Cell

Extracting Insight from Noisy Cellular Networks

Christian R. Landry,[1,5] Emmanuel D. Levy,[2,5] Diala Abd Rabbo,[3] Kirill Tarassov,[4] and Stephen W. Michnick[3,*]
[1]Département de Biologie, IBIS and PROTEO, Pavillon Charles-Eugene-Marchand, 1030 Avenue de la Medecine, Laval University, Québec City, QC G1V 0A6, Canada
[2]Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel
[3]Département de Biochimie, Université de Montréal, C.P. 6128, Succursale Centre-ville, Montréal, QC H3C 3J7, Canada
[4]Zora Biosciences Oy, Biologinkuja 1, Espoo 02150, Finland
[5]These authors contributed equally to this work
*Correspondence: stephen.michnick@umontreal.ca
http://dx.doi.org/10.1016/j.cell.2013.11.003

Network biologists attempt to extract meaningful relationships among genes or their products from very noisy data. We argue that what we categorize as noisy data may sometimes reflect noisy biology and therefore may shield a hidden meaning about how networks evolve and how matter is organized in the cell. We present practical solutions, based on existing evolutionary and biophysical concepts, through which our understanding of cell biology can be enormously enriched.

The spandrels of San Marco is an architectural analogy that Stephen J. Gould and Richard C. Lewontin used to explain the fundamental flaw in systematically ascribing individual traits of an organism to adaptation rather than to a possible coincidental evolution of some other characteristic (Gould and Lewontin, 1979). More than three decades on, such adaptationist tendencies remain common to the interpretation of biological data and no less in network biology. The intrinsically beautiful and elegant structure embedded in interpretations, such as functional modularity, can mask important details and understanding of what the data tell us about the organization and evolution of networks. With the accelerating accumulation of data gathered at all layers of the cell, it is useful to return to first principles and ask precisely what we measure and what assumptions we make when analyzing the large-scale data that populate networks. The recent debate around ENCODE regarding how much of the human genome is functional is a clear example of why we need to address these issues (Doolittle, 2013; Graur et al., 2013; Maher, 2012). Interpretations of gene function in a project like ENCODE requires the integration of a number of different types of large- and small-scale experimental data (Gerstein et al., 2012). To discuss all of the issues involved in interpreting these different types of data is beyond the scope of this Perspective. Instead, we take a fresh look at the raw details of one type of data, protein-protein interactions (PPI), and we ask what the experiments upon which they are based measure and take an alternative approach to their interpretation.

PPIs constitute the physical link among gene products and thus provide us with essential clues to how biological processes are organized and integrated in cells and organisms (Babu et al., 2012; Gerstein et al., 2012; Havugimana et al., 2012; Zhang et al., 2012). PPI networks are, however, largely difficult to interpret functionally and appear to be both poorly conserved across organisms and immensely large. Statistical strategies for interpreting large data sets have aided greatly in our attempts to understand PPI networks and continue to advance (Collins et al., 2007), but for nonspecialists, results of such analyses are abstractions of the physical results that can obscure hidden and important details about how PPIs are organized. As sometimes happens in science, the object of interest becomes the abstract representation itself and not the underlying data. Here, we discuss key problems that may hinder clear understanding of PPIs, PPI networks, and their evolutionary history, and we propose solutions for each of these problems (Box 1).

Problem 1: When Is an Interaction Not an Interaction?

We begin by asking how, at an essential level, large-scale PPI data are interpreted (Figures 1A and 1B). For the sake of brevity, we discuss the results of PPI screens for the model eukaryote budding yeast, *Saccharomyces cerevisiae*, for which there is the greatest amount of data available (Gavin et al., 2002, 2006; Ho et al., 2002; Ito et al., 2000; Krogan et al., 2006; Tarassov et al., 2008; Uetz et al., 2000; Yu et al., 2008). The reader may be surprised to learn that large-scale PPI detection methods do not necessarily detect direct interactions between proteins. Three families of methods have produced the bulk of large-scale PPIs, including first those based on affinity purification followed by mass spectroscopy (AP-MS). This approach provides evidence, largely of stable complexes that can survive conditions of cell lysis and purification (Babu et al., 2012; Gavin et al., 2002, 2006; Ho et al., 2002; Krogan et al., 2006; Zhang et al., 2012). Yeast two-hybrid (Y2H) methods are performed in vivo and may yield direct, binary information albeit in an unnatural compartment for most proteins (the nucleus), and proteins are typically expressed under nonnative promoters (Ito et al., 2000; Uetz et al., 2000; Yu et al., 2008). Finally, protein-fragment

CrossMark

Box 1. Solutions for Common Problems in Interpreting Protein Interaction Data

Problem 1: Different methods produce different types of data.

Solution 1: Different data need to be conceptualized and assessed differently, and models of reference (ideal PPIs) need to reflect the breadth of methods used to probe PPIs and to reflect the biological diversity of PPIs as well. Reference PPIs should thus be tailored for each method.

Problem 2: PPIs often appear as having a poor functional relevance.

Solution 2: This observation has a biological explanation—promiscuity. A substantial number of PPIs that we observe serve no discernible function in the cell. Key cellular and chemical parameters, namely protein abundance, complex stoichiometry, and interaction conservation need to be taken into account to single out functional interactions and understand the biology behind networks.

Problem 3: Proteins do not always follow rules of organization commonly depicted as molecular modules.

Solution 3: An open mind with a combination of the two above-mentioned points. Models of how PPIs are organized should be extracted from the data rather than imposed on the data.

complementation assays (PCA) are in the middle; they do not provide unambiguous evidence of direct binary PPI but rather provide an indication of spatial proximity between two proteins. An advantage of this method is that proteins are expressed at endogenous levels and within relevant cellular compartments in living cells (Tarassov et al., 2008).

Importantly, applications of criteria to access one type of data can be wholly misleading if applied to another. For instance, a gold standard such as reference protein complexes would include many interactions within complexes that cannot be captured by PCA or Y2H because the proteins are not physically close or in contact. Thus, different standards should be used to assess different data sets.

Solution 1: Always Compare the Comparable

Admittedly, false-positives and biases derived from experimental errors must be eliminated statistically—for instance, based on their reproducibility (Mellacheruvu et al., 2013). However, care should be taken to choose appropriate reference PPIs for each particular experimental approach, and ideally, these methods should use information that is orthogonal and based on as many different methods as possible to the PPI detection approach. This would allow for correct assessment of the reliability of the data without biases toward one method or another. A better understanding and consideration of the methods used and their shortcomings may also help explain why so many interactions are not detected. In turn, such understanding could help raise the confidence that a lack of interaction in the data reflects the genuine absence of an interaction in the cell. For instance, some reporters may destabilize the fusion proteins and make interactions impossible to see or may hinder binding interfaces. Some proteins may be unable to work in a particular cell compartment where the reporter is reconstituted. Some screening methods may have a high rate of failure at some point in their procedure. In all cases, better controls on the experimental procedures (e.g., measurement of reproducibility) and of molecular constructs (e.g., confirmation of expression of fusion proteins) may alleviate these shortcomings. Furthermore, the lack of overlap among the current data sets may be relevant to a more fundamental question than those based solely on the reproducibility of a given technique (Bader and Hogue, 2002; von Mering et al., 2002). This question relates to the incompleteness of our current model of the interactome. If current methods fail to uncover the same relationships among proteins, we expect that many more relationships may have been missed and thus that new technological developments are needed. For instance, approaches that would allow single-cell analysis of interactions based on fluorescent or luminescent reporters or approaches that would allow resolving spatiotemporal dependencies of PPIs could help fill the current gap.

Regardless of how data might be validated, an important question is: what are we measuring? In other words, do we understand why a given signal from a given experiment maximizes our ability to predict biologically meaningful PPIs (Balaji et al., 2008; Jensen and Bork, 2008; Wodak et al., 2013)? As we describe below, to answer this question, we must first consider how PPIs have evolved.

Problem 2: Measurable Does Not Mean Functional

In the past few years, we and others have made several observations suggesting that many PPIs, regardless of whether they are reproducible by different techniques, could have no function in the cell (Landry et al., 2009; Levy et al., 2009). What do we mean by a nonfunctional PPI? From an evolutionary perspective, we mean that such PPIs appeared as the product of evolutionary processes but were not selected for imparting any benefit to the organism in which they arose (or later on) during evolution. Accordingly, such PPIs are not currently maintained by purifying selection. Consequently, disruption of a nonfunctional PPI is not expected to result in any biochemical or phenotypic deleterious consequence. It is also important to distinguish nonfunctional from nonspecific interactions. We usually think of nonspecific interactions as those that arise from, for instance, binding between hydrophobic surfaces of proteins. A nonfunctional interaction, however, could have the hallmarks of a specific interaction, including stereospecificity and shape complementarity, but again, the interaction may impart no beneficial functional consequences to either of the two proteins. This is a difficult idea to fathom, but not without precedent. As Gould and Lewontin (1979) argued, not all features of a biological system have evolved because they provide a favorable function to the organism, and the same argument applies to molecular phenotypes. For instance, we know that transcription factors bind to hundreds of sites in a genome, but few are involved in regulating gene expression (Biggin, 2011; Euskirchen and Snyder, 2004; Hahn et al., 2003). Formal predictions of nonfunctional transcription factor binding have been explored, and we have extended the same analyses to PPIs (Hahn et al., 2003; Levy et al., 2009). For one type of PPI, that of protein kinases with their substrates, we estimate that nonfunctional PPIs may compose higher than 50% of observables (Landry et al., 2009). Following these studies, investigators have used evolutionary and structural information as a means of prioritizing posttranslational modifications for functional studies, for instance for modifications that regulate PPIs (Beltrao et al., 2012).

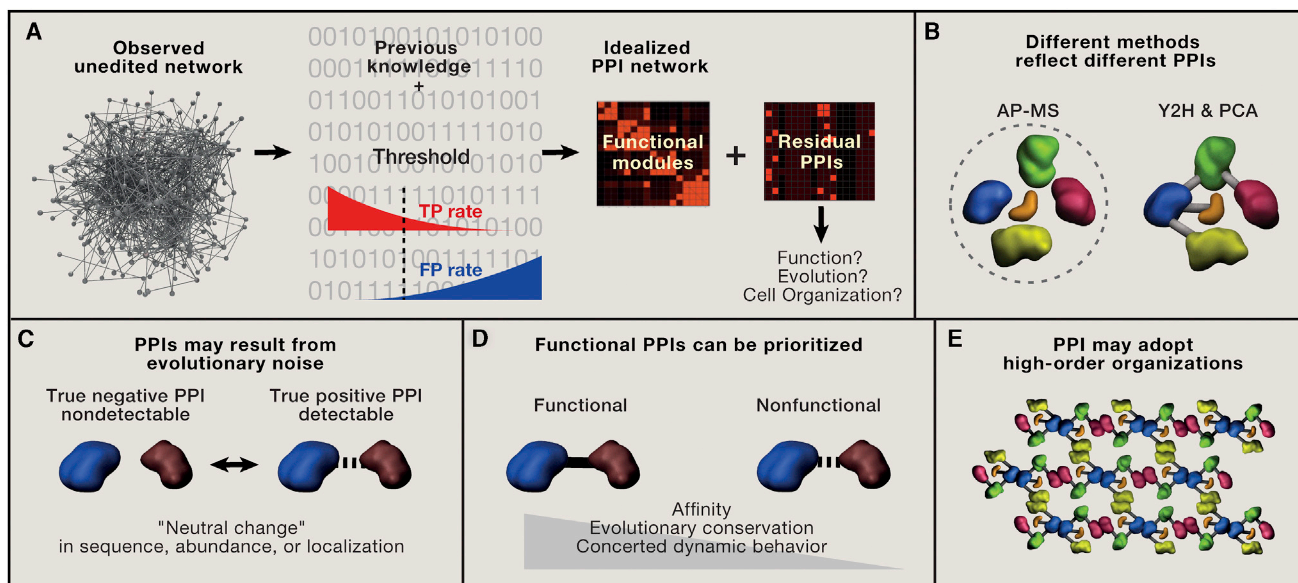

Figure 1. Challenges in Interpreting Protein-Protein Interaction Networks

(A) General outline of PPI network interpretation. Typical large-scale network observations produce raw data that are filtered to eliminate experimental errors and biases. This incidentally also tends to produce networks with the properties of idealized PPIs. Previous data consisting of a set of experimentally verified interactions judged to be irrefutable and a larger set of hypothetical noninteracting proteins are used to set a threshold for the experimental data signal (broken line). The threshold typically maximizes the recovery of gold standard data (true positives, TP) while minimizing the recovery of interactions assumed impossible (false positives, FP). A final PPI network is interpreted in a number of ways, including various types of clustering to reveal structural and functional organization. There is, however, an equal or even larger subset of residual data that do not fit any desirable model of network organization.

(B) Different methods for detecting PPI capture different information. AP-MS captures associations of proteins without reference to which pairs are in direct contact, Y2H captures mostly direct interactions, and PCA captures proteins that are spatially restrained to specific distances from each other.

(C) New, nonfunctional PPIs can be formed or broken through simple changes in expression or point mutations.

(D) Physicochemical, evolutionary, and dynamic properties of PPIs may help to distinguish functional from nonfunctional PPIs.

(E) Super-organizations of proteins (e.g., hydrogels) may have an important function but would not fit standard models of PPI network organization.

One might expect that nonfunctional interactions should be eliminated by natural selection. This, however, is likely to occur only if nonfunctional PPIs are deleterious to the cell and if the appropriate mutational and population genetics requirements are met (Figures 1C and 1D) (Fernández and Lynch, 2011; Levy et al., 2009). A PPI could arise from point mutations (Grueninger et al., 2008)—or perhaps even due to a change in expression level (Gagnon-Arsenault et al., 2013) or subcellular localization of one of the partners (Kuriyan and Eisenberg, 2007)—but have no functional consequence. Furthermore, a particular PPI may be an inevitable consequence of a function of another PPI in which one of the partners is involved. What we observe then may be a tradeoff between the specificity of PPIs and the ability of proteins to perform specific functions (Pechmann et al., 2009).

Nonfunctional PPIs may appear to be an obstacle to our understanding of how the cell works. We argue the opposite: understanding nonfunctional PPIs provide a window into the past, the present, and the future of evolving PPI networks. For instance, the birth and death of PPIs may contribute to the evolution of biochemical networks and to speciation of organisms (Tawfik, 2010). What will be functional in the future is impossible to predict. However, one could argue that nonfunctional interactions may provide templates for the accumulation of beneficial mutations in the future. Accordingly, the wandering of PPI networks in the nonfunctional space may allow cells to explore configurations not directly available to beneficial mutations or modify the functional space so as to affect the neutrality of future mutations (Doolittle, 2013). Thus, on the one hand, nonfunctional PPIs can be a source of annoyance to those trying to understand PPI data, but on the other, they may represent a feature of ongoing evolution of cellular networks (Levy et al., 2009, 2010; Lynch, 2007a, 2007b; Zhang et al., 2008). Furthermore, functional and nonfunctional PPIs can be separated based on simple biophysical and evolutionary concepts.

Solution 2: Orthogonal Measures of Functionality

Chemical principles of PPI may provide important clues of functionality (Figure 1D) (Schreiber and Keating, 2011). Existing PPI data analyses implicitly test chemical parameters. For instance, intensity or frequency of an observable can be thought of as measuring the affinities or rates of association or dissociations of complexes. Recently, we have demonstrated that the proportion of protein phosphorylation on specific residues (or stoichiometry) can provide meaningful predictions of their functionality (Landry et al., 2009; Levy et al., 2012b). Whether this principle applies to PPIs or other types of biomolecular interactions remains to be explored but, if true, could provide strong evidence of functionality. In addition, the thermodynamics of PPIs could be even more useful to distinguishing functional versus nonfunctional PPIs. For instance, it has been recently demonstrated that functional transcription factor binding in a genome could be distinguished from nonfunctional interactions by virtue

that transcription factors exchange slower at functional sites, where transcription occurs, than at sites where no transcription is initiated (Lickwar et al., 2012). Similar conclusions have been reached regarding the occupancy of transcription factors during development (Fisher et al., 2012). Whether this principle applies to PPIs or other types of biomolecular interactions remains to be explored but, if true, could provide strong evidence of functionality. One particular challenge in using these principles is that many PPIs, including functional ones, are weak. In that respect, the occurrence frequency of a PPI cannot be used by itself as an indicator of functionality. However, experimental work could be used to measure the distribution of effects that mutations have on PPIs affinity to ultimately estimate how binding can evolve through neutral mutations. Such knowledge could indeed help us infer a confidence level for an interaction given a measured affinity. Most importantly, the consideration of what may be functional and what may not be functional requires that we move beyond the simple definitions of false-positive and false-negative hits in large-scale interactome studies. Accordingly, efforts should be made so that parameters such as the stoichiometry or the affinity of interactions can be considered in the analysis of networks.

Other approaches reside in correlating the behavior of PPIs with specific cellular responses. In principle, one should expect nonfunctional interactions to exhibit no specific dynamic responses, whereas functional PPIs should display coordinated dynamics within pathways or cellular processes. For instance, there is evidence that dynamic changes in phosphorylation following perturbations of cells are more likely functional than static sites, as others have intuitively surmised (Olsen et al., 2006). We have recent evidence suggesting that PPIs that are dynamic in response to a perturbation are also more likely functional (Messier et al., 2013). As methods and tools for perturbing interactomes on a large-scale are developing, it should now be feasible to exploit this principle to further investigate the functional elements of these networks (Diss et al., 2013).

Finally, the most straightforward solution is to use an approach that all biochemists and geneticists intuitively use when dissecting the function of a gene or of a protein: comparative analyses. Functional features of genes and genomes indeed tend to be under purifying selection and thus tend to be conserved within and between species. The systematic use of orthogonal information, such as PPI conservation among closely related species, is thus an obvious solution that has already been applied successfully to map out gene regulatory networks (Harbison et al., 2004) and is being developed for protein interactomes (Leducq et al., 2012). In the case of proteins of known structure, comparative approaches have also been successful in discriminating interaction interfaces from solvent-accessible protein surface (Armon et al., 2001; Elcock and McCammon, 2001; Valdar and Thornton, 2001), even among weak interactions (Dey et al., 2010). Comparative approaches will also require an understanding of the contributions of nonadaptive forces and how these do link to chemical constraints that shape them (Fernández and Lynch, 2011; Levy et al., 2012a; Lynch, 2007a). This requires quantifying mutation rates for gains and losses of interactions as well as estimating the costs of nonfunctional PPIs. In vivo PPI studies so far have been mostly descriptive, as there have been very few attempts to manipulate PPIs through perturbations such as mutations on binding interfaces or changes in protein abundance (Gagnon-Arsenault et al., 2013) and mutations of protein residues (Dreze et al., 2009; Ear and Michnick, 2009). With these parameters in hand, it may be possible to estimate how many nonfunctional interactions may populate protein networks. In order to reach this goal, we will need methods to study the role of PPIs independently of the other functions of the proteins involved. It is currently difficult to examine the function of a single PPI because most genetic approaches involve gene deletions that eliminate both the proteins and all of its interactions altogether. Current experimental approaches have been developed to manipulate PPIs without completely eliminating the proteins involved (Dreze et al., 2009; Ear and Michnick, 2009), and these could be used for this purpose. These PPI-centered experiments will also be important for testing hypotheses regarding the role of PPIs that are indirect consequences of another function. In addition, it remains difficult, if not impossible, to define a putative function (let alone define what a function is) for any given PPI and thus to design the appropriate experiments. However, with the development of the tools described above that allow us to specifically dissect PPIs, one could assay the fitness of point mutants in diverse environmental conditions to uncover combinations where network failure may decrease fitness—for instance, using approaches developed for gene knockouts (Hillenmeyer et al., 2008) or for genetic interaction mapping (Braberg et al., 2013).

There are many characteristics that are associated with functional PPIs, and an optimal approach will be to consider them jointly. With the accumulation of experimental data on context-dependent PPIs and on interspecies comparisons, one will eventually be able to integrate these features into a scoring scheme that will allow us to make predictions as to which PPIs are most likely functional.

Problem 3: Proteins Are Rule Breakers

In addition to considering nonfunctional PPIs, assumptions about how PPI networks should be organized into complexes, for example, could obscure additional or even alternative explanations for how matter is organized in the cell (Figure 1E). There is evidence of such modules that can work independently from the rest of the network. Groups of proteins that are hypothesized to reflect functional modules can indeed be artificially assembled and maintain their functional dynamics, even in isolation. For instance, mammalian MAP kinases have been shown to display predictable behavioral response to external stimuli when reconstituted in yeast (O'Shaughnessy et al., 2011). In addition, the cell-cycle control in fission yeast has been reduced to a monomolecular Cdk-cyclin engine that can drive major cell-cycle transitions (Coudreuse and Nurse, 2010). However, any disagreeing results should also be considered. For instance, several recent studies suggest that proteins and nucleic acids can exist in a number of potential states, including as assemblies of different dimensions, from nano- to micrometer scales and can have unique physical properties. Structures seen in cells, such as nucleoli, PML, Cajal bodies, and ribonucleoprotein granules are just a few examples of what are likely metascale (hundreds of nanometers) liquid states in which groups of molecules may

be organized and physical properties may be maintained to carry out specific functions (Brangwynne et al., 2009, 2011; Han et al., 2012; Hyman and Brangwynne, 2011; Kato et al., 2012; Li et al., 2012; Narayanaswamy et al., 2009). The discovery of these alternative states of protein assembly is prompting a re-evaluation of how matter is organized in cells and how these organizations may affect everything from transcription and translation of genes to signal transduction and morphogenesis. Importantly, the physical properties, such as viscosity, of these assemblies are quite different from the surrounding cytoplasm (Brangwynne, 2011). Consequently, interactions that might be weak in the cytoplasm could be stronger inside of these bodies, and there may be no apparent functional logic to the PPIs involved. These may simply be a nonfunctional consequence of localization within the bodies to which they may contribute certain properties, such as mRNA storage or regulation with P-bodies (Brangwynne et al., 2009). The proteins within the body may have no apparent functional relationships to each other but that of aggregating under certain conditions. Furthermore, other proteins may freely exchange between the cytosol and the body over time, resulting in many interactions—none making any particular sense. Such PPIs could be considered outliers in large-scale PPI studies, the data likely shuttled into a supplementary file and forgotten (Figure 1A). PPI data therefore need to be examined without any a priori judgments or models regarding how they should be organized in the cell.

Conclusions

Network biologists have many tools at hand to discern meaning from existing data, but much of the models they use are based on intuitive, teleological assumptions. We view PPI data as they exist today as capturing a snapshot of evolutionary wiring and rewiring of a PPI network in which much of the information may be superfluous to the contemporary function of an organism. Sorting out the meaningful from the superfluous may be a matter of distinguishing chemical parameters that are shaped through natural selection. We think that PPI networks could be telling us a good deal more about the organization and history of matter in the cell. Taking a fresh view that is equally objective, biophysical, and comparative will provide a more meaningful understanding of where PPI networks came from and where they are going.

ACKNOWLEDGMENTS

Many of the ideas discussed in this manuscript were inspired by the results of research funded by the Canadian Institutes of Health Research (CIHR) grants MOP-GMX-152556 and MOP-GMX-231013 to S.W.M. and the Natural Sciences and Engineering Research Council of Canada (NSERC) grant 194582 to S.W.M. C.R.L. is a CIHR New Investigator, and his research on PINs is funded by CIHR grants (GMX-299432 and GMX-191597) and by NSERC. E.D.L. acknowledges the I-CORE Program of the Planning and Budgeting Committee and the Israel Science Foundation (grant 1775/12).

REFERENCES

Armon, A., Graur, D., and Ben-Tal, N. (2001). ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. J. Mol. Biol. *307*, 447–463.

Babu, M., Vlasblom, J., Pu, S., Guo, X., Graham, C., Bean, B.D.M., Burston, H.E., Vizeacoumar, F.J., Snider, J., Phanse, S., et al. (2012). Interaction landscape of membrane-protein complexes in Saccharomyces cerevisiae. Nature *489*, 585–589.

Bader, G.D., and Hogue, C.W. (2002). Analyzing yeast protein-protein interaction data obtained from different sources. Nat. Biotechnol. *20*, 991–997.

Balaji, S., Iyer, L.M., Babu, M.M., and Aravind, L. (2008). Comparison of transcription regulatory interactions inferred from high-throughput methods: what do they reveal? Trends Genet. *24*, 319–323.

Beltrao, P., Albanèse, V., Kenner, L.R., Swaney, D.L., Burlingame, A., Villén, J., Lim, W.A., Fraser, J.S., Frydman, J., and Krogan, N.J. (2012). Systematic functional prioritization of protein posttranslational modifications. Cell *150*, 413–425.

Biggin, M.D. (2011). Animal transcription networks as highly connected, quantitative continua. Dev. Cell *21*, 611–626.

Braberg, H., Jin, H., Moehle, E.A., Chan, Y.A., Wang, S., Shales, M., Benschop, J.J., Morris, J.H., Qiu, C., Hu, F., et al. (2013). From structure to systems: high-resolution, quantitative genetic analysis of RNA polymerase II. Cell *154*, 775–788.

Brangwynne, C.P. (2011). Soft active aggregates: mechanics, dynamics and self-assembly of liquid-like intracellular protein bodies. Soft Matter *7*, 3052–3059.

Brangwynne, C.P., Eckmann, C.R., Courson, D.S., Rybarska, A., Hoege, C., Gharakhani, J., Jülicher, F., and Hyman, A.A. (2009). Germline P granules are liquid droplets that localize by controlled dissolution/condensation. Science *324*, 1729–1732.

Brangwynne, C.P., Mitchison, T.J., and Hyman, A.A. (2011). Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. Proc. Natl. Acad. Sci. USA *108*, 4334–4339.

Collins, S.R., Kemmeren, P., Zhao, X.C., Greenblatt, J.F., Spencer, F., Holstege, F.C., Weissman, J.S., and Krogan, N.J. (2007). Toward a comprehensive atlas of the physical interactome of Saccharomyces cerevisiae. Mol. Cell. Proteomics *6*, 439–450.

Coudreuse, D., and Nurse, P. (2010). Driving the cell cycle with a minimal CDK control network. Nature *468*, 1074–1079.

Dey, S., Pal, A., Chakrabarti, P., and Janin, J. (2010). The subunit interfaces of weakly associated homodimeric proteins. J. Mol. Biol. *398*, 146–160.

Diss, G., Dube, A.K., Boutin, J., Gagnon-Arsenault, I., and Landry, C.R. (2013). A systematic approach for the genetic dissection of protein complexes in living cells. Cell Rep. *3*, 2155–2167.

Doolittle, W.F. (2013). Is junk DNA bunk? A critique of ENCODE. Proc. Natl. Acad. Sci. USA *110*, 5294–5300.

Dreze, M., Charloteaux, B., Milstein, S., Vidalain, P.O., Yildirim, M.A., Zhong, Q., Svrzikapa, N., Romero, V., Laloux, G., Brasseur, R., et al. (2009). 'Edgetic' perturbation of a C. elegans BCL2 ortholog. Nat. Methods *6*, 843–849.

Ear, P.H., and Michnick, S.W. (2009). A general life-death selection strategy for dissecting protein functions. Nat. Methods *6*, 813–816.

Elcock, A.H., and McCammon, J.A. (2001). Identification of protein oligomerization states by analysis of interface conservation. Proc. Natl. Acad. Sci. USA *98*, 2990–2994.

Euskirchen, G., and Snyder, M. (2004). A plethora of sites. Nat. Genet. *36*, 325–326.

Fernández, A., and Lynch, M. (2011). Non-adaptive origins of interactome complexity. Nature *474*, 502–505.

Fisher, W.W., Li, J.J., Hammonds, A.S., Brown, J.B., Pfeiffer, B.D., Weiszmann, R., MacArthur, S., Thomas, S., Stamatoyannopoulos, J.A., Eisen, M.B., et al. (2012). DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in Drosophila. Proc. Natl. Acad. Sci. USA *109*, 21330–21335.

Gagnon-Arsenault, I., Marois Blanchet, F.C., Rochette, S., Diss, G., Dube, A.K., and Landry, C.R. (2013). Transcriptional divergence plays a role in the

rewiring of protein interaction networks after gene duplication. J. Proteomics *81*, 112–125.

Gavin, A.C., Bösche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M., et al. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature *415*, 141–147.

Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dümpelfeld, B., et al. (2006). Proteome survey reveals modularity of the yeast cell machinery. Nature *440*, 631–636.

Gerstein, M.B., Kundaje, A., Hariharan, M., Landt, S.G., Yan, K.K., Cheng, C., Mu, X.J., Khurana, E., Rozowsky, J., Alexander, R., et al. (2012). Architecture of the human regulatory network derived from ENCODE data. Nature *489*, 91–100.

Gould, S.J., and Lewontin, R.C. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. Proc. R. Soc. Lond. B Biol. Sci. *205*, 581–598.

Graur, D., Zheng, Y., Price, N., Azevedo, R.B., Zufall, R.A., and Elhaik, E. (2013). On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE. Genome Biol. Evol. *5*, 578–590.

Grueninger, D., Treiber, N., Ziegler, M.O., Koetter, J.W., Schulze, M.S., and Schulz, G.E. (2008). Designed protein-protein association. Science *319*, 206–209.

Hahn, M.W., Stajich, J.E., and Wray, G.A. (2003). The effects of selection against spurious transcription factor binding sites. Mol. Biol. Evol. *20*, 901–906.

Han, T.W., Kato, M., Xie, S., Wu, L.C., Mirzaei, H., Pei, J., Chen, M., Xie, Y., Allen, J., Xiao, G., and McKnight, S.L. (2012). Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. Cell *149*, 768–779.

Harbison, C.T., Gordon, D.B., Lee, T.I., Rinaldi, N.J., Macisaac, K.D., Danford, T.W., Hannett, N.M., Tagne, J.B., Reynolds, D.B., Yoo, J., et al. (2004). Transcriptional regulatory code of a eukaryotic genome. Nature *431*, 99–104.

Havugimana, P.C., Hart, G.T., Nepusz, T., Yang, H., Turinsky, A.L., Li, Z., Wang, P.I., Boutz, D.R., Fong, V., Phanse, S., et al. (2012). A census of human soluble protein complexes. Cell *150*, 1068–1081.

Hillenmeyer, M.E., Fung, E., Wildenhain, J., Pierce, S.E., Hoon, S., Lee, W., Proctor, M., St Onge, R.P., Tyers, M., Koller, D., et al. (2008). The chemical genomic portrait of yeast: uncovering a phenotype for all genes. Science *320*, 362–365.

Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L., Adams, S.L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., et al. (2002). Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature *415*, 180–183.

Hyman, A.A., and Brangwynne, C.P. (2011). Beyond stereospecificity: liquids and mesoscale organization of cytoplasm. Dev. Cell *21*, 14–16.

Ito, T., Tashiro, K., Muta, S., Ozawa, R., Chiba, T., Nishizawa, M., Yamamoto, K., Kuhara, S., Sakaki, Y., Giaever, G., et al. (2000). Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. Proc. Natl. Acad. Sci. USA *97*, 1143–1147.

Jensen, L.J., and Bork, P. (2008). Biochemistry. Not comparable, but complementary. Science *322*, 56–57.

Kato, M., Han, T.W., Xie, S., Shi, K., Du, X., Wu, L.C., Mirzaei, H., Goldsmith, E.J., Longgood, J., Pei, J., et al. (2012). Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. Cell *149*, 753–767.

Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P., et al. (2006). Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature *440*, 637–643.

Kuriyan, J., and Eisenberg, D. (2007). The origin of protein interactions and allostery in colocalization. Nature *450*, 983–990.

Landry, C.R., Levy, E.D., and Michnick, S.W. (2009). Weak functional constraints on phosphoproteomes. Trends Genet. *25*, 193–197.

Leducq, J.B., Charron, G., Diss, G., Gagnon-Arsenault, I., Dubé, A.K., and Landry, C.R. (2012). Evidence for the robustness of protein complexes to inter-species hybridization. PLoS Genet. *8*, e1003161.

Levy, E.D., Landry, C.R., and Michnick, S.W. (2009). How perfect can protein interactomes be? Sci. Signal. *2*, pe11.

Levy, E.D., Landry, C.R., and Michnick, S.W. (2010). Cell signaling. Signaling through cooperation. Science *328*, 983–984.

Levy, E.D., De, S., and Teichmann, S.A. (2012a). Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. Proc. Natl. Acad. Sci. USA *109*, 20461–20466.

Levy, E.D., Michnick, S.W., and Landry, C.R. (2012b). Protein abundance is key to distinguish promiscuous from functional phosphorylation based on evolutionary information. Philos. Trans. R. Soc. Lond. B Biol. Sci. *367*, 2594–2606.

Li, P., Banjade, S., Cheng, H.-C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J.V., King, D.S., Banani, S.F., et al. (2012). Phase transitions in the assembly of multivalent signalling proteins. Nature *483*, 336–340.

Lickwar, C.R., Mueller, F., Hanlon, S.E., McNally, J.G., and Lieb, J.D. (2012). Genome-wide protein-DNA binding dynamics suggest a molecular clutch for transcription factor function. Nature *484*, 251–255.

Lynch, M. (2007a). The evolution of genetic networks by non-adaptive processes. Nat. Rev. Genet. *8*, 803–813.

Lynch, M. (2007b). The frailty of adaptive hypotheses for the origins of organismal complexity. Proc. Natl. Acad. Sci. USA *104* (*Suppl 1*), 8597–8604.

Maher, B. (2012). Fighting about ENCODE and Junk. http://blogs.nature.com/news/2012/09/fighting-about-encode-and-junk.html.

Mellacheruvu, D., Wright, Z., Couzens, A.L., Lambert, J.P., St-Denis, N.A., Li, T., Miteva, Y.V., Hauri, S., Sardiu, M.E., Low, T.Y., et al. (2013). The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. Nat. Methods *10*, 730–736.

Messier, V., Zenklusen, D., and Michnick, S.W. (2013). A nutrient-responsive pathway that determines M phase timing through control of B-cyclin mRNA stability. Cell *153*, 1080–1093.

Narayanaswamy, R., Levy, M., Tsechansky, M., Stovall, G.M., O'Connell, J.D., Mirrielees, J., Ellington, A.D., and Marcotte, E.M. (2009). Widespread reorganization of metabolic enzymes into reversible assemblies upon nutrient starvation. Proc. Natl. Acad. Sci. USA *106*, 10147–10152.

O'Shaughnessy, E.C., Palani, S., Collins, J.J., and Sarkar, C.A. (2011). Tunable signal processing in synthetic MAP kinase cascades. Cell *144*, 119–131.

Olsen, J.V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. (2006). Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell *127*, 635–648.

Pechmann, S., Levy, E.D., Tartaglia, G.G., and Vendruscolo, M. (2009). Physicochemical principles that regulate the competition between functional and dysfunctional association of proteins. Proc. Natl. Acad. Sci. USA *106*, 10159–10164.

Schreiber, G., and Keating, A.E. (2011). Protein binding specificity versus promiscuity. Curr. Opin. Struct. Biol. *21*, 50–61.

Tarassov, K., Messier, V., Landry, C.R., Radinovic, S., Serna Molina, M.M., Shames, I., Malitskaya, Y., Vogel, J., Bussey, H., and Michnick, S.W. (2008). An in vivo map of the yeast protein interactome. Science *320*, 1465–1470.

Tawfik, D.S. (2010). Messy biology and the origins of evolutionary innovations. Nat. Chem. Biol. *6*, 692–696.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., et al. (2000). A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature *403*, 623–627.

Valdar, W.S., and Thornton, J.M. (2001). Conservation helps to identify biologically relevant crystal contacts. J. Mol. Biol. *313*, 399–416.

von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S., and Bork, P. (2002). Comparative assessment of large-scale data sets of protein-protein interactions. Nature *417*, 399–403.

Wodak, S.J., Vlasblom, J., Turinsky, A.L., and Pu, S. (2013). Protein-protein interaction networks: the puzzling riches. Curr. Opin. Struct. Biol. Published online September 2, 2013. http://dx.doi.org/10.1016/j.sbi.2013.08.002.

Yu, H., Braun, P., Yildirim, M.A., Lemmens, I., Venkatesan, K., Sahalie, J., Hirozane-Kishikawa, T., Gebreab, F., Li, N., Simonis, N., et al. (2008). High-quality binary protein interaction map of the yeast interactome network. Science *322*, 104–110.

Zhang, J., Maslov, S., and Shakhnovich, E.I. (2008). Constraints imposed by non-functional protein-protein interactions on gene expression and proteome size. Mol. Syst. Biol. Published online August 5, 2008. http://dx.doi.org/10.1038/msb.2008.48.

Zhang, Q.C., Petrey, D., Deng, L., Qiang, L., Shi, Y., Thu, C.A., Bisikirska, B., Lefebvre, C., Accili, D., Hunter, T., et al. (2012). Structure-based prediction of protein-protein interactions on a genome-wide scale. Nature *490*, 556–560.