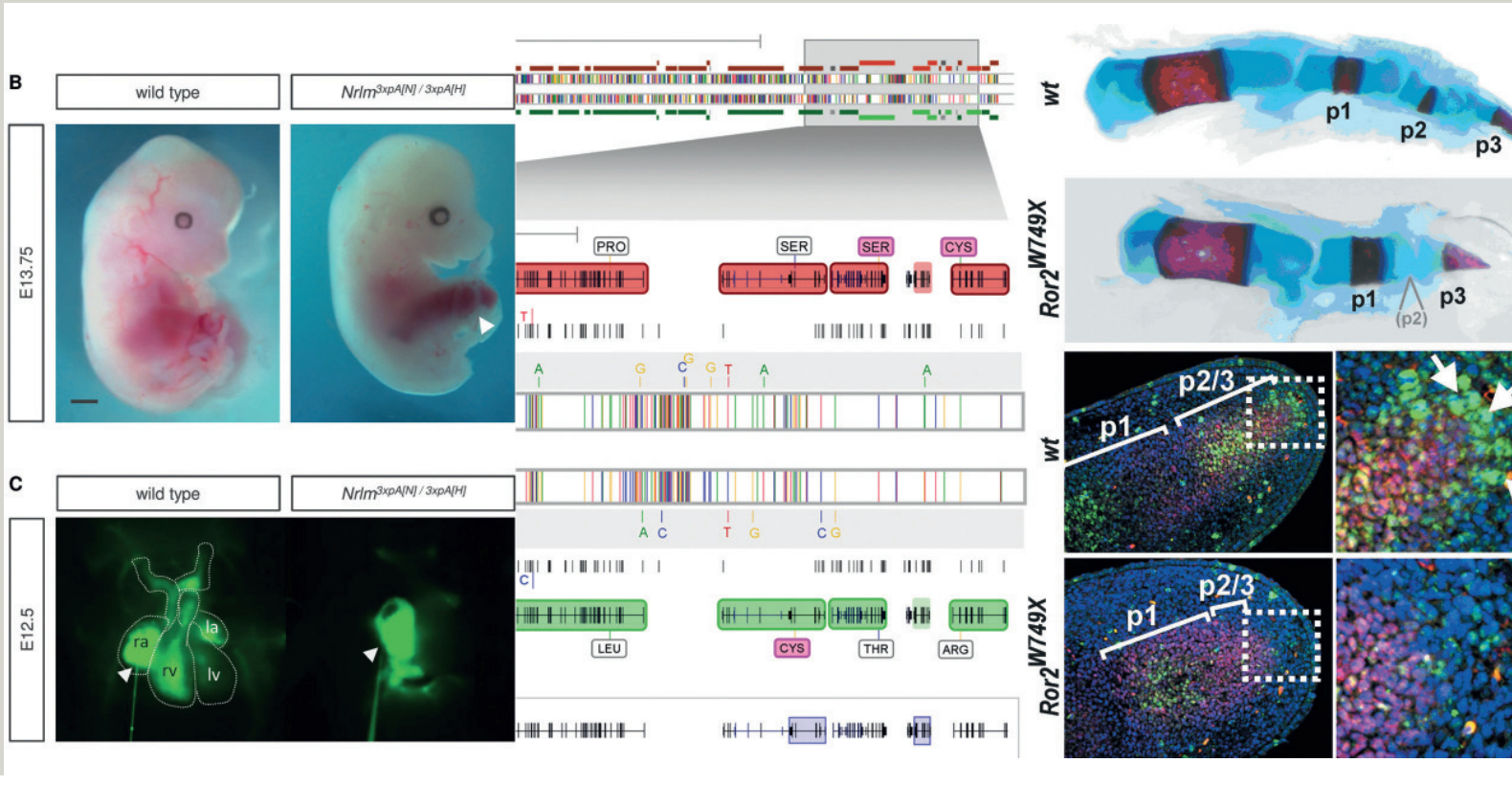
MPIMG

MAX-PLANCK-GESELLSCHAFT

Research Report 2012

Max Planck Institute for Molecular Genetics, Berlin



Imprint | Research Report 2012

Published by the Max Planck Institute for Molecular Genetics (MPIMG), Berlin, Germany, August 2012

Editorial Board:	Bernhard G. Herrmann, Hans Lehrach, H.-Hilger Ropers, Martin Vingron
Coordination:	Patricia Marquardt
Photography:	Katrin Ullrich, David Ausserhofer (p 14), Norbert Michalke (p 211)
Scientific Illustrations:	MPIMG
Production:	Thomas Didier, Meta Druck
Copies:	1,000

Contact:
Max Planck Institute for Molecular Genetics
Ihnestr. 63 – 73
14195 Berlin
Germany

Phone:	+49 (0)30 8413-0
Fax:	+49 (0)30 8413-1207
Email:	info@molgen.mpg.de

For further information about the MPIMG, see http://www.molgen.mpg.de

Otto Warburg Laboratory

Max Planck Research Group Molecular Interaction Networks

(Established: 06/2007)

Head
Ulrich Stelzl (since 06/07)
Phone: +49 (0)30 8413-1264
Fax: +49 (0)30 8413-1960
Email: stelzl@molgen.mpg.de

Secretary of the OWL
Cordula Mancini
Phone: +49 (0)30 8413-1691
Fax: +49 (0)30 8413-1960
Email: mancini@molgen.mpg.de

Scientists
Jonathan Woodsmith (since 01/11)
Nouhad Benlasfer (since 05/10)
Anna Hegele* (06/07-04/12)
Petra Birth (06/08-12/11)
Reynaldo López-Mirabal* (07/08-06/10)

PhD students
Stefanie Jehle (since 05/12)
Thomas Corwin (since 09/10)
Luise Apelt (since 07/10)
Mareike Weimann (12/07-02/12)
Josphine Worseck* (09/07-11/11)
Atanas Kamburov* (01/10-12/11, part time)
Arndt Grossmann* (07/07-12/11)

Undergraduate students
Federico Apelt (since 11/10, part time)
Ziya Özkan (since 08/09, part time)
Franziska Wachsmuth (10/11-06/12)
Chrysovalantis Sourlis (10/09-06/10)
Sylvia Wowro (07/08-05/09, part time)

Scientific overview

Research concept

A major goal in current genome research is to predict the influence of human genetic variation on disease phenotypes. One idea is that large sequencing endeavors, e.g. whole genome sequencing of multiple individuals or large GWAS studies, will provide enough information to make better predictions for risk, cause, pathogenesis or medication of patients vs. control groups (Figure 1a). Molecular interaction networks, such as protein-protein interaction (PPI) networks, are very

* externally funded

useful for studying genotype to phenotype relationships. Integrative computational analyses that largely rely on molecular interaction data result in a more accurate interpretation of genomic variation but remains probabilistic (Figure 1b). However, we want to go beyond improving statistical predictions more towards specific information about an individual. This means we need to consider different sources of molecular, environmental and behavioral variation in addition to individual genomic sequence information. Thus we need to measure – e.g. at the level of cellular networks – to generate additional molecular information about the individual reflecting this variation. To do so, high quality molecular network information will be a necessity (Figure 1c). Differential network analysis will be informative on sets of decisive molecules such as drivers in cancer or transcriptional "master-regulator" proteins and their connections and then guide measurements to better understand and classify individual phenotypes in model systems and ultimately humans.

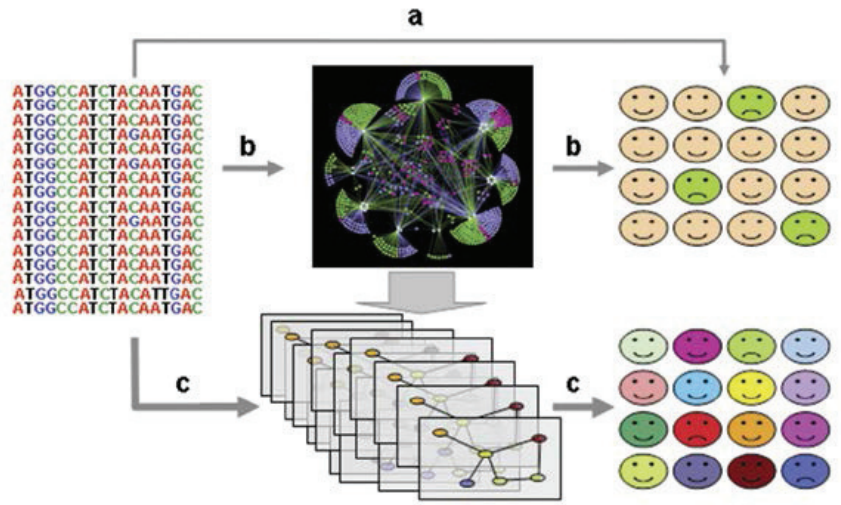


Figure 1: Molecular network information is required to predict genotype – phenotype relationships. To strengthen predictions about the phenotype from genomic information (a) cellular interaction networks will be useful (b) but remain probabilistic over groups of phenotypes. Differential network analyses will discover relevant sets of key molecules that reflect molecular, environmental and behavioral variation and can be measured (c) and will be necessary to predict phenotypes for individual cells /organisms from genomic information.

In the OWL group *Molecular Interaction Networks* we do not work specifically on the interpretation of genetic variation, but aim to provide the network information to better describe disease relevant cellular processes and thus contribute to genotype to phenotype predictions. We work on the generation and analysis of human interaction data and study interaction network dynamics. The later point is particular important as interactome networks are extensively re-wired during a cellular response e.g. during development or during the processing of internal or environmental cues. Differential interaction patterns imply mechanistic changes that are the result of these responses and will thus be most informative when studying genotype to phenotype relationships (Figure 1c).
Specifically, our work aims at i) improving data generation and analysis of human protein-protein interaction networks, ii) integrative approaches to analyze

protein interaction dynamics and iii) experimental approaches to directly investigate conditional protein-protein interactions, such as interactions that e.g. require triggered phosphorylation of one interaction partner mediating the response to changing conditions.

Scientific methods and achievements / findings

Systematic generation of high quality human protein-protein interaction networks

Several studies to systematically map protein-protein interaction (PPI) networks on a large scale have been successful and proven very useful in further studies. Nevertheless, for most species including human only a small fraction of all possible interactions has been mapped today. High quality PPI data collection and independent assessment of data quality remain important tasks.

To provide independent measures of interaction confidence, we have developed a cluster-based method for the assessment of protein-protein interaction confidence and implemented this as a web tool. The method (CAPPIC) exploits the modular network architecture independently of prior parameters or reference sets for confidence scoring of interactions.

In an international collaboration led by the Vidal Lab (Harvard/CCSB, Boston) we have assessed protein interaction data empirically demonstrating that systematic Y2H interaction data, including those generated with our setup are of high precision. The study also revealed that the coverage of the data is low due to relatively low sensitivity of the method. To overcome this limitation, we developed a Y2H-seq approach which enables very high PPI sampling through a second generation short read sequencing readout. Importantly, the method has significantly improved sensitivity and provides a quantitative readout that is indicative of the quality of the PPI information. It will accelerate large-scale interactome mapping efforts.

As Y2H-seq test case, we comprehensively screened proteins involved in methylation and demethylation, i.e. protein methyltransferases and demethylases such as AOF2/LSD1, for interacting partners. Protein methylation of non-histone proteins is a largely unexplored posttranslational modification. We report 523 interactions between 22 methyltransferases or demethylases and 324 interacting proteins. The methyltransferase network is experimentally validated, comprehensively annotated and defines novel cellular roles of non-histone protein methylation. It will thus serve as a major informational resource to the scientific community and is the basis to study methylation dependent protein interactions in the lab in more detail (see below).

Focusing on neurodegenerative diseases, we generated a PPI network connecting proteins implicated in Alzheimer's disease (AD) with the Aloy Lab (IRB, Barcelona). The study suggests novel roles for central proteins in the network that link between oxidative stress, inflammation, and mitochondrial dysfunction in AD. With the Beyer Lab (TU Dresden) a map of human protein interactions was inferred using combined random forest / Bayesian networks to distinguish functional from physical interactions. The map was in part experimentally validated and used to explore the relationships of candidate genes from GWAS of neurodegenerative diseases, such as AD.

Dynamic alterations in protein-protein interaction networks: integrative approaches

The goal here is to reveal differential network states that describe changing cellular processes *in vivo*. Successful analysis of network dynamics through data integration will focus on a specific biological process with interesting dynamic behavior and requires data of very high quality.

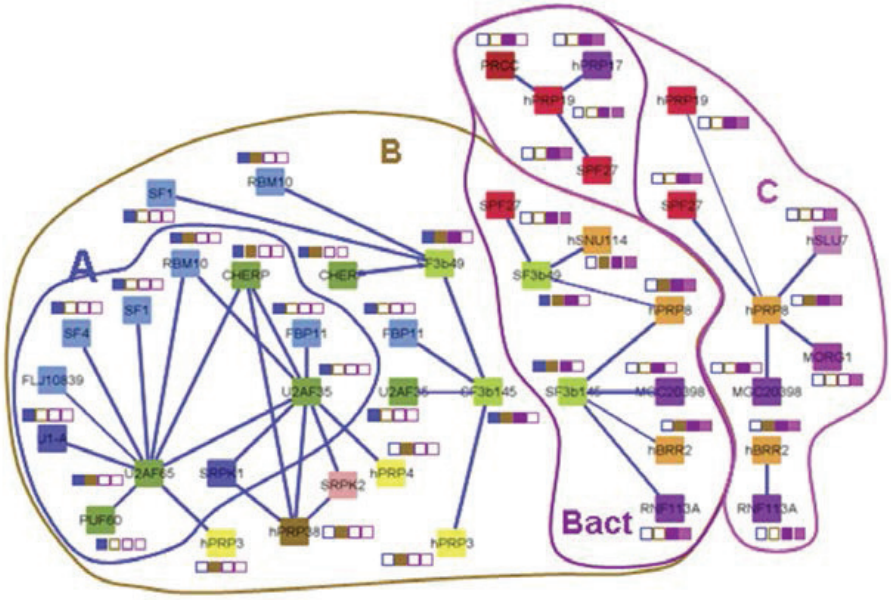


Figure 2: PPI dynamics involving SF3b-complex proteins and hPRP8 (Hegele, Mol Cell 2012). Selected interactions from the (U2AF35,U2AF65), the (SF3b145,SF3b49), the hPRP19 and the hPRP8 modules are shown. Distinct PPI patterns for proteins are suggested for different stages (i.e. A, B, Bact and C complexes) of the spliceosomal assembly cycle.

In a recent study, we focused on PPI dynamics of the splicing cycle. Pre-mRNA splicing is catalyzed by the spliceosome, a highly complex, dynamic and protein rich ribonucleoprotein complex that assembles *de novo* on each intron to be spliced. During spliceosome assembly, activation, catalysis and disassembly, defined large complexes are formed in an ordered, stepwise manner. A data set describing 632 interactions between 200 human spliceosomal proteins was generated including e.g. the first contact sites between the U5 proteins and specific U2-SF3b components at the heart of the spliceosome. We then integrated co-complex purification information from 76 purifications of active spliceosomal complexes with our data and performed PPI clustering. This approach revealed several interesting dynamic PPI patterns with relevance for a better understanding of the splicing cycle. For example, changing PPIs during B to C transition (Figure 2) with one of the most central proteins, hPRP8, are found. Together with interaction competition experiments, these data suggest that during step 1 of splicing, hPRP8 interactions with the SF3b49 protein is replaced by hSLU7, positioning this essential second step factor close to the active site and that the DEAH-box helicases hPRP2 and hPRP16 cooperate through ordered interactions with the G-patch protein GPKOW.

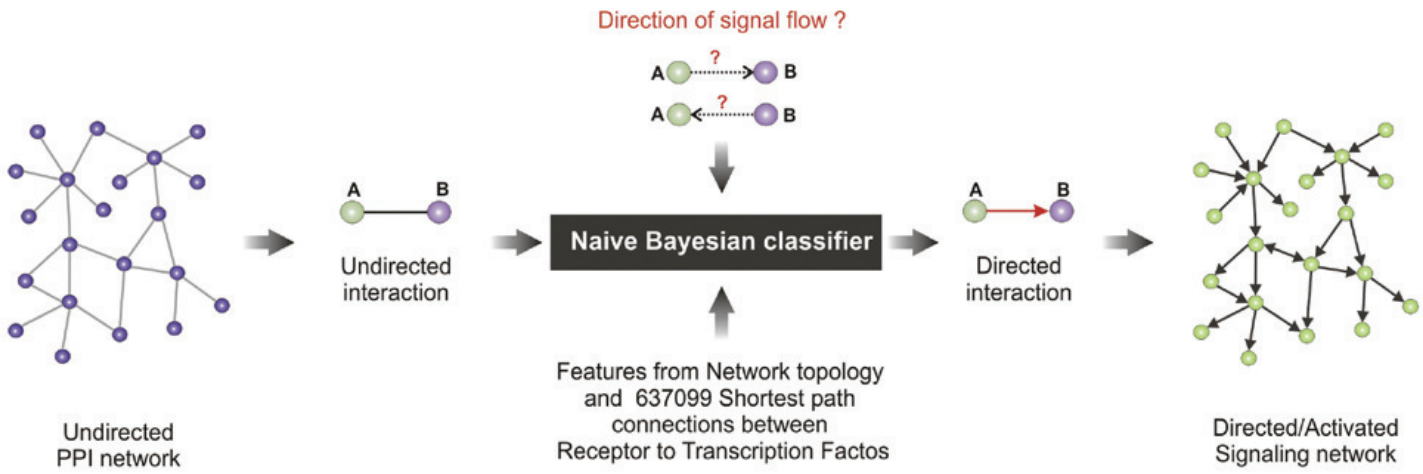


Figure 3: Inferring edge directions from PPI data (Vinayagam, Sci Signal 2011). For each interaction in the undirected PPI network, a naïve Bayesian classifier was used to predict the edge direction from topological network properties as well as shortest PPI paths connecting membrane receptors and transcription factors. An activated signaling network was assembled from all interactions that had a direction assigned.

In another study, performed together with the Wanker Lab (MDC, Berlin, Buch) we identified >2500 PPIs among human proteins broadly involved in cellular signaling. To provide information about how the signal is transmitted through the network, we developed a Bayesian learning strategy to assign direction to the interactions reflecting the potential signal flow among the proteins (Figure 3).
The resulting directed network is a unique resource for various modeling approaches. For example, we used the model to identify previously unknown modulators of the EGF/ERK pathway, of which 18 were validated with cell-based assays. It also enabled us to model EGF-induced protein phosphorylation dynamics. We could correlate *in vivo* phosphorylation dynamics with the output distance from the EGF/ERK pathway in our network resolving global protein phosphorylation events in a time-dependent manner.

In the two projects described, we exemplarily addressed PPI dynamics through combined experimental and computational approaches and successfully modeled how a signal spreads from an activated signaling pathway through a dense PPI network to the very distant proteins as well as identified crucial sites of changing PPI patterns that contribute to the exceptional compositional dynamics (and thus function) of the human spliceosome. In a next step, we want to take a direct experimental approach to analyze alterations of PPI patterns.

Planned developments

Conditional / modification-dependent protein interactions

In ongoing projects, we want to elucidate the role of posttranslational protein modifications (PTMs), such as phosphorylation (P) and methylation, for these dynamic processes and investigate how genetic variations, e.g. SNPs, may change protein-protein interaction patterns.

We have established a modified Y2H setup employing kinases to screen for P-dependent PPIs. We identified a novel P-dependent interaction between ADAP and Nck adaptor molecules that alter adhesion and migration of Jurkat T cells. We also contributed to the identification of a phosphorylation–triggered interaction between neuronal Fez1 and Munc18, mediating axonal transport of Syntaxin. The characterization of isoform-specific and P-dependent protein interactions between tumor suppressor protein NF2 (merlin) and AOF2, EMILIN1 and PIK3R3 suggests novel regulatory loops influencing NF2 conformation and function of the protein.

In a proteome wide approach, we have identified more than 300 novel pY-dependent PPIs that show high specificity with respect to human kinases and interacting proteins. P-dependent interactions are further analyzed in mammalian cell culture systems using e.g. co-immunoprecipitation, protein complementation and functional reporter readouts. The more detailed characterization of selected P-dependent GRB2 and PIK3R3 interactions exemplarily demonstrate how these PPIs are dynamically and spatially constrained to separate simultaneously triggered growth signals which are often altered in oncogenic conditions. Our screening approach is extended to other posttranslational protein modification such as methylation. Methyltransferase-substrate relationships discovered through Y2H-seq mapping of the methyltransferase interactome provide a reliable basis to exploit cellular functions of non-histone protein methylation. Finally, we integrate genetic variation data in our interaction studies and investigate how disease causing missense mutations change protein interaction profiles.

Cooperation within the institute

Within the institute, the Molecular Interaction Networks group cooperates with the following people and their groups: Ralf Herwig, Dept. of Vertebrate Genomics, on computational tools and network algorithms; Sebastiaan Meijsing, Dept. of Computational Molecular Biology, on splice variant specific protein interaction studies of GR; Hermann Bauer, Phillip Grote, and Heiner Schrewe, all from the Dept. of Developmental Genetics, on targeted protein interaction studies for proteins involved in mesoterm formation and of t-complex distorters; with Sascha Sauer, OWL, and David Meierhofer, Mass Spec group, on protein modification analyses, and with Bernd Timmermann (next generation sequencing).

General information

Complete list of publications (2009-2012)

2012

Chua JJ, Butkevich E, Worseck JM, Kittelmann M, Grønborg M, Behrmann E, Stelzl U, Pavlos NJ, Lalowski MM, Eimer S, Wanker EE, Klopfenstein DR, Jahn R (2012). *Phosphorylation-regulated axonal dependent transport of syntaxin 1 is mediated by a Kinesin-1 adapter.* Proc Natl Acad Sci U S A 109(15):5862-7

Hegele A*, Kamburov A*, Grossmann A, Sourlis C, Wowro S, Weimann M, Will CL, Pena V, Lührmann R, Stelzl U (2012). *Dynamic protein-protein interaction wiring of the human spliceosome*. Mol Cell 45(4):567-80

Kamburov A [#], Stelzl U, Herwig R (2012). *IntScore: a web tool for confidence scoring of biological interactions*. Nucleic Acids Res 2012 (40):W140-6

Worseck JM*, Grossmann A*, Weimann M*, Hegele A, Stelzl U (2012). *A stringent yeast two-hybrid matrix screening approach for protein-protein interaction discovery.* Methods Mol Biol 812:63-87

2011

Elefsinioti A, Saraç ÖS, Hegele A, Plake C, Hubner NC, Poser I, Sarov M, Hyman A, Mann M, Schroeder M, Stelzl U, Beyer A (2011). *Large-scale de novo prediction of physical protein-protein association*. Mol Cell Proteomics 10(11):M111.010629

Soler-López M*, Zanzoni A*, Lluís R, Stelzl U, Aloy P (2011). *Interactome mapping suggests new mechanistic details underlying Alzheimer's disease*. Genome Res 21(3):364-76

Vinayagam A*, Stelzl U*,#, Foulle R, Plassmann S, Zenkner M, Timm J, Assmus HE, Andrade-Navarro MA, Wanker EE[#](2011). *A directed protein interaction network for investigating intracellular signal transduction*. Sci Signal 4(189):rs8

2010

Sylvester M, Kliche S, Lange S, Geithner S, Klemm C, Schlosser A, Grossmann A, Stelzl U, Schraven B, Krause E, Freund C (2010). *Adhesion and degranulation promoting adapter protein (ADAP) is a central hub for phosphotyrosine-mediated interactions in T cells*. PLoS One 5(7):e11708

Vinayagam A, Stelzl U, Wanker EE (2010). *Repeated two-hybrid screening detects transient protein-protein interactions*. Theor Chem Acc 125(3-6):613-19

2009

Palidwor GA, Shcherbinin S, Huska MR, Rasko T, Stelzl U, Arumughan A, Foulle R, Porras P, Sanchez-Pulido L, Wanker EE, Andrade-Navarro MA (2009). *Detection of alpha-rod protein repeats using a neural network and application to huntingtin*. PLoS Comput Biol 5(3):e1000304

Venkatesan K*, Rual JF*, Vazquez A*, Stelzl U*, Lemmens I*, Hirozane-Kishikawa T, Hao T, Zenkner M, Xin X, Goh KI, Yildirim MA, Simonis N, Heinzmann K, Gebreab F, Sahalie JM, Cevik S, Simon C, de Smet AS, Dann E, Smolyar A, Vinayagam A, Yu H, Szeto D, Borick H, Dricot A, Klitgord N, Murray RR, Lin C, Lalowski M, Timm J, Rau K, Boone C, Braun P, Cusick ME, Roth FP, Hill DE, Tavernier J, Wanker EE, Barabási AL, Vidal M (2009). *An empirical framework for binary interactome mapping*. Nat Methods 6(1):83-90