

## **Chapter 5**

# **Proteome Dynamics: Identification of Newly Synthesized Proteins in Mammalian Cells using Bioorthogonal Non-Canonical Amino Acid Tagging (BONCAT)**

The text in this chapter is from a manuscript by Daniela C. Dieterich, A. James Link, Johannes Graumann, David A. Tirrell, and Erin M. Schuman, submitted July 2005.

**Abstract**

In both normal and pathological states, cells respond rapidly to environmental cues by altering protein synthesis. Current techniques provide little information about proteome dynamics. We describe here a new technology, based on the co-translational introduction of azide groups into proteins and the chemo-selective tagging of azide-labeled proteins with an alkyne-affinity tag, to identify, acutely, the newly synthesized proteins in mammalian cells. Incorporation of the azide-bearing amino acid azidohomoalanine (AHA) is not toxic to cells and does not increase protein degradation. Following tagging of AHA-labeled proteins with an alkyne-affinity tag, we demonstrate the selective affinity-purification and identification of 194 metabolically-labeled proteins with multidimensional liquid chromatography in-line to tandem mass spectrometry (LC-MS/MS). Furthermore, we demonstrate that, in combination with leucine-based mass tagging, candidates can be immediately validated as newly-synthesized proteins. The identified proteins, synthesized in a 2-hour window, possess a broad range of biochemical properties and span most functional gene ontology categories. This technology can be used to address the temporal and spatial characteristics of any newly synthesized proteome in any cell type.

## Introduction

The proteome is a dynamic entity, tightly regulated by protein synthesis and degradation to maintain homeostasis in a cell, tissue or whole organism. Quantitative knowledge of the particular set of proteins expressed in different cellular locales at different times arguably brings one close to a knowledge of a cell's phenotype. Proteomic approaches, including 2D gel electrophoresis,<sup>1,2</sup> isotope-coded affinity tags (ICAT),<sup>3</sup> quantitative proteomic analysis using samples from cells grown in <sup>14</sup>N or <sup>15</sup>N-media,<sup>3</sup> and stable isotope labeling by amino acids in cell culture (SILAC)<sup>4,5</sup> have been developed to compare the protein expression profiles of cells in these different states. There are, however, limitations associated with these approaches. For example, 2D gel electrophoresis possesses a limited ability to detect proteins of low abundance, membrane proteins, or proteins with unusual isoelectric points or molecular weights. ICAT is limited to small sample sizes, and valuable protein information (posttranslational modifications such as phosphorylation or glycosylation) might be lost due the fact that only cysteine-containing peptides are analysed. Recent developments both in instrumentation and analysis of peptide phosphorylation<sup>6-8</sup> and glycosylation<sup>9</sup> by mass spectrometry address this latter issue by elucidating the "phosphoproteome" and "glycoproteome" of a cell. SILAC critically depends on the ideally complete incorporation of a given "light" or "heavy" form of an essential amino acid into two cell pools, a process that is both expensive and time-consuming. All of these approaches lack temporal resolution and the selective enrichment for newly synthesized proteins.

## Results and Discussion

To circumvent these problems we have developed a new approach to identify, specifically, newly synthesized proteins, endowing them with a novel chemical functionality that enables their selective separation from the existing proteome via affinity-purification (Fig 1). Previous experiments in *E. coli* have shown that the methionyl-tRNA synthetase accepts the non-natural amino acid AHA as a substrate and that AHA can be efficiently incorporated into bacterial proteins.<sup>10,11</sup> In a different approach by Cravatt and colleagues, the chemo-selective potential of the azide group was used to label proteins based on their catalytic properties (e.g., activity-based protein profiling, ABPP).<sup>12</sup> We examined first the specificity of AHA and its potential toxicity in mammalian cells. Incubation with methionine was used as a general control. HEK293 cells were incubated for 2 hrs at 37 °C in methionine (4 mM), AHA (4 mM), or saline (see Materials and Methods). Subsequently, propidium iodide staining (to visualize dead cells) revealed no difference in cell viability between AHA and methionine-exposed cells (Fig 2A). In order to visualize their morphology, neurons were infected with a destabilized and myristoylated form of the fluorescent protein EGFP<sup>13</sup> and incubated for 1.5 hours with either AHA or methionine. We found that AHA was not toxic to neurons, as indicated by healthy neuronal processes and the absence of abnormal varicosities in the dendrites (Fig 2B).

To examine whether AHA is incorporated into mammalian proteins we tagged lysates prepared from AHA-treated HEK 293 cells with the biotin-bearing alkyne reagent biotin-PEO-propargylamide.<sup>11</sup> Subsequent Western blot analysis revealed the successful incorporation of AHA (Fig. 2C). To examine the specificity of AHA incorporation into

newly synthesized proteins, HEK293 cells were incubated for just 2 hrs with AHA in the presence or absence of a protein synthesis inhibitor (anisomycin or cycloheximide). After incubation, cells were lysed and subjected to [3+2]-cycloaddition with the alkyne-linker, followed by Western blot analysis with an anti-biotin antibody. While abundant signal was detected in AHA lanes, no signal was detected in either the methionine or the protein synthesis inhibitor lanes (Fig 2C), indicating that this procedure specifically labels newly synthesized proteins. Lastly, no increased ubiquitination was observed in AHA treated cells when compared to buffer or methionine controls (Fig 2D), indicating that the modified amino acid does not cause severe protein misfolding or degradation.

Can we purify AHA-labeled proteins for subsequent shotgun mass spectrometry? For this purpose, we designed an alkyne tag (Fig 3A) with a biotin moiety at the N-terminus followed by the FLAG-antibody epitope covalently linked to propargylglycine. This tag can be cleaved by trypsin to allow direct proteolysis of proteins on the affinity matrix, bypassing the need for an elution step. Following tag cleavage, the mass gain of tagged AHA (over methionine) is 107 AMU, which can be detected in the mass spectrometrical analysis. To achieve an inner validation of the mass spectrometrical data we also administered tenfold deuterated leucine ( $d_{10}$ -Leu) along with AHA. To test the capacity of our technique to detect newly synthesized proteins, we transiently expressed a hemagglutinin-(HA) tagged version of the brain-specific<sup>14,15</sup> mammalian huntingtin-associated protein 1A (HAP1A) in HEK293 cells. Proteins were purified from either AHA- or methionine-treated HEK293 cells expressing HA-HAP1A. While abundant biotin signal was detected in lysates prepared from AHA-treated cells, negligible biotin signal was detected in the lysates prepared from methionine-treated cells (Fig 3B). In

addition, HA-HAP1A was exclusively present in the Neutravidin affinity matrix bound fraction of AHA-treated lysates (Fig 3B). HA-HAP1A contains 14 methionine residues. Thus, a complete exchange of methionine with AHA would lead to a mass gain of 23 kDa. We observed tagged HA-HAP1A migrating as a “smear” about 7 kDa higher than untagged HA-HAP1A, indicating successful incorporation and tagging of up to 4 AHA residues (Fig. 3B). We performed on-resin trypsination of the extensively washed affinity matrix and used these samples in all of the following experiments. Using Western blot analysis we observed that HA-HAP1A immunoreactivity was completely absent after trypsination, and the levels of biotin immunoreactivity were reduced to background levels (Fig 3B).

To identify the newly synthesized, affinity-purified proteins we used multidimensional protein identification technology (MudPIT)<sup>16</sup> (see Materials and Methods). Consistent with our biochemical data (Fig. 3B), peptides for HA-HAP1A were identified in all four AHA experiments but not in the control samples (Fig 4A). Representative fragmentation spectra for a peptide without a modification (Fig. 4B), a peptide with several d<sub>10</sub>-Leu-leucine substitutions (Fig. 4C), and an AHA substitution (Fig. 4D) show that modifications can be reliably detected using AHA/d<sub>10</sub>-Leu mass tagging. Overall, ten different valid peptide sequences for HA-HAP1A were detected, yielding a sequence coverage of 27.3%. Six peptides contained at least one modification (Table S1). The combination of affinity purification of alkyne-tagged AHA-labeled proteins and d<sub>10</sub>-Leu mass tagging provides excellent inner validation of newly synthesized proteins.

We next explored the identity and representation of all HEK 293 cell proteins

synthesized in a 2-hour window by comparing AHA experiments with each other and with control (methionine) experiments, examining the frequency with which individual peptides were identified.<sup>17</sup> A total of 194 unique proteins were identified in AHA samples (Table 1): over half of the proteins were identified in more than one experiment. In spite of the fact that our procedure may be expected to enrich for proteins with large numbers of methionine residues, the methionine content of our 194 candidates is 2.40%, which is only slightly higher than the methionine content of the whole database (2.13%). One candidate identified, Histone H4, possesses only 2 methionine residues in its ORF and yet it was identified by our technique. In control samples, only 2 proteins were identified, and each protein was identified in a single experiment. One identification was shared by AHA and control samples.

To independently confirm our data, a Western blot analysis on two of the identified candidates, EF2 and actin, was performed (Fig. S1). Both proteins were detected in the Neutravidin-bound fraction of AHA- but not of methionine-treated cells, indicating the specificity of the purification and our ability to enrich and identify newly synthesized proteins in a complex mixture of proteins.

To assess the rate of false-positive peptide identifications, we used the methodology of Peng *et al.*<sup>18</sup> Peptide spectra were searched against a sequence database supplemented with a randomized form of every ORF using different sets of search constraints; peptide matches to randomized proteins were considered false-positives. On the single peptide level, we found no matches to the randomized database among the AHA samples (total of 4096 peptides) compared to 6 matches out of a total of 33 peptides for control samples, when a fully tryptic status<sup>19</sup> and a single modified peptide was

required. Using the above peptide data together with a requirement for at least two valid peptides for protein identification, we found no false-positive proteins for either the AHA or control samples (Table S2).

The 193 endogenous HEK293 cell proteins were examined to determine their isoelectric points (pI) and molecular mass. As indicated in Table 2, the isoelectric points of the identified proteins spanned a broad range from 4.6 (Nucleophosmin) to 11.4 (Histone H4). In addition, a broad range of molecular masses, from 11 kDa (Histone H4) to 466 kDa (Splice isoform 2 of DNA-dependent protein kinase catalytic subunit) was also represented in our protein group, showing that this technique can detect proteins with unusual biochemical properties.

Next, we examined the candidates for their biological association to Gene Ontology (GO) categories using GoMiner.<sup>20</sup> Annotations representing the majority of protein categories could be assigned to 158 of the 193 proteins identified (Fig. S2). The proteins identified included soluble cytoplasmic proteins, membrane-associated, and insoluble proteins, as well as proteins that are typically expressed at low levels like transcriptional and translational regulators. Taken together, these data indicate that we can enrich and identify newly synthesized proteins within a 2-hour window that cover a wide range of biochemical properties.

A major endeavor in biology is the comparison of two or more states in biological systems, for example the cancerous vs. non-cancerous state, the addicted vs. non-addicted brain circuit, or the physiology of a genetically altered mouse vs. a wild-type littermate. Currently, changes in the protein composition of a cell can only be identified by differential approaches that compare two proteomes with one another, without enriching

for newly synthesized proteins. All proteins, new and old, share the same limited pool of 20 amino acids and thus are chemically indistinguishable. Here we have described the use of non-canonical amino acids to introduce chemically unique groups to proteins, thus selectively delivering novel functionality to newly synthesized proteins and enabling their subsequent enrichment and identification. The subsequent tandem mass spectrometrical analysis of those affinity-purified protein mixtures revealed the identification of 194 proteins including the overexpressed protein HA-HAP1A. Further analysis of the 194 identified candidates confirmed that a broad and wide-spanning range of proteins was identified, including small and large proteins, protein with low and high isoelectric points, as well as membrane associated proteins. A complete understanding of cellular function requires a dynamic view of the proteome in which one can selectively identify changes in protein synthesis that occur on the time scale of minutes. Temporal resolution on this scale can begin to describe the immediate responses of cells to environmental changes, rather than responses that may be secondary or tertiary. This technique, combined with subcellular fractionation or microdissection of tissue,<sup>21</sup> can provide a comprehensive picture of spatial and temporal aspects of cellular proteomes. Particularly unique to our approach is the selective enrichment for newly synthesized proteins. A series of these “snapshots” promises an unbiased picture of the temporal and spatial dynamics of proteomes.

## **Materials and Methods**

**Reagents.** Inorganic salts, Tris, HEPES, TCEP, iodoacetamide, and detergents were purchased from Sigma. Urea was purchased from ICN; acetonitrile (ACN) and formic

acid were purchased from Mallinckrodt.

**Cell Culture.** HEK293 cells were cultivated in DMEM plus 10% FCS and penicillin / streptomycin (all from Invitrogen). For cell viability assays with propidium iodide, HEK293 cells were grown on poly-D-lysine-coated coverslips, washed with HEPES-buffered saline (HBS; 10 mM HEPES, 119 mM NaCl, 2 mM CaCl<sub>2</sub>, 2 mM MgCl<sub>2</sub>, 5 mM KCl, 30 mM glucose, pH 7.35), and incubated with equimolar concentrations of methionine or AHA in HBS or just HBS for 2 hours. After brief incubation with 3  $\mu$ M propidium iodide (PI, Molecular Probes) in HBS, coverslips were rinsed with HBS and immediately imaged on an Olympus AX70 fluorescence microscope and a Hamamatsu Digital camera through a 10x lens. PI was excited at 568 nm. For measurement of ubiquitination levels, cells were incubated for 2 hours with different concentrations of methionine or AHA, washed briefly with PBS-MC (1 mM MgCl<sub>2</sub>, 0.1 mM CaCl<sub>2</sub> in PBS), and solubilized in 2x SDS sample buffer (0.5% SDS, 20% glycerol, 10% 2-mercaptoethanol, 125 mM Tris pH 6.8, 0.002% Bromophenol Blue) for subsequent Western blot analysis. PolyFect (Qiagen) was used to transiently transfect cells with mmHAP1A-HA plasmid<sup>22</sup> (generous gift from Edoardo Marcora). 16-18 hours post transfection, cells were washed with HBS and incubated for 1.5-2 hours with either 4 mM methionine + 2.86 mM d<sub>10</sub>-Leu or AHA + 2.86 mM d<sub>10</sub>-Leu in HBS. Cells were briefly rinsed in PBS-MC to remove excess AHA and methionine and stored at -80 °C until further use. Dissociated hippocampal neuron cultures were prepared from newborn rat pups (P0) as published previously.<sup>13</sup> In viral infection experiments, neurons were infected with the pSinRep5-5'myrdGFP3' viroids<sup>13</sup> at div 12-14 in growth medium

containing the Sindbis virus. Nine hours post initial infection, neurons were rinsed in HBS and incubated for 2 hours with either 2.86 mM AHA or 4 mM methionine in HEPES-buffered solution. Images were immediately acquired in HBS with an Olympus AX70 fluorescence microscope and a Hamamatsu Digital camera through a 10x lens. GFP was excited at 488 nm.

We tested concentrations of AHA ranging from 1 to 10 mM with best results obtained for 2.86 to 4 mM AHA considering biotin signal strength with no change in the level of ubiquitination (data not shown).

**[3+2] Cycloaddition Chemistry and Purification of Tagged Proteins.** AHA,<sup>23</sup> biotin-PEO-propargylamide,<sup>11</sup> and the triazole ligand<sup>24</sup> were prepared as described previously. The cleavable biotin-FLAG-alkyne tag was synthesized by GenScript Corporation. d<sub>10</sub>-Leu was purchased from Sigma. Cell pellets were lysed in 1% SDS in PBS plus EDTA-free complete protease inhibitors (Roche), with genomic DNA sheared with a syringe and a needle and boiled for 10 min at 96-100°C. Lysates were diluted to 0.2% SDS, 0.2% Triton X-100 in PBS plus complete protease inhibitors before addition of 200 μM triazole ligand, 50 μM alkyne tag, and 75 μg/ml CuBr. The reaction was allowed to proceed for 6 hours at RT and excess reagents were removed by gel filtration through PD-10 columns (Amersham Bioscience). The amount of biotinylated protein was estimated by dot blot analysis with biotinylated BSA (Pierce) as standard and a rabbit anti-biotin antibody (Bethyl Laboratories). Protein extracts were adjusted to 1% Nonidet P40 (Sigma), 0.1% SDS in PBS, pH 7.4 plus complete EDTA-free protease inhibitors and tagged proteins were purified on Immobilized Neutravidin affinity resin (Pierce). After extensive

washing in incubation buffer, followed by washes in 1% Nonidet P40 in PBS as well as 50 mM ammonium bicarbonate buffer, resin suspensions were incubated for 10 min at 70 °C and adjusted to 2 M urea as 25-33 % slurries. After cooling to RT, the suspensions were incubated for 30 min at RT with 3.125 µM TCEP followed by incubation with 15 µM iodoacetamide for 30 min at RT in the dark, both under constant agitation. Proteins were digested on-resin with 0.5 µg/ml endoproteinase Lys-C for 4 hours at 37 °C and trypsin (both Roche) after adjusting to 1 mM CaCl<sub>2</sub> overnight at 37°C. Peptides were separated from resin using Economy Mini-Spin Columns (Pierce). The resulting eluates were adjusted to 3.5% formic acid and centrifuged at 20000 x g, 45 min, 4 °C prior loading onto the LC columns for MS analysis. In a typical experiment 400 µl peptide solution was loaded.

**Western Blot Analysis.** For protein analysis in polyacrylamide gels, proteins were separated in 10% Tris-glycine gels under reducing conditions. For Western blot analysis, proteins in polyacrylamide gels were transferred to PVDF membranes (BioRad). The membranes were blocked with 5% nonfat dry milk in PBS containing 0.1% Tween-20 (T-milk) and incubated with primary antibodies in milk. Secondary antibodies were conjugated with horseradish-peroxidase (HRP) and detection was achieved using the ECL system from Amersham Bioscience. The following primary antibodies were used: rabbit anti-Biotin (1:10000, Bethyl Laboratories), mouse anti-HA (1:1000, Covance), mouse anti-Actin (IgM, 1:10000, Oncogene), rabbit anti-Ubiquitin (1:1000, Stressgen), rabbit anti-eEF2 (1:500, Cell Signaling). Secondary HRP-conjugated antibodies were purchased from Jackson Immuno Research Laboratories (for IgG) or from Calbiochem

(for IgM) and used at 1:10000 in T-milk.

**Mass Spectrometry and Data Analysis.** Analysis of peptide mixtures by MudPIT was done essentially as described in Graumann *et al.*<sup>25</sup> using a HP-1100 quaternary HPLC pump (Agilent, Palo Alto, CA) and a LCQ-DecaXP electrospray ion trap mass spectrometer (ThermoElectron, Palo Alto, CA). Briefly, proteolytically digested samples were separated on a triphasic microcapillary column as described<sup>26</sup> with the following modifications allowing loading of a higher sample volume: 3.5 cm of SCX resin (Partisphere SCX, Whatman, Clifton, NJ) was packed into a 100 µm internal diameter fused silica column (PolyMicro Technology, Phoenix, AZ) with 2.5 cm Aqua C18 reverse phase (RP) material (Phenomenex, Ventura, CA) placed upstream of the SCX. This biphasic column was fitted into an Inline MicroFilter Assembly unit with a 0.5 µm PEEK Frit (all from Upchurch Scientific, Oak Harbor, WA). Samples were loaded directly onto this column using a pressure cell. After sample loading, a column head packed with 6.5 cm Aqua C18 resin was attached at the other end of the MicroFilter Assembly unit, creating a triphasic RP-SCX-RP column. Sample separation was achieved with a six-step chromatography program, and the column eluate was analyzed as described.<sup>25</sup> Centroided fragmentation spectra acquired by Xcalibur 1.3 software (ThermoElectron, Palo Alto, CA) were evaluated for spectrum quality and charge state using 2to3 software<sup>27</sup> and searched against the translated open reading frames of the human IPI database ipi.HUMAN.v3.06.fasta (<http://www.ebi.ac.uk/IPI/IPIhelp.html>) by Sequest, version 27, rev. 9,<sup>28</sup> utilizing unified input and output files.<sup>29</sup> Relevant Sequest parameters used were: 1) peptide tolerance of 3.0 AMU, 2) parent ion masses were

treated as monoisotopic, 3) fragmentation ion masses were treated as averaged, and 4) a 57.0 AMU static modification of cysteines accounted for alkylation. Additionally, to search for modifications due to AHA and d<sub>10</sub>-Leu incorporation into proteins, two sets of dynamic modification searches were performed to increase the confidence of identification of newly synthesized proteins using: +10 AMU for d<sub>10</sub>-Leu and +107.0 AMU in case of successful tagging of methionine-replacing AHA versus +10 AMU for d<sub>10</sub>-Leu and -5.1 AMU in case of failed tagging of AHA. Protein matches from these two sets of searches were compared and unified manually. The majority of identified modifications resulted from incorporation of d<sub>10</sub>-Leu; therefore only minor differences between the two search sets existed.

Sequest results were filtered using DTASelect 1.9 and Contrast<sup>17</sup> with the following requirements for peptide and locus identifications considered valid: minimum XCorrs of 1.8, 2.5, and 3.5 for singly, doubly, and triply charged ions, respectively, a minimum of DeltCN of 0.08, a minimum of two valid peptides per locus, and a minimum of one peptide containing a AHA- or d<sub>10</sub>-Leu-based modification per locus.

To assess the false positive rate, peptide spectra were searched against a randomized sequence database supplemented with peptide matches to randomized proteins considered as false positives. The percentage of false positives was calculated according to Peng *et al.*<sup>18</sup> Using the DTASelect default settings for valid peptide and locus identifications (minimum XCorrs, minimum DeltCN and varying requirements for the tryptic status of a peptide: 1) any status accepted, 2) partially or 3) fully tryptic status required as well as at least one modification present in at least one valid peptide, we determined the mean rates of false positives of the 4 experimental sets.

To classify the identified proteins according to Gene Ontology categories (<http://www.godatabase.org/dev>), the program GoMiner was used.<sup>20</sup> The Python scripts IPIWEBConverter and GeneNameFinder fetched and extracted for each identified protein its gene name and saved it in a text file. This text file was then uploaded into the GoMiner program. Additional information on proteins was obtained from the GeneCards® database.

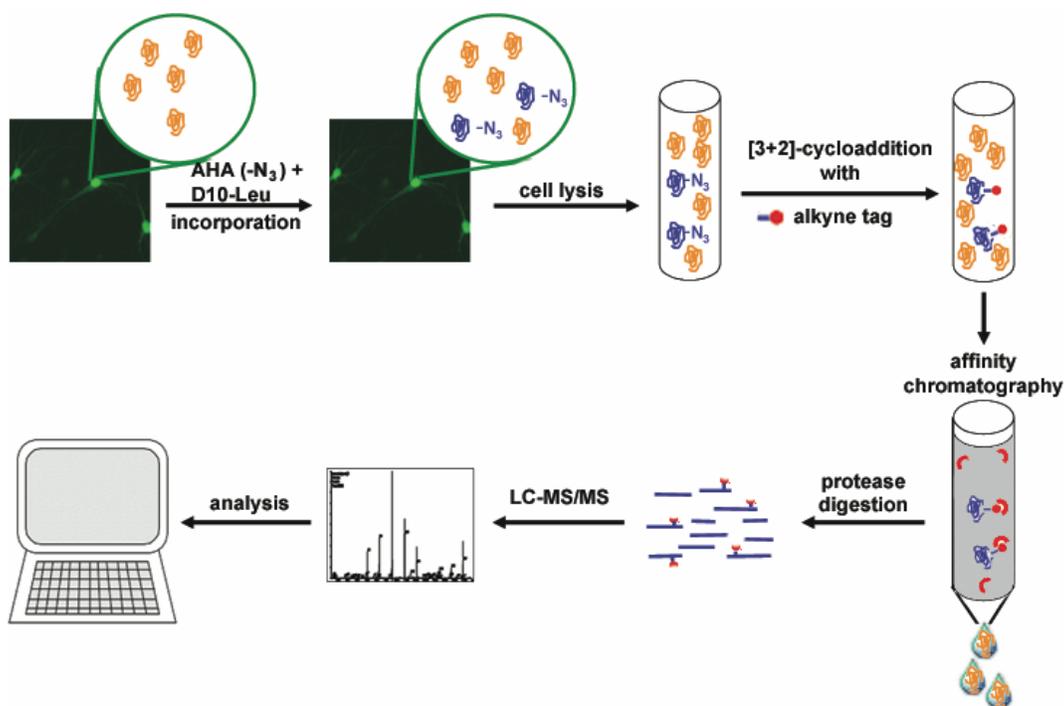
## References

- (1) Freeman, W. M.; Hemby, S. E. *Neurochemical Research* **2004**, *29*, 1065-1081.
- (2) Lilley, K. S.; Friedman, D. B. *Expert Review of Proteomics* **2004**, *1*, 401-409.
- (3) Washburn, M. P.; Ulaszek, R.; Deciu, C.; Schieltz, D. M.; Yates, J. R. *Analytical Chemistry* **2002**, *74*, 1650-1657.
- (4) Andersen, J. S.; Lam, Y. W.; Leung, A. K. L.; Ong, S. E.; Lyon, C. E.; Lamond, A. I.; Mann, M. *Nature* **2005**, *433*, 77-83.
- (5) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. *Molecular & Cellular Proteomics* **2002**, *1*, 376-386.
- (6) Chang, E. J.; Archambault, V.; McLachlin, D. T.; Krutchinsky, A. N.; Chait, B. T. *Analytical Chemistry* **2004**, *76*, 4472-4483.
- (7) Garcia, B. A.; Shabanowitz, J.; Hunt, D. F. *Methods* **2005**, *35*, 256-264.
- (8) Peters, E. C.; Brock, A.; Ficarro, S. B. *Mini-Reviews in Medicinal Chemistry* **2004**, *4*, 313-324.
- (9) Tai, H. C.; Khidekel, N.; Ficarro, S. B.; Peters, E. C.; Hsieh-Wilson, L. C. *J. Am. Chem. Soc.* **2004**, *126*, 10500-10501.

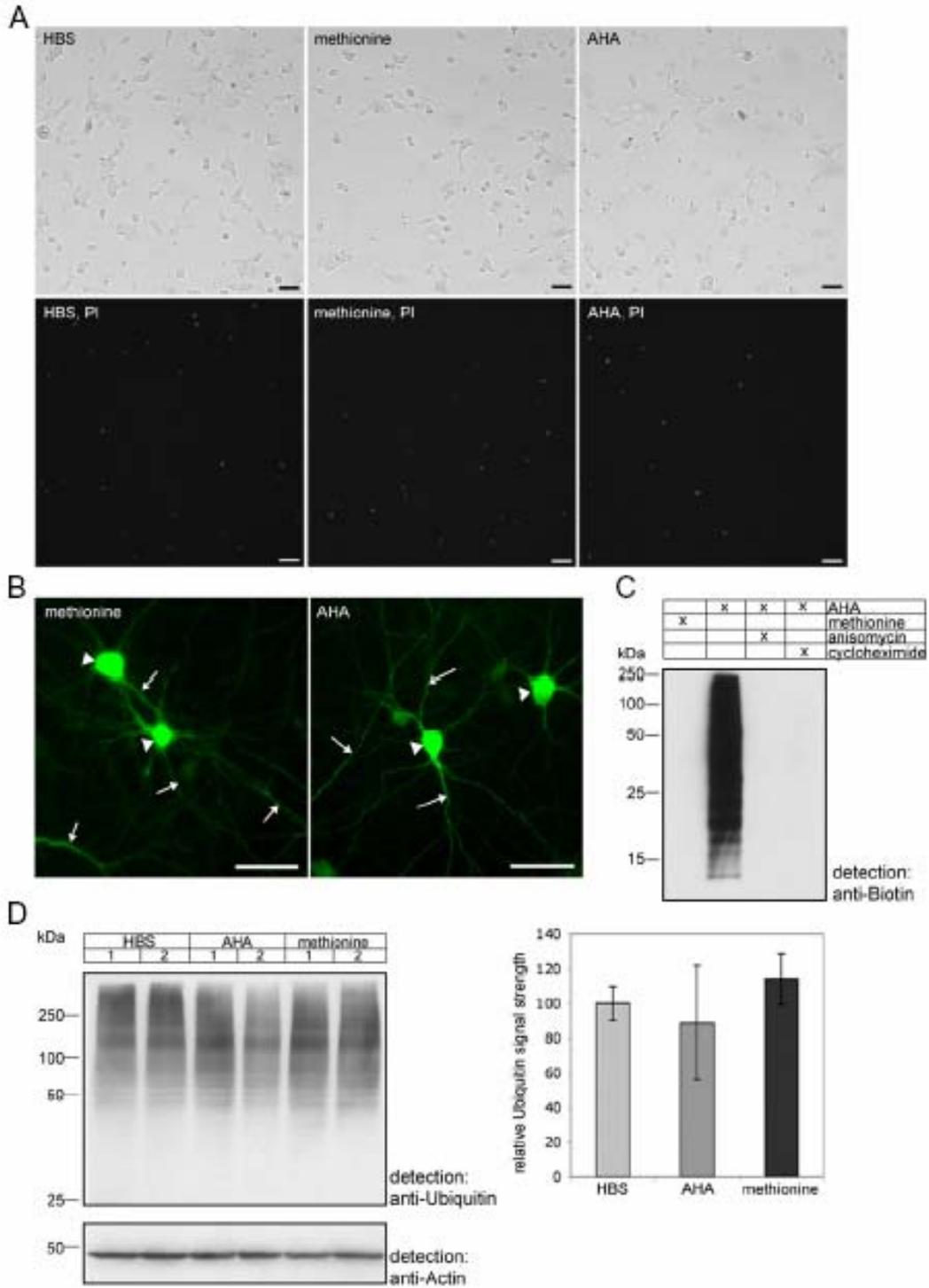
- (10) Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 19-24.
- (11) Link, A. J.; Tirrell, D. A. *J. Am. Chem. Soc.* **2003**, *125*, 11164-11165.
- (12) Speers, A. E.; Adam, G. C.; Cravatt, B. F. *J. Am. Chem. Soc.* **2003**, *125*, 4686-4687.
- (13) Aakalu, G.; Smith, W. B.; Nguyen, N.; Jiang, C. G.; Schuman, E. M. *Neuron* **2001**, *30*, 489-502.
- (14) Bertaux, F.; Sharp, A. H.; Ross, C. A.; Lehrach, H.; Bates, G. P.; Wanker, E. *FEBS Lett.* **1998**, *426*, 229-232.
- (15) Li, X. J.; Li, S. H.; Sharp, A. H.; Nucifora, F. C.; Schilling, G.; Lanahan, A.; Worley, P.; Snyder, S. H.; Ross, C. A. *Nature* **1995**, *378*, 398-402.
- (16) Washburn, M. P.; Wolters, D.; Yates, J. R. *Nat. Biotechnol.* **2001**, *19*, 242-247.
- (17) Tabb, D. L.; McDonald, W. H.; Yates, J. R. *Journal of Proteome Research* **2002**, *1*, 21-26.
- (18) Peng, J. M.; Elias, J. E.; Thoreen, C. C.; Licklider, L. J.; Gygi, S. P. *Journal of Proteome Research* **2003**, *2*, 43-50.
- (19) Olsen, J. V.; Ong, S. E.; Mann, M. *Molecular & Cellular Proteomics* **2004**, *3*, 608-614.
- (20) Zeeberg, B. R.; Feng, W. M.; Wang, G.; Wang, M. D.; Fojo, A. T.; Sunshine, M.; Narasimhan, S.; Kane, D. W.; Reinhold, W. C.; Lababidi, S.; Bussey, K. J.; Riss, J.; Barrett, J. C.; Weinstein, J. N. *Genome Biology* **2003**, *4*.
- (21) Torre, E. R.; Steward, O. *Journal of Neuroscience* **1992**, *12*, 762-772.
- (22) Marcora, E.; Gowan, K.; Lee, J. E. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*,

9578-9583.

- (23) Mangold, J. B.; Mischke, M. R.; Lavelle, J. M. *Mutation Research* **1989**, *216*, 27-33.
- (24) Wang, Q.; Chan, T. R.; Hilgraf, R.; Fokin, V. V.; Sharpless, K. B.; Finn, M. G. *J. Am. Chem. Soc.* **2003**, *125*, 3192-3193.
- (25) Graumann, J.; Dunipace, L. A.; Seol, J. H.; McDonald, W. H.; Yates, J. R.; Wold, B. J.; Deshaies, R. J. *Molecular & Cellular Proteomics* **2004**, *3*, 226-237.
- (26) McDonald, W. H.; Yates, J. R. *Disease Markers* **2002**, *18*, 99-105.
- (27) Sadygov, R. G.; Eng, J.; Durr, E.; Saraf, A.; McDonald, H.; MacCoss, M. J.; Yates, J. R. *Journal of Proteome Research* **2002**, *1*, 211-215.
- (28) Eng, J. K.; McCormack, A. L.; Yates, J. R. *Journal of the American Society for Mass Spectrometry* **1994**, *5*, 976-989.
- (29) McDonald, W. H.; Tabb, D. L.; Sadygov, R. G.; MacCoss, M. J.; Venable, J.; Graumann, J.; Johnson, J. R.; Cociorva, D.; Yates, J. R. *Rapid Communications in Mass Spectrometry* **2004**, *18*, 2162-2168.



**Figure 1:** Overview of the protein identification procedure using azidohomoalanine (AHA). Cells are challenged (+/- a modulator of activity) in the presence of AHA to allow for protein synthesis with AHA-incorporation. After incubation, cells are lysed *or* undergo a subcellular fractionation for biochemical enrichment of specific cellular compartments followed by lysis. Lysates are then coupled to an alkyne-bearing affinity tag, followed by affinity chromatography to enrich for AHA-incorporated proteins. Purified proteins are digested with a protease, and the resulting peptides are analyzed by tandem mass spectrometry to obtain experimental spectra. Different search programs are used to match the acquired spectra to protein sequences.



**Figure 2:** Caption on following page.

**Figure 2:** Use of AHA to label newly synthesized proteins: testing toxicity, membrane permeability, specificity, possible increased protein degradation, and purification of AHA-incorporated proteins. (A) HEK293 cells were incubated for 2 hrs with either HBS, 4 mM methionine in HBS, or 4 mM AHA in HBS and were then incubated with propidium iodide (PI) to stain for dead cells. Upper image row shows Nemarski, lower row PI signals. Scalebar = 100  $\mu\text{m}$ . (B) Dissociated hippocampal cultured neurons (12 DIV) were infected with a destabilized and myristoylated variant of EGFP, whose mRNA is targeted into dendrites (6). Nine hours post infection cells were incubated for 2 hrs with equimolar concentrations of AHA or methionine. Images show representative neurons expressing the GFP reporter indicating no change in the gross morphology of AHA incubated neurons compared to methionine controls. Arrows indicate dendrites, arrowheads point to somata. Scalebar = 50  $\mu\text{m}$ . (C) Western blot analysis for biotinylated AHA-incorporated proteins in cell lysates from HEK293 cells incubated with AHA in the absence or presence of the protein synthesis inhibitor Anisomycin or Cycloheximide as compared to methionine control. The biotin-immunoreactivity was completely dependent on protein synthesis and the presence of the AHA. (D) Western blot analysis of ubiquitination levels of whole cell lysates in AHA-treated HEK293 cells as compared to methionine or buffer (HBS) control samples. Quantification of ubiquitin signals in relation to actin signal is shown in the right graph; n=4.

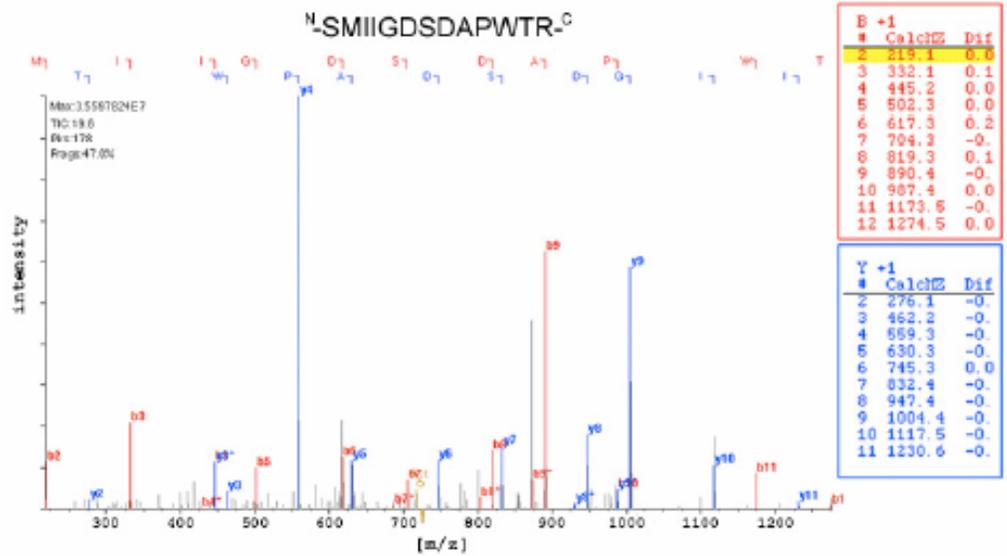


**Figure 3:** Purification of AHA-incorporated proteins after azide-alkyne ligation with a Biotin-FLAG-alkyne tag. (A) Structure of the Trypsin-cleavable Biotin-FLAG-alkyne tag. Biotin (red square), and alkyne (green square), as well as the tryptic cleavage sites (blue scissors), are indicated. The FLAG epitope DYKDDDDK is separated from the biotin moiety by a short linker (GGA). (B) Western blot analysis for tandem purified biotinylated proteins using the Biotin-FLAG-alkyne tag. Cell lysates from both AHA and methionine incubated HEK293 cells were subjected to [3+2]-cycloaddition with the Biotin-FLAG-alkyne tag and subsequently purified using Neutravidin affinity matrices. Save for the non-specific protein staining of samples containing Neutravidin affinity resin, control samples show no biotin signal. Note the higher migration level of alkyne-tagged HA-HAP1A protein compared to the untagged one in the methionine control and in the supernatant of the AHA sample. Sizes of marker proteins are indicated on the left margin.

**A**

MYPYDVPDYARPKEQVQSGAGDGTGSGDPAAGTPTTQPAVGPAPPEP  
 SAEPKPAPAQGTGSGQKSGSRTKTGSFCRSMIIGDSDAPWTRYVFQG  
 #  
 PYGPRATGLGTGKAEGIWKTPAAYIGRRPGVSGPERAAFIRELQEALCP  
 NPPPTKKITEDDVKVMLYLLEEKERDLNTAARIGQSLVKQNSVLMEENN  
 KLETMLGSAREEILHLRKQVNLRDDLLQLYSDSDDDDEEDEDEEEEG  
 EEEEREGQRDQDQQHDPYGAPKPHPKAETAHRCQPLETLQQLRLL  
 EEENDHLREEASHLDNLEDEEQMLILECVEQFSEASQQMAELSEVLVL  
 RLEGYERQQKEITQLQAEITKLQQRCSYGAQTEKLQQLASEKGIHS  
 ESLRAGSYMQDYGSRPRDRQEDGKSHRQRSSMPAGSVTHYGYSVPL  
 #  
 # # # \*  
 DALPSFPETLAEELRTSLRKFITDPAYFMERRDTHCREGRKKEQRAMP  
 PPPAQDLKPPEDFEAPEELVPEEELGAIEEVGTAEDGQAEENEQASEE  
 TEAWEEVEPEVDETTRMNVVVSALASGLGPSHLDMKYVLQQLSNWQ  
 # # # #  
 DAHSKRQKQKQVVPKGECSRGRHPPASGTSFRSSTI

**B**



**Figure 4:** Continued on following page.

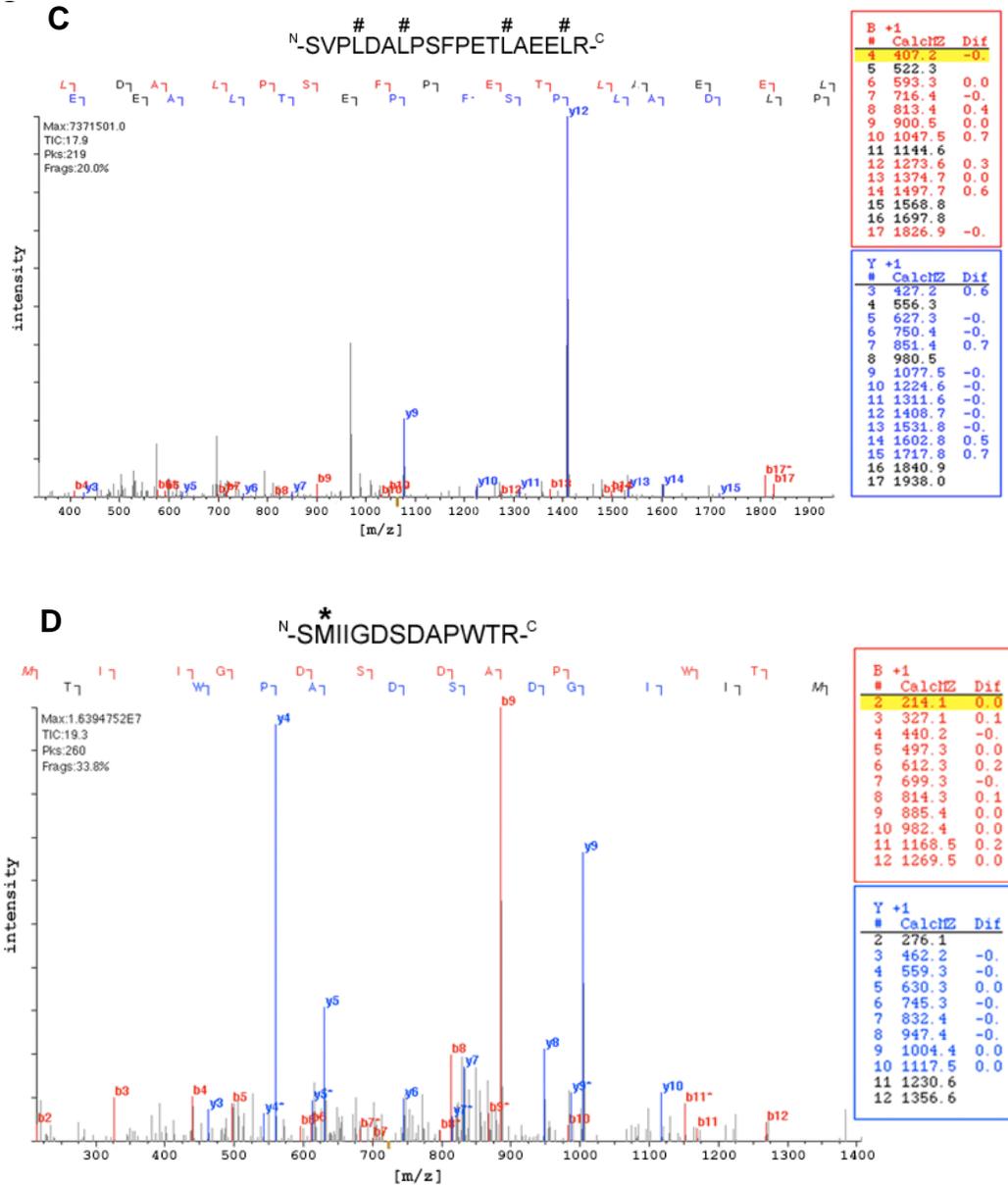


Figure 4: Caption on following page.

**Figure 4:** Sequence coverage and detection of modifications for HA-HAP1A. (A) For the 4 independent sets of experiments, identified peptides for HA-HAP-1A are indicated by lines in four different colors under the covering sequence. (B) to (D) Representative MS/MS spectra of identified HA-HAP-1A peptides. The b-series ions are labeled in red; the y-series ions are labeled in blue. The sequence for each peptide is given at the top of the spectrum in black. Below the sequence, a series of letters show how the peptide sequence aligns to the spectrum. In the spectrum itself, highlighted fragment ions are the most intense, falling within 0.75 m/z of a calculated fragment ion position. To the right, the correspondence between observed and predicted fragment ion m/z for a particular ion series is given (y-series in blue, b-series in red). Ions that do not match the spectrum are in black. “\*” indicates exchange of methionine for AHA and “#” indicates d<sub>10</sub>-leu for (A) to (D).

<b>Proteins identified in</b>	<b>Number of proteins</b>	<b>Percentage [%]</b>
4 AHA exp.	30	15.2
3 AHA exp.	24	12.2
2 AHA exp.	54	43.7
single AHA exp.	86	43.7
2 AHA & 1 Met exp.	1	0.5
single Met exp.	2	1.0

**Table 1:** Protein identifications among four independent experimental sets.

Listed counts for identified proteins were acquired under the following criteria: minimum of two valid peptides and one modified peptide per locus, partial tryptic status required, and minimal XCorr and DeltCN as indicated in material and methods. Proteins that are subsets of each other (a subset protein is one for which all peptides are found in another protein) were combined, counting as 1 identification in the report.

Locus	Description	SeqCov	AA	Mw	pl	Redundancy
IPI00010740.1 SWI	Splice Isoform Long of Splicing factor proline-and glutamine-rich	19.1	707	76150	9.4	4
IPI00000816.1 SWI	14-3-3 protein epsilon	23.5	255	29174	4.7	4
IPI00003865.1 SWI	Splice Isoform 1 of Heat shock cognate 71 kDa protein	27.4	646	70898	5.5	4
IPI00007765.5 SWI	Stress-70 protein mitochondrial precursor	13	679	73681	6.2	4
IPI00013881.5 SWI	Heterogeneous nuclear ribonucleoprotein H1	19.2	449	49229	6.3	4
IPI00479484.1 ENS	48 kDa protein	19.2	442	47927	7.1	4
IPI00479191.1 TRE	HNRPH1 protein	19.2	472	51230	6.8	4
IPI00477925.1 ENS	50 kDa protein	19.2	465	50455	7.6	4
IPI00477457.1 TRE	Hypothetical protein DKFZp686A15170	19.2	420	46092	7.1	4
IPI00017297.1 SWI	Matrin 3	15.0	847	94623	6.3	4
IPI00549755.1 H-I	Matrin 3	15.0	848	95197	5.9	4
IPI00018465.1 SWI	T-complex protein 1 eta subunit	4.1	543	59367	7.6	4
IPI00021263.3 SWI	14-3-3 protein zeta/delta	29.8	245	27745	4.8	4
IPI00180776.2 ENS	29 kDa protein	28.0	261	29420	4.7	4
IPI00021439.1 SWI	Actin cytoplasmic 1	42.7	375	41737	5.5	4
IPI00021440.1 SWI	Actin cytoplasmic 2	42.7	375	41793	5.5	4
IPI00027499.1 SWI	Splice Isoform 1 of Nucleophosmin	44.2	294	32575	4.8	4
IPI00220740.1 SWI	Splice Isoform 2 of Nucleophosmin	49.1	265	29465	4.6 <sup>1</sup>	4
IPI00081836.1 TRE	OTTHUMP00000016173	55.5	128	13906	10.9	4
IPI00552873.1 SWI	Histone H2A.e	55.9	127	13805	10.9	4
IPI00291764.2 SWI	Histone H2A.c/d/i/n/p	55.0	129	13960	10.9	4
IPI00255316.4 SWI	Histone H2A.g	55.0	129	13976	10.9	4
IPI00220855.3 TRE	H2A histone family member J isoform 2	55.0	129	14019	10.9	4
IPI00102165.1 TRE	Hypothetical protein FLJ10903	45.8	155	16610	10.9	4
IPI00215638.2 SWI	ATP-dependent RNA helicase A	22.9	1270	140877	6.8	4
IPI00216049.1 SWI	Splice Isoform 1 of Heterogeneous nuclear ribonucleoprotein K	26.1	463	50976	5.5	4
IPI00514561.1 TRE	Heterogeneous nuclear ribonucleoprotein K	28.3	428	47557	6.6	4
IPI00216746.1 SWI	Splice Isoform 2 of Heterogeneous nuclear ribonucleoprotein K	26.1	464	51028	5.3	4
IPI00216457.6 SWI	Histone H2A.o	55.0	129	13964	10.9	4
IPI00339274.4 SWI	Histone H2A.q	55.5	128	13857	10.9	4
IPI00217465.4 SWI	Histone H1.2	19.3	212	21234	10.9	4
IPI00217466.2 SWI	Histone H1.3	18.6	220	22219	11.0	4
IPI00217467.2 SWI	Histone H1.4	18.8	218	21734	11.0	4
IPI00219038.4 SWI	Histone H3.3	38.5	135	15197	11.3	4
IPI00376838.2 REF	PREDICTED: similar to Histone H3.3	28.3	184	20846	10.7	4
IPI00290566.1 SWI	T-complex protein 1 alpha subunit	6.5	556	60344	6.1	4
IPI00550591.1 REF	T-complex protein 1 isoform b	9.0	401	43886	7.7	4
IPI00301277.1 SWI	Heat shock 70 kDa protein 1L	18.6	641	70375	6	4
IPI00304596.1 SWI	NONO protein	28.9	471	54232	8.9	4
IPI00304925.1 SWI	Heat shock 70 kDa protein 1	20.4	641	70052	5.6	4
IPI00514377.3 TRE	Heat shock 70kDa protein 1A	20.4	641	70038	5.6	4
IPI00329745.4 SWI	130 kDa leucine-rich protein	8.8	1402	158925	6.3	4
IPI00477140.1 TRE	LRPPRC protein	9.6	1278	145768	5.8	4
IPI00334775.3 TRE	Hypothetical protein DKFZp761K0511	29.9	737	84843	5.4	4
IPI00414676.5 SWI	Heat shock protein HSP 90-beta	30.4	723	83133	5	4
IPI00382470.2 SWI	Heat shock protein HSP 90-alpha	23.9	723	83133	5	4
IPI00387144.4 SWI	Tubulin alpha-ubiquitous chain	34.8	451	50152	5.1	4
IPI00440493.2 SWI	ATP synthase alpha chain mitochondrial precursor	12.7	553	59751	9.1	4
IPI00549805.2 TRE	ATP5A1 protein	12.7	553	59809	9	4
IPI00471928.3 TRE	ATP synthase H+ transporting mitochondrial F1 complex alpha subunit isoform b	13.9	503	54494	8.2	4
IPI00449049.2 SWI	Poly [ADP-ribose] polymerase-1	30.5	1013	112952	8.9	4
IPI00453473.3 SWI	Histone H4	59.8	102	11236 <sup>3</sup>	11.4 <sup>2</sup>	4
IPI00465248.4 SWI	Alpha enolase	40.4	433	47038	7.4	4
IPI00472102.1 SWI	60 kDa heat shock protein mitochondrial precursor	24.7	575	61213	5.9	4

gi 6754158 ref NP	huntingtin-associated protein 1 isoform A	27.3	607	67887	4.8	4
IPI00003881.3 SWI	Heterogeneous nuclear ribonucleoprotein F	18.3	415	45672	5.6	3
IPI00003935.5 SWI	Histone H2B.q	38.4	125	13789	10.3	3
IPI00515061.2 SWI	Histone H2B.r	38.4	125	13773	10.3	3
IPI00220403.2 SWI	Histone H2B.f	38.4	125	13819	10.3	3
IPI00166293.4 SWI	Histone H2B type 12	38.4	125	13777	10.3	3
IPI00152785.2 SWI	Histone H2B.n	38.4	125	13775	10.3	3
IPI00005705.1 SWI	Splice Isoform Gamma-1 of Serine/threonine protein phosphatase PP1- gamma catalytic subunit	10.8	323	36984	6.5	3
IPI00550451.1 SWI	Serine/threonine protein phosphatase PP1-alpha catalytic subunit	10.6	330	37512	6.3	3
IPI00218236.4 SWI	Serine/threonine protein phosphatase PP1-beta catalytic subunit	10.7	327	37187	6.2	3
IPI00218187.1 SWI	Splice Isoform Gamma-2 of Serine/threonine protein phosphatase PP1- gamma catalytic subunit	10.4	337	38518	6.1	3
IPI00027423.2 TRE	Protein phosphatase type 1 catalytic subunit	10.3	341	38631	6.6	3
IPI00011257.1 SWI	Eukaryotic translation initiation factor 3 subunit 5	5.3	357	37564	5.4	3
IPI00240909.1 REF	PREDICTED: similar to eukaryotic translation initiation factor 3 subunit 5 epsilon 47kDa	5.3	361	37974	5.6	3
IPI00011654.2 SWI	Tubulin beta-2 chain	41	444	49671	4.9	3
IPI00017617.1 SWI	Probable RNA-dependent helicase p68	9.4	337	38518	6.1	3
IPI00018534.3 SWI	Histone H2B.c	38.4	125	13821	10.3	3
IPI00554798.1 SWI	Histone H2B.e	38.4	125	13858	10.3	3
IPI00554445.1 SWI	Histone H2B.d	38.4	125	13791	10.3	3
IPI00477495.2 SWI	Histone H2B.s	38.4	125	13813	10.4	3
IPI00419833.1 SWI	HIST1H2BM protein	34.8	138	15157	10.2	3
IPI00377199.4 TRE	OTTHUMP00000039500	28.9	166	18804	10.5	3
IPI00329665.5 TRE	PREDICTED: similar to Hist1h2bc protein	37.5	128	14166	10.3	3
IPI00303133.7 SWI	Histone H2B.j	38.4	125	13761	10.3	3
IPI00152906.5 SWI	Histone H2B.b	38.4	125	13805	10.3	3
IPI00020101.4 SWI	Histone H2B.a/g/h/k/l	38.4	125	13775	10.3	3
IPI00022334.1 SWI	Ornithine aminotransferase mitochondrial precursor	12.3	439	48535	7	3
IPI00024993.3 SWI	Enoyl-CoA hydratase mitochondrial precursor	18.3	290	31371	8.1	3
IPI00026119.4 SWI	Ubiquitin-activating enzyme E1	10.7	506	56852	6.2	3
IPI00179964.5 SWI	Splice Isoform 1 of Polypyrimidine tract-binding protein 1	26.2	531	57221	9.2	3
IPI00334175.3 SWI	Splice Isoform 2 of Polypyrimidine tract-binding protein 1	25.3	550	59031	9.2	3
IPI00183626.7 TRE	PTBP1 protein	25.0	557	59633	9.2	3
IPI00217507.4 SWI	Neurofilament triplet M protein	6.4	915	102317	4.9	3
IPI00220834.7 SWI	ATP-dependent DNA helicase II 80 kDa subunit	12.2	731	82573	5.8	3
IPI00383237.3 TRE	Pyruvate kinase M2	25.1	530	57781	7.7	3
IPI00550724.2 H-I	Similar to Actin cytoplasmic 2	45.6	305	34088	6.2	3
IPI00018140.2 SWI	Splice Isoform 1 of Heterogeneous nuclear ribonucleoprotein Q	10.1	623	69633	8.6	3
IPI00402182.1 SWI	Splice Isoform 2 of Heterogeneous nuclear ribonucleoprotein Q	10.7	588	65712	8.6	3
IPI00018278.1 TRE	Histone H2A.F/Z variant	53.9	128	13509	10.6	3
IPI00218448.3 SWI	Histone H2A.z	54.3	127	13422	10.6	3
IPI00021700.1 SWI	Proliferating cell nuclear antigen	15.3	261	28769	4.7	3
IPI00025252.1 SWI	Protein disulfide-isomerase A3 precursor	6.3	505	56782	6.4	3
IPI00215637.2 SWI	DEAD-box protein 3 X-chromosomal	3.3	661	73112	7.2	3
IPI004195856 SWI	Peptidylprolyl isomerase A-like	25.6	165	17998	7.8	3
IPI00013877.2 SWI	Splice Isoform 1 of Heterogeneous nuclear ribonucleoprotein H3	4.9	346	36926	6.9	3
IPI00216496.1 SWI	Splice Isoform 6 of Heterogeneous nuclear ribonucleoprotein H3	12.2	139	15413	6.4	3
IPI00216495.1 SWI	Splice Isoform 5 of Heterogeneous nuclear ribonucleoprotein H3	11.7	145	16084	6.4	3
IPI00216494.1 SWI	Splice Isoform 4 of Heterogeneous nuclear ribonucleoprotein H3	7.9	215	22322	5.7	3
IPI00216493.1 SWI	Splice Isoform 3 of Heterogeneous nuclear ribonucleoprotein H3	5.7	297	31525	7.3	3
IPI00216492.1 SWI	Splice Isoform 2 of Heterogeneous nuclear ribonucleoprotein H3	5.1	331	35239	6.9	3

IPI00022774.2 SWI	Transitional endoplasmic reticulum ATPase	6.7	805	89191	5.3	3
IPI00478540.1 ENS	89 kDa protein	6.7	806	89396	5.3	3
IPI00478348.2 ENS	46 kDa protein	14.3	421	45575	7.8	3
IPI00002966.1 SWI	Heat shock 70 kDa protein 4	7.5	840	94300	5.3	2
IPI00008274.3 SWI	Adenylyl cyclase-associated protein 1	4.9	474	51542	8	2
IPI00012074.2 SWI	HNRPR protein	11.2	636	71214	8.1	2
IPI00013122.1 SWI	Hsp90 co-chaperone Cdc37	5.6	378	44468	5.2	2
IPI00013894.1 SWI	Stress-induced-phosphoprotein 1	3.5	543	62639	6.8	2
IPI00479946.2 ENS	68 kDa protein	3.2	590	68080	7.7	2
IPI00017299.3 SWI	B lymphoma Mo-MLV insertion region	5.2	326	36997	8.7	2
IPI00514181.1 TRE	B lymphoma Mo-MLV insertion region	9.2	184	21262	8.8	2
IPI00018146.1 SWI	14-3-3 protein tau	13.5	245	27764	4.8	2
IPI00020984.1 SWI	Calnexin precursor	14.5	592	67568	4.6	2
IPI00021428.1 SWI	Actin alpha skeletal muscle	16.4	287	32048	5.4	2
IPI00023006.1 SWI	Actin alpha cardiac	16.4	377	42019	5.4	2
IPI00022827.1 TRE	CTCL tumor antigen se20-9	1.7	1235	142695	5.1	2
IPI00247439.2 TRE	HSLK	1.7	1204	138968	5.1	2
IPI00022977.1 SWI	Creatine kinase B chain	22	381	42644	5.6	2
IPI00024057.2 SWI	Transgelin 2	24.1	220	24454	8.2	2
IPI00026089.3 SWI	Splicing factor 3B subunit 1	10.2	1304	145815	7	2
IPI00027230.3 SWI	Endoplasmic precursor	7	803	92469	4.8	2
IPI00027834.3 SWI	Heterogeneous nuclear ribonucleoprotein L isoform a	28.2	589	64133	8.2	2
IPI00031556.5 SWI	Splicing factor U2AF 65 kDa subunit	16.2	475	53501	9.1	2
IPI00552483.1 TRE	U2 (RNU2) small nuclear RNA auxiliary factor 2 isoform b	16.3	471	53121	9.1	2
IPI00063234.1 TRE	PRKAR2A protein	4.5	382	43067	5.1	2
IPI00219774.2 SWI	cAMP-dependent protein kinase type II-alpha regulatory subunit	4.2	403	45387	5.1	2
IPI00163505.2 SWI	Splice Isoform 1 of RNA-binding region containing protein 2	12.5	530	59380	10.1	2
IPI00513959.1 TRE	RNA-binding region	19.6	336	36585	5.6	2
IPI00377114.2 REF	RNA-binding region containing protein 2 isoform c	17.7	373	40541	6.3	2
IPI00377112.1 TRE	Hypothetical protein DKFZp686A11192	18.0	367	39818	6.6	2
IPI00215801.1 SWI	Splice Isoform 2 of RNA-binding region containing protein 2	12.6	524	58657	10.1	2
IPI00165434.2 SWI	Splice Isoform 1 of YLP motif containing protein 1	0.9	1951	219983	6.6	2
IPI00514265.1 SWI	Splice Isoform 3 of YLP motif containing protein 1	1.0	1862	209481	6.6	2
IPI00456742.1 TRE	PREDICTED: YLP motif containing 1	0.8	2140	240735	6.6	2
IPI00181641.5 SWI	Splice Isoform 2 of YLP motif containing protein 1	1.0	1822	204947	6.7	2
IPI00186290.5 SWI	Elongation factor 2	16	857	95207	6.8	2
IPI00217506.4 SWI	Nucleolin	27.8	706	76213	4.7	2
IPI00219037.4 SWI	Histone H2A.x	52.8	143	15013	10.7	2
IPI00219156.2 SWI	60S ribosomal protein L30	24.6	114	12653	9.6	2
IPI00219217.2 SWI	L-lactate dehydrogenase B chain	31.8	333	36507	6.1	2
IPI00384026.1 TRE	Full-length cDNA 5-PRIME end of clone CS0DN005YI08 of Adult brain of Homo sapiens	20.6	262	29031	5.2	2
IPI00419373.1 SWI	Splice Isoform 1 of Heterogeneous nuclear ribonucleoprotein A3	25.7	378	39595	9	2
IPI00465216.3 TRE	MRPL1 protein	6.9	303	34475	8.4	2
IPI00549381.2 TRE	MRPL1 protein	6.9	303	34420	8.4	2
IPI00001639.2 SWI	Importin beta-1 subunit	3.4	876	97170	4.8	2
IPI00005614.3 SWI	Splice Isoform Long of Spectrin beta chain brain 1	2.8	2364	274630	5.6	2
IPI00333015.3 TRE	Beta-spectrin 2 isoform 2	3.1	2155	251417	5.6	2
IPI00328230.1 SWI	Splice isoform short of spectrin beta chain brain 1	3.1	2168	253111	5.6	2
IPI00008248.2 SWI	Anaphase promoting complex subunit 7	2.7	565	63161	5.8	2
IPI00024067.1 SWI	Clathrin heavy chain 1	9.8	1675	191613	5.7	2
IPI00455383.1 SWI	Splice isoform 2 of Clathrin heavy chain	10.0	1638	187758	5.7	2
IPI00030275.4 SWI	Heat shock protein 75 kDa mitochondrial precursor	13.4	704	80011	8	2

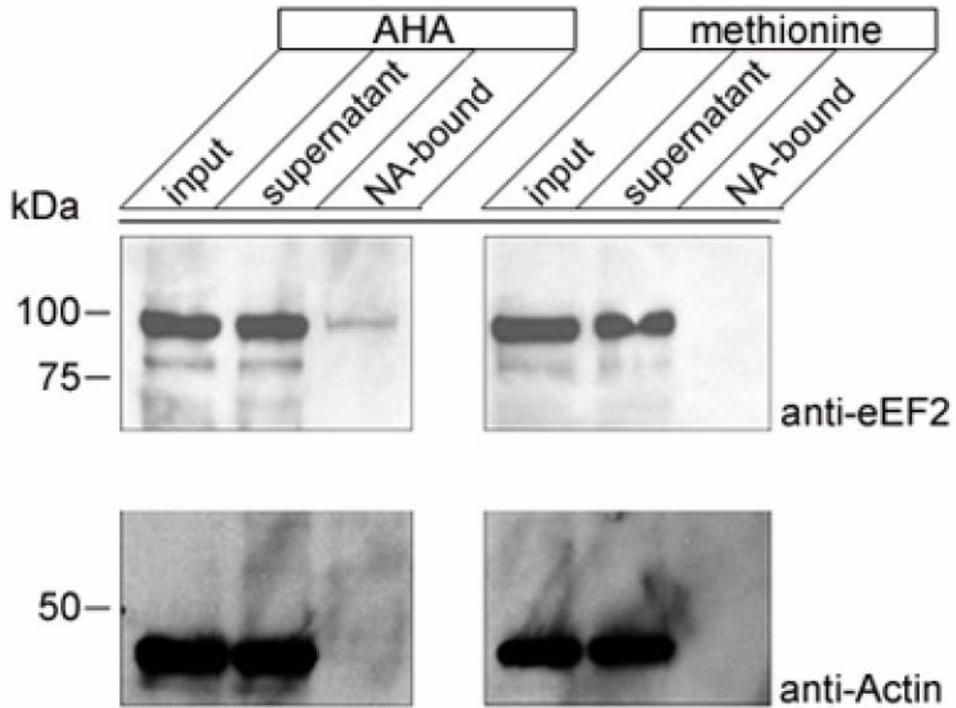
IPI00376215.2 SWI	Splice Isoform 2 of DNA-dependent protein kinase catalytic subunit	6.5	4097	465505 <sup>4</sup>	7.2	2
IPI00554482.1 TRE	Nucleophosmin	33.4	293	32460	4.8	2
IPI00008982.1 SWI	Splice Isoform Long of Delta 1-pyrroline-5-carboxylate synthetase	8.4	795	87302	7.1	2
IPI00218547.1 SWI	Splice Isoform Short of Delta 1-pyrroline-5-carboxylate Synthetase	8.4	793	87089	7.1	2
IPI00033130.3 SWI	Ubiquitin-like 1 activating enzyme E1A	5.8	346	38450	5.3	2
IPI00217966.4 SWI	L-lactate dehydrogenase A chain	23.3	331	36558	8.3	2
IPI00218343.4 SWI	Tubulin alpha-6 chain	31.8	449	49895	5.1	2
IPI00293655.3 SWI	ATP-dependent helicase DDX1	5.9	740	82432	7.2	2
IPI00402183.1 SWI	Splice Isoform 3 of Heterogeneous nuclear ribonucleoprotein Q	8.4	562	62686	7.6	2
IPI00402184.3 SWI	Splice Isoform 4 of Heterogeneous nuclear ribonucleoprotein Q	8.9	527	58766	7.6	2
IPI00554720.1 TRE	H2A histone family member V isoform 1	45.4	130	13752	10.4	2
IPI00012442.1 SWI	Ras-GTPase-activating protein binding protein 1	10.3	466	52164	5.5	2
IPI00023785.3 SWI	Hypothetical protein DKFZp761H2016	8	652	72542	8.6	2
IPI00218753.1 SWI	Splice Isoform 3 of DNA topoisomerase II alpha isozyme	12.8	1567	178711	8.8	2
IPI00478232.2 SWI	Splice Isoform 1 of DNA topoisomerase II alpha isozyme	12.9	1563	178352	8.7	2
IPI00414101.2 SWI	Splice Isoform 2 of DNA topoisomerase II alpha isozyme	12.9	1557	177500	8.8	2
IPI00218754.1 SWI	Splice Isoform 4 of DNA topoisomerase II alpha isozyme	12.5	1612	182680	8.6	2
IPI00413614.2 TRE	SYMPK protein	5.3	533	58929	5.5	2
IPI00413728.2 SWI	Splice Isoform 1 of Spectrin alpha chain brain	4.5	2472	284526	5.3	2
IPI00513786.1 SWI	Splice Isoform 3 of Spectrin alpha chain brain	4.6	2452	282268	5.3	2
IPI00478292.1 SWI	Splice Isoform 2 of Spectrin alpha chain brain	4.5	2480	285480	5.4	2
IPI00413826.1 ENS	15 kDa protein	28.7	136	15226	11	2
IPI00465294.2 TRE	Pombe Cdc5-related protein	6.6	802	92251	8.2	2
IPI00477805.1 TRE	RIF1 isoform 4	2.1	2472	274561	5.5	2
IPI00012998.1 TRE	HPBRII-4 mRNA	4.5	551	59209	7.4	2
IPI00553042.1 TRE	CPSF6 protein	5.2	478	52326	6.4	2
IPI00030654.1 TRE	CPSF6 protein	4.3	588	63471	7.7	2
IPI00303476.1 SWI	ATP synthase beta chain mitochondrial precursor	9.8	529	56560	5.4	2
IPI00031522.2 SWI	Trifunctional enzyme alpha subunit mitochondrial precursor	2.4	763	83000	9	2
IPI00037070.1 SWI	Splice Isoform 2 of Heat shock cognate 71 kDa protein	19.5	497	53921	6	2
IPI00328163.1 TRE	K-ALPHA-1 protein	37.9	335	37218	5	2
IPI00003519.1 SWI	116 kDa U5 small nuclear ribonucleoprotein component	12.4	972	109436	5	1
IPI00005024.2 TRE	MYBBP1A protein	16.9	1328	148873	9.3	1
IPI00007928.3 TRE	PRP8 protein	6	2335	273783	8.8	1
IPI00008433.2 SWI	40S ribosomal protein S5	7.4	204	22876	9.7	1
IPI00009471.1 SWI	WD-repeat protein 3	4.2	943	106099	6.6	1
IPI00164857.1 ENS	105 kDa protein	4.3	935	105149	6.7	1
IPI00010810.1 SWI	Electron transfer flavoprotein alpha-subunit mitochondrial precursor	8.1	333	35080	8.4	1
IPI00015602.1 SWI	Mitochondrial precursor proteins import receptor	3	608	67455	7.1	1
IPI00015953.2 SWI	Nucleolar RNA helicase II	26.3	783	87357	9.3	1
IPI00477179.1 TRE	Hypothetical protein DKFZp686F21172	28.8	515	79657	9.4	1
IPI00024638.2 SWI	Similar to Creatine kinase ubiquitous mitochondrial precursor	5	419	47264	8.3	1
IPI00025054.1 SWI	Splice Isoform Long of Heterogenous nuclear ribonucleoprotein U	17.6	824	90480	6	1
IPI00479217.1 TRE	Heterogeneous nuclear ribonucleoprotein U isoform b	18	806	88980	5.8	1
IPI00386491.4 SWI	Splice Isoform Short of Heterogenous nuclear ribonucleoprotein U	18	806	88946	5.8	1
IPI00027280.2 SWI	Splice Isoform Beta-2 of DNA topoisomerase II beta isozyme	4.2	1626	183266	8	1
IPI00217709.1 SWI	Splice Isoform Beta-1 of DNA topoisomerase II beta isozyme	4.2	1621	182661	8.1	1
IPI00028955.4 SWI	Ribosome biogenesis protein BOP1	12.3	746	83630	6.2	1

IPI00029079.5 SWI	GMP synthase	2.7	693	76715	6.9	1
IPI00031461.1 SWI	Rab GDP dissociation inhibitor beta	8.5	445	50663	6.5	1
IPI00140420.4 SWI	Staphylococcal nuclease domain containing protein 1	2.3	910	101997	7.2	1
IPI00141938.3 REF	H2A histone family member V isoform 2	39.5	114	12146	10.5	1
IPI00171611.5 TRE	Histone H3	38.2	135	15197	11.3	1
IPI00419900.2 SWI	PREDICTED: similar to CG31613-PA	37.7	184	20846	10.7	
IPI00217081.1 TRE	FUN14 domain containing 1	11	155	17178	8.6	1
IPI00220766.3 SWI	Lactoylglutathione lyase	10.9	183	20588	5.5	1
IPI00246058.3 SWI	PDCD6IP protein	1.9	873	96818	6.5	1
IPI00291175.3 SWI	Vinculin isoform VCL	2.0	1066	116722	6.1	1
IPI00307162.2 SWI	Vinculin isoform meta-VCL	1.9	1134	123799	5.7	
IPI00292975.2 TRE	PREDICTED: RNA binding motif protein 27	2.1	1060	118718	9.2	1
IPI00297211.1 SWI	SWI/SNF related matrix associated actin dependent regulator of chromatin subfamily A member 5	3.8	1052	121905	8.1	1
IPI00297579.3 SWI	Chromobox protein homolog 3	15.8	183	20823	5.3	1
IPI00298308.5 TRE	PREDICTED: similar to RIKEN cDNA D330038I09	2.3	910	100248	6.6	1
IPI00300078.5 SWI	Periodic tryptophan protein 2 homolog	1.7	919	102451	6.2	1
IPI00300127.3 SWI	UPF0202 protein KIAA1709	12.2	1025	115704	8.3	1
IPI00329389.5 SWI	60S ribosomal protein L6	21.6	287	32597	10.6	1
IPI00551069.2 TRE	Ribosomal protein L6	21.5	288	32726	10.6	1
IPI00550604.1 TRE	RPL6 protein	21.5	288	32742	10.6	1
IPI00550404.2 TRE	DNA-binding protein TAXREB107	21.5	289	32891	10.6	1
IPI00334190.4 SWI	Stomatin-like protein 2	5.9	356	38534	7.4	1
IPI00477195.2 ENS	40 kDa protein	5.7	369	39943	7.9	1
IPI00442543.1 TRE	Hypothetical protein FLJ27036	11.6	181	19744	7.3	1
IPI00396378.3 SWI	Splice Isoform B1 of Heterogeneous nuclear ribonucleoproteins A2/B1	34.6	353	37430	8.9	1
IPI00477522.1 ENS	37 kDa protein	34.7	351	37298	9.1	1
IPI00414696.1 SWI	Splice Isoform A2 of Heterogeneous nuclear ribonucleoproteins A2/B1	35.8	341	36006	8.6	1
IPI00418471.5 SWI	Vimentin	20.9	465	53521	5.1	1
IPI00448095.3 SWI	L-xylulose reductase	8.2	244	25913	8.1	1
IPI00465260.1 SWI	GARS protein	5.5	751	84648	7.8	1
IPI00465430.3 SWI	ATP-dependent DNA helicase II 70 kDa subunit	17.6	608	69712	6.6	1
IPI00478017.1 ENS	70 kDa protein	17.6	609	69870	6.6	1
IPI00470585.2 TRE	RNA-binding region RNP-1 (RNA recognition motif) domain containing protein	5.0	503	55548	8.4	1
IPI00550821.2 TRE	FLJ12529 protein	5.3	471	52050	8	1
IPI00515047.1 TRE	Actin alpha 1 skeletal muscle	18.8	287	32048	5.4	1
IPI00003362.1 SWI	78 kDa glucose-regulated protein precursor	5.2	654	72333	5.2	1
IPI00008477.1 SWI	Targeting protein for Xklp2	2.8	747	85653	9.2	1
IPI00102661.1 TRE	Hepatocellular carcinoma-associated antigen 90	2.7	783	89393	9.2	1
IPI00010153.1 SWI	60S ribosomal protein L23	14.3	140	14865	10.5	1
IPI00012011.3 SWI	Cofilin non-muscle isoform	17	165	18371	8.1	1
IPI00013808.1 SWI	Alpha-actinin 4	3.1	911	104854	5.4	1
IPI00015148.3 SWI	Ras-related protein Rap-1b	6.5	184	20825	5.8	1
IPI00550325.1 TRE	RAP1A member of RAS oncogene family	6.5	184	20953	6.6	1
IPI00412525.1 ENS	21 kDa protein	6.5	184	20925	5.5	1
IPI00019345.1 SWI	Ras-related protein Rap-1A	6.5	184	20987	6.6	1
IPI00018120.1 SWI	Mitochondrial 28S ribosomal protein S29	5	398	45566	8.9	1
IPI00020944.1 SWI	Farnesyl-diphosphate farnesyltransferase	4.3	417	48115	6.5	1
IPI00021417.2 TRE	SART-1	4.6	805	90701	6.3	1
IPI00022305.3 TRE	MSTP017	4.5	419	48120	6.7	1
IPI00064431.3 TRE	Hypothetical protein FLJ14931	9.5	201	22819	8.8	1
IPI00026230.1 SWI	Heterogeneous nuclear ribonucleoprotein H'	13.4	449	49264	6.3	1

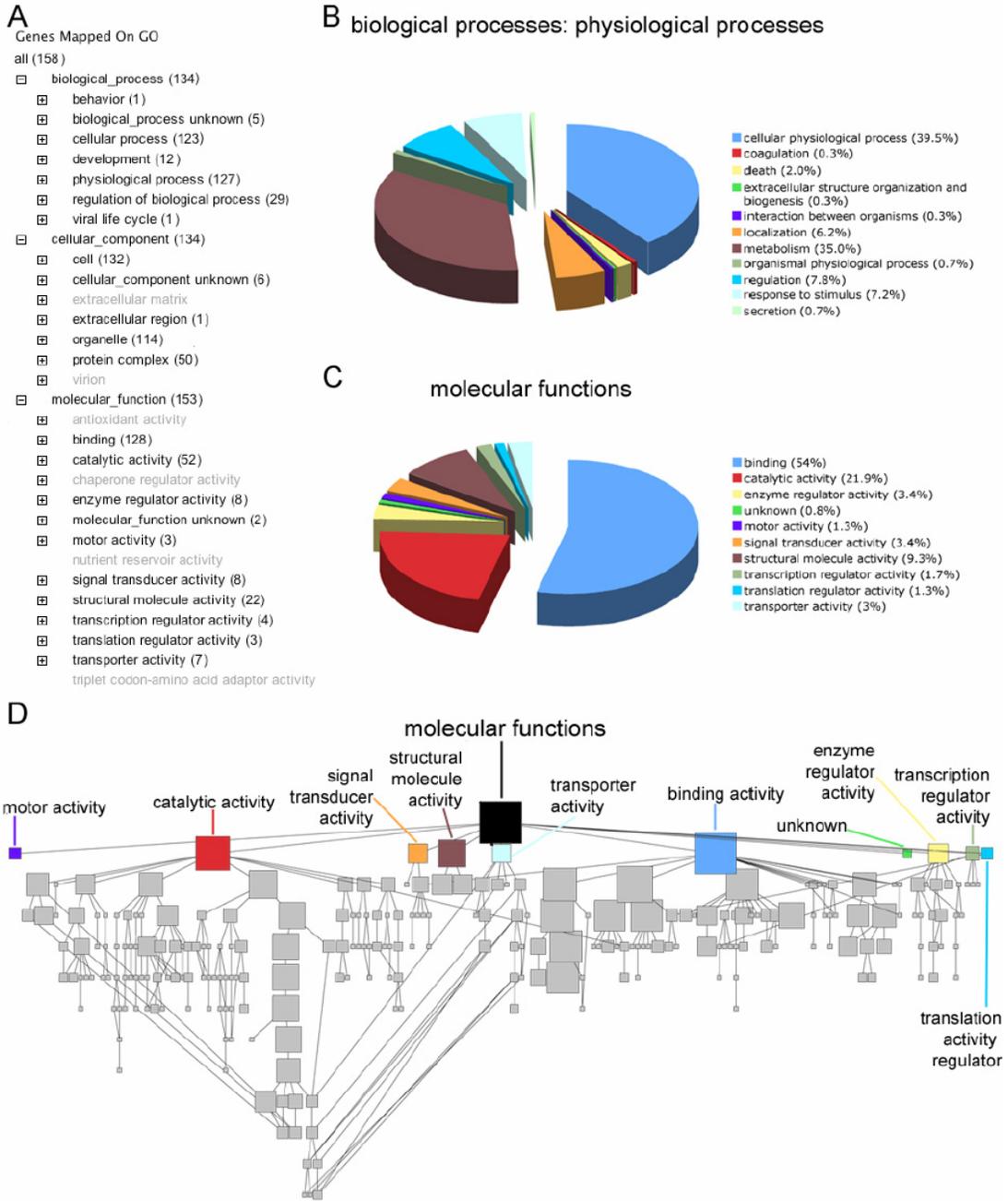
IPI00028888.1 SWI	Splice Isoform 1 of Heterogeneous nuclear ribonucleoprotein D0	20.3	355	38434	7.8	1
IPI00220684.1 SWI	Splice Isoform 3 of Heterogeneous nuclear ribonucleoprotein D0	23.5	306	32835	8.2	1
IPI00045109.1 TRE	Histone 1 H2aa	24.4	131	14233	10.9	1
IPI00216730.1 TRE	Histone H2A	24.6	130	13995	10.9	1
IPI00059366.3 TRE	H2A histone family member Y isoform 2	21.3	371	39489	9.8	1
IPI00304171.2 SWI	Core histone macro-H2A.1	23.3	339	36113	10	1
IPI00148096.1 SWI	H2A histone family member Y isoform 1	21.4	369	39183	9.8	1
IPI00100460.2 TRE	FLJ10514 protein	3.1	645	73563	8	1
IPI00514846.2 ENS	DJ383J4.4	3.0	665	75593	8	1
IPI00169383.2 SWI	Phosphoglycerate kinase 1	10.8	416	44484	8.1	1
IPI00179330.5 SWI	Ubiquitin and ribosomal protein S27a	19.9	156	17905	9.6	1
IPI00550025.2 TRE	Ubiquitin	5.1	609	68492	7.6	1
IPI00514655.1 ENS	17 kDa protein	20.9	148	16532	9.1	1
IPI00456429.2 SWI	Ubiquitin and ribosomal protein L40 0 precursor	24.2	128	14728	9.8	1
IPI00418813.1 TRE	Hypothetical protein FLJ46113	16.7	186	21127	9.6	1
IPI00387164.1 TRE	Hypothetical protein FLJ32377	8.0	388	43649	6.4	1
IPI00328348.4 ENS	16 kDa protein	22.1	140	16028	9.9	1
IPI00215893.8 SWI	Heme oxygenase 1	8.3	288	32819	8.2	1
IPI00293078.1 SWI	Splice Isoform 1 of Probable ATP-dependent RNA helicase DDX27	5.7	796	89835	9.3	1
IPI00329801.9 SWI	Annexin A5	8.8	319	35806	5	1
IPI00384342.1 SWI	Splice Isoform 2 of Retinoid X receptor interacting protein 110	3.6	553	61325	5.6	1
IPI00552059.1 SWI	Splice Isoform 3 of Retinoid X receptor interacting protein 110	3.1	641	70744	5.1	1
IPI00396071.3 SWI	Splice Isoform 1 of Retinoid X receptor interacting protein 110	2.8	719	79728	5.4	1
IPI00397526.1 SWI	Myosin heavy chain nonmuscle type B	3.7	1976	228937	5.5	1
IPI00410128.1 REF	Protein Phosphatase 1 catalytic subunit alpha isoform 2	6.3	286	32595	6.1	1
IPI00410693.3 SWI	Splice Isoform 1 of Plasminogen activator inhibitor 1 RNA-binding protein	3.9	408	44965	8.6	1
IPI00470498.1 SWI	Splice Isoform 3 of Plasminogen activator inhibitor 1 RNA-binding protein	4.1	393	43135	8.4	1
IPI00470497.3 SWI	Splice Isoform 2 of Plasminogen activator inhibitor 1 RNA-binding protein	4.0	402	44257	8.7	1
IPI00412714.2 SWI	Hypothetical protein DKFZp686P17171	4.1	387	42441	8.4	1
IPI00414481.2 SWI	Hypothetical protein DKFZp686O0870	1.4	2112	238179	7.5	1
IPI00442863.1 TRE	Hypothetical protein FLJ26493	9.8	307	34576	5.4	1
IPI00470525.2 SWI	Glycogen phosphorylase liver	1.7	847	97222	6.7	1
IPI00479359.5 SWI	Ezrin	2.9	585	69268	6.3	1
IPI00549319.2 TRE	Villin 2	2.9	586	69242	6.3	1
IPI00005198.1 TRE	NF45 protein	8.4	406	44697	8.1	1
IPI00514514.1 TRE	Interleukin enhancer binding factor 2 45kDa	8.7	390	43062	5.3	1
IPI00007334.1 SWI	Splice Isoform 1 of Apoptotic chromatin condensation inducer in the nucleus	4.5	1341	151887	6.4	1
IPI00010720.1 SWI	T-complex protein 1 epsilon subunit	3.1	541	59671	5.6	1
IPI00017339.1 SWI	Splicing factor 3B subunit 4	4.7	424	44386	8.6	1
IPI00017726.1 SWI	Splice Isoform 1 of 3-hydroxyacyl-CoA dehydrogenase type II	18.8	261	26923	7.8	1
IPI00336094.3 SWI	Splice Isoform 2 of 3-hydroxyacyl-CoA dehydrogenase type II	19.4	252	25984	7.2	1
IPI00026215.1 SWI	Flap endonuclease-1	4.2	380	42593	8.6	1
IPI00176590.3 REF	PREDICTED: similar to Heterogeneous nuclear ribonucleoprotein A1 (Helix-destabilizing protein) (Single-strand binding protein) (hnRNP core protein A1) (HDP-1) (Topoisomerase-inhibitor suppressed)	23.1	321	34257	8.4	1
IPI00549923.2 SWI	HNRPA1 protein	23.1	320	34180	9.2	1
IPI00465365.2 REF	Heterogeneous nuclear ribonucleoprotein A1 isoform a	23.1	320	34196	9.2	1
IPI00411329.1 ENS	29 kDa protein	28.2	262	29059	9.1	1
IPI00215965.1 SWI	Heterogeneous nuclear ribonucleoprotein A1 isoform b	19.9	372	38747	9.1	1
IPI00177817.4 SWI	Splice Isoform SERCA2A of Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	3.0	997	109691	5.4	1
IPI00219078.5 SWI	Splice Isoform SERCA2B of Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	2.9	1042	114757	5.3	1

IPI00217975.3 SWI	Lamin B1	7	585	66277	5.2	1
IPI00220642.4 SWI	14-3-3 protein gamma	15.4	246	28171	4.9	1
IPI00249267.1 REF	PREDICTED: similar to H2A histone family member Z	29.7	128	13449	10.4	1
IPI00293260.1 TRE	ER-resident protein ERdj5	2.8	793	91108	7.6	1
IPI00550334.1 TRE	Hypothetical protein FLJ14741	2.8	793	91080	7.2	1
IPI00394788.2 TRE	OTTHUMP00000039883	2.2	985	107340	6.3	1
IPI00003886.1 TRE	E2IG3	4.3	560	63569	9.5	1
IPI00477497.2 ENS	66 kDa protein	4.2	576	65504	9.6	1
IPI00005719.1 SWI	Splice Isoform 1 of Ras-related protein Rab-1A	7.8	205	22678	6.2	1
IPI00374519.1 TRE	PREDICTED: similar to RAB1B member RAS oncogene family	8.0	201	22017	5.4	1
IPI00334174.1 SWI	Splice Isoform 2 of Ras-related protein Rab-1A	11.3	141	15331	8.1	1
IPI00185217.2 SWI	Splice Isoform 3 of Ras-related protein Rab-1A	12.4	129	13903	5.6	1
IPI00008964.3 SWI	Ras-related protein Rab-1B	8.0	201	22171	5.7	1
IPI00019880.1 SWI	47 kDa heat shock protein precursor	6.2	417	46267	8.3	1
IPI00455917.1 REF	PREDICTED: similar to Collagen-binding protein 2 precursor (Colligin 2) (Rheumatoid arthritis related antigen RA-A47)	6.6	392	43537	7.5	1
IPI00442080.1 TRE	Hypothetical protein FLJ16712	10.0	261	27769	9.6	1
IPI00032140.2 SWI	Collagen-binding protein 2 precursor	6.2	418	46441	8.7	1
IPI00023972.3 SWI	DEAD-box-protein 47	4.0	455	50647	9.1	1
IPI00397372.1 REF	DEAD (Asp-Glu-Ala-Asp) box polypeptide 47 isoform 2	4.4	406	45169	9.2	1
IPI00029628.1 SWI	Reticulocalbin 2 precursor	13.2	317	36876	4.4	1
IPI00031691.1 SWI	60S ribosomal protein L9	9.9	192	21863	10	1
IPI00106509.2 SWI	Splice Isoform 4 of Heterogeneous nuclear ribonucleoprotein A/B	14.4	284	31233	9.3	1
IPI00334713.1 SWI	Splice Isoform 3 of Heterogeneous nuclear ribonucleoprotein A/B	14.4	285	30588	7.9	1
IPI00334587.1 SWI	Splice Isoform 2 of Heterogeneous nuclear ribonucleoprotein A/B	12.3	332	35968	6.9	1
IPI00329355.2 SWI	Splice Isoform 1 of Heterogeneous nuclear ribonucleoprotein A/B	12.4	331	36613	9	1
IPI00221093.6 SWI	40S ribosomal protein S17	16.4	134	15419	9.8	1
IPI00414603.2 REF	PREDICTED: similar to 40S ribosomal protein S17	16.3	135	15602	9.6	1
IPI00549248.3 TRE	Nucleophosmin	23.8	298	32945	4.8	1

**Table 2:** Proteins identified in HA-HAP-1A transiently transfected HEK293 cells. Listed identified proteins were acquired under the same criteria as outlined in Table 1. Proteins that are subsets of each other are grouped within one cell. Given are the locus, description, sequence coverage in % (SeqCov), amino acid length of the protein (AA), molecular weight of the protein in Daltons (Mw) and the isoelectric point pI. Lowest pI: <sup>1</sup>; highest pI: <sup>2</sup>; lowest Mw: <sup>3</sup> and highest Mw: <sup>4</sup> are indicated.



**Figure S1:** Western blot analysis of Neutravidin-purified newly synthesized proteins in transiently HA-HAP-1A transfected HEK293 cells. For immunoblot analysis, equal amounts of FLAG-Biotin-alkyne tagged whole cell lysates of AHA and methionine-treated HEK293 cells before (“input”) and after (“supernatant”) incubation with Neutravidin affinity matrix and equal amounts of protein-bound matrix (“NA-bound”) were subjected to Western blot analysis using antibodies as indicated. Sizes of marker proteins are indicated on the left margin.



**Figure S2:** Caption on following page.

**Figure S2:** Gene ontology analysis of newly synthesized proteins in transiently HA-HAP1A transfected HEK293 cells. Proteins identified in this study covered all of the “biological process” or most of the “cellular components” and “molecular function” categories at GO level 2. GO analysis was performed following the procedure described in ref 20. Briefly, the gene symbol for every candidate was loaded in the GoMiner program to display coverage of the proteomics-identified proteins with respect to GO category. Of the 194 candidates, 158 finally contributed to the coverage. (A) Tree-like representation of GO categories at level 2. Numbers in brackets indicate the number of proteins represented in this category. Note that a protein can be represented in more than one category. (B) and (C) Pie representation of GO category distribution for the biological process subcategory “physiological process” (B) and molecular functions (C). 12 from 16 possible “physiological process” categories are covered and indicated on the right side of the diagram (except “bioluminescence,” “photosynthesis,” “rhythmic process,” and “seed dormancy”). (D) The 194 proteins identified were imported into the GoMiner program to generate a directed acyclic graph (DAG) based on current annotations in the GO database. Each node of the DAG represents one GO category at various levels, and the size of each node corresponds to the number of proteins associated to this category. Several GO categories are labeled, and the best represented ones are “catalytic activity” and “binding” containing generally highly expressed “housekeeping” genes like Pyruvate kinase M2, Alpha enolase, members of the ribonucleoprotein family for mRNA-binding (heterogeneous nuclear ribonucleoprotein K), or the structural proteins Actin and beta-Tubulin.

Position on aa level	z	CalcM+H+ w/o modifications	CalcM+H+ w/ modifications	ObsM+H+	XCorr	DeltCN	Sequence (N → C)	Red
76-88	2	1449.62	-	1449.43	4.2756	0.3131	SMIIGDSDAPWTR	2
76-88	2	1449.62	1444.52	1445.11	3.3831	0.3836	SM*IIGDSDAPWTR	2
89-98	2	1184.34	-	1184.49	2.8274	0.2844	YVFQGPYGP	3
99-112	2	1389.59	1399.59	1399.26	2.8304	0.3966	ATGL#GTGKAEGIW K	1
121-129	2	955.06		955.77	2.85	0.3	RPGVSGPER	1
249-266	2	2096.22	-	2096.17	3.9956	0.2415	DQDQQHDHPYGA PKPHPK	2
412-443	3	3423.81	3463.81	3462.54	3.6797	0.1732	SSMPAGSVTHYGY SVPL#DAL#PSFPE TL#AEEL#R	2
448-459	2	1518.77	-	1520.58	2.8989	0.1197	KFITDPAYFMER	1
449-459	2	1390.59	-	1390.36	3.0694	0.2201	FITDPAYFMER	2
449-459	2	1380.59	1385.49	1385.34	2.6006	0.2537	FITDPAYFM*ER	1
541-561	2	2156.52	-	2155.38	4.5354	0.5281	MNVVVSAL#EASGL GPSHLDK	1
541-561	2	2156.52	2176.32	2175.38	4.1054	0.3858	M*NVVVSAL#EASGL L#GPSHL#DM*K	2
562-576	2	1817.99	-	1816.86	3.7434	0.3708	YVLQQLSNWQDAH SK	2
562-576	2	1817.99	1837.99	1835.87	2.9319	0.2198	YVL#QQL#SNWQD AHSK	1
592-603	2	1270.39	-	1270.66	2.9664	0.3495	RGHPASGTSFR	2

**Table S1:** Peptides identified in HA-HAP1A. For each peptide the position within the amino acid (aa) sequence, the charge state (z), the calculated mass without (CalcM+H+ w/o) or with (CalcM+H+ w/ ) modifications, as well as the observed mass (ObsM+H+) in Dalton are given. Furthermore, XCorr and DeltCN values, the peptide sequence, and the redundancy (Red.) of the identification are indicated. “\*” indicates exchange of methionine for AHA and “#” indicates d<sub>10</sub>-Leu.

experiment	no tryptic status required single peptide peptides not purged at least one modified peptide			partially tryptic status required single peptide peptides not purged at least one modified peptide			fully tryptic status required single peptide peptides not purged at least one modified peptide		
	n <sub>real</sub> +n <sub>rand</sub>	n <sub>rand</sub>	%fal	n <sub>real</sub> +n <sub>rand</sub>	n <sub>rand</sub>	%fal	n <sub>real</sub> +n <sub>rand</sub>	n <sub>rand</sub>	%fal
042705_AHA	656	104	31.7	442	29	13.1	352	0	0
050405_AHA	1392	112	16.1	1189	26	4.4	922	0	0
050605_AHA	1913	138	14.4	1625	24	3.0	1373	0	0
051505_AHA	1813	125	13.8	1625	33	4.1	1422	0	0
042605_MET	176	112	127.3	40	23	115.0	1	1	200.0
050805_MET	418	148	70.8	114	43	75.4	10	2	40.0
050905_MET	329	91	55.3	98	30	61.2	22	3	27.3
051605_MET	251	84	66.9	40	21	105.0	0	0	0

**Table S2:** False positive rates for protein identification (according to Peng et al., ref. 18). Peptide spectra from the four sets of experiments of AHA and methionine were searched against an extended database containing, in addition to the real database, a randomized one using different search constraints for DTASelect as stated in the first row of each table. Number of real matches (n<sub>real</sub>) and matches to the randomized database (n<sub>rand</sub>) are used to determine the false positive rate (%fal) according to  $\%fal = 2 * n_{rand} / (n_{real} + n_{rand})$ .