

Using systems and structure biology tools to dissect cellular phenotypes

Aris Floratos,^{1,2} Barry Honig,^{2,3} Dana Pe'er,^{2,4} Andrea Califano^{1,2}

¹Department of Biomedical Informatics, Columbia University, New York, USA

²Center for Computational Biology and Bioinformatics, Columbia University, New York, USA

³Department of Biochemistry and Molecular Biophysics, Columbia University, New York, USA

⁴Department of Biological Sciences, Columbia University, New York, USA

Correspondence to

Dr Andrea Califano, Center for Computational Biology and Bioinformatics, Columbia University, 1130 St Nicholas Avenue, New York, NY 10032, USA; califano@c2b2.columbia.edu

Received 18 July 2011
Accepted 31 July 2011
Published Online First
10 November 2011

ABSTRACT

The Center for the Multiscale Analysis of Genetic Networks (MAGNet, <http://magnet.c2b2.columbia.edu>) was established in 2005, with the mission of providing the biomedical research community with Structural and Systems Biology algorithms and software tools for the dissection of molecular interactions and for the interaction-based elucidation of cellular phenotypes. Over the last 7 years, MAGNet investigators have developed many novel analysis methodologies, which have led to important biological discoveries, including understanding the role of the DNA shape in protein–DNA binding specificity and the discovery of genes causally related to the presentation of malignant phenotypes, including lymphoma, glioma, and melanoma. Software tools implementing these methodologies have been broadly adopted by the research community and are made freely available through geWorkbench, the Center's integrated analysis platform. Additionally, MAGNet has been instrumental in organizing and developing key conferences and meetings focused on the emerging field of systems biology and regulatory genomics, with special focus on cancer-related research.

INTRODUCTION

Cell phenotypes are determined by the concerted activity of thousands of gene products. These activities are coordinated by a complex network of molecular interactions that regulate the expression of genetic programs. Understanding the organization of this network is crucial to elucidating physiological mechanisms that are dysregulated in human disease. Historically, scientists have proceeded by painstakingly dissecting molecular event cascades, or pathways, using hypothesis-driven experimental approaches to test individual interactions—for example, a protein kinase phosphorylating a specific substrate. More recently, the availability of high-throughput genomic measurement technologies has made it possible to accelerate the rate of discovery by leveraging computational methods to reconstruct, model, and interrogate gene interaction networks on a whole-genome scale.

The MAGNet Center was created to support and systematically pursue such computational investigations. A basic tenet of our strategy has been the notion that by bridging and integrating molecular interactions across multiple levels of granularity (from the atomic to the functional level) and by using multiple data modalities (from structural coordinates to genetic and epigenetic variability), significant progress can be achieved, both within and across research domains. To that end, MAGNet

brings together interdisciplinary scientists from a wide range of domains to develop novel Structural and Systems Biology tools (table 1) for the interaction-based elucidation of cellular phenotypes and to validate these tools in the context of Driving Biological Projects (DBPs) (table 2).

Our efforts have resulted in several high-impact accomplishments. For instance, we have assembled the first experimentally validated, context-specific regulatory networks of cancer cells.^{1–4} Network interrogation has helped elucidate synergistic Master Regulators of the most aggressive subtype of Glioblastoma⁵ and driver genes in Melanoma.⁵ MAGNet investigators have also demonstrated that nucleotide sequence determines DNA minor groove width, which in turn determines the DNA-binding specificity of individual Hox transcription factors.⁶ The discovery that DNA shape plays a role in protein–DNA binding specificity sets the stage for a deeper integration of Systems and Structural Biology approaches. These examples are only a representative sample of a much broader research portfolio, which, since 2005, has resulted in more than 200 papers, including 56 in journals with Impact Factor ≥ 9.7 .

We have systematically promoted the dissemination of scientific knowledge produced by the Center by organizing and developing key conferences and meetings that have attracted hundreds of scientists each year. These include (a) the DREAM (Dialog for Reverse Engineering Assessments and Methods) conference and competition, to establish objective, community-based benchmarks to test reverse-engineering algorithms, (b) the RECOMB Systems Biology conference, and (c) the New York Academy of Science Systems Biology Interest Group. In addition, MAGNet investigators have played a key role in highlighting the increasing body of work in Cancer Systems Biology by co-chairing the 2011 Annual Meeting of the AACR and the 2011 AACR Conference on Cancer Systems Biology (Confronting the Complexity of Cancer), both with special focus on this emerging field.

To streamline the dissemination of analytical methods and data resources produced by the Center we have developed geWorkbench⁷ (<http://www.geworkbench.org/>), an open source, component-based bioinformatics platform that provides integrated access to MAGNet tools as well as many third-party analysis and visualization modules. geWorkbench leverages standards-based middleware grid technologies to offer seamless access to remote data, annotation, and computational servers, thus, enabling researchers with limited local resources to benefit from available public infrastructure.

Table 1 Software tools developed by MAGNet Center investigators

Tool name	Description	Lab
ARACNe*	Reconstruction of cellular networks from gene-expression data using information theoretic approaches; http://wiki.c2b2.columbia.edu/califanolab/index.php/Software/ARACNE	Andrea Califano
CNKB*	Cellular Networks Knowledge Base, a database of gene-interaction networks across multiple cellular phenotypes; http://wiki.c2b2.columbia.edu/workbench/index.php/Cellular_Networks_KnowledgeBase	Andrea Califano
CONNEXIC	Integration of copy-number variation and gene expression to identify driving cancer mutations and the processes they influence; http://www.c2b2.columbia.edu/danapeerlab/html/conexic.html	Dana Pe'er
Genatomy	Visualization and analysis of biological data, such as gene expression, genotypes, growth curves, copy number variation, etc; http://www.c2b2.columbia.edu/danapeerlab/html/genatomy.html	Dana Pe'er
IDEA*	Interactome dysregulation enrichment analysis	Andrea Califano
MARINA*	Inference of master regulators of cellular phenotypes; http://wiki.c2b2.columbia.edu/workbench/index.php/Master_Regulator_Analysis	Andrea Califano
MarkUs*	Structure-based function annotations of proteins; http://wiki.c2b2.columbia.edu/honiglab_public/index.php/Software:Mark-Us	Barry Honig
MatrixREDUCE*	Weight matrix discovery from gene expression or ChIP-chip data based on a biophysical model; http://bussemaker.bio.columbia.edu/software/MatrixREDUCE/	Harmen Bussemaker
MEDUSA*	Machine learning-based reconstruction of regulatory networks that predict the differential expression of target genes; http://cbio.mskcc.org/leslielab/software/medusa	Chris Wiggins, Christina Leslie
MINDy*	Genome-wide discovery of post-translational modulators of transcriptional interactions; http://wiki.c2b2.columbia.edu/califanolab/index.php/Software	Andrea Califano
MutaGeneSys	Use of genome-wide genotype data to estimate individual disease susceptibility; http://www1.cs.columbia.edu/~jds1/MutaGeneSys/	Itzik Pe'er
NetBoost*	Prediction of generative evolutionary models for gene-interaction networks; http://www.columbia.edu/~chw2/code.html	Chris Wiggins
Pudge*	Protein-structure prediction; http://wiki.c2b2.columbia.edu/honiglab_public/index.php/Software:PUDGE	Barry Honig
Skybase*	Database of computed protein-structure models, produced by the SkyLine pipeline; http://skybase.c2b2.columbia.edu/nescg3/nescg.php	Diana Murray
SkyLine*	Pipeline for reverse homology-based protein structure modeling	Diana Murray
Xplorigin	Population ancestry deciphering of different regions along an individual's genome; http://www.cs.columbia.edu/~itsik/Xplorigin/Xplorigin.htm	Itzik Pe'er

*Available as a geWorkbench plug-in. ARACNe, Algorithm for the Reconstruction of Accurate Cellular Networks; MINDy, Modulator Inference by Network Dynamics.

COLLABORATIVE PROJECTS HIGHLIGHTS

The methodologies and tools developed by the Center were validated through DBPs and other collaborative projects, coupling computational and experimental research. Seven DBPs have been completed to date, and three new ones have been initiated following the MAGNet renewal (table 2). Furthermore, providing strong evidence for the type of collaborative projects that can originate from the computational advances created by the center, MAGNet investigators initiated 41 collaborations

based on MAGNet tools, of which 31 have resulted in National Institutes of Health-funded activities. The following provide a few highlights of these collaborative research efforts.

Regulatory modules in normal and transformed B cells

Our goal was to understand the regulatory modules that determine B cell transition in the Germinal Center and transformation, with a focus on targets of the MYC and BCL6 proto-oncogenes. We used Bayesian evidence integration methods to

Table 2 Driving biological projects

DBP title	Investigators
Structural and energetic basis of cadherin binding specificity; http://magnet.c2b2.columbia.edu/?q=node/34	Larry Shapiro (E) Barry Honig (C)
Regulatory modules in normal and transformed B cells; http://magnet.c2b2.columbia.edu/?q=node/20	Riccardo Dalla-Favera (E) Andrea Califano (C) Conrad Gilliam (E) Andrey Rzhetsky (C)
Genomic and bioinformatics solutions to the search for genetic determinants of common, heritable disorders; http://magnet.c2b2.columbia.edu/?q=node/21	Richard Mann (E) Barry Honig (C)
Understanding and predicting transcription factor specificity; http://magnet.c2b2.columbia.edu/?q=node/35	Andrea Califano (C) Riccardo Dalla-Favera (E) Andrea Califano (C)
microRNA analysis in normal and neoplastic human B cell phenotypes; http://magnet.c2b2.columbia.edu/?q=node/36	Levi Garraway (E) Dana Pe'er (C)
Computational and functional dissection of drug targets in melanoma; http://magnet.c2b2.columbia.edu/?q=node/37	Richard Losick (E) Dennis Vitkup (C)
An integrated analysis of structural organization, design principles, and evolution across multiple genomes of a model developmental network; http://magnet.c2b2.columbia.edu/?q=node/38	Tom Tullius (E) Richard Mann (E) Barry Honig (C) Harmen Bussemaker (C)
Identifying Hox protein-specific DNA-binding sites and probing their shapes; http://magnet.c2b2.columbia.edu/?q=node/41	Peter Sorger (E) Dennis Vitkup (C) Cory Abate-Shen (E) Michael Shen (E) Andrea Califano (C)
Probabilistic dynamic modeling of the ErbB signaling pathways; http://magnet.c2b2.columbia.edu/?q=node/42	
Master regulators of tumorigenesis and drug sensitivity in prostate cancers; http://magnet.c2b2.columbia.edu/?q=node/43	

The first seven projects have been completed; the last three are ongoing. The designation next to an investigator's name indicates their role in the project (C, computational; E, experimental).

assemble a human B cell interactome (hBCi)^{8–10} comprising transcriptional and post-translational molecular interactions (inferred by Algorithm for the Reconstruction of Accurate Cellular Networks (ARACNe)^{11–13} and Modulator Inference by Network Dynamics (MINDy),^{2, 10, 14} respectively), literature evidence (inferred by the GeneWays algorithm¹⁵), and interactions from database sources. The hBCi, which is accessible via the Cellular Network Knowledge Base component in geWorkbench, includes a first map of the interface between signal-transduction and transcriptional networks in a eukaryotic cell;¹⁶ see figure 1. Interrogation of the hBCi with MAGNet algorithms such as IDEA¹⁰ and MARINA⁹ led to identification of regulatory modules and master regulators causally related to physiologic, pathologic, and drug-induced phenotypes. Algorithms developed by MAGNet were also used to identify activation of the NF- κ B pathway in the ABC (aggressive) subtype of diffuse large B cell lymphoma, as well as the specific spectrum of genetic alterations contributing to its activation, such as those for A20 and CARD11.¹⁷ These approaches have been critical in the elucidation of a variety of additional pathologic phenotypes, including glioma,³ T-ALL,¹³ breast-cancer progression,⁴ brain morphogenesis defects,⁴ and drug-induced phenotypes,¹⁰ and in the study of Germinal Center formation.⁹

Understanding and predicting transcription-factor specificity

Hox proteins are a family of transcription factors that define the body axis of developing embryos from fly to human. The source of their ability to read out specific DNA sequences had remained unknown for many years. Crystal structures of the Hox protein Scr and a second co-factor protein, Exd, revealed that the source of specificity resided in the minor groove of DNA. Specifically, while all Hox proteins share a common major-groove recognition mechanism, it is the minor groove interaction by an N-terminal region of the homeodomain that determines binding specificity¹⁸ (see figure 2). Remarkably, sequence readout is not accomplished via direct hydrogen bonds between proteins and DNA bases, as in the major groove, but rather through recognition of minor groove width, which is sequence-dependent. The biophysical basis of such recognition lies in the fact that sequence determines width, and width determines electrostatic potential, which is then recognized by specific positively charged arginines on the protein. This mode of protein–DNA recognition is quite general, as we established in a recent paper in

Nature.⁶ We are now continuing this work in the context of a new DBP aimed at discovering the full range of sequences recognized by different Hox proteins and then using experimental and computational tools aimed at determining their shape and electrostatic properties, with the ultimate goal of understanding how the members of an entire protein family recognize DNA in a sequence-specific fashion.

Computational and functional dissection of drug targets in Melanoma

Systematic characterization of cancer genomes has revealed a staggering number of diverse alterations that differ among individuals, so that their functional importance and physiological impact remains poorly defined. In order to identify which genetic alterations are functional, we combined computational modeling and functional genomics to provide an integrated view of tumor-specific alterations, using melanoma to demonstrate our approach. Specifically, we developed the Bayesian probabilistic CONEXIC algorithm⁵ to integrate copy number and gene-expression data in order to identify tumor-specific ‘driver’ aberrations, as well as the cellular processes they affect (figure 3). Using data from melanoma samples, we identified a list of 64 putative ‘drivers’ and the core processes affected by them. This list includes many known driver genes (eg, MITF), which CONEXIC correctly identified and paired with their known targets. This list also includes novel ‘driver’ candidates including Rab27a and TBC1D16, both involved in protein trafficking. ShRNA-mediated silencing of these genes in short-term tumor-derived cultures determined that they are tumor dependencies and validated their computationally predicted role in melanoma (including target identification), suggesting that protein trafficking may play an important role in this malignancy.⁵ CONEXIC and its extensions are being successfully applied to additional cancers including glioblastoma, ovarian cancer, and breast cancer.

OUTPUTS

The methodological and algorithmic innovations of the Center have generated a large number of software tools and databases (table 1) which have been instrumental in supporting biomedical research, resulting in more than 200 scientific publications (<http://magnet.c2b2.columbia.edu/?q=node/11>). MAGNet

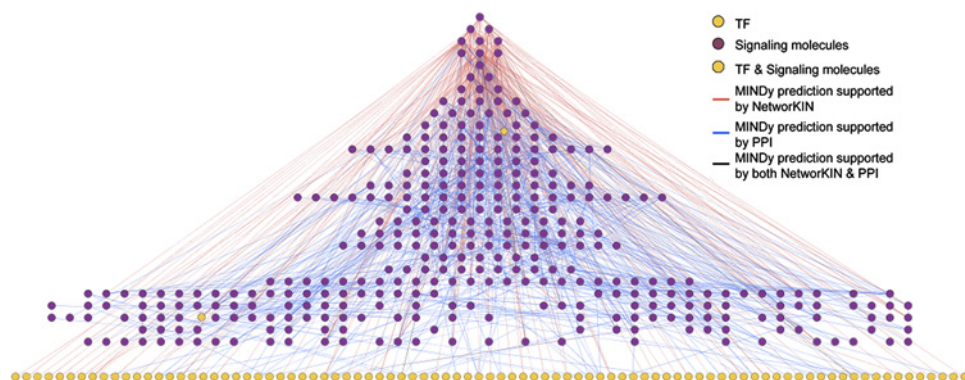
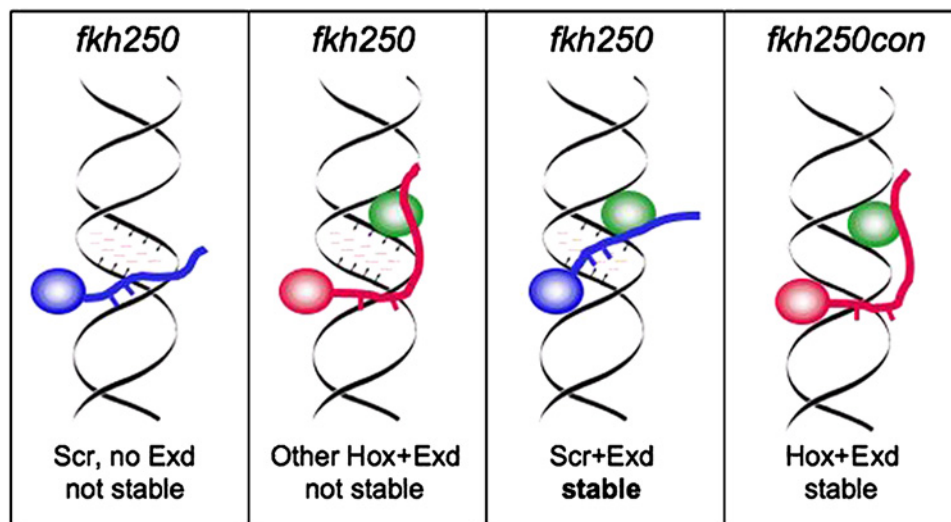


Figure 1 Visualization of the signalome-transfactome network by integrating Modulator Inference by Network Dynamics (MINDy) predictions with the Protein–Protein Interface Database (PPIDB) and Kinase–Substrate Interactions (KSIDB) Database. Two types of interactions are represented in the network: (1) MINDy predicted signaling protein (SP)–transcription factor (TF) interactions supported by PPIDB or KSIDB (ie, modulatory interactions predicted by MINDy that have physical interaction evidence); (2) MINDy predicted SP–SP interactions supported by PPIDB or KSIDB (ie, between modulators predicted by MINDy of the same TF, and are supported by physical interaction evidence). Depending on the source of evidence, these interactions can be either undirected (supported by known PPIs) or directed (supported by NetworkKIN, ie, kinase \rightarrow substrate).

Figure 2 First three panels: Hox/Exd dimers in the presence of the *fkh250*, an Scr/Exd specific binding site, which has a narrow minor groove (small arrows) and negative electrostatic potential (pink dashes) in the center of the binding site. Fourth panel: *fkh250con*, a non-specific Hox/Exd binding site, which does not have these characteristics. Scr/Exd, but not other Hox/Exd dimers, can effectively bind to the *fkh250* binding site because Exd positions a normally unstructured peptide so that the basic side chains (short blue lines) can insert into the negative pocket formed by the narrow minor groove. In contrast, because *fkh250con* does not have this negative pocket, it is less selective and can bind multiple Hox/Exd dimers.



investigators make their tools freely available to the academic community from their lab websites, through code downloads and access to web applications. Additionally, we have developed geWorkbench,⁷ an integrative analysis platform to serve as a one-stop, uniform framework for making the center resources available to the research community.

geWorkbench uses a plug-in framework to integrate computational resources and to make them available through a unified user interface. It leverages the caGrid infrastructure¹⁹ to expose computationally demanding analyses as grid services that can be

executed on the Center's compute cluster. In addition to MAGNet tools, the more than 70 modules in geWorkbench provide access to many popular third-party applications and databases such as Cytoscape,²⁰ GenePattern,²¹ BLAST,²² and the caArray microarray data repository.²³ Its open-source, component-based architecture enables software developers to easily add new functionality in the form of plug-in modules and to use available caGrid tooling (Introduce Toolkit²⁴) to streamline access to server side components. Finally, genSpace,²⁵ an innovative logging framework that collects and mines data about

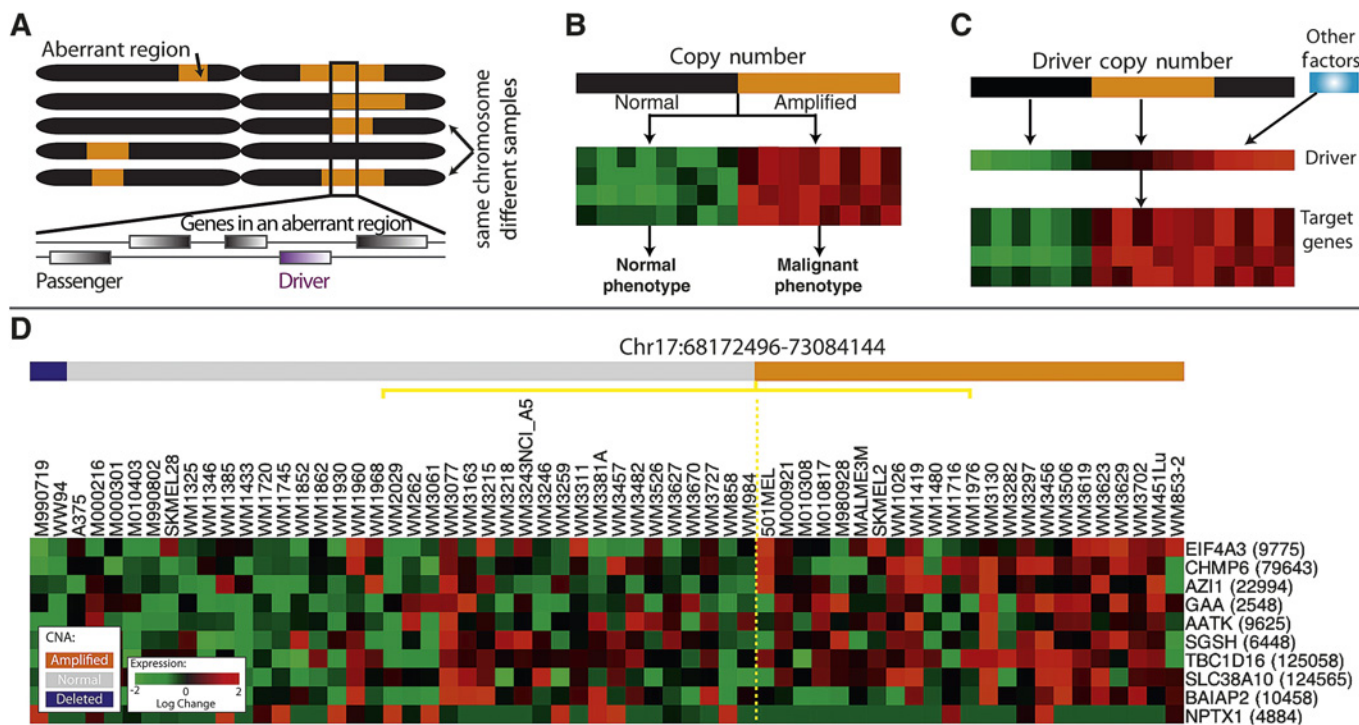


Figure 3 ‘Driver mutation’, a genetic alteration that provides the tumor cell with a growth advantage during carcinogenesis or tumor progression. We reasoned that driver mutations might leave a genomic ‘footprint’ that can assist in distinguishing between driver and passenger mutations based on the following assumptions: (A) a driver mutation should occur in multiple tumors more often than would be expected by chance; (B) a driver mutation may be associated (correlated) with the expression of a group of genes that form a ‘module’; (C) copy number aberrations often influence the expression of genes in the module via changes in expression of the driver; (D) gene expression is particularly useful for identifying candidate drivers within large amplified or deleted regions of a chromosome, whereas genes located in a region of genomic copy gain/loss are indistinguishable in copy-number aberrations, expression permits the ranking of genes based on how well they correspond with the phenotype. Reprinted with permission from Akavia *et al.*⁵

how various modules are being utilized, offers the possibility for novice users to learn from their more advanced peers.

The geWorkbench project site (<http://www.geworkbench.org>) includes extensive documentation (manuals, tutorials, and code samples) to facilitate adoption and to encourage open-source development. Additional support is available through the caBIG Molecular Analysis Tools Knowledge Center (<https://cabig-kc.nci.nih.gov/Molecular/KC/>), which offers access to user and developer forums, a comprehensive knowledge base, and many movies, training tutorials, and web demos. Adoption by the research community (one of the most important goals of the NCBC program) has been robust: there have been more than 8000 geWorkbench downloads since February 2006, corresponding to more than 400 unique users (statistics collected through NCI's GForge download site and the genSpace logs). Finally, we note that geWorkbench is included in the caBIG Life Sciences Distribution (LSD),²⁶ one of three 'bundles' of tools that have been designed to help support and streamline clinical trials, imaging, tissue banking, and integrative cancer research.

CURRENT AND FUTURE GOALS

The long-range goal of our center is the creation of detailed, context-specific molecular interaction networks that can be interrogated to elucidate cellular mechanisms and their dysregulation in disease. During the past few years, we have been able to make significant progress toward this goal through the development of many novel algorithms for the dissection of transcriptional, post-transcriptional, and post-translational interactions.^{1 2 10 11 14 15 27} The network models generated by these algorithms have undergone extensive experimental validation, and their interrogation has helped elucidate novel mechanisms and master regulators of drug-induced, developmental, and tumor-related phenotypes.^{1 3 4 13 28 29}

As we move forward, we will continue refining our network analysis methods by including additional sources of information. First and foremost, we aim to systematically incorporate three-dimensional protein structure information, where available from databases, as well as those from modeling approaches. For instance, we are currently developing methods (PREdicting Protein Interactions; PREPPI) that allow the inference of protein-protein interactions from structural data. Another thread of investigation will focus on integrating genetic alteration data into our regulatory models. As we have already demonstrated,⁵ this approach can be extremely effective in identifying driver mutations in cancer. Our plan is to combine genetic mutation and copy-number variation information with the transcriptional, post-transcriptional, post-translational, and cell-cell interaction layers of gene interaction networks to generate global and computationally uniform regulatory models.

Eventually, our objective is to apply our computational tools to systematically model and explore the regulatory and signaling networks of a wide range of normal and pathological cellular phenotypes, in order to understand and characterize the molecular machinery that determines their behavior. The feasibility of this approach is strongly supported by the results of our DBPs and collaborative projects, and has led to seminal findings in the field of cancer (B cell lymphoma, prostate, glioblastoma, melanoma, etc). As multidimensional genomic data are becoming available for a number of disease-related and normal conditions, MAGNet tools are poised to play an important role across all major biomedical research areas. Furthermore, we are committed to making our technologies accessible to the widest possible scientific audience by freely distributing the Center's tools, models, and data through geWorkbench.

Funding National Institutes of Health, grant number U54CA121852.

Competing interests None.

Provenance and peer review Commissioned; internally peer reviewed.

REFERENCES

1. Basso K, Margolin AA, Stolovitzky G, *et al*. Reverse engineering of regulatory networks in human B cells. *Nat Genet* 2005;**37**:382–90.
2. Wang K, Saito M, Bisikirska BC, *et al*. Genome-wide identification of post-translational modulators of transcription factor activity in human B cells. *Nat Biotechnol* 2009;**27**:829–39.
3. Carro MS, Lim WK, Alvarez MJ, *et al*. The transcriptional network for mesenchymal transformation of brain tumours. *Nature* 2010;**463**:318–25.
4. Zhao X, D'Arca D, Lim WK, *et al*. The N-Myc-DLL3 cascade is suppressed by the ubiquitin ligase Huwe1 to inhibit proliferation and promote neurogenesis in the developing brain. *Dev Cell* 2009;**17**:210–21.
5. Akavia UD, Litvin O, Kim J, *et al*. An integrated approach to uncover drivers of cancer. *Cell* 2010;**143**:1005–17.
6. Rohs R, West SM, Sosinsky A, *et al*. The role of DNA shape in protein-DNA recognition. *Nature* 2009;**461**:1248–53.
7. Floratos A, Smith K, Ji Z, *et al*. geWorkbench: an open source platform for integrative genomics. *Bioinformatics* 2010;**26**:1779–80.
8. Lefebvre C, Lim WK, Basso K, *et al*. A context-specific network of protein-DNA and protein-protein interactions reveals new regulatory motifs in human B cells. RECOMB Satellite Workshop on Systems Biology, San Diego, CA, December 2006. *Lect Notes Comput Sci* 2007;**4532**:42–56.
9. Lefebvre C, Rajbhandari P, Alvarez MJ, *et al*. A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. *Mol Syst Biol* 2010;**6**:377.
10. Mani KM, Lefebvre C, Wang K, *et al*. A systems biology approach to prediction of oncogenes and molecular perturbation targets in B-cell lymphomas. *Mol Syst Biol* 2008;**4**:169.
11. Margolin AA, Nemenman I, Basso K, *et al*. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC bioinformatics* 2006;**7**(Suppl 1):S7.
12. Margolin AA, Wang K, Lim WK, *et al*. Reverse engineering cellular networks. *Nat Protoc* 2006;**1**:662–71.
13. Palomero T, Lim WK, Odom DT, *et al*. NOTCH1 directly regulates c-MYC and activates a feed-forward-loop transcriptional network promoting leukemic cell growth. *Proc Natl Acad Sci U S A* 2006;**103**:18261–6.
14. Wang K, Banerjee N, Margolin AA, *et al*. Genome-wide discovery of modulators of transcriptional interactions in human B lymphocytes. RECOMB 2006. *Lect Notes Comput Sci* 2006;**3909**:348–62.
15. Zhetsky A, Iossifov I, Koike T, *et al*. GeneWays: a system for extracting, analyzing, visualizing, and integrating molecular pathway data. *J Biomed Inform* 2004;**37**:43–53.
16. Wang K, Alvarez MJ, Bisikirska BC, *et al*. Dissecting the interface between signaling and transcriptional regulation in human B cells. *Pac Symp Biocomput* 2009:264–75.
17. Compagno M, Lim WK, Grunn A, *et al*. Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. *Nature* 2009;**459**:717–21.
18. Joshi R, Passner JM, Rohs R, *et al*. Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell* 2007;**131**:530–43.
19. Saltz J, Oster S, Hastings S, *et al*. caGrid: design and implementation of the core architecture of the cancer biomedical informatics grid. *Bioinformatics* 2006;**22**:1910–16.
20. Shannon P, Markiel A, Ozier O, *et al*. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;**13**:2498–504.
21. Reich M, Liefeld T, Gould J, *et al*. GenePattern 2.0. *Nat Genet* 2006;**38**:500–1.
22. Altschul SF, Gish W, Miller W, *et al*. Basic local alignment search tool. *J Mol Biol* 1990;**215**:403–10.
23. Heiskanen M, Lorenz J, Bian X, *et al*. Cancer microarray informatics (caArray) data management and analysis tools at the National Cancer Institute (NCI) Center for Bioinformatics [abstract]. *Proc Am Assoc Canc Res* 2005;**46**. <http://aacrmeetingabstracts.org/cgi/content/abstract/2005/1/7-a>.
24. Hastings S, Oster S, Langella S, *et al*. Introduce: an open source toolkit for rapid development of strongly typed grid services. *Journal of Grid Computing* 2007;**5**:407–27. <http://www.springerlink.com/content/u301u225wg5356w3/>.
25. Murphy C, Sheth S, Kaiser G, *et al*. genSpace: exploring social networking metaphors for knowledge sharing and scientific collaborative work. *1st International Workshop on Social Software Engineering and Applications*; 2008. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?reload=true&arnumber=4686308.
26. caBIG. Life Sciences Distribution (LSD). 2008. <https://cabig.nci.nih.gov/adopt/LSD/>.
27. Foat BC, Morozov AV, Bussemaker HJ. Statistical mechanical modeling of genome-wide transcription factor occupancy data by MatrixREDUCE. *Bioinformatics* 2006;**22**:e141–9.
28. Chen BJ, Causton HC, Mancenido D, *et al*. Harnessing gene expression to identify the genetic basis of drug resistance. *Mol Syst Biol* 2009;**5**:310.
29. Basso K, Saito M, Sumazin P, *et al*. Integrated biochemical and computational approach identifies BCL6 direct target genes controlling multiple pathways in normal germinal-center B cells. *Blood* 2010;**115**:975–84.