-97.
112. 112   Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* (2003) **13**, 11:2498-2504.
113. 113   Bindea, G. *et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* (2009) **25**, 8:1091-1093.
114. 114   Bindea, G., Galon, J. & Mlecnik, B. CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. *Bioinformatics* (2013) **29**, 5:661-663.
115. 115   Hu, Z. *et al.* VisANT 4.0: Integrative network platform to connect genes, drugs, diseases and therapies. *Nucleic Acids Res.* (2013) **41**, Web Server issue:W225-231.

Appendix I: Upregulated genes in OGT cKO hippocampi at 2 months of age

Gene Symbol	Description	log2 FC
Ccl3	chemokine (C-C motif) ligand 3	5.0
Clec7a	C-type lectin domain family 7, member a	4.9
Cst7	cystatin F (leukocystatin)	4.9
Cxcl10	chemokine (C-X-C motif) ligand 10	4.3
Ccl4	chemokine (C-C motif) ligand 4	3.9
Itgax	integrin alpha X	3.6
Isg15	ISG15 ubiquitin-like modifier	3.2
Ccl12	chemokine (C-C motif) ligand 12|c-C motif chemokine 12-like	3.2
Glycam1	glycosylation dependent cell adhesion molecule 1	3.1
Ccl5	chemokine (C-C motif) ligand 5	3.0
Usp18	ubiquitin specific peptidase 18	3.0
Ifi27l2a	interferon, alpha-inducible protein 27 like 2A	2.9
Ccl6	chemokine (C-C motif) ligand 6	2.9
Cd52	CD52 antigen	2.9
Taar3	trace amine-associated receptor 3	2.8
Ifit1	interferon-induced protein with tetratricopeptide repeats 1	2.7
Irf7	interferon regulatory factor 7	2.7
Timp1	tissue inhibitor of metalloproteinase 1	2.7
Pyhin1	pyrin and HIN domain family, member 1	2.7
Zc3h12d	zinc finger CCCH type containing 12D	2.7
Gfap	glial fibrillary acidic protein	2.6
Trem2	triggering receptor expressed on myeloid cells 2	2.6
Rtp4	receptor transporter protein 4	2.5
Cxcl9	chemokine (C-X-C motif) ligand 9	2.5
C3ar1	complement component 3a receptor 1	2.5
I830012O16Rik	RIKEN cDNA I830012O16 gene	2.5
Mpeg1	macrophage expressed gene 1	2.4
Lag3	lymphocyte-activation gene 3	2.3
Lgals3bp	lectin, galactoside-binding, soluble, 3 binding protein	2.3
Bst2	bone marrow stromal cell antigen 2	2.3
Slc15a3	solute carrier family 15, member 3	2.3
Siglec5	sialic acid binding Ig-like lectin 5	2.3
Endou	endonuclease, polyU-specific	2.2
Oasl2	2'-5' oligoadenylate synthetase-like 2	2.2
Ch25h	cholesterol 25-hydroxylase	2.2
Aspg	asparaginase homolog (S. cerevisiae)	2.2
Tlr2	toll-like receptor 2	2.2
Ly86	lymphocyte antigen 86	2.2
Cd48	CD48 antigen	2.1

Gpr84	G protein-coupled receptor 84	2.1
Gpnmb	glycoprotein (transmembrane) nmb	2.1
Gbp3	guanylate binding protein 3	2.0
C1qb	complement component 1, q subcomponent, beta polypeptide	2.0
Ccl2	chemokine (C-C motif) ligand 2	2.0
Ip6k3	inositol hexaphosphate kinase 3	2.0
Ch25h	cholesterol 25-hydroxylase	2.0
Tyrobp	TYRO protein tyrosine kinase binding protein	2.0
Osmr	oncostatin M receptor	2.0
Lat2	linker for activation of T cells family, member 2	2.0
Slc11a1	solute carrier family 11 (proton-coupled divalent metal ion transporters), member 1	2.0
Cd14	CD14 antigen	2.0
Csf3r	colony stimulating factor 3 receptor (granulocyte)	2.0
Ddx60	DEAD (Asp-Glu-Ala-Asp) box polypeptide 60	1.9
Mpeg1	macrophage expressed gene 1	1.9
Capg	capping protein (actin filament), gelsolin-like	1.9
Cfb	complement factor B	1.9
Hvcn1	hydrogen voltage-gated channel 1	1.9
Ctsz	cathepsin Z	1.9
Rnf213	ring finger protein 213	1.9
Plek	pleckstrin	1.9
Tap1	transporter 1, ATP-binding cassette, sub-family B (MDR/TAP)	1.8
Parp14	poly (ADP-ribose) polymerase family, member 14	1.8
Sstr5	somatostatin receptor 5	1.8
1700112E06Rik	RIKEN cDNA 1700112E06 gene	1.8
C1qc	complement component 1, q subcomponent, C chain	1.8
LOC68395	histocompatibility 2, Q region locus 6-like	1.8
Psmb9	proteasome (prosome, macropain) subunit, beta type 9 (large multifunctional peptidase 2)	1.8
Cd68	CD68 antigen	1.8
C1qa	complement component 1, q subcomponent, alpha polypeptide	1.8
Cd72	CD72 antigen|b-cell differentiation antigen CD72-like	1.8
Gvin1	GTPase, very large interferon inducible 1|predicted gene 4070|very large inducible GTPase 1 pseudogene|very large inducible GTPase 1 pseudogene	1.8
Fyb	FYN binding protein	1.8
Psmb8	proteasome (prosome, macropain) subunit, beta type 8 (large multifunctional peptidase 7)	1.8
Cxcl5	chemokine (C-X-C motif) ligand 5	1.8
Fcgr4	Fc receptor, IgG, low affinity IV	1.8
Trim30d	tripartite motif-containing 30D|tripartite motif-containing 30A	1.7
Itgb2	integrin beta 2	1.7
Tbxas1	thromboxane A synthase 1, platelet	1.7

A2m	alpha-2-macroglobulin	1.7
Myo1f	myosin IF	1.7
Igtp	interferon gamma induced GTPase	1.7
Trim34a	tripartite motif-containing 34A	1.6
Ctss	cathepsin S	1.6
Irf8	interferon regulatory factor 8	1.6
Socs3	suppressor of cytokine signaling 3	1.6
Mx2	myxovirus (influenza virus) resistance 2	1.6
Gbp2	guanylate binding protein 2	1.6
Fcer1g	Fc receptor, IgE, high affinity I, gamma polypeptide	1.6
Ifitm3	interferon induced transmembrane protein 3	1.6
Fcgr2b	Fc receptor, IgG, low affinity IIb	1.6
Naip5	NLR family, apoptosis inhibitory protein 5	1.6
Cyba	cytochrome b-245, alpha polypeptide	1.6
Phf11	PHD finger protein 11|predicted gene 4902|predicted gene 6904	1.5
Ms4a6c	membrane-spanning 4-domains, subfamily A, member 6C	1.5
Fcgr1	Fc receptor, IgG, high affinity I	1.5
Havcr2	hepatitis A virus cellular receptor 2	1.5
Ggta1	glycoprotein galactosyltransferase alpha 1, 3	1.5
Zc3hav1	zinc finger CCCH type, antiviral 1	1.5
Ddx58	DEAD (Asp-Glu-Ala-Asp) box polypeptide 58	1.5
Irf9	interferon regulatory factor 9	1.5
Cd74	CD74 antigen (invariant polypeptide of major histocompatibility complex, class II antigen-associated)	1.5
Trim25	tripartite motif-containing 25	1.5
Icam1	intercellular adhesion molecule 1	1.5
Gbp6	guanylate binding protein 6	1.5
D14Ertd668e	DNA segment, Chr 14, ERATO Doi 668, expressed	1.5
AF251705	cDNA sequence AF251705	1.5
Laptm5	lysosomal-associated protein transmembrane 5	1.4
Dnase2a	deoxyribonuclease II alpha	1.4
Fcgr3	Fc receptor, IgG, low affinity III	1.4
Samsn1	SAM domain, SH3 domain and nuclear localization signals, 1	1.4
Irgm2	immunity-related GTPase family M member 2	1.4
Ptpn6	protein tyrosine phosphatase, non-receptor type 6	1.4
Gusb	glucuronidase, beta	1.4
Aif1	allograft inflammatory factor 1	1.4
Iigp1	interferon inducible GTPase 1	1.4
Ifi44	interferon-induced protein 44	1.4
Pld4	phospholipase D family, member 4	1.4
Mlxipl	MLX interacting protein-like	1.4
Tlr7	toll-like receptor 7	1.4

Mettl11b	methyltransferase like 11B	1.4
Spp1	secreted phosphoprotein 1	1.4
Slc43a3	solute carrier family 43, member 3	1.4
Fcrls	Fc receptor-like S, scavenger receptor	1.4
Lgals9	lectin, galactose binding, soluble 9	1.4
Il4i1	interleukin 4 induced 1|Nup62-Il4i1 protein	1.4
Irgm1	immunity-related GTPase family M member 1	1.4
Olfml3	olfactomedin-like 3	1.4
Tlr13	toll-like receptor 13	1.3
Lair1	leukocyte-associated Ig-like receptor 1	1.3
Fcgr2b	Fc receptor, IgG, low affinity IIb	1.3
Herc6	hect domain and RLD 6	1.3
Vav1	vav 1 oncogene	1.3
Ptpn18	protein tyrosine phosphatase, non-receptor type 18	1.3
Parp9	poly (ADP-ribose) polymerase family, member 9	1.3
Cd53	CD53 antigen	1.3
Slfn2	schlafen 2	1.3
Grn	granulin	1.3
Samd9l	sterile alpha motif domain containing 9-like	1.3
Tcfcp2l1	transcription factor CP2-like 1	1.3
Hexb	hexosaminidase B	1.3
A430084P05Rik	RIKEN cDNA A430084P05 gene	1.3
Btk	Bruton agammaglobulinemia tyrosine kinase	1.3
Cd74	CD74 antigen (invariant polypeptide of major histocompatibility complex, class II antigen-associated)	1.3
Ctsh	cathepsin H	1.3
Parp3	poly (ADP-ribose) polymerase family, member 3	1.3
Tnfaip8l2	tumor necrosis factor, alpha-induced protein 8-like 2	1.3
Wdfy4	WD repeat and FYVE domain containing 4	1.3
Plin4	perilipin 4	1.3
Irf5	interferon regulatory factor 5	1.3
Igsf6	immunoglobulin superfamily, member 6	1.3
Trim12a	tripartite motif-containing 12A	1.3
Slamf9	SLAM family member 9	1.3
Hk3	hexokinase 3	1.2
Cd44	CD44 antigen	1.2
Batf3	basic leucine zipper transcription factor, ATF-like 3	1.2
Oasl1	2'-5' oligoadenylate synthetase-like 1	1.2
Ifit2	interferon-induced protein with tetratricopeptide repeats 2	1.2
Pyroxd2	pyridine nucleotide-disulphide oxidoreductase domain 2	1.2
Hpse	heparanase	1.2
Gpr65	G-protein coupled receptor 65	1.2

Tgfbr2	transforming growth factor, beta receptor II	1.2
Arl11	ADP-ribosylation factor-like 11	1.2
Tnfrsf1a	tumor necrosis factor receptor superfamily, member 1a	1.2
Cd274	CD274 antigen	1.2
Hcst	hematopoietic cell signal transducer	1.2
Cd86	CD86 antigen	1.2
Ctsd	cathepsin D	1.2
Ifi47	interferon gamma inducible protein 47	1.2
Cx3cr1	chemokine (C-X3-C) receptor 1	1.2
Uba7	ubiquitin-like modifier activating enzyme 7	1.2
Hck	hemopoietic cell kinase	1.2
Inpp5d	inositol polyphosphate-5-phosphatase D	1.2
Nckap1l	NCK associated protein 1 like	1.2
Isg20	interferon-stimulated protein	1.2
Emr1	EGF-like module containing, mucin-like, hormone receptor-like sequence 1	1.2
Parp12	poly (ADP-ribose) polymerase family, member 12	1.2
Fermt3	fermitin family homolog 3 (Drosophila)	1.2
Pros1	protein S (alpha)	1.2
Rac2	RAS-related C3 botulinum substrate 2	1.2
Il1a	interleukin 1 alpha	1.2
Sh3rf2	SH3 domain containing ring finger 2	1.2
Tubb6	tubulin, beta 6	1.2
Msn	moesin	1.1
Slc14a1	solute carrier family 14 (urea transporter), member 1	1.1
Lair1	leukocyte-associated Ig-like receptor 1	1.1
Pfkfb3	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3	1.1
Hcls1	hematopoietic cell specific Lyn substrate 1	1.1
Serpina3n	serine (or cysteine) peptidase inhibitor, clade A, member 3N	1.1
Ncf1	neutrophil cytosolic factor 1	1.1
Lrmp	lymphoid-restricted membrane protein	1.1
Hcst	hematopoietic cell signal transducer	1.1
Csf1r	colony stimulating factor 1 receptor	1.1
Dtx3l	deltex 3-like (Drosophila)	1.1
Il1r2	interleukin 1 receptor, type II	1.1
Itgb5	integrin beta 5	1.1
Arpc1b	actin related protein 2/3 complex, subunit 1B	1.1
Blnk	B-cell linker	1.1
Glrp1	glutamine repeat protein 1	1.1
Parp12	poly (ADP-ribose) polymerase family, member 12	1.1
Cd84	CD84 antigen	1.1
Arhgap30	Rho GTPase activating protein 30	1.1

Slfn5	schlafen 5	1.1
B2m	beta-2 microglobulin	1.1
Tor3a	torsin family 3, member A	1.1
Neat1	nuclear paraspeckle assembly transcript 1 (non-protein coding)	1.1
S100a6	S100 calcium binding protein A6 (calcyclin)	1.1
Abca1	ATP-binding cassette, sub-family A (ABC1), member 1	1.1
Snx20	sorting nexin 20	1.1
Thbs4	thrombospondin 4	1.1
Trim21	tripartite motif-containing 21	1.1
Plscr2	phospholipid scramblase 2	1.1
Apobec1	apolipoprotein B mRNA editing enzyme, catalytic polypeptide 1	1.1
Was	Wiskott-Aldrich syndrome homolog (human)	1.1
Apbb1ip	amyloid beta (A4) precursor protein-binding, family B, member 1 interacting protein	1.1
Tex11	testis expressed gene 11	1.1
Parp10	poly (ADP-ribose) polymerase family, member 10	1.1
C3	complement component 3	1.1
Serping1	serine (or cysteine) peptidase inhibitor, clade G, member 1	1.0
Glipr2	GLI pathogenesis-related 2	1.0
Ifi30	interferon gamma inducible protein 30	1.0
Tlr12	toll-like receptor 12	1.0
Gpsm3	G-protein signalling modulator 3 (AGS3-like, C. elegans)	1.0
Ms4a6d	membrane-spanning 4-domains, subfamily A, member 6D	1.0
Lpxn	leupaxin	1.0
Kcnj2	potassium inwardly-rectifying channel, subfamily J, member 2	1.0
Slc7a7	solute carrier family 7 (cationic amino acid transporter, y+ system), member 7	1.0
Trim56	tripartite motif-containing 56	1.0
Axl	AXL receptor tyrosine kinase	1.0
Hspb6	heat shock protein, alpha-crystallin-related, B6	1.0
Tnni2	troponin I, skeletal, fast 2	1.0
Dhx58	DEXH (Asp-Glu-X-His) box polypeptide 58	1.0
H2-Aa	histocompatibility 2, class II antigen A, alpha	1.0
Spint1	serine protease inhibitor, Kunitz type 1	1.0
Hpgds	hematopoietic prostaglandin D synthase	1.0
Ctla2b	cytotoxic T lymphocyte-associated protein 2 beta	1.0
Ptprc	protein tyrosine phosphatase, receptor type, C	1.0
Guca1a	guanylate cyclase activator 1a (retina)	1.0
Wnk4	WNK lysine deficient protein kinase 4	0.99
Tap2	transporter 2, ATP-binding cassette, sub-family B (MDR/TAP)	0.99
Rasa4	RAS p21 protein activator 4	0.99
Sh3bp2	SH3-domain binding protein 2	0.99

Ctla2b	cytotoxic T lymphocyte-associated protein 2 beta	0.99
Baiap2l2	BAI1-associated protein 2-like 2	0.99
Cchcr1	coiled-coil alpha-helical rod protein 1	0.98
Trim21	tripartite motif-containing 21	0.98
Ifi27l1	interferon, alpha-inducible protein 27 like 1	0.97
Gadl1	glutamate decarboxylase-like 1	0.97
Hmox1	heme oxygenase (decycling) 1	0.97
Fblim1	filamin binding LIM protein 1	0.97
P2ry6	pyrimidinergic receptor P2Y, G-protein coupled, 6	0.97
Neurl3	neuralized homolog 3 homolog (Drosophila)	0.97
Gngt2	guanine nucleotide binding protein (G protein), gamma transducing activity polypeptide 2	0.97
Adora3	adenosine A3 receptor	0.97
Il1r1	interleukin 1 receptor, type I	0.97
Gbp5	guanylate binding protein 5	0.96
Asap3	ArfGAP with SH3 domain, ankyrin repeat and PH domain 3	0.96
Eif2ak2	eukaryotic translation initiation factor 2-alpha kinase 2	0.96
Cd9	CD9 antigen	0.96
Naprt1	nicotinate phosphoribosyltransferase domain containing 1	0.96
H2-DMb1	histocompatibility 2, class II, locus Mb1|class II histocompatibility antigen, M beta 1 chain-like	0.96
Dok1	docking protein 1	0.96
Selplg	selectin, platelet (p-selectin) ligand	0.95
Padi2	peptidyl arginine deiminase, type II	0.95
S100a4	S100 calcium binding protein A4	0.95
Itgb5	integrin beta 5	0.95
Hk2	hexokinase 2	0.95
F11r	F11 receptor	0.94
Fmod	fibromodulin	0.94
Vim	vimentin	0.93
Aldh1l2	aldehyde dehydrogenase 1 family, member L2	0.93
Rasgrp3	RAS, guanyl releasing protein 3	0.93
Rasal3	RAS protein activator like 3	0.93
Spata13	spermatogenesis associated 13	0.93
Hist1h4h	histone cluster 1, H4h	0.93
Flnc	filamin C, gamma	0.92
Tmem176a	transmembrane protein 176A	0.92
Tmem176b	transmembrane protein 176B	0.92
Ms4a7	membrane-spanning 4-domains, subfamily A, member 7	0.92
LOC635918	beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase-like	0.92
Itgam	integrin alpha M	0.92
Tspo	translocator protein	0.92

Arhgap9	Rho GTPase activating protein 9	0.92
Ttc28	tetratricopeptide repeat domain 28	0.91
H2-Ab1	histocompatibility 2, class II antigen A, beta 1	0.90
Sash3	SAM and SH3 domain containing 3	0.90
Npas4	neuronal PAS domain protein 4	0.90
Prrx2	paired related homeobox 2	0.90
Syngr2	synaptogyrin 2	0.89
Naip6	NLR family, apoptosis inhibitory protein 6	0.89
Ifi35	interferon-induced protein 35	0.89
Arhgap25	Rho GTPase activating protein 25	0.89
Renbp	renin binding protein	0.89
4632428N05Rik	RIKEN cDNA 4632428N05 gene	0.89
Hpgd	hydroxyprostaglandin dehydrogenase 15 (NAD)	0.89
Cd63	CD63 antigen|CD63 antigen-like	0.88
Epsti1	epithelial stromal interaction 1 (breast)	0.88
Pmp22	peripheral myelin protein 22	0.88
Bcl3	B-cell leukemia/lymphoma 3	0.88
Zfp36	zinc finger protein 36	0.88
Phyhd1	phytanoyl-CoA dioxygenase domain containing 1	0.87
Clic1	chloride intracellular channel 1	0.87
P4ha3	procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), alpha polypeptide III	0.87
Tapbp	TAP binding protein	0.87
Kcnk6	potassium inwardly-rectifying channel, subfamily K, member 6	0.86
Unc93b1	unc-93 homolog B1 (C. elegans)	0.86
Runx1	runt related transcription factor 1	0.86
Il10ra	interleukin 10 receptor, alpha	0.86
Gpr34	G protein-coupled receptor 34	0.86
Gal3st4	galactose-3-O-sulfotransferase 4	0.86
Aqp4	aquaporin 4	0.86
F9	coagulation factor IX	0.86
Tep1	telomerase associated protein 1	0.86
Tgif1	TGFB-induced factor homeobox 1	0.85
5430435G22Rik	RIKEN cDNA 5430435G22 gene	0.85
Pik3cg	phosphoinositide-3-kinase, catalytic, gamma polypeptide	0.85
Mafb	v-maf musculoaponeurotic fibrosarcoma oncogene family, protein B (avian)	0.85
Siglech	sialic acid binding Ig-like lectin H	0.84
Itgb5	integrin beta 5	0.84
Hlx	H2.0-like homeobox	0.84
Dock8	dedicator of cytokinesis 8	0.84
Csf2rb	colony stimulating factor 2 receptor, beta, low-affinity (granulocyte-macrophage)	0.84

Syngr2	synaptogyrin 2	0.84
Lamp2	lysosomal-associated membrane protein 2	0.83
Ly9	lymphocyte antigen 9	0.83
Hexa	hexosaminidase A	0.83
Nfkb2	nuclear factor of kappa light polypeptide gene enhancer in B-cells 2, p49/p100	0.83
Rhoj	ras homolog gene family, member J	0.83
Ms4a6d	membrane-spanning 4-domains, subfamily A, member 6D	0.83
Lgi4	leucine-rich repeat LGI family, member 4	0.83
BC026585	cDNA sequence BC026585	0.83
Apoc1	apolipoprotein C-I	0.82
Tcn2	transcobalamin 2	0.82
Rtp4	receptor transporter protein 4	0.82
Rrbp1	ribosome binding protein 1	0.82
Anxa3	annexin A3	0.81
Smoc1	SPARC related modular calcium binding 1	0.81
Cmtm3	CKLF-like MARVEL transmembrane domain containing 3	0.81
Pbxip1	pre-B-cell leukemia transcription factor interacting protein 1	0.80
Bgn	biglycan	0.80
Pdpn	podoplanin	0.80
Tcirg1	T-cell, immune regulator 1, ATPase, H+ transporting, lysosomal V0 protein A3	0.80
H2-M3	histocompatibility 2, M region locus 3	0.80
Kcne1l	potassium voltage-gated channel, Isk-related family, member 1-like, pseudogene	0.80
Sox9	SRY-box containing gene 9	0.80
Pik3ap1	phosphoinositide-3-kinase adaptor protein 1	0.79
2810459M11Rik	RIKEN cDNA 2810459M11 gene	0.78
Lrig1	leucine-rich repeats and immunoglobulin-like domains 1	0.78
Rgs1	regulator of G-protein signaling 1	0.78
Tmem173	transmembrane protein 173	0.78
Myo1g	myosin IG	0.78
Slc25a45	solute carrier family 25, member 45	0.77
Fxyd1	FXYD domain-containing ion transport regulator 1	0.77
Mrc2	mannose receptor, C type 2	0.77
Tlr3	toll-like receptor 3	0.76
A830007P12Rik	RIKEN cDNA A830007P12 gene	0.76
Slc1a4	solute carrier family 1 (glutamate/neutral amino acid transporter), member 4	0.76
Slc13a3	solute carrier family 13 (sodium-dependent dicarboxylate transporter), member 3	0.76
Fgl2	fibrinogen-like protein 2	0.76
Ptplad2	protein tyrosine phosphatase-like A domain containing 2	0.76

Ddit3	DNA-damage inducible transcript 3	0.76
Plce1	phospholipase C, epsilon 1	0.75
Lgmn	legumain	0.75
Gm885	predicted gene 885	0.74
Man2b1	mannosidase 2, alpha B1	0.74
Asph	aspartate-beta-hydroxylase	0.74
Itpr2	inositol 1,4,5-triphosphate receptor 2	0.74
Oas2	2'-5' oligoadenylate synthetase 2	0.74
Plxnb2	plexin B2	0.74
C1ra	complement component 1, r subcomponent A|complement component 1, r subcomponent B	0.73
Sdc4	syndecan 4	0.73
Thbs1	thrombospondin 1	0.73
Tmc6	transmembrane channel-like gene family 6	0.73
Tchh	trichohyalin	0.73
Samhd1	SAM domain and HD domain, 1	0.72
Cmtm7	CKLF-like MARVEL transmembrane domain containing 7	0.72
Col16a1	collagen, type XVI, alpha 1	0.72
Wfdc3	WAP four-disulfide core domain 3	0.71
Mvp	major vault protein	0.71
Dap	death-associated protein	0.71
Ahnak	AHNAK nucleoprotein (desmoyokin)	0.71
Anxa4	annexin A4	0.71
Anxa2	annexin A2	0.71
Oas1c	2'-5' oligoadenylate synthetase 1C	0.71
Atp1b2	ATPase, Na+/K+ transporting, beta 2 polypeptide	0.70
Tnfrsf1b	tumor necrosis factor receptor superfamily, member 1b	0.70
Gmfg	glia maturation factor, gamma	0.70
Plcd4	phospholipase C, delta 4	0.70
Ncf4	neutrophil cytosolic factor 4	0.69
Nmi	N-myc (and STAT) interactor	0.69
Cd37	CD37 antigen	0.69
Ccr5	chemokine (C-C motif) receptor 5	0.69
Gmip	Gem-interacting protein	0.69
Cdt1	chromatin licensing and DNA replication factor 1	0.69
Pik3r5	phosphoinositide-3-kinase, regulatory subunit 5, p101	0.69
Slc29a3	solute carrier family 29 (nucleoside transporters), member 3	0.69
Casp8	caspase 8	0.69
Hrsp12	heat-responsive protein 12	0.69
S100a10	S100 calcium binding protein A10 (calpactin)	0.69
Sparc	secreted acidic cysteine rich glycoprotein	0.68
Angpt1	angiopoietin 1	0.68

Hepacam	hepatocyte cell adhesion molecule	0.68
Npc2	Niemann Pick type C2	0.68
Trim66	tripartite motif-containing 66	0.68
Ehd4	EH-domain containing 4	0.67
Entpd4	ectonucleoside triphosphate diphosphohydrolase 4	0.67
Fgfrl1	fibroblast growth factor receptor-like 1	0.67
Fgfrl1	fibroblast growth factor receptor-like 1	0.67
Gem	GTP binding protein (gene overexpressed in skeletal muscle)	0.67
Eif4ebp1	eukaryotic translation initiation factor 4E binding protein 1	0.67
Arhgdib	Rho, GDP dissociation inhibitor (GDI) beta	0.67
Ikzf1	IKAROS family zinc finger 1	0.67
Cnn3	calponin 3, acidic	0.67
Plekhg2	pleckstrin homology domain containing, family G (with RhoGef domain) member 2	0.66
Rcsd1	RCSD domain containing 1	0.66
Fam46a	family with sequence similarity 46, member A	0.66
Kif23	kinesin family member 23	0.66
Gcnt1	glucosaminyl (N-acetyl) transferase 1, core 2	0.66
Dna2	DNA replication helicase 2 homolog (yeast)	0.66
Ahnak	AHNAK nucleoprotein (desmoyokin)	0.66
Map3k1	mitogen-activated protein kinase kinase kinase 1	0.65
Naglu	alpha-N-acetylglucosaminidase (Sanfilippo disease IIIB)	0.65
Tmem119	transmembrane protein 119	0.65
Ltbp3	latent transforming growth factor beta binding protein 3	0.65
Mtmr11	myotubularin related protein 11	0.65
Lmo2	LIM domain only 2	0.65
Rgs10	regulator of G-protein signalling 10	0.65
Ang	angiogenin, ribonuclease, RNase A family, 5|angiogenin, ribonuclease A family, member 3|angiogenin, ribonuclease A family, member 4	0.65
Anxa5	annexin A5	0.65
Sbno2	strawberry notch homolog 2 (Drosophila)	0.65
Asb2	ankyrin repeat and SOCS box-containing 2	0.65
P2ry13	purinergic receptor P2Y, G-protein coupled 13	0.65
Tor1aip1	torsin A interacting protein 1	0.65
Gpr160	G protein-coupled receptor 160	0.65
Tnfrsf11a	tumor necrosis factor receptor superfamily, member 11a	0.65
H2-Eb1	histocompatibility 2, class II antigen E beta	0.65
2610528A11Rik	RIKEN cDNA 2610528A11 gene	0.64
Mfsd1	major facilitator superfamily domain containing 1	0.64
Abcc3	ATP-binding cassette, sub-family C (CFTR/MRP), member 3	0.64
Unc93b1	unc-93 homolog B1 (C. elegans)	0.64
Csf1	colony stimulating factor 1 (macrophage)	0.64

Pdlim4	PDZ and LIM domain 4	0.64
Pde4d	phosphodiesterase 4D, cAMP specific	0.64
Rps6ka1	ribosomal protein S6 kinase polypeptide 1	0.64
Sdc3	syndecan 3	0.64
Il6ra	interleukin 6 receptor, alpha	0.64
Il20rb	interleukin 20 receptor beta	0.64
Srebf1	sterol regulatory element binding transcription factor 1	0.64
Hmha1	histocompatibility (minor) HA-1	0.64
Tagln2	transgelin 2	0.63
Reep3	receptor accessory protein 3	0.63
Abhd4	abhydrolase domain containing 4	0.63
Cebpa	CCAAT/enhancer binding protein (C/EBP), alpha	0.63
Trib3	tribbles homolog 3 (Drosophila)	0.63
Igfbp5	insulin-like growth factor binding protein 5	0.63
Pnpla7	patatin-like phospholipase domain containing 7	0.63
Rapgef3	Rap guanine nucleotide exchange factor (GEF) 3	0.63
Stom	stomatin	0.62
Amotl1	angiomotin-like 1	0.62
Sfxn5	sideroflexin 5	0.62
Zfp703	zinc finger protein 703	0.62
Tal1	T-cell acute lymphocytic leukemia 1	0.62
Lrrfip1	leucine rich repeat (in FLII) interacting protein 1	0.62
Epb4.1l2	erythrocyte protein band 4.1-like 2	0.62
Scamp2	secretory carrier membrane protein 2	0.61
Zfp36l1	zinc finger protein 36, C3H type-like 1	0.61
Gpr37l1	G protein-coupled receptor 37-like 1	0.61
Foxn3	forkhead box N3	0.61
Gna15	guanine nucleotide binding protein, alpha 15	0.61
Prcp	prolylcarboxypeptidase (angiotensinase C)	0.61
Tln1	talin 1	0.61
Rhbdf1	rhomboid family 1 (Drosophila)	0.61
Chchd2	coiled-coil-helix-coiled-coil-helix domain containing 2	0.60
Gns	glucosamine (N-acetyl)-6-sulfatase	0.60
Trim47	tripartite motif-containing 47	0.60
Cd83	CD83 antigen	0.60
Skap2	src family associated phosphoprotein 2	0.60
Susd3	sushi domain containing 3	0.59
Iqgap1	IQ motif containing GTPase activating protein 1	0.59
3110049J23Rik	RIKEN cDNA 3110049J23 gene	0.59
AW112010	expressed sequence AW112010	0.59
Lect1	leukocyte cell derived chemotaxin 1	0.59

Gnao1	guanine nucleotide binding protein, alpha O	0.59
Gltp	glycolipid transfer protein	0.58
Ptgs1	prostaglandin-endoperoxide synthase 1	0.58
Slc9a3r1	solute carrier family 9 (sodium/hydrogen exchanger), member 3 regulator 1	0.58
Cmtm6	CKLF-like MARVEL transmembrane domain containing 6	0.58
Itga6	integrin alpha 6	0.57
Serpine2	serine (or cysteine) peptidase inhibitor, clade E, member 2	0.57
Gsdmd	gasdermin D	0.57
Lamb2	laminin, beta 2	0.57
Liph	lipase, member H	0.57
Serpinf1	serine (or cysteine) peptidase inhibitor, clade F, member 1	0.57
Necap2	NECAP endocytosis associated 2	0.57
Foxd1	forkhead box D1	0.56
Socs1	suppressor of cytokine signaling 1	0.56
Sept11	septin 11	0.56
Ccrl2	chemokine (C-C motif) receptor-like 2	0.56
Nfe2l2	nuclear factor, erythroid derived 2, like 2	0.56
Scarf2	scavenger receptor class F, member 2	0.55
Hsd3b7	hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 7	0.55
Gnai2	guanine nucleotide binding protein (G protein), alpha inhibiting 2	0.55
Ppnr	per-pentamer repeat gene	0.55
Sepn1	selenoprotein N, 1	0.55
Plcg2	phospholipase C, gamma 2	0.55
Myd88	myeloid differentiation primary response gene 88	0.55
Stat2	signal transducer and activator of transcription 2	0.55
Rhog	ras homolog gene family, member G	0.55
Ptprf	protein tyrosine phosphatase, receptor type, F	0.54
Amz1	archaelysin family metallopeptidase 1	0.54
Cyth4	cytohesin 4	0.54
Hpn	hepsin	0.54
Arhgap18	Rho GTPase activating protein 18	0.54
Elavl1	ELAV (embryonic lethal, abnormal vision, Drosophila)-like 1 (Hu antigen R)	0.54
Slc39a1	solute carrier family 39 (zinc transporter), member 1	0.54
Tprn	taperin	0.54
Aqp4	aquaporin 4	0.54
Ctsl	cathepsin L	0.54
Shisa5	shisa homolog 5 (Xenopus laevis)	0.54
S100a1	S100 calcium binding protein A1	0.54
Fgfbp1	fibroblast growth factor binding protein 1	0.53
Cnn3	calponin 3, acidic	0.53

Inppl1	inositol polyphosphate phosphatase-like 1	0.53
Sgpl1	sphingosine phosphate lyase 1	0.53
Ppapdc1a	phosphatidic acid phosphatase type 2 domain containing 1A	0.53
Cdk2	cyclin-dependent kinase 2	0.53
Lrrc33	leucine rich repeat containing 33	0.52
Cdk5rap2	CDK5 regulatory subunit associated protein 2	0.52
Cyp4v3	cytochrome P450, family 4, subfamily v, polypeptide 3	0.52
Apod	apolipoprotein D	0.52
Rab13	RAB13, member RAS oncogene family	0.52
Fam129b	family with sequence similarity 129, member B	0.52
AA388235	expressed sequence AA388235	0.52
Scpep1	serine carboxypeptidase 1	0.52
Hist1h1c	histone cluster 1, H1c	0.51
Nfatc1	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 1	0.51
Rab31	RAB31, member RAS oncogene family	0.51
Slc7a11	solute carrier family 7 (cationic amino acid transporter, y+ system), member 11	0.51
S100a16	S100 calcium binding protein A16	0.51
St6gal1	beta galactoside alpha 2,6 sialyltransferase 1	0.51
Cspg5	chondroitin sulfate proteoglycan 5	0.51
Htra3	HtrA serine peptidase 3	0.50
S100a13	S100 calcium binding protein A13	0.50
Triobp	TRIO and F-actin binding protein	0.50
Tirap	toll-interleukin 1 receptor (TIR) domain-containing adaptor protein	0.50
Rbms2	RNA binding motif, single stranded interacting protein 2	0.50
Atp1a2	ATPase, Na+/K+ transporting, alpha 2 polypeptide	0.50
Scrg1	scrapie responsive gene 1	0.50
P2ry1	purinergic receptor P2Y, G-protein coupled 1	0.50
Cklf	chemokine-like factor	0.50
Vasp	vasodilator-stimulated phosphoprotein	0.50
Sft2d2	SFT2 domain containing 2	0.50
Plcb3	phospholipase C, beta 3	0.50
Slc7a3	solute carrier family 7 (cationic amino acid transporter, y+ system), member 3	0.49
Ptbp1	polypyrimidine tract binding protein 1	0.49
Rabgap1l	RAB GTPase activating protein 1-like	0.49
Cd151	CD151 antigen	0.49
Suclg2	succinate-Coenzyme A ligase, GDP-forming, beta subunit	0.49
Myo10	myosin X	0.49
Dock1	dedicator of cytokinesis 1	0.49
Vsig4	V-set and immunoglobulin domain containing 4	0.49
Gria2	glutamate receptor, ionotropic, AMPA2 (alpha 2)	0.49

Ucma	upper zone of growth plate and cartilage matrix associated	0.48
Igdcc4	immunoglobulin superfamily, DCC subclass, member 4	0.48
Timeless	timeless homolog (Drosophila)|protein timeless homolog	0.48
Rab7l1	RAB7, member RAS oncogene family-like 1	0.48
Hyi	hydroxypyruvate isomerase homolog (E. coli)	0.48
Vcam1	vascular cell adhesion molecule 1	0.48
Gpr56	G protein-coupled receptor 56	0.48
Lgals1	lectin, galactose binding, soluble 1	0.48
Zcwpw1	zinc finger, CW type with PWWP domain 1	0.48
Scara3	scavenger receptor class A, member 3	0.47
Adam17	a disintegrin and metallopeptidase domain 17	0.47
Sh3glb1	SH3-domain GRB2-like B1 (endophilin)	0.47
Fastk	Fas-activated serine/threonine kinase	0.47
Ddah1	dimethylarginine dimethylaminohydrolase 1	0.47
Kank2	KN motif and ankyrin repeat domains 2	0.47
Tspan4	tetraspanin 4	0.47
Phka1	phosphorylase kinase alpha 1	0.47
Entpd2	ectonucleoside triphosphate diphosphohydrolase 2	0.47
Itih3	inter-alpha trypsin inhibitor, heavy chain 3	0.47
Npc2	Niemann Pick type C2	0.47
Spc25	SPC25, NDC80 kinetochore complex component, homolog (S. cerevisiae)	0.47
Bak1	BCL2-antagonist/killer 1	0.47
Sft2d2	SFT2 domain containing 2	0.47
Tsc22d4	TSC22 domain family, member 4	0.46
Rhou	ras homolog gene family, member U	0.46
Ripk1	receptor (TNFRSF)-interacting serine-threonine kinase 1	0.46
Il13ra1	interleukin 13 receptor, alpha 1	0.46
Zfp521	zinc finger protein 521	0.46
Tmod3	tropomodulin 3	0.46
Phactr4	phosphatase and actin regulator 4	0.46
Prc1	protein regulator of cytokinesis 1	0.46
Lpcat2	lysophosphatidylcholine acyltransferase 2	0.46
Fcgrt	Fc receptor, IgG, alpha chain transporter	0.46
Csf2ra	colony stimulating factor 2 receptor, alpha, low-affinity (granulocyte-macrophage)	0.46
Angptl6	angiopoietin-like 6	0.45
Yap1	yes-associated protein 1	0.45
Itpripl2	inositol 1,4,5-triphosphate receptor interacting protein-like 2	0.45
Zcchc24	zinc finger, CCHC domain containing 24	0.45
Ssfa2	sperm specific antigen 2	0.45
Fli1	Friend leukemia integration 1	0.45

Fmnl3	formin-like 3	0.45
Mertk	c-mer proto-oncogene tyrosine kinase	0.45
Rassf2	Ras association (RalGDS/AF-6) domain family member 2	0.45
Ttyh2	tweety homolog 2 (Drosophila)|protein tweety homolog 2-like	0.44
Map3k14	mitogen-activated protein kinase kinase kinase 14	0.44
Gjb6	gap junction protein, beta 6	0.44
Id3	inhibitor of DNA binding 3	0.44
2310014H01Rik	RIKEN cDNA 2310014H01 gene	0.44
Mfng	MFNG O-fucosylpeptide 3-beta-N-acetylglucosaminyltransferase	0.44
X99384	cDNA sequence X99384	0.44
Heatr5a	HEAT repeat containing 5A	0.44
Sun2	Sad1 and UNC84 domain containing 2	0.44
Dapp1	dual adaptor for phosphotyrosine and 3-phosphoinositides 1	0.44
Vcl	vinculin	0.43
Atf5	activating transcription factor 5	0.43
Cotl1	coactosin-like 1 (Dictyostelium)	0.43
Sepp1	selenoprotein P, plasma, 1	0.43
Grin2c	glutamate receptor, ionotropic, NMDA2C (epsilon 3)	0.43
Tmem184b	transmembrane protein 184b	0.43
Hist2h3c2-ps	histone cluster 2, H3c2, pseudogene	0.43
Ccna2	cyclin A2	0.43
Srpk3	serine/arginine-rich protein specific kinase 3	0.43
Samd4	sterile alpha motif domain containing 4	0.43
Abca9	ATP-binding cassette, sub-family A (ABC1), member 9	0.43
Atp13a4	ATPase type 13A4	0.43
2310007A19Rik	RIKEN cDNA 2310007A19Rik	0.43
Adipor2	adiponectin receptor 2	0.42
Rcan3	regulator of calcineurin 3	0.42
Gm2a	GM2 ganglioside activator protein	0.42
Rhoc	ras homolog gene family, member C	0.42
Cyp4f14	cytochrome P450, family 4, subfamily f, polypeptide 14	0.42
Sirpa	signal-regulatory protein alpha	0.42
Myo6	myosin VI	0.42
Nagpa	N-acetylglucosamine-1-phosphodiester alpha-N-acetylglucosaminidase	0.42
Acot11	acyl-CoA thioesterase 11	0.41
Cxadr	coxsackie virus and adenovirus receptor	0.41
Tead1	TEA domain family member 1	0.41
Rgs20	regulator of G-protein signaling 20	0.40
Stk10	serine/threonine kinase 10	0.40
Gba	glucosidase, beta, acid	0.40
Sept8	septin 8	0.40

Bmpr1b	bone morphogenetic protein receptor, type 1B	0.40
Klc1	kinesin light chain 1	0.40
Fubp1	far upstream element (FUSE) binding protein 1	0.40
Psd2	pleckstrin and Sec7 domain containing 2	0.40
Erbb2ip	Erbb2 interacting protein	0.40
Ctsb	cathepsin B	0.39
Tpcn1	two pore channel 1	0.39
Cd24a	CD24a antigen	0.39
Ly6e	lymphocyte antigen 6 complex, locus E	0.39
Notch1	Notch gene homolog 1 (Drosophila)	0.38
Fbxw4	F-box and WD-40 domain protein 4	0.38
Arrdc3	arrestin domain containing 3	0.38
Itsn1	intersectin 1 (SH3 domain protein 1A)	0.38
Reln	reelin	0.38
Slc20a1	solute carrier family 20, member 1	0.38
Slc44a2	solute carrier family 44, member 2	0.38
Glb1	galactosidase, beta 1	0.38
Itfg3	integrin alpha FG-GAP repeat containing 3	0.37
Tst	thiosulfate sulfurtransferase, mitochondrial	0.37
Dcx	doublecortin	0.36
Tns3	tensin 3	0.36
Dazap2	DAZ associated protein 2	0.36
Nln	neurolysin (metallopeptidase M3 family)	0.36
Ednrb	endothelin receptor type B	0.36
Carhsp1	calcium regulated heat stable protein 1	0.36
Lrrc8a	leucine rich repeat containing 8A	0.35
Plekhb1	pleckstrin homology domain containing, family B (evectins) member 1	0.35
Aldh1a1	aldehyde dehydrogenase family 1, subfamily A1	0.35
Trf	transferrin	0.35
Wsb1	WD repeat and SOCS box-containing 1	0.35
Dhrs1	dehydrogenase/reductase (SDR family) member 1	0.35
Kcnj10	potassium inwardly-rectifying channel, subfamily J, member 10	0.35
Mrps6	mitochondrial ribosomal protein S6	0.35
Rela	v-rel reticuloendotheliosis viral oncogene homolog A (avian)	0.34
Heatr7a	HEAT repeat containing 7A	0.34
Ndp	Norrie disease (pseudoglioma) (human)	0.34
Pcsk6	proprotein convertase subtilisin/kexin type 6	0.34
Klc4	kinesin light chain 4	0.34
Asph	aspartate-beta-hydroxylase	0.34
Ankrd44	ankyrin repeat domain 44	0.34
Mcl1	myeloid cell leukemia sequence 1	0.34

Slc1a3	solute carrier family 1 (glial high affinity glutamate transporter), member 3	0.34
Inpp5a	inositol polyphosphate-5-phosphatase A	0.34
Appbp2	amyloid beta precursor protein (cytoplasmic tail) binding protein 2	0.33
Bcas1	breast carcinoma amplified sequence 1	0.33
Bcan	brevican	0.33
Paox	polyamine oxidase (exo-N4-amino)	0.32
Rlbp1	retinaldehyde binding protein 1	0.32
Pla2g15	phospholipase A2, group XV	0.32
Mfap3l	microfibrillar-associated protein 3-like	0.32
Cdc42	cell division cycle 42 homolog (S. cerevisiae)	0.32
Cav1	caveolin 1, caveolae protein	0.32
Ak3	adenylate kinase 3	0.31
Uap1l1	UDP-N-acteylglucosamine pyrophosphorylase 1-like 1	0.31
Dbp	D site albumin promoter binding protein	0.31
Inf2	inverted formin, FH2 and WH2 domain containing	0.30
Ephx2	epoxide hydrolase 2, cytoplasmic	0.30
Patl1	protein associated with topoisomerase II homolog 1 (yeast)	0.30
St3gal4	ST3 beta-galactoside alpha-2,3-sialyltransferase 4	0.30
Capn2	calpain 2	0.30
Megf10	multiple EGF-like-domains 10	0.29
D19Wsu162e	DNA segment, Chr 19, Wayne State University 162, expressed	0.29
Noxo1	NADPH oxidase organizer 1	0.29
Oat	ornithine aminotransferase	0.29
Phc3	polyhomeotic-like 3 (Drosophila)	0.29
Snx5	sorting nexin 5	0.28
Appl2	adaptor protein, phosphotyrosine interaction, PH domain and leucine zipper containing 2	0.28
Aaas	achalasia, adrenocortical insufficiency, alacrimia	0.26
Samd14	sterile alpha motif domain containing 14|sterile alpha motif domain-containing protein 14-like	0.26

Appendix I: Listed above are the differentially-expressed genes in hippocampi in the OGT cKO mice compared to their WT littermates at 2 months of age. Reported *P*-values are Bonferroni corrected (*P*-values < 0.001).

Appendix II: Downregulated genes in OGT cKO hippocampi at 2 months

Gene Symbol	Description	log2 FC
Creld1	cysteine-rich with EGF-like domains 1	-0.23
Adss	adenylosuccinate synthetase, non muscle	-0.24
Rps6kl1	ribosomal protein S6 kinase-like 1	-0.25
Nkrf	NF-kappaB repressing factor	-0.25
Tmem145	transmembrane protein 145	-0.25
Trappc1	trafficking protein particle complex 1	-0.26
Dbc1	deleted in bladder cancer 1 (human)	-0.26
Prosapip1	ProSAPiP1 protein	-0.26
Porcn	porcupine homolog (Drosophila)	-0.26
Aqp11	aquaporin 11	-0.28
Hecw1	HECT, C2 and WW domain containing E3 ubiquitin protein ligase 1	-0.28
Ankmy2	ankyrin repeat and MYND domain containing 2	-0.29
Hk1	hexokinase 1	-0.29
Sult4a1	sulfotransferase family 4A, member 1	-0.29
Fam174a	family with sequence similarity 174, member A	-0.30
Hmg20a	high mobility group 20A	-0.30
0610007P14Rik	RIKEN cDNA 0610007P14 gene	-0.30
Sirt5	sirtuin 5 (silent mating type information regulation 2 homolog) 5 (S. cerevisiae)	-0.30
Ppapdc2	phosphatidic acid phosphatase type 2 domain containing 2	-0.30
A930017M01Rik	Smg-5 homolog, nonsense mediated mRNA decay factor pseudogene	-0.30
Enox2	ecto-NOX disulfide-thiol exchanger 2	-0.31
Zfp235	zinc finger protein 235	-0.31
Rin1	Ras and Rab interactor 1	-0.31
Cdhr1	cadherin-related family member 1	-0.31
Napepld	N-acyl phosphatidylethanolamine phospholipase D	-0.31
Kbtbd7	kelch repeat and BTB (POZ) domain containing 7	-0.32
Aimp2	aminoacyl tRNA synthetase complex-interacting multifunctional protein 2	-0.32
Ypel1	yippee-like 1 (Drosophila)	-0.32
Slc35f4	solute carrier family 35, member F4	-0.32
Fam164a	family with sequence similarity 164, member A	-0.32
Gzf1	GDNF-inducible zinc finger protein 1	-0.33
Rgmb	RGM domain family, member B	-0.33
Abhd6	abhydrolase domain containing 6	-0.33
Mpp3	membrane protein, palmitoylated 3 (MAGUK p55 subfamily member 3)	-0.33
Ydjc	YdjC homolog (bacterial)	-0.33
Ankrd33b	ankyrin repeat domain 33B	-0.33
Pcdhb12	protocadherin beta 12	-0.33

Mcrs1	microspherule protein 1|microspherule protein 1-like	-0.34
Zfp94	zinc finger protein 94	-0.34
Arhgef25	Rho guanine nucleotide exchange factor (GEF) 25	-0.34
Copg2	coatomer protein complex, subunit gamma 2	-0.34
Esd	esterase D/formylglutathione hydrolase	-0.34
AI593442	expressed sequence AI593442	-0.34
Icam5	intercellular adhesion molecule 5, telencephalin	-0.35
Strbp	spermatid perinuclear RNA binding protein	-0.35
Nap1l3	nucleosome assembly protein 1-like 3	-0.35
Sept6	septin 6	-0.35
AW209491	expressed sequence AW209491	-0.35
Acss2	acyl-CoA synthetase short-chain family member 2	-0.35
A330050F15Rik	RIKEN cDNA A330050F15 gene	-0.36
Dffa	DNA fragmentation factor, alpha subunit	-0.36
Arhgap20	Rho GTPase activating protein 20	-0.36
Stxbp5l	syntaxin binding protein 5-like	-0.36
Gspt2	G1 to S phase transition 2	-0.36
2200002K05Rik	RIKEN cDNA 2200002K05 gene	-0.36
Exoc6	exocyst complex component 6	-0.37
Fam92a	family with sequence similarity 92, member A	-0.37
Tars2	threonyl-tRNA synthetase 2, mitochondrial (putative)	-0.37
Ppapdc3	phosphatidic acid phosphatase type 2 domain containing 3	-0.37
Tram1l1	translocation associated membrane protein 1-like 1	-0.37
Dnajc4	DnaJ (Hsp40) homolog, subfamily C, member 4	-0.37
Ccdc30	coiled-coil domain containing 30	-0.37
Txndc9	thioredoxin domain containing 9	-0.37
Lrrc20	leucine rich repeat containing 20	-0.37
Papolg	poly(A) polymerase gamma	-0.37
Myadml2	myeloid-associated differentiation marker-like 2	-0.37
Rundc3b	RUN domain containing 3B	-0.38
Ola1	Obg-like ATPase 1	-0.38
Fam57b	family with sequence similarity 57, member B	-0.38
Tmem8b	transmembrane protein 8B	-0.38
Adcy1	adenylate cyclase 1	-0.38
Prkcc	protein kinase C, gamma	-0.38
Syt16	synaptotagmin XVI	-0.38
Lmbr1	limb region 1	-0.38
Rpl23	ribosomal protein L23	-0.38
Inpp4b	inositol polyphosphate-4-phosphatase, type II	-0.39
Gm11818	predicted gene 11818	-0.39
Pfkp	phosphofructokinase, platelet	-0.39

Fuk	fucokinase	-0.39
Zfp27	zinc finger protein 27	-0.39
Rasd2	RASD family, member 2	-0.39
Zfand2b	zinc finger, AN1 type domain 2B	-0.39
Mkx	mohawk homeobox	-0.39
Brms1l	breast cancer metastasis-suppressor 1-like	-0.39
B3galnt1	UDP-GalNAc:betaGlcNAc beta 1,3-galactosaminyltransferase, polypeptide 1	-0.40
Pcdhb9	protocadherin beta 9	-0.40
Gpi1	glucose phosphate isomerase 1	-0.40
Rerg	RAS-like, estrogen-regulated, growth-inhibitor	-0.40
Zmat2	zinc finger, matrin type 2	-0.40
Krt10	keratin 10	-0.40
Yrdc	yrdC domain containing (E.coli)	-0.40
Pacrgl	PARK2 co-regulated-like	-0.41
Tdrkh	tudor and KH domain containing protein	-0.41
Dapk3	death-associated protein kinase 3	-0.41
Fam131b	family with sequence similarity 131, member B	-0.41
Fam116a	family with sequence similarity 116, member A	-0.41
Cep70	centrosomal protein 70	-0.41
Fam120b	family with sequence similarity 120, member B	-0.42
Zfp799	zinc finger protein 799	-0.42
Wbscr27	Williams Beuren syndrome chromosome region 27 (human)	-0.42
Reep6	receptor accessory protein 6	-0.42
B3galt4	UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase, polypeptide 4	-0.42
Mas1	MAS1 oncogene	-0.42
4930572J05Rik	RIKEN cDNA 4930572J05 gene	-0.42
Fbxl2	F-box and leucine-rich repeat protein 2	-0.43
Snx19	sorting nexin 19	-0.43
N6amt1	N-6 adenine-specific DNA methyltransferase 1 (putative)	-0.43
Dcbld1	discoidin, CUB and LCCL domain containing 1	-0.43
Col11a1	collagen, type XI, alpha 1	-0.44
Fkrp	fukutin related protein	-0.44
Nupl1	nucleoporin like 1	-0.44
Fam84a	family with sequence similarity 84, member A	-0.45
Srebf2	sterol regulatory element binding factor 2	-0.45
Nanos1	nanos homolog 1 (Drosophila)	-0.45
Acat2	acetyl-Coenzyme A acetyltransferase 2	-0.45
Galnt13	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 13	-0.46
Nov	nephroblastoma overexpressed gene	-0.46
Bnip3	BCL2/adenovirus E1B interacting protein 3	-0.47

Zfp940	zinc finger protein 940	-0.47
2810032G03Rik	RIKEN cDNA 2810032G03 gene	-0.47
Mrps33	mitochondrial ribosomal protein S33	-0.48
Pcdhb17	protocadherin beta 17	-0.48
3830406C13Rik	RIKEN cDNA 3830406C13 gene	-0.48
Wdr17	WD repeat domain 17	-0.48
Dhcr7	7-dehydrocholesterol reductase	-0.48
Sstr3	somatostatin receptor 3	-0.48
Mpped2	metallophosphoesterase domain containing 2	-0.48
1700001L05Rik	RIKEN cDNA 1700001L05 gene	-0.48
AW209491	expressed sequence AW209491	-0.49
Gp1bb	glycoprotein Ib, beta polypeptide	-0.49
Mast3	microtubule associated serine/threonine kinase 3	-0.49
Adnp2	ADNP homeobox 2	-0.49
Rab26	RAB26, member RAS oncogene family	-0.50
E130012A19Rik	RIKEN cDNA E130012A19 gene	-0.50
Agbl4	ATP/GTP binding protein-like 4	-0.50
Tarsl2	threonyl-tRNA synthetase-like 2	-0.51
Ptpn3	protein tyrosine phosphatase, non-receptor type 3	-0.51
Gm9885	predicted gene 9885	-0.51
Ccdc106	coiled-coil domain containing 106	-0.51
Nkrf	NF-kappaB repressing factor	-0.51
Hmgn3	high mobility group nucleosomal binding domain 3	-0.51
Zfp239	zinc finger protein 239	-0.51
Smoc2	SPARC related modular calcium binding 2	-0.51
Ankrd37	ankyrin repeat domain 37	-0.52
Rps16	ribosomal protein S16	-0.52
D4Bwg0951e	DNA segment, Chr 4, Brigham & Women's Genetics 0951 expressed	-0.52
Fbxo27	F-box protein 27	-0.52
Zfp277	zinc finger protein 277	-0.52
Zfp189	zinc finger protein 189	-0.53
Polr2k	polymerase (RNA) II (DNA directed) polypeptide K	-0.53
Cdh9	cadherin 9	-0.53
Plekha2	pleckstrin homology domain-containing, family A (phosphoinositide binding specific) member 2	-0.54
Neurod6	neurogenic differentiation 6	-0.54
Utp23	UTP23, small subunit (SSU) processome component, homolog (yeast)	-0.54
Hspa1a	heat shock protein 1A	-0.54
Pex11b	peroxisomal biogenesis factor 11 beta	-0.55
Ftl1	ferritin light chain 1	-0.55
Tmem29	transmembrane protein 29	-0.55

Rnf8	ring finger protein 8	-0.55
Sebox	SEBOX homeobox	-0.55
2900060B14Rik	RIKEN cDNA 2900060B14 gene	-0.55
Sertad4	SERTA domain containing 4	-0.55
Shroom2	shroom family member 2	-0.56
Insig1	insulin induced gene 1	-0.56
3632451O06Rik	RIKEN cDNA 3632451O06 gene	-0.57
Gnas	GNAS (guanine nucleotide binding protein, alpha stimulating) complex locus	-0.57
Kcnj3	potassium inwardly-rectifying channel, subfamily J, member 3	-0.57
Zfp108	zinc finger protein 108	-0.57
Gls	glutaminase	-0.57
Rapgef4	Rap guanine nucleotide exchange factor (GEF) 4	-0.57
Fam19a1	family with sequence similarity 19, member A1	-0.58
Grm8	glutamate receptor, metabotropic 8	-0.60
Eif4a2	eukaryotic translation initiation factor 4A2	-0.61
Ripply2	ripply2 homolog (zebrafish)	-0.61
Fxr1	fragile X mental retardation gene 1, autosomal homolog	-0.63
A830031A19Rik	RIKEN cDNA A830031A19 gene	-0.63
Pgam2	phosphoglycerate mutase 2	-0.63
Arhgap32	Rho GTPase activating protein 32	-0.63
Lrrc10b	leucine rich repeat containing 10B	-0.63
Pcdhb16	protocadherin beta 16	-0.63
Adamtsl2	ADAMTS-like 2	-0.64
Sebox	SEBOX homeobox	-0.64
Ccdc85a	coiled-coil domain containing 85A	-0.64
Ralyl	RALY RNA binding protein-like	-0.64
Ccdc32	coiled-coil domain containing 32	-0.64
Cpne9	copine family member IX	-0.65
Ywhag	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, gamma polypeptide	-0.65
Hsd17b7	hydroxysteroid (17-beta) dehydrogenase 7	-0.65
Glt8d2	glycosyltransferase 8 domain containing 2	-0.66
Rwdd2a	RWD domain containing 2A	-0.66
Kbtbd3	kelch repeat and BTB (POZ) domain containing 3	-0.66
Acot5	acyl-CoA thioesterase 5	-0.68
Glyctk	glycerate kinase	-0.68
Fdft1	farnesyl diphosphate farnesyl transferase 1	-0.68
Trpc6	transient receptor potential cation channel, subfamily C, member 6	-0.69
Kcnk4	potassium channel, subfamily K, member 4	-0.69
Limd2	LIM domain containing 2	-0.70
Dusp4	dual specificity phosphatase 4	-0.70

Adra1d	adrenergic receptor, alpha 1d	-0.71
Arhgap15	Rho GTPase activating protein 15	-0.72
Chst9	carbohydrate (N-acetylgalactosamine 4-0) sulfotransferase 9	-0.74
Rwdd2a	RWD domain containing 2A	-0.75
Acot3	acyl-CoA thioesterase 3	-0.76
Cyp51	cytochrome P450, family 51	-0.76
Zcchc5	zinc finger, CCHC domain containing 5	-0.77
Mvd	mevalonate (diphospho) decarboxylase	-0.78
Dtd1	D-tyrosyl-tRNA deacylase 1 homolog (S. cerevisiae)	-0.80
Cyp51	cytochrome P450, family 51	-0.82
Stard4	StAR-related lipid transfer (START) domain containing 4	-0.86
Tm7sf2	transmembrane 7 superfamily member 2	-0.88
Sc4mol	sterol-C4-methyl oxidase-like	-0.95
Capn3	calpain 3	-0.96
Scgn	secretagogin, EF-hand calcium binding protein	-1.1
Cldn22	claudin 22	-1.1
Icam4	intercellular adhesion molecule 4, Landsteiner-Wiener blood group	-1.1
Fdps	farnesyl diphosphate synthetase	-1.2
Tdo2	tryptophan 2,3-dioxygenase	-1.2
Idi1	isopentenyl-diphosphate delta isomerase	-1.2

Appendix II: Listed above are the downregulated genes in hippocampi in the OGT cKO mice compared to their WT littermates at 2 months of age. Reported *P*-values are Bonferroni corrected (*P*-values < 0.001)

Appendix III- qPCR Primers

Gene	#	Forward Primer	Reverse Primer
2010111I 01Rik	5	GGATGATCTGCCTCTCATGGC	AAGAAAAGCACTATGTTGCCCT
Actr5	5	GGAGGCTCGATGCCAAAAA	CATGCGGCTGAGTGTGATG
Advyap1	9	CTCGGACGGCATCTTCACA	CCTCTGTTTATACCTTTTCCCTAG CA
Apod	4	TCCCACCCTCTTCAAGACAC	GGTTCTGGCTTGCTACTGG
Apoe	4	TGAACCGCTTCTGGGATTAC	TGTGTGACTTGGGAGCTCTG
Arc	1	TGGAGCAGCTTATCCAGAGG	TATTCAGGCTGGGTCCTGTC
B2m	3	ATGGCTCGCTCGGTGACCCT	TTCTCCGGTGGGTGGCGTGA
Bcl6	3	TGAGGTCGTGGAGAACAATATG	GAGATGGCTGTACATGGGATAAG
Bdnf	1	CAGAGCAGCTGCCTTGATGTT	GCCTTGTCCGTGGACGTTTA
Brsk2	5	ACCTGCTGCTAGATGAGAGGA	CTCGCCCCGAATCACTTCC
C1qa	5	CCAGGAGAGTCCATACCAGAA	GTCCCACTTGGAGATCACTTG
C1qb	3	TGAGCCACGCAACGGCAAGT	TCGCGGCCACGAACGAGATT
C1qc	3	CAGAGGCCAACGCCCTCGTC	AGAGGCCCGGCACTTCACAG
C3	3	ACCCCCGATGGCATTCCTGTCA	GCTCGGATCTTCCACTGCCCC
Cd14	3	ATTCGGAGCCCTGCGTGTGC	ATCGGGTCCGGTGGCTTCCA
Cd52	5	ATCCTTGGGACAAGCCACTAC	GGCACATTAAGGTATTGGCAAAG
Cd63	4	TAACTGTGGGCTGTGGGAAT	TCCCAAGACCTCCACAAAAG
Cdk5	3	CACCGTTGGTGCAATGAATAC	TACACTAGGCTTACCTCCTACC
Cnot3	5	TGAGATCAAGGACAAAAGGCAG	CACAACTTTGAACCGTTCCATTT
Cre	11	TTGGGCCAGCTAAACATGCT	GCATTGCTGTCACTTGGTCG
Creb1 (α/β)	4	ACTAAATGACCATGGAATCTGGA G	ACCTGGGCTAATGTGGCAAT
Creb1 (β)	4	AGTTATCCAGTCTCCACAAGTCC	CTGAGTTCCGGAGAAAAGTCTT
Crem	8	CAGAGGAAGAAGGGACACCA	TTGTATTGCCCCGTGCTAGT
Ctnnb1	1	TTATGGACTGCCTGTTGTGG	AGTCGTGGAATAGCACCCTG
Ctsd	3	CTCCCGGCGTCTTGCTGCTC	AGCCGCCCACCTCCGTCATA
Ctsh	4	GAGATGGGGACAGACAGGAA	CCTGTGGCCATTACACTCCT
Cyba	4	GTGGACTCCCATTGAGCCTA	CTCCTCTTCACCCTCACTCG
Egr1	4	GACGAGTTATCCCAGCCAAA	GGTTCAGGCCACAAAGTGTT
Egr2	1	TGCGGGCATCTTGCAAGGGG	GTCCCGTGGCCAGTGGGTTG
Fcgr1	4	CCCCAAACTCCCACACTCTA	AAAGCTCTCACCACCCAATG
Fdps	4	CTGAGAAGGAGCTGGGACAC	CCGGTTGTACTTGCCTCCTA
Fos	1	CCGACTCCTTCTCCAGCAT	TCACCGTGGGGATAAAGTTG
Gap43	4	GCTGTAGACGAAGCCAAACC	CAACGTGGAAAGCCATTTCT
Gfp	4	AAGCTGACCCTGAAGTTCATC	CGTCCTGGACGTAGCCTCG
Gfpt1	10	CCTCGTGATGTTTGCTCTCA	GGACCGACTTCTGGTGGTAA
Gfpt2	10	TCGGGGTACGAAGCAAATAC	TAAGAAGATGACCCGGTTGG
Grin2b	4	AGGTCGTTTCCAGAAGGACA	ATTGCTGGAGCCATTGAAAG

Gene	#	Forward Primer	Reverse Primer
Grm1	6	TCCACCTCATAGCCTTCGAT	GGGGGTTACTGAGTGCCAT
Gusb	3	TCCGTTGGCTCGGGGCAAAT	TGCCCACACCGGGACACTCA
Igsf6	4	ACGGTCCAAGAAGCAGAGAA	GTGCCCTCTTGCTCAGGTAG
Il1b	2	TTGTGGCTGTGGAGAAGCTGT	ACCTGCTGGTGTGTGACGTT
Irak1	5	ACTCCAGAGAAGTCCCAACCA	CAGGAATGCAGGGTAGCAGAG
Itgb5	4	CAACGAGGAAGTGAGGAAGC	CCGATCTTCTCCTTGCAGAC
Kif1c	5	GGAGCCTCCGTGAAAGTTG	CCGAAGTATGCGACCAGTAAGA
Lgals3bp	4	ACTGCCCTGGACACCAATAG	GTAGAAGGGGCGTATGACCA
Man2b1	3	CCCGGGCGACACCAAAAACG	ACTGGTGCTCCAAGCGCAGC
Mgea5	10	TGGAAGACCTTGGGTTATGG	TGCTCAGCTTCTTCCACTGA
Nr4a1	1	TGATGTTCCCGCCTTTGC	CAATGCGATTCTGCAGCTCTT
Nr4a3	1	GATCACAGAGCGACATGGGTTA	GAGCCTGTCCCTTCCTCTGG
Ogt	5	TATCAGGCAGGAGATTTTGAGGC	ACACCAGTATTGTCAGGCTCT
Opa3	1	GCAAAGGCAAAAGATGGAAC	GTGTTCACCGAAGGAAGGAG
Parp1	5	GCTTTATCGAGTGGAGTACGC	GGAGGGAGTCCTTGGGAATAC
Pdia4	5	ACAGGCCGTTGACTATGATGG	GTGGAGGTGTCCAATCAGGC
Per1	1	CAGGCTAACCAGGAATATTACCAGC	CACAGCCACAGAGAAGGTGTCCTGG
Pou2f2	5	TCTGTGGATTCAAGCCTACCC	GGACATTCGTCTCGATGCTG
Rttn	5	TGGGATGCAGTGTCTTCCAG	GTGCCACGTCCAAAGGAGA
Spp1	3	CCCGGTGAAAGTGACTGATTCTGGC	GGGTCAGGCACCAGCCATGTG
Syn1	4	CTGAGCCCTTCATTGATGCT	GGTCTTCCAGTTACCCGACA
Synpo2	4	CCAGTGAGTCGGAAGTGGAT	TGTGGTGTCTGGCAACATTT
Synt	4	CATCAAGCTGGAGAACAGCA	CGCGTGTTCCACATTGTACT
Syp	4	CTCCTCGGCTGAATTCTTTG	CATTGGCCCTTTGTTGTTCT
Syt1	4	ATGCAGAACGGCAAGAGACT	CTCGAACGGAACTTCAAAGC
Tlr1	5	TGAGGGTCCTGATAATGTCCTAC	AGAGGTCCAAATGCTTGAGGC
Trem2	7	GGGAGCAGGAATACTGGTGT	TTGGTGTGTGGAGAATGTTT
Tyrobp	3	TGAGCCCTGGTGTACTGGCTGG	TTGACCTCGGGAGACCAGGCG
Wnt2	1	CATAGCCCCCCACCACTGT	AGTTCCTTCGCTATGTGATGTTTCT

Reference Primers:

Gene	#	Forward Primer	Reverse Primer
Rpl3	1	TCATTGACACCACCTCCAAA	GCACAAAGTGGTCCTGGAAT
Gapdh	1	CTGAGTATGTCGTGGAGTCTACTGG	GTCATATTTCTCGTGGTTCACACC
Rps17	5	CCGGCTATGTCACGCATCTG	ATGATCTCCTGATCTAGGGCTG
a4Tub	5	ATGCGCGAGTGCATTTCAG	CACCAATGGTCTTATCGCTGG

References (#):

1.   Rexach, et al. Dynamic O-GlcNAc modification regulates CREB-mediated gene expression and memory formation. *Nat Chem Biol*. 2012. 8 (3) 253-61.

2.   Kim, D.; You, B.; Jo, E.; Han, S.; Simon, M.I.; Lee, S.J. NADPH oxidase 2-derived reactive oxygen species in spinal cord microglia contribute to peripheral nerve injury-induced neuropathic pain. *PNAS*. 2010. 107 (33) 14851-14856.

3.   Primer Blast. http://www.ncbi.nlm.nih.gov/tools/primer-blast/

4.   OligoPerfect Designer. https://tools.lifetechnologies.com/content.cfm?pageid=9716&icid=fr-oligo-6?CID=fl-oligoperfect

5.   PrimerBank. http://pga.mgh.harvard.edu/primerbank/

6.   Primer Depot. http://mouseprimerdepot.nci.nih.gov/

7.   Wirz, K.T.S.; Bossers, K.; Stargardt, A.; Kamphuis, W.; Swaab, D.F.; Hol, E.M.; Verhaagen, J. Cortical beta amyloid protein triggers an immune response, but no synaptic changes in the APPswe/PS1dE9 Alzheimer's disease mouse model. *Neurobiology of Aging*. 34 (5) 1328-1342.

8.   Gundersen, B.B.; Briand, L.A.; Onksen, J.L.; Lelay, J.; Kaestner, K.H.; Blendy, J.A. Increased hippocampal neurogenesis and accelerated response to antidepressants in mice with specific deletion of CREB in the hippocampus: role of cAMP response-element modulator tau. *J Neurosci*. 2013. 33 (34) 13673-85.

9.   Wu, Z. L.; Ciallella, J. R.; Flood, D. G.; O'Kane, T. M.; Bozyczko-Coyne, D.; Savage, M. J. Comparative analysis of cortical gene expression in mouse models of Alzheimer's disease. *Neurobiol Aging*. 27 (3) 377-86.

10.   Watson, L. J.; Facundo, H. T.; Ngoh, G. A.; Ameen, M.; Brainard, R. E.; Lemma, K. M.; Long, B. W.; Prabhu, S. D.; Xuan, Y. T.; Jones, S. P. O-linked β-N-acetylglucosamine transferase is indispensable in the failing heart. *PNAS*. 2011. 107 (41) 17797-802.

11.   Zhang, Y.; Huang, G.; Shornick, L.P.; Roswit, W.T.; Shipley, J.M.; Brody, S.L.; Holtzman, M.J. A Transgenic FOXJ1-Cre System for Gene Inactivation in Ciliated Epithelial Cells. *Am J Respir Cell Mol Biol*. 2007; 36(5):515-519.