Principles of Addressing Specificity in Promiscuous Ligand-Receptor Systems

Thesis by
Christina Su

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

2022
Defended June 14, 2021

Christina Su
ORCID: 0000-0002-9223-9777

Acknowledgments

Writing this thesis was done alone in my room, but the work described in it was anything but solitary. First and foremost, I would like to thank my advisor, Michael Elowitz. Even now, I still sometimes find myself in disbelief that I have had the opportunity to learn from such a passionate, brilliant, and innovative scientist. He has been an incredible role model for conducting interesting, rigorous research; communicating questions and results clearly; and fostering a collaborative, supportive atmosphere. I am forever grateful for his unfailing patience and kindness in guiding my progress as a scientist.

I have been incredibly lucky to have not one but two wonderful mentors. Yaron Antebi was my mentor throughout my early years in the lab and somehow found the time to continue providing guidance even after starting his own lab and having his own students to mentor. I could not have asked for a more intelligent, unflappable, and fun person to work with. Regardless of what technical problem or reviewer concern came up, I could always count on him for an efficient and effective plan of attack. I also want to thank the other members of Team BMP. Heidi Klumpe has been my partner in crime throughout our graduate school adventures. She is a phenomenal scientist with more skills than I can describe, but what stands out most to me is how she brightens every community. James Linton has been a constant source of knowledge and guidance, and his breadth and depth of expertise, not to mention the quality and quantity of his science, are truly inspirational. Arvind Murugan has brought a completely new perspective to the project and has been a joy to work with. Finally, it is wonderful to see that the BMP torch will be carried on by someone just as dedicated and positive, Matthew Langley.

I want to thank everyone else in the Elowitz Lab for their camaraderie, support, and many wonderful discussions. They have made the lab an always welcoming and happy place to be. I am particularly grateful to Rachael Kuintzle and Leah Santat for listening patiently to the same story and providing insightful feedback during subgroup meetings; Jo Leonardo, James, and Leah for keeping the lab running smoothly; Grace Chow, Duncan Chadly, Fangyuan Ding, and Nico Pelaez for being great office-mates; and Mark Budde for joining me in delightful discussions and Yogurtland trips. I am also grateful to the other students in my lab, Ph.D. program, and MSTP, who have helped make the challenges of graduate school less intimidating and the good times more fun.

My committee members—Lea Goentoro, Matt Thomson, and David Chan—have provided excellent feedback and guidance throughout the process. I have had the privilege to work with undergraduate students Jacob Bradley and Akshay Yeluri as well as rotation student John Marken. I am particularly grateful to Akshay. He was exceptionally hard-working, intelligent, and engaged, and it was such a pleasure to work with him.

Finally, I am grateful to the many people outside the lab who have helped make this work possible. Grace, Dennis Edmonds, and Joe Chen have been there through the successes and failures of my Ph.D. journey. My parents, Tom and Shirley Su, and my brother, Jeffrey Su, have always been unconditionally supportive of my academic pursuits. Last but not least, I am so grateful to Darvin Yi, whose humor, openness, and drive have been an inspiration to me. Thank you for being the best partner I could imagine.

Abstract

In multicellular organisms, a relatively small number of highly conserved signaling pathways are used to enable intercellular communication. While the underlying molecular components and interactions are increasingly well understood, a fundamental mystery is how the diverse cell types of the body can be so precisely coordinated by so few pathways. It has long been known that different cell types exhibit varied responses to molecular signals, and it is unclear how this cell type specificity arises. In this work, we take a different perspective on this question and explore how cell type specificity can be generated at the level of intracellular signal. We refer to this ability to selectively activate different cell types as "addressing." By eliminating the complexity of considering downstream pathway effectors, we are able to more comprehensively understand how cell type specificity can arise in spite of—or because of—promiscuity in ligand-receptor interactions. We focus on the bone morphogenetic protein (BMP) pathway as an ideal example. This pathway is essential in development, is of therapeutic interest in an array of pathologies, and has proven amenable to theoretical and experimental analysis. We first describe a minimal model of the pathway and identify what types of response functions can be achieved. We show that each layer of computation, from the formation of signaling complexes to the activation of downstream second messenger, can provide nontrivial integrations of ligand inputs. We then extend this analysis to systems with multiple cell types that may vary in receptor expression profile. The diverse response functions of this pathway enable systems in which different cell types or sets of cell types may be addressed with high specificity. In particular, the BMP pathway can address multiple cell types with high capacity, flexibility, and robustness. Taken together, these results provide a framework for understanding how molecular promiscuity in signaling pathways can, in fact, enable cellular specificity in pathway responses.

Published Content and Contributions

Su, C.J., Murugan, A., Linton, J.M., Yeluri, A., Bois, J., Klumpe, H., Langley, M.A., Antebi, Y.E., and Elowitz, M.B. (2022). Ligand-receptor promiscuity enables cellular addressing. *Cell Syst.* 13, 408–425. DOI 10.1016/j.cels.2022.03.001.

C.J.S. conceptualized the research, designed the study, developed the mathematical models, performed the computational analysis, and wrote the manuscript. The authors have a patent related to this work (U.S. patent number 10,527,631).

Klumpe, H.E., Langley, M.A., Linton, J.M., Su, C.J., Antebi, Y.E., and Elowitz, M.B. (2022). The context-dependent, combinatorial logic of BMP signaling. *Cell Syst.* 13, 388–407. DOI 10.1016/j.cels.2022.03.002.

C.J.S. developed mathematical models, performed parameter fitting, and gave feedback on the work.

Ding, F., Su, C., Chow, K.-H.K., Liang, G., and Elowitz, M.B. (2022). Dynamics and functional roles of splicing factor autoregulation. *Cell Rep.*, accepted. *bioRxiv*. DOI 10.1101/2020.07.22.216887.

C.S. developed mathematical models, performed parameter fitting, and participated in writing of the manuscript.

Antebi, Y.E., Linton, J.M., Klumpe, H., Bintu, B., Gong, M., Su, C., McCardell, R., and Elowitz, M.B. (2017). Combinatorial Signal Perception in the BMP Pathway. *Cell* 170, 1184–1196. DOI 10.1016/j.cell.2017.08.015.

C.S. performed modeling and simulations, gave feedback on the work, and participated in revision of the manuscript.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Intercellular Communication

The emergence of life is remarkable, the evolution of multicellular life even more so. Every human living today comprises approximately 40 trillion cells that originated from a single fertilized egg (Bianconi et al., 2013). These cells represent hundreds of different cell types with specialized functions (Vickaryous and Hall, 2006), yet all work in tandem to enable development and homeostasis. Intercellular communication systems allow these distinct cell fates to act in concert, from the coordinated contraction of myocytes for movement to the synchronized release of digestive enzymes after eating. As such, analyzing how signals can be directed to appropriate cell types is essential to understanding how these communication systems function in vivo.

These biological communication systems can be divided into two main types: signaling pathways and neuronal transmission. In intercellular signaling pathways, sending cells produce signaling molecules known as ligands that can interact with receptors on receiving cells to transduce responses (Nair et al., 2019). Most cells can both send and receive signals, and these chemical signals can be transmitted at different length scales (Alberts et al., 2008). At the largest or most "public" scale, signals can be secreted into the bloodstream and "broadcast" across the entire body. These molecules, known as hormones, mediate endocrine signaling. At a smaller scale, ligands can diffuse within their local environment, performing paracrine signaling (Francis and Palsson, 1997). Alternatively, the ligand may be expressed

directly on the plasma membrane of the signaling cell and activate receptors on an adjacent cell that is in direct contact with the sender cell; this contact-dependent signaling is known as juxtacrine signaling. Finally, cells can respond to their own signals through autocrine signaling.

In addition to signaling pathways, many multicellular organisms feature a specialized nervous system that processes inputs and coordinates outputs across different parts of the body. This system uses electrochemical signals to transmit information. Briefly, specialized cells known as neurons fire electrical impulses along their axons. An impulse induces the release of chemicals called neurotransmitters at the axon terminal, where the signaling neuron forms synapses with the receiving cell(s). These neurotransmitters diffuse across the gap between the sending cell's axon and the target cell's membrane. The target cell can then respond by producing its own electrical signal or otherwise enacting a desired response (Cotman and McGaugh, 1980). However, intercellular communication remains an essential function across all cells, most of which are not neurons. For typical cells, signal transduction is largely implemented by signaling pathways using chemical mediators.

1.2 Signaling Pathways

Although these chemical signals are utilized across hundreds of cell types, they represent a relatively small number of signaling pathways that are highly conserved in development. These key signaling pathways (typically named for a component of the pathway) include the Notch, Hippo, transforming growth factor β (TGF-β), Wnt, Hedgehog, receptor tyrosine kinase (RTK), Janus kinase (JAK)-signal transducer and activator of transcription (STAT) (JAK-STAT), Jun kinase (JNK), nuclear factor-κB (NF-κB), and nuclear receptor pathways (Basson, 2012; Housden and Perrimon, 2014; Perrimon et al., 2012). Among these, the Notch and Hippo pathways are juxtacrine, while the remaining are paracrine. Each pathway may have

multiple variants of ligands and receptors, along with a variety of other components that modulate or regulate signaling activity.

While the importance of cellular communication is clear, untangling the complex underlying interactions is a major challenge and an abiding goal in biology. The Hodgkin-Huxley model describing how action potentials are propagated in neurons was first published in 1952 (Hodgkin and Huxley, 1952). Initial evidence of chemical signaling pathways was discovered around the same time. While the concept that chemicals secreted into blood could affect distant cells had emerged earlier (Nair et al., 2019; Starling, 1914), discoveries of the underlying molecular components were only reported starting in the 1950s. Cohen and Levi-Montalcini discovered epidermal growth factor (EGF) and nerve growth factor (NGF) (Cohen, 1962; Levi-Montalcini and Hamburger, 1951). These studies identified particular examples of "first messengers" (now called ligands), or the extracellular molecules that initiate signaling by binding to cell surface receptors. Concurrently, Sutherland discovered that epinephrine and glucagon result in the production of cyclic AMP (Rall et al., 1956; Sutherland and Rall, 1957). This result represented the first identification of "second messengers," or the intracellular molecules that act within a cell to mediate the cellular response to ligands. The development of this conceptual framework enabled the emergence of signal transduction as a field (Hunter, 2000).

Subsequent technological advances in molecular biology have greatly expanded our understanding of the molecular players in signaling pathways. These discoveries include not only the ligands and receptors for the various pathways mentioned above but also their overall architectures, which may encompass many other molecules that transduce or modify ligand signals. For example, the Wnt pathway involves binding of Wnt proteins to the Frizzled (Fz)/low-density lipoprotein (LDL) receptor-related protein (LRP) complex at the cell surface. Canonical Wnt signaling results in in-

hibition of the degradation of β-catenin, which then accumulates and translocates to the nucleus to act as a transcriptional regulator (Logan and Nusse, 2004). In the JAK-STAT pathway, ligand-receptor binding leads to conformational changes in the receptor and activation of receptor-associated JAKs. They are activated through trans-phosphorylation and can in turn phosphorylate select tyrosine residues on the cytokine receptor, allowing docking of STATs and other signaling components. STATs are then themselves phosphorylated by JAKs and, upon activation, translocate to the nucleus to effect cellular responses (Kisseleva et al., 2002; O'Shea et al., 2002). These examples illustrate how signaling pathways can be understood within an overarching conceptual framework: ligand-receptor binding at the cell surface triggers activation of second messengers that enable the intracellular response (here, by translocating to the nucleus and regulating gene expression). They also demonstrate some of the variations in implementation across different pathways, such as inhibition of degradation (Wnt) or phosphorylation-induced association (JAK-STAT). Nonetheless, these mechanisms often recur across pathways (Krebs and Beavo, 1979).

Knowledge about intracellular signaling pathways has advanced concurrently with our understanding of signal-initiating events. One essential class of signaling proteins is the set of guanine nucleotide-binding proteins (G proteins), both the small monomeric G proteins and the heterotrimeric G proteins. These proteins bind guanine nucleotides and can cycle between inactive guanosine diphosphate (GDP)-bound and active guanosine triphosphate (GTP)-bound states through their GTPase activity (Hamm, 1998; Takai et al., 2001). Upon activation by G protein-coupled receptors, heterotrimeric G proteins undergo GDP release and then bind GTP at the Gα subunit, allowing dissociation of the Gβγ subunits. The activated $G\alpha^*$ subunits interact with effector enzymes and are divided into families based on these interactions. For example, G_s activates adenylyl cyclase, G_i inhibits adenylyl

cyclase, and G_q activates phospholipase C (PLC) (Neves et al., 2002). Ras, the best-characterized member of the small G proteins, is well known for its role in activating the intracellular mitogen-activated protein (MAP) kinase cascade (Mor and Philips, 2006).

While G proteins are ubiquitous and highly evolutionarily conserved signaling components, other work has revealed that signaling transducers are not limited to proteins. For example, lipids were identified early on as part of the response to signaling (Hokin and Hokin, 1953). Subsequent work unraveled more details about phosphoinositide (PI) pathways in signal transduction. In one, phosphatidylinositol is phosphorylated to phosphatidylinositol 4,5-bisphosphate, which is then hydrolyzed by PLC to two second messengers: inositol 1,4,5-trisphosphate (IP_3) and diacylglycerol (DAG) (Kapeller and Cantley, 1994). IP_3 acts to release calcium ions (Ca^{+2}) from intracellular stores, and Ca^{+2} itself can serve as an important second messenger for a variety of cellular processes, often acting through Ca^{+2}-sensing proteins (Cheung, 1980; Chin and Means, 2000; Klee et al., 1980; Means and Dedman, 1980). DAG acts by activating protein kinase C (Bell and Burns, 1991). An alternate pathway of PI metabolism involves PI 3-kinase, which generates 3-phosphoinositides that are not hydrolyzed by PLC (Kapeller and Cantley, 1994). In addition to lipids and ions, even gases like nitric oxide have emerged as important second messengers (Stamler et al., 1992).

This increased molecular understanding of signaling pathways has also enabled major advancements in synthetic control and therapeutic applications. Genetic mutations in signaling pathway components have been linked with various pathologies (Wang et al., 2014). Compounds to activate or inhibit important effectors have been developed, many of significant clinical impact (Cohen et al., 2021). Nonetheless, a deeper systems-level understanding of these pathways has not yet been achieved.

It remains unclear what communication capabilities these pathways provide for an individual cell type or for a tissue composed of multiple cell types.

1.3 Signaling Pleiotropy

While the molecular understanding of signaling pathways has broadened dramatically, the mystery of how so many cell types can be precisely coordinated by so few pathways remains a key question. Any given cell may be exposed to numerous different signaling molecules in its environment, and each cell must selectively respond to particular signals based on its particular function.

The relationship between signals and downstream cellular responses is complex and context-dependent. A given molecular signal can induce a range of cellular responses across cell types. For example, acetylcholine acts to decrease the rate and strength of contraction in cardiac muscle but stimulates contraction in skeletal muscle; in salivary gland, it induces glandular secretion (Alberts et al., 2008). Conversely, a single cell type can show a range of cellular responses to different signals. In a given cell, activation of the PI 3-kinase pathway by the insulin receptor may induce metabolic responses while activation of the same pathway by growth factor receptor does not (Hunter, 2000). These observations demonstrate specificity of cellular responses depending on the particular signal or cell type. However, how these differential responses arise is still unclear.

Various models have been proposed to explain these phenomena. In models of intrinsic specificity, distinct responses are programmed within the receiving cells. For example, two cell types might express distinct intracellular pathway components or downstream effectors (Tan and Kim, 1999). In the former, a receptor capable of activating multiple intracellular signaling pathways can activate different pathways and therefore different responses if each cell type expresses components for

a different pathway. In the latter, cell types might express different transcription factors, thus generating different changes in gene expression upon activation. These hypotheses suggest that different cellular responses can be generated in a permissive manner by leveraging intrinsic differences in the responding cells.

Alternatively, models of extrinsic specificity imply that the signals themselves control the resulting responses. For example, different signaling kinetics or combinatorial integration of multiple signaling pathways could allow for different responses (Tan and Kim, 1999). With the former, cell types might respond differently depending on whether signals are transient or sustained. With the latter, multiple signaling pathways could synergize (or show other combinatorial interactions) to provide a larger set of responses from a smaller number of signals. These hypotheses imply that signals act in an instructive manner to induce different cellular responses.

It is likely that each of these models may hold in different contexts. For example, the LET-23 RTK provides support for intrinsically generated specificity. This receptor functions in multiple tissues in *Caenorhabditis elegans* development (Aroian and Sternberg, 1991). Several of these tissues rely on the Ras-mediated MAP kinase pathway (Church et al., 1995; Sundaram et al., 1996), but one instead expresses and requires components of the IP_3 pathway (Clandinin et al., 1998). Conversely, the Notch signaling pathway gives one example of extrinsically generated specificity. The ligand Dll1 activates Notch1 in pulses and leads to upregulation of Hes1, while the ligand Dll4 yields sustained activation of Notch1 and induces upregulation of Hey1 and HeyL (Nandagopal et al., 2018). These models all provide possible mechanisms for the outstanding question of how different cell types exhibit varied responses to molecular signals.

1.4 Cellular Addressing

In this work, we approach the question of cellular specificity from a unique angle by asking how signals can generate different levels of pathway activation across different cell types. In other words, we ask how cell type specificity can be realized at the level of intracellular signal rather than overall phenotypic response, eliminating the complexity added through downstream pathway effectors. While analyzing further layers of computation is essential for a complete understanding of how cell types respond differentially to signals, focusing on the level of the signal itself enables us to better understand the operational principles of signal transduction and determine what functions can be achieved at each level of complexity. We refer to this ability to selectively activate different cell types as "addressing."

As an analogy, any communication system, from mail to email, must provide this ability to address the desired recipient(s). In some cases, this specificity can be achieved through physical proximity. Consider, for example, early telephone systems, where switchboard operators physically connected the caller to the recipient. A similar situation applies to neuronal transmission, where a given neuron can only signal to cells with which it forms synaptic contacts. Thus, mapping the complex connectivity of the nervous system is a monumental task (Micheva et al., 2010), but it is relatively straightforward to infer the cell(s) that a cell may send information to or receive information from. Similarly, understanding how signals may be addressed in a juxtacrine pathway is relatively clear; a cell may only send or respond to its immediate neighbors. An analogy might be a conversation at a loud party, where any given individual can only converse with someone immediately adjacent.

However, it is unclear whether and how secreted signals can be similarly targeted. In the many signaling pathways involving secreted ligands, signals enter the local environment and/or bloodstream and can affect a large number of cells. One possibility

is that different cells can simply express a limited set of receptors in order to respond to a limited set of ligands, similar to listening to a particular frequency on the radio. However, ligands and receptors are often expressed in combinations (Diez-Roux et al., 2011; Faber et al., 2002; Glister et al., 2004; Onagbesan et al., 2003; Simic and Vukicevic, 2005). Furthermore, many pathways exhibit ligand-receptor promiscuity, where ligands can activate multiple different receptors to varying extents. These pathways include BMP (Heldin et al., 1997; Massagué, 1998; Mueller and Nickel, 2012; Nickel and Mueller, 2019; Schmierer and Hill, 2007), Notch (Shimizu et al., 2000a, 2000b), Wnt (Llimargas and Lawrence, 2001; Wodarz and Nusse, 1998), Eph-Ephrin (Dai et al., 2014), and FGF (Ornitz et al., 1996; Zhang et al., 2006). As such, the intuitive expectation might be that widespread pathway activation would be observed. Therefore, it remains unclear whether cell type specificity can emerge at the level of pathway activation or whether such specificity requires the additional layer of downstream signal transducers.

In this work, we investigate whether and how cellular addressing can be achieved in a prototypical paracrine pathway. Given the ubiquity of promiscuous ligand-receptor interactions in key signaling pathways, we focus on analyzing the potential for addressing in a promiscuous pathway architecture. It has remained unclear whether molecular promiscuity in ligand-receptor interactions is compatible with addressing at all and, if so, what advantages it might provide. Understanding how specific addressing of particular cell types can be achieved would not only improve understanding of intercellular signaling in physiological contexts but also facilitate development of therapeutic applications with enhanced specificity for desired cell types.

1.5 BMP Signaling

To analyze this question, we use the BMP signaling pathway as a prototypical example of a paracrine signaling pathway with promiscuous ligand-receptor interactions. Briefly, BMP ligands (part of the TGF-β superfamily) are secreted dimers that signal by interacting with a heterotetrameric receptor complex formed from dimeric type I and type II serine-threonine kinase receptors. In mammals, there are ten or more ligands that can interact promiscuously with four type I receptors and three type II receptors (Miyazono et al., 2010). Upon ligand binding, the constitutively active type II receptor phosphorylates type I receptor, which can in turn phosphorylate the receptor-regulated SMADs (R-SMADs). For BMP signaling, these second messengers are SMAD1, SMAD5, and SMAD8 (with SMAD2 and SMAD3 primarily activated by receptors for the TGF-β subfamily). These R-SMADs form a complex with the co-mediator SMAD (co-SMAD) SMAD4, which then translocates to the nucleus and acts as a transcription factor to regulate expression of target genes (Massagué, 2000; Shi and Massagué, 2003). While the specific downstream effects depend on cell type, the level of phosphorylated SMAD can be used as a cell type-independent measure of pathway activation.

First identified as inducers of bone formation (Urist, 1965), BMPs are not limited to bone and cartilage but also play vital roles in the development and normal function of heart, lung, liver, kidney, and many other systems, where they mediate a host of downstream responses (David and Massagué, 2018; Wagner et al., 2010; Wang et al., 2014).

Furthermore, dysregulation of the BMP pathway is known to contribute to a range of diseases. For example, several classes of skeletal disorders that manifest with severe longitudinal growth defects in the limbs are associated with reduced BMP signaling, such as mutations reducing secretion of the ligand GDF5 or impairing

ligand binding by the receptor BMPR1B. Conversely, excessive BMP signaling also leads to skeletal disorders; for instance, mutations of the receptor *ACVR1* that increase activation potential can produce the extraskeletal ossification seen in fibrodysplasia ossificans progressiva (FOP) (Salazar et al., 2016). Mouse transgenic models have further demonstrated the broad array of phenotypic defects resulting from mutations in TGF-β and BMP pathway components (Wu et al., 2016). BMP-related pathologies are not limited to the skeletal system. For instance, BMP signaling plays a role in inducing cardiac mesoderm formation and cardiomyocyte differentiation (Morrell et al., 2016). BMP ligands are also upregulated at sites of systemic vascular injury and in vessel walls in animal models of hypertension (Lowery and de Caestecker, 2010). Mutations in the receptor *ACVRL1* are linked to a subtype of hereditary hemorrhagic telangiectasia (HHT), which manifests with arteriovenous malformations in the lungs and other organs. Again, animal models of mutations in BMP pathway components have shown a variety of cardiovascular phenotypes (Morrell et al., 2016).

These examples illustrate a few of the essential roles that BMP signaling plays in physiology and pathology. Unsurprisingly, modulation of the BMP pathway is an area of therapeutic interest. For instance, small-molecule BMP inhibitors show potential use in regenerative medicine, such as inducing cardiomyocyte differentiation from pluripotent stem cells (Hong and Yu, 2009). BMP7 has shown effective suppression of renal fibrosis in animal models (Liu, 2006), while BMP9 has demonstrated promise in reducing glycemia in mouse models of diabetes (Chen et al., 2003). BMP signaling is of particular interest in bone injuries, and recombinant BMP2 and BMP7 have both shown promise in treating fractures (Friedlaender et al., 2001; Govender et al., 2002). However, effective delivery is an essential consideration (Anitua et al., 2008), and it remains an outstanding challenge to deliver BMP signals to desired targets while minimizing off-target activity. As such, un-

derstanding how cell type specificity can be achieved will be vital for optimizing signaling-based therapeutic approaches.

Together, these observations indicate that BMP signaling is a system of great biological importance. Beyond its significance, however, previous work suggests that the pathway represents an ideal choice for systematically investigating how signaling specificity can be achieved. Experimental measurements of pathway responses in multiple cell types exposed to multiple ligand combinations revealed that cells can "compute" combinatorial functions of ligand inputs. Different computations can be generated for a single cell type with different ligand combinations or for multiple cell types with the same ligand combination. A minimal mathematical model describing only ligand-receptor interactions (neglecting further pathway components) successfully described the types of responses observed experimentally in cell lines (Antebi et al., 2017; Klumpe et al., 2022). These results demonstrate that the BMP pathway exhibits a range of capabilities that are amenable to experimental and mathematical modeling.

1.6 Present Work

In this work, we build on this body of knowledge to model what computations are possible in the BMP pathway and how they can enable cellular addressing. To better understand the functional capabilities of this architecture, we develop a simplified mathematical model of the pathway that describes promiscuous ligand-receptor interactions and their enzymatic activation of second messenger. In Chapter 2, we perform a comprehensive analysis of the model to identify what types of cellular response functions can be achieved. This study enables deeper understanding of the functional capabilities of a promiscuous ligand-receptor architecture. In Chapter 3, we then generalize this approach to analyze systems of diverse cells, each with potentially distinct receptor expression profiles. Specifically, we show that these

diverse responses allow high-capacity, flexible, and robust addressing of cell types. Together, this work provides a framework for understanding the unique functions of a promiscuous ligand-receptor architecture and how molecular promiscuity can, counterintuitively, enable cellular specificity.

1.7 References

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2008). *Molecular Biology of the Cell*.

Anitua, E., Sánchez, M., Orive, G., and Andia, I. (2008). Delivering growth factors for therapeutics. *Trends Pharmacol. Sci.* 29, 37–41.

Antebi, Y.E., Linton, J.M., Klumpe, H., Bintu, B., Gong, M., Su, C., McCardell, R., and Elowitz, M.B. (2017). Combinatorial Signal Perception in the BMP Pathway. *Cell* 170, 1184–1196.

Aroian, R.V., and Sternberg, P.W. (1991). Multiple Functions of *let-23*, a *Caenorhabditis elegans* Receptor Tyrosine Kinase Gene Required for Vulval Induction. *Genetics* 128, 251–267.

Basson, M.A. (2012). Signaling in Cell Differentiation and Morphogenesis. *Cold Spring Harb. Perspect. Biol.* 4, a008151.

Bell, R.M., and Burns, D.J. (1991). Lipid Activation of Protein Kinase C. *J. Biol. Chem.* 266, 4661–4664.

Bianconi, E., Piovesan, A., Facchin, F., Beraudi, A., Casadei, R., Frabetti, F., Vitale, L., Pelleri, M.C., Tassani, S., Piva, F., et al. (2013). An estimation of the number of cells in the human body. *Ann. Hum. Biol.* 40, 463–471.

Chen, C., Grzegorzewski, K.J., Barash, S., Zhao, Q., Schneider, H., Wang, Q., Singh, M., Pukac, L., Bell, A.C., Duan, R., et al. (2003). An integrated functional genomics screening program reveals a role for BMP-9 in glucose homeostasis. *Nat. Biotechnol.* 21, 294–301.

Cheung, W.Y. (1980). Calmodulin Plays a Pivotal Role in Cellular Regulation. *Science* 207, 19–27.

Chin, D., and Means, A.R. (2000). Calmodulin: a prototypical calcium sensor. *Trends Cell Biol.* 10, 322–328.

Church, D.L., Guan, K.L., and Lambie, E.J. (1995). Three genes of the MAP kinase cascade, *mek-2*, *mpk-1/sur-1* and *let-60 ras*, are required for meiotic cell cycle progression in *Caenorhabditis elegans*. *Development* 121, 2525–2535.

Clandinin, T.R., DeModena, J.A., and Sternberg, P.W. (1998). Inositol Trisphosphate Mediates a RAS-Independent Response to LET-23 Receptor Tyrosine Kinase Activation in *C. elegans*. *Cell* 92, 523–533.

Cohen, P., Cross, D., and Jänne, P.A. (2021). Kinase drug discovery 20 years after imatinib: progress and future directions. *Nat. Rev. Drug Discov.* 20, 551-569.

Cohen, S. (1962). Isolation of a Mouse Submaxillary Gland Protein Accelerating Incisor Eruption and Eyelid Opening in the New-born Animal. *J. Biol. Chem.* 237, 1555–1562.

Cotman, C.W., and McGaugh, J.L. (1980). *Behavioral Neuroscience*.

Dai, D., Huang, Q., Nussinov, R., and Ma, B. (2014). Promiscuous and specific recognition among ephrins and Eph receptors. *Biochim. Biophys. Acta — Proteins Proteom.* 1844, 1729–1740.

David, C.J., and Massagué, J. (2018). Contextual determinants of TGFβ action in development, immunity and cancer. *Nat. Rev. Mol. Cell Biol.* 19, 419–435.

Diez-Roux, G., Banfi, S., Sultan, M., Geffers, L., Anand, S., Rozado, D., Magen, A., Canidio, E., Pagani, M., Peluso, I., et al. (2011). A High-Resolution Anatomical Atlas of the Transcriptome in the Mouse Embryo. *PLoS Biol.* 9, e1000582.

Faber, S.C., Robinson, M.L., Makarenkova, H.P., and Lang, R.A. (2002). Bmp signaling is required for development of primary lens fiber cells. *Development* 129, 3727–3737.

Francis, K., and Palsson, B.O. (1997). Effective intercellular communication distances are determined by the relative time constants for cyto/chemokine secretion and diffusion. *Proc. Natl. Acad. Sci. U.S.A.* 94, 12258–12262.

Friedlaender, G.E., Perry, C.R., Cole, J.D., Cook, S.D., Cierny, G., Muschler, G.F., Zych, G.A., Calhoun, J.H., LaForte, A.J., and Yin, S. (2001). Osteogenic Protein-1 (Bone Morphogenetic Protein-7) in the Treatment of Tibial Nonunions. *J. Bone Joint Surg.* 83, S151–S158.

Glister, C., Kemp, C.F., and Knight, P.G. (2004). Bone morphogenetic protein (BMP) ligands and receptors in bovine ovarian follicle cells: actions of BMP-4, -6 and -7 on granulosa cells and differential modulation of Smad-1 phosphorylation by follistatin. *Reproduction* 127, 239–254.

Govender, S., Csimma, C., Genant, H.K., Valentin-Opran, A., Amit, Y., Arbel, R., Aro, H., Atar, D., Bishay, M., Börner, M.G., et al. (2002). Recombinant Human Bone Morphogenetic Protein-2 for Treatment of Open Tibial Fractures: A Prospective, Controlled, Randomized Study of Four Hundred and Fifty patients. *J. Bone Joint Surg.* 84, 2123–2134.

Hamm, H.E. (1998). The Many Faces of G Protein Signaling. *J. Biol. Chem.* 273, 669–672.

Heldin, C.-H., Miyazono, K., and ten Dijke, P. (1997). TGF-β signalling from cell membrane to nucleus through SMAD proteins. *Nature* 390, 465–471.

Hodgkin, A.L., and Huxley, A.F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* 117, 500–544.

Hokin, M.R., and Hokin, L.E. (1953). Enzyme secretion and the incorporation of P32 into phospholipides of pancreas slices. *J. Biol. Chem.* 203, 967–977.

Hong, C.C., and Yu, P.B. (2009). Applications of small molecule BMP inhibitors in physiology and disease. *Cytokine Growth Factor Rev.* 20, 409–418.

Housden, B.E., and Perrimon, N. (2014). Spatial and temporal organization of signaling pathways. *Trends Biochem. Sci.* 39, 457–464.

Hunter, T. (2000). Signaling—2000 and Beyond. *Cell* 100, 113–127.

Kapeller, R., and Cantley, L.C. (1994). Phosphatidylinositol 3-Kinase. *BioEssays* 16, 565–576.

Kisseleva, T., Bhattacharya, S., Braunstein, J., and Schindler, C.W. (2002). Signaling through the JAK/STAT pathway, recent advances and future challenges. *Gene* 285, 1–24.

Klee, C.B., Crouch, T.H., and Richman, P.G. (1980). Calmodulin. *Annu. Rev. Biochem.* 49, 489–515.

Klumpe, H.E., Langley, M.A., Linton, J.M., Su, C.J., Antebi, Y.E., and Elowitz, M.B. (2022). The context-dependent, combinatorial logic of BMP signaling. *Cell Syst.* 13, 388–407.

Krebs, E.G., and Beavo, J.A. (1979). Phosphorylation-Dephosphorylation of Enzymes. *Annu. Rev. Biochem.* 48, 923–959.

Levi-Montalcini, R., and Hamburger, V. (1951). Selective growth stimulating effects of mouse sarcoma on the sensory and sympathetic nervous system of the chick embryo. *J. Exp. Zool.* 116, 321–361.

Liu, Y. (2006). Renal fibrosis: New insights into the pathogenesis and therapeutics. *Kidney Int.* 69, 213–217.

Llimargas, M., and Lawrence, P.A. (2001). Seven Wnt homologues in *Drosophila*: A case study of the developing tracheae. *Proc. Natl. Acad. Sci. U.S.A.* 98, 14487–14492.

Logan, C.Y., and Nusse, R. (2004). The Wnt Signaling Pathway in Development and Disease. *Annu. Rev. Cell Dev. Biol.* 20, 781–810.

Lowery, J.W., and de Caestecker, M.P. (2010). BMP signaling in vascular development and disease. *Cytokine Growth Factor Rev.* 21, 287–298.

Massagué, J. (1998). TGF-β signal transduction. *Annu. Rev. Biochem.* 67, 753–791.

Massagué, J. (2000). How cells read TGF-β signals. *Nat. Rev. Mol. Cell Biol.* 1, 169–178.

Means, A.R., and Dedman, J.R. (1980). Calmodulin—an intracellular calcium receptor. *Nature* 285, 73–77.

Micheva, K.D., Busse, B., Weiler, N.C., O'Rourke, N., and Smith, S.J. (2010). Single-Synapse Analysis of a Diverse Synapse Population: Proteomic Imaging Methods and Markers. *Neuron* 68, 639–653.

Miyazono, K., Kamiya, Y., and Morikawa, M. (2010). Bone morphogenetic protein receptors and signal transduction. *J. Biochem.* 147, 35–51.

Mor, A., and Philips, M.R. (2006). Compartmentalized Ras/MAPK Signaling. *Annu. Rev. Immunol.* 24, 771–800.

Morrell, N.W., Bloch, D.B., ten Dijke, P., Goumans, M.-J.T.H., Hata, A., Smith, J., Yu, P.B., and Bloch, K.D. (2016). Targeting BMP signalling in cardiovascular disease and anaemia. *Nat. Rev. Cardiol.* 13, 106–120.

Mueller, T.D., and Nickel, J. (2012). Promiscuity and specificity in BMP receptor activation. *FEBS Lett.* 586, 1846–1859.

Nair, A., Chauhan, P., Saha, B., and Kubatzky, K.F. (2019). Conceptual Evolution of Cell Signaling. *Int. J. Mol. Sci.* 20, 3292.

Nandagopal, N., Santat, L.A., LeBon, L., Sprinzak, D., Bronner, M.E., and Elowitz, M.B. (2018). Dynamic Ligand Discrimination in the Notch Signaling Pathway. *Cell* 172, 869–880.

Neves, S.R., Ram, P.T., and Iyengar, R. (2002). G Protein Pathways. *Science* 296, 1636–1639.

Nickel, J., and Mueller, T.D. (2019). Specification of BMP Signaling. *Cells* 8, 1579.

Onagbesan, O.M., Bruggeman, V., Van As, P., Tona, K., Williams, J., and Decuypere, E. (2003). BMPs and BMPRs in chicken ovary and effects of BMP-4 and -7 on granulosa cell proliferation and progesterone production in vitro. *Am. J. Physiol. Endocrinol. Metab.* 285, E973–E983.

Ornitz, D.M., Xu, J., Colvin, J.S., McEwen, D.G., MacArthur, C.A., Coulier, F., Gao, G., and Goldfarb, M. (1996). Receptor Specificity of the Fibroblast Growth Factor Family. *J. Biol. Chem.* 271, 15292–15297.

O'Shea, J.J., Gadina, M., and Schreiber, R.D. (2002). Cytokine Signaling in 2002: New Surprises in the Jak/Stat Pathway. *Cell* 109, S121–S131.

Perrimon, N., Pitsouli, C., and Shilo, B.-Z. (2012). Signaling Mechanisms Controlling Cell Fate and Embryonic Patterning. *Cold Spring Harb. Perspect. Biol.* 4, a005975.

Rall, T.W., Sutherland, E.W., and Wosilait, W.D. (1956). The relationship of epinephrine and glucagon to liver phosphorylase: III. Reactivation of liver phosphorylase in slices and in extracts. *J. Biol. Chem.* 218, 483–495.

Salazar, V.S., Gamer, L.W., and Rosen, V. (2016). BMP signalling in skeletal development, disease and repair. *Nat. Rev. Endocrinol.* 12, 203–221.

Schmierer, B., and Hill, C.S. (2007). TGFβ-SMAD signal transduction: molecular specificity and functional flexibility. *Nat. Rev. Mol. Cell Biol.* 8, 970–982.

Shi, Y., and Massagué, J. (2003). Mechanisms of TGF-β Signaling from Cell Membrane to the Nucleus. *Cell* 113, 685–700.

Shimizu, K., Chiba, S., Hosoya, N., Kumano, K., Saito, T., Kurokawa, M., Kanda, Y., Hamada, Y., and Hirai, H. (2000a). Binding of Delta1, Jagged1, and Jagged2 to Notch2 Rapidly Induces Cleavage, Nuclear Translocation, and Hyperphosphorylation of Notch2. *Mol. Cell. Biol.* 20, 6913–6922.

Shimizu, K., Chiba, S., Saito, T., Kumano, K., and Hirai, H. (2000b). Physical Interaction of Delta1, Jagged1, and Jagged2 with Notch1 and Notch3 Receptors. *Biochem. Biophys. Res. Commun.* 276, 385–389.

Simic, P., and Vukicevic, S. (2005). Bone morphogenetic proteins in development and homeostasis of kidney. *Cytokine Growth Factor Rev.* 16, 299–308.

Stamler, J.S., Singel, D.J., and Loscalzo, J. (1992). Biochemistry of Nitric Oxide and Its Redox-Activated Forms. *Science* 258, 1898–1902.

Starling, E.H. (1914). Discussion on the Therapeutic Value of Hormones. *Proc. R. Soc. Med.* 7, 29–31.

Sundaram, M., Yochem, J., and Han, M. (1996). A Ras-mediated signal transduction pathway is involved in the control of sex myoblast migration in *Caenorhabditis elegans*. *Development* 122, 2823–2833.

Sutherland, E.W., and Rall, T.W. (1957). The properties of an adenine ribonucleotide produced with cellular particles, ATP, Mg++, and epinephrine or glucagon. *J. Am. Chem. Soc.* 79, 3608.

Takai, Y., Sasaki, T., and Matozaki, T. (2001). Small GTP-Binding Proteins. *Physiol. Rev.* 81, 153–208.

Tan, P.B.O., and Kim, S.K. (1999). Signaling specificity: the RTK/RAS/MAP kinase pathway in metazoans. *Trends Genet.* 15, 145–149.

Urist, M.R. (1965). Bone: Formation by Autoinduction. *Science* 150, 893–899.

Vickaryous, M.K., and Hall, B.K. (2006). Human cell type diversity, evolution, development, and classification with special reference to cells derived from the neural crest. *Biol. Rev. Camb. Philos. Soc.* 81, 425–455.

Wagner, D.O., Sieber, C., Bhushan, R., Börgermann, J.H., Graf, D., and Knaus, P. (2010). BMPs: From Bone to Body Morphogenetic Proteins. *Sci. Signal.* 3, mr1.

Wang, R.N., Green, J., Wang, Z., Deng, Y., Qiao, M., Peabody, M., Zhang, Q., Ye, J., Yan, Z., Denduluri, S., et al. (2014). Bone Morphogenetic Protein (BMP) signaling in development and human diseases. *Genes Dis.* 1, 87–105.

Wodarz, A., and Nusse, R. (1998). Mechanisms of Wnt signaling in development. *Annu. Rev. Cell Dev. Biol.* 14, 59–88.

Wu, M., Chen, G., and Li, Y.-P. (2016). TGF-β and BMP signaling in osteoblast, skeletal development, and bone formation, homeostasis and disease. *Bone Res.* 4, 16009.

Zhang, X., Ibrahimi, O.A., Olsen, S.K., Umemori, H., Mohammadi, M., and Ornitz, D.M. (2006). Receptor Specificity of the Fibroblast Growth Factor Family: The Complete Mammalian FGF Family. *J. Biol. Chem.* 281, 15694–15700.

Chapter 2

Ligand-Receptor Promiscuity Enables Complex Computations

2.1 Abstract

Signaling pathways represent an integral mode of intercellular communication in multicellular organisms. As such, it is essential to understand the capabilities they provide. The bone morphogenetic protein (BMP) pathway is an ideal example for study. This pathway is used in many biological contexts, has potential therapeutic applications in a range of pathologies, and has previously been shown to allow cells to perform complex computations on ligand inputs. In this work, we utilize a mathematical model of ligand-receptor interactions in the pathway to systematically interrogate how these computations arise. We show that nontrivial integrations of inputs can occur even at the level of formation of ligand-receptor signaling complexes. Receptor multiplicity is required to achieve these computations. Further diversity of responses is added when considering the combined output of these signaling complexes at the level of pathway activity. These results provide a more systematic understanding of the computations achievable in this pathway.

2.2 Introduction

Multicellular organisms require precise coordination of diverse cell types during development and normal functioning. Core communication pathways allow for signal transmission between cells and are used recurrently in diverse biological processes. Many of these pathways exhibit promiscuous, many-to-many interactions between multiple ligand and receptor variants. These pathways include not only the

BMP (Heldin et al., 1997; Massagué, 1998; Mueller and Nickel, 2012; Nickel and Mueller, 2019; Schmierer and Hill, 2007) but also the Wnt (Llimargas and Lawrence, 2001; Wodarz and Nusse, 1998), Notch (Shimizu et al., 2000a, 2000b), Eph-Ephrin (Dai et al., 2014), FGF (Ornitz et al., 1996; Zhang et al., 2006), and other signaling pathways. However, the model of promiscuous interaction systems can also be applied more generally to other input-output systems, such as basic helix-loop-helix (bHLH) transcription factors (de Martin et al., 2021).

Despite the ubiquity of such systems in biology, we have lacked a general, quantitative framework for understanding the signal processing capabilities of these architectures. Specifically, it has remained unclear what types of computations such systems can implement and how this functional repertoire depends on the molecular architecture of the system.

One prototypical example of a signaling pathway with a promiscuous ligand-receptor architecture is the BMP pathway. In mammals, the pathway contains more than ten distinct ligands that dimerize and interact promiscuously with four type I and three type II receptor variants (Massagué, 2000; Miyazono et al., 2010; Shi and Massagué, 2003). Formation of a full signaling complex results in phosphorylation of the SMAD1/5/8 proteins, regardless of which ligands and receptors are involved. However, despite their similar molecular functions in the pathway, different ligands have been observed to have distinct and even opposite effects on developmental processes (Klammert et al., 2015; Piscione et al., 1997). Similarly, ligands can have non-additive effects and can synergize or antagonize each other (Açil et al., 2014; Ying and Zhao, 2001; Ying et al., 2000, 2001).

Recently, a systematic analysis of how the pathway integrates multiple ligands revealed that it can respond combinatorially to ligands, effectively "computing" functions of multiple ligands. A single cell type can respond differently to different

combinations of ligands; conversely, cell types with distinct receptor profiles can compute different functions of the same ligand inputs. These computational effects of promiscuous ligand-receptor interactions could be recapitulated using a minimal mathematical model describing the promiscuous interactions of ligands and receptors and their enzymatic activation of a downstream target (Antebi et al., 2017; Klumpe et al., 2022).

In this work, we perform a more complete analysis of this model of the BMP pathway, aiming to develop a more fundamental understanding of what types of computations could theoretically be achieved with such an architecture. We utilize two complementary approaches to study this question. We first simulate many randomly generated systems to build intuition for the spectrum of possible responses. We then systematically enumerate all possible targets in a more constrained system and optimize for each one. Together, these results reveal the diverse spectrum of computations that can be achieved and provide a better understanding of how they arise.

2.3 Results

2.3.1 A simple one-step model yields a similar computational repertoire as a more complex two-step model

To model the BMP pathway, we consider a general framework with n_L ligands, n_A type I receptors, and n_B type II receptors, denoted as a (n_L, n_A, n_B) model. We have previously described a model of sequential two-step binding in which the ith ligand, jth type I receptor, and kth type II receptor interact to form active signaling complex T_{ijk} in a two-step process (Methods 2.5.1). Ligand L_i first binds with receptor A_j to form dimer D_{ij}, which then interacts with receptor B_k to form trimeric complex T_{ijk}.

To understand the signal processing capabilities of this system for two inputs, we simulated the system for many random parameter sets and found that the resulting functions interpolated among four archetypal functions: ratiometric, additive, imbalance, and balance (Antebi et al., 2017). Briefly, ratiometric responses feature reduction of activity of one ligand by the second, such that the overall response approximates the ratio of the two concentrations; these responses can be generated by competitive inhibition, where the "denominator" competes for receptors needed to generate signaling activity but produces inactive complexes. Additive responses approximate the sum of the two ligand concentrations, as the ligands increase pathway activity either alone or together; they are readily generated when both ligands activate receptors similarly. Imbalance detection responses, where cells respond maximally to imbalances in the levels of the two ligands, can arise if, for instance, competition between two ligands favors complexes with low signaling activity. Conversely, balance detection responses, where cells respond maximally when both ligands are present at a specific ratio, can be generated when ligand binding favors formation of high-activity signaling complexes.

Having identified these archetypal functions in a two-step model, we next considered a simplified one-step model in which T_{ijk} forms from L_i, A_j, and B_k in a single reaction (Figure 2.1; Methods 2.5.2). This model is described by $n_L n_A n_B$ affinity parameters and $n_L n_A n_B$ efficiency parameters for each complex, as well as the initial levels of ligands and receptors. We have shown theoretically that this model can give rise to the same steady-state responses as the previously described two-step model (Methods 2.5.3). We have further demonstrated that the one-step model can recapitulate the four previously identified archetypal functions of ratiometric, additive, imbalance, and balance responses (Figure 3.S1). Since the one-step model is sufficient to capture the computational complexity of the two-step model but reduces the number of parameters to consider, we chose to proceed with this model.

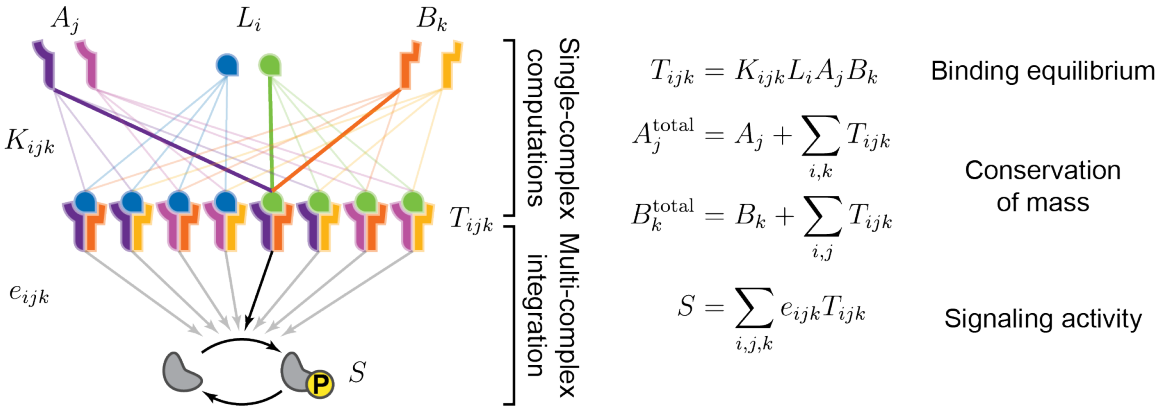

Figure 2.1: The BMP pathway can be represented by a minimal mathematical model describing promiscuous ligand-receptor interactions.

We developed a model that describes the interactions among n_L ligands, n_A type I receptors, and n_B type II receptors in the BMP pathway (Methods 2.5.2). All ligands L_i and receptors A_j and B_k can potentially interact, allowing formation of any of the possible signaling complexes T_{ijk}. For simplicity, we model this interaction as a one-step process with an affinity K_{ijk}. Each signaling complex can then phosphorylate SMAD proteins to generate second messenger with a corresponding activity e_{ijk}. This model provides two layers of computation. In the first, the affinity parameters dictate the levels at which different signaling complexes form. In the second, the activity parameters represent the weights with which each single-complex response is combined to generate the total single-cell response. Equations describe the steady-state levels of each component and the total signal S.

This model effectively represents two layers of computation. The first layer describes the formation of specific combinatorial complexes from the initial concentrations of the individual components. The computations that determine which complexes are formed and at what amount are controlled by the affinity parameters, K_{ijk}. The second layer specifies how the formed complexes combine together to generate the single-cell pathway response. In this level, the computations are governed by the activity parameters, e_{ijk}. Together, the two layers form a directed network, where each layer can perform specific computations on its inputs and generate the input of the next layer.

2.3.2 Analysis of responses at the level of signaling complexes identifies several archetypal "basis functions"

To understand the repertoire of computations possible in the BMP pathway, we sought to gain a deeper understanding of how pathway activity arises as an integration of the levels of multiple signaling complexes. Pathway activity represents a linear combination of signaling complex levels, so we started by analyzing the profiles of signaling complexes to determine whether we could identify fundamental archetypes or basis functions of single-complex responses, similar to the ratiometric, additive, imbalance, and balance functions observed previously for single-cell behaviors.

Certain responses are inherently impossible at the level of signaling complexes; since complex T_{ijk} comprises ligand i, formation of T_{ijk} requires the presence of L_i, and there cannot be a response in the absence of that ligand. Intuitively, complex levels would rise in conjunction with increased concentrations of the associated ligand, but increasing the concentration of a different ligand would decrease complex levels through competition for a limited set of receptors. Thus, the simplest single-complex responses would exhibit a ratiometric-like behavior. However, the presence of a different ligand can potentially modulate the overall response and enable other types of computations.

We first sought to identify what single-complex functions could arise in a model with two variants each of ligand, type I receptor, and type II receptor, as this $(2, 2, 2)$ model represents the minimal model size that demonstrates multiplicity of each component. Therefore, we randomly generated many different parameter sets and simulated the resulting steady-state levels of each of the $n_L n_A n_B$ signaling complexes for a high-resolution (10×10) ligand titration (Methods 2.5.4). Upon visualizing the resulting single-complex responses, we identified many examples that matched the expected

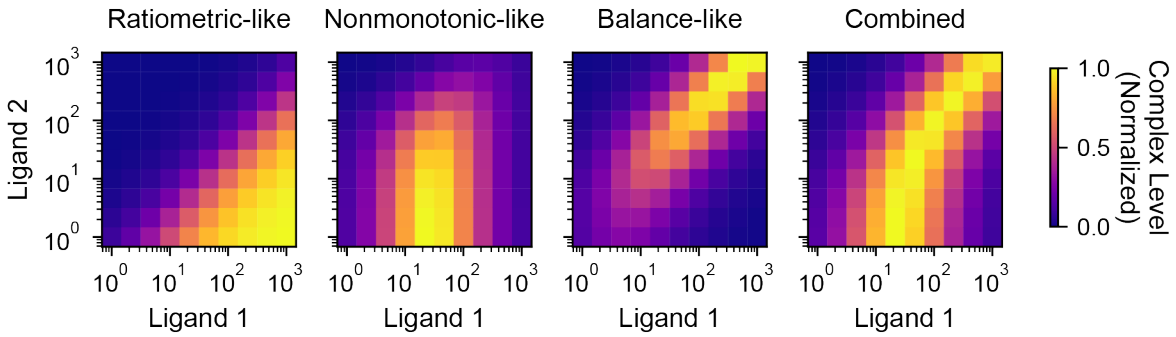

Figure 2.2: Single-complex responses can be classified into archetypal basis functions.

To identify possible single-complex responses, we performed many random simulations of a $(2, 2, 2)$ model (Methods 2.5.4). This analysis revealed several archetypal responses: ratiometric-like, nonmonotonic-like, and balance-like. We also saw responses that combined the features of nonmonotonic-like and balance-like behaviors.

ratiometric-like behavior; however, we also found response profiles that showed more complex responses (Figure 2.2). Some responses exhibited nonmonotonic-like behavior, in which signaling complex levels are highest when the associated ligand is present alone but in moderate concentration. We also observed a response type in which levels of signaling complex are maximal at a specific ratio of the two ligands. We referred to such responses as balance-like, due to their similarity to balance detection responses seen at the level of pathway activity. Finally, we observed responses that combined features of the nonmonotonic-like and balance-like profiles. These results illustrate that complex computations can arise even at the level of signaling complex formation.

2.3.3 Multiplicity of both receptor types is necessary to produce more complex computations at the level of signaling complexes

To understand how the types of computations observed relate to overall model complexity, we repeated this procedure for a range of receptor multiplicities. We started from a $(2, 1, 1)$ model, which has no receptor multiplicity, and increased

the number of receptor variants to a $(2, 4, 3)$ model, which reflects the receptor multiplicity in humans and other mammals. Specifically, we evaluated $(2, 1, 1)$, $(2, 2, 1)$, $(2, 2, 2)$, $(2, 3, 2)$, $(2, 3, 3)$, and $(2, 4, 3)$ models.

To more systematically evaluate the types of responses seen across these hundreds of thousands of examples, we reasoned that we could analyze the qualitative behavior of any given response by examining where local maxima occurred. For simplicity, we symmetrized responses by always labeling the ligand associated with a given complex as ligand 1 and considering the other ligand to be ligand 2. The simple ratiometric-like responses, for instance, would yield a single local maximum, occurring where the concentration of ligand 1 is high and the concentration of ligand 2 low. We identified all such local maxima for each response profile and then computed the frequency distribution for their locations for every choice of model complexity (Figure 2.3; Methods 2.5.5).

These results revealed that multiplicity of both receptors is required to achieve computations beyond the predicted ratiometric-like response. In the $(2, 1, 1)$ and $(2, 2, 1)$ models, local maxima are observed only when ligand 1 is present at high concentrations and ligand 2 at low concentrations, consistent with the ratiometric-like behavior. By contrast, models with multiplicity of both receptors began showing the presence of local maxima at intermediate concentrations of ligand 1, corresponding to nonmonotonic-like responses, or at specific ratios of the two ligands, corresponding to balance-like responses. However, all models with at least two variants of each receptor showed the same general classes of responses. Although models with additional receptor variants showed a slight increase in the proportion of more sophisticated computations (such as the balance-like function), no new classes of integration functions were seen beyond those already observed in the $(2, 2, 2)$ model. Thus, promiscuous ligand-receptor interactions generate a relatively small

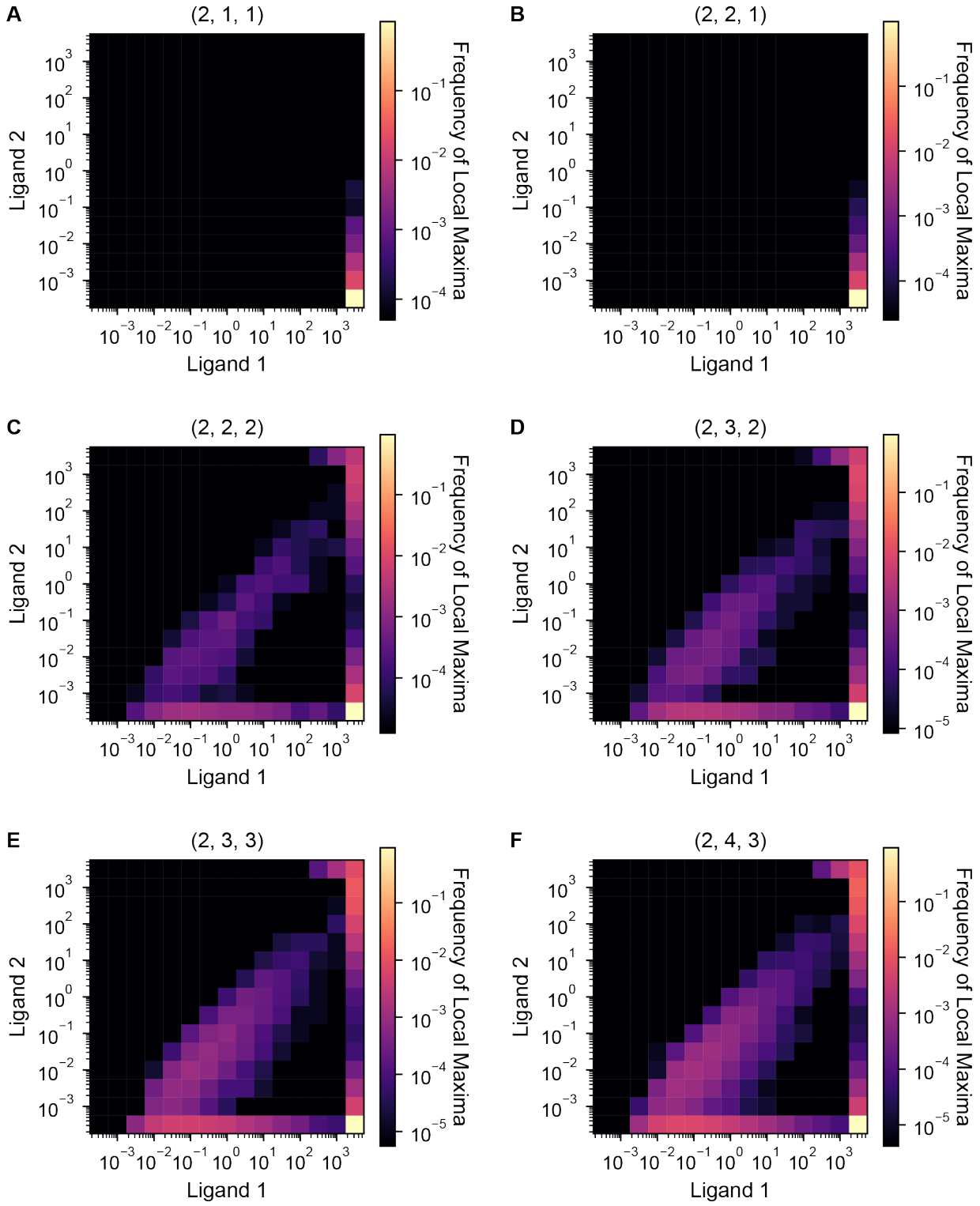

Figure 2.3: Multiplicity of both receptor subunits is required to generate more sophisticated single-complex responses. Continued on next page.

Figure 2.3: Multiplicity of both receptor subunits is required to generate more sophisticated single-complex responses. Continued from previous page.

For different model sizes, we randomly generated 10,000 parameter sets and identified the local maxima of each single-complex response. We then computed the overall frequency distribution for each model size (Methods 2.5.5). Specifically, we considered (A) $(2, 1, 1)$, (B) $(2, 2, 1)$, (C) $(2, 2, 2)$, (D) $(2, 3, 2)$, (E) $(2, 3, 3)$, and (F) $(2, 4, 3)$ models. These choices range from models with no receptor multiplicity to models capturing the mammalian pathway size. Models without multiplicity in both receptor types showed only local maxima consistent with ratiometric-like responses. All models with multiplicity in both receptors yielded additional regions of local maxima consistent with a broader spectrum of possible responses, including the more complex nonmonotonic-like and balance-like functions.

set of response types at the level of signaling complexes, and multiplicity of both receptor variants is necessary and sufficient to achieve those functions beyond the expected ratiometric-like behavior.

2.3.4 Systematic screen for single-complex responses identifies no additional functions

Analyzing many randomly generated parameter sets revealed the existence of a limited number of classes of single-complex computations. However, given the large number of parameters even in our one-step model, the number of parameter sets sampled represents a relatively small subset of parameter space. Therefore, we sought to develop a more systematic approach to evaluating the full spectrum of possible computations in the BMP pathway architecture.

Specifically, we asked what computations could be generated by systematically enumerating all response profiles in a discretized space and seeking to optimize parameters to achieve each target (Figure 2.4; Methods 2.5.6). To comprehensively sample the space of possible responses, we quantized the input ligand concentrations to three levels (low, medium, and high) spanning a 1,000-fold dynamic range, and we considered binarized response values (off and on). We specified "off" responses

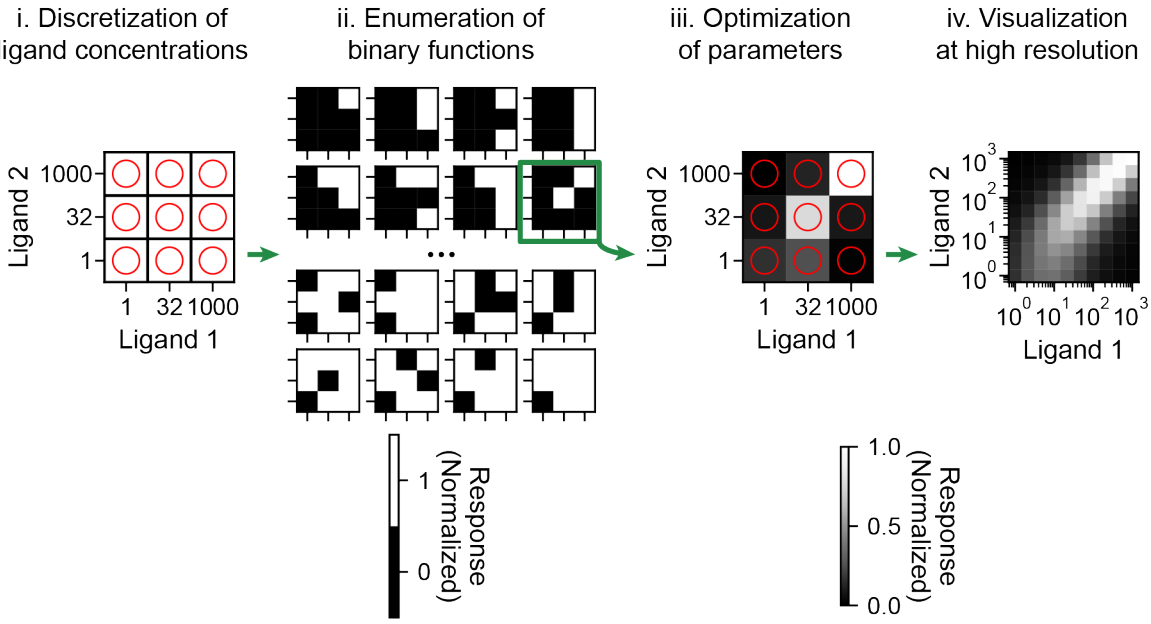

Figure 2.4: Enumerate-optimize approach enables systematic analysis of possible responses.

To comprehensively analyze all possible responses and identify which can be achieved, we used an enumerate-optimize approach (Methods 2.5.6). i. We first discretized ligand space and considered three levels of ligand concentrations. ii. This discretization enabled us to enumerate all possible binary responses in the low-resolution space. iii. We then sought to optimize parameters that yielded responses best matching the targeted function. iv. Finally, we simulated the responses at higher resolution to better visualize the computations.

when all ligand concentrations are "low." This process generated $2^8 = 256$ possible 3×3 binary response functions to analyze. For each of these functions, we used least-squares optimization to optimize both biochemical parameters (affinities and activities) and initial conditions (receptor expression levels) such that the resulting output would best match the target function. We then used these parameters to simulate the response at higher resolution.

We first used this strategy to analyze single-complex computations in a $(2, 2, 2)$ model (Methods 2.5.7). For each targeted response, we visualized a high-resolution simulated response for the best parameter set identified (Figure 2.5). These results

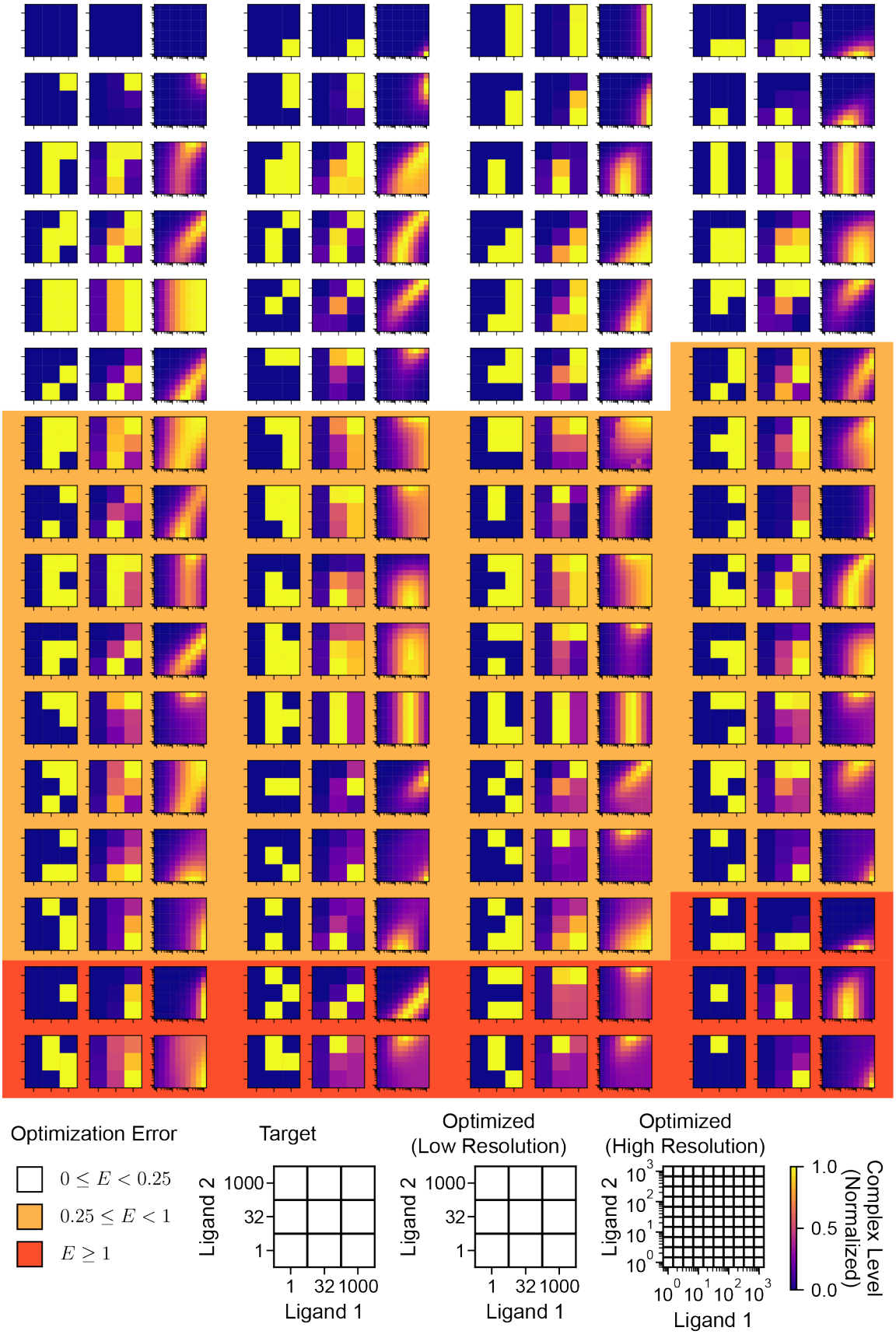
Optimization Error
$0 \leq E < 0.25$
$0.25 \leq E < 1$
$E \geq 1$
Target
Optimized
(Low Resolution)
Optimized
(High Resolution)
Ligand 2
Ligand 1
1000
32
1
Complex Level
(Normalized)
1.0
0.5
0.0

Figure 2.5: Systematic screening of single-complex functions reveals consistent repertoire of archetypes. Continued from previous page.

To evaluate the types of single-complex response functions achievable in the BMP pathway, optimization was performed to identify parameters best achieving each of the possible binary 3×3 functions (Methods 2.5.7). These parameters represent the affinity and activity of each signaling complex as well as the expression level of each receptor. For each function, the performance of the parameter set with the lowest optimization error, defined as the sum of squared errors, is shown. In particular, each result shows the binary 3×3 target function (left), the corresponding low-resolution 3×3 matrix (center) obtained from the optimized parameters, and the high-resolution 10×10 simulated matrix (right). Shading indicates low (white), moderate (yellow), and high (red) optimization error. The resulting response functions can be classified qualitatively into a relatively small number of archetypes.

largely agreed with those from our analysis of random parameter sets. We again identified ratiometric-like, nonmonotonic-like, and balance-like responses, as well as responses combining nonmonotonic-like and balance-like behavior. Overall, the classes of computations identified by both approaches were in good agreement.

These results revealed that two of the previously established response functions, ratiometric and balance detection behavior, can already arise at the level of single complexes. However, additive and imbalance behaviors (or any previously unseen computations) would require the integration of multiple complexes.

2.3.5 Systematic screen for single-cell responses reveals a wide array of possible computations

Therefore, we next extended our analysis to include multi-complex integration. By activating the same intracellular signal with different rates, the pathway can integrate multiple complexes using distinct weights. In this way, multiple single-complex response functions could combine to generate a diverse set of computations. We applied the same enumerate-optimize strategy to identify possible responses at the level of total pathway output (Figure 2.6; Methods 2.5.8). In analyzing the results

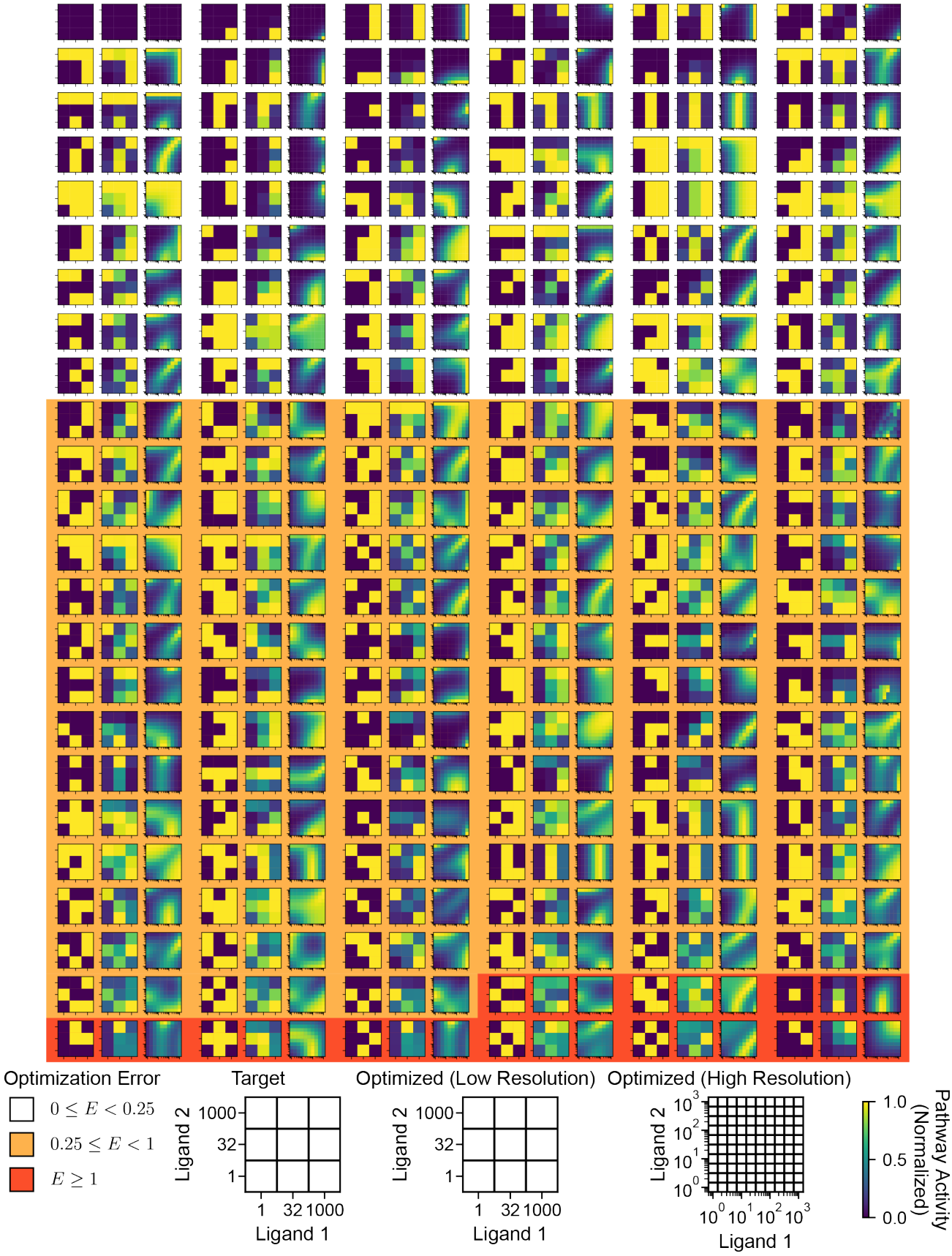
Optimization Error
$0 \leq E < 0.25$
$0.25 \leq E < 1$
$E \geq 1$
Target
Optimized (Low Resolution)
Optimized (High Resolution)
Ligand 2
Ligand 1
1000
32
1
Pathway Activity (Normalized)
1.0
0.5
0.0

Figure 2.6: Promiscuous ligand-receptor interactions generate a larger repertoire of response functions at the level of pathway activity. Continued from previous page.

The enumerate-optimize approach of Figure 2.4 was applied to analyze the types of response functions achievable at the level of pathway activity in a $(2, 2, 2)$ model (Methods 2.5.8). For each possible binary 3×3 function, optimization was performed to identify parameters best achieving that function, and the result from the parameter set with the lowest optimization error is shown. For each target function (left), the corresponding low-resolution 3×3 matrix (center) and high-resolution 10×10 matrix (right) are shown, as simulated from the optimized parameters. Shading indicates low (white), moderate (yellow), and high (red) optimization error. The resulting responses span a diverse repertoire of input-output functions.

of the top parameter sets for each target function, we found that we were able to reproduce many of the targets. These responses included the four response functions identified previously (Antebi et al., 2017) as well as many others. However, it is interesting to note that some target functions were not generated at all. For example, no parameters were found that give rise to a response profile that is only active when both ligands are used together at an intermediate level. These results suggest that many, but not all, two-ligand functions can be generated within the $(2, 2, 2)$ model.

2.3.6 Single-cell responses represent linear combinations of single-complex responses

Having identified a diverse array of responses achievable in the $(2, 2, 2)$ model, we sought to understand them further. In particular, we decomposed select examples as linear combinations of the underlying single-complex computations, ordered by their relative weights (Figure 2.7). We chose the additive and imbalance archetypes that had previously been identified as well as two new response functions that had never previously been observed experimentally or theoretically. One demonstrates a nonmonotonic responses for both ligands, while the other is a "multimodal" response that combines imbalance and balance detection. Intriguingly, each example could

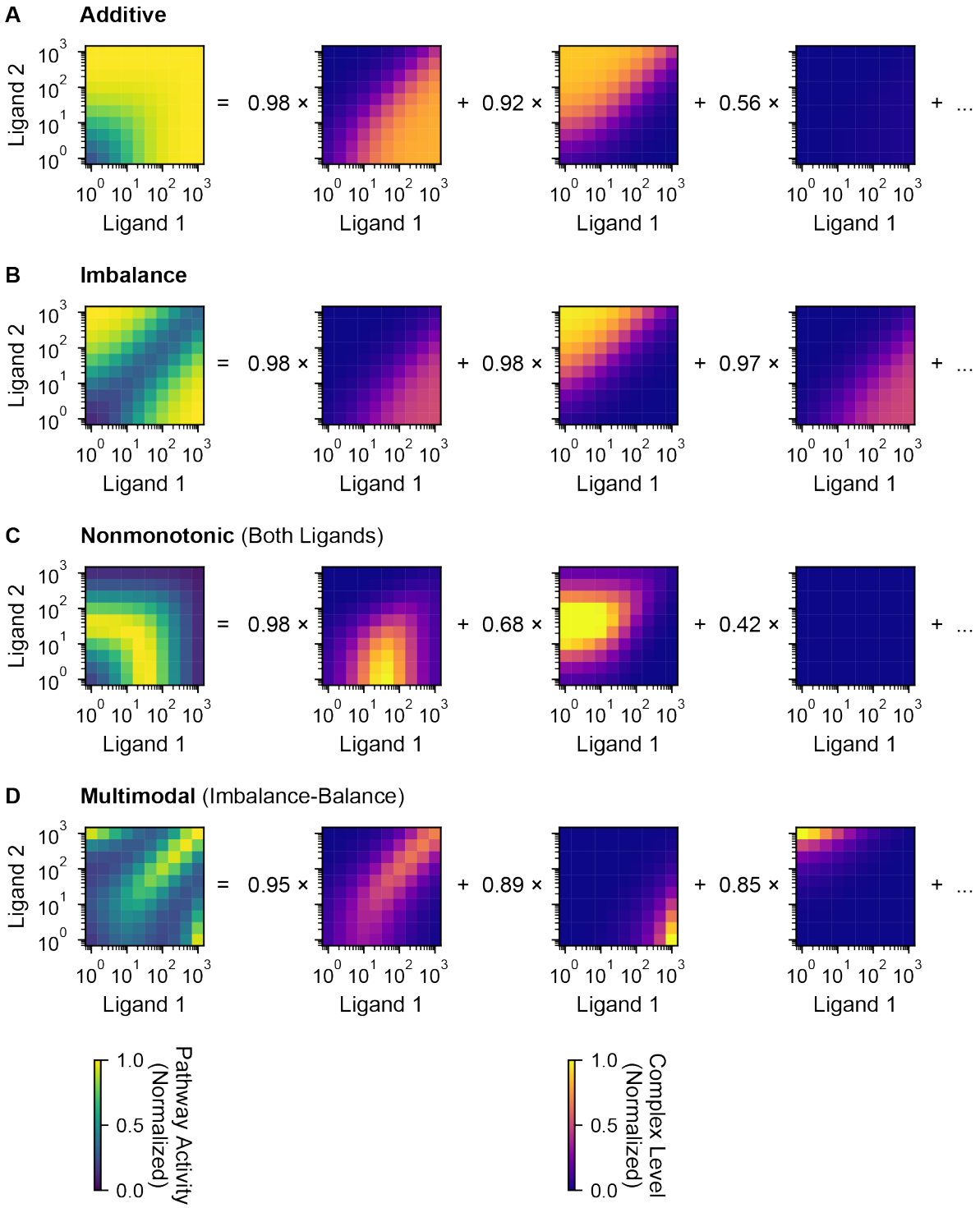

Figure 2.7: Diverse single-cell responses can be understood as linear combinations of simpler single-complex responses. Continued on next page.

Figure 2.7: Diverse single-cell responses can be understood as linear combinations of simpler single-complex responses. Continued from previous page.

Each single-cell response represents a linear combination of the underlying single-complex responses. For selected single-cell responses, the corresponding single-complex profiles with the highest contributions are shown, together with their respective weightings. These responses represent previously known response types such as (A) additive or (B) imbalance behaviors as well as newly identified computations such as (C) nonmonotonic or (D) multimodal functions. In each case, the top three single-complex functions are sufficient to approximate the full single-cell response.

be accurately approximated by as few as three of the underlying signaling complex profiles. This finding indicates that complex computations can arise even if many signaling complexes have relatively low activities.

2.4 Discussion

In this work, we have described a minimal mathematical model for the BMP pathway and performed a systematic analysis to understand the theoretical capabilities of this architecture. In particular, our model involves two key types of parameters and two associated levels of computations. Affinity parameters describe ligand-receptor binding and thereby dictate the levels at which different signaling complexes form. Activity parameters quantify how efficiently signaling complexes produce second messenger and thus dictate how signaling complexes integrate to generate the total level of pathway activation.

To better understand the computational capabilities of the pathway, we have analyzed both single-complex and single-cell responses. We use two different approaches: (1) analysis of many randomly generated parameter sets and (2) optimization of systematically enumerated target functions. These complementary approaches revealed that single-complex responses can be divided into relatively few archetypes. Ratiometric-like responses arise at all levels of model complexity,

whereas nonmonotonic-like and balance-like responses (as well as a response combining features of both) require multiplicity of both receptor subunits. However, integration of multiple complexes produces a larger diversity of response types. We reproduce the previously observed ratiometric, additive, imbalance, and balance responses as well as many other responses, such as nonmonotonic responses or a multimodal response that combines imbalance and balance detection.

Together, these results provide a fundamental understanding of the limits of the BMP pathway architecture. They also reveal the existence of additional responses beyond those previously observed experimentally (Antebi et al., 2017; Klumpe et al., 2022). It remains unclear how difficult they may be to achieve in practice, as they may require a narrow range of biochemical parameters or be sensitive to natural biological variation in ligand and receptor levels.

This work lays a foundation for understanding higher-level principles of the BMP pathway. Our results have helped elucidate how a single cell may respond to ligand combinations in its environment. The approach could be generalized to understand how multiple cells, potentially expressing different sets of receptors, respond to their environment. This question is fundamental for analyzing how signaling may occur in natural biological contexts, where signals may diffuse locally to affect multiple cell types.

This framework for analyzing ligand-receptor interactions can also be generalized to the range of signaling pathways that likewise demonstrate a promiscuous pathway architecture. It will be interesting to assess whether other pathways yield a similar range of computations and evaluate the unique functions that their different architectures may enable. Ultimately, enhancing our ability to predict and control biological signaling will enable better understanding of developmental and disease processes and guide development of therapeutics to rationally modulate signaling.

2.5 Methods

All data and code are publicly accessible at the CaltechDATA research data repository (`https://doi.org/10.22002/D1.20181`). Code is also publicly accessible at GitHub (`https://github.com/christinasu/PromiSys`).

2.5.1 Two-Step Model for Promiscuous Interactions

Many signaling pathways demonstrate promiscuous interactions between multiple ligand and receptor variants, which can bind with varying affinities to form many distinct signaling complexes. The BMP pathway represents a canonical example of such an architecture. Previously, we have described a mathematical model that captures key features of this pathway and recapitulates experimentally observed responses (Antebi et al., 2017). This model considers formation of the heterotrimeric complexes T_{ijk} in a two-step process. Briefly, ligand L_i and receptor A_j form an intermediate dimer D_{ij}, which then binds to receptor B_k to form trimer T_{ijk}. We assume that the reactions are reversible and follow first-order kinetics, with forward and reverse reaction rates $k^D_{f_{ij}}$ and $k^D_{r_{ij}}$ for formation of the dimers and $k^T_{f_{ijk}}$ and $k^T_{r_{ijk}}$ for formation of the trimers. Defining $K^D_{ij} \equiv k^D_{f_{ij}}/k^D_{r_{ij}}$ and $K^T_{ijk} \equiv k^T_{f_{ijk}}/k^T_{r_{ijk}}$, the steady-state solutions for T_{ijk} in the two-step model are as follows:

$$T_{ijk} = K^T_{ijk} K^D_{ij} L_i \left(\frac{A^0_j - \sum_{i'=1}^{n_L} \sum_{k'=1}^{n_B} T_{i'jk'}}{1 + \sum_{i'=1}^{n_L} K^D_{i'j} L_{i'}} \right) \left(B^0_k - \sum_{i'=1}^{n_L} \sum_{j'=1}^{n_A} T_{i'j'k} \right) \tag{2.1}$$

Each complex T_{ijk} phosphorylates the second messenger at some rate ε_{ijk} to generate intracellular signal S, which degrades at rate γ. The rate of change of the total signal is given by the following differential equation:

$$\frac{dS}{dt} = \sum_{i=1}^{n_L} \sum_{j=1}^{n_A} \sum_{k=1}^{n_B} \varepsilon_{ijk} T_{ijk} - \gamma S \tag{2.2}$$

Defining activities $e_{ijk} \equiv \varepsilon_{ijk}/\gamma$, Equation 2.2 can be solved at steady state as below:

$$S = \sum_{i=1}^{n_L} \sum_{j=1}^{n_A} \sum_{k=1}^{n_B} e_{ijk} T_{ijk} \tag{2.3}$$

2.5.2 One-Step Model for Promiscuous Interactions

Here, we consider a simplified version of the model that captures equivalent behaviors at steady state while reducing the number of parameters to be considered. Briefly, we describe a system in which active signaling complexes, composed of a ligand and two receptor subunits, form in a single reaction, which we refer to as a one-step model. A system with n_L ligands, n_A type A receptors, and n_B type B receptors is denoted as a (n_L, n_A, n_B) model. Since binding and unbinding of ligands and receptors occur on fast time scales relative to the time scales of reporter detection, we analyze the behavior of this system at steady state (Methods 3.6.1). In particular, this model is described by the following set of steady-state equations:

$$L_i^0 = L_i \tag{2.4}$$

$$A_j^0 = A_j + \sum_{i=1}^{n_L} \sum_{k=1}^{n_B} T_{ijk} \tag{2.5}$$

$$B_k^0 = B_k + \sum_{i=1}^{n_L} \sum_{j=1}^{n_A} T_{ijk} \tag{2.6}$$

$$T_{ijk} = K_{ijk} L_i A_j B_k \tag{2.7}$$

$$S = \sum_{i=1}^{n_L} \sum_{j=1}^{n_A} \sum_{k=1}^{n_B} e_{ijk} T_{ijk} \tag{2.8}$$

We can solve this system of equations to find an expression for the values of T_{ijk} at steady state, which we can then use to compute the total signaling response S. We obtain a system of $n_T = n_L n_A n_B$ quadratic equations for T_{ijk}:

$$T_{ijk} = K_{ijk} L_i \left(A_j^0 - \sum_{i'=1}^{n_L} \sum_{k'=1}^{n_B} T_{i'jk'} \right) \left(B_k^0 - \sum_{i'=1}^{n_L} \sum_{j'=1}^{n_A} T_{i'j'k} \right) \tag{2.9}$$

As in the two-step model, the solutions for T_{ijk} from this system of equations can then be substituted into Equation 2.3 to compute the total signal S.

To solve the model efficiently, we used Equilibrium Toolkit (EQTK), an optimized Python-based numerical solver for biochemical reaction systems (Bois, 2020). EQTK casts the coupled equilibrium problem as an unconstrained convex dual optimization problem and employs a globally convergent trust region algorithm to solve it (Bois, 2020; Dirks et al., 2007). This method accelerated computation by approximately 600-fold compared to standard nonlinear least-squares optimization used previously (Antebi et al., 2017).

2.5.3 Comparison of One-Step and Two-Step Models

Comparing Equation 2.1 and Equation 2.9, the steady-state solutions for T_{ijk} in the two-step model can be mapped to the one-step model under the following parameter choice:

$$K_{ijk} = \frac{K^T_{ijk} K^D_{ij}}{1 + \sum_{i'=1}^{n_L} K^D_{i'j} L_{i'}} \tag{2.10}$$

Since S is defined by the values of T_{ijk} and is given by Equation 2.3 in both the one-step and two-step models, the steady-state behavior of the two-step model with any set of parameters can also be represented in the one-step model. However, the two-step model requires $N_p^{\text{two-step}} = n_A + n_B + n_L n_A + 2 n_L n_A n_B$ parameters, while the one-step model involves $N_p^{\text{one-step}} = n_A + n_B + 2 n_L n_A n_B$ parameters. Thus, the one-step model enables us to simplify the system while preserving all possible behaviors of T_{ijk} and S at steady state.

2.5.4 Simulation of Single-Complex Responses

To understand what types of single-complex responses are possible with the promiscuous pathway architecture, we started by generating many random parameter sets

in a $(2, 2, 2)$ model. In particular, we considered 10,000 randomly generated parameter sets and the resulting steady-state profiles of each of the $n_L n_A n_B = 2 \cdot 2 \cdot 2 = 8$ signaling complexes, or 80,000 responses in total. We then visualized these profiles to identify qualitatively different response types.

2.5.5 Analysis of Single-Complex Responses

We next sought to generalize this analysis of single-complex responses to multiple different model sizes and therefore developed an approach to capture key features across large numbers of responses. For each model size, we again generated 10,000 random parameter sets and simulated the steady-state levels of each signaling complex for an extensive ligand titration, ranging from $10^{-3.5}$ to $10^{3.5}$. This procedure yielded varying numbers of responses, from 20,000 for the $(2, 1, 1)$ model (with 2 possible signaling complexes) to 240,000 for the $(2, 4, 3)$ model (with 24 possible signaling complexes). In each profile, we identified all local maxima, defined as any element with value greater than all adjacent elements. To allow comparison of the results from different profiles, we relabeled ligands such that ligand 1 always denotes the ligand that is present in the signaling complex of interest. (In particular, profiles for complexes T_{2jk} were transposed.) For each ligand concentration, we counted the total number of occurrences M of a local maximum and compared it to the total number of profiles P. We then visualized $\log\left[\frac{1+M}{P}\right]$ to analyze the distribution of local maxima in ligand concentration space.

2.5.6 Enumeration of Target Responses

To systematically analyze the computational repertoire of the BMP pathway, we used a discretized ligand concentration space to enable a comprehensive screening. We considered a 1,000-fold dynamic range and discretized ligand concentrations to three levels ($10^0 = 1$, $10^{1.5} \approx 32$, and $10^3 = 1{,}000$), yielding $3^2 = 9$ possible ligand

words. By considering the resulting responses to be binary (off or on), we could enumerate $2^9 = 512$ total targets to represent all possible binary functions in this ligand space. However, the combination with all ligands at the lowest concentration is assumed to yield negligible activation, leaving a total of $2^8 = 256$ targets to consider.

2.5.7 Optimization of Single-Complex Responses

We then sought to optimize all possible single-complex responses. For a given T_{ijk}, there cannot be a response to any ligand combination in which the concentration of ligand i is low, as ligand i is required for the formation of T_{ijk}. Therefore, we remove any target functions which include an active response when ligand i is present at low concentration. We also filter out targets that are redundant under changes in ligand labels. This procedure yielded a total of 64 possible targets.

We then performed constrained least-squares optimization to search for parameters achieving each target. Specifically, we optimized for $n_L n_A n_B$ affinity parameters K_{ijk}, $n_L n_A n_B$ activity parameters e_{ijk}, and $n_A + n_B$ receptor expression levels, with all parameters required to be nonnegative. We sought to minimize the residuals between the target responses and the simulated responses at all ligand combinations. (To allow comparison with the binary target responses, we normalized the simulated responses by the maximum value.) As this optimization procedure is not guaranteed to converge to a global minimum, we optimized repeatedly with different initial conditions. At each trial, we chose which target to optimize based on both the lowest error E achieved and the number of trials T attempted. Specifically, we optimized for the target with maximal value of E/T^2 and repeated this process for at least an average of 50 optimizations per target to ensure adequate testing of all targets.

2.5.8 Optimization of Single-Cell Responses

We applied the same optimization approach to study all possible single-cell responses. In this context, the ligand combination with both ligands present at low concentration should not generate any response, but all other combinations could yield pathway activation. Therefore, we only excluded targets invariant under changes in ligand labels, leaving a total of 144 possibilities. We then systematically optimized parameters for all targets using the procedure described above (Methods 2.5.7).

2.6 References

Açil, Y., Ghoniem, A.-A., Wiltfang, J., and Gierloff, M. (2014). Optimizing the osteogenic differentiation of human mesenchymal stromal cells by the synergistic action of growth factors. *J. Cranio-Maxillofac. Surg.* 42, 2002–2009.

Antebi, Y.E., Linton, J.M., Klumpe, H., Bintu, B., Gong, M., Su, C., McCardell, R., and Elowitz, M.B. (2017). Combinatorial Signal Perception in the BMP Pathway. *Cell* 170, 1184–1196.

Bois, J.S. (2020). Equilibrium Toolkit. *CaltechDATA*. DOI 10.22002/D1.1430.

Dai, D., Huang, Q., Nussinov, R., and Ma, B. (2014). Promiscuous and specific recognition among ephrins and Eph receptors. *Biochim. Biophys. Acta — Proteins Proteom.* 1844, 1729–1740.

de Martin, X., Sodaei, R., and Santpere, G. (2021). Mechanisms of Binding Specificity among bHLH Transcription Factors. *Int. J. Mol. Sci.* 22, 9150.

Dirks, R.M., Bois, J.S., Schaeffer, J.M., Winfree, E., and Pierce, N.A. (2007). Thermodynamic Analysis of Interacting Nucleic Acid Strands. *SIAM Rev.* 49, 65–88.

Heldin, C.-H., Miyazono, K., and ten Dijke, P. (1997). TGF-β signalling from cell membrane to nucleus through SMAD proteins. *Nature* 390, 465–471.

Klammert, U., Mueller, T.D., Hellmann, T.V., Wuerzler, K.K., Kotzsch, A., Schliermann, A., Schmitz, W., Kuebler, A.C., Sebald, W., and Nickel, J. (2015). GDF-5 can act as a context-dependent BMP-2 antagonist. *BMC Biol.* 13, 77.

Klumpe, H.E., Langley, M.A., Linton, J.M., Su, C.J., Antebi, Y.E., and Elowitz, M.B. (2022). The context-dependent, combinatorial logic of BMP signaling. *Cell Syst.* 13, 388–407.

Llimargas, M., and Lawrence, P.A. (2001). Seven Wnt homologues in *Drosophila*: A case study of the developing tracheae. *Proc. Natl. Acad. Sci. U.S.A.* 98, 14487–14492.

Massagué, J. (1998). TGF-β signal transduction. *Annu. Rev. Biochem.* 67, 753–791.

Massagué, J. (2000). How cells read TGF-β signals. *Nat. Rev. Mol. Cell Biol.* 1, 169–178.

Miyazono, K., Kamiya, Y., and Morikawa, M. (2010). Bone morphogenetic protein receptors and signal transduction. *J. Biochem.* 147, 35–51.

Mueller, T.D., and Nickel, J. (2012). Promiscuity and specificity in BMP receptor activation. *FEBS Lett.* 586, 1846–1859.

Nickel, J., and Mueller, T.D. (2019). Specification of BMP Signaling. *Cells* 8, 1579.

Ornitz, D.M., Xu, J., Colvin, J.S., McEwen, D.G., MacArthur, C.A., Coulier, F., Gao, G., and Goldfarb, M. (1996). Receptor Specificity of the Fibroblast Growth Factor Family. *J. Biol. Chem.* 271, 15292–15297.

Piscione, T.D., Yager, T.D., Gupta, I.R., Grinfeld, B., Pei, Y., Attisano, L., Wrana, J.L., and Rosenblum, N.D. (1997). BMP-2 and OP-1 exert direct and opposite effects on renal branching morphogenesis. *Am. J. Physiol.* 273, F961–F975.

Schmierer, B., and Hill, C.S. (2007). TGFβ-SMAD signal transduction: molecular specificity and functional flexibility. *Nat. Rev. Mol. Cell Biol.* 8, 970–982.

Shi, Y., and Massagué, J. (2003). Mechanisms of TGF-β Signaling from Cell Membrane to the Nucleus. *Cell* 113, 685–700.

Shimizu, K., Chiba, S., Hosoya, N., Kumano, K., Saito, T., Kurokawa, M., Kanda, Y., Hamada, Y., and Hirai, H. (2000a). Binding of Delta1, Jagged1, and Jagged2 to Notch2 Rapidly Induces Cleavage, Nuclear Translocation, and Hyperphosphorylation of Notch2. *Mol. Cell. Biol.* 20, 6913–6922.

Shimizu, K., Chiba, S., Saito, T., Kumano, K., and Hirai, H. (2000b). Physical Interaction of Delta1, Jagged1, and Jagged2 with Notch1 and Notch3 Receptors. *Biochem. Biophys. Res. Commun.* 276, 385–389.

Wodarz, A., and Nusse, R. (1998). Mechanisms of Wnt signaling in development. *Annu. Rev. Cell Dev. Biol.* 14, 59–88.

Ying, Y., and Zhao, G.-Q. (2001). Cooperation of Endoderm-Derived BMP2 and Extraembryonic Ectoderm-Derived BMP4 in Primordial Germ Cell Generation in the Mouse. *Dev. Biol.* 232, 484–492.

Ying, Y., Liu, X.-M., Marble, A., Lawson, K.A., and Zhao, G.-Q. (2000). Requirement of Bmp8b for the Generation of Primordial Germ Cells in the Mouse. *Mol. Endocrinol.* 14, 1053–1063.

Ying, Y., Qi, X., and Zhao, G.-Q. (2001). Induction of primordial germ cells from murine epiblasts by synergistic action of BMP4 and BMP8B signaling pathways. *Proc. Natl. Acad. Sci. U.S.A.* 98, 7858–7862.

Zhang, X., Ibrahimi, O.A., Olsen, S.K., Umemori, H., Mohammadi, M., and Ornitz, D.M. (2006). Receptor Specificity of the Fibroblast Growth Factor Family: The Complete Mammalian FGF Family. *J. Biol. Chem.* 281, 15694–15700.

Chapter 3

Ligand-Receptor Promiscuity Enables Cellular Addressing

Su, C.J., Murugan, A., Linton, J.M., Yeluri, A., Bois, J., Klumpe, H., Langley, M.A., Antebi, Y.E., and Elowitz, M.B. (2022). Ligand-receptor promiscuity enables cellular addressing. *Cell Syst.* 13, 408–425.

3.1 Abstract

In multicellular organisms, secreted ligands selectively activate, or "address," specific target cell populations to control cell fate decision-making and other processes. Key cell-cell communication pathways use multiple promiscuously interacting ligands and receptors, provoking the question of how addressing specificity can emerge from molecular promiscuity. To investigate this issue, we developed a general mathematical modeling framework based on the bone morphogenetic protein (BMP) pathway architecture. We find that promiscuously interacting ligand-receptor systems allow a small number of ligands, acting in combinations, to address a larger number of individual cell types, defined by their receptor expression profiles. Promiscuous systems outperform seemingly more specific one-to-one signaling architectures in addressing capability. Combinatorial addressing extends to groups of cell types, is robust to receptor expression noise, grows more powerful with increasing numbers of receptor variants, and is maximized by specific biochemical parameter relationships. Together, these results identify design principles governing cellular addressing by ligand combinations.

3.2 Introduction

During development, a handful of core communication pathways control a huge range of cell fate decisions and other processes across diverse tissues and contexts. These pathways include the BMP and the broader transforming growth factor β (TGF-β) pathways, as well as Wnt, fibroblast growth factor (FGF), Hedgehog, and Notch. Each of these pathways comprises multiple ligand and receptor variants that are expressed in different combinations in different cell types. Expression of these pathway components is generally widespread, with receptors for most pathways expressed in most cell types and ligands for most pathways present in most tissues. Puzzlingly, despite the ubiquitous expression of their signaling components, activation of pathways tends to be tightly restricted, occurring only in specific cell types within particular spatiotemporal contexts. If we understood the principles that naturally restrict signaling to specific cell types, we could potentially apply them to control pathways with greater cell type specificity in therapeutic applications.

Multiple mechanisms have been shown to restrict pathway activation. First, modulation of extracellular ligand concentrations through formation of morphogenetic gradients, secreted inhibitors, and factors in the extracellular matrix allows spatial and temporal control of signaling (Bier and De Robertis, 2015; Rogers and Schier, 2011). Second, intracellularly, cells can regulate the amplitude and dynamics of their individual pathway responses by controlling phosphorylation of effector proteins, subcellular localization of signaling components, or selective silencing of particular target gene sets (Axelrod et al., 1998; Lim-Tio and Fuller, 1998; Shaul and Seger, 2007). Third, different ligand variants could bind to and interact with different receptor variants with different strengths. This mechanism would allow the organism to use different ligands to preferentially activate different cell types, based on the receptor variants they express. These three mechanisms could operate individually or in combination.

Here, we focus on this third class of mechanism. It is well known that different ligands can preferentially bind to and activate different receptors, and that different ligands can activate different downstream target genes in the same cell type (Nandagopal et al., 2018; Wootten et al., 2018). However, the features that determine the number of distinct cell types or cell type combinations that can be selectively activated using a given set of ligands are not understood. This level of specificity describes the encoding of information about which cell types, among the constellation of cell types in a complex tissue or entire body, will activate in response to the ligand-encoded "message." We therefore introduce the term "addressing" to denote the ability of ligands to selectively activate, or "address," a pathway in different cell types or cell type combinations.

The simplest conceivable implementation of addressing uses specific, one-to-one ligand-receptor interactions, where each ligand variant interacts exclusively with a single cognate receptor variant (Figure 3.1A, left). This architecture is conceptually straightforward, has been implemented synthetically in the synNotch system (Morsut et al., 2016), and is extendable, as new orthogonal ligand-receptor pairs can provide additional communication channels without disrupting existing ones. However, most natural signaling pathways do not exhibit one-to-one ligand-receptor interactions. Instead, they employ a many-to-many, or promiscuous, architecture, in which each ligand variant interacts with multiple receptor variants and vice versa (Figure 3.1A, right). Pathways such as BMP (Heldin et al., 1997; Massagué, 1998; Mueller and Nickel, 2012; Nickel and Mueller, 2019; Schmierer and Hill, 2007), Wnt (Llimargas and Lawrence, 2001; Wodarz and Nusse, 1998), Notch (Shimizu et al., 2000a, 2000b), Eph-Ephrin (Dai et al., 2014), and FGF (Ornitz et al., 1996; Zhang et al., 2006) all exhibit promiscuous interactions among their multiple ligand and receptor variants. It has generally remained unclear whether molecular promiscuity in ligand-receptor interactions is compatible with addressing at all and, if so, whether

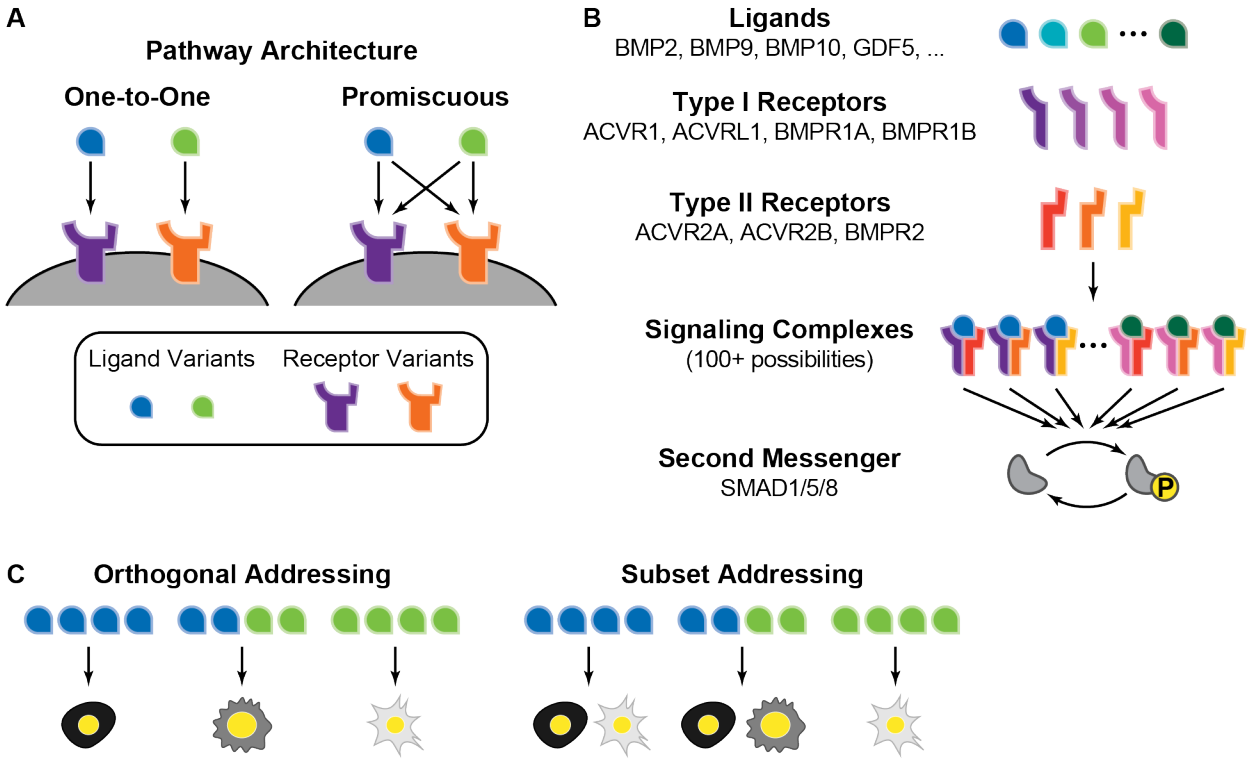

Figure 3.1: Promiscuous ligand-receptor interactions in the BMP pathway may allow combinatorial addressing.

(A) In a one-to-one ligand-receptor architecture (left), each ligand interacts exclusively with a single receptor, while in a promiscuous architecture (right), ligands interact with multiple receptor variants.

(B) In this simplified schematic of the BMP pathway, ligands interact combinatorially with type I and type II receptors at the cell membrane to form signaling complexes, which then activate SMAD1/5/8 effector proteins.

(C) Signaling pathways could enable different forms of addressing. In orthogonal addressing (left), different combinations of ligands each activate a distinct cell type. More generally, subset addressing (right) could allow activation of different groups of cell types by different ligand combinations.

it might counterintuitively provide potential advantages compared to simpler one-to-one architectures. More generally, a concise set of principles governing the design of multi-ligand, multi-receptor interaction systems has not been identified.

The BMP pathway provides an ideal system to study these questions. BMP plays diverse roles in most tissues and has demonstrated therapeutic potential (David and

Massagué, 2018; Massagué, 2000; Miyazono et al., 2010; Wagner et al., 2010; Wang et al., 2014). In well studied systems, individual cells co-express multiple receptor variants and are simultaneously exposed to multiple ligand variants, suggesting that the pathway could function combinatorially (Diez-Roux et al., 2011; Dudley and Robertson, 1997; Godin et al., 1999; Graham et al., 2014; Kapushesky et al., 2010; Li and Ge, 2011; Liem et al., 1995; Simic and Vukicevic, 2005; Zhang et al., 1998). These ligand and receptor variants have been shown to interact promiscuously. In mammals, the pathway comprises more than ten distinct homodimeric and heterodimeric ligand variants as well as four type I and three type II receptor variants (Massagué, 2000; Miyazono et al., 2010; Shi and Massagué, 2003). Signaling complexes, comprising a covalent ligand dimer and two type I and two type II receptor subunits, phosphorylate SMAD1/5/8 effectors, which translocate to the nucleus and act as transcription factors to control the expression of target genes (Figure 3.1B). Overall, this pathway architecture uses combinations of receptors to integrate information from combinations of ligands.

Previous observations suggest that the BMP system can generate complex ligand- and cell type-dependent pathway activation patterns (Baur et al., 2000; Chen et al., 2013; Grassinger et al., 2007; Lind et al., 1996; Varley and Maxwell, 1996; Yu et al., 2008). For example, during neural tube development, different BMP ligands, expressed in overlapping combinations, direct distinct dorsal interneuron identities in neural progenitors, with each ligand showing specific effects on a subset of interneuron identities but not others (Andrews et al., 2017). This behavior could result from addressing of different progenitor states by distinct ligand combinations and/or by ligand-specific activation of different target programs.

Recently, mathematical modeling, together with in vitro experiments, showed that competitive formation of distinct BMP signaling complexes with different ligands

and receptors effectively generates a set of "computations," in which pathway activity depends on the identities and relative concentrations of multiple ligands (Antebi et al., 2017; Klumpe et al., 2022; Martinez-Hackert et al., 2021). These response functions include ratiometric and additive responses as well as imbalance and balance detection responses that are minimal or maximal, respectively, at defined ligand ratios (Figure 3.S1). Further, the pathway can perform different computations on the same ligands depending on the combinations of receptors expressed by individual cells. These results suggest that promiscuous ligand-receptor interactions might allow addressing to function combinatorially, with different ligand combinations addressing particular cell types based on their receptor expression profiles.

Here, we aim to understand how molecular promiscuity in ligand-receptor interactions could potentially enable addressing of specific cell types based on their receptor expression profiles. To this end, we developed a minimal mathematical model of promiscuous ligand-receptor interactions. While additional biochemical mechanisms could further augment addressing in natural biological systems, focusing on ligand-receptor interactions allowed us to explore and understand the specific capabilities that are introduced by this aspect of the pathway. Using this model, we found that promiscuous ligand-receptor interactions alone are sufficient to generate an extensive repertoire of orthogonal communication channels (Figure 3.1C, left), with higher specificity than that of the same number of ligands in the simpler one-to-one architecture. Modest increases in the number of receptor variants increase the number and orthogonality of these addressing channels. Furthermore, the promiscuous architecture allows ligand combinations to address not only individual cell types but also more complex groups of cell types (Figure 3.1C, right). Experimentally, similar types of addressing can be observed in cell lines with differing receptor expression profiles. Finally, using an information theoretic framework, we show how biochemical features, such as anticorrelations between

affinity and activity parameters, maximize the information content that can be transmitted through promiscuous ligand-receptor interactions. These results highlight a potentially general biological design principle—promiscuous ligand-receptor interactions enable ligand combinations to address cell types based on the receptor combinations they express—that should be useful for understanding and designing natural and synthetic communication systems.

3.3 Results

3.3.1 A minimal model allows analysis of promiscuous BMP ligand-receptor interactions

To explore the addressing capacity of promiscuous ligand-receptor systems, we developed a minimal mathematical model based on the architecture of the BMP pathway (Methods 3.6.1). Briefly, the model describes a set of n_L ligands, n_A type I receptors, and n_B type II receptors. A ligand L_i binds simultaneously to type I and type II receptor subunits A_j and B_k to form an active signaling complex T_{ijk} (Figure 3.2A, left). A set of effective interaction strengths, denoted K_{ijk}, represents the strength of binding between a ligand, a type I receptor subunit, and a type II receptor subunit. We further assume that each signaling complex has its own specific activity, denoted e_{ijk}, controlling the efficiency or rate at which it phosphorylates downstream SMAD effector proteins. The overall activity of the pathway is then the sum of the concentrations of the signaling complexes, each weighted by its own activity parameter. We assume steady state in ligand-receptor binding and unbinding, which occur at fast time scales relative to the response to signaling. Under this assumption, the model can be described by one set of equations representing binding and unbinding interactions, a second set of equations representing conservation of total receptor levels, and an expression for total pathway activity, S (Figure 3.2A, right).

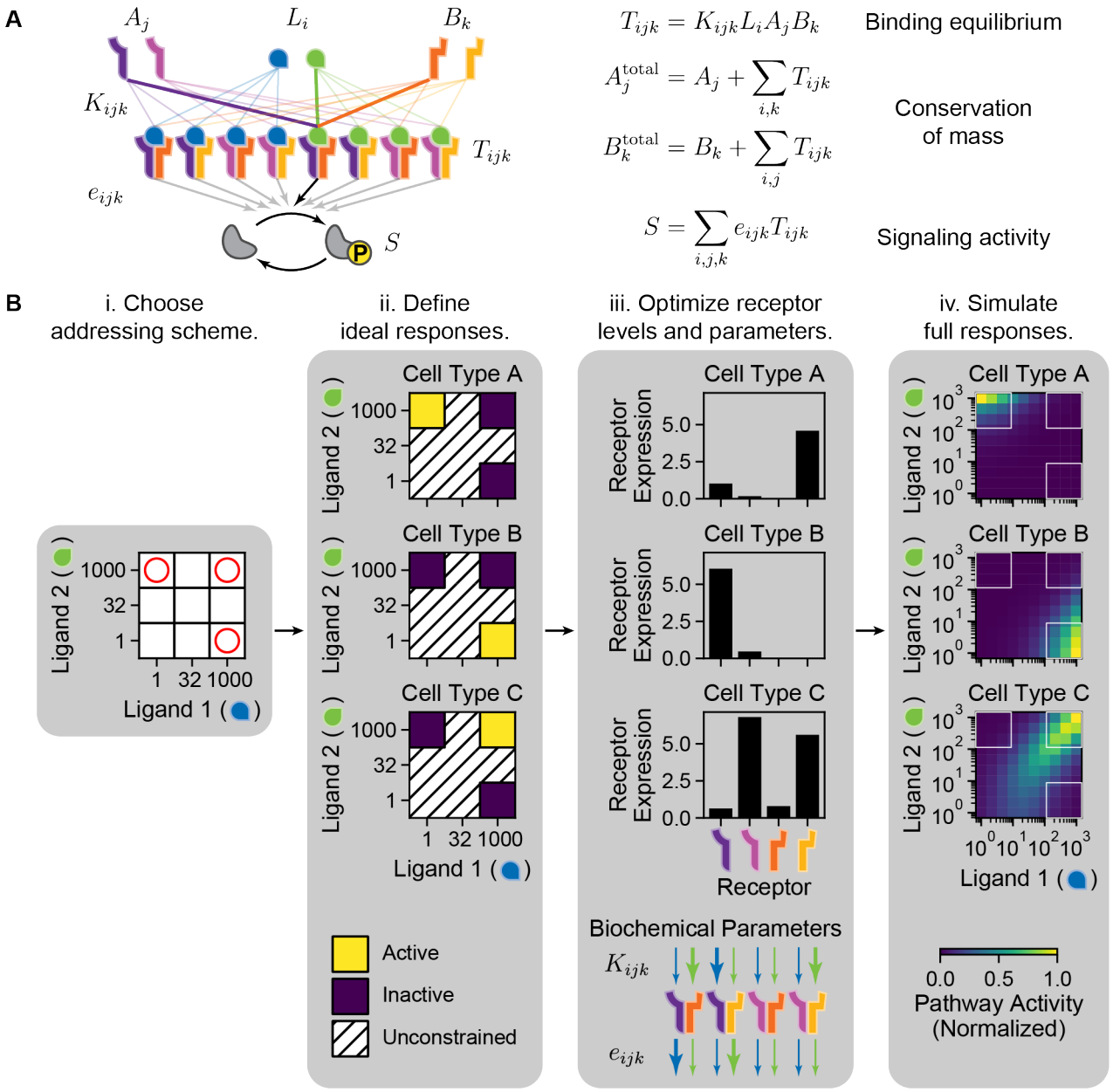
A
A_j
L_i
B_k
K_{ijk}
T_{ijk}
e_{ijk}
S
$T_{ijk} = K_{ijk} L_i A_j B_k$
Binding equilibrium
$A_j^{\text{total}} = A_j + \sum_{i,k} T_{ijk}$
$B_k^{\text{total}} = B_k + \sum_{i,j} T_{ijk}$
Conservation of mass
$S = \sum_{i,j,k} e_{ijk} T_{ijk}$
Signaling activity
B
i. Choose addressing scheme.
ii. Define ideal responses.
iii. Optimize receptor levels and parameters.
iv. Simulate full responses.
Ligand 2 ( )
Ligand 1 ( )
1000
32
1
Cell Type A
Cell Type B
Cell Type C
Active
Inactive
Unconstrained
Receptor Expression
5.0
2.5
0.0
Receptor
Biochemical Parameters
K_{ijk}
e_{ijk}
10^3
10^2
10^1
10^0
0.0 0.5 1.0
Pathway Activity (Normalized)
C
Crosstalk Matrix
Cell Type
A B C
Ligand Word
1 2 3
On-Target
Off-Target
0.0 0.5 1.0
Pathway Activity (Normalized)
D
Performance of different 2-ligand, 3-channel addressing schemes
Pathway Activity (Normalized)
1.00
0.75
0.50
0.25
0.00
On-Target
Off-Target
Distinguishability
Minimum On−Target / Maximum Off−Target
10^1
10^0
Promiscuous
One-to-One
Not addressable
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Addressing Scheme

Figure 3.2: A mathematical model of promiscuous ligand-receptor interactions allows systematic optimization of addressing capabilities. Continued from previous page.

(A) A minimal model of the BMP signaling pathway includes ligand variants (L_i, blue and green), which interact with type I receptors (A_j, purple and pink) and type II receptors (B_k, orange and yellow) to form a combinatorial set of trimeric signaling complexes (T_{ijk}) with varying affinities (K_{ijk}). Active signaling complexes phosphorylate the SMAD effector with varying efficiencies (e_{ijk}). Equations describe the steady-state levels of each component and the total signal S (Methods 3.6.1).

(B) Optimization systematically identifies potential combinatorial addressing schemes in four steps. i. An orthogonal addressing scheme is specified as orthogonal activation by a set of desired ligand words (red circles). Discretization of ligand space (3×3 grid) enables enumeration of all such addressing schemes. ii. A given orthogonal addressing scheme can be translated into target response functions, in which each cell type is activated by exactly one ligand word (yellow) and not by others (blue). Responses to other ligand words (hatched) are unconstrained. iii. Least-squares optimization identifies a global set of affinity (K_{ijk}) and activity (e_{ijk}) parameters, along with a set of receptor expression levels for each cell type, that yield responses similar to the target functions. Upper and lower arrows represent affinity and activity parameters, respectively, for each receptor dimer complexed with each of the two ligands (blue and green arrows). Thin and thick arrows correspond to low and high values, respectively. iv. Responses can be simulated at higher resolution for visualization and further analysis.

(C) After optimization, the crosstalk matrix represents the responses of each cell type at the selected ligand words (orthogonal channels). For orthogonal addressing, this matrix should ideally be diagonal, with each ligand word activating only its target cell type (orange border) with no off-target activation (blue border).

(D) Best optimization results are shown for all 31 possible three-channel orthogonal addressing schemes (Methods 3.6.4). (Top) Distributions of on-target (orange) and off-target (blue) activation levels are plotted, representing all elements in the crosstalk matrix. Shaded regions span all activity values. (Bottom) The corresponding distinguishability value for each addressing scheme is shown (black). Distinguishability values below 1 (gray region) indicate that the corresponding scheme cannot be successfully addressed. For comparison, the best distinguishability achieved in a one-to-one architecture is shown (red). Addressing schemes (x-axis) are shown in order of decreasing distinguishability.

See also Figures 3.S1 and 3.S2.

To solve the model efficiently, we used Equilibrium Toolkit (EQTK), an optimized Python-based numerical solver for biochemical reaction systems (Bois, 2020; Dirks et al., 2007). For simplicity, the model neglects some specific features of the natural BMP pathway, including sequential binding of ligands to receptors and the hexameric nature of the full BMP signaling complexes (Massagué, 2000; Shi and Massagué, 2003). These features could enable even greater complexity in pathway behavior beyond that described for this minimal model (Methods 3.6.2). This model was capable of reproducing different response functions that were previously observed experimentally and in a more complex model (Antebi et al., 2017), including ratiometric, additive, imbalance, and balance behaviors (Figure 3.S1).

3.3.2 An optimization approach identifies possible addressing schemes

Here, using the model, we sought to identify mixtures of ligands at specific concentrations, or "ligand words," that preferentially activate, or address, specific "cell types," defined here and throughout the paper as a group of cells sharing a common receptor expression profile. (An overview of addressing terminology is provided in Appendix A.) We started by searching for instances of "orthogonal addressing," where each ligand word exclusively activates a single cell type, providing one communication channel per cell type. Intuitively, increasing the number of variants of ligand (n_L) and receptors (n_A and n_B) should expand the number N of possible channels by allowing greater diversity of ligand words and cell types. However, it remains unclear how the number and quality of channels in a promiscuous architecture compares to that possible in a one-to-one architecture, how the number of addressable channels grows with increasing ligand and receptor multiplicity, and what biochemical properties enable optimal orthogonal addressing.

To systematically identify parameters that generate orthogonal channels, we used an optimization approach (Figure 3.2B). We considered discrete ligand concentrations,

allowing each ligand to take on one of three logarithmically spaced concentrations, $10^0 = 1$, $10^{1.5} \approx 32$, and $10^3 = 1{,}000$ arbitrary units (AU), reflecting the experimentally observed input dynamic range for BMP signaling (Antebi et al., 2017; Bradford et al., 2019; Hatsell et al., 2015). This discretization defines a finite set of 3^{n_L} possible ligand words. To identify a system with N channels, we chose a subset of N ligand words (Figure 3.2Bi). Each such choice defines an "addressing scheme." Achieving an addressing scheme requires identifying N cell types that are each individually activated by one word (Figure 3.2Bii). We then used least-squares optimization to identify biochemical parameters (affinities, K_{ijk}, and activities, e_{ijk}) and N receptor expression profiles (one for each cell type) that best implement the target addressing scheme (Figure 3.2Biii; Methods 3.6.3). To obtain a more complete view of the functional behavior, we then computed the responses of each cell type on a higher-resolution (10×10) grid of ligand levels (Figure 3.2Biv).

To quantify the channel structure of the resulting communication system, we computed the crosstalk matrix (Figure 3.2C), where each row is a ligand word, each column is a cell type, and each value represents the normalized response of that cell type to the corresponding ligand word. Diagonal elements of this matrix represent "on-target" signaling, which ideally approach 1. Off-diagonal elements represent "off-target" signaling, ideally 0.

3.3.3 Two ligands can orthogonally address five distinct cell types

To test whether the promiscuous architecture can improve on a one-to-one system, we applied our optimization approach to search for two-ligand systems ($n_L = 2$) that generate three orthogonal channels in a model with two type I and two type II receptor subunits ($n_A = 2$ and $n_B = 2$), reflecting the receptor multiplicity seen in *Drosophila*. We enumerated all 31 possible discrete addressing schemes, optimized parameters for each scheme, and analyzed the resulting responses (Figure 3.2D; Methods 3.6.4).

In 28 of the 31 possible schemes, all on-target activity levels (orange shaded regions) exceeded all off-target activity levels (blue shaded regions), giving rise to orthogonal addressing. To quantify the addressing specificity, we computed a distinguishability score, defined as the fold difference between the lowest on-target activity and the highest off-target activity (Methods 3.6.5). The best scheme, based on using each ligand individually as well as a word with both ligands at their maximal level, produced a distinguishability greater than 45 (Figure 3.2D, scheme 1). (We note that these results represent a lower bound on the potential addressing capacity and specificity, as global optima are not guaranteed.) By contrast, one-to-one systems achieved distinguishability values of only ~1.4 for three channels (Figure 3.2D; Methods 3.6.6).

Inspection of the addressing schemes showed that they typically used combinations of archetypal response functions previously observed in the BMP signaling pathway (Figure 3.S1) (Antebi et al., 2017). In most schemes, two cell types produced opposite ratiometric responses to the two ligands, with the third cell type exhibiting a variety of other responses (Figure 3.S2). These included a balance detector, in which the combination of the two ligands synergistically activated the pathway more than either ligand alone (Figure 3.S2, e.g., schemes 1 and 2); a nonmonotonic response, in which the pathway was most highly activated at intermediate concentrations of a given ligand (e.g., schemes 3 and 4); a distinct ratiometric response (e.g., schemes 18 and 19); and an additive response to the two ligands (e.g., scheme 17). Many of these response types require ligand-receptor promiscuity. For example, ratiometric responses cannot occur in a one-to-one architecture, as an additional ligand that signals through a different receptor cannot decrease the response to the activating ligand. Thus, the ability of cells to access a variety of multi-ligand response functions with different receptor configurations facilitates addressing.

We extended this analysis to systems with up to eight channels. (The value of eight channels reflects the discretization of ligand concentration space and is not inherent in the system.) With a fly-like model ($n_L = 2$, $n_A = 2$, $n_B = 2$), up to seven orthogonal channels could be addressed with distinguishability greater than that possible in a corresponding one-to-one model (Figures 3.3A–B; Methods 3.6.6). Fewer channels (lower bandwidth) could be achieved with greater distinguishability. For instance, a five-channel scheme exhibited a distinguishability of 3.6 through a combination of ratiometric, balance detection, and nonmonotonic responses (Figure 3.S3A). Taken together, these results demonstrate that two ligands with promiscuous ligand-receptor interactions can address a larger number of cell types, albeit at varying levels of distinguishability.

3.3.4 Addressing can occur despite gene expression noise

Stochastic fluctuation, or noise, in gene expression presents a challenge for addressing (Elowitz et al., 2002; Raser and O'Shea, 2005). On the one hand, signaling must be sensitive to receptor expression in order for cell types to have different responses to the same ligand words. On the other hand, if sensitivity is too high, receptor expression noise could disrupt addressing. Here, we asked whether addressing could occur despite correlated (extrinsic noise) and uncorrelated (intrinsic noise) fluctuations in receptor expression, each assumed to have a physiologically reasonable coefficient of variation (ratio of the standard deviation to the mean) of 0.5 (Elowitz et al., 2002; Raj et al., 2006; Suter et al., 2011).

To characterize the extent to which each type of noise degrades addressing, we computed receiver operating characteristic (ROC) curves and corresponding area under the curve (AUC) values (Figure 3.3C; Methods 3.6.7), which characterize the proportion of on- and off-target cells that are correctly classified (Hanley and McNeil, 1982). (AUC values range from 0.5 for a random system to 1.0 for an ideal system.)

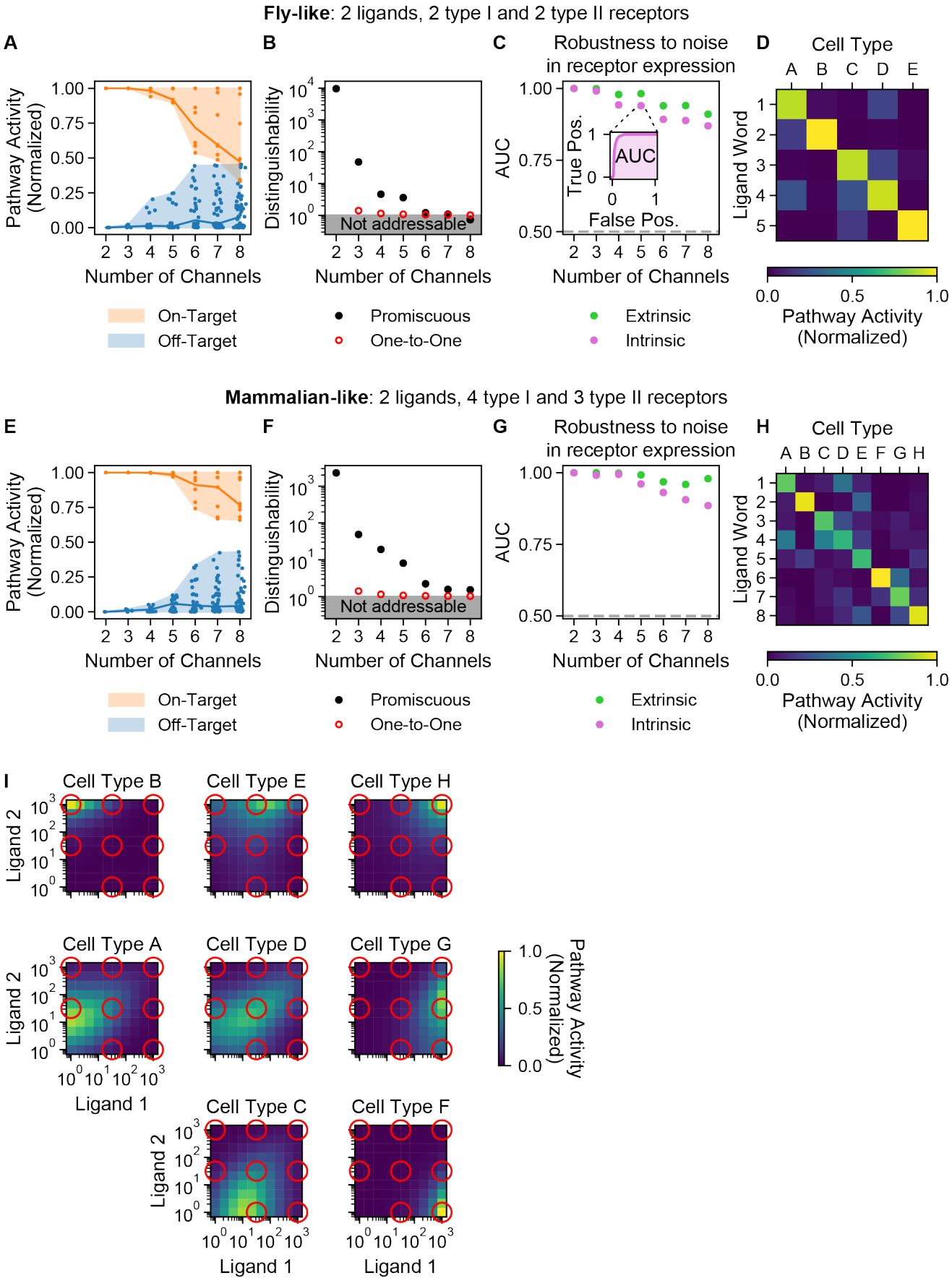
Fly-like: 2 ligands, 2 type I and 2 type II receptors
A
Pathway Activity (Normalized)
Number of Channels
On-Target
Off-Target
B
Distinguishability
Not addressable
Number of Channels
Promiscuous
One-to-One
C
Robustness to noise in receptor expression
AUC
True Pos.
False Pos.
Number of Channels
Extrinsic
Intrinsic
D
Cell Type
Ligand Word
Pathway Activity (Normalized)
Mammalian-like: 2 ligands, 4 type I and 3 type II receptors
E
F
G
H
I
Cell Type B
Cell Type E
Cell Type H
Cell Type A
Cell Type D
Cell Type G
Cell Type C
Cell Type F
Ligand 1
Ligand 2

Figure 3.3: Two ligand variants can independently address eight cell types with high specificity and robustness. Continued from previous page.

(A) In the fly-like model with two type I and two type II receptor subunits, the pathway activities of each cell type in response to each ligand word (y-axis) are plotted for varying numbers of channels (x-axis), using the optimal parameters for each bandwidth. Shaded regions span full distribution of on-target (orange) and off-target (blue) activities, and lines indicate median values.

(B) Distinguishability values are plotted for each number of channels (black), together with the optimal values achieved for the same bandwidths in a one-to-one architecture (red). The five-channel system is further analyzed in (D).

(C) Robustness to receptor expression fluctuations was evaluated for the top-performing system of each bandwidth. Optimized receptor expression levels were perturbed in a correlated or uncorrelated way to represent, respectively, extrinsic (green) or intrinsic (purple) noise, with a coefficient of variation of 0.5. The resulting receiver operating characteristic (ROC) curves are computed by comparing true and false positive rates for classifying on- and off-target values at different thresholds (inset), and the corresponding area under the curve (AUC) values are plotted for each bandwidth. A perfect classifier has AUC 1, and a random classifier has AUC 0.5 (gray dashed line).

(D) The crosstalk matrix shows the response of each cell type at each ligand word of interest for the five-channel example from (A–C). Perfect orthogonal specificity would yield a diagonal matrix.

(E) The pathway activities for a mammalian-like model with four type I and three type II receptors are shown, as in (A).

(F) As in (B), distinguishability values are plotted for the mammalian-like model from (E) (black), along with the optimal values achieved for the same bandwidths in a one-to-one architecture (red). The eight-channel system is further analyzed in (H–I).

(G) AUC values for the top parameter set of each bandwidth are shown, as in (C).

(H) The crosstalk matrix for the eight-channel system in the mammalian-like model is shown, as in (D).

(I) The full responses of each cell type are shown for the eight-channel system analyzed in (H). Red circles correspond to the eight ligand words, and cell types are spatially arranged according to the ligand word to which they preferentially respond. For example, the bottom right cell type (cell type F) is orthogonally activated by high levels of ligand 1 only, while the top right cell type (cell type H) would be activated by combining high levels of ligand 1 and 2 together. The bottom left ligand word, with low levels of both ligands, is non-activating and therefore omitted.

See also Figures 3.S3, 3.S4, and 3.S5.

The five-channel system showed separation between channels (Figure 3.3D) but also demonstrated high AUC values of 0.9820 and 0.9400 with extrinsic and intrinsic noise sources, respectively. The more stringent metric of distinguishability, which is sensitive to incorrect activation of even a single cell type, was impacted more by intrinsic than extrinsic noise (Figure 3.S3B, left). These results suggest that minimizing intrinsic noise is important for maximizing addressing capacity.

The addressing capacity of a system depends on the minimum acceptable distinguishability level. Five-channel addressing could be achieved with a distinguishability threshold of 2, which could be physiologically reasonable given that two-fold changes in signaling pathway activation have been shown to alter cell fate decisions (Dessaud et al., 2007; Falo-Sanjuan et al., 2019; Van de Walle et al., 2013; Zagorski et al., 2017). Greater values of distinguishability, such as 4 or 10, were achieved with four or three channels, respectively (Figure 3.S4). In general, higher distinguishability thresholds translated to reduced noise sensitivity (Figures 3.S4A–C). In particular, the three-channel system (distinguishability greater than 10) exhibited AUC values of 0.9992 and 0.9914 for extrinsic and intrinsic noise, respectively. These results show that systems with just two ligands and only two variants of each receptor type can provide multiple channels with reasonable levels of distinguishability.

3.3.5 The number of addressable channels increases with the number of receptor variants

BMP receptor multiplicity has varied during evolution, leading to different numbers of receptor variants in *Drosophila* (two type I and two type II), humans (four type I and three type II), and other species (Massagué, 1998; Newfeld et al., 1999; O'Connor et al., 2006). What additional addressing capabilities emerge with this increase in receptors? A mammalian-like model with four type I and three type II receptor subunit variants outperformed the fly-like model (Figures 3.3E–F), achieving better specificity at any given number of channels (cf. Figures 3.3A–B). In fact, in this model, two ligands were able to address as many as eight orthogonal channels with 1.5-fold distinguishability between on- and off-target activity (Figure 3.3F) and high AUC values (Figure 3.3G), resulting in a generally diagonal crosstalk matrix (Figure 3.3H). Six-, five-, and four-channel systems could be achieved with distinguishability values greater than 2, 4, and 10, respectively (Figures 3.S4D–F). Five-channel addressing yielded AUCs of 0.9924 and 0.9608 for extrinsic and intrinsic noise, while four-channel addressing gave near-perfect AUCs of 0.9983 and 0.9944 (Figures 3.S4E–F).

Eight-channel addressing was more robust to extrinsic than intrinsic noise in receptor expression levels, with AUC values of 0.9794 and 0.8853 for extrinsic and intrinsic noise, respectively (Figure 3.3G). Distinguishability values remained above 1 for correlated fluctuations of receptor expression but not for uncorrelated noise (Figure 3.S3B, right). The overall addressing scheme resulted from diverse single-cell responses, including ratiometric, balance detection, and nonmonotonic behaviors (Figure 3.3I). Taken together, these results show that a modest increase in the number of receptor variants generates a substantial expansion in addressing capacity, achieved through a variety of single-cell responses.

While eight is the maximum number of channels in the three-level ligand discretization scheme, more channels may be possible with higher-resolution grids. For instance, a four-level ligand discretization scheme allows up to fifteen channels. At this level, it was no longer computationally feasible to systematically test all possible addressing schemes. Instead, we sought to optimize increasing bandwidths, choosing random schemes with a given number of channels until successfully optimized (Methods 3.6.3).

We performed this analysis for both the fly-like and mammalian-like models. With a fly-like model, we successfully optimized systems with up to five orthogonal channels (Figures 3.S5A–B). This five-channel system exhibited distinguishability of 2.3 (Figure 3.S5C) and, like the system obtained using a lower-resolution grid, exhibited a combination of ratiometric, balance detection, and nonmonotonic responses (Figure 3.S5D; cf. Figure 3.S3A). In the mammalian-like model, we identified a seven-channel system with distinguishability of 2.2 as well as a more weakly addressable eight-channel system with distinguishability of 1.4 (Figures 3.S5E–G). The eight-channel system used similar types of responses to those in the three-level discretization scheme (Figure 3.S5H; cf. Figure 3.3I). Overall, the bandwidths achievable and the responses observed remained qualitatively similar using the higher-resolution ligand grid compared to the three-level discretization. Allowing ligand concentrations to vary continuously or exploring parameter space more comprehensively could reveal greater addressing capacity.

3.3.6 Promiscuous architectures enable subset addressing

Combinatorial addressing can extend beyond the addressing of individual cell types, as explored thus far, to generate more complex, multi-cell type response patterns. In such "subset addressing," each ligand word activates a specific subset of cell types. In the olfactory system, for example, odorants activate specific subsets of olfactory

receptor neurons, giving rise to a combinatorial representation of odors (Hallem and Carlson, 2006; Malnic et al., 1999). Subset addressing systems can be characterized by an "addressing repertoire," defined as all unique subsets of cell types (channels) that can be addressed across all possible ligand words (Figure 3.4A). For example, a system with three cell types that can only be orthogonally activated would have three channels (Figure 3.4A, top). The highest bandwidth of seven addressable subsets occurs when all cell types can be activated in any required combination using some ligand word (Figure 3.4A, bottom).

We first asked what addressing repertoires are possible in mammalian-like systems with three cell types, or seven possible channels. Using the optimization approach, we identified parameters that achieve the fully addressable seven-channel system (Figure 3.4B; Methods 3.6.8). We then generalized this approach to all 32 possible addressing repertoires (Methods 3.6.9), successfully identifying parameter sets that generated every repertoire with distinguishability greater than 1 and 29 repertoires with distinguishability greater than 2 (Figure 3.4C). These results show that two ligand variants can generate any addressing repertoire of three cell types, most with high distinguishability.

Achieving such a broad set of addressing repertoires requires promiscuous ligand-receptor interactions. In a one-to-one model, having high concentrations of all ligands will activate all cell types; therefore, any addressing repertoire in which the three cell types cannot be simultaneously co-activated requires a promiscuous architecture (Figure 3.4C, orange stars; Methods 3.6.10). Taken together, these results demonstrate that the promiscuous ligand-receptor architecture allows diverse addressing repertoires, beyond those achievable in a one-to-one model.

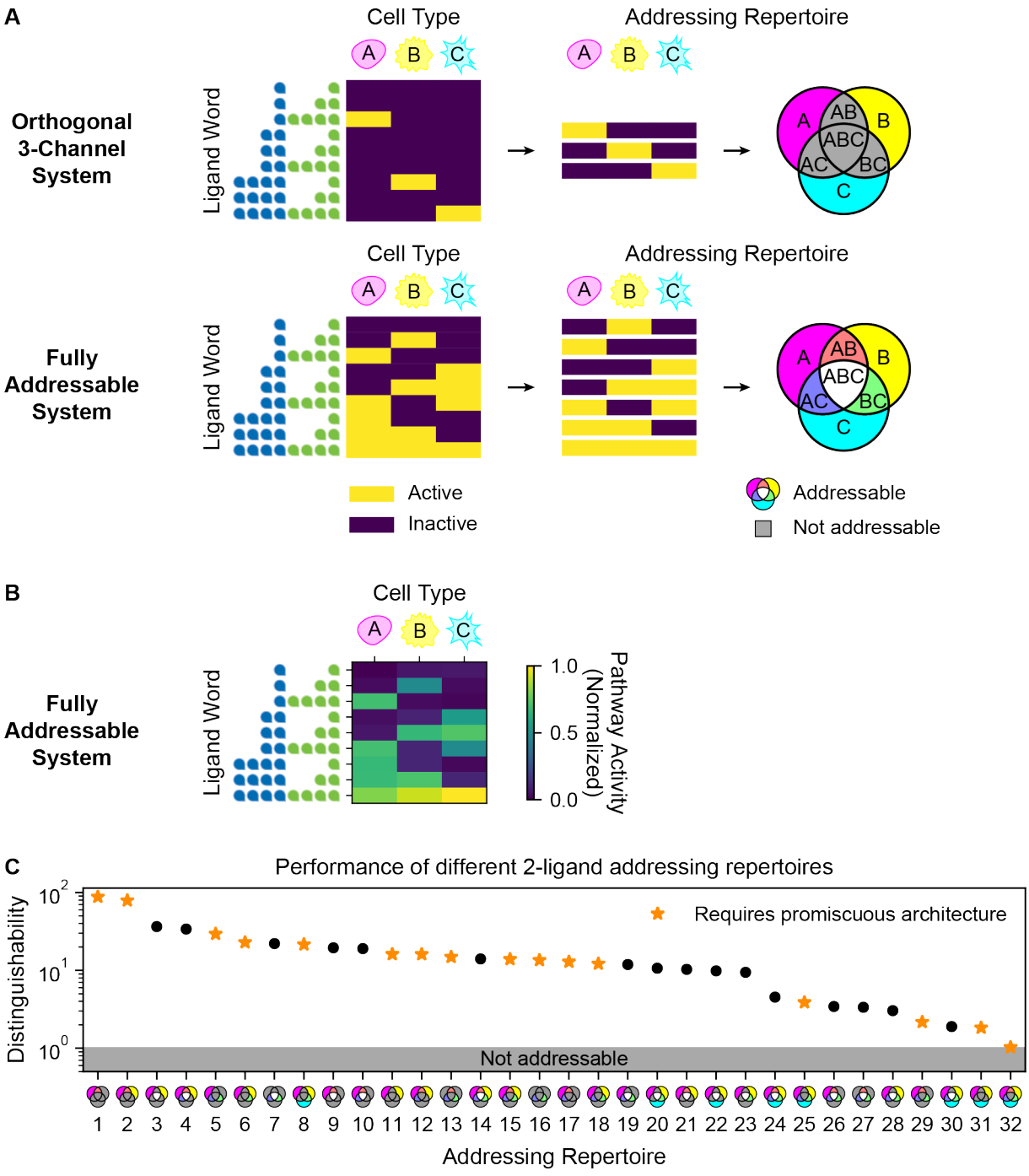

Figure 3.4: Promiscuous architecture enables diverse addressing repertoires. Continued on next page.

Figure 3.4: Promiscuous architecture enables diverse addressing repertoires. Continued from previous page.

(A) For different parameter sets, the responses of three cell types (A, magenta; B, yellow; and C, cyan) to a titration of two ligands (blue and green) are shown (left). Unique rows reveal the subsets of cell types that can be activated across all ligand words (center). Addressable subsets can also be represented as a Venn diagram (right), where colored regions represent subsets that are activated by at least one ligand combination and gray regions represent subsets that cannot be addressed by any ligand combination. These subsets constitute the "addressing repertoire" of a system. Addressing capability can vary widely. Examples include purely orthogonal activation (top) and all possible subsets (bottom).

(B) We optimized parameters to achieve the fully addressable system of (A). Simulating the responses of the three cell types to each ligand word confirms that any of the seven possible subsets can be successfully addressed.

(C) We generalized the optimization approach to identify parameters achieving each possible addressing repertoire of three cell types in a mammalian-like model with four type I and three type II receptors. The optimal distinguishability value for each repertoire is plotted. Orange stars indicate addressing repertoires that cannot be achieved in the one-to-one architecture (Methods 3.6.10).

3.3.7 Cell lines show combinatorial addressing in vitro

Having analyzed the theoretical conditions that permit addressing, we next asked whether addressing could occur in living cells. Previous work revealed that individual cell lines exhibit complex responses to ligand combinations that can be altered by perturbing the expression of specific receptors (Antebi et al., 2017). However, it is unclear to what extent the responses generated in this way could allow differential activation of distinct cell types using different ligand words. Experimentally analyzing the responses of multiple cell lines with differing receptor expression profiles to the same panel of ligand combinations could reveal the potential for addressing.

To this end, we engineered cell lines with different receptor expression profiles and analyzed their responses across two-dimensional titrations of two different ligand

pairs—BMP2+BMP9 and BMP9+BMP10—in which combined signaling activity was shown to depend on receptor expression profiles (Klumpe et al., 2022). To read out pathway activity, we used a transcriptional fluorescent reporter for SMAD1/5/8 containing BMP response elements from the Id1 promoter (Korchynskyi and ten Dijke, 2002). We stably integrated the reporter into cell lines with different receptor expression profiles, then analyzed their responses to a range of BMP ligand combinations by flow cytometry 24 hours after ligand addition (Methods 3.6.11). We used ligand concentrations up to 1,000 ng/mL to broadly survey physiologically relevant levels. While measured serum levels are around 0.1–10 ng/mL (Albilia et al., 2013; David et al., 2008; Herrera and Inman, 2009; Penn et al., 2017), effective levels at the cell surface are likely to be higher due to local production, consistent with the higher concentrations of 10–500 ng/mL used for in vitro studies of various BMP-dependent processes (Blackwell et al., 2009; Grassinger et al., 2007; Kim et al., 2013; LaVaute et al., 2009; Valera et al., 2010; Zhao et al., 2003; Zhu et al., 2017). Finally, BMP ligands are used clinically in concentrations on the order of 1,000 ng/mL (Gupta and Khan, 2005; Kim et al., 2015). Thus, our chosen titration range effectively represents these varied conditions.

We started with a previously characterized epithelial cell line, NAMRU mouse mammary gland (NMuMG) cells, that robustly responds to a variety of BMP ligands (Antebi et al., 2017). NMuMG cells responded additively to BMP2 and BMP9 (Figure 3.5A, first), consistent with previous results (Antebi et al., 2017). Since opposing ratiometric responses are a common "motif" in the addressing schemes identified above (Figures 3.2 and 3.3), we sought to generate additional cell lines that would exhibit such responses by knocking down receptors with known preferences for each specific ligand. In the NMuMG background, knockdown of ACVR1, which directly interacts with BMP9 (Luo et al., 2010), resulted in a minimal response to BMP9 but a strong response to BMP2, thereby generating a ratiometric response

profile (Figure 3.5A, second). By contrast, knockdown of BMPR2, the major BMP2 receptor (Xia et al., 2007), gave rise to a reduced responsiveness to BMP2 with a strong BMP9 response, producing a complementary ratiometric response (Figure 3.5A, third). In contrast to receptor knockdown, which increases competition for a limited receptor pool and therefore could increase the complexity of multi-ligand responses, we anticipated that ectopic receptor expression might relieve receptor competition and thereby generate more additive responses (Klumpe et al., 2022). Indeed, ectopic ACVRL1 expression increased sensitivity to ligands without qualitatively altering the combinatorial response (Figure 3.5A, fourth). Finally, for comparison with a more distantly related cell type, we also analyzed E14 mouse embryonic stem cells (mESCs), which differ in receptor expression profile from NMuMG cells in at least three receptors (Antebi et al., 2017). mESCs responded maximally to combinations of BMP2 and BMP9 together (Figure 3.5A, fifth).

Comparing the responses of these cell lines across double titrations of BMP2 and BMP9 showed that the two ligands could be used at different concentrations to preferentially activate certain cell types individually or in groups (Figure 3.5B). For example, moderate levels of BMP9 alone predominantly activated NMuMG with ectopic ACVRL1 (Figure 3.5B, word 1), while higher levels additionally activated first NMuMG (word 2) and then NMuMG with BMPR2 knockdown (word 3). Intermediate levels of both ligands abolished activation of NMuMG with BMPR2 knockdown and allowed activation of NMuMG with ACVR1 knockdown (word 4). High levels of BMP2 activated all cell types except NMuMG with BMPR2 knockdown (word 5), and high levels of both ligands activated all cell types (word 6). In this way, distinct combinations of BMP2 and BMP9 enabled preferential activation of six distinct combinations of five cell types (Figure 3.5C), establishing that the BMP pathway has combinatorial addressing capability.

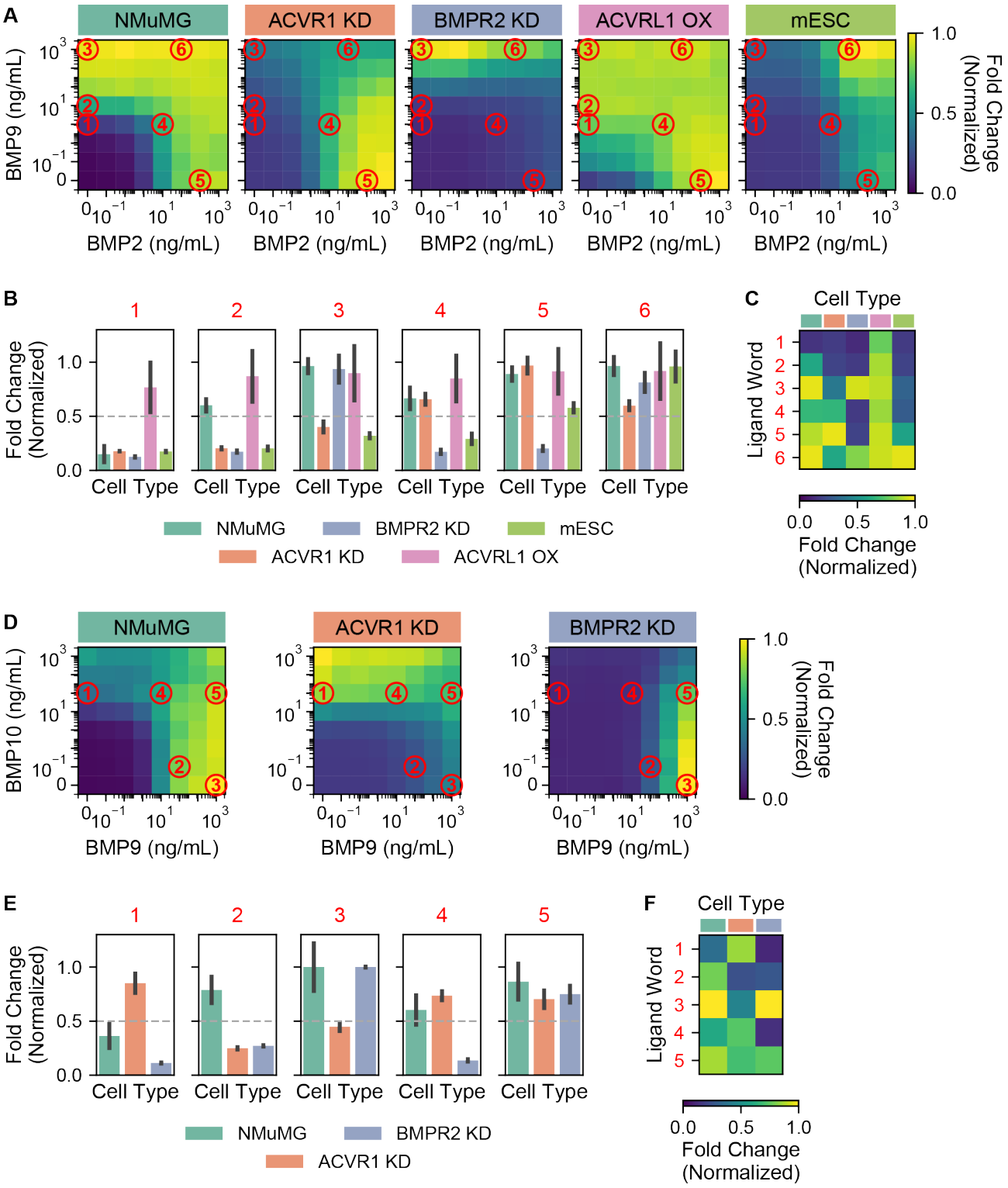

Figure 3.5: Cell lines preferentially respond to different ligand combinations.
Continued on next page.

Figure 3.5: Cell lines preferentially respond to different ligand combinations. Continued from previous page.

(A) Responses were measured for, from left to right, NMuMG cells, NMuMG cells with ACVR1 knockdown (KD), NMuMG cells with BMPR2 KD, NMuMG cells with ACVRL1 overexpression (OX), and mESCs, using flow cytometry of an integrated fluorescent protein reporter (Methods 3.6.11). Each cell line was exposed to a double titration of BMP2 and BMP9, and responses were quantified by taking the mean of at least 3 replicates. For each cell line, fold change is calculated relative to the baseline fluorescence with no added ligand and then normalized by the maximum value. Responses at select ligand words (red circles) are analyzed further in (B).

(B) For select ligand words from (A), the responses of each cell line are shown. Error bars indicate standard deviation of at least 3 repeats. Ligand words were chosen by fixing a threshold of 0.5 (gray dashed line) and identifying those ligand combinations yielding unique on- and off-target activation patterns.

(C) Data from (B) are summarized by showing the response of each cell type (columns) to each ligand word (rows), illustrating that distinct ligand words can activate different subsets of cell types.

(D) Responses of NMuMG, NMuMG with ACVR1 KD, and NMuMG with BMPR2 KD to BMP9 and BMP10 are shown, as in (A).

(E) As in (B), the responses of each cell type at selected ligand words are shown.

(F) As in (C), the responses of each cell type (columns) to each ligand word (rows) confirm that distinct ligand words preferentially activate distinct groups of cell types.

See also Table 3.S1.

To extend this analysis to another ligand pair, we evaluated the responses of three NMuMG-derived lines to BMP9 and BMP10 (Figure 3.5D). In this pairwise titration, NMuMG and NMuMG with BMPR2 knockdown responded more strongly to BMP9 than to BMP10 alone, with BMP10 reducing activation by BMP9 when present in combination. BMP10 inhibition of BMP9 signaling was stronger when BMPR2 was knocked down. By contrast, ACVR1 knockdown cells exhibited the opposite response, responding more strongly to BMP10 than to BMP9. We were able to identify five ligand words that activated distinct combinations of these three cell

types (Figures 3.5E–F). BMP10 alone activated only ACVR1 knockdown cells (Figure 3.5E, word 1). Intermediate levels of BMP9 activated the wild-type cells (word 2), while higher levels additionally activated BMPR2 knockdown cells (word 3). NMuMG and ACVR1 knockdown cells could be simultaneously activated with intermediate levels of both ligands (word 4), while additional BMP9 enabled activation of all three cell types simultaneously (word 5). These results provide additional evidence that the BMP pathway could potentially support addressing.

The ability to achieve addressing in vitro does not demonstrate that addressing occurs in physiological contexts. However, single-cell gene expression atlases reveal that the receptor profiles of NMuMG cells and their perturbed derivatives resemble those in some natural cell types (Table 3.S1) (Tabula Muris Consortium, 2020; Tabula Muris Consortium et al., 2018). It will be interesting to determine whether the profiles analyzed here play natural addressing roles in vivo.

3.3.8 Response function diversity increases addressability

The values of key biochemical parameters—affinities and activities—ultimately determine the addressing bandwidth of a promiscuous ligand-receptor system. What is the distribution of addressing bandwidth across parameter sets? Are there design rules that allow tuning of those values, in absolute or relative terms, to optimize addressing? Information theory provides a natural framework to answer these questions (Huntley et al., 2016; Itzkovitz et al., 2006). More specifically, the concept of mutual information can be used to quantify the addressing power of a promiscuous ligand-receptor system without assuming any particular choice of ligand words or cell types, or any particular mapping between them (Methods 3.6.12).

To identify parameter sets that maximize mutual information, we systematically analyzed the diversity of responses across a set of cell types to a set of ligand

words for different biochemical parameter sets (Figure 3.6A). Mutual information measures information communicated by the optimal subset of ligand words to the optimal subset of cell types, allowing the use of comprehensive libraries. In a fly-like model, we constructed a discrete ligand word library in which each of two ligands takes on one of three concentration values (3^{n_L} ligand words, or 9); a cell type library, in which each of the two type I and two type II receptors is expressed at one of two values ($2^{n_A+n_B}$ cell types, or 16); and a biochemical parameter library, in which each K_{ijk} and e_{ijk} takes on one of two values ($2^{2n_Ln_An_B}$ parameter sets, or 65,536). We then simulated the response of each cell type to each ligand word for each biochemical parameter set and computed the mutual information between the sets of ligand words and pathway activities across the library of cell types (Figures 3.6A–B; Methods 3.6.13). Random, rather than grid-based, sampling of K_{ijk} and e_{ijk} produced similar results (Figure 3.S6A). Mutual information values varied broadly across parameter sets, from 0.32 to 1.91 bits, with a median value of 1.36 bits (Figure 3.6B). By refining our search over biochemical parameters, we were able to identify parameters with values as high as 2.38 bits (Methods 3.6.14).

To assess whether mutual information correlates with addressing, we defined an addressability metric, which quantifies how strongly activation patterns differ for different ligand words without requiring specific targeted profiles (Methods 3.6.15). For every pair of ligand words, we identified the largest fold difference of activation levels across all cell types. This value is high when two ligand words induce distinct responses in at least one cell type. We defined the addressability metric as the lowest such value across all ligand word pairs and calculated this value for a given number of channels N by taking the best choice of all possible subsets of N ligand words. Using this metric, we analyzed addressability for systems with low, intermediate, and high mutual information (Figure 3.6C). For the parameter set of highest mutual information (2.38 bits), each of the eight ligand words activated a distinct cell type

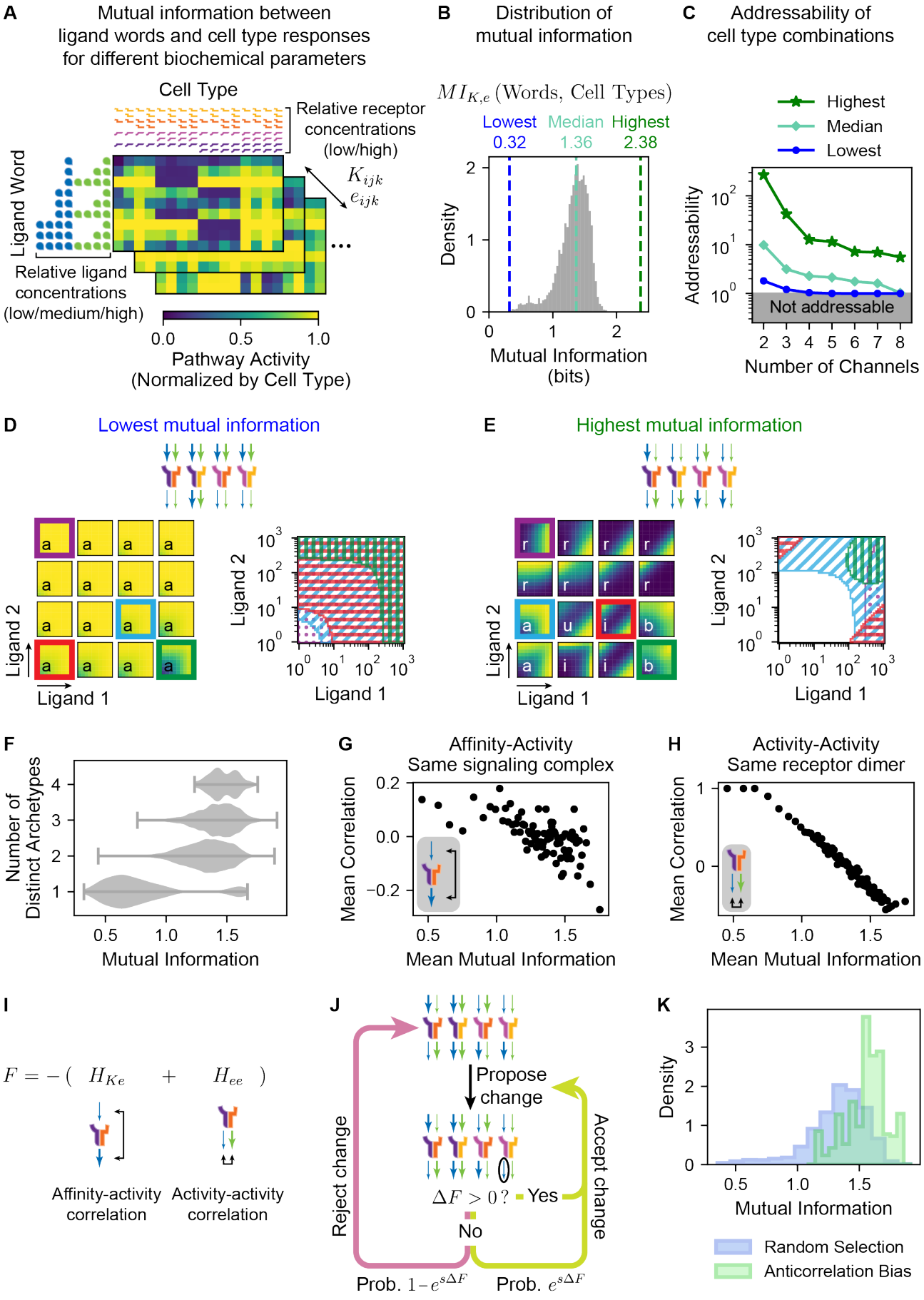
A
Mutual information between ligand words and cell type responses for different biochemical parameters
Cell Type
Relative receptor concentrations (low/high)
K_{ijk}
e_{ijk}
Ligand Word
Relative ligand concentrations (low/medium/high)
0.0
0.5
1.0
Pathway Activity (Normalized by Cell Type)
B
Distribution of mutual information
$MI_{K,e}$ (Words, Cell Types)
Lowest Median Highest
0.32 1.36 2.38
Density
Mutual Information (bits)
C
Addressability of cell type combinations
Highest
Median
Lowest
Addressability
Not addressable
Number of Channels
D
Lowest mutual information
Ligand 2
Ligand 1
E
Highest mutual information
Ligand 2
Ligand 1
F
Number of Distinct Archetypes
Mutual Information
G
Affinity-Activity
Same signaling complex
Mean Correlation
Mean Mutual Information
H
Activity-Activity
Same receptor dimer
Mean Correlation
Mean Mutual Information
I
$F = -($ H_{Ke} $+$ H_{ee} $)$
Affinity-activity correlation
Activity-activity correlation
J
Propose change
$\Delta F > 0$?
Yes
No
Reject change
Accept change
Prob. $1-e^{s\Delta F}$
Prob. $e^{s\Delta F}$
K
Density
Mutual Information
Random Selection
Anticorrelation Bias

Figure 3.6: Information theoretic analysis reveals design principles for combinatorial addressing. Continued from previous page.

(A) Mutual information between a comprehensive library of ligand words (rows) and the corresponding activation patterns across a library of cell types (columns) can be computed across a systematic grid-based sampling of the biochemical parameters K and e (matrices). For each row, one, two, and four ligand symbols indicate low (10^0), medium ($10^{1.5}$), or high (10^3) concentrations of the indicated ligand. Similarly, one or two receptor symbols indicate low (1) or high (100) levels of the indicated receptor for each column.

(B) The distribution of mutual information across biochemical parameters is shown. Dashed lines indicate the lowest (blue), median (cyan), and highest (green) values. High mutual information indicates that many distinct cell type combinations can be specifically activated by distinct ligand words.

(C) The addressability values of activated subsets are shown for different numbers of channels. The addressability reflects the minimal fold difference in the response of at least one cell type when exposed to any two distinct ligand words (Methods 3.6.15). Results are shown for three sets of biochemical parameters generating the lowest, median, and highest mutual information values.

(D) The parameter set with the lowest mutual information is represented schematically (top), as in Figure 3.2Biii. For these parameters, the responses for the library of 16 cell types are shown as a 4×4 grid (bottom left). In each response, the x- and y-axes represent logarithmic titrations of ligands 1 and 2, respectively. All show the same qualitative response of additive ("a") behavior, differing only in their quantitative sensitivity. Schematically, overlaying four differing responses (highlighted in purple, cyan, red, and green) reveals that different ligand words largely address similar combinations of cell types (bottom right), with relatively few distinct subsets represented.

(E) For the parameter set with the highest mutual information (top), the cell types in the library show a variety of response patterns (bottom left): ratiometric ("r"), additive ("a"), imbalance ("i"), and balance ("b"), matching the response archetypes (Figure 3.S1) previously observed experimentally (Antebi et al., 2017). One response not fully matching any archetype is unclassified ("u"). Schematically, overlaying four differing responses (purple, cyan, red, and green) reveals that different ligand words can address many distinct subsets of cell types (bottom right). Note that complexes tend to have opposite values of affinity and activity parameters as well as other parameter anticorrelations, as analyzed in (G–H).

(F) Violin plots indicate the distribution of mutual information values for systems with different numbers of distinct archetypes represented among individual cell response functions. Note that greater archetype diversity enriches for high mutual information.

(G) Anticorrelation of affinity and activity parameters for the same complex is associated with higher mutual information. We analyzed average properties across bins of 800 parameter sets. To measure the correlation between affinity and activity of complexes, we represented low and high values as -1 and 1 and computed the dot product between K and e vectors. The average correlation and mutual information across bins are plotted.

(H) Parameter sets with high mutual information show anticorrelation in the activities of complexes with the same receptor but different ligands. Analysis was done analogous to (G).

(I) We defined a fitness function F that rewards parameter sets exhibiting the anticorrelations observed in (G–H).

(J) An evolutionary algorithm identifies parameter sets that maximize F. At each iteration, a random parameter value is flipped from low to high or vice versa. Changes that increase F are accepted. Changes that decrease F are accepted with indicated probability (bottom), which depends on a selection pressure parameter s. This process is repeated iteratively (Methods 3.6.18).

(K) An evolutionary algorithm enriches for high mutual information. We ran the algorithm with $s > 0$ to favor anticorrelations or with $s = 0$ to randomly sample parameters. For each case, we randomly initialized 2,000 parameter sets and performed 200 iterations. We then evaluated the mutual information for the final value of the parameter set and visualized the resulting distributions. Random selection ($s = 0$, blue) led to a similar distribution of values as the systematically sampled parameter sets (cf. Figure 3.6B), while favoring anticorrelations ($s > 0$, green) resulted in an overall increase in mutual information.

See also Figure 3.S6.

combination with over 5.5-fold addressability. The median parameter set (1.36 bits) addressed up to seven distinct cell type combinations at an addressability of 1.6, while the parameter set with lowest mutual information (0.32 bits) addressed only three distinct cell type combinations with addressability of 1.2. Overall, a 1-bit difference in mutual information can increase addressing specificity as well as bandwidth, enabling diverse responses to different ligand words.

We next sought to understand how high addressing bandwidth arises from the individual response functions of each cell type by comparing the parameter sets with the lowest and highest mutual information. The parameter set with the lowest mutual information generated a homogeneous spectrum of responses across all cell types (Figure 3.6D; Methods 3.6.16). These responses predominantly varied quantitatively in their sensitivity to ligand. By failing to fully exploit the two-dimensional nature of ligand concentration space, this parameter set exhibited limited addressing potential. By contrast, the parameter set with the highest mutual information generated a broad diversity of ligand response functions across the cell types, reproducing the experimentally observed ratiometric, additive, imbalance detection, and balance detection "archetypal" functions (Figure 3.6E; Methods 3.6.16) (Antebi et al., 2017). By generating diverse two-dimensional response functions, this parameter set allowed each ligand word to activate a distinctive combination of cell types. In fact, such a correlation between the diversity of response functions and mutual information is seen across the full library of parameter sets (Figure 3.6F).

3.3.9 Affinity-activity relationships control addressing bandwidth

How do parameter sets with high mutual information generate the varied response functions associated with addressing? Inspection of the parameter set with the highest mutual information revealed two notable relationships between binding affinities K_{ijk} and signaling activities e_{ijk} (Figure 3.6E). First, complexes that formed with strong affinity (large K_{ijk}) often had low signaling efficiency (small e_{ijk}). Second, the activity of a given receptor pair strongly depended on the identity of the bound ligand, producing opposite values for e_{1jk} or e_{2jk}. Systematic analysis of these relationships revealed their dependence on mutual information (Methods 3.6.17). In particular, anticorrelations between the affinity and activity (K_{ijk} and e_{ijk}; Figure 3.6G) and between the activities of complexes with distinct ligands

(e_{1jk} and e_{2jk}; Figure 3.6H) were associated with higher mutual information. These results suggest that such anticorrelations could predict high addressing capacity.

To test whether the anticorrelated structure of the parameters is sufficient to produce high mutual information, we developed an evolutionary algorithm that evolves the biochemical parameters to maximize the above anticorrelations (Figures 3.6I–K). The algorithm starts with an initial parameter set, proposes a random change to one K_{ijk} or e_{ijk} value, and accepts that change with probability 1 if the change increases the fitness function F and with probability $e^{s\Delta F}$ if it does not, where s is a parameter that controls the strength of the selection (Methods 3.6.18). Iteration of this procedure increased mutual information between ligand words and cell type responses to values comparable to the strongest ones identified in the systematic screen (Figure 3.6K; cf. Figure 3.6B).

These results indicate that strong addressing is not rare. It is realized to varying degrees across all of parameter space and enhanced by specific parameter anticorrelations. Further, experimental analysis of ligand-receptor interactions suggests that the natural BMP system may exhibit such anticorrelations (Klumpe et al., 2022). When systematic measurements of responses to pairwise ligand combinations were fit to the same model used here, four of the five ligands analyzed (BMP4, BMP7, BMP9, and BMP10) exhibited properties consistent with formation of strong-affinity, low-activity complexes (e.g., with the BMPR1A/ACVR2B receptors) and weak-affinity, high-activity complexes (e.g., with the ACVR1/ACVR2A receptors). Overall, the parameter fits from this study are consistent with a broad range of addressing capabilities (Figure 3.S6B).

3.4 Discussion

Communication systems such as email enable one to address messages to specific recipients and groups of recipients. Similarly, in multicellular organisms, it is crucial to activate the right cells at the right time and place. A fundamental mystery in cell-cell communication is how freely diffusing ligands can precisely target, or address, specific cell types. The promiscuity of ligand-receptor interactions in BMP and other communication pathways makes this question especially perplexing, since it appears to reduce rather than enhance communication specificity. However, promiscuous architectures are employed for specificity in other biological contexts. For example, promiscuous ligand-receptor interactions in the olfactory system enable a limited number of receptors to sense a great diversity of odorants through a combinatorial population code (Duchamp-Viret et al., 1999; Goldman et al., 2005; Hallem and Carlson, 2006; Malnic et al., 1999). Such architectures also appear analogous to simple neural networks, which can compute complex functions of multi-dimensional inputs (Bray, 1995). This computational ability could allow different cell types to respond to different ligand combinations, as observed experimentally (Figure 3.5) (Antebi et al., 2017; Klumpe et al., 2022).

Our results show that promiscuity could potentially allow ligand combinations to address different cell types or groups of cell types with remarkable specificity (Figures 3.7A–B). Compared to one-to-one architectures that achieve perfect specificity for a limited number of channels, promiscuous signaling pathways can target a large number of cell types at higher specificity (Figure 3.3) as well as enable greater flexibility in addressing arbitrary subsets of cell types (Figure 3.4). High addressing capacity can be a robust feature of promiscuous ligand-receptor systems, withstanding correlated noise in receptor expression levels (Figures 3.3C,G) and emerging across a broad range of biochemical parameter values. A more general mutual information framework identified design principles that maximize addressing capacity (Figure

3.6). In particular, these include anticorrelation between the affinity and activity of a given ligand-receptor complex and anticorrelation between the activities of two ligands interacting with the same receptor dimer. Together, these results show how addressing specificity can emerge from molecular promiscuity in a canonical cell-cell communication system.

Are the biochemical parameters of the natural BMP pathway compatible with addressing? Quantification of BMP ligand-receptor interaction parameters has been done for select components, although direct measurements of the activity of specific signaling complexes have not yet been achieved (Karim et al., 2021). Systematic analysis of pairwise ligand combinations showed complex responses to ligand combinations and revealed their dependence on specific receptors (Klumpe et al., 2022). When fit to the same model used here (Figure 3.2A), these data provide estimated values for K_{ijk} and e_{ijk}. These values exhibit the types of anticorrelations that favor addressing, suggesting that the BMP pathway may have evolved to facilitate high-capacity addressing. However, experimental studies of specific developmental systems will be necessary to establish to what extent combinatorial addressing functions in natural contexts.

The addressing principles described here do not reflect the full complexity of cellular signaling systems. We have focused on computations that can arise through promiscuous ligand-receptor binding and activation of intracellular second messengers. However, our model omits myriad additional processes that can modulate and regulate cell signaling. Examples include nonhomogeneous spatial distribution of ligands and receptors, such as in polarized cells (DeWitt et al., 2002; Kuwada et al., 1998); trafficking of receptors (Burke et al., 2001; Resat et al., 2003; Shankaran et al., 2012) and second messengers (Schmierer et al., 2008); and signaling-induced feedback loops (Shankaran et al., 2007). These processes can enrich the com-

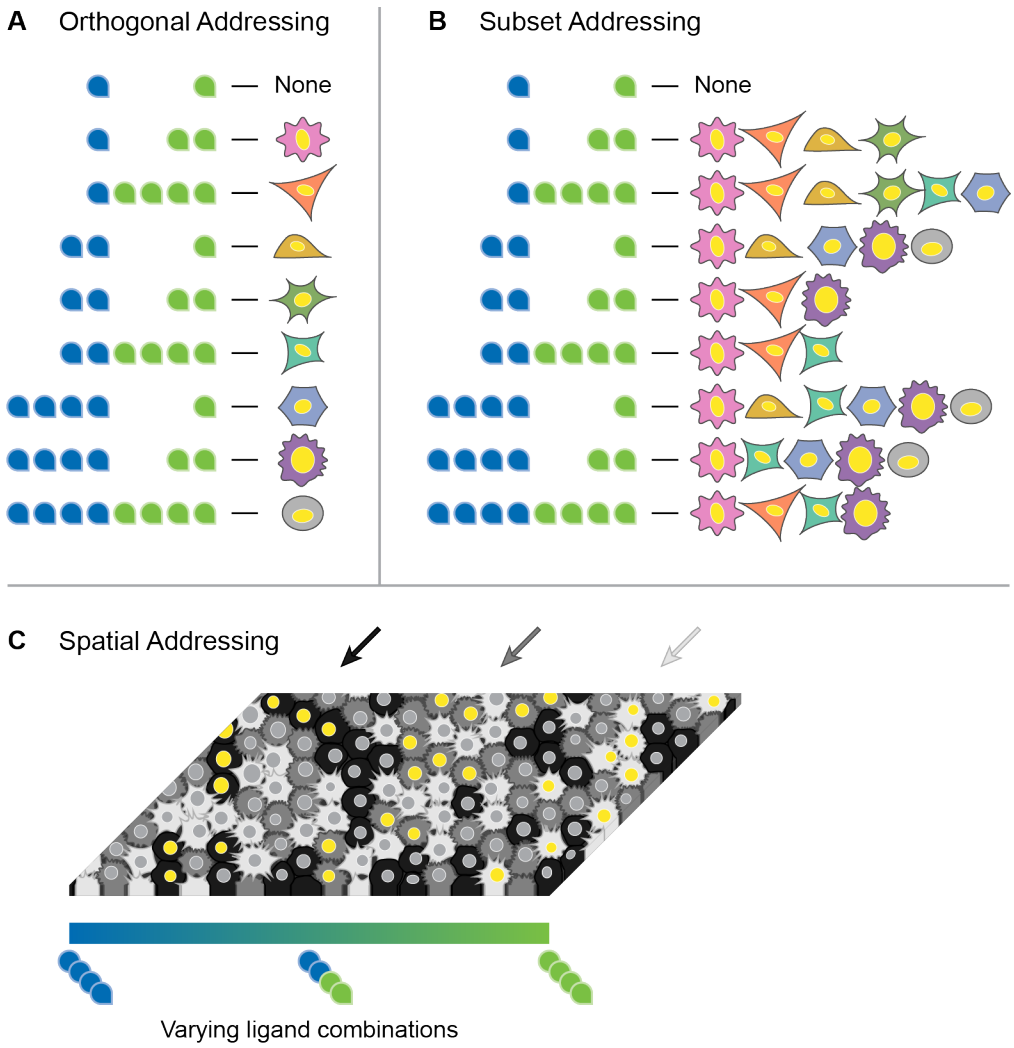

Figure 3.7: Promiscuous ligand-receptor interactions allow for flexible and high-bandwidth addressing.

(A) Promiscuous ligand-receptor interactions enable orthogonal addressing, in which individual cell types can be specifically activated using combinations of only two different ligand variants (cf. Figures 3.3E–I).

(B) Promiscuous ligand-receptor interactions enable subset addressing, in which different ligand words address diverse cell type combinations (cf. Figure 3.6E).

(C) This notional schematic shows how two antiparallel morphogen gradients could address different cell types (black, dark gray, and light gray) in specific spatial regions. Yellow nuclei indicate activation. In this example, high levels of blue ligand activate the black cell type (left), the combination of both ligands (blue and green) activates the dark gray cell type (center), and high levels of green ligand activate the light gray cell type (right).

plexity of computations beyond those analyzed here and thereby potentially further enhance the number and distinguishability of addressing channels. While smaller, more focused models have been essential for developing our current understanding of signaling, fully describing biological systems will require building integrated models that span multiple scales and processes (Wells and Wiley, 2018; Wiley et al., 2003).

BMPs function as morphogens, provoking the question of whether and how addressing plays out in a dynamic, spatially extended "heterocellular" tissue context (Wells and Wiley, 2018). BMP-dependent developmental patterning processes typically use multiple BMP ligands in spatially and temporally overlapping gradients that can be further shaped by shuttling and other extracellular processes. For example, during early *Xenopus* embryo development, an antiparallel gradient of BMP ligands is formed between ventral and dorsal centers (Ben-Zvi et al., 2008; Reversade and De Robertis, 2005). Similarly, overlapping expression patterns of GDF5 and multiple BMP ligands, together with distinct receptor expression patterns, play a key role in activation (and suppression) of BMP signaling in specific cell populations during joint formation (Lyons and Rosen, 2019; Salazar et al., 2016). In such overlapping gradients, addressing could allow different cell types, all with functional BMP pathways, to each selectively respond in distinct regions based on the concentrations of multiple ligands (Figure 3.7C).

Additionally, temporal changes in receptor expression are common during development (Danesh et al., 2009; Dewulf et al., 1995; Erickson and Shimasaki, 2003; Sanyal et al., 2002). For instance, in neural precursors, *Bmpr1a* is expressed early and ubiquitously; subsequent treatment with BMP2 induces activation of BMPR1A and expression of *Bmpr1b* (Panchision et al., 2001). These different receptor expression states could preferentially respond to different ligand combinations and

therefore be addressable. Spatiotemporal addressing could be tested experimentally by genetically modifying the expression of BMP variants in developmental contexts and analyzing the effects on different cell types. In vitro reconstitution of multi-ligand gradients could allow a complementary, systematic analysis of spatial addressing (Li et al., 2018).

An increasing amount of expression data is available from cell atlas projects, enabling analysis of expression profiles of ligands and receptors across diverse cell types. Together with quantitative measurements of effective biochemical parameters, these data could potentially be used to design ligand combinations that selectively address target cell populations. The ability to design selective targeting would be useful in biomedical applications such as directed differentiation and targeted therapy. For example, recombinant BMP2 has been tested in a variety of therapeutic applications, largely related to promoting bone healing and regrowth. However, there are substantial risks, such as ectopic bone formation, respiratory failure, tissue inflammation, and others (Epstein, 2011; Poon et al., 2016). If these complications result from undesired activation of off-target cell types, using a combination of ligands could potentially provide more specific addressing of the appropriate cell type(s). Other potential therapeutic applications for modulators of BMP signaling include cardiac fibrosis, where BMP2 and BMP7 have both shown promise in animal models (Flevaris et al., 2017; Wang et al., 2012); Parkinson disease, where BMP2 and GDF5 both appear to promote survival of dopaminergic neurons (Hegarty et al., 2014; O'Keeffe et al., 2017; O'Sullivan et al., 2010); and cancer, where inhibition of BMP signaling reduces tumor formation in mice (Yokoyama et al., 2017). As the range of clinical applications targeting BMP signaling continues to grow, it will be essential to determine whether combinations of ligands could provide greater specificity than individual ligands.

The principles elucidated here in the context of BMP signaling could apply to other pathways that exhibit promiscuous ligand-receptor interactions, including the broader TGF-β pathway as well as the Wnt, Eph-Ephrin, FGF, and JAK-STAT pathways. The principle of addressing suggests that beyond sensing the concentration of a given set of ligands, these pathways may serve more broadly as computational devices that exploit promiscuous interactions, enabling cells to tune in to specific ligand words and thereby receive information specifically addressed to them.

3.5 Supplemental Information

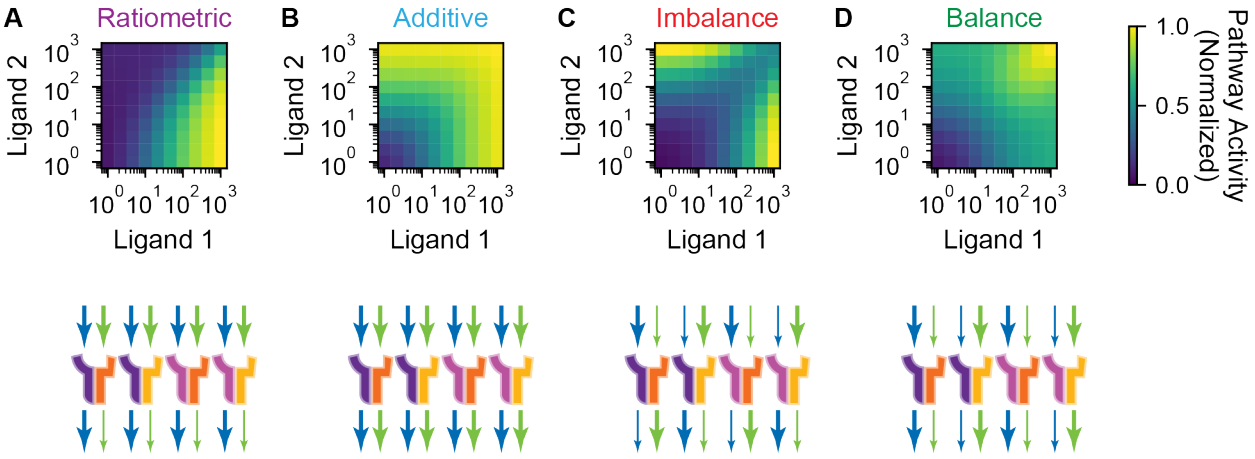

Figure 3.S1: Promiscuous ligand-receptor interactions generate a repertoire of archetypal response functions.

Four archetypal response functions—ratiometric, additive, imbalance, and balance—appeared in a more complex model of the BMP pathway (Antebi et al., 2017). Here, we show that similar archetypes appear in the model used here (top) and illustrate the model parameters that generate them (bottom). Parameter diagrams represent the affinity (top arrows) and activity (bottom arrows) parameters associated with each signaling complex. Arrow width indicates relative magnitude. Thin arrows correspond to values of 0.1, while thick arrows represent values of 1.

(A) In ratiometric responses, one ligand reduces the activity of the other, such that the overall response approximates the ratio of the two concentrations. Such responses can arise through competitive inhibition, where a second ligand binds the receptors that are needed to generate signaling activity but produces inactive signaling complexes.

(B) Additive responses approximate the sum of the two ligand concentrations, as the ligands increase pathway activity either alone or together. Ligands that activate receptors equivalently can generate such responses.

(C) In imbalance detection, the pathway is most active when there is a large imbalance in the levels of the two ligands. These responses can arise if, for instance, competition between two ligands favors complexes with low signaling activity.

(D) Balance detection responses show most activity when both ligands are present simultaneously at a particular ratio. One mode for generating them is when ligand binding favors formation of high-activity signaling complexes.

Related to Figure 3.2.

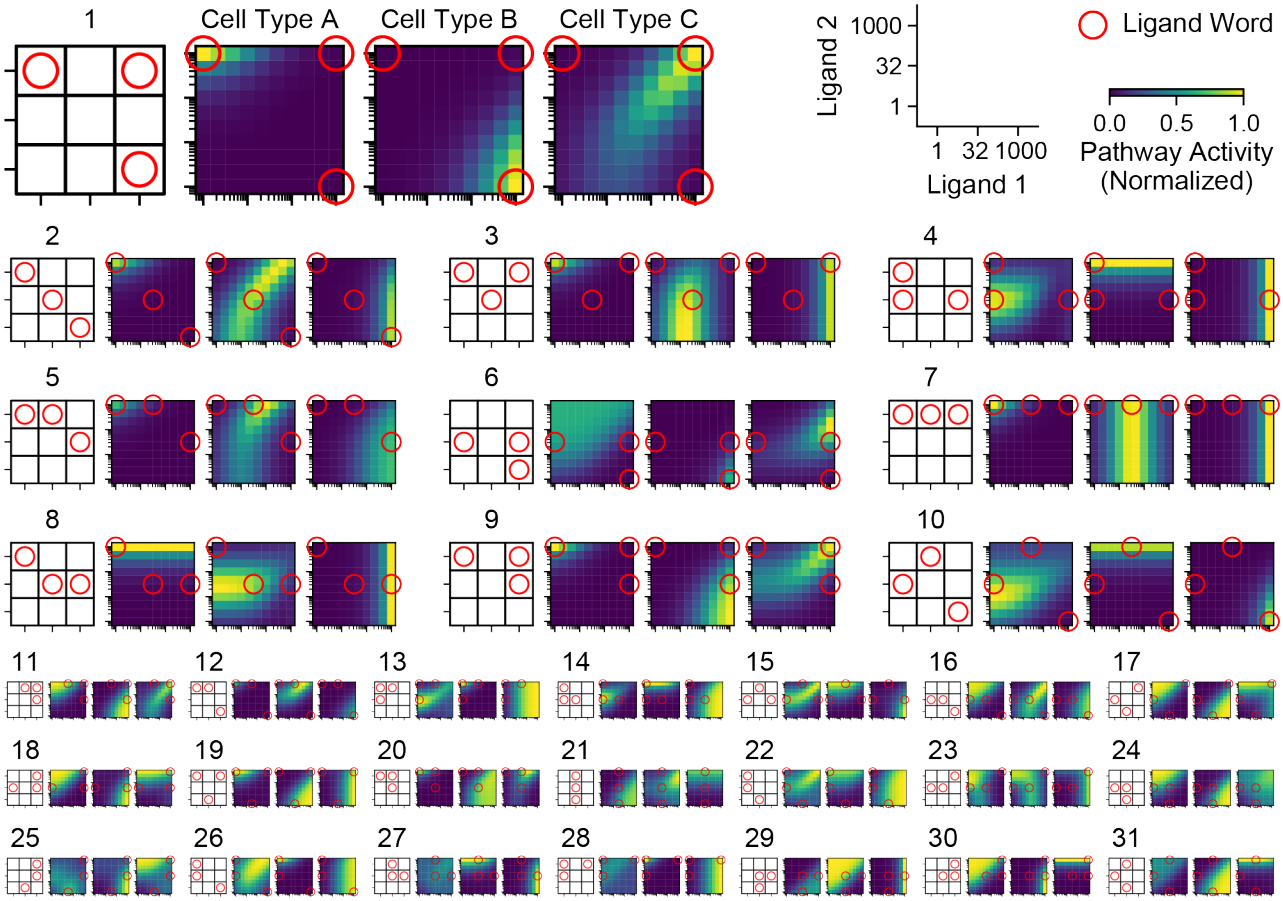

Figure 3.S2: Orthogonal addressing can arise from a variety of different response types.

For the parameter sets represented in Figure 3.2D, the responses of each cell type are shown. Parameter sets are ordered by distinguishability, from best to worst. These responses illustrate that three-channel addressing can be achieved in a variety of ways, although common patterns do emerge (for example, schemes 1–2 and schemes 3–4). Different scales are used to focus on the strongest examples.

Related to Figure 3.2.

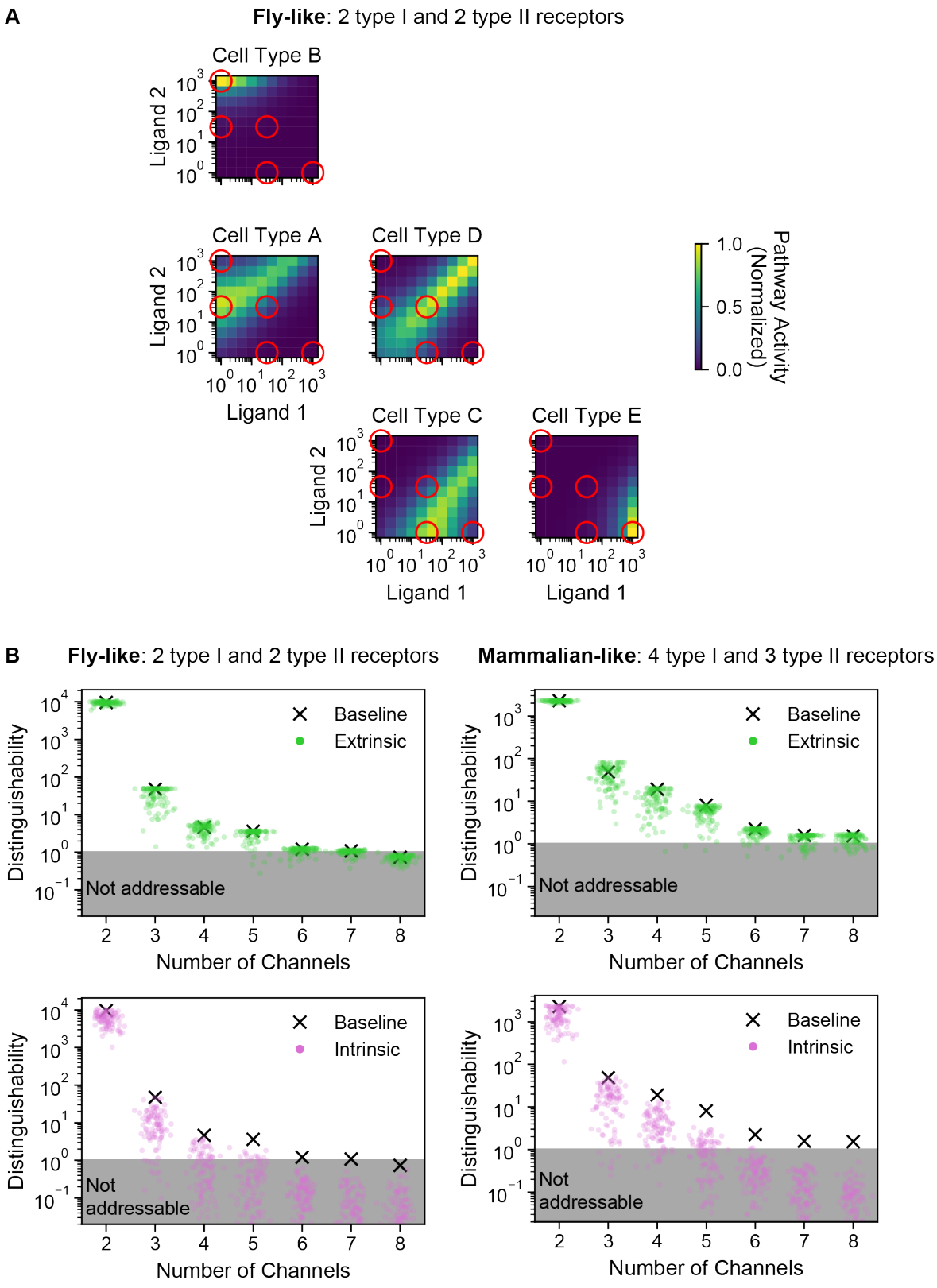

Figure 3.S3: Orthogonal addressing schemes are robust to extrinsic noise in receptor expression levels. Continued on next page.

Figure 3.S3: Orthogonal addressing schemes are robust to extrinsic noise in receptor expression levels. Continued from previous page.

(A) The responses of each cell type in the five-channel system analyzed in Figure 3.3D are shown. As in Figure 3.3I, ligand words corresponding to orthogonally activating channels are shown as red circles. Responses have been rearranged such that the response of a given cell type is shown in the relative position of its orthogonally activating ligand word. For example, the top left cell type is activated by high levels of ligand 2 only, while the bottom right cell type is orthogonally activated by high levels of ligand 1 only.

(B) Top parameter sets from (left) the fly-like model (two type I and two type II receptor variants) of Figures 3.3A–C and (right) the mammalian-like system (four type I and three type II receptor variants) of Figures 3.3E–G were evaluated for addressing specificity in the presence of noise. Receptor expression levels were perturbed with correlated extrinsic (top, green) or uncorrelated intrinsic (bottom, purple) noise (Methods 3.6.7), and distinguishability values were computed with all other parameters held constant. For each condition, the results of 100 perturbations are shown, along with the baseline value (black crosses).

Related to Figure 3.3.

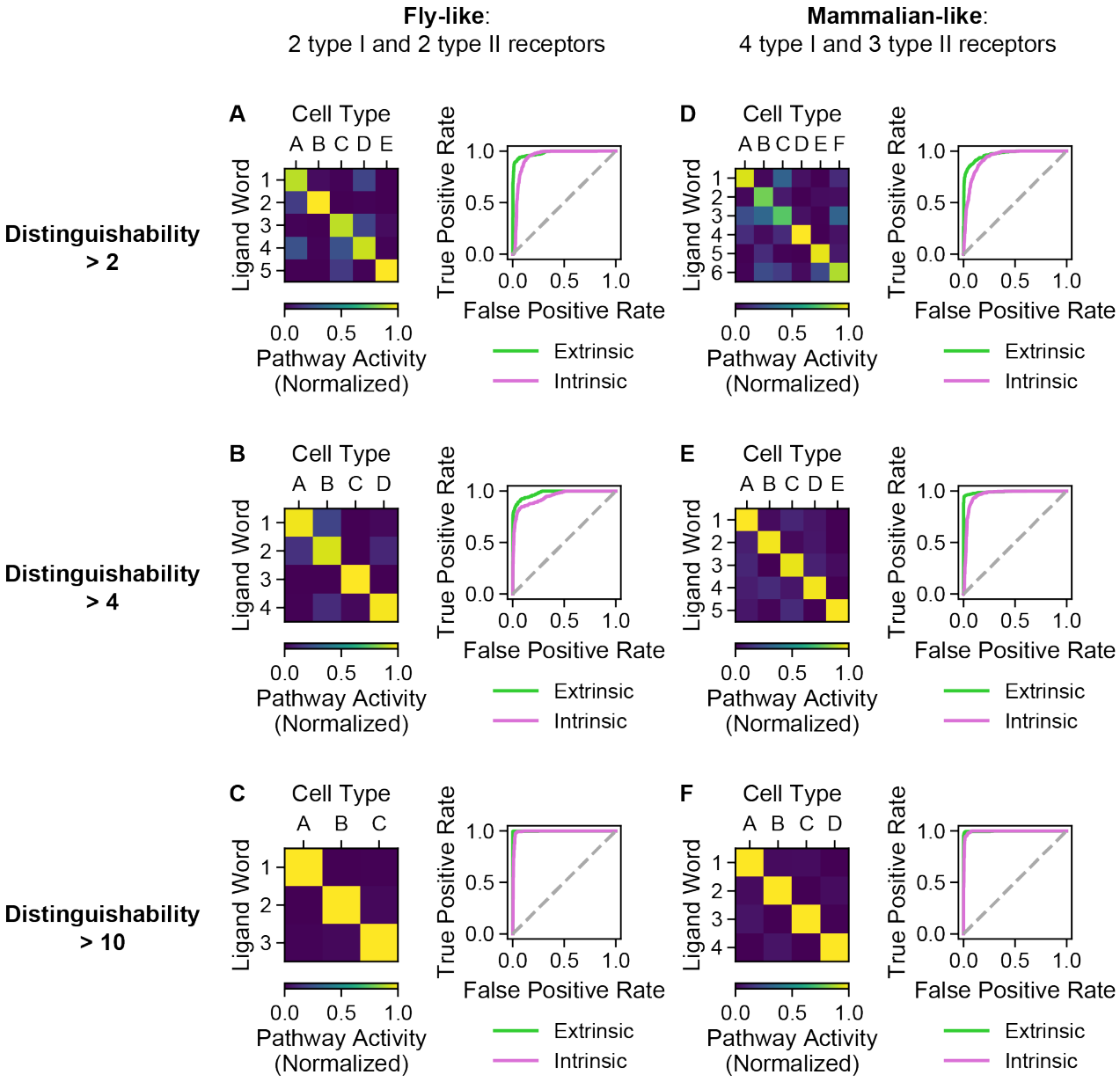

Figure 3.S4: Varying distinguishability thresholds reveal a tradeoff between the capacity and robustness of addressing systems. Continued on next page.

Figure 3.S4: Varying distinguishability thresholds reveal a tradeoff between the capacity and robustness of addressing systems. Continued from previous page.

To understand the tradeoff between addressing capacity and robustness, we show the system of highest bandwidth that exceeds the indicated distinguishability thresholds. Specifically, we highlight the crosstalk matrix as well as the ROC curves in response to extrinsic (green) or intrinsic (purple) noise.

(A) In the fly-like model, five channels can be orthogonally addressed at a distinguishability of at least 2.

(B) In the fly-like model, four channels can be orthogonally addressed at a distinguishability of at least 4.

(C) In the fly-like model, three channels can be orthogonally addressed at a distinguishability of at least 10.

(D) In the mammalian-like model, six channels can be orthogonally addressed at a distinguishability of at least 2.

(E) In the mammalian-like model, five channels can be orthogonally addressed at a distinguishability of at least 4.

(F) In the mammalian-like model, four channels can be orthogonally addressed at a distinguishability of at least 10.

Related to Figure 3.3.

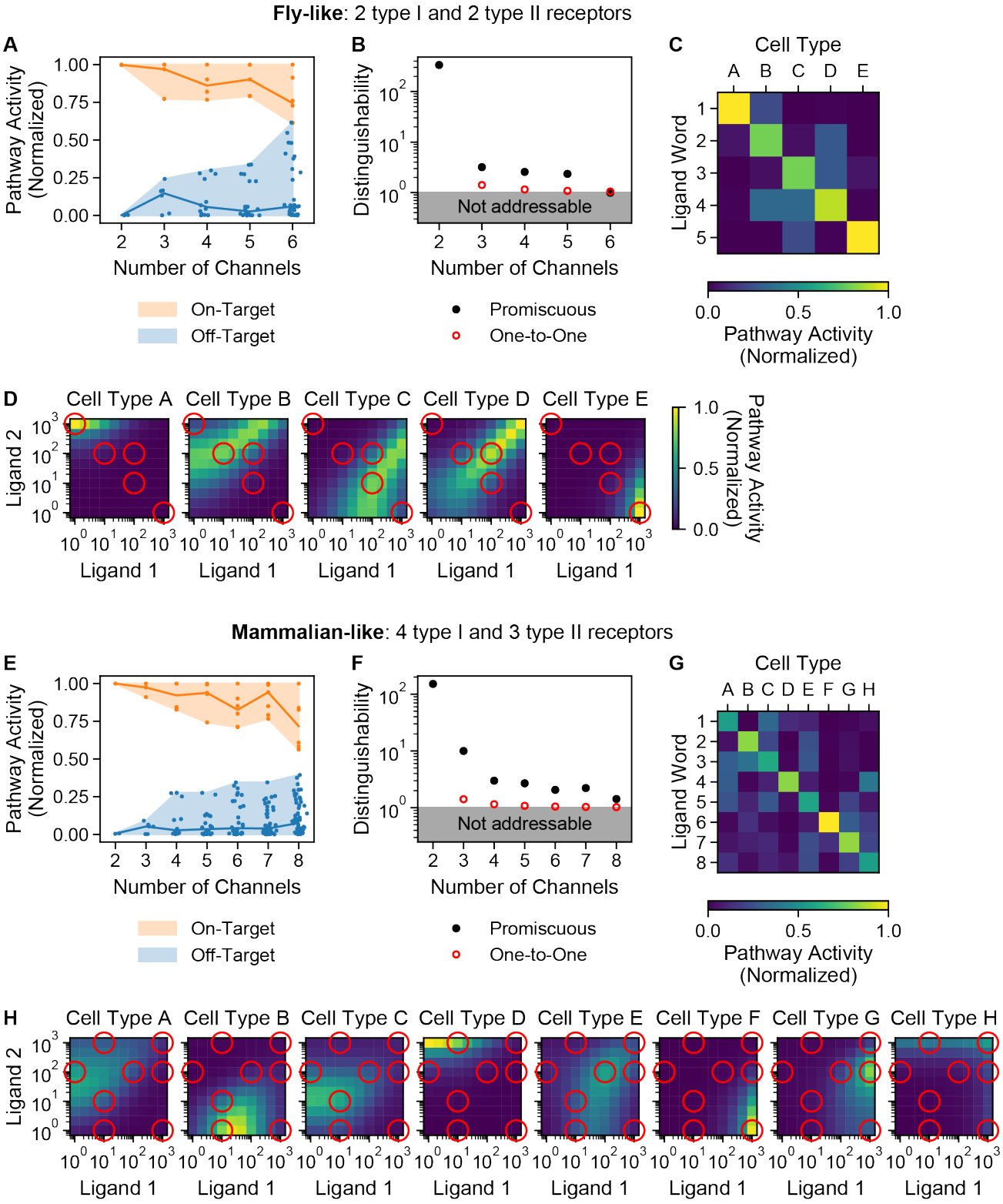

Figure 3.S5: Orthogonal addressing capacity remains generally consistent when considering higher-resolution ligand discretization. Continued on next page.

Figure 3.S5: Orthogonal addressing capacity remains generally consistent when considering higher-resolution ligand discretization. Continued from previous page.

(A) We sought to optimize for orthogonal addressing systems using a four-level ligand grid, compared to the three concentration levels used throughout the rest of this work. We optimized randomly generated addressing schemes for each given number of channels, moving to a higher bandwidth once a parameter set had been successfully optimized. On-target and off-target responses are shown for the best parameter set for each number of channels in a fly-like model, as in Figure 3.3A.

(B) Distinguishability values are plotted for each number of channels, as in Figure 3.3B. The five-channel system, which reflects the highest addressable bandwidth and has a distinguishability of 2.3, is further analyzed in (C–D).

(C) The crosstalk matrix is shown for the five-channel system, as in Figure 3.3D.

(D) The full responses of each cell type are shown for the five-channel system. Red circles correspond to the five ligand words.

(E) The pathway activities for the top parameter set of each tested bandwidth are shown for a mammalian-like model with four type I and three type II receptors, as in (A).

(F) Distinguishability values are plotted for the mammalian-like model, as in (B). The eight-channel system, which reflects the highest addressable bandwidth and has a distinguishability of 1.4, is further analyzed in (G–H).

(G) The crosstalk matrix for the example in the mammalian-like model is shown, as in (C).

(H) The full responses for the example in the mammalian-like model are shown, as in (D).

Related to Figure 3.3.

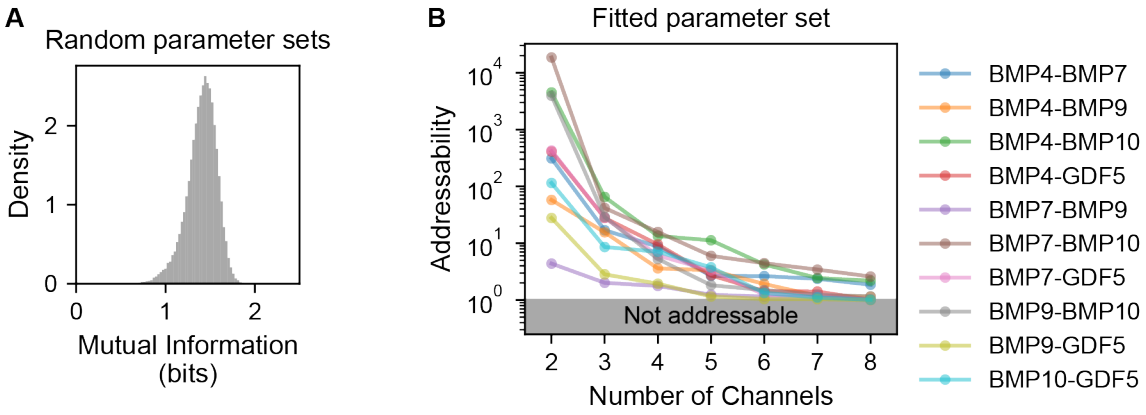

Figure 3.S6: Addressing properties vary across parameter sets.

(A) The approach of Figures 3.6A–B was applied to an equivalent number of randomly generated parameter sets, rather than a grid of parameter values. The resulting distribution of mutual information values is similar, indicating that the result is robust to the method of parameter sampling (cf. Figure 3.6B).

(B) Parameters for five ligands (indicated in legend), two type I receptors, and three type II receptors were fitted to the experimental measurements of BMP responses in multiple cell lines with differing receptor expression profiles described in (Klumpe et al., 2022) and analyzed for their addressing potential. Specifically, we computed the addressability for every pair of ligands using the same libraries of ligand combinations and cell types as in (A). Different pairs show varying levels of addressing potential, with some pairs (e.g., BMP4-BMP7, BMP4-BMP10, and BMP7-BMP10) exhibiting addressing of different cell type groups for every ligand combination and others (e.g., BMP7-BMP9) showing lower bandwidths.

Related to Figure 3.6.

Table 3.S1: Experimentally analyzed cell lines have receptor expression profiles resembling those of biological cell types.

Cell lines analyzed in Figure 3.5 were compared to cell types from annotated single-cell RNA-seq expression atlases (Tabula Muris Consortium, 2020; Tabula Muris Consortium et al., 2018). Receptor profiles for NMuMG cells were measured directly through bulk RNA-seq, while receptor profiles for perturbed NMuMG lines were simulated by estimating the perturbation magnitudes from qPCR data (Klumpe et al., 2022). For each cell line, a cell type with similar receptor expression profile is identified, along with the relevant developmental timepoint or age.

Related to Figure 3.5.

Cell Line	**Similar Cell Type**
NMuMG	Tongue keratinocyte (24 months)
ACVR1 KD	Tongue basal cell of epidermis (3 months)
BMPR2 KD	Tongue keratinocyte (18 months)
ACVRL1 OX	Limb muscle Schwann cell (18 months)

3.6 Methods

All data are publicly accessible at the CaltechDATA research data repository (`https://doi.org/10.22002/D1.1692`). All code is publicly accessible at GitHub (`https://github.com/christinasu/PromiSys`) as well as at the CaltechDATA research data repository (`https://doi.org/10.22002/D1.20047`).

3.6.1 One-Step Model for Promiscuous Interactions

Ligand-Receptor Interactions

Many signaling pathways demonstrate promiscuous interactions between multiple ligand and receptor variants, which can bind with varying affinities to form many distinct signaling complexes. The BMP pathway represents a canonical example of such an architecture. Previously, we have described a mathematical model that captures key features of this pathway and recapitulates experimentally observed responses (Antebi et al., 2017). Here, we develop a simplified version of the model that captures equivalent behaviors at steady state while reducing the number of parameters to be considered.

In the model, we describe binding of a ligand to a heterodimer of type I and type II receptors. Specifically, we consider n_L ligand variants, n_A type I or A receptor variants, and n_B type II or B receptor variants, where ligand L_i can interact with A receptor A_j and B receptor B_k to form the heterotrimeric signaling complex T_{ijk}. We assume that this process occurs as a one-step reaction with an effective three-way interaction, with forward rate $k_{f_{ijk}}$ and reverse rate $k_{r_{ijk}}$. This reaction can be summarized as

$$L_i + A_j + B_k \underset{k_{r_{ijk}}}{\overset{k_{f_{ijk}}}{\rightleftharpoons}} T_{ijk} \tag{3.1}$$

Differential Equations and Constraints

Letting L_i denote the concentration of ligand in a volume V and letting A_j, B_k, and T_{ijk} denote the absolute numbers of receptors and complexes on the cell surface, we can then write the differential equations that describe the dynamics of these reactions:

$$\frac{dL_i}{dt} = \frac{1}{V}\sum_{j=1}^{n_A}\sum_{k=1}^{n_B}\left(-k_{f_{ijk}}L_iA_jB_k + k_{r_{ijk}}T_{ijk}\right) \tag{3.2}$$

$$\frac{dA_j}{dt} = \sum_{i=1}^{n_L}\sum_{k=1}^{n_B}\left(-k_{f_{ijk}}L_iA_jB_k + k_{r_{ijk}}T_{ijk}\right) \tag{3.3}$$

$$\frac{dB_k}{dt} = \sum_{i=1}^{n_L}\sum_{j=1}^{n_A}\left(-k_{f_{ijk}}L_iA_jB_k + k_{r_{ijk}}T_{ijk}\right) \tag{3.4}$$

$$\frac{dT_{ijk}}{dt} = k_{f_{ijk}}L_iA_jB_k - k_{r_{ijk}}T_{ijk} \tag{3.5}$$

Each complex T_{ijk} phosphorylates the second messenger at some rate ε_{ijk} to generate intracellular signal S, which degrades at rate γ. The rate of change of the total signal is given by the following differential equation:

$$\frac{dS}{dt} = \sum_{i=1}^{n_L}\sum_{j=1}^{n_A}\sum_{k=1}^{n_B}\varepsilon_{ijk}T_{ijk} - \gamma S \tag{3.6}$$

We assume that the volume for the ligands is large, or $V \to \infty$. In this regime, there are significantly more ligand molecules than receptors, as is the case for experimental conditions in which ligands are dissolved in an excess of media. Under this assumption, ligand concentrations remain constant. We further assume that production and consumption of the various molecular species are in steady state. By conservation of mass, the total number of each type of molecule, alone or in complex with other species, must remain constant. Letting L_i^0, A_j^0, and B_k^0 denote the initial values of the respective species, we obtain the following constraints:

$$L_i^0 = L_i \tag{3.7}$$

$$A_j^0 = A_j + \sum_{i=1}^{n_L} \sum_{k=1}^{n_B} T_{ijk} \tag{3.8}$$

$$B_k^0 = B_k + \sum_{i=1}^{n_L} \sum_{j=1}^{n_A} T_{ijk} \tag{3.9}$$

Steady-State Equations

Since binding and unbinding of ligands and receptors occur on fast time scales relative to the time scales of reporter detection, we focus on characterizing the behavior of this system at steady state. Here, all time derivatives in Equations Equations 3.2–3.6 vanish. Defining affinities $K_{ijk} \equiv k_{f_{ijk}}/k_{r_{ijk}}$ and activities $e_{ijk} \equiv \varepsilon_{ijk}/\gamma$, Equations 3.5–3.6 can be solved as follows:

$$T_{ijk} = K_{ijk} L_i A_j B_k \tag{3.10}$$

$$S = \sum_{i=1}^{n_L} \sum_{j=1}^{n_A} \sum_{k=1}^{n_B} e_{ijk} T_{ijk} \tag{3.11}$$

Together, Equations 3.7–3.11 describe the behavior of the model at steady state.

$$\begin{aligned}
L_i^0 &= L_i \\
A_j^0 &= A_j + \sum_{i=1}^{n_L} \sum_{k=1}^{n_B} T_{ijk} \\
B_k^0 &= B_k + \sum_{i=1}^{n_L} \sum_{j=1}^{n_A} T_{ijk} \\
T_{ijk} &= K_{ijk} L_i A_j B_k \\
S &= \sum_{i=1}^{n_L} \sum_{j=1}^{n_A} \sum_{k=1}^{n_B} e_{ijk} T_{ijk}
\end{aligned}$$

We can solve this system of equations to find the values of T_{ijk} at steady state, which we can then use to compute the total signaling response S. By rearranging

Equations 3.8–3.9, the steady-state values of the receptors can be derived as the following expressions:

$$A_j = A_j^0 - \sum_{i=1}^{n_L} \sum_{k=1}^{n_B} T_{ijk} \tag{3.12}$$

$$B_k = B_k^0 - \sum_{i=1}^{n_L} \sum_{j=1}^{n_A} T_{ijk} \tag{3.13}$$

Substituting into Equation 3.10, we have a system of $n_T = n_L n_A n_B$ quadratic equations for T_{ijk}:

$$T_{ijk} = K_{ijk} L_i \left(A_j^0 - \sum_{i'=1}^{n_L} \sum_{k'=1}^{n_B} T_{i'jk'} \right) \left(B_k^0 - \sum_{i'=1}^{n_L} \sum_{j'=1}^{n_A} T_{i'j'k} \right) \tag{3.14}$$

The solutions for T_{ijk} from this system of equations can then be substituted into Equation 3.11 to compute the total signal S.

To solve the model efficiently, we used Equilibrium Toolkit (EQTK), an optimized Python-based numerical solver for biochemical reaction systems (Bois, 2020). EQTK casts the coupled equilibrium problem as an unconstrained convex dual optimization problem and employs a globally convergent trust region algorithm to solve it (Bois, 2020; Dirks et al., 2007). This method accelerated computation by approximately 600-fold compared to standard nonlinear least-squares optimization used previously (Antebi et al., 2017).

3.6.2 Comparison with Alternative Models

Promiscuous vs. One-to-One Model

We compare a promiscuous signaling architecture to a simple model for signaling with one-to-one ligand-receptor interactions, where each ligand variant L_i binds with a single cognate receptor R_i to form an active dimer D_i. We can derive the equations describing this system analogous to the analysis done above for the promiscuous architecture. Assuming that this binding has forward rate k_{f_i} and reverse rate k_{r_i},

the chemical reactions can be expressed as follows:

$$L_i + R_i \underset{k_{r_i}}{\overset{k_{f_i}}{\rightleftharpoons}} D_i \tag{3.15}$$

We measure L_i as the concentration of ligand in a volume V and R_i and D_i as the absolute numbers of receptors or complexes on the cell surface. The differential equations describing their dynamics are then as below:

$$\frac{dL_i}{dt} = \frac{1}{V}\left(-k_{f_i} L_i R_i + k_{r_i} D_i\right) \tag{3.16}$$

$$\frac{dR_i}{dt} = -k_{f_i} L_i R_i + k_{r_i} D_i \tag{3.17}$$

$$\frac{dD_i}{dt} = k_{f_i} L_i R_i - k_{r_i} D_i \tag{3.18}$$

We again assume that the volume for ligands is large, such that there are significantly more ligand molecules than receptors. This assumption holds for our experimental setting, where ligands are dissolved in an excess of media. Ligand concentrations thus remain constant. We enforce conservation of mass for the receptor subunits. Finally, we define affinities $K_i \equiv k_{f_i}/k_{r_i}$ and activities $e_i \equiv \varepsilon_i/\gamma$ as before. At steady state, we have the following equations to describe the behavior of a model with n ligands and receptors:

$$L_i^0 = L_i \tag{3.19}$$

$$R_i^0 = R_i + D_i \tag{3.20}$$

$$D_i = K_i L_i R_i \tag{3.21}$$

$$S = \sum_{i=1}^{n} e_i D_i \tag{3.22}$$

The steady-state solutions for D_i can then be derived as

$$\begin{aligned} D_i &= K_i L_i^0 \left(R_i^0 - D_i\right) \\ &= \frac{K_i L_i^0 R_i^0}{1 + K_i L_i^0} \end{aligned} \tag{3.23}$$

One-Step vs. Two-Step Model

We have previously considered a mathematical model that describes the promiscuous architecture of the BMP pathway, which considers formation of the heterotrimeric complexes T_{ijk} in a two-step process (Antebi et al., 2017). Briefly, ligand L_i and receptor A_j form an intermediate dimer D_{ij}, which then binds to receptor B_k to form trimer T_{ijk}. Again, we assume that the reactions are reversible and follow first-order kinetics, with forward and reverse reaction rates $k^D_{f_{ij}}$ and $k^D_{r_{ij}}$ for formation of the dimers and $k^T_{f_{ijk}}$ and $k^T_{r_{ijk}}$ for formation of the trimers. Defining $K^D_{ij} \equiv k^D_{f_{ij}}/k^D_{r_{ij}}$ and $K^T_{ijk} \equiv k^T_{f_{ijk}}/k^T_{r_{ijk}}$, the steady-state solutions for T_{ijk} in the two-step model, analogous to Equation 3.14 in the one-step model, are as follows:

$$T_{ijk} = K^T_{ijk} K^D_{ij} L_i \left(\frac{A^0_j - \sum_{i'=1}^{n_L} \sum_{k'=1}^{n_B} T_{i'jk'}}{1 + \sum_{i'=1}^{n_L} K^D_{i'j} L_{i'}} \right) \left(B^0_k - \sum_{i'=1}^{n_L} \sum_{j'=1}^{n_A} T_{i'j'k} \right) \tag{3.24}$$

Comparing Equation 3.14 and Equation 3.24, the steady-state solutions for T_{ijk} in the two-step model can be mapped to the one-step model under the following parameter choice:

$$K_{ijk} = \frac{K^T_{ijk} K^D_{ij}}{1 + \sum_{i'=1}^{n_L} K^D_{i'j} L_{i'}} \tag{3.25}$$

Since S is defined by the values of T_{ijk} and is given by Equation 3.11 in both the one-step and two-step models, the steady-state behavior of the two-step model with any set of parameters can also be represented in the one-step model. However, the number of parameters is reduced from $N_p^{\text{two-step}} = n_A + n_B + n_L n_A + 2 n_L n_A n_B$ to $N_p^{\text{one-step}} = n_A + n_B + 2 n_L n_A n_B$. Thus, the one-step model enables us to simplify the system while preserving all possible behaviors of T_{ijk} and S at steady state.

Trimeric vs. Hexameric Model

We have developed a simplified model in which a ligand binds to type I and type II receptor subunits to form a trimeric signaling complex. However, the BMP signaling

pathway is known to involve hexameric signaling complexes, where a dimeric ligand interacts with two type I and two type II receptors. This model captures reactions of the following form:

$$L_i^1 + L_j^2 + A_k^1 + A_l^2 + B_m^1 + B_n^2 \underset{k_{r_{ijklmn}}}{\overset{k_{f_{ijklmn}}}{\rightleftharpoons}} H_{ijklmn} \tag{3.26}$$

This model can essentially be reduced to a trimeric model by setting reaction rates to 0 for any reaction with $i \neq j$, $k \neq l$, or $m \neq n$. As such, responses in the trimeric model represent a subset of the functions that could be possible in the hexameric model.

3.6.3 Optimization of Orthogonal Addressing Schemes

Given a target orthogonal addressing scheme of N channels, we optimized for parameters that would yield matching responses. Specifically, we used constrained least-squares optimization for $n_L n_A n_B$ affinity parameters K_{ijk}, $n_L n_A n_B$ activity parameters e_{ijk}, and $N\,(n_A + n_B)$ receptor expression levels. We bounded affinity and activity parameters in $[0, 1]$ and receptor levels in $[0, \infty)$. We sought to minimize the residuals between the target responses and the simulated responses at the N ligand words of interest. Since the simulated responses have arbitrary units, we normalized all responses for a given parameter set. In particular, we normalized by the maximum value in any cell type over the full ligand titration, not only the ligand words of interest. This normalization ensures that all cell types share a relatively similar level of activation and that the activation in the orthogonal channels is distinguishable from activation by other ligand combinations.

As this optimization procedure is not guaranteed to converge to a global minimum, we optimized repeatedly with different initial conditions. Biochemical parameters were chosen in a uniform random distribution over $[0, 1]$, and receptor levels were initialized to 1. To evaluate the potential capacity of promiscuous ligand-receptor

systems for orthogonal addressing, we sought to optimize progressively higher bandwidths without requiring any particular scheme. For N channels, we randomly selected N ligand words as orthogonally activating inputs and sought to optimize parameters as described. We iterated this process with randomly chosen ligand words until parameters had been identified to generate N channels successfully. Once this criterion was met, we then proceeded to optimize $N + 1$ channels, up to the limit derived from the number of possible ligand words. We performed at least an average of 500 optimizations per bandwidth.

3.6.4 Enumeration of Orthogonal Addressing Schemes

To analyze the possibility for orthogonal addressing, we used a discretized ligand concentration space to enable a comprehensive screening. We reasoned that we could systematically test for all possible orthogonal addressing schemes by selecting a subset of the possible ligand combinations to be orthogonally activating and defining a set of targeted response functions accordingly. For a set of N chosen ligand words, we enumerated N targeted response functions, where each ligand word activates exactly one cell type and, conversely, each cell type is activated by exactly one ligand word. Having discretized ligand concentrations to three levels, there are $3^2 = 9$ possible ligand words. The combination with all ligands at the lowest level is assumed to yield negligible activation in any cell type, so this ligand word cannot generate a channel. Thus, there can be one to eight possible communication channels.

For each possible number of channels N, we took all possible subsets and sought to achieve these addressing schemes. There are $\binom{8}{N}$ possible addressing schemes for a given bandwidth and $2^8 = 256$ possible addressing schemes overall. However, addressing schemes that are identical under changes in ligand labels were removed, leaving 144 total possibilities. We systematically tested for the ability to achieve

each addressing scheme by performing a search over all schemes. At each trial, we chose which addressing scheme to optimize based on both the lowest error E achieved and the number of trials T attempted. Specifically, we optimized for the addressing scheme with maximal value of E/T^2 and repeated this process for at least an average of 50 optimizations per scheme to ensure that all schemes would be tested adequately. This strategy of systematically optimizing every possible addressing scheme rather than randomly selecting schemes of a given bandwidth provided a complementary approach for analyzing addressing capability.

3.6.5 Distinguishability of Channels

We optimized parameters for each addressing scheme based on the squared error between the targeted and simulated responses at each ligand word of interest. However, this error does not necessarily guarantee specificity of addressing, where a given ligand combination should activate only a single cell type and not the others. To quantify the performance of each system, we analyzed the distributions of on-target and off-target activation levels. We defined the distinguishability as the fold difference between the minimum on-target and maximum off-target activities, which measures the ability to differentiate between specific and nonspecific signals in the worst case.

3.6.6 Orthogonal Addressing in One-to-One Model

Lower Bound for Distinguishability

For a one-to-one architecture, we can readily use Equations 3.22–3.23 to calculate the steady-state signal for any set of initial ligand and receptor concentrations under a given set of biochemical parameters. With a two-component system, we obtain the following:

$$S = \frac{K_1 e_1 L_1^0 R_1^0}{1 + K_1 L_1^0} + \frac{K_2 e_2 L_2^0 R_2^0}{1 + K_2 L_2^0} \tag{3.27}$$

Suppose that $K_1 = K_2$ and $K_i \gg L_i^0$ for any i. Define $Q_i = e_i R_i^0$. In this regime, the signal (subject to an arbitrary normalization factor) is approximately as below:

$$S \approx L_i^0 Q_i + L_2^0 Q_2 \tag{3.28}$$

To obtain an addressing system of N orthogonal channels, we must define N ligand words and N cell types. Let $\mathbf{W}_i$ denote the ith word, or a length-2 vector representing a ligand expression profile as $\mathbf{W}_i = \left(L_1^0, L_2^0\right)^{(i)}$. Likewise, let $\mathbf{C}'_j$ represent the jth cell type, or a length-2 vector representing a receptor expression profile as $\mathbf{C}'_j = \left(R_1^0, R_2^0\right)^{(j)}$. For mathematical convenience, we can instead consider $\mathbf{C}_j = (Q_1, Q_2)^{(j)}$, which encompasses receptor expression as well as activity. Letting $\mathbf{S}_{ij}$ denote the steady-state signal for the ith ligand word and the jth cell type, this value is simply $\mathbf{S}_{ij} = \mathbf{W}_i \cdot \mathbf{C}_j$. As such, we wish to maximize $\mathbf{W}_i \cdot \mathbf{C}_i$ and minimize $\mathbf{W}_i \cdot \mathbf{C}_j$ (for $i \neq j$) in order to maximize distinguishability. Since the dot product between any two vectors $\mathbf{a}$ and $\mathbf{b}$ with an angle θ between them is given by $\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\|\|\mathbf{b}\| \cos\theta$, we should choose $\mathbf{W}_i$ and $\mathbf{C}_i$ to be directly proportional to one another (i.e., $\theta = 0$, thus maximizing $\cos\theta$), and we should choose the different $\mathbf{W}_i$ to be equidistantly spaced across the "unit arc" or first quadrant of the unit circle (thus maximizing θ and minimizing $\cos\theta$). Specifically, we can define $\mathbf{W}_i = \mathbf{C}_i = (\cos\theta_i, \sin\theta_i)$ for $\theta_i = \frac{i-1}{N-1} \cdot \frac{\pi}{2}$ radians. Each on-target signal will then simply be $\mathbf{S}_{ii} = 1$, and each off-target signal will be $\cos\left(\frac{\pi}{2(N-1)}\right)$. Thus, we can guarantee implementation of N orthogonal channels with a distinguishability of at least

$$D = \frac{1}{\cos\left(\frac{\pi}{2(N-1)}\right)} \tag{3.29}$$

Optimization

The above analysis guarantees implementation of arbitrarily many orthogonal channels with distinguishability greater than 1. However, it makes several assumptions. Therefore, we also sought to identify parameters that could improve upon the class

of solutions derived above in the absence of such simplifications. We used an optimization approach, optimizing for distinguishability and searching over a single set of affinity and activity parameters as well as N sets of ligand and receptor expression values. For the same bandwidths as considered for the promiscuous system, we performed 100 optimization trials each. In all cases, the resulting distinguishability values approached but did not exceed the theoretical value. Thus, our theoretical solutions, while not proven to be optimal, are able to outperform computationally optimized parameters.

3.6.7 Analysis of Robustness

Biological systems are subject to noise. In particular, cellular systems show both extrinsic noise, or correlated changes such as during cell growth or changes in expression machinery, and intrinsic noise, or independent stochastic variation in each element. To assess whether the optimized parameters are robust to noise in receptor expression levels, we evaluated whether on-target and off-target signals could be correctly distinguished across many random perturbations, using the receiver operating characteristic (ROC). In particular, we computed the area under the ROC curve (AUC), which represents the probability of successfully classifying on-target from off-target activations. We considered both purely extrinsic and purely intrinsic noise. For a given coefficient of variation (CV) ν (here, $\nu = 0.5$), we simulated extrinsic noise by generating a scale factor from a gamma distribution with shape parameter $1/\nu^2$ and scale parameter ν^2 (giving a mean of 1 and a variance of ν^2) and multiplied all receptor levels by this scale factor. For intrinsic noise, we instead drew scale factors i.i.d. for each receptor. With each form of noise, we generated 100 random perturbations, simulated the resulting activity levels, and computed the corresponding ROC and AUC (Figures 3.3C,G). We also plotted the distinguishability values for these perturbations (Figure 3.S3B).

3.6.8 Optimization of Subset Addressing Repertoires

We next considered more general addressing systems, targeting activation not just of individual cell types but also of groups of cell types. Specifically, an addressing repertoire encompasses all subsets of cell types that can be co-activated by any ligand word across a complete titration of ligand concentrations. For instance, titrating two ligand variants with three concentrations yields $3^2 = 9$ ligand words, each of which activates some subset of the cell types considered. Every distinct group of cell types constitutes an achievable channel. The set of channels resulting from the 9 ligand words considered constitutes the addressing repertoire for that set of parameters.

To characterize the specificity of addressing different subsets of cell types, we generalized the distinguishability metric defined above. Each ligand word activates a particular subset of cell types; the corresponding response(s) of the cell type(s) would represent on-target signaling, while the response(s) of any other cell type(s) would represent off-target signaling. Therefore, as in the case of orthogonal addressing, distinguishability can be calculated as the fold difference between the minimum on-target activity and the maximum off-target activity.

Similar to a specific number of orthogonal channels, a given addressing repertoire can potentially be implemented in many ways. In other words, many different sets of responses can generate the same addressing repertoire. Unlike the orthogonal case, however, the responses for all ligand combinations are relevant. Thus, enumerating the ways to achieve a given addressing repertoire requires considering any possible response for every cell type.

To generalize our optimization approach to analyze addressing repertoires, we first set out to define what sets of responses could yield a given repertoire. Therefore, we started by enumerating all possible binary response matrices for a single cell type. The number of possible responses then reduces to the number of ways to choose

"on" signals. Assuming that cells are always inactive for the ligand combination where both ligands are present at low levels and ignoring the case where the cell is entirely nonresponsive, there are $2^8 - 1 = 255$ possible responses.

By considering all combinations of three responses from this set, we were able to map all addressing repertoires to the potential sets of three responses. Due to the large number of possibilities for a given repertoire, we sought to prioritize sets of responses that were more likely to be achievable. Therefore, we individually optimized each of the 255 responses and quantified the quality of each response using the sum of squared distances to the target, after normalizing the simulated response to have a maximum response of 1 (Figure 2.6). We ranked sets of three responses based on the sum of the scores of each response individually. Since parameter sets were individually optimized, a response that can be achieved with high quality independently may not be possible in the same biochemical parameter regime as another; however, this scoring should reduce consideration of responses that are challenging to optimize individually, let alone together with others.

Having selected candidate sets of responses, we could then perform least-squares optimization as done previously. We also complemented this optimization approach by reasoning that any given set of responses matches some addressing repertoire, depending only on how the threshold between "inactive" and "active" pathway response is defined. Therefore, we simulated a random set of responses, chose the threshold that yielded the greatest distinguishability between the lowest on-target and highest off-target responses, and associated those parameters with the resulting addressing repertoire. We iteratively optimized for this distinguishability, stopping if the resulting addressing repertoire was one for which a valid parameter set had not yet been identified.

3.6.9 Enumeration of Subset Addressing Repertoires

We focused on analyzing addressing repertoires for three cell types, denoted A, B, and C. With three cell types, there are eight possible channels: one with no cell types activated, three with a single cell type activated, three with two cell types activated, and one with all cell types activated. Since the channel with no cell types activated is always achieved in the absence of any ligand, we neglect this from further consideration. Each of the remaining seven channels may or may not be present, for a total of up to $2^7 = 128$ addressing repertoires. We discard repertoires that are redundant with respect to relabeling of ligands as well as repertoires in which two cell types are indistinguishable by any ligand combination (such that the addressing repertoire could be mapped to a repertoire for two cell types). As discussed in more detail below, these simplifications leave us with 32 addressing repertoires of three cell types, corresponding to those shown in Figure 3.4C.

We first seek to eliminate repertoires that are invariant with respect to relabeling of ligands. The subset with all three cell types activated, or triple, does not change when ligands are relabeled; however, singles or doubles may. For example, the addressing repertoire consisting of "A" and "BC" is equivalent to that comprising "B" and "AC," simply by swapping the labels of cell types A and B. As such, we consider the unique ways to include singles or doubles. Consider each single with its complementary double (for example, "A" with "BC"). There are $2^2 = 4$ possible ways to include this pair in an addressing repertoire: both absent, only single present, only double present, and both present. The three pairs can then encompass three distinct choices of these four possibilities (4 combinations), two distinct choices ($4 \cdot 3 = 12$ combinations), or the same choice (4 possibilities). There are then $4 + 12 + 4 = 20$ ways to choose combinations of singles or doubles, and the triple may be either present or absent. Thus, considering ligand relabeling reduces the total number of addressing repertoires to consider to 40.

Note, however, that some of these repertoires may only represent two distinct cell types, rather than three. For example, the addressing repertoire with channels "A" and "BC" indicates that cell types B and C are indistinguishable across all ligand combinations and are therefore functionally equivalent. Let B and C be indistinguishable, without loss of generality. The only possible channels are then "A," "BC," and "ABC." Thus, the $2^3 = 8$ repertoires that only contain these channels can be reduced to two distinct cell types and are therefore omitted from our analysis of addressing three cell types. This correction yields our final set of 32 addressing repertoires.

3.6.10 Subset Addressing in One-to-One Model

To understand how subset addressing in a promiscuous pathway compares with that in a one-to-one architecture, we note that all responses in a one-to-one pathway must be monotonic, meaning that responses never decrease with added ligand. As such, a given cell's response is maximal when exposed to the ligand combination where all ligands are present at highest concentration. Therefore, every cell type will be active in response to this ligand combination. (Otherwise, there would be no response across the entire ligand titration, and there would be no addressing.) Consequently, the subset "ABC" will always be addressable in the one-to-one architecture. Conversely, any repertoire where "ABC" is absent cannot be achieved in the one-to-one architecture.

3.6.11 Addressing of Cell Lines

Cell Lines

NAMRU mouse mammary gland (NMuMG) cells (female) were acquired from ATCC (CRL-1636). Mouse embryonic stem cells (mESCs; E14Tg2a.4, male) were obtained from the laboratory of Bill Skarnes and Peri Tate. Reporter cell

lines were cultured as in (Antebi et al., 2017). Briefly, cells were cultured in a humidity-controlled chamber at 37°C with 5% CO_2. NMuMG cells were cultured in DMEM supplemented with 10% FBS (Clontech #631367), 1 mM sodium pyruvate, 1 unit/mL penicillin, 1 µg/mL streptomycin, 2 mM L-glutamine, and 1× MEM nonessential amino acids. mESCs were plated on tissue culture plates pre-coated with 0.1% gelatin and cultured using DMEM supplemented with 15% FBS (Gibco #16141), 1 mM sodium pyruvate, 1 unit/mL penicillin, 1 µg/mL streptomycin, 2 mM L-glutamine, 1× MEM nonessential amino acids, 55 mM β-mercaptoethanol, and 1,000 units/mL leukemia inhibitory factor (LIF).

Receptor Knockdown and Ectopic Expression

To analyze the potential for addressing in living cells, we engineered cell lines with differing receptor profiles. Using an NMuMG reporter line, individual BMP receptors were knocked down (KD) or overexpressed (OX), as described in (Klumpe et al., 2022). Briefly, receptor knockdown lines were engineered by transducing lentiviral particles containing constructs for constitutive shRNA expression reported by mCherry with a puromycin resistance gene (SMARTvector, Dharmacon). ACVR1 KD cells were a clonal population selected by limiting dilution of cells transduced with a pool of three shRNAs, while BMPR2 KD cells were a polyclonal population generated by a single shRNA. Cells were selected 48 hours after transduction and continuously maintained in 3 µg/mL puromycin. For ectopic expression of BMP receptors, a construct for constitutive expression of mouse receptor cDNA reported by mTurquoise and co-expressed with a geneticin resistance gene was integrated by PiggyBac integration (System Biosciences) using previously reported plasmids (Antebi et al., 2017). ACVRL1 OX cells were selected and maintained in 500 µg/mL geneticin.

BMP Response and Flow Cytometry

Responses of cell lines to BMP ligands were quantified by flow cytometry as described in (Klumpe et al., 2022). Recombinant BMP ligands were acquired from R&D Systems (BMP2, catalog #355-BM; BMP9, catalog #5566-BP; BMP10, catalog #6038-BP). Cells were plated at 20–30% confluency in 96-well plates and cultured under standard conditions for 12 hours. Media was then replaced, and ligands were added at specified concentrations. 24 hours after ligand addition, cells were prepared for flow cytometry by washing with PBS and lifting from the plate using trypsin (NMuMG) or Accutase (mESC) for 5 minutes at 37°C. Protease activity was quenched by resuspending the cells in HBSS with 2.5 mg/mL bovine serum albumin (BSA). Cells were then filtered with a 40 µm mesh and analyzed by flow cytometry (MACSQuant VYB, Miltenyi Biotec; CytoFLEX, Beckman Coulter).

Single-cell flow cytometry data were analyzed by taking the population median. For measured experimental responses (Figure 3.5), responses were quantified by taking the mean of at least 3 repeats. Fold change was measured compared to response with no ligand present and then normalized by the maximum fold change for each cell type.

3.6.12 Computation of Mutual Information

We use mutual information between ligand words and activation patterns across a library of cell types to quantify the combinatorial addressing power of the ligand-receptor system. Mutual information was initially developed to quantify the capacity of a noisy channel to transmit information, or the extent to which distinct input messages can be resolved by the receiver after passing through the channel. Here, we view the ligand words as input messages and the resulting activation pattern across cell types as the received message. Then, the communication system's capacity is determined by the biochemical constants K_{ijk} and e_{ijk}.

One important benefit of using an information theoretic framework is that we do not need to assume a particular set of ligand words and cell types and then optimize over them. Instead, we can use extensive libraries of input ligand words and cell types; mutual information will reflect the best subset of each with no penalty (or benefit) for redundancies. Thus, in our framework, mutual information reflects a property of the biochemical constants K and e alone; ligand inputs and cell types are implicitly assumed to be optimally chosen. (In information theoretic language, we do not need to know optimal error-correcting codes to compute the capacity of a channel.)

Let $\mathbf{W}$ represent a library of n_{LW} ligand words, where the ith input $\mathbf{W}_i$ is a vector of n_L ligand concentrations. Given a library of n_{CT} cell types, let $\mathbf{S}(\mathbf{W})$ represent the resulting activation profiles of these cell types, or a set of $n_{LW} \times n_{CT}$ responses. Earlier sections have presented a way to compute $\mathbf{S}(\mathbf{W})$ deterministically by solving quadratic equations. Here, we assume that the activation is probabilistic due to a Gaussian error bar of size σ around the deterministic solution $\mathbf{S}_{\text{determ}}(\mathbf{W})$. The standard deviation σ can represent molecular fluctuations upstream of SMAD (e.g., in receptor levels or activity) that result in fluctuations of SMAD phosphorylation. Thus, the distribution of signaling activities can be represented as follows:

$$P(\mathbf{S} \mid \mathbf{W}) = \text{Normal}\left(\mathbf{S}_{\text{determ}}(\mathbf{W}), \sigma^2\right) \tag{3.30}$$

We compute mutual information $I(\mathbf{S}, \mathbf{W})$ using the formula below:

$$I(\mathbf{S}, \mathbf{W}) = H(\mathbf{S}) - H(\mathbf{S} \mid \mathbf{W}) \tag{3.31}$$

The second term can be expressed as follows:

$$H(\mathbf{S} \mid \mathbf{W}) = \sum_{i=1}^{n_{LW}} p(\mathbf{W}_i) H(\mathbf{S} \mid \mathbf{W} = \mathbf{W}_i) \tag{3.32}$$

Each $H(\mathbf{S} \mid \mathbf{W} = \mathbf{W}_i)$ is the entropy of a n_{CT}-dimensional Gaussian with covariance matrix $\Sigma = \sigma^2 I_{n_{CT}}$, where $I_{n_{CT}}$ is the identity matrix of size n_{CT}. This entropy (in

bits) is

$$\begin{aligned} H\left(\mathbf{S} \mid \mathbf{W}=\mathbf{W}_i\right) &= \frac{1}{2} \lg\left[\det\left(2\pi e \Sigma\right)\right] \\ &= \frac{1}{2} \lg\left[\left(2\pi e \sigma^2\right)^{n_{CT}}\right] \\ &= \frac{n_{CT}}{2} \lg\left[2\pi e \sigma^2\right] \end{aligned} \tag{3.33}$$

Assuming input probabilities are uniformly distributed, or $p\left(\mathbf{W}_i\right) = \frac{1}{n_{LW}}$, this conditional entropy is simply given by

$$H\left(\mathbf{S} \mid \mathbf{W}\right) = \frac{n_{CT}}{2} \lg\left[2\pi e \sigma^2\right] \tag{3.34}$$

Thus, this term is constant regardless of choice of biochemical parameters.

The entropy $H\left(\mathbf{S}\right)$ in Equation 3.31 is the entropy of $P\left(\mathbf{S}\right)$, which is a sum of Gaussians, one at each of the activation patterns corresponding to each ligand input $\mathbf{W}_i$. This entropy is a measure of the distinguishability of activation patterns $\mathbf{S}\left(\mathbf{W}_i\right)$ for different inputs $\mathbf{W}_i$; the entropy will be small if the Gaussians are overlapping and large otherwise. Intuitively, this entropy is a measure of how well separated the activation patterns for different ligand inputs are.

The problem of determining the entropy of a normalized sum of Gaussians (i.e., a Gaussian mixture) in high dimensions is complex; however, simple analytic approximations have been developed in a recent advance (Kolchinsky and Tracey, 2017). We use the approximation to the kernel density estimator presented therein for a sum of n_{LW} Gaussians $p\left(x\right) = \frac{1}{n_{LW}} \sum_{i=1}^{n_{LW}} p_i\left(x\right)$ in n_{CT} dimensions:

$$H_{KL}\left(p\left(x\right)\right) = \frac{n_{CT}}{2} - \sum_{i=1}^{n_{LW}} m_i \ln\left[\sum_{j=1}^{n_{LW}} m_j p_j\left(\boldsymbol{\mu}_i\right)\right] \tag{3.35}$$

Here, p_j is the jth Gaussian component (normalized to 1, individually), $\boldsymbol{\mu}_i$ the mean of the ith component, and m_i the mixture weight of the ith component. In this case,

we assume uniform mixture weights, or $m_i = \frac{1}{n_{LW}}$ for all i. Further, $p_j\left(\boldsymbol{\mu}_i\right)$ is

$$p_j\left(\boldsymbol{\mu}_i\right) = \frac{1}{\sqrt{(2\pi)^{n_{CT}} \det \Sigma}} e^{-\frac{1}{2}\left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right)^T \Sigma^{-1} \left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right)} \tag{3.36}$$

Thus, the mutual information can be evaluated by simply evaluating each Gaussian at the mean of all other Gaussian components, or by using the matrix $\mathbf{D}_{ij}$ of distances between activation patterns $\mathbf{S}_{\text{determ}}\left(\mathbf{W}_i\right)$ for different ligand inputs $\mathbf{W}_i$:

$$\mathbf{D}_{ij} = \left\|\mathbf{S}_{\text{determ}}\left(\mathbf{W}_i\right) - \mathbf{S}_{\text{determ}}\left(\mathbf{W}_j\right)\right\|^2 \tag{3.37}$$

We can therefore simplify $p_j\left(\boldsymbol{\mu}_i\right)$ to

$$p_j\left(\boldsymbol{\mu}_i\right) = \frac{1}{\sqrt{\left(2\pi\sigma^2\right)^{n_{CT}}}} e^{-\frac{\mathbf{D}_{ij}}{2\sigma^2}} \tag{3.38}$$

Substituting into the approximation, we find an entropy (in nats) of

$$\begin{aligned}
H_{KL}\left(P\left(\mathbf{S}\right)\right) &= \frac{n_{CT}}{2} - \sum_{i=1}^{n_{LW}} \frac{1}{n_{LW}} \ln\left[\sum_{j=1}^{n_{LW}} \frac{1}{n_{LW}} \cdot \frac{1}{\sqrt{\left(2\pi\sigma^2\right)^{n_{CT}}}} e^{-\frac{\mathbf{D}_{ij}}{2\sigma^2}}\right] \\
&= \frac{n_{CT}}{2} - \frac{1}{n_{LW}} \sum_{i=1}^{n_{LW}} \ln\left[\frac{1}{n_{LW}} \cdot \frac{1}{\sqrt{\left(2\pi\sigma^2\right)^{n_{CT}}}} \sum_{j=1}^{n_{LW}} e^{-\frac{\mathbf{D}_{ij}}{2\sigma^2}}\right] \\
&= \frac{n_{CT}}{2} - \frac{1}{n_{LW}} \sum_{i=1}^{n_{LW}} \left(-\ln\left[n_{LW}\right] - \frac{n_{CT}}{2} \ln\left[2\pi\sigma^2\right] + \ln\left[\sum_{j=1}^{n_{LW}} e^{-\frac{\mathbf{D}_{ij}}{2\sigma^2}}\right]\right) \\
&= \frac{n_{CT}}{2} + \ln\left[n_{LW}\right] + \frac{n_{CT}}{2} \ln\left[2\pi\sigma^2\right] - \frac{1}{n_{LW}} \sum_{i=1}^{n_{LW}} \left(\ln\left[\sum_{j=1}^{n_{LW}} e^{-\frac{\mathbf{D}_{ij}}{2\sigma^2}}\right]\right) \\
&= \ln\left[n_{LW}\right] + \frac{n_{CT}}{2} \ln\left[2\pi e\sigma^2\right] - \frac{1}{n_{LW}} \sum_{i=1}^{n_{LW}} \left(\ln\left[\sum_{j=1}^{n_{LW}} e^{-\frac{\mathbf{D}_{ij}}{2\sigma^2}}\right]\right)
\end{aligned} \tag{3.39}$$

We can convert this expression to bits by multiplying by $\lg 2$ and then combine with Equation 3.34 to estimate mutual information. We note that these derivations omit a correction factor $-n_{CT} \lg \Delta S$ arising from binning with bin width ΔS to make the entropy of a continuous distribution well defined; however, as the same correction

applies to both $H(\mathbf{S})$ and $H(\mathbf{S} \mid \mathbf{W})$, these terms cancel out. Our estimator of mutual information is therefore

$$\begin{aligned} I(\mathbf{S}, \mathbf{W}) &= H(\mathbf{S}) - H(\mathbf{S} \mid \mathbf{W}) \\ &= \lg e \left(\ln [n_{LW}] + \frac{n_{CT}}{2} \ln \left[2\pi e \sigma^2 \right] - \frac{1}{n_{LW}} \sum_{i=1}^{n_{LW}} \left(\ln \left[\sum_{j=1}^{n_{LW}} e^{-\frac{\mathbf{D}_{ij}}{2\sigma^2}} \right] \right) \right) \\ &\quad - \frac{n_{CT}}{2} \lg \left[2\pi e \sigma^2 \right] \\ &= \lg [n_{LW}] - \frac{1}{n_{LW}} \sum_{i=1}^{n_{LW}} \lg \left[\sum_{j=1}^{n_{LW}} e^{-\frac{\mathbf{D}_{ij}}{2\sigma^2}} \right] \end{aligned} \tag{3.40}$$

We use this expression to estimate mutual information in this paper. From the form, it is clear that mutual information rewards large values of $\mathbf{D}_{ij}$, or distinct activation patterns for different ligand inputs.

The mutual information framework above can be naturally extended to scenarios not considered here. For example, not all ligand inputs might be equally likely or of equal physiological significance. In this case, the map of ligand inputs to activation profiles (i.e., the coding scheme) can separate the activation patterns of more important ligand words at the expense of more similar activation patterns for less important words. The mutual information framework can account for such weighting of different inputs easily through unequal $p(\mathbf{W}_i)$ above.

Finally, note that mutual information naturally rewards robustness, since mutual information is higher when each activation pattern is realized over equally sized regions of input space. For example, if an output $\mathbf{S}_1$ is only obtained for a sliver of ligand input space while another output pattern $\mathbf{S}_2$ is realized over the rest of input space, mutual information will be lower than if both outputs are realized over half of input space.

3.6.13 Analysis of Mutual Information

Libraries of Ligand Words and Cell Types

To provide the broadest information theoretic characterization, we first constructed comprehensive libraries of ligand words, or input messages, and cell types, or receptor expression profiles. Each ligand can independently take on three distinct concentrations sampled logarithmically over three orders of magnitude, or $10^0 = 1$, $10^{1.5} \approx 32$, and $10^3 = 1{,}000$. This library of $3^2 = 9$ words is a representative sampling of all possible ligand inputs. Similarly, we constructed a library of cell types by varying each receptor level independently over two distinct concentration levels $\{1, 100\}$. For a system with two type I receptor variants and two type II receptor variants, this library comprises $2^{2+2} = 16$ possible cell types.

Systematic Sampling of Biochemical Parameters

In our model, a promiscuous ligand-receptor system is defined by its interaction affinities K and signaling activities e. To comprehensively sample all possible biochemical parameters, each parameter was allowed to be either of $\{0.1, 1\}$, giving a total of $2^{16} = 65{,}536$ qualitatively distinct parameter sets. We then evaluated mutual information between the ligand words and the corresponding cell type activation profiles for each possible choice of biochemical parameters (K, e). The resulting data characterize the combinatorial addressing power across a comprehensive set of promiscuous ligand-receptor systems.

Random Sampling of Biochemical Parameters

To ensure that the grid-based sampling procedure did not introduce any artifacts, we also repeated this analysis for an identical number of randomly generated parameter sets. Specifically, we chose each value independently and randomly with a log-uniform distribution over $\left[10^{-1.5}, 1\right]$. The resulting distribution of mutual information values is shown in Figure 3.S6A.

Choice of Variance

Computing mutual information requires choosing the Gaussian fluctuation or variance σ^2 for activation levels. For all results here, we choose $\sigma^2 = 0.5$, based on testing a range of values. Very large or small choices of σ lead to the same value of mutual information for all biochemical parameters, either low or high, respectively. Intermediate choices of σ discriminate between different (K, e). While the precise value of mutual information depends on σ, different choices of σ do not qualitatively change the relative ordering of biochemical parameter sets.

3.6.14 Optimization of Mutual Information

These sampling procedures enable us to comprehensively analyze mutual information across parameter space. However, they are likely to miss extremes of mutual information. Therefore, we chose the 16 parameter sets from the systematic grid-based sampling with the highest starting mutual information and further refined (K, e) to maximize mutual information, using least-squares optimization.

3.6.15 Addressability of Ligand Words

To determine which ligand words in the library activate distinct combinations of cell types, we define the overall addressability of a set of ligand words by evaluating all pairs. To compare a pair of ligand words, we compute the ratio of activation levels for each cell type and take the separation r as the largest such fold change (inverted if needed, such that $r \geq 1$). If two ligand words have a separation of 5, then at least one cell type's activation is different by a factor of at least 5 in the two conditions. We extend this pairwise separation to a set of N ligand words by forming a $N \times N$ addressability matrix, where element (i, j) corresponds to the pairwise separation of ligand words i and j. This matrix has 1s along the diagonal. We define the smallest off-diagonal value, which represents the minimum pairwise

separation between different ligand words, to be the overall addressability of that set of N ligand words.

3.6.16 Analysis of Archetypal Responses

Response Classes

We next analyzed the responses generated by the full library of cell types for high-performing and low-performing parameter sets. As expected, parameter sets giving rise to low mutual information showed relatively little diversity in responses (Figure 3.6D); parameter sets that generated high mutual information showed distinct activation patterns among cell types (Figure 3.6E). Furthermore, these response types appeared qualitatively different and were similar to experimentally observed patterns reported previously (Antebi et al., 2017). We therefore further analyzed the presence of these archetypal responses across parameter sets.

Examples of these archetypes were generated by simulating responses to parameters reflecting our understanding of the underlying design principles (Figure 3.S1). These parameters are not specifically tuned, with all affinity and activity values set to either 0.1 or 1 and all receptor levels fixed at $10^{-1.5}$. Thus, they reflect qualitative differences rather than finely tuned quantitative ones. Briefly, ratiometric responses feature reduction of activity of one ligand by the second, such that the overall response approximates the ratio of the two concentrations. Competitive inhibition, where the "denominator" competes for receptors needed to generate signaling activity but produces inactive complexes, can produce such responses (Figure 3.S1A). Additive responses approximate the sum of the two ligand concentrations, as the ligands increase pathway activity either alone or together, and are readily generated when both ligands activate receptors similarly (Figure 3.S1B). Imbalance detection responses, where cells respond maximally to imbalances in the levels of the two ligands, can arise if, for instance, competition between two ligands favors complexes

with low signaling activity (Figure 3.S1C). Conversely, balance detection responses, where cells respond maximally when both ligands are present at a specific ratio, can be generated when ligand binding favors formation of high-activity signaling complexes (Figure 3.S1D).

Phenotypical Parameters

We characterized the spectrum of responses as described previously (Antebi et al., 2017). Briefly, we use the relative ligand strength (RLS), which represents the ratio of activation produced by the weaker ligand to that produced by the stronger ligand, and the ligand interference coefficient (LIC), which measures the degree to which two ligands positively or negatively synergize. We computed these values for each of the 16 responses of the cell type library for each set of biochemical parameters and determined what response classes each fell into, adapting previously described criteria (Antebi et al., 2017). Ratiometric responses were defined by RLS < 0.2, additive responses by RLS > 0.8 and $|\text{LIC}| < 0.05$, imbalance responses by RLS > 0.8 and LIC < -0.1, and balance responses by RLS > 0.8 and LIC > 0.1. Responses outside these ranges were considered to be intermediate variants and were not classified as a particular archetype.

Relationship with Mutual Information

Having identified the response classes represented for each set of biochemical parameters, we plotted the distribution of mutual information values associated with a given number of response classes (Figure 3.6F).

3.6.17 Analysis of Parameter Correlations

Based on observations from parameter sets with high mutual information, we computed two correlation measures for the biochemical parameters. Since each parameter could only take on two values, we transformed them to -1 and 1. In particular,

we defined $K'_{ijk} = -1$ if $K_{ijk} = 0.1$ (low) and $K'_{ijk} = 1$ if $K_{ijk} = 1$ (high), with e'_{ijk} defined analogously. (Equivalently, we defined $K'_{ijk} = 1 + 2 \log K_{ijk}$.) We computed $H_{Ke} = \sum_{i,j,k} K'_{ijk} e'_{ijk}$ to measure correlation between binding and signaling activity for each signaling complex. We also computed $H_{ee} = \sum_{j,k} e'_{1jk} e'_{2jk}$ to measure the correlation between activities of the two signaling complexes with the same receptor dimer but different ligands.

To analyze potential relationships between parameters, we calculated these two correlation metrics for each of the 65,536 parameter sets (Methods 3.6.17). To evaluate the association of these correlations with mutual information, parameter sets were sorted by mutual information and binned into sets of 800 (apart from the last bin, with a bin size of 736). These bins were then analyzed by computing the mean correlation and the mean mutual information (Figures 3.6G–H).

3.6.18 Evolutionary Algorithm as a Generative Model

While the observed anticorrelations of K and e appear to be predictive of mutual information, it is not clear if these relationships are sufficient to fully describe the criteria for high addressing power and can thus serve as a design principle. Therefore, we developed a generative algorithm that systematically evolves a given set of parameters (K, e) according to these principles and asked whether favoring anticorrelations yields higher addressing power.

We first formed a fitness function $F(K, e) = -(H_{Ke} + H_{ee})$, where H_{Ke} and H_{ee} are as above. Intuitively, a choice of (K, e) that has strong affinity-activity or activity-activity anticorrelations would have high fitness. Our algorithm is then a simple "evolutionary" algorithm that performs noisy gradient ascent in this fitness landscape. Starting with a given (K, e), each iteration involves choosing a random element of(K, e) and proposing a flip (changing it to high if currently low or vice

versa). We then compute the resulting change in fitness ΔF. If such a flip increases fitness ($\Delta F > 0$), we immediately implement it. If the proposed flip decreases fitness ($\Delta F < 0$), we accept it with a probability $e^{s\Delta F}$, where s represents the selection pressure. Such moves towards lower fitness allow dynamics to escape local fitness maxima; the frequency of such moves towards lower fitness is controlled by the selection pressure s (or, equivalently, temperature in Monte Carlo algorithms). We repeat this process over many iterations and track the addressing power of the resulting (K, e) configuration.

We ran our algorithm on 2,000 randomly initialized choices of (K, e). For each initialization, we performed 200 iterations of the evolutionary algorithm and quantified the addressing capacity of the final (K, e) using mutual information. We then visualized the resulting distribution of mutual information values (Figure 3.6K). With $s = 0$ (i.e., no selection for particular parameter relationships), there is a wide histogram equivalent to random sampling of parameter space; indeed, only a few parameter sets show substantial addressing power. However, with $s = 1$ and therefore selection for parameter anticorrelations, the resulting histogram is notably shifted towards higher addressing power, despite starting from similar randomly chosen initial conditions. Longer runs of the evolutionary algorithm did not change the resulting histograms, indicating equilibration within 200 iterations.

3.7 References

Albilia, J.B., Tenenbaum, H.C., Clokie, C.M.L., Walt, D.R., Baker, G.I., Psutka, D.J., Backstein, D., and Peel, S.A.F. (2013). Serum Levels of BMP-2, 4, 7 and AHSG in Patients With Degenerative Joint Disease Requiring Total Arthroplasty of the Hip and Temporomandibular Joints. *J. Orthop. Res.* 31, 44–52.

Andrews, M.G., del Castillo, L.M., Ochoa-Bolton, E., Yamauchi, K., Smogorzewski, J., and Butler, S.J. (2017). BMPs direct sensory interneuron identity in the developing spinal cord using signal-specific not morphogenic activities. *eLife* 6, e30647.

Antebi, Y.E., Linton, J.M., Klumpe, H., Bintu, B., Gong, M., Su, C., McCardell, R., and Elowitz, M.B. (2017). Combinatorial Signal Perception in the BMP Pathway. *Cell* 170, 1184–1196.

Axelrod, J.D., Miller, J.R., Shulman, J.M., Moon, R.T., and Perrimon, N. (1998). Differential recruitment of Dishevelled provides signaling specificity in the planar cell polarity and Wingless signaling pathways. *Genes Dev.* 12, 2610–2622.

Baur, S.T., Mai, J.J., and Dymecki, S.M. (2000). Combinatorial signaling through BMP receptor IB and GDF5: shaping of the distal mouse limb and the genetics of distal limb diversity. *Development* 127, 605–619.

Ben-Zvi, D., Shilo, B.-Z., Fainsod, A., and Barkai, N. (2008). Scaling of the BMP activation gradient in *Xenopus* embryos. *Nature* 453, 1205–1211.

Bier, E., and De Robertis, E.M. (2015). BMP gradients: A paradigm for morphogen-mediated developmental patterning. *Science* 348, aaa5838.

Blackwell, K.A., Hortschansky, P., Sanovic, S., Choudhary, S., Raisz, L.G., and Pilbeam, C.C. (2009). Bone morphogenetic protein 2 enhances PGE_2-stimulated osteoclast formation in murine bone marrow cultures. *Prostaglandins Other Lipid Mediat.* 90, 76–80.

Bois, J.S. (2020). Equilibrium Toolkit. *CaltechDATA*. DOI 10.22002/D1.1430.

Bradford, S.T.J., Ranghini, E.J., Grimley, E., Lee, P.H., and Dressler, G.R. (2019). High-throughput screens for agonists of bone morphogenetic protein (BMP) signaling identify potent benzoxazole compounds. *J. Biol. Chem.* 294, 3125–3136.

Bray, D. (1995). Protein molecules as computational elements in living cells. *Nature* 376, 307–312.

Burke, P., Schooler, K., and Wiley, H.S. (2001). Regulation of Epidermal Growth Factor Receptor Signaling by Endocytosis and Intracellular Trafficking. *Mol. Biol. Cell* 12, 1897–1910.

Chen, H., Ridgway, J.B., Sai, T., Lai, J., Warming, S., Chen, H., Roose-Girma, M., Zhang, G., Shou, W., and Yan, M. (2013). Context-dependent signaling defines

roles of BMP9 and BMP10 in embryonic and postnatal development. *Proc. Natl. Acad. Sci. U.S.A.* 110, 11887–11892.

Dai, D., Huang, Q., Nussinov, R., and Ma, B. (2014). Promiscuous and specific recognition among ephrins and Eph receptors. *Biochim. Biophys. Acta — Proteins Proteom.* 1844, 1729–1740.

Danesh, S.M., Villasenor, A., Chong, D., Soukup, C., and Cleaver, O. (2009). BMP and BMP receptor expression during murine organogenesis. *Gene Expr. Patterns* 9, 255–265.

David, C.J., and Massagué, J. (2018). Contextual determinants of TGFβ action in development, immunity and cancer. *Nat. Rev. Mol. Cell Biol.* 19, 419–435.

David, L., Mallet, C., Keramidas, M., Lamandé, N., Gasc, J.-M., Dupuis-Girod, S., Plauchu, H., Feige, J.-J., and Bailly, S. (2008). Bone Morphogenetic Protein-9 Is a Circulating Vascular Quiescence Factor. *Circ. Res.* 102, 914–922.

Dessaud, E., Yang, L.L., Hill, K., Cox, B., Ulloa, F., Ribeiro, A., Mynett, A., Novitch, B.G., and Briscoe, J. (2007). Interpretation of the sonic hedgehog morphogen gradient by a temporal adaptation mechanism. *Nature* 450, 717–720.

DeWitt, A., Iida, T., Lam, H.-Y., Hill, V., Wiley, H.S., and Lauffenburger, D.A. (2002). Affinity Regulates Spatial Range of EGF Receptor Autocrine Ligand Binding. *Dev. Biol.* 250, 305–316.

Dewulf, N., Verschueren, K., Lonnoy, O., Morén, A., Grimsby, S., Vande Spiegle, K., Miyazono, K., Huylebroeck, D., and Ten Dijke, P. (1995). Distinct spatial and temporal expression patterns of two type I receptors for bone morphogenetic proteins during mouse embryogenesis. *Endocrinology* 136, 2652–2663.

Diez-Roux, G., Banfi, S., Sultan, M., Geffers, L., Anand, S., Rozado, D., Magen, A., Canidio, E., Pagani, M., Peluso, I., et al. (2011). A High-Resolution Anatomical Atlas of the Transcriptome in the Mouse Embryo. *PLoS Biol.* 9, e1000582.

Dirks, R.M., Bois, J.S., Schaeffer, J.M., Winfree, E., and Pierce, N.A. (2007). Thermodynamic Analysis of Interacting Nucleic Acid Strands. *SIAM Rev.* 49, 65–88.

Duchamp-Viret, P., Chaput, M.A., and Duchamp, A. (1999). Odor Response Properties of Rat Olfactory Receptor Neurons. *Science* 284, 2171–2174.

Dudley, A.T., and Robertson, E.J. (1997). Overlapping expression domains of bone morphogenetic protein family members potentially account for limited tissue defects in *BMP7* deficient embryos. *Dev. Dynam.* 208, 349–362.

Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic Gene Expression in a Single Cell. *Science* 297, 1183–1186.

Epstein, N.E. (2011). Pros, cons, and costs of INFUSE in spinal surgery. *Surg. Neurol. Int.* 2, 10.

Erickson, G.F., and Shimasaki, S. (2003). The spatiotemporal expression pattern of the bone morphogenetic protein family in rat ovary cell types during the estrous cycle. *Reprod. Biol. Endocrinol.* 1, 9.

Falo-Sanjuan, J., Lammers, N.C., Garcia, H.G., and Bray, S.J. (2019). Enhancer Priming Enables Fast and Sustained Transcriptional Responses to Notch Signaling. *Dev. Cell* 50, 411–425.

Flevaris, P., Khan, S.S., Eren, M., Schuldt, A.J.T., Shah, S.J., Lee, D.C., Gupta, S., Shapiro, A.D., Burridge, P.W., Ghosh, A.K., et al. (2017). Plasminogen Activator Inhibitor Type I Controls Cardiomyocyte Transforming Growth Factor-β and Cardiac Fibrosis. *Circulation* 136, 664–679.

Godin, R.E., Robertson, E.J., and Dudley, A.T. (1999). Role of BMP family members during kidney development. *Int. J. Dev. Biol.* 43, 405–411.

Goldman, A.L., Van der Goes van Naters, W., Lessing, D., Warr, C.G., and Carlson, J.R. (2005). Coexpression of Two Functional Odor Receptors in One Neuron. *Neuron* 45, 661–666.

Graham, S.J.L., Wicher, K.B., Jedrusik, A., Guo, G., Herath, W., Robson, P., and Zernicka-Goetz, M. (2014). BMP signalling regulates the pre-implantation development of extra-embryonic cell lineages in the mouse embryo. *Nat. Commun.* 5, 5667.

Grassinger, J., Simon, M., Mueller, G., Drewel, D., Andreesen, R., and Hennemann, B. (2007). Bone morphogenetic protein (BMP)-7 but not BMP-2 and BMP-4 improves maintenance of primitive peripheral blood-derived hematopoietic progenitor cells (HPC) cultured in serum-free medium supplemented with early acting cytokines. *Cytokine* 40, 165–171.

Gupta, M.C., and Khan, S.N. (2005). Application of bone morphogenetic proteins in spinal fusion. *Cytokine Growth Factor Rev.* 16, 347–355.

Hallem, E.A., and Carlson, J.R. (2006). Coding of Odors by a Receptor Repertoire. *Cell* 125, 143–160.

Hanley, J.A., and McNeil, B.J. (1982). The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. *Radiology* 143, 29–36.

Hatsell, S.J., Idone, V., Alessi Wolken, D.M., Huang, L., Kim, H.J., Wang, L., Wen, X., Nannuru, K.C., Jimenez, J., Xie, L., et al. (2015). *ACVR1*^{R206H} receptor mutation causes fibrodysplasia ossificans progressiva by imparting responsiveness to activin A. *Sci. Transl. Med.* 7, 303ra137.

Hegarty, S.V., Collins, L.M., Gavin, A.M., Roche, S.L., Wyatt, S.L., Sullivan, A.M., and O'Keeffe, G.W. (2014). Canonical BMP-Smad Signalling Promotes Neurite Growth in Rat Midbrain Dopaminergic Neurons. *Neuromol. Med.* 16, 473–489.

Heldin, C.-H., Miyazono, K., and ten Dijke, P. (1997). TGF-β signalling from cell membrane to nucleus through SMAD proteins. *Nature* 390, 465–471.

Herrera, B., and Inman, G.J. (2009). A rapid and sensitive bioassay for the simultaneous measurement of multiple bone morphogenetic proteins. Identification and quantification of BMP4, BMP6 and BMP9 in bovine and human serum. *BMC Cell Biol.* 10, 20.

Huntley, M.H., Murugan, A., and Brenner, M.P. (2016). Information capacity of specific interactions. *Proc. Natl. Acad. Sci. U.S.A.* 113, 5841–5846.

Itzkovitz, S., Tlusty, T., and Alon, U. (2006). Coding limits on the number of transcription factors. *BMC Genomics* 7, 239.

Kapushesky, M., Emam, I., Holloway, E., Kurnosov, P., Zorin, A., Malone, J., Rustici, G., Williams, E., Parkinson, H., and Brazma, A. (2010). Gene Expression Atlas at the European Bioinformatics Institute. *Nucleic Acids Res.* 38, D690–D698.

Karim, M.S., Madamanchi, A., Dutko, J.A., Mullins, M.C., and Umulis, D.M. (2021). Heterodimer-heterotetramer formation mediates enhanced sensor activity in a biophysical model for BMP signaling. *PLoS Comput. Biol.* 17, e1009422.

Kim, H.J., Chung, J.H., Shin, S.Y., Shin, S.I., Kye, S.B., Kim, N.K., Kwon, T.G., Paeng, J.Y., Kim, J.W., Oh, O.H., et al. (2015). Efficacy of rhBMP-2/Hydroxyapatite on Sinus Floor Augmentation: A Multicenter, Randomized Controlled Clinical Trial. *J. Dent. Res.* 94, 158S–165S.

Kim, H.K.W., Oxendine, I., and Kamiya, N. (2013). High-concentration of BMP2 reduces cell proliferation and increases apoptosis via DKK1 and SOST in human primary periosteal cells. *Bone* 54, 141–150.

Klumpe, H.E., Langley, M.A., Linton, J.M., Su, C.J., Antebi, Y.E., and Elowitz, M.B. (2022). The context-dependent, combinatorial logic of BMP signaling. *Cell Syst.* 13, 388–407.

Kolchinsky, A., and Tracey, B.D. (2017). Estimating Mixture Entropy with Pairwise Distances. *Entropy* 19, 361.

Korchynskyi, O., and ten Dijke, P. (2002). Identification and Functional Characterization of Distinct Critically Important Bone Morphogenetic Protein-specific Response Elements in the Id1 Promoter. *J. Biol. Chem.* 277, 4883–4891.

Kuwada, S.K., Lund, K.A., Li, X.F., Cliften, P., Amsler, K., Opresko, L.K., and Wiley, H.S. (1998). Differential signaling and regulation of apical vs. basolateral EGFR in polarized epithelial cells. *Am. J. Physiol.* 275, C1419–C1428.

LaVaute, T.M., Yoo, Y.D., Pankratz, M.T., Weick, J.P., Gerstner, J.R., and Zhang, S.-C. (2009). Regulation of Neural Specification from Human Embryonic Stem Cells by BMP and FGF. *Stem Cells* 27, 1741–1749.

Li, C.W., and Ge, W. (2011). Spatiotemporal Expression of Bone Morphogenetic Protein Family Ligands and Receptors in the Zebrafish Ovary: A Potential Paracrine-Signaling Mechanism for Oocyte-Follicle Cell Communication. *Biol. Reprod.* 85, 977–986.

Li, P., Markson, J.S., Wang, S., Chen, S., Vachharajani, V., and Elowitz, M.B. (2018). Morphogen gradient reconstitution reveals Hedgehog pathway design principles. *Science* 360, 543–548.

Liem, K.F., Jr., Tremml, G., Roelink, H., and Jessell, T.M. (1995). Dorsal differentiation of neural plate cells induced by BMP-mediated signals from epidermal ectoderm. *Cell* 82, 969–979.

Lim-Tio, S.S., and Fuller, P.J. (1998). Intracellular Signaling Pathways Confer Specificity of Transactivation by Mineralocorticoid and Glucocorticoid Receptors. *Endocrinology* 139, 1653–1661.

Lind, M., Eriksen, E.F., and Bünger, C. (1996). Bone Morphogenetic Protein-2 but not Bone Morphogenetic Protein-4 and -6 Stimulates Chemotactic Migration of Human Osteoblasts, Human Marrow Osteoblasts, and U2-OS Cells. *Bone* 18, 53–57.

Llimargas, M., and Lawrence, P.A. (2001). Seven Wnt homologues in *Drosophila*: A case study of the developing tracheae. *Proc. Natl. Acad. Sci. U.S.A.* 98, 14487–14492.

Luo, J., Tang, M., Huang, J., He, B.-C., Gao, J.-L., Chen, L., Zuo, G.-W., Zhang, W., Luo, Q., Shi, Q., et al. (2010). TGFβ/BMP Type I Receptors ALK1 and ALK2 Are Essential for BMP9-induced Osteogenic Signaling in Mesenchymal Stem Cells. *J. Biol. Chem.* 285, 29588–29598.

Lyons, K.M., and Rosen, V. (2019). BMPs, TGFβ, and border security at the interzone. *Curr. Top. Dev. Biol.* 133, 153–170.

Malnic, B., Hirono, J., Sato, T., and Buck, L.B. (1999). Combinatorial Receptor Codes for Odors. *Cell* 96, 713–723.

Martinez-Hackert, E., Sundan, A., and Holien, T. (2021). Receptor binding competition: A paradigm for regulating TGF-β family action. *Cytokine Growth Factor Rev.* 57, 39–54.

Massagué, J. (1998). TGF-β signal transduction. *Annu. Rev. Biochem.* 67, 753–791.

Massagué, J. (2000). How cells read TGF-β signals. *Nat. Rev. Mol. Cell Biol.* 1, 169–178.

Miyazono, K., Kamiya, Y., and Morikawa, M. (2010). Bone morphogenetic protein receptors and signal transduction. *J. Biochem.* 147, 35–51.

Morsut, L., Roybal, K.T., Xiong, X., Gordley, R.M., Coyle, S.M., Thomson, M., and Lim, W.A. (2016). Engineering Customized Cell Sensing and Response Behaviors Using Synthetic Notch Receptors. *Cell* 164, 780–791.

Mueller, T.D., and Nickel, J. (2012). Promiscuity and specificity in BMP receptor activation. *FEBS Lett.* 586, 1846–1859.

Nandagopal, N., Santat, L.A., LeBon, L., Sprinzak, D., Bronner, M.E., and Elowitz, M.B. (2018). Dynamic Ligand Discrimination in the Notch Signaling Pathway. *Cell* 172, 869–880.

Newfeld, S.J., Wisotzkey, R.G., and Kumar, S. (1999). Molecular Evolution of a Developmental Pathway: Phylogenetic Analyses of Transforming Growth Factor-β Family Ligands, Receptors and Smad Signal Transducers. *Genetics* 152, 783–795.

Nickel, J., and Mueller, T.D. (2019). Specification of BMP Signaling. *Cells* 8, 1579.

O'Connor, M.B., Umulis, D., Othmer, H.G., and Blair, S.S. (2006). Shaping BMP morphogen gradients in the *Drosophila* embryo and pupal wing. *Development* 133, 183–193.

O'Keeffe, G.W., Hegarty, S.V., and Sullivan, A.M. (2017). Targeting bone morphogenetic protein signalling in midbrain dopaminergic neurons as a therapeutic approach in Parkinson's disease. *Neuronal Signal.* 1, NS20170027.

Ornitz, D.M., Xu, J., Colvin, J.S., McEwen, D.G., MacArthur, C.A., Coulier, F., Gao, G., and Goldfarb, M. (1996). Receptor Specificity of the Fibroblast Growth Factor Family. *J. Biol. Chem.* 271, 15292–15297.

O'Sullivan, D.B., Harrison, P.T., and Sullivan, A.M. (2010). Effects of GDF5 overexpression on embryonic rat dopaminergic neurones in vitro and in vivo. *J. Neural Transm.* 117, 559–572.

Panchision, D.M., Pickel, J.M., Studer, L., Lee, S.-H., Turner, P.A., Hazel, T.G., and McKay, R.D.G. (2001). Sequential actions of BMP receptors control neural precursor cell production and fate. *Genes Dev.* 15, 2094–2110.

Penn, M., Mausner-Fainberg, K., Golan, M., and Karni, A. (2017). High serum levels of BMP-2 correlate with BMP-4 and BMP-5 levels and induce reduced neuronal phenotype in patients with relapsing-remitting multiple sclerosis. *J. Neuroimmunol.* 310, 120–128.

Poon, B., Kha, T., Tran, S., and Dass, C.R. (2016). Bone morphogenetic protein-2 and bone therapy: successes and pitfalls. *J. Pharm. Pharmacol.* 68, 139–147.

Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA Synthesis in Mammalian Cells. *PLoS Biol.* 4, e309.

Raser, J.M., and O'Shea, E.K. (2005). Noise in Gene Expression: Origins, Consequences, and Control. *Science* 309, 2010–2013.

Resat, H., Ewald, J.A., Dixon, D.A., and Wiley, H.S. (2003). An Integrated Model of Epidermal Growth Factor Receptor Trafficking and Signal Transduction. *Biophys. J.* 85, 730–743.

Reversade, B., and De Robertis, E.M. (2005). Regulation of ADMP and BMP2/4/7 at Opposite Embryonic Poles Generates a Self-Regulating Morphogenetic Field. *Cell* 123, 1147–1160.

Rogers, K.W., and Schier, A.F. (2011). Morphogen gradients: from generation to interpretation. *Annu. Rev. Cell Dev. Biol.* 27, 377–407.

Salazar, V.S., Gamer, L.W., and Rosen, V. (2016). BMP signalling in skeletal development, disease and repair. *Nat. Rev. Endocrinol.* 12, 203–221.

Sanyal, A., Oursler, M.J., Clemens, V.R., Fukumoto, T., Fitzsimmons, J.S., and O'Driscoll, S.W. (2002). Temporal expression patterns of BMP receptors and collagen II (B) during periosteal chondrogenesis. *J. Orthop. Res.* 20, 58–65.

Schmierer, B., and Hill, C.S. (2007). TGFβ-SMAD signal transduction: molecular specificity and functional flexibility. *Nat. Rev. Mol. Cell Biol.* 8, 970–982.

Schmierer, B., Tournier, A.L., Bates, P.A., and Hill, C.S. (2008). Mathematical modeling identifies Smad nucleocytoplasmic shuttling as a dynamic signal-interpreting system. *Proc. Natl. Acad. Sci. U.S.A.* 105, 6608–6613.

Shankaran, H., Wiley, H.S., and Resat, H. (2007). Receptor downregulation and desensitization enhance the information processing ability of signalling receptors. *BMC Syst. Biol.* 1, 48.

Shankaran, H., Zhang, Y., Chrisler, W.B., Ewald, J.A., Wiley, H.S., and Resat, H. (2012). Integrated experimental and model-based analysis reveals the spatial aspects of EGFR activation dynamics. *Mol. Biosyst.* 8, 2868–2882.

Shaul, Y.D., and Seger, R. (2007). The MEK/ERK cascade: From signaling specificity to diverse functions. *Biochim. Biophys. Acta — Mol. Cell Res.* 1773, 1213–1226.

Shi, Y., and Massagué, J. (2003). Mechanisms of TGF-β Signaling from Cell Membrane to the Nucleus. *Cell* 113, 685–700.

Shimizu, K., Chiba, S., Hosoya, N., Kumano, K., Saito, T., Kurokawa, M., Kanda, Y., Hamada, Y., and Hirai, H. (2000a). Binding of Delta1, Jagged1, and Jagged2 to Notch2 Rapidly Induces Cleavage, Nuclear Translocation, and Hyperphosphorylation of Notch2. *Mol. Cell. Biol.* 20, 6913–6922.

Shimizu, K., Chiba, S., Saito, T., Kumano, K., and Hirai, H. (2000b). Physical Interaction of Delta1, Jagged1, and Jagged2 with Notch1 and Notch3 Receptors. *Biochem. Biophys. Res. Commun.* 276, 385–389.

Simic, P., and Vukicevic, S. (2005). Bone morphogenetic proteins in development and homeostasis of kidney. *Cytokine Growth Factor Rev.* 16, 299–308.

Suter, D.M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., and Naef, F. (2011). Mammalian Genes Are Transcribed with Widely Different Bursting Kinetics. *Science* 332, 472–474.

Tabula Muris Consortium. (2020). A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature* 583, 590–595.

Tabula Muris Consortium, Overall coordination, Logistical coordination, Organ collection and processing, Library preparation and sequencing, Computational data analysis, Cell type annotation, Writing group, Supplemental text writing group, and Principal investigators. (2018). Single-cell transcriptomics of 20 mouse organs creates a *Tabula Muris*. *Nature* 562, 367–372.

Valera, E., Isaacs, M.J., Kawakami, Y., Izpisúa Belmonte, J.C., and Choe, S. (2010). BMP-2/6 Heterodimer Is More Effective than BMP-2 or BMP-6 Homodimers as Inductor of Differentiation of Human Embryonic Stem Cells. *PLoS ONE* 5, e11167.

Van de Walle, I., Waegemans, E., De Medts, J., De Smet, G., De Smedt, M., Snauwaert, S., Vandekerckhove, B., Kerre, T., Leclercq, G., Plum, J., et al. (2013). Specific Notch receptor-ligand interactions control human TCR-$\alpha\beta$/$\gamma\delta$ development by inducing differential Notch signal strength. *J. Exp. Med.* 210, 683–697.

Varley, J.E., and Maxwell, G.D. (1996). BMP-2 and BMP-4, but Not BMP-6, Increase the Number of Adrenergic Cells Which Develop in Quail Trunk Neural Crest Cultures. *Exp. Neurol.* 140, 84–94.

Wagner, D.O., Sieber, C., Bhushan, R., Börgermann, J.H., Graf, D., and Knaus, P. (2010). BMPs: From Bone to Body Morphogenetic Proteins. *Sci. Signal.* 3, mr1.

Wang, R.N., Green, J., Wang, Z., Deng, Y., Qiao, M., Peabody, M., Zhang, Q., Ye, J., Yan, Z., Denduluri, S., et al. (2014). Bone Morphogenetic Protein (BMP) signaling in development and human diseases. *Genes Dis.* 1, 87–105.

Wang, S., Sun, A., Li, L., Zhao, G., Jia, J., Wang, K., Ge, J., and Zou, Y. (2012). Up-regulation of BMP-2 antagonizes TGF-β1/ROCK-enhanced cardiac fibrotic signalling through activation of Smurf1/Smad6 complex. *J. Cell. Mol. Med.* 16, 2301–2310.

Wells, A., and Wiley, H.S. (2018). A systems perspective of heterocellular signaling. *Essays Biochem.* 62, 607–617.

Wiley, H.S., Shvartsman, S.Y., and Lauffenburger, D.A. (2003). Computational modeling of the EGF-receptor system: a paradigm for systems biology. *Trends Cell Biol.* 13, 43–50.

Wodarz, A., and Nusse, R. (1998). Mechanisms of Wnt signaling in development. *Annu. Rev. Cell Dev. Biol.* 14, 59–88.

Wootten, D., Christopoulos, A., Marti-Solano, M., Babu, M.M., and Sexton, P.M. (2018). Mechanisms of signalling and biased agonism in G protein-coupled receptors. *Nat. Rev. Mol. Cell Biol.* 19, 638–653.

Xia, Y., Yu, P.B., Sidis, Y., Beppu, H., Bloch, K.D., Schneyer, A.L., and Lin, H.Y. (2007). Repulsive Guidance Molecule RGMa Alters Utilization of Bone Morphogenetic Protein (BMP) Type II Receptors by BMP2 and BMP4. *J. Biol. Chem.* 282, 18129–18140.

Yokoyama, Y., Watanabe, T., Tamura, Y., Hashizume, Y., Miyazono, K., and Ehata, S. (2017). Autocrine BMP-4 Signaling Is a Therapeutic Target in Colorectal Cancer. *Cancer Res.* 77, 4026–4038.

Yu, P.B., Deng, D.Y., Beppu, H., Hong, C.C., Lai, C., Hoyng, S.A., Kawai, N., and Bloch, K.D. (2008). Bone Morphogenetic Protein (BMP) Type II Receptor Is Required for BMP-mediated Growth Arrest and Differentiation in Pulmonary Artery Smooth Muscle Cells. *J. Biol. Chem.* 283, 3877–3888.

Zagorski, M., Tabata, Y., Brandenberg, N., Lutolf, M.P., Tkačik, G., Bollenbach, T., Briscoe, J., and Kicheva, A. (2017). Decoding of position in the developing neural tube from antiparallel morphogen gradients. *Science* 356, 1379–1383.

Zhang, D., Mehler, M.F., Song, Q., and Kessler, J.A. (1998). Development of Bone Morphogenetic Protein Receptors in the Nervous System and Possible Roles in Regulating trkC Expression. *J. Neurosci.* 18, 3314–3326.

Zhang, X., Ibrahimi, O.A., Olsen, S.K., Umemori, H., Mohammadi, M., and Ornitz, D.M. (2006). Receptor Specificity of the Fibroblast Growth Factor Family: The Complete Mammalian FGF Family. *J. Biol. Chem.* 281, 15694–15700.

Zhao, M., Berry, J.E., and Somerman, M.J. (2003). Bone Morphogenetic Protein-2 Inhibits Differentiation and Mineralization of Cementoblasts *in vitro*. *J. Dent. Res.* 82, 23–27.

Zhu, X., Zhang, H., Zhang, X., Ning, C., and Wang, Y. (2017). *In vitro* study on the osteogenesis enhancement effect of BMP-2 incorporated biomimetic apatite coating on titanium surfaces. *Dent. Mater. J.* 36, 677–685.

Chapter 4

Conclusion

4.1 Summary

In this work, we have developed a framework for modeling signaling pathways with promiscuous ligand-receptor interactions and applied this approach to the BMP pathway to better understand what features emerge from this promiscuous pathway architecture. We first show that ligand-receptor multiplicity and promiscuity enable cells to perform complex computations on ligand combinations and explore how these computations arise at the level of ligand-receptor signaling complexes as well as at the level of total pathway activity (Chapter 2). Building on these findings, we then show how cell types expressing different receptor combinations can implement distinct computations on the same ligand inputs, allowing ligand combinations to selectively activate or "address" individual cell types or groups of cell types (Chapter 3). Taken together, these results provide a deeper systems-level understanding of intercellular signaling pathways and their capabilities as communication systems.

4.2 Future Directions

This work illustrates how promiscuously interacting ligand and receptor variants can enable increased computational complexity and greater communication specificity compared to a seemingly simpler one-to-one architecture. While these findings provide a foundation for understanding how signaling pathways function in intercellular communication, our results focus on the application of a simplified model of the BMP pathway to a limited set of target behaviors. Further research to expand the scope of this work will be valuable for better understanding the many intercellular

signaling pathways with this promiscuous architecture. Here, we suggest possible future directions toward this goal and discuss key considerations for these areas.

4.2.1 Improvements in Methodology

Our work leveraged a systematic optimization approach to comprehensively identify target behaviors and test which behavior(s) can be achieved. However, both aspects—a "systematic" screening using an "optimization" method—have limitations. We largely focused on analyzing responses to two ligands using a three-level ligand discretization scheme, with $3^2 = 9$ total ligand words. More generally, analyzing n_L ligands with a d-level ligand discretization scheme would involve d^{n_L} ligand words. Determining which single-cell response functions are achievable would require testing on the order of $2^{d^{n_L}}$ possible binary targets, and optimizing for multi-cell addressing schemes would require even larger numbers, given that each cell type considered will have its own output. While these numbers can be modestly reduced (such as by eliminating targets that are invariant to ligand relabeling), it is clear that the number of targets increases superexponentially as more complex applications are considered, such as higher-resolution ligand discretization or increased ligand multiplicity. As a concrete example, our analysis of all possible single-cell responses to two ligands with a three-level ligand discretization has a baseline of $2^9 = 512$ possible targets. Increasing to a four-level ligand discretization would require $2^{4^2} = 2^{16} = 65{,}536$ targets, and increasing to three ligands would create an immense set of $2^{3^3} = 2^{27} = 134{,}217{,}728$ targets. Thus, systematically enumerating all possible outputs rapidly becomes intractable as larger problems are studied, given the superexponential growth in possible targets.

In addition to increasing the number of targets to consider, such applications generally make each optimization problem more challenging as well, since the number of variables to optimize—one set of affinity and activity parameters for the system as

well as one set of receptor expression levels for every cell type considered—typically increases concurrently. While our optimization approach successfully identified parameters for most response functions and addressing schemes that reproduced the target with high fidelity, the optimization problem is not guaranteed to yield global optima. We addressed this issue by performing repeated trials with many different random initializations, which should allow for better exploration of local optima. However, we observed that high-quality solutions were identified with decreasing frequency as the number of parameters under consideration increased, consistent with the fact that the parameter space to be explored expands exponentially with additional parameters. As such, it is challenging to assess whether more complex targets truly cannot be achieved or simply have not been identified with optimization.

Given these challenges, improvements in the optimization algorithm would be valuable in approaching further extensions of this work. Better optimization methods could increase the quality and/or speed of solutions. For example, methods that more accurately estimate global optima will allow for better solutions and increase our ability to determine which targets can or cannot be achieved, while methods that more efficiently explore parameter space or more frequently identify quality solutions will allow for faster solutions and enable us to apply this approach to more complex applications that may have previously been computationally intractable. Apart from pursuing improvements in the optimization method, developing new approaches entirely could allow for new avenues of study. For instance, incorporating an information theoretic approach enabled us to analyze information capacity across a comprehensive enumeration of parameter sets, although analyzing larger systems remained challenging due to the exponential increase in the number of parameter sets to consider. Developing additional methods could expand the questions we can answer about the communication capabilities of signaling pathways. Possibilities for new ways to approach questions of interest are discussed further below.

4.2.2 Spectrum of Computations

In this work, we performed a systematic optimization of all possible 3×3 binary response functions and revealed a broad range of achievable computations, including multimodal functions that have not previously been described (Figure 2.6). Given these diverse responses, testing 3×3 targets may not be sufficient to describe all possible functions, and further analysis could reveal additional response types. As a natural extension of our approach, we could use higher-resolution targets to investigate the existence of other types of functions. While increasing the number of levels of ligand discretization may yield an impractical number of targets, we could selectively identify a smaller subset to focus on; for example, we could filter out those that are essentially equivalent to targets from the low-resolution screen. However, other approaches could provide a complementary way to explore the spectrum of possible response functions. For example, we could randomly generate many different parameter sets, simulate their responses, and classify the results. This classification could be guided by specific quantitative metrics computed for each response (such as the phenotypical parameters of RLS and LIC used in this work), which would allow us to readily identify responses with particular features of interest, or similarity to the response types we have already identified, which would enable us to use our existing data to screen for responses that are more likely to represent new functions. We could also combine unsupervised learning methods such as clustering, principal component analysis (PCA), *t*-distributed stochastic neighbor embedding (t-SNE), and autoencoders to automatically learn the structure of response types without further input. These techniques could enhance our understanding of the full spectrum of computations that can be achieved with two ligands.

While our analysis focused exclusively on responses to pairs of ligands, the BMP pathway includes ten or more ligand variants in mammals, and combinations of three or more ligands have been identified in various biological contexts. As such,

analyzing how response complexity increases with ligand multiplicity will provide insight into the potential computations achievable in vivo. Since additional receptor variants can increase the types of possible responses, we expect that increased ligand multiplicity will likewise enable more complex computations. The number of possible three-ligand functions at the same three-level ligand discretization scheme is vastly larger than the number of two-ligand functions and is computationally impractical to analyze with our current approach, and refinement of the existing optimization algorithm is unlikely to provide enough improvement in performance to allow scaling up to a comprehensive screen. As such, it will likely be most effective to start from a simulation-based approach, or simulating many responses from randomly generated parameter sets and using unsupervised learning methods to better understand the structure of possible response types. As with our two-ligand systems, exploring the possible responses at the level of signaling complexes first should then provide a better understanding of how they integrate to generate the full spectrum of computations at the level of total pathway activity. Once different response types have been identified, we could then use optimization to identify additional functions of particular interest. Together, these results will enable us to map the possible computations on ligand inputs and better understand the complexity of BMP signaling in vivo, where multiple ligands are often expressed in combination.

4.2.3 Scalability of Addressing

We have shown that a fly-like model with two type I and two type II receptor subunits allows addressing cell types or groups of cell types with high specificity, flexibility, and robustness (Figures 3.3 and 3.6). A mammalian-like model with four type I and three type II receptor subunits further expanded the capacity for orthogonal addressing, which should likewise translate into higher addressing capacity overall. Quantifying this increase in information could be done by extending our analysis of

mutual information to the mammalian-like model and comparing the distributions of mutual information across parameter sets.

A complete understanding of how addressing scales with pathway complexity involves analyzing different numbers of variants for not only receptors but also ligands. The complexity of possible computations is likely to grow with more ligand variants, which would allow for more types of responses to be used in any given addressing system. However, many of the high-quality addressing systems we identified did not require diverse response types but instead incorporated multiple responses of the same type, such as ratiometric responses with different ligand preferences or sensitivities (Figure 3.S2). Even if computational complexity does not increase with the number of ligands, exponential growth in the space of ligand combinations still allows for higher bandwidth, due to the increased volumetric space for ligand combinations to differentially activate cell types. Initial analysis reveals that three-ligand systems can indeed expand on the capabilities of two-ligand systems—for example, increasing the number of orthogonal channels—by incorporating a small number of response classes that are used repeatedly with varying ligand preferences or sensitivities (Figure 4.1). However, identification of high-capacity addressing systems with our optimization approach is technically challenging. Increased ligand multiplicity yields an exponential increase in the number of ligand combinations and responses to consider, precluding a systematic screen; concurrently, the higher number of parameters to optimize reduces the quality and speed of optimization. New approaches circumventing these challenges would be valuable for analyzing how addressing capacity scales. For example, we could accept additional assumptions in the model, such that the model can be sufficiently simplified to allow for mathematically derived solutions (rather than computationally obtained ones). Developing improved or new approaches will allow us to mathematically and computationally describe how addressing scales with increasing numbers of pathway components.

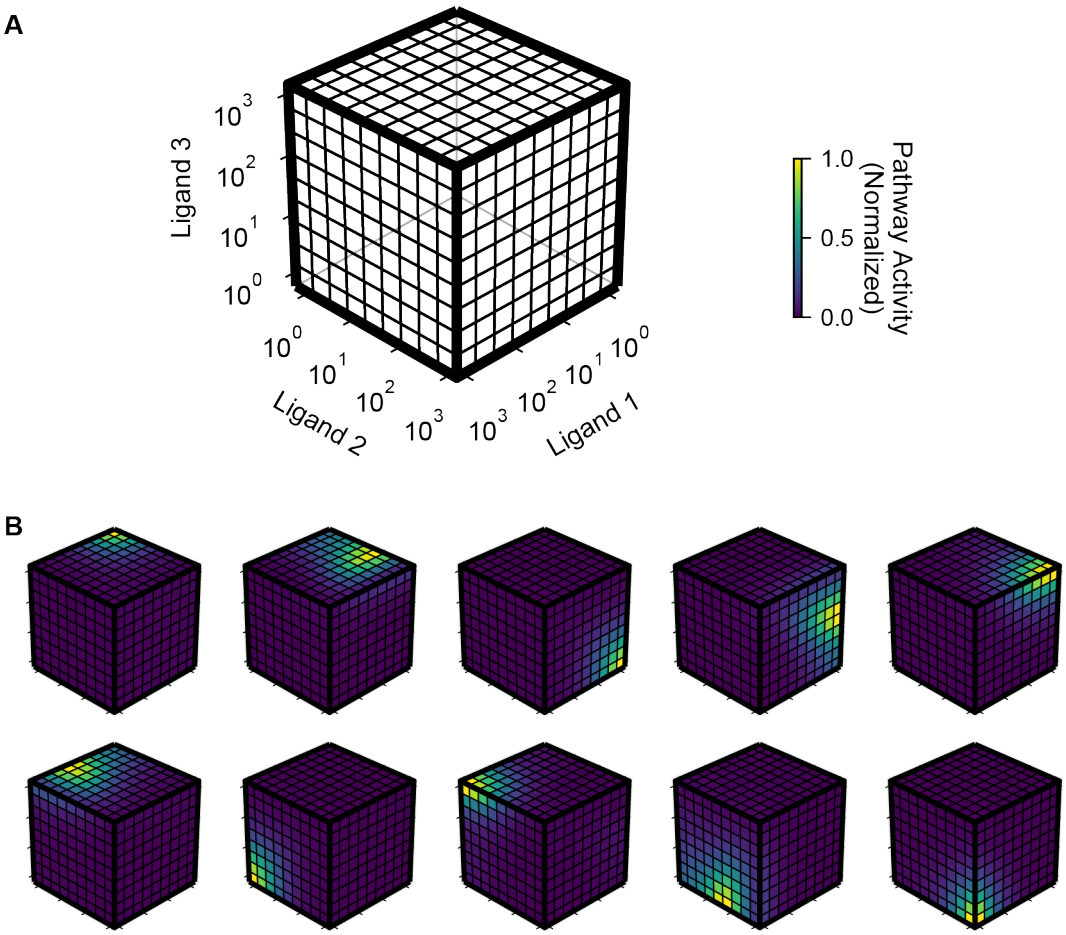

Figure 4.1: Addressing systems with three ligand variants can provide greater bandwidth than two-ligand systems.

(A) In systems with three ligands, responses can be visualized as cubes.

(B) For an example addressing system achieving ten orthogonal channels, the full responses of each cell type are shown, following the representation depicted in (A).

To experimentally evaluate the anticipated increase in addressing capacity between the fly and mammalian BMP systems, we could analyze a library of *Drosophila* cell types with different receptor profiles, interrogate their responses to different ligand pairs, and compare the results with the subset addressing profiles identified in mammalian cells (Figure 3.5). These analyses can also provide a different perspective on the evolutionary history of the BMP pathway; increasing organism

complexity has generally been associated with higher numbers of variants of pathway components, which may also enable greater capacity for addressing.

4.2.4 Generalizability to Biological Contexts

Ultimately, our goal is to develop a framework for addressing that can be used to understand organismal physiology and facilitate synthetic or therapeutic applications. Our results have shown that addressing can be realized in vitro (Figure 3.5), and we have preliminary evidence suggesting that the receptor profiles allowing for in vitro addressing can also be identified in vivo (Table 3.S1). However, it remains unclear whether and how addressing is used in vivo. Further work in this direction could include systematically analyzing receptor profiles of known cell types, using the rapidly growing atlases of single-cell expression data. Ideally, we would be able to use these known receptor expression profiles alongside fitted model parameters to be able to predict and design ligand combinations with desired cell type specificity. This ability to address signals to specific cell types would have broad utility in both synthetic and clinical applications.

Generalizing our modeling framework to biological applications will require validated estimates of model parameters that can capture the behavior of known BMP ligands and receptors. We have identified various parameter sets that are consistent with responses to pairs of BMP ligands in a comprehensive experimental screen. While improving the parameter fitting process could provide better estimates of the model parameters, we can also refine the possibilities by evaluating the predictive power of these parameter sets. We could simulate the responses to unmeasured ligand combinations predicted by each parameter set, identify those combinations that best differentiate between the parameter sets, and experimentally measure the responses to the chosen ligand combinations. Similarly, we can use the fitted parameters to predict responses of a new cell type with known receptor expression

profile, measure the responses experimentally, and validate or refine model parameters accordingly.

Alternatively, we could aim to directly measure the parameter values. One approach would be to generate cell lines expressing single receptor subunits and then measure the response of every type I and type II receptor pair with different ligands, thus providing an estimate for the affinity and activity parameters of each signaling complex. Identifying model parameters with full predictive power for BMP pathway responses would enable us to use the model to identify ligand combinations and cell types that exhibit desired addressing behaviors.

Our model focuses on a relatively limited aspect of BMP signaling, describing only ligand-receptor binding and activation of second messenger. As such, it may be impossible to fully predict cellular responses to BMP signaling without incorporating other biochemical processes that we have omitted. Including other mechanisms in our model will not only enhance the predictive power of the model but also reveal the ways in which those features modulate the function of the pathway as a communication system.

This work has focused on the capabilities of the BMP pathway, but multiple signaling pathways that are likewise of great biological and clinical importance also exhibit a promiscuous pathway architecture. The exact molecular details of each pathway vary, and these variations may result in different types of computations and addressing systems across pathways. Translating our modeling framework to other pathways and characterizing their responses experimentally will provide a broader understanding of how addressing is achieved in promiscuous signaling pathways as well as what unique capabilities are provided by each pathway.

Appendix A

Terminology

Combinatorial addressing involves mappings between combinations of ligands and responses of cell types defined by their receptor subunit expression. Here, we define some of the terminology introduced in this work to describe these relationships between ligand inputs and cell type responses.

Ligand word
: A set of specific concentration values for each ligand variant in a combination. For example, a concentration of 10 μM for ligand 1 and 100 μM for ligand 2 constitutes a ligand word (10 μM, 100 μM).

Cell type
: A set of specific receptor subunit expression levels. For example, a cell expressing receptor subunits 1 and 3 would represent a different cell type than a cell expressing subunits 1, 2, and 4 or a cell expressing more subunit 1 and less subunit 3.

Channel
: A set of one or more cell types that can be selectively activated (without activating other cell types) by some ligand word. For example, if ligand word 1 activates cell type A, while ligand word 2 activates cell types B and C, then "A" and "BC" constitute distinct channels.

Bandwidth
: The number of unique channels in a given system. As an example, suppose that ligand words 1 and 2 both activate only cell type A, while ligand word 3 activates cell type B. This system would have a bandwidth of two channels, as ligand words 1 and 2 yield the same activation profile.

Combinatorial addressing (or simply **addressing**)
: A mapping between ligand words and the corresponding cell type(s) activated by those words.

Orthogonal addressing
: A particular form of combinatorial addressing in which each ligand word activates a single, unique cell type. An example of three-channel orthogonal addressing is shown in Figure 3.2B.

Addressing repertoire

The combinations of cell types (each combination representing a channel) that can be activated across all possible ligand words for a given set of cell types and biochemical parameters. Examples of addressing repertoires are shown with the Venn diagrams in Figure 3.4A.

The next two terms define quantitative metrics used in this paper.

Distinguishability (Figures 3.2, 3.3, and 3.4; Methods 3.6.5)

A measure of the specificity of a given addressing scheme, defined as the ratio of the lowest on-target activity to the highest off-target activity, or (lowest on-target activity)/(highest off-target activity). As an example, consider a system where ligand word 1 activates on-target cell type A and off-target cell type B at levels of 0.8 and 0.1 units, respectively, while ligand word 2 activates off-target cell type A and on-target cell type B at levels of 0.4 and 0.9. The distinguishability for orthogonal addressing of "A" and "B" would then be $0.8/0.4 = 2$. As another example, if cell types A and B are both on-target for ligand word 2, addressing "A" and "AB" would have a distinguishability of $0.4/0.1 = 4$.

Addressability (Figure 3.6; Methods 3.6.15)

A measure of the diversity of the addressing repertoire for a set of ligand words. The separation of two ligand words is first quantified as the largest fold change of their resulting activation levels in any cell type. Addressability is then defined as the separation of the least separable pair of ligand words.