



HHS Public Access

Author manuscript

Cell Syst. Author manuscript; available in PMC 2024 June 21.

Published in final edited form as:

Cell Syst. 2023 June 21; 14(6): 430–446. doi:10.1016/j.cels.2023.05.001.

The computational capabilities of many-to-many protein interaction networks

Heidi E. Klumpe¹, Jordi Garcia-Ojalvo^{2,*}, Michael B. Elowitz^{3,4,*}, Yaron E. Antebi^{5,*}

¹Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA

²Department of Medicine and Life Sciences, Pompeu Fabra University, 08003 Barcelona, Spain

³Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA

⁴Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA

⁵Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel

Summary

Many biological circuits comprise sets of protein variants that interact with one another in a many-to-many, or promiscuous, fashion. These architectures can provide powerful computational capabilities that are especially critical in multicellular organisms. Understanding the principles of biochemical computations in these circuits could allow more precise cellular control of cellular behaviors. However, these systems are inherently difficult to analyze, due to their large number of interacting molecular components, partial redundancies, and cell context dependence. Here, we discuss recent experimental and theoretical advances that are beginning to reveal how promiscuous circuits compute, what roles those computations play in natural biological contexts, and how promiscuous architectures can be applied for the design of synthetic multicellular behaviors.

TOC blurb:

Biological pathways comprise sets of protein variants that interact with one another in a many-to-many fashion to biochemically “compute” complex functions of protein inputs. Here, Klumpe et al. discuss recent experimental and theoretical advances that reveal how these circuit architectures compute, what roles those computations play in natural biological contexts, and how they can be applied to the design of synthetic multicellular behaviors.

*co-corresponding authors. jordi.g.ojalvo@upf.edu, melowitz@caltech.edu, and ayaron@weizmann.ac.il.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Declaration of Interests

M.B.E. is a scientific advisory board member or consultant at TeraCyte, Primordium, and Spatial Genomics. Y.E.A. is a scientific advisory board member or consultant at TeraCyte. The authors declare no competing interests with this work.

Introduction

With the evolutionary transition from single cells to multicellular life, cells faced an expanded set of information-processing challenges. They had to support large numbers of distinct cell fates, decipher complex signals from other cells, and tailor their behavior depending on their own state, history, and local environment. Many of the protein circuits, or pathways, that evolved to address these challenges share a particular feature. They employ families of homologous, but subtly different, protein variants interacting in a many-to-many, or promiscuous, manner with one another, and with members of other protein families. These variants likely arose from duplications of simpler ancestral pathways¹⁻⁵ (Figure 1A). Typically, different cell types express different combinations of variants, generating a set of related, but distinct, versions of any given circuit⁶ (Figure 1B). While specific components and interactions within these circuits are often well-studied, it is still generally unknown how their components collectively respond to diverse combinations of inputs, and how those responses change depending on the specific components expressed in each cell context. Recent work is beginning to shed light on these issues, revealing how promiscuous circuits can support the complex information processing requirements of multicellular life.

Promiscuous protein interaction circuits are prevalent throughout biology. Developmental cell-cell communication pathways such as Bone Morphogenetic Protein (BMP),⁷ Notch,⁸ JAK/STAT,^{9,10} Wnt,^{11,12} Fibroblast Growth Factor (FGF),¹³ and Epidermal Growth Factor (EGF),¹⁴ comprise sets of distinct ligands, each of which interacts promiscuously with multiple receptor variants (Figure 1C). Promiscuity is also conspicuous in the interactions of receptors with downstream effectors. For example, individual G-protein coupled receptors (GPCRs) interact with multiple effector G proteins in an overlapping, promiscuous manner¹⁵ (Figure 1D). Transcriptional regulation, too, is replete with promiscuity. Families of transcription factor variants, such as NF- κ B, bHLH, bZIP, and POU/Sox, interact promiscuously to form diverse homo- and heterodimeric complexes (Figure 1E) that vary in their DNA binding specificity and activate or repress overlapping sets of target genes.¹⁶⁻¹⁸ Transcription factors also interact promiscuously with co-factors, as has been analyzed extensively for the Mediator complex¹⁹ and the homeobox genes.²⁰ Extracellularly, families of Cadherin variants interact promiscuously with one another to control adhesion between adjacent cells²¹⁻²⁴ (Figure 1F). In many of these examples, proteins may form higher order, oligomeric complexes with more than two components, adding additional complexity to the system. These examples show that combinatorial protein networks occur at multiple levels with multiple protein families.

It can be tempting to regard protein interaction promiscuity more as a nuisance—perhaps an artifact of gene duplication in evolutionary history—than a feature. On the other hand, complexity science, neurobiology and artificial neural networks show that simple elements, connected together in a many-to-many fashion, can act as powerful computational systems.²⁵ By computation, we mean the ability to process input information encoded in multiple molecular signals in flexible and complex ways. Computations are usually understood to comprise three distinct levels: a function to be computed, an algorithm to implement that function, and the physical media that can execute that algorithm.²⁶ The first two are abstract descriptions of how inputs become outputs and are agnostic of

physical implementation. By contrast, the properties of a particular physical medium may constrain the functions or algorithms it can implement. Here, we consider protein-protein dimerization networks as physical media for computation in cells, and discuss the algorithms and functions they can compute. We are interested in both what higher-level functions these networks support, as well as how this logic can be physically realized by a variety of protein families. In all these respects, we consider protein dimerization networks as a powerful architecture for computation in cells.

The interconnectedness of protein and metabolic networks within the cell has long been recognized as providing the potential for computations in cells.²⁷⁻²⁹ Previous works have largely focused on how interacting biomolecules could implement aspects of digital computing, such as Boolean logic gates^{27,30} or neural networks.^{28,31} It has generally been more challenging to determine what computations are implemented natively by natural biomolecular circuits in living cells. Better understanding the range of native computations that arise in these settings could explain how cells, with many distinct properties from silicon chips, compute complex responses to their diverse input signals and, more specifically, how these computational capabilities address the unique challenges of multicellularity.

When many protein complexes can form through combinatorial dimerization of a smaller number of monomeric components, their distribution will in general depend on the abundances of each protein, their pairwise interaction strengths (e.g. affinities), as well as other inputs. These dependencies can be nonlinear and indirect, such that perturbing one protein level can affect the concentrations of complexes in which it does not appear. Further, distinct complexes can, in general, have distinct target specificities or activity levels. As a result, inputs that modulate one or more protein components can be processed by the combinatorial dimerization network in non-trivial ways to produce computations. These computations can support higher level functions in signaling, adhesion, and transcriptional regulation.^{7,18,32,33}

This type of computation occurs at two levels. First, within an individual cell or cell type, systems can compute a specific function of multi-protein inputs. Second, the function that is computed can vary between different cell states (i.e. be context-dependent). This is possible because cell states, in general, express different subsets of the circuit's molecular components, as single-cell atlas data sets have revealed.⁶ Each configuration (and thus each cell state) may compute a distinct functional response to the same environmental inputs. A single combinatorial system then provides not just one cellular computation, but a whole repertoire of different computations in different cell states.

Here, we argue that combinatorial protein dimerization networks represent a general biological strategy to implement context-dependent computations in cells. We further show how this viewpoint provides a unifying framework to explain a diverse array of biological processes, including signaling, gene regulation, and cell-cell adhesion. More specifically, we first highlight recent work that reveals computations in promiscuous protein-protein interaction systems and how they arise. We then discuss possible biological functions of

these computations. Finally, we explore experimental and computational approaches to understand and manipulate them.

Throughout, we focus on computations that emerge from networks of combinatorial protein-protein binding interactions within pathways. We omit additional complexity that can emerge through interactions between different pathways,³⁴ combinatorial allosteric control of individual protein activities,³⁵⁻³⁷ as well as computations within enzymatic and metabolic networks. We also leave out important work in neural circuits such as olfaction, which depends on many-to-many interactions between odorants and receptors, as these systems rely on neural circuitry for most signal processing and receptor expression is limited to one per cell, precluding formation of diverse protein complexes. Lastly, because we primarily focus on protein-protein dimerization as a mechanism, we do not discuss the fascinating computational capabilities produced by other circuit features, such as the regulation of protein activity by co-factor binding, expression-mediated feedback, and post-translational modifications.

Promiscuous protein-protein interaction systems compute functions of multiple input proteins

Combinatorial protein networks are based on protein complex formation (Figure 1C-F, black and white illustrations). These systems consist of one or more sets of protein variants that can combinatorially assemble into a zoo of different protein complexes. In some cases, the complexes contain only one class of protein, of which there are many variants. For example, different bHLH transcription factors can promiscuously dimerize with one another to produce many distinct homo- and heterodimers (Figure 1E). In other cases, complexes contain two or more types of proteins, each of which may comprise many variants. For instance, a set of ligand variants can form many potential signaling complexes with a set of receptor variants (Figure 1C). In either case, each of the resulting complexes can exhibit a different level of activity or affect a different set of molecular targets. Further, when certain protein concentrations are limiting, competition to form complexes can lead to non-intuitive behaviors. Perturbing the abundance of one or more individual input protein components can directly affect the concentrations of complexes and indirectly affect the concentrations of other complexes. Considering certain proteins as “inputs” and the activity or effects of the complexes as “outputs,” the overall network can thereby perform a variety of input-output computations, depending on the concentrations of each component, and the binding energies or affinities for each possible interaction (Figure 2A).

Mathematical models of competitive complex formation can capture many of the effects described above, allowing us to explore what kinds of computations are possible within different combinatorial protein network architectures. In addition, they reveal how those computations depend on the overall interaction architecture and biochemical parameters (Box 1). For example, models of receptor-ligand interactions have revealed how signaling pathways can compute Boolean and other types of combinatorial responses to multi-ligand inputs, two examples which we explore below. These examples show how diverse ligand-

receptor affinities and signaling complex activities, defining features of combinatorial protein networks, can produce complex functions.

Computation of Boolean logic for two ligand inputs

An early theoretical study of combinatorial receptor-ligand interactions was performed by De Ronde *et al.*³⁸ The authors explored a set of ligand-receptor interaction architectures and compared their ability to produce different Boolean logic gates. To do this, they adapted the equilibrium statistical mechanics description of allostery, provided by the classical Monod-Wyman-Changeux model,³⁹ to describe how ligand binding alters the conformation of receptors to activate a pathway. In one scheme, they considered a dimer of two receptor subunits, in which one subunit could bind promiscuously to two ligand variants, while the other could bind only to one. By tuning the affinities of each ligand for each receptor, and the activities of the different possible complexes, the authors were able to generate all possible two-input Boolean logic functions. The most complex functional responses, such as XOR, required competition of two ligands to bind one of the receptor subunits. In this case, each ligand, presented individually, produced an activating receptor conformation. But when mixed together, one ligand outcompeted the other for its preferred subunit. The outcompeted ligand then bound its less preferred subunit and produced an inhibitory complex, leading to reduced activity when both ligands were present compared to the activity of each ligand alone. Simpler schemes, based on a single receptor or limited to homodimerization of receptors, exhibited more limited computational repertoires. These results demonstrated how ligands that promiscuously and competitively bind to heterodimeric receptors can perform a variety of multi-input computations.

Computation in the BMP signaling system

One example of computation in combinatorial protein networks occurs in the BMP pathway. This core cell-cell signaling pathway plays pivotal roles in a broad range of developmental, physiological, and disease processes. In mammals, BMP ligands are secreted by cells, diffuse through tissues to form morphogenetic gradients, and activate cognate receptors on signal-receiving cells. Critically, in most biological processes, cells are exposed to combinations of multiple BMP ligand variants, suggesting that the BMP signaling process could be inherently combinatorial.

To decipher these combinatorial input signals, cells express multiple variants of type I and type II receptor subunits, which assemble together with ligands to form signaling complexes, which activate downstream Smad transcription factors.⁴⁰ In humans, there are approximately 15 different ligand variants and 7 receptor variants (four type I and three type II), potentially generating thousands of distinct signaling complexes.^{7,41-43} Moreover, these complexes likely vary quantitatively in their rates of Smad phosphorylation,⁷ as non-signaling (i.e. inactive) complexes have been shown to occur in the closely related TGF β pathway.⁴⁴

This combinatorial complexity allows the BMP pathway to compute complex responses to multi-ligand combinations.⁴¹ For example, we found that in a mouse epithelial cell line, the ligand BMP4 strongly activates the BMP pathway, but its activity is blocked in a dose-dependent manner by another ligand, GDF5, such that pathway activity is

approximately proportional to the ratio of BMP4 and GDF5 concentrations. This type of ratiometric response can occur when two ligands bind to the same receptor subunits with similar affinity, but one (in this case, BMP4) forms an active signaling complex while the other (GDF5) forms a non-activating partial complex, or a weakly-activating full complex. Similar types of antagonism have been reported for TGF β ligands^{45,46} and a related form of ratiometric sensing has been observed in yeast.⁴⁷

Experiments also revealed more complex computations. For example, BMP4 and BMP10 produce an “imbalance detector” response, in which either ligand can efficiently activate on its own, but the two ligands inhibit each other’s activity, neutralizing the response at a particular concentration ratio. (In this way the pathway is most active when the ligand’s concentrations are “imbalanced.”) Modeling showed that imbalance detection occurs in “incoherent” parameter regimes, where complexes that form with high affinity exhibit weak specific activity, while less preferred complexes are more active. In this regime, mixtures of the two ligands preferentially form their high affinity, but low activity complexes. By contrast, when either of the ligands is present alone, a mixture of strong and weak signaling complexes form, producing greater total activity. In the opposite “coherent” regime, where high affinity complexes also have high activity, the model predicts the opposite “balance detector” response, which was experimentally observed in a different cell line.⁴¹ Together, these results demonstrate that the promiscuous BMP pathway performs complex computations on multi-ligand inputs that can be explained in terms of an interplay between the affinity and activity of the various signaling complexes. These explanations for experimentally observed behaviors were consistent with a simplified model. Taken together, these modeling and experimental studies show how combinatorial ligand-receptor networks can compute responses to multiple ligands. However, further experimental validation will be necessary to definitively establish the underlying mechanisms that generate the observed responses.

Making computation contextual

Single-cell atlas projects have revealed the transcriptional expression profiles of diverse cell states.⁴⁸⁻⁵² Analysis of these profiles showed that the components of many combinatorial protein networks are expressed in distinct and often recurring configurations in different cell types.⁶ For example, in pathways such as BMP, Wnt, and Notch, multiple receptor variants are co-expressed in specific combinations. This observation provokes the question of whether cells that express distinct, but overlapping, sets of pathway components can perform different computations (Figure 2A).

Modeling of the BMP pathway showed that receptor expression levels could strongly impact the computation performed by the pathway (Box 1). Consistent with this prediction, perturbing the expression of individual receptors experimentally changed the computational response of the cell to different ligand combinations.^{41,43} For example, knocking down the type II receptor BMPR2 changed an additive response to BMP4 and BMP9 into a ratiometric response. Similarly, ectopically expressing the type I receptor ACVRL1 converted an imbalance detector response function into an additive response. Analysis of cell lines also suggested that receptor expression profile dictates computation. Two otherwise

unrelated cell lines with similar receptor profiles, one epithelial and one fibroblastic, exhibited similar multi-ligand integration responses. These differed qualitatively from responses of an embryonic stem cell line with a different receptor profile. Together, these results suggest that receptor expression profiles can (and likely do) control the type of computation that the BMP pathway performs on multiple ligand inputs.

This specific feature observed in signaling pathways suggests a more general feature of promiscuous protein-protein interaction networks. Indeed, earlier theoretical work on such networks showed that, in particular regimes, the steady-state concentrations of components depended more on protein abundance than the parameters of the protein-protein interactions themselves, suggesting that new network behaviors could be flexibly accessed by merely changing the concentration of different components.⁵³ Therefore, in addition to performing complex computations on multiple inputs, protein-protein dimerization networks also allow cells to reprogram those computations by varying the expression levels of pathway components. This ability to perform a range of computations on the same inputs allows different cell types to extract different types of information from the same environment, or a given cell type to change its behavior in different developmental or physiological contexts. Returning to the neural network analogy, this behavior roughly corresponds to different cell lines operating the same network topology with different weights to compute distinct functions.

Molecular promiscuity enables cell-cell specificity in communication pathways

One of the most basic capabilities of any communication system is the ability to address messages to particular recipients. For example, in the familiar case of email, we can send messages to individuals or groups. Cell-cell communication presents an analogous addressing challenge: how to use signals to selectively activate specific target cell types.

The simplest way to achieve such cell-specific “addressing” is a one-to-one ligand-receptor system in which each ligand variant activates a distinct receptor variant exclusively expressed by a single target cell type. This architecture reflects the strategy used in human engineered communication systems, where the goal is to avoid undesired crosstalk between communication channels.^{54,55} It is also the basis for synthetic biological signaling systems such as synNotch, where each ‘ligand’ (cell surface protein) is recognized by a cognate antibody fused to its corresponding synNotch receptor.⁵⁶ By contrast, in many natural cell-cell communication pathways, each ligand variant activates multiple receptor variants, each cell type expresses multiple receptor variants, and each environment contains multiple ligand variants. Is such a system compatible with communication specificity in the presence of promiscuous interactions?

In a recent computational study, we sought to answer this question using a mathematical model of a BMP-inspired promiscuous signaling pathway.⁵⁷ This work revealed that molecular promiscuity counterintuitively generates a powerful, fundamentally combinatorial, addressing capability, in which ligand combinations activate cells expressing corresponding receptor *combinations*. In this system, just two ligand variants, present together at different concentrations, could orthogonally activate at least eight different cell types (Figure 2A). This occurs because cells with different receptor profiles can respond to

different localized regions of ligand concentration space. These localized responses are in turn enabled by the computational capacity generated through promiscuous ligand-receptor interactions.

Increasing the number of receptor variants in the system allows more complex computational responses, and thereby increases the number of cell types that can be uniquely addressed by a given number of ligands. Key features that facilitate combinatorial addressing are (1) diverse activities for a given receptor, depending on the ligand that activates it, and (2) the existence of signaling complexes that form with high affinity but have low activity, and vice versa,⁵⁷ similar to the interplay observed in other models of BMP signaling described above.⁴¹

The addressing capabilities in the many-to-many architecture can far exceed what is possible in the seemingly simpler and more efficient one-to-one architecture, assuming the same number of ligand and receptor variants. This advantage becomes even stronger when one considers schemes in which a specific ligand combination activates multiple cell types (analogous to an email “mailing list”). For example, a particular combination of 2 ligands could selectively activate a subset of 3 cell types, showing that promiscuous interactions can produce versatile addressing schemes. Overall, these results suggest a picture in which different cell types “tune in” to specific combinations of ligands by expressing different receptor combinations. Different environments, with distinct combinations of ligands, can then preferentially activate one cell type, another, or both. Nonetheless, direct experimental tests will be necessary to confirm if evolved signaling capacities make use of this addressing capability.

Molecular promiscuity and cellular specificity in adhesion

Context-dependent computation also enables another important biological function: the ability of one cell to distinguish itself from other cells. For example, neural circuit assembly requires self-avoidance: individual neurons must physically contact other neurons but avoid contacting themselves.⁵⁸ To that end, neurons use unique sets of surface adhesion proteins to distinguish their own identity from that of other cells. In *Drosophila*, each neuron stochastically expresses a single isoform of the cell adhesion protein Dscam1 out of a large set (~19,000) of distinct isoforms. These isoforms interact homophilically (in a one-to-one manner), leading to contact-dependent repulsion for same-cell interactions. In this way, every isoform directly represents a single address.

By contrast, vertebrates use a combinatorial approach to generate a much larger set of addresses for unique neurons.⁵⁹ Each cell stochastically expresses ~15 different protocadherin variants from a much smaller set of 58 (in the case of the mouse) variants. Within the same cell (i.e. in *cis*), different variants interact promiscuously,⁶⁰ resulting in multimeric, heterogeneous protocadherin clusters. Each combination of expressed proteins represents a unique cellular identity. These protein variants interact in a many-to-many manner in *cis*, with each protein variant able to bind to all others. However, the multi-protein clusters exhibit a combinatorially homophilic interaction between juxtaposed cell membranes (in *trans*), such that a perfect match of *all* components is required for *trans* interactions, whether on the same or different cells (Figure 2B). By experimentally co-

expressing sets of up to 5 different protocadherin variants, Thu *et al.* showed that only cells with identical expression patterns exhibit strong *trans* binding. Mismatch of a single isoform located on two juxtaposed membranes was sufficient to greatly diminish homophilic adhesion strength.^{60,61} In this way, combinations of a few promiscuously interacting protocadherin isoforms in vertebrates provide neurons with a vast number of orthogonally interacting addresses and enable them to discriminate self from non-self. Recent work has shown that this mechanism might play more general roles in the mechanical organization of tissues during morphogenesis.⁶²

Transcription factor dimerization networks allow complex gene regulation

In a multicellular organism, combinatorial transcription factor dimerization networks could play a key role in enabling cell type specific gene regulatory responses to signals. For example, members of the bHLH transcription factor family share similar helix-loop-helix domains that allow the formation of many different dimers, as many as 30 in some systems.⁶³ These dimers can exhibit different transcriptional activities and DNA binding site preferences. Further, some bHLH factors lack DNA binding domains altogether, and therefore generate transcriptionally inactive dimers. Consistent with the mathematical models and other examples above, this combinatorial complex formation could allow cell context-dependent computational gene regulation responses. For example, key developmental pathways activate the expression of bHLH transcription factors, which can then form various homo- and heterodimers with themselves and one another. In this way, they provide a layer of combinatorial dimerization between signaling pathways and their genomic targets. For example, the *Hes* and *Her* genes are bHLH transcription factors that are activated by Notch, FGF, and other signaling pathways.⁶⁴ Once expressed, these transcription factors can then form dimers regulating the expression of downstream targets, as well as their own expression.

In many developmental processes, *Hes* and *Her* proteins negatively regulate their own expression, producing complex regulation of downstream gene targets. In somitogenesis, this autoregulation leads to synchronized oscillations in expression, which are necessary for the repetitive process of somitogenesis.⁶⁵ In neurogenesis, additional factors controlling the strength of this feedback can produce different outcomes in embryonic and adult stem cells.⁶⁶ However, though the negative feedback of *Hes* and *Her* genes plays a clear role in both somitogenesis and neurogenesis, it was insufficient to explain all the observed patterns of gene expression in these two processes.^{33,67} This suggests that combinatorial interactions between bHLHs could play a key role in determining gene activation and dynamics. Below, we describe two examples where explicitly accounting for bHLH dimerization was necessary to explain the combinatorial and context-dependent effects of *Hes* and *Her* on their target genes.

In somitogenesis, knocking out *Her1* or *Her7* produces distinct effects, even though they promiscuously bind similar targets and have similar effects when overexpressed.³³ A careful study of *Her1* and *Her7* in zebrafish development revealed that they form many dimers, binding to themselves, each other, and a third binding partner, *Hes6*. This confirmed an earlier prediction of the existence of *Her1-Her7* heterodimers to explain observed effects in

zebrafish.⁶⁸ Most of the formed dimers are transcriptionally inactive. Only the *Her7:Hes6* heterodimer and the *Her1:Her1* homodimer have strong DNA binding activity and can regulate downstream expression of *Her1* and *Her7*. Thus, while *Her1* and *Her7* bind similar DNA regions, their sequestration in a “dimer cloud” and competition for shared binding partners give them distinct roles in controlling gene expression, as *Her7*, but not *Her1*, is required for certain oscillations.

Dimerization of bHLHs can also be used to explain their context-dependent effects on their gene targets. Neural stem cells (NSCs) readily proliferate during embryonic development but are largely quiescent during adult neurogenesis. The *Hes* bHLH genes have two regulatory roles in this transition to quiescence: inhibiting their own expression and inhibiting the expression of proneural factors. Combinatorial bHLH dimerization is key to explaining how this circuit can produce different degrees of quiescence between the adult and embryonic states.⁶⁷ Specifically, Inhibitor of DNA-binding factors (IDs) have a helix-loop-helix domain that allows them to dimerize with bHLHs, but lack the basic domain that promotes DNA binding. Thus, IDs can dimerize with *Hes* proteins and alter their DNA binding properties. As a result, *Hes*-ID dimers can perform only one of *Hes*'s two regulatory roles, as they can inhibit expression of proneural factors but do not inhibit expression of *Hes* itself. This suggests that the expression of ID factors in adult NSCs stabilizes *Hes* expression, removing oscillations produced by its negative autoregulation, while more strongly repressing proneural factors, increasing the degree of quiescence.

Together, these examples highlight the complex gene regulation requirements for mammalian development and how bHLH dimerization supports those higher level functions. But key questions remain unclear: What computations are possible with bHLH and other combinatorial transcription factor dimerization architectures, and how do those computations enable developmental and physiological cell behaviors? What particular functions do these computations support, outside of achieving diverse patterns of gene expression? For example, dimerization networks can decouple the control of different aspects of circuit output, such as the period or amplitude of oscillations in somitogenesis.³³ Moreover, control of bHLH gene expression is also crucial to buffering these oscillations from the effects of bursty transcription.^{69,70} How do these promiscuous dimerization networks integrate with other complex circuit features, such as autoregulation to produce oscillations of shared components? Experimental work is needed to systematically map bHLH dimerization networks, identify the functions they compute, and determine how those functions change depending on which factors are expressed in any given cell context. Moreover, the role of competitive dimerization in tuning DNA-binding and transcription factor activity can be applied to other transcription factor families, such as nuclear hormone receptors, which also exhibit complex dimerization patterns.^{71,72} Such studies would allow us to understand transcription factor dimerization networks as predictable computational devices, and better control their behavior.

Synthetic combinatorial protein networks enable scalable multistability

Previously, most synthetic biological circuits have operated at the level of individual cells. However, recent advances have begun to enable the design of multicellular circuits that

allow cells to operate in multiple states, communicate specifically with one another, and spatially organize into tissue-like structures. These multicellular behaviors will require increased computational capabilities, paralleling those that arose during the corresponding evolutionary transition to multicellularity. The same principles of computation by combinatorial protein networks could enable such synthetic circuits.

For example, we recently sought to create a synthetic cell fate control system that would generate multiple stable states.⁷³ To choose a design, we were inspired by two prevalent features of natural cell fate control systems: First, natural systems use promiscuous dimerization among transcription factors, such as bHLH factors, to generate a variety of dimers with different DNA binding specificities and activities, including inactive complexes^{18,74,75} (Figure 2C). Second, they often include positive autoregulatory feedback loops in which dimers directly or indirectly activate expression of their own components. A minimal circuit design, termed MultiFate, embodies those two principles. It is based on a set of engineered zinc finger transcription factors that can promiscuously homo- and heterodimerize to form all possible dimers (Figure 2D, left panel). Homodimers positively regulate their own expression while heterodimers are inactive. Mathematical modeling showed that this circuit design could produce multiple stable attractors in the space of transcription factor concentrations (Figure 2D, right panel). Further, the number of states initially scales exponentially as $2^N - 1$, where N is the number of factors. This scaling is enabled in part by the use of dimerization for cross-inhibition among transcription factors. Eventually, this scaling becomes limited by the more rapid growth of inactive heterodimers compared to active homodimers. Consistent with the model, adding additional transcription factors to an existing experimental circuit expanded the number of states without re-engineering the existing system, with 2- and 3-factor circuits respectively generating 3 and 7 stable states. The synthetic system also recapitulated other aspects of natural cell fate control systems, such as progressive and irreversible differentiation. This work shows how combinatorial dimerization can play a key computational role in enabling multicellularity.

Studying promiscuous protein-protein interaction systems

Many biological systems possess the elements necessary for computation through combinatorial binding (Figure 3A, cf. Box 1). However, predicting and controlling the capabilities of these systems poses particular challenges and opportunities. First, these systems can contain many paralogs, whose subtle functional differences are not apparent in their highly similar structures and sequences. Second, the large number of combinatorial interactions, even for modest numbers of components, requires commensurately large datasets for qualitative mapping and quantitative measurement. Third, design of new modified protein variants could potentially expand network computations, but requires models that relate component sequences to their interaction parameters. Initial work has begun to address these three issues.

Classifying inputs by their effective interactions

The maintenance of multiple protein variants within networks suggests that they could perform unique functions, but differences between these components can be subtle and

difficult to determine. Often, they have similar sequences and structures that do not provide obvious clues to the subtle differences in binding or activity that allow them to produce distinct effects. In other fields, comprehensive measurements of pairwise interactions have allowed functional systems-level classification of molecular components. For example, measuring the effects of a set of antibiotics, alone and in all pairwise combinations, on the growth of *E. coli* identified antagonistic and synergistic interactions.⁷⁶ These interactions were then used to classify antibiotics into equivalence groups, where drugs exhibiting the same pattern of interactions with other drugs are grouped together. Remarkably, this epistatic classification scheme—lacking any direct molecular information—matched that based on the drugs' known biochemical mechanisms of action.

Inspired by such studies, we asked whether a similar functional classification of BMP ligands could be obtained by systematically and quantitatively analyzing BMP pathway responses in several cell lines to 10 BMP ligands, both individually and in all pairwise combinations.⁴³ In mouse embryonic stem cells (mESCs), many of the 10 ligands studied combined additively with one another, as if they were functionally interchangeable. However, even ligands that combined additively with one another sometimes interacted differently with other ligands. For example, BMP7 and BMP9 combined additively with each other, but either additively or synergistically, respectively, with BMP4, putting them into distinct equivalence groups. Altogether, this approach classified the full set of 10 ligands into 5 equivalence groups, with ligands within a group sharing both a similar individual potency and a similar pattern of effective interactions with other ligands (Figure 3B).

Are these equivalence groups intrinsic properties of the ligands, or are they contextual, depending on properties of the cell line and its receptor expression profile? In a second cell line, NMuMG, which exhibits a different BMP receptor expression profile, the ligands produced different pairwise responses. For example, BMP9 and BMP4 synergized in mESCs, but combined additively in NMuMG cells. More generally, NMuMG cells exhibited more antagonistic interactions between ligands. While the ligands could again be classified into five equivalence groups, these groups differed from those obtained in mES cells. More generally, expanding this analysis to a panel of five additional cell lines differing in their receptor expression profiles revealed the strong effect of receptor expression on ligand equivalence, with each of the 7 cell lines, some differing only in their receptor expression profiles, producing a distinct equivalence group classification. Evidently, cell context strongly impacts ligand equivalence.

This large dataset also provided insight into the effects of specific receptors on multi-ligand computation. For example, consistent with known ligand-receptor interactions, the BMPRII receptor mediates signaling by the ligands GDF5, GDF6, and GDF7 by converting their antagonistic interactions to additive ones. Similarly, the ACVR1 and ACVRL1 receptors relieve non-additive interactions of BMP10 and BMP9, respectively. This phenotypic classification also proved useful in explaining previous results in the literature. For example, BMP9 was shown to functionally replace BMP10 during vasculature formation but not heart development.⁷⁷ This difference in the equivalence of BMP9 and BMP10 is consistent with observations that only when the receptor ACVRL1 is expressed, as it is in endothelial

cells, do the two ligands fall into the same equivalence group. Taken together, these results underscore that BMP signaling is both combinatorial and contextual, as well as the power of effective interactions for revealing the spectrum of computational power in a given pathway and for individual ligands.

Mapping and quantifying protein-protein interactions

A key step in building a predictive model of combinatorial computations requires identifying which protein-protein complexes can form. However, even small changes in amino acid sequence can drastically affect protein binding affinity.⁷⁸ As a result, one cannot necessarily predict how strongly one pair of protein variants based on the binding strengths of their paralogs.

For example, Wnt signaling provides powerful instructive cues during development. The pathway includes 10 ligand variants, each of which could potentially bind and activate any of 9 Fzd receptor variants. In fact, however, the signaling activities of the ligands differ from one another and vary across different cell contexts.^{11,12} To build a mechanistic understanding of the functional differences between Wnt ligands and Fzd receptors, Voloshanenko *et al.* engineered cells to express each Fzd receptor variant individually and then measured the activation of each receptor by each Wnt ligand.⁷⁹ This map showed that each ligand activated multiple receptors and that most receptors were activated by multiple ligands (Figure 3C). (Note that two receptors were not activated by any ligand and may provide other, possibly inhibitory, functions.) Moreover, many structurally homologous ligands, such as Wnt9a and Wnt9b, activated different sets of receptors. Thus, this approach revealed counter-intuitive interactions and revealed the diverse capabilities of these apparently similar ligand and receptor variants.

An accurate prediction of combinatorial computations requires knowing not only which complexes can form, but their relative affinities of formation and their output-producing activities (k_{ij} and e_{ij} in Box 1 and Figure 3A). Lucarelli *et al.* built such a model to predict the gene-specific dynamics of multiple targets of the TGF β pathway. In TGF β signaling, three variants of the Smad effector, Smad2/3/4, promiscuously assemble in different combinations to generate a family of 10 possible trimeric transcription factor complexes that differ in their genomic targets and activities. Predicting how the distribution of these trimers depends on the expression levels of the three components is essential for understanding computation within the pathway. Lucarelli *et al.* used mass spectrometry and co-immunoprecipitation to quantify the phosphorylation dynamics of Smad monomers.⁸⁰ The authors then used these data to fit a model in which monomers undergo pairwise association and dissociation (Figure 3D). Because each Smad2 and Smad3 monomer can be phosphorylated on two sites, many molecularly distinct complexes are possible. In fitting the model, they therefore used a standard regularization technique^{81,82} to limit the number of nonzero parameters and focus on key Smad complexes. The resulting model predicted that three crucial Smad complexes, out of the possible 10, drove most target gene expression. Critically, the fit model could predict TGF β activation of gene expression in other cell lines based solely on Smad expression levels. It could also make reverse predictions, correctly inferring that a tumor sample had elevated levels of Smad activation overall based on changes in its

gene expression. These results provide hope that limited numbers of measurements can be sufficient to enable predictive, mechanistic modeling of combinatorial protein networks.

Redesigning component sequence to produce new behaviors

Machine learning methods show promise for identifying subtle patterns in the behavior of combinatorial protein networks and guiding the design of new signaling components. A clear example of this is recent work with GPCRs.¹⁵ These receptors promiscuously activate a variety of G-proteins, via interactions with the C-terminal tail of the $G\alpha$ subunit. The authors quantified pathway activation by all possible pairs of 148 human GPCR and 11 $G\alpha$ C-terminal domain variants, each fused to a common backbone. The sigmoidal signaling response to varying ligand concentrations allowed the authors to quantify the coupling between each receptor and each $G\alpha$ variant. Training a machine learning model with these data identified sequence-specific coupling features, which the authors then used to forward design GPCR variants that coupled with a particular $G\alpha$ variant of interest (Figure 3E). These results demonstrated how quantitative input-output measurements contain sufficient information to guide the design of new signaling components. Further, even though the underlying system is promiscuous, these results show that comprehensive analysis of promiscuous interactions, together with machine learning approaches, can enable the design of components with greater specificity.

Conclusions and future directions

In his 1995 review, “Protein molecules as computational elements in living cells,” Dennis Bray wrote, “Because proteins...integrate inputs and produce outputs it seems inescapable that the highly interconnected network of protein-based pathways in living cells will share some of the properties of neural nets.”²⁸ Here, we have seen how combinatorial protein networks can provide a versatile architecture for realizing neural-like computation at the protein level. Mounting experimental evidence shows that these systems can compute combinatorial functions of multiple input proteins. Furthermore, the computation can be reprogrammed by modulating the expression levels of circuit components. This ability allows cells to tailor computations to cell state and context, and support higher level functions such as addressing and self-avoidance. Critically, the elements necessary for combinatorial computation are present in a huge range of systems, the vast majority of which have not yet been systematically analyzed from this point of view.

In computer science, different neural network architectures have been developed to perform different kinds of computational tasks.⁸³ In a similar way, different combinatorial protein network architectures may be optimized to perform different classes of computation. For example, bHLH dimers often autoregulate their own expression, introducing direct feedback loops that allow the generation of stable attractors, while developmental signaling systems like BMP can use computation to selectively respond to different ligand combinations. Similarly, biological pathways may also operate over multiple timescales. For example, BMP can transmit signals on timescales of minutes to hours. Over the longer timescales of differentiation, cells can alter receptor expression, dynamically changing the ligand combinations they sense. Analysis of single-cell trajectories in developmental time courses

could shed more light on such dynamic behaviors and on their potential to alter computational behavior.

Combinatorial protein networks differ in fundamental ways from their artificial neural network analogs. Because proteins have finite concentrations in cells, competition for binding can lead to indirect (or nonlocal) interactions across a set of promiscuously dimerizing proteins. That is, an increase in the level of one complex necessarily reduces the concentrations of alternative complexes. It would be interesting to understand how this architectural difference impacts the functional expressivity of the network, and whether the principles of combinatorial protein networks could be usefully imported into artificial neural networks for engineering applications.

Above, we focused on a single set, or “layer,” of promiscuously interacting proteins. Superficially, this contrasts with deep multi-layer neural network architectures. However, biological circuits can also contain multiple distinct layers of combinatorial dimerization, with one layer regulating the next. For example, the promiscuous ligand-receptor interactions in the BMP signaling system (layer 1) lead to promiscuous phosphorylation of a set of distinct Smad proteins by distinct signaling complexes (layer 2). These in turn promiscuously multimerize to form a variety of transcription factor complexes (layer 3). Finally, these complexes exhibit many-to-many binding relations with different target sites on the chromosome (layer 4). In this way, biological circuits achieve multiple layers of computation. It will be interesting to determine which aspects of the overall computation are performed at each level, and why. It may also be important to consider how computations are deployed across multicellular systems. Contextuality potentially allows each cell type to tailor the computation of a given pathway to its own needs. Experiments and models will both be central to exploring how computational modules within individual cells compose to generate more complex, and spatially extended, tissue level functions at a larger scale.⁸⁴

The paradigm of combinatorial protein networks can support at least two types of potential applications. First, the ability to develop predictive quantitative models of these systems could open up the ability to control cell behaviors with greater precision. Cell atlas projects have revealed the receptor expression profiles of most cell types. Promiscuous pathways are often involved in disease and targets for drug development. For example, mutations in the ACVR1 BMP receptor cause the disease fibrodysplasia ossificans progressiva (FOP). The discovery that these mutations cause disease by changing a non-signaling ligand-receptor complex into an signaling complex introduced important new approaches in the development of FOP therapeutics^{44,85}. Understanding how potential perturbations of this and other pathways in each cell type and tissue environment change ligand perception could enable more rational design of therapeutic interventions. A second class of applications of this paradigm will undoubtedly unfold in synthetic biology, where systems of ever greater computational capacity are needed to enable more specific and controllable cell therapies.⁸⁶⁻⁹¹ The case of the MultiFate architecture discussed above shows a relatively simple example in which combinatorial protein networks can enable a complex dynamical behavior—multistability. But to program more complex multicellular systems, eventually rivaling those of the natural immune system, for instance, many more types of rapid protein-level computation will be needed.

The principal challenge in both analyzing and synthetically implementing combinatorial protein networks is the explosion of interactions and component expression contexts. Analytically, one approach is based on minimal, effective biophysical models that represent the essential components and interactions (Box 1), and are parameterized with empirical measurements. This approach requires systematic characterization of components, and relies on approximations or simplifications that may not be universally valid across all cell contexts. Alternatively, machine learning approaches based on high throughput measurements could potentially provide a mechanism-independent way to predict behaviors. These, too, suffer from the need for large data sets and may be limited in their predictive ability by the contents of the training data set. While such statistical models may not be immediately interpretable, they could nevertheless be valuable for design of experiments and therapeutics. For both modeling approaches, a key challenge is grappling with the strong contextuality of dimerization networks. Different cell types express different levels of circuit components, potentially generating a whole repertoire of different functional behaviors. It will therefore be critical to ensure that models can accurately predict circuit behavior across cell contexts.

Synthetically, designing combinatorial protein networks faces the challenges of imposing defined many-to-many interactions on a set of engineered protein components, and then regulating those components in different cell types. Engineering of combinatorial protein networks will benefit from growing libraries of well-characterized protein interaction domains,⁹²⁻⁹⁴ and emerging methods to engineer complex protein-level circuits.^{31,95-98} However, methods to assemble larger circuits in a more systematic way will be pivotal in expanding synthetic mammalian biology to multicellularity.

How “special” is the exact architecture and parameter values that characterize any given combinatorial protein network? At one extreme, it could be that in order to perform the required range of signal processing tasks within the organism, a combinatorial protein network needs to fine-tune its structure, affinities, activities, and other biochemical parameters. If one could replay the tape of evolution, in Stephen Jay Gould’s hypothetical metaphor, a closely equivalent network would have evolved with nearly identical parameters. On the other hand, it is possible that many different network topologies could provide the full diversity of computational functions for many or most possible parameter values. In this case, evolution might have stumbled upon a very different network design or parameter regime that nevertheless performs equivalent computations. In this case, the particular details of the pathways we observe would represent a contingent evolutionary artifact. Either way, it would be interesting to know whether the computational capacity of the system was more conserved or constrained than the molecular implementation.

Unlocking the full potential of promiscuous protein computation would provide a new lens to understand, predict, control, and ultimately program biological systems. Achieving this goal will require a combination of mathematical modeling, quantitative experimental analysis, and synthetic biology approaches. It will also require deeper analysis of the systems discussed above and extension of the paradigm to other systems that remain poorly characterized from this point of view. In the longer term, it will be interesting to see to what

extent a unified paradigm can explain a broad diversity of promiscuous molecular circuits, built out of proteins and potentially other biological molecules as well.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

H.E.K. is a Damon Runyon Fellow supported by the Damon Runyon Cancer Research Foundation (DRG-2472-22). J.G.-O. is supported by project PID2021-127311NB-I00 financed by the Spanish Ministry of Science and Innovation, the Spanish State Research Agency and FEDER (MICIN/AEI/10.13039/501100011033/FEDER), by the Maria de Maeztu Programme for Units of Excellence in R&D (project CEX2018-000792-M), and by the Generalitat de Catalunya (ICREA Academia programme). Work in the lab of MBE was supported by the National Institutes of Health grants R01 HD075335A and R01 MH116508, by the Paul G. Allen Frontiers Group and Prime Awarding Agency under Award No. UWSC10142, and by the National Science Foundation grant EF-2021552 under subaward UWSC10278. Y.E.A. is supported by the Israel Science Foundation (grant No. 1105/20) and is the incumbent Sygnet Career Development Chair for Bioinformatics.

References:

- Carretero-Paulet L, Galstyan A, Roig-Villanova I, Martínez-García JF, Bilbao-Castro JR, and Robertson DL (2010). Genome-wide classification and evolutionary analysis of the bHLH family of transcription factors in Arabidopsis, poplar, rice, moss, and algae. *Plant Physiol.* 153, 1398–1412. 10.1104/pp.110.153593. [PubMed: 20472752]
- Yan J, Ma Z, Xu X, and Guo A-Y (2014). Evolution, functional divergence and conserved exon–intron structure of bHLH/PAS gene family. *Mol. Genet. Genomics* 289, 25–36. 10.1007/s00438-013-0786-0. [PubMed: 24202550]
- Wotton KR, Alcaine Colet A, Jaeger J, and Jimenez-Guri E (2013). Evolution and expression of BMP genes in flies. *Dev. Genes Evol* 223, 335–340. 10.1007/s00427-013-0445-9. [PubMed: 23595982]
- Marques CL, Fernández I, Viegas MN, Cox CJ, Martel P, Rosa J, Cancela ML, and Laizé V (2016). Comparative analysis of zebrafish bone morphogenetic proteins 2, 4 and 16: molecular and evolutionary perspectives. *Cell. Mol. Life Sci* 73, 841–857. 10.1007/s00018-015-2024-x. [PubMed: 26341094]
- Gul IS, Hulpiau P, Saeys Y, and van Roy F (2017). Evolution and diversity of cadherins and catenins. *Exp. Cell Res* 358, 3–9. 10.1016/j.yexcr.2017.03.001. [PubMed: 28268172]
- Granados AA, Kanrar N, and Elowitz MB (2022). Combinatorial expression motifs in signaling pathways. *bioRxiv*, 2022.08.21.504714. 10.1101/2022.08.21.504714.
- Nickel J, and Mueller TD (2019). Specification of BMP Signaling. *Cells* 8. 10.3390/cells8121579.
- Kakuda S, LoPilato RK, Ito A, and Haltiwanger RS (2020). Canonical Notch ligands and Fringes have distinct effects on NOTCH1 and NOTCH2. *J. Biol. Chem* 295, 14710–14722. 10.1074/jbc.RA120.014407. [PubMed: 32820046]
- Delgoffe GM, Murray PJ, and Vignali DAA (2011). Interpreting mixed signals: the cell's cytokine conundrum. *Curr. Opin. Immunol* 23, 632–638. 10.1016/j.coi.2011.07.013. [PubMed: 21852079]
- Morris R, Kershaw NJ, and Babon JJ (2018). The molecular details of cytokine signaling via the JAK/STAT pathway. *Protein Sci.* 27, 1984–2009. 10.1002/pro.3519. [PubMed: 30267440]
- Eubelen M, Bostaille N, Cabochette P, Gauquier A, Tebabi P, Dumitru AC, Koehler M, Gut P, Alsteens D, Stainier DYR, et al. (2018). A molecular mechanism for Wnt ligand-specific signaling. *Science* 361. 10.1126/science.aat1178.
- Nusse R, and Clevers H (2017). Wnt/ β -Catenin Signaling, Disease, and Emerging Therapeutic Modalities. *Cell* 169, 985–999. 10.1016/j.cell.2017.05.016. [PubMed: 28575679]
- Ornitz DM, and Itoh N (2015). The Fibroblast Growth Factor signaling pathway. *Wiley Interdiscip. Rev. Dev. Biol* 4, 215–266. 10.1002/wdev.176. [PubMed: 25772309]

14. Trenker R, and Jura N (2020). Receptor tyrosine kinase activation: From the ligand perspective. *Curr. Opin. Cell Biol* 63, 174–185. 10.1016/j.ceb.2020.01.016. [PubMed: 32114309]
15. Inoue A, Raimondi F, Kadji FMN, Singh G, Kishi T, Uwamizu A, Ono Y, Shinjo Y, Ishida S, Arang N, et al. (2019). Illuminating G-Protein-Coupling Selectivity of GPCRs. *Cell* 177, 1933–1947.e25. 10.1016/j.cell.2019.04.044. [PubMed: 31160049]
16. Rodríguez-Martínez JA, Reinke AW, Bhimsaria D, Keating AE, and Ansari AZ (2017). Combinatorial bZIP dimers display complex DNA-binding specificity landscapes. *Elife* 6. 10.7554/eLife.19272.
17. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, Chen X, Taipale J, Hughes TR, and Weirauch MT (2018). The Human Transcription Factors. *Cell* 172, 650–665. 10.1016/j.cell.2018.01.029. [PubMed: 29425488]
18. Neuhold LA, and Wold B (1993). HLH forced dimers: tethering MyoD to E47 generates a dominant positive myogenic factor insulated from negative regulation by Id. *Cell* 74, 1033–1042. 10.1016/0092-8674(93)90725-6. [PubMed: 7691411]
19. Balemans W, and Van Hul W (2002). Extracellular regulation of BMP signaling in vertebrates: a cocktail of modulators. *Dev. Biol* 250, 231–250. [PubMed: 12376100]
20. Feng S, Rastogi C, Loker R, Glassford WJ, Tomas Rube H, Bussemaker HJ, and Mann RS (2022). Transcription factor paralogs orchestrate alternative gene regulatory networks by context-dependent cooperation with multiple cofactors. *Nat. Commun* 13, 3808. 10.1038/s41467-022-31501-2. [PubMed: 35778382]
21. Sotomayor M, Gaudet R, and Corey DP (2014). Sorting out a promiscuous superfamily: towards cadherin connectomics. *Trends Cell Biol.* 24, 524–536. 10.1016/j.tcb.2014.03.007. [PubMed: 24794279]
22. Wojtowicz WM, Vielmetter J, Fernandes RA, Siepe DH, Eastman CL, Chisholm GB, Cox S, Klock H, Anderson PW, Rue SM, et al. (2020). A Human IgSF Cell-Surface Interactome Reveals a Complex Network of Protein-Protein Interactions. *Cell* 182, 1027–1043.e17. 10.1016/j.cell.2020.07.025. [PubMed: 32822567]
23. Carrillo RA, Özkan E, Menon KP, Nagarkar-Jaiswal S, Lee P-T, Jeon M, Birnbaum ME, Bellen HJ, Garcia KC, and Zinn K (2015). Control of Synaptic Connectivity by a Network of Drosophila IgSF Cell Surface Proteins. *Cell* 163, 1770–1782. 10.1016/j.cell.2015.11.022. [PubMed: 26687361]
24. Tsai TY-C, Sikora M, Xia P, Colak-Champollion T, Knaut H, Heisenberg C-P, and Megason SG (2020). An adhesion code ensures robust pattern formation during tissue morphogenesis. *Science* 370, 113–116. 10.1126/science.aba6637. [PubMed: 33004519]
25. Hertz J, Krogh A, and Palmer RG (2018). Introduction. *Introduction to the Theory of Neural Computation*, 1–10. 10.1201/9780429499661-1.
26. Marr D (1982). *Vision: A computational investigation into the human representation and processing of visual information* (W. H. Freeman and Company).
27. Arkin A, and Ross J (1994). Computational functions in biochemical reaction networks. *Biophys. J* 67, 560–578. 10.1016/S0006-3495(94)80516-8. [PubMed: 7948674]
28. Bray D (1995). Protein molecules as computational elements in living cells. *Nature* 376, 307–312. 10.1038/376307a0. [PubMed: 7630396]
29. Gerhart J, and Kirschner M (1997). *Cells, Embryos and Evolution* (John Wiley & Sons, Incorporated).
30. Jones TS, Oliveira SMD, Myers CJ, Voigt CA, and Densmore D (2022). Genetic circuit design automation with Cello 2.0. *Nat. Protoc* 17, 1097–1113. 10.1038/s41596-021-00675-2. [PubMed: 35197606]
31. Chen Z, Linton JM, Zhu R, and Elowitz MB (2022). A synthetic protein-level neural network in mammalian cells. *bioRxiv*, 2022.07.10.499405. 10.1101/2022.07.10.499405.
32. Steinbacher T, Kummer D, and Ebnet K (2018). Junctional adhesion molecule-A: functional diversity through molecular promiscuity. *Cell. Mol. Life Sci* 75, 1393–1409. 10.1007/s00018-017-2729-0. [PubMed: 29238845]

33. Schröter C, Ares S, Morelli LG, Isakova A, Hens K, Soroldoni D, Gajewski M, Jülicher F, Maerkl SJ, Deplancke B, et al. (2012). Topology and dynamics of the zebrafish segmentation clock core circuit. *PLoS Biol.* 10, e1001364. 10.1371/journal.pbio.1001364. [PubMed: 22911291]
34. Ammeux N, Housden BE, Georgiadis A, Hu Y, and Perrimon N (2016). Mapping signaling pathway cross-talk in *Drosophila* cells. *Proc. Natl. Acad. Sci. U. S. A* 113, 9940–9945. 10.1073/pnas.1610432113. [PubMed: 27528688]
35. Galstyan V, Funk L, Einav T, and Phillips R (2019). Combinatorial Control through Allosterity. *J. Phys. Chem. B* 123, 2792–2800. 10.1021/acs.jpcc.8b12517. [PubMed: 30768906]
36. Yan SF, D'Agati V, Schmidt AM, and Ramasamy R (2007). Receptor for Advanced Glycation Endproducts (RAGE): a formidable force in the pathogenesis of the cardiovascular complications of diabetes & aging. *Curr. Mol. Med* 7, 699–710. [PubMed: 18331228]
37. Agliari E, Altavilla M, Barra A, Dello Schiavo L, and Katz E (2015). Notes on stochastic (bio)-logic gates: computing with allosteric cooperativity. *Sci. Rep* 5, 9415. 10.1038/srep09415. [PubMed: 25976626]
38. de Ronde W, Rein ten Wolde P, and Mugler A (2012). Protein logic: a statistical mechanical study of signal integration at the single-molecule level. *Biophys. J* 103, 1097–1107. 10.1016/j.bpj.2012.07.040. [PubMed: 23009860]
39. Marzen S, Garcia HG, and Phillips R (2013). Statistical Mechanics of Monod–Wyman–Changeux (MWC) Models. *J. Mol. Biol* 425, 1433–1460. 10.1016/j.jmb.2013.03.013. [PubMed: 23499654]
40. Massagué J (1990). The transforming growth factor-beta family. *Annu. Rev. Cell Biol* 6, 597–641. 10.1146/annurev.cb.06.110190.003121. [PubMed: 2177343]
41. Antebi YE, Linton JM, Klumpe H, Bintu B, Gong M, Su C, McCardell R, and Elowitz MB (2017). Combinatorial Signal Perception in the BMP Pathway. *Cell* 170, 1184–1196.e24. 10.1016/j.cell.2017.08.015. [PubMed: 28886385]
42. Mueller TD, and Nickel J (2012). Promiscuity and specificity in BMP receptor activation. *FEBS Lett.* 586, 1846–1859. 10.1016/j.febslet.2012.02.043. [PubMed: 22710174]
43. Klumpe H, Langley MA, Linton JM, Su CJ, Antebi YE, and Elowitz MB (2022). The context-dependent, combinatorial logic of BMP signaling. *Cell Systems.* 10.1016/j.cels.2022.03.002.
44. Aykul S, Huang L, Wang L, Das NM, Reisman S, Ray Y, Zhang Q, Rothman N, Nannuru KC, Kamat V, et al. (2022). Anti-ACVR1 antibodies exacerbate heterotopic ossification in fibrodysplasia ossificans progressiva (FOP) by activating FOP-mutant ACVR1. *J. Clin. Invest* 132. 10.1172/JCI153792.
45. Olsen OE, Wader KF, Hella H, Mylin AK, Turesson I, Nesthus I, Waage A, Sundan A, and Holien T (2015). Activin A inhibits BMP-signaling by binding ACVR2A and ACVR2B. *Cell Commun. Signal* 13, 27. 10.1186/s12964-015-0104-z. [PubMed: 26047946]
46. Hiepen C, Jatzlau J, Hildebrandt S, Kampfrath B, Goktas M, Murgai A, Cuellar Camacho JL, Haag R, Ruppert C, Sengle G, et al. (2019). BMP2 acts as a gatekeeper to protect endothelial cells from increased TGFβ responses and altered cell mechanics. *PLoS Biol.* 17, e3000557. 10.1371/journal.pbio.3000557. [PubMed: 31826007]
47. Escalante-Chong R, Savir Y, Carroll SM, Ingraham JB, Wang J, Marx CJ, and Springer M (2015). Galactose metabolic genes in yeast respond to a ratio of galactose and glucose. *Proc. Natl. Acad. Sci. U. S. A* 112, 1636–1641. 10.1073/pnas.1418058112. [PubMed: 25605920]
48. Grosswendt S, Kretzmer H, Smith ZD, Kumar AS, Hetzel S, Wittler L, Klages S, Timmermann B, Mukherji S, and Meissner A (2020). Epigenetic regulator function through mouse gastrulation. *Nature* 584, 102–108. 10.1038/s41586-020-2552-x. [PubMed: 32728215]
49. He P, Williams BA, Trout D, Marinov GK, Amrhein H, Berghella L, Goh S-T, Plajzer-Frick I, Afzal V, Pennacchio LA, et al. (2020). The changing mouse embryo transcriptome at whole tissue and single-cell resolution. *Nature* 583, 760–767. 10.1038/s41586-020-2536-x. [PubMed: 32728245]
50. The COVID Tissue Atlas Consortium, Granados AA, Bucher S, Agrawal A, Song H, Chen AT, Peng T, Neff N, Pisco AO, Huang F, et al. (2022). Comprehensive characterization of the transcriptional response to COVID-19 in multiple organs reveals shared signatures across tissues. *bioRxiv*, 2022.05.31.493925. 10.1101/2022.05.31.493925.

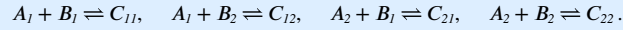
51. Consortium TTM, The Tabula Muris Consortium, Coordination, O., Coordination, L., Organ collection and processing, Library preparation and sequencing, Analysis, C.D., Annotation, C.T., Writing group, Supplemental text writing group, et al. (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* 562, 367–372. 10.1038/s41586-018-0590-4. [PubMed: 30283141]
52. Travaglini KJ, Nabhan AN, Penland L, Sinha R, Gillich A, Sit RV, Chang S, Conley SD, Mori Y, Seita J, et al. (2020). A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* 587, 619–625. 10.1038/s41586-020-2922-4. [PubMed: 33208946]
53. Maslov S (2008). Topological and Dynamical Properties of Protein Interaction Networks. In *Protein-protein Interactions and Networks: Identification, Computer Analysis, and Prediction*. A. Panchenko and Przytycka T, eds. (Springer London), pp. 115–137. 10.1007/978-1-84800-125-1_7.
54. Pickholtz RL, Milstein LB, and Schilling DL (1991). Spread spectrum for mobile communications. *IEEE Trans. Veh. Technol* 40, 313–322. 10.1109/25.289412.
55. Stüber GL *Principles of Mobile Communication* (Springer International Publishing) 10.1007/978-3-319-55615-4.
56. Morsut L, Roybal KT, Xiong X, Gordley RM, Coyle SM, Thomson M, and Lim WA (2016). Engineering Customized Cell Sensing and Response Behaviors Using Synthetic Notch Receptors. *Cell* 164, 780–791. 10.1016/j.cell.2016.01.012. [PubMed: 26830878]
57. Su CJ, Murugan A, Linton JM, Yeluri A, Bois J, Klumpe H, Antebi YE, and Elowitz MB (2022). Ligand-receptor promiscuity enables cellular addressing. *Cell Systems*. 10.1016/j.cels.2022.03.001.
58. Lawrence Zipursky S, and Grueber WB (2013). The Molecular Basis of Self-Avoidance. *Annu. Rev. Neurosci* 36, 547–568. 10.1146/annurev-neuro-062111-150414. [PubMed: 23841842]
59. Schreiner D, and Weiner JA (2010). Combinatorial homophilic interaction between γ -protocadherin multimers greatly expands the molecular diversity of cell adhesion. *Proc. Natl. Acad. Sci. U. S. A* 107, 14893–14898. 10.1073/pnas.1004526107. [PubMed: 20679223]
60. Thu CA, Chen WV, Rubinstein R, Chevee M, Wolcott HN, Felsovalyi KO, Tapia JC, Shapiro L, Honig B, and Maniatis T (2014). Single-cell identity generated by combinatorial homophilic interactions between α , β , and γ protocadherins. *Cell* 158, 1045–1059. 10.1016/j.cell.2014.07.012. [PubMed: 25171406]
61. Mountoufaris G, Canzio D, Nwakeze CL, Chen WV, and Maniatis T (2018). Writing, Reading, and Translating the Clustered Protocadherin Cell Surface Recognition Code for Neural Circuit Assembly. *Annu. Rev. Cell Dev. Biol* 34, 471–493. 10.1146/annurev-cellbio-100616-060701. [PubMed: 30296392]
62. Vu VH, Rahil Z, Sullivan BG, and Leckband DE (2020). Cadherin Complexes are Combinatorial Mechano-Switches that Differentially Regulate Cell Mechanics. *Biophys. J* 118, 250a.
63. Grove CA, De Masi F, Barrasa MI, Newburger DE, Alkema MJ, Bulyk ML, and Walhout AJM (2009). A multiparameter network reveals extensive divergence between *C. elegans* bHLH transcription factors. *Cell* 138, 314–327. 10.1016/j.cell.2009.04.058. [PubMed: 19632181]
64. Pourquié O (2011). Vertebrate segmentation: from cyclic gene networks to scoliosis. *Cell* 145, 650–663. 10.1016/j.cell.2011.05.011. [PubMed: 21620133]
65. Riedel-Kruse IH, Müller C, and Oates AC (2007). Synchrony dynamics during initiation, failure, and rescue of the segmentation clock. *Science* 317, 1911–1915. 10.1126/science.1142538. [PubMed: 17702912]
66. Ochi S, Imaizumi Y, Shimojo H, Miyachi H, and Kageyama R (2020). Oscillatory expression of *Hes1* regulates cell proliferation and neuronal differentiation in the embryonic brain. *Development* 147. 10.1242/dev.182204.
67. Boareto M, Iber D, and Taylor V (2017). Differential interactions between Notch and ID factors control neurogenesis by modulating *Hes* factor autoregulation. *Development* 144, 3465–3474. 10.1242/dev.152520. [PubMed: 28974640]
68. Cinquin O (2007). Repressor dimerization in the zebrafish somitogenesis clock. *PLoS Comput. Biol* 3, e32. 10.1371/journal.pcbi.0030032. [PubMed: 17305423]

69. Zinani OQH, Kesero lu K, Dey S, Ay A, Singh A, and Özbudak EM (2022). Gene copy number and negative feedback differentially regulate transcriptional variability of segmentation clock genes. *iScience* 25, 104579. 10.1016/j.isci.2022.104579. [PubMed: 35789861]
70. Zinani OQH, Kesero lu K, Ay A, and Özbudak EM (2021). Pairing of segmentation clock genes drives robust pattern formation. *Nature* 589, 431–436. 10.1038/s41586-020-03055-0. [PubMed: 33361814]
71. Amoutzias GD, Pichler EE, Mian N, De Graaf D, Imsiridou A, Robinson-Rechavi M, Bornberg-Bauer E, Robertson DL, and Oliver SG (2007). A protein interaction atlas for the nuclear receptors: properties and quality of a hub-based dimerisation network. *BMC Syst. Biol* 1, 34. 10.1186/1752-0509-1-34. [PubMed: 17672894]
72. Amoutzias GD, Robertson DL, Van de Peer Y, and Oliver SG (2008). Choose your partners: dimerization in eukaryotic transcription factors. *Trends Biochem. Sci* 33, 220–229. 10.1016/j.tibs.2008.02.002. [PubMed: 18406148]
73. Zhu R, Del Rio-Salgado JM, Garcia-Ojalvo J, and Elowitz MB (2022). Synthetic multistability in mammalian cells. *Science* 375, eabg9765. 10.1126/science.abg9765. [PubMed: 35050677]
74. Osborn DPS, Li K, Hinitz Y, and Hughes SM (2011). Cdkn1c drives muscle differentiation through a positive feedback loop with Myod. *Dev. Biol* 350, 464–475. 10.1016/j.ydbio.2010.12.010. [PubMed: 21147088]
75. Bhattacharya A, and Baker NE (2011). A network of broadly expressed HLH genes regulates tissue-specific cell fates. *Cell* 147, 881–892. 10.1016/j.cell.2011.08.055. [PubMed: 22078884]
76. Yeh P, Tschumi AI, and Kishony R (2006). Functional classification of drugs by properties of their pairwise interactions. *Nat. Genet* 38, 489–494. 10.1038/ng1755. [PubMed: 16550172]
77. Chen H, Brady Ridgway J, 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. 10.1073/pnas.1306074110. [PubMed: 23812757]
78. Marti-Solano M, Crilly SE, Malinverni D, Munk C, Harris M, Pearce A, Quon T, Mackenzie AE, Wang X, Peng J, et al. (2020). Combinatorial expression of GPCR isoforms affects signalling and drug responses. *Nature* 587, 650–656. 10.1038/s41586-020-2888-2. [PubMed: 33149304]
79. Voloshanenko O, Gmach P, Winter J, Kranz D, and Boutros M (2017). Mapping of Wnt-Frizzled interactions by multiplex CRISPR targeting of receptor gene families. *FASEB J.* 31, 4832–4844. 10.1096/fj.201700144R. [PubMed: 28733458]
80. Lucarelli P, Schilling M, Kreutz C, Vlasov A, Boehm ME, Iwamoto N, Steiert B, Lattermann S, Wäsch M, Stepath M, et al. (2018). Resolving the Combinatorial Complexity of Smad Protein Complex Formation and Its Link to Gene Expression. *Cell Syst* 6, 75–89.e11. 10.1016/j.cels.2017.11.010. [PubMed: 29248373]
81. Merkle R, Steiert B, Salopiata F, Depner S, Raue A, Iwamoto N, Schelker M, Hass H, Wäsch M, Böhm ME, et al. (2016). Identification of Cell Type-Specific Differences in Erythropoietin Receptor Signaling in Primary Erythroid and Lung Cancer Cells. *PLoS Comput. Biol* 12, e1005049. 10.1371/journal.pcbi.1005049. [PubMed: 27494133]
82. Steiert B, Timmer J, and Kreutz C (2016). L1 regularization facilitates detection of cell type-specific parameters in dynamical systems. *Bioinformatics* 32, i718–i726. 10.1093/bioinformatics/btw461. [PubMed: 27587694]
83. LeCun Y, Bengio Y, and Hinton G (2015). Deep learning. *Nature* 521, 436–444. 10.1038/nature14539. [PubMed: 26017442]
84. Regot S, Macia J, Conde N, Furukawa K, Kjellén J, Peeters T, Hohmann S, de Nadal E, Posas F, and Solé R (2011). Distributed biological computation with multicellular engineered networks. *Nature* 469, 207–211. 10.1038/nature09679. [PubMed: 21150900]
85. Williams E, Bagarova J, Kerr G, Xia D-D, Place ES, Dey D, Shen Y, Bocobo GA, Mohedas AH, Huang X, et al. (2021). Saracatinib is an efficacious clinical candidate for fibrodysplasia ossificans progressiva. *JCI Insight* 6. 10.1172/jci.insight.95042.
86. Wu M-R, Jusiak B, and Lu TK (2019). Engineering advanced cancer therapies with synthetic biology. *Nat. Rev. Cancer* 19, 187–195. 10.1038/s41568-019-0121-0. [PubMed: 30837696]

87. Kitada T, DiAndreth B, Teague B, and Weiss R (2018). Programming gene and engineered-cell therapies with synthetic biology. *Science* 359. 10.1126/science.aad1067.
88. Cubillos-Ruiz A, Guo T, Sokolovska A, Miller PF, Collins JJ, Lu TK, and Lora JM (2021). Engineering living therapeutics with synthetic biology. *Nat. Rev. Drug Discov* 20, 941–960. 10.1038/s41573-021-00285-3. [PubMed: 34616030]
89. Scheller L, and Fussenegger M (2019). From synthetic biology to human therapy: engineered mammalian cells. *Curr. Opin. Biotechnol* 58, 108–116. 10.1016/j.copbio.2019.02.023. [PubMed: 30933864]
90. Johnson MB, March AR, and Morsut L (2017). Engineering multicellular systems: using synthetic biology to control tissue self-organization. *Curr Opin Biomed Eng* 4, 163–173. 10.1016/j.cobme.2017.10.008. [PubMed: 29308442]
91. Markson JS, and Elowitz MB (2014). Synthetic biology of multicellular systems: new platforms and applications for animal cells and organisms. *ACS Synth. Biol* 3, 875–876. 10.1021/sb500358y. [PubMed: 25524091]
92. Boyken SE, Chen Z, Groves B, Langan RA, Oberdorfer G, Ford A, Gilmore JM, Xu C, DiMaio F, Pereira JH, et al. (2016). De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. *Science* 352, 680–687. 10.1126/science.aad8865. [PubMed: 27151862]
93. Chen Z, Boyken SE, Jia M, Busch F, Flores-Solis D, Bick MJ, Lu P, VanAernum ZL, Sahasrabudhe A, Langan RA, et al. (2019). Programmable design of orthogonal protein heterodimers. *Nature* 565, 106–111. 10.1038/s41586-018-0802-y. [PubMed: 30568301]
94. Lebar T, Lainš ek D, Merljak E, Aupi J, and Jerala R (2020). A tunable orthogonal coiled-coil interaction toolbox for engineering mammalian cells. *Nat. Chem. Biol* 16, 513–519. 10.1038/s41589-019-0443-y. [PubMed: 31907374]
95. Chen Z, and Elowitz MB (2021). Programmable protein circuit design. *Cell* 184, 2284–2301. 10.1016/j.cell.2021.03.007. [PubMed: 33848464]
96. Gao XJ, Chong LS, Kim MS, and Elowitz MB (2018). Programmable protein circuits in living cells. *Science* 361, 1252–1258. 10.1126/science.aat5062. [PubMed: 30237357]
97. Chung HK, Zou X, Bajar BT, Brand VR, Huo Y, Alcludia JF, Ferrell JE Jr, and Lin MZ (2019). A compact synthetic pathway rewires cancer signaling to therapeutic effector release. *Science* 364. 10.1126/science.aat6982.
98. Fink T, and Jerala R (2022). Designed protease-based signaling networks. *Curr. Opin. Chem. Biol* 68, 102146. 10.1016/j.cbpa.2022.102146. [PubMed: 35430555]

Box 1:**Mathematical model of computation by promiscuous protein dimerization networks**

Promiscuous protein dimerization networks can implement diverse computations. To see how this works, consider a minimal system, inspired by signaling pathways, where one set of protein variants, A_i , analogous to ligands, acting as inputs, can bind to a second set of components, B_j , analogous to receptors, to form a combinatorial set of binary signaling complexes, C_{ij} . A similar formalism can represent other combinatorial dimerization and more complex multimerization networks. More specifically, we consider a system with two A and two B variants, generating four binding reactions (Box Figure 1A):



At equilibrium, these reactions lead to the equations $A_i B_j = k_{ij} C_{ij}$ where k_{ij} denotes the affinity of binding of A_i with B_j . We also assume that the total abundances of the components are at steady-state, i.e. that complex formation and dissociation are fast compared to the timescales of protein production and removal. Additionally, for analytical simplicity, we consider a regime in which the A variants are supplied at a fixed concentration in a large extracellular volume, as could occur in the context of signaling, so that their concentrations are negligibly perturbed by binding to B variants. However, similar functional behaviors can be obtained without this assumption. We then have the following conservation laws for the total B abundances, B_{1T} and B_{2T} :

$$B_1 + C_{11} + C_{21} = B_{1T} \quad B_2 + C_{12} + C_{22} = B_{2T}.$$

From these equations, we obtain expressions for the concentrations of the free B proteins as a function of the A concentrations:

$$B_1 = \frac{B_{1T}}{1 + k_{11}A_1 + k_{21}A_2}, \quad B_2 = \frac{B_{2T}}{1 + k_{12}A_1 + k_{22}A_2}. \quad (1)$$

Assuming each complex, C_{ij} signals with a specific activity, e_{ij} , the total output signal, S , is the sum of contributions from each complex,

$$S = e_{11}C_{11} + e_{12}C_{12} + e_{21}C_{21} + e_{22}C_{22}.$$

Finally, using the conservation laws (1), the expression for free receptors (2), and the equilibrium equations above, we can express the system output, S , in terms of the concentrations of the A variants and the total abundances of B_1 and B_2 .

$$S = \frac{e_{11}B_{1T}k_{11}A_1}{I + k_{11}A_1 + k_{21}A_2} + \frac{e_{12}B_{2T}k_{12}A_1}{I + k_{12}A_1 + k_{22}A_2} + \frac{e_{21}B_{1T}k_{21}A_2}{I + k_{11}A_1 + k_{21}A_2} + \frac{e_{22}B_{2T}k_{22}A_2}{I + k_{12}A_1 + k_{22}A_2}. \quad (2)$$

This simple model can represent at least four different classes of computation: additive, ratiometric, balance and imbalance responses (Box Figure 1B). For instance, when the activities of C_{12} and C_{21} are small compared to those of C_{11} and C_{22} (when each A component strongly activates a distinct B component), the system response can be approximated by the sum of two increasing functions, each hyperbolically increasing with one A variant and cross-inhibited by the other:

$$S \approx \frac{e_{11}B_{1T}k_{11}A_1}{I + k_{11}A_1 + k_{21}A_2} + \frac{e_{22}B_{2T}k_{22}A_2}{I + k_{12}A_1 + k_{22}A_2}$$

In this limit, when the affinity coefficients have similar values, the system performs an *additive* computation (see Box Figure 1B, left). On the other hand, when the values of these coefficients differ from each other, and are chosen appropriately, the system can exhibit a *balance* or an *imbalance* response (Box Figure 1B, center). When e_{21} and e_{22} are small, the response of the circuit depends approximately on the ratio between A_2 and A_1 :

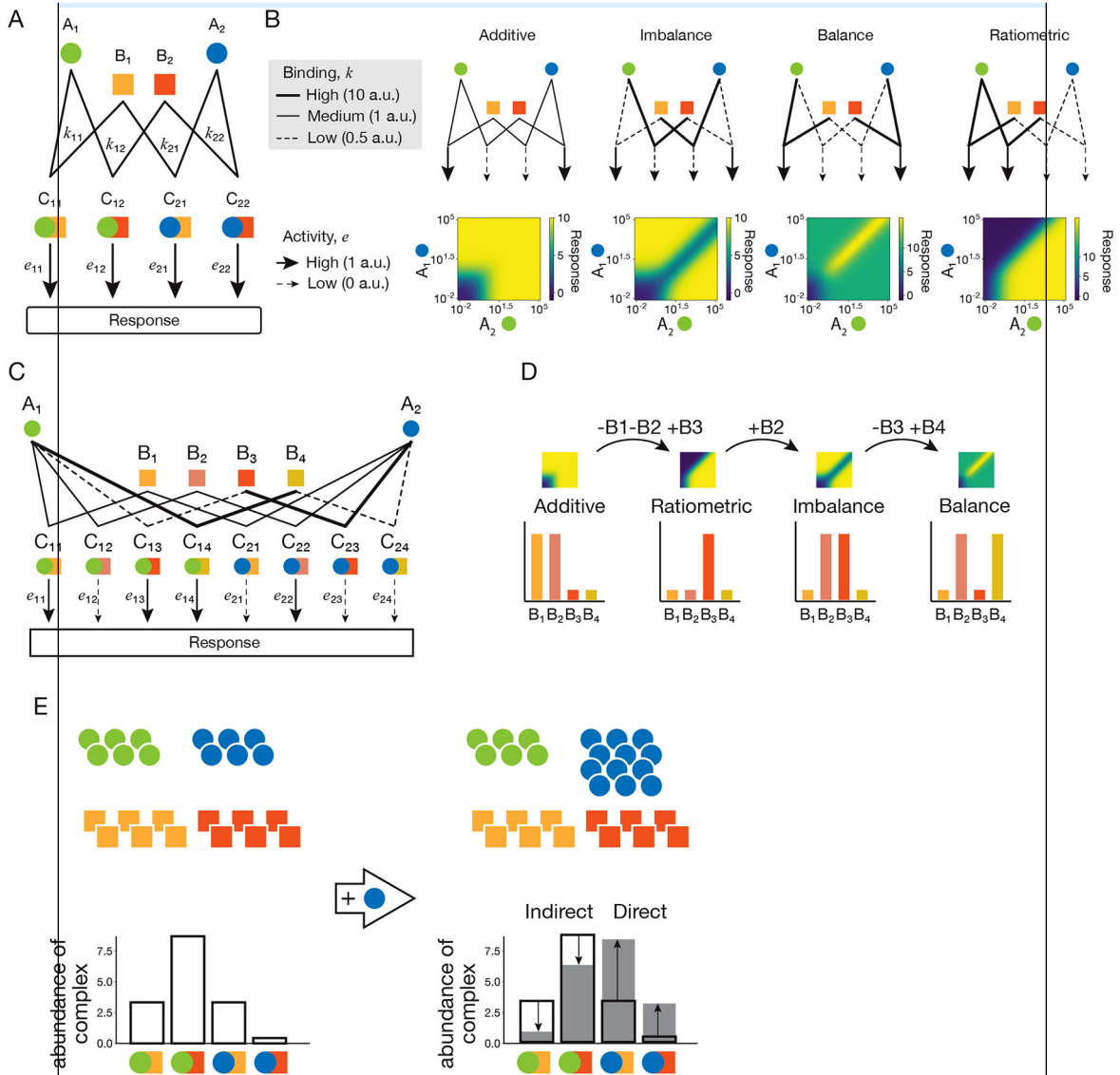
$$S \approx \frac{e_{11}B_{1T}k_{11}A_1}{I + k_{11}A_1 + k_{21}A_2} + \frac{e_{12}B_{2T}k_{12}A_1}{I + k_{12}A_1 + k_{22}A_2} \approx \frac{e_{11}B_{1T}k_{11} + e_{12}B_{2T}k_{12}}{k_{11} + k_{21}(A_2 / A_1)}$$

where we have considered that $k_{11} = k_{12}$ and $k_{21} = k_{22}$, and have assumed $k_{ij}A_k \gg I$. This behavior corresponds to a *ratiometric* computation (see Box Figure 1B, right).

In summary, this minimal model shows that specific sets of parameter values enable even simple promiscuous protein networks to operate in different computational modes. In particular, the four computational modes described above (additive, ratiometric, balance, and imbalanced) can be reached by selecting subsets of B components from a pool of only four variants, which interact with a single pair of A components (Box Figure 1C,D). This feature can be visualized by smoothly varying B levels and observing the resulting changes in computation (Supplementary Movie 1). In that way, computations in the A space depend on the context defined by the B levels. This model highlights the central importance of differences in parameters such as affinity, activity, and expression levels, which can be mediated by diverse mechanisms such as post-translational modification and expression-mediated feedbacks. Importantly, our model is not sensitive to the exact mechanism determining these parameters and can be expanded to explore more complicated protein binding architectures that lead to an even wider array of functions.

This simple model also reveals a feature of promiscuous protein networks that sets these computational systems apart from traditional neural networks. Specifically, because of the stoichiometric nature of the biochemical reactions, reflected in the conservation laws above, a change in the total abundance of one individual component can lead to a redistribution of other complexes (Box Figure 1E). This effect can be direct,

affecting complexes containing the component whose abundance was perturbed, or indirect, impacting complexes in which the perturbed component does not appear. The resulting redistribution of components across complexes is absent in neural networks, whose components are not subject to conservation laws, but functions to strengthen the computational capabilities of promiscuous protein interaction networks.



Box figure: Mathematical model of computation by promiscuous protein dimerization networks

(A) In a minimal promiscuous protein dimerization network, two A-type components (circles) interact with two B-type components (squares) to form complexes that induce a response. The binding affinities of the components A_i and B_j to form the complex C_{ij} are denoted by k_{ij} , and their response activity levels are denoted by e_{ij} .

(B) Heatmaps representing the response of the minimal network as a function of the A concentrations. Parameter values are indicated by line weights, as indicated in the legend (a.u. = arbitrary units). All levels of B-type components are set at 10 a.u. The code to generate these plots can be found at <https://github.com/dsb-lab/MinimalPromiscuousCircuit>.

(C) All four distinct classes of computations shown in (B) can be generated by a single model with specific biochemical parameters using only four B-type components. Parameter values are indicated by line weights as in (B).

(D) By changing the expression levels of the B-type components the model can transition smoothly between the four computation classes. See also Movie S1.

(E) When the concentration of a ligand is increased (blue circles), the stoichiometric nature of the biochemical reaction results in a global redistribution of complexes. Outline bars and shaded bars show the amount of each complex before and after the increase in ligand, respectively. The increase results in direct effects on complexes containing the varying ligand (marked 'Direct'). Additionally, the redistribution of complexes has indirect effects on complexes that do not contain the varying ligand (marked 'Indirect').

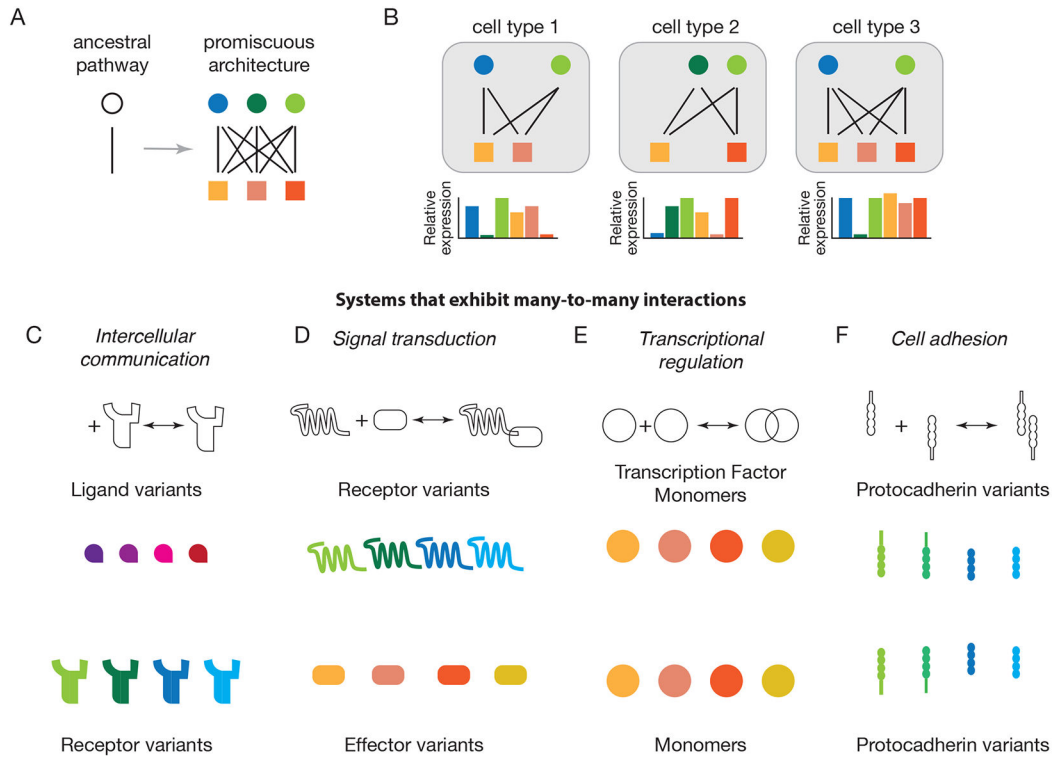


Figure 1: Diverse cellular pathways exhibit many-to-many interactions among protein families.
 (A) Extant pathways, such as signaling from one component (circle) to another (square) can evolve from simpler ancestral pathways through gene duplications, resulting in many-to-many interaction networks (right).
 (B) Different cell types (gray) typically express different protein variant profiles (schematic bar plots).
 (C) Intercellular communication systems often comprise multiple ligand (upper) and receptor (lower) variants that interact in a many-to-many fashion, with each ligand binding to multiple receptors and each receptor binding to multiple ligand variants.
 (D) In signal transduction systems, receptor variants (upper) interact with variant intracellular signal transducers, or effectors, in a many-to-many fashion.
 (E) In cell adhesion processes, protocadherin variants interact with other protocadherins in adjacent cells, also in a many-to-many fashion.
 (F) Eukaryotic transcription factor variants can often bind to one another to form a large repertoire of distinct dimers, each with distinct DNA binding specificities.

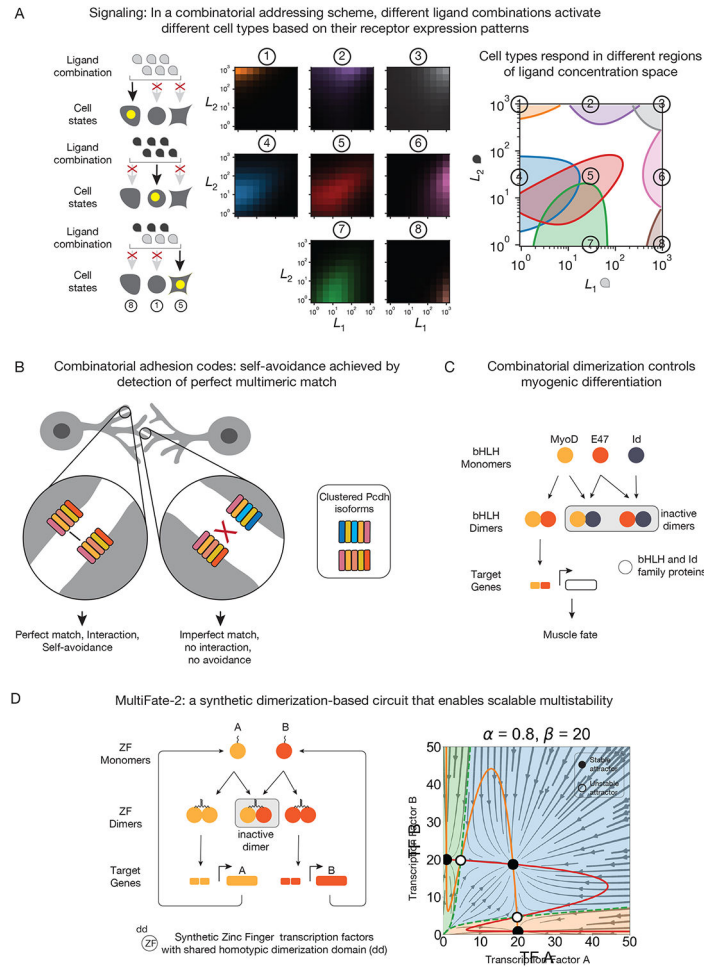


Figure 2: Many-to-many interaction systems can provide key functional capabilities.

(A) In cell-cell signaling, many-to-many interactions between ligands and receptors allow ligands mixtures to “address” messages to specific cell types, rather than broadcasting messages to any cell expressing receptors (left). This works by generating responses in specific regions of a multi-dimensional ligand concentration space. A model of BMP signaling showed that particular parameters for receptor-ligand interactions allow a single ligand pair to activate up to 8 distinct receptor configurations, depending on the ligand concentrations (middle, right). In each plot, the x- and y-axes show concentrations of two ligands, L_1 and L_2 .

(B) In cell-cell adhesion, combinations of Pcdh isoforms (different colors) can encode cellular identity, allowing neurons to distinguish self from non-self.

(C) In gene regulation circuits, dimerization of bHLH transcription factors partitions the protein variants into active and inactive dimers. Perturbations of monomer expression can thus alter dimer abundance to produce nuanced effects on target gene expression.

(D) Naturally occurring protein-protein dimerization networks can guide the design of synthetic circuits for more complex computations. For example, the MultiFate-2 system includes a synthetic transcription mechanism that, like bHLH transcription factors, includes dimerization, DNA binding, and transcriptional activation domains. These programmed

interactions (left) give rise to diverse stable cell states (right) defined by concentrations of active transcription factor dimers, analogous to differentiated cell states.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

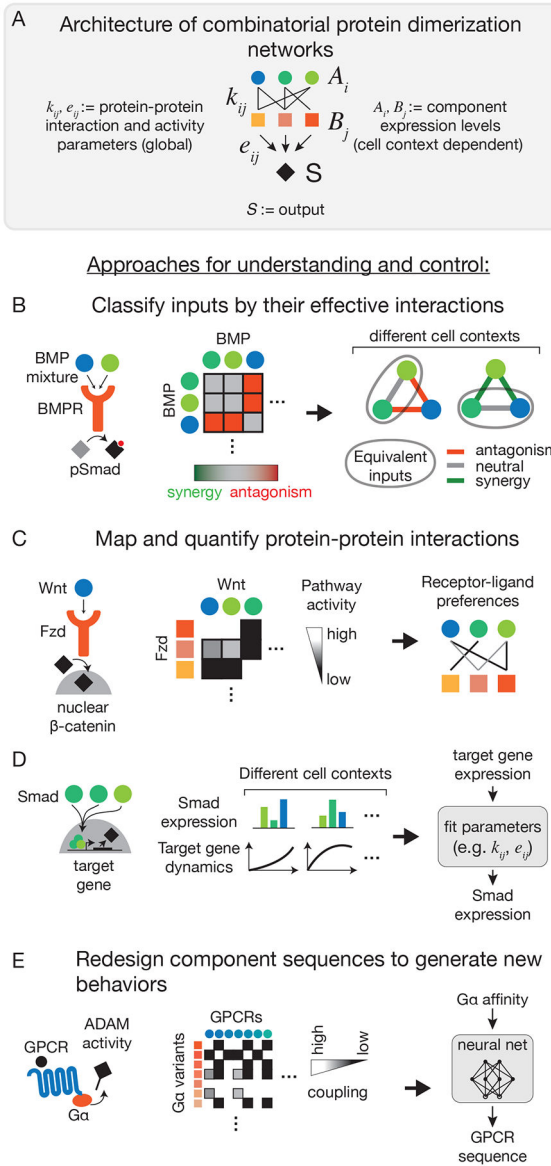


Figure 3: Strategies to understand and control combinatorial complexity of protein dimerization networks.

(A) In combinatorial protein dimerization networks, one set of protein variants (A_i) binds combinatorially to another set of protein variants (B_j); the resulting complexes produce a common output (S). Global parameters describe the protein-protein binding (k_{ij}) and complex activity (e_{ij}) parameters. The output of the network depends on the complex levels, and thus on the abundance of various components (A_i, B_j), which can vary between cell contexts.

(B) The functional differences between protein variants are often unclear because of the variants' high sequence similarity. For example, the highly homologous BMP ligands activate a common output (pSmad), but can produce distinct effects in combinations. Mapping pairwise responses allows individual ligands to be classified by their effective

interactions, highlighting functional differences between the ligands⁴³. These differences are not global, but can vary between cell contexts.

(C) While all dimerizations are possible in principle, they do not necessarily produce output. For example, in the Wnt pathway, not all Wnt ligands bind and activate the associated receptors (Fzd). Exposing cell lines engineered to express only a single Fzd receptor to each Wnt ligand revealed patterns of receptor-ligand preferences that were not clear from protein sequence alone⁷⁹.

(D) Quantitatively predicting the output produced by protein-protein interactions requires careful parameterization of binding affinity and overall activity (i.e. k_{ij} , e_{ij}). Combining measurements of component abundance (A_i , B_j) with output dynamics (S) allows the inference of these parameters. For example, quantifying expression of Smad proteins and the expression of their downstream target genes fit a model that predicted target gene expression from Smad expression, and vice versa⁸⁰.

(E) Redesigning a given component's sequence can alter its protein-protein interactions and, in some cases, produce entirely new responses and computations. The coupling interactions of 148 GPCRs with 11 $G\alpha$ subunits were used to train a machine learning predictor of coupling strength. This model then guided the design of a less promiscuous GPCR, by generating GPCR sequences predicted to couple uniquely to a given $G\alpha$ subunit¹⁵.