Department of Systems and Computational Biology

Albert Einstein College of Medicine of Yeshiva University

Faculty Research Interests
2015-2016

Introduction

The Department of Systems and Computational Biology, through innovative research and education, focuses its efforts on advancing the understanding of living systems as a whole by promoting a new approach to biology. Our research combines theoretical and experimental approaches aimed at explaining how the higher-level properties of complex biological systems materialize from the interactions among their parts.

The faculty develop research and education programs that embrace engineering, computational, mathematical and physical sciences as an integral part of the Biological and Biomedical sciences, leading to the foundation of a Systems and Computational Biology discipline. We seek to form an academic environment that benefits from and respects the value of these existing disciplines. An important part of this fusion entails serious, sustained and multidisciplinary research programs in pursuit of fundamental questions in Biology, ranging from the function of biological systems to an understanding of the evolution of life's diversity.

The main challenges currently facing Biology involve the understanding of biological phenomena at a holistic, systems level without neglecting the reductionist, detailed information approach related to its individual components. This is in contrast to the "massively parallel reductionism" view. That is, despite the availability of a vast amount of information, the focus, nonetheless, remains on the individual components. The academic community has recognized, once again, the need for integrated, systems-level strategies that combine computational and experimental tools, as well as evolutionary-based inferences in addressing these challenges. It is this integration that defines the department's activity

Faculty, Research Fellows and Graduate Students

Faculty

Aviv Bergman, Ph.D., Professor and Founding Chairman
Eduardo Fajardo, Ph.D., Associate
Andras Fiser, Ph.D., Professor
Libusha Kelly, Ph.D., Assistant Professor
Jessica Mar, Ph.D., Assistant Professor
Parsa Mirajhi, M.D., Ph.D., Research Associate Professor
Ian Willis, Ph.D.*, Professor (Department of Biochemistry)
Yinghao Wu, Ph.D., Assistant Professor

Research Fellows

Julia Brown, Ph.D. (Kelly Lab)
William Chang, Ph.D. (Kelly Lab)
Jiawen Chen, Ph.D. (Wu Lab)
Laurence de Torrente, Ph.D. (Mar Lab)
Eng Hui Yap, Ph.D. (Fiser Lab)
Zhong Ru Xie, Ph.D. (Wu Lab)

Graduate Students

Daniel Biro, MSTP Student (Advisor: Bergman)
Raymund Bueno, Graduate Student (Advisor: Mar)
Kelly Burke, Graduate Student (Advisor: Bergman/Fox)
Nelson Gil, MSTP Student (Advisor: Fiser)
Leah Guthrie, Graduate Student (Advisor: Kelly)
Joaquin Pechuan Jorge, Graduate Student (Advisor: Bergman)

Bioinformatic Analysts

Carlos Madrid-Aliste (Fiser Lab)
Al Tucker (Fiser Lab)

**Secondary appointment in the Department of Systems and Computational Biology*

Requirements for a Ph.D. in Systems and Computational Biology

A. Courses

Students who wish to pursue a Ph.D. in Systems and Computational Biology must successfully complete seven courses. Introduction to Systems Biology: Theory and Case Studies and the Systems Biology Seminar must both be successfully completed. The other five courses will be determined with the help of the student's mentor and advisory committee, although Graduate Biochemistry and either Molecular Genetics or Gene Expression are strongly recommended courses for all students.

B. Advisors

By the end of the first year, all graduate students choose a faculty member to serve as their primary research advisor, as well as an advisory committee made up of three to five faculty members (all will not necessarily be from the Department of Systems and Computational Biology).

C. Qualifying Examination

Each candidate for the Ph.D. degree must satisfactorily complete a qualifying examination, typically taken after the second year of study. By this time, the student should have fulfilled the bulk of his or her coursework. The exam includes both an oral presentation and written proposal on the student's proposed thesis project, and will be assessed by a Qualifying Exam Committee.

D. Thesis Dissertation

1. First Author Manuscript: The Graduate School requires that at least one first-author manuscript must be published before a student can defend his or her thesis. If a manuscript is not published, then the final draft of a Submitted, In Revision, or In Press first-author manuscript must be appended to the thesis.
2. Thesis Defense Committee: According to designated criteria established by the Graduate Committee, all Thesis Defense Committees must be approved by the Assistant Dean. The Committee must include a minimum of five faculty, one of which will be designated as the Committee Chair and two of which must be from the Department of Systems and Computational Biology. The thesis must be presented to this committee at least 3 weeks prior to the thesis defense.
3. Public Seminar: The presentation of a public seminar at the College of Medicine is required for successful completion of the Ph.D. degree.
4. Thesis Defense: All candidates for the Ph.D. degree must submit a satisfactory dissertation and pass an oral examination (defense) of their thesis.

E. Other Requirements

The department Journal Club/Works-in-Progress meets once a month. Seminars are held monthly during the academic year. Additional seminars in specific areas are also held throughout the semester. Attendance at these activities is required.

Please see <www.einstein.yu.edu/phd> for a more detailed guide to requirements and policies.

Aviv Bergman, Ph.D.
Professor and Founding Chairman, Department of Systems and Computational Biology
Price Center, Room 153C
aviv@einstein.yu.edu; (718) 678-1063

Selective Research Topics from the Bergman Lab:

Topology of biological networks

We study the relationship between the topology of biological networks and their functional (e.g. robustness) and evolutionary (e.g. polymorphism and divergence) properties. It has been conjectured that genes with a large number of downstream targets are more highly conserved, and when compromised, will tend to have a larger effect on network functioning than sparsely connected genes. However, we have shown that 'topdown' inferences of biological properties based on simple measures such as number of targets, are of limited utility. We argue that such lack of predictive power is the result of a composite effect in which certain sub-networks obeying a strong correlation between biological function and simple measures, coexist with other sub-networks having no correlation at all. We have demonstrated that more detailed information, e.g., dynamic gene-expression data, and the specifics of the genetic background, are needed to make meaningful functional and evolutionary inferences.

Investigations with an evolutionary perspective, such as these, can also be extended to biomedical research of phenotypic traits resulting from complex genetic interactions, including Cancer, Diabetes, Hypertension and Aging, as well as mechanistic models of the immune system. Indeed, we have successfully applied methodologies adopted from evolutionary theory to identify genes associated with extreme longevity as well as their targets, age-related disease genes.

Computational Immunology and somatic hypermutation

Somatic hypermutation (SHM) is a key process in the generation of antibody diversity that normally operates in antibody-forming B cells by introducing point mutations into the variable regions of immunoglobulin (Ig) heavy and light chain genes. SHM is initiated when the highly mutagenic enzyme activation-induced deaminase (AID) generates C→U mutations by deaminating cytosines preferentially at WRC hotspot motifs (where W=A/T, R=G/A and C is the mutated base). In collaboration with Matthew Scharff (Department of Cell Biology, Albert Einstein College of Medicine), we use computational and statistical methods together with relevant experimental data to improve our understanding of the molecular mechanisms underlying SHM.

How does the target sequence affect AID activity? To study the behavior of AID and the role of the target sequence, we have used computational methods to compare mutated sequences from three different models of AID activity: (a) an *in vivo* mouse model, (b) an *in*

vitro model which captures essential biochemical activity of AID on DNA, and (c) an *in silico* model which simulates only hotspot targeting. This analysis suggests that there is considerably more complexity involved in the mutation process than can be described by simple of WRC hotspot motifs. We have also found strong differences between the two strands (transcribed and non-transcribed) in terms of the similarity between the models. A potential clue comes from differences in the profile of inter-mutational distances between the two strands, which suggest the existence of a complex interplay between the enzyme structure and the sequence.

Evolution of gene regulatory networks

There is little doubt that plasticity in gene regulatory networks plays a key role in evolution, particularly in developmental networks. We use computational and mathematical models of gene networks to investigate key evolutionary questions and generate novel hypotheses. Where possible we also use relevant biological data to confirm theoretical findings.

How does degeneracy in transcription factor binding motifs affect evolution of *cis*-regulatory regions? In collaboration with Andras Fiser (SCB, Albert Einstein College of Medicine) we are developing structural models of transcription factor – DNA interactions in which we predict binding affinities for all possible interactions. The predicted binding affinities have been integrated with existing evolutionary models, enabling us to address questions concerning the evolution of regulatory motifs. Turnover of transcription factor binding sites is widespread in both insects and mammals, yet is poorly understood. Using our modeling framework we aim to understand what factors (e.g. motif degeneracy or selection) influence turnover rates.

What is fate of duplicated genes in networks? Several explanations have been proposed to explain the unexpectedly high retention of duplicate genes. One popular theory is the duplication-degeneration-complementation (DDC) model, which proposes that following gene duplication the two gene copies degenerate to perform complementary functions that jointly match that of the single ancestral gene, a process also known as subfunctionalization. However, the DDC model is gene-centric, and does not take into account the network context. Using computational models of evolving gene networks we have analyzed the fate of duplicate genes and found that network plasticity undermines the relevance of subfunctionalization, and that neofunctionalization (recruitment of novel interactions) plays a more predominant role than was previously thought.

How did sexual reproduction evolve? The prevalence of sexual reproduction, as opposed to asexual reproduction, remains one of the most perplexing phenomena in evolutionary biology. We have used computational modeling to explore the role played by epistasis, a condition in which mutations cause a greater change in fitness when combined than would be expected from their individual effects.

Caloric Restriction

Caloric restriction (CR) is a major intervention conclusively shown to extend lifespan in many organisms including mammals, birds, nematodes, flies and even unicellular species. Besides extending lifespan, CR also has been shown to prevent age-associated diseases and keeps organisms in a relatively youthful and healthy state compared to the ad libitum fed counterparts. These observations suggest that the somatic maintenance functions (e.g., cellular error-checking and damage repair) may be up-regulated in animals under CR conditions.

We try to derive a general quantitative and predictive theory, from physical energetic viewpoints, for understanding CR's effects on retarding aging and maintenance. We hypothesize that the longevity of an organism is correlated to biological pathways of maintenance of its integrities (e.g., repairing damage and error-checking), which are energetically costly, and that CR, counter-intuitively, by suppressing organisms' caloric energy supply and biosynthesis, alters the organisms' energy allocation strategy and channels additional energy/resource to the maintenance pathways, therefore retarding aging and extending the lifespan.

Based on principles of mass and energy balance and allometric scaling of metabolism and biosynthesis, we have developed an empirically grounded theoretical model that correctly predicts how organisms allocate energy between the synthesis of new biomass and the maintenance of existing biomass with normal food supply. We then try to extend the model and apply it to animals under CR. During growth, organisms need to do override work (indirect metabolic work) to store energy in new biomass. Our preliminary study suggests that because CR suppresses the energy storage in biomass, organisms do not devote as much metabolic work to do this storage as with normal food supply, therefore this amount of metabolic work can possibly be channeled to maintenance. We take the elongated lifespan as a measurement of enhanced maintenance during CR, and focus on four longstanding questions regarding CR's effects on maintaining organisms' integrity. (1) How does body temperature drop in CR animals influence CR's effect? (2) What is the relationship between intensity of CR and its effect? (3) With the same intensity and period of CR, how is an organism's adult body size correlated to CR's effect? (4) How does the age at which CR begins influence the CR's effect?

Besides the theoretical development, we also test hypotheses by a meta-analysis of empirical data on a diverse set of species from published literatures.

References:

Kunisaki, Y., Bruns, I., Scheiermann, C., Ahmed, J., Pinho1, S., Zhang, D., Mizoguchi, T., Wei1, Q., Lucas, D., Ito, K., Mar, J, C., Bergman, A., Frenette, P. S. Arteriolar niches maintain haematopoietic stem cell quiescence. *Nature.* 2013 September doi: 10.1038/nature12612

Pujato, M., MacCarthy, T., Fiser, A., and Bergman, A. The underlying molecular and network level mechanisms in the evolution of robustness in gene regulatory networks. *PLOS Computational Biology.* 2013 Jan. 9(1):e1002865

Conrad, T.M., Lewis, N.E. & Palsson, B.Ø. Microbial laboratory evolution in the era of genome-scale science. *Molecular Systems Biology* **7**, (2011).

MacCarthy T, Kalis SL, Roa S, Pham P, Goodman MF, Scharff MD, Bergman A. V-region mutation in vitro, in vivo, and in silico reveal the importance of the enzymatic properties of AID and the sequence environment. *Proc Natl Acad Sci USA* 106(21): 8629-8634, 2009.

MacCarthy T, Roa S, Scharff, MD, Bergman A. SHMTool: a webserver for comparative analysis of somatic hypermutation datasets. *DNA Repair* 8(1): 137-141, 2009.

Buckling A, Craig Maclean R, Brockhurst MA, Colegrave N. The Beagle in a bottle. *Nature* 457, 824-9(2009).

West GB, Bergman A. Toward a systems biology framework for understanding aging and health span. *J Gerontol A Biol Sci* 64(2):205-208, 2009.

Roa S, Avdievich E, Peled JU, MacCarthy T, Werling U, Kuang FL, Kan R, Zhao C, Bergman A, Cohen PE, Edelmann W, Scharff MD. Ubiquitylated PCNA plays a role in somatic hypermutation and class-switch recombination and is required for meiotic progression. *Proc Natl Acad Sci USA* 105(42): 16248-53, 2008.

MacCarthy T, Bergman A. The limits of subfunctionalization. *BMC Evol Biol* 7: 213, 2007.

Bergman A, Atzmon G, Ye K, MacCarthy T, Barzilai N. Buffering mechanisms in aging: a systems approach toward uncovering the genetic component of aging. *PLoS Comput Biol* 3(8): e170, 2007.

MacCarthy T, Bergman A. Coevolution of robustness, epistasis, and recombination favors asexual reproduction. *Proc Natl Acad Sci USA* 104(31), 12801-12806, 2007.

Siegal ML, Promislow DE, Bergman A. Functional and evolutionary inference in gene networks: does topology matter? *Genetica* 129(1): 83-103, 2006.

MacCarthy T, Pomiankowski A, Seymour R. Using large-scale perturbations in gene network reconstruction, *BMC Bioinformatics*, 6:11, 2005.

Masal J, Bergman A. The evolution of the evolvability properties of the yeast prion [PSI+]. *Evolution* 57(7): 1498-1512, 2003.

Bergman A, Siegal ML. Evolutionary capacitance as a general feature of complex gene networks. *Nature* 424(6948): 549-552, 2003.

MacCarthy T, Seymour R, Pomiankowski A. The evolutionary potential of the Drosophila sex determination gene network, *J Theor Biol* 225: 461-468, 2003.

Elena, S.F. & Lenski, R.E. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature Reviews Genetics* **4**, 457-69 (2003).

Siegal ML, Bergman A. Waddington's canalization revisited: developmental stability and evolution. *Proc Natl Acad Sci USA* 99(16): 10528-10532, 2002.

Mac Lane, S. *Categories for the Working Mathematician (Graduate Texts in Mathematics)*. 326(Springer: 1998).4.

Gould, S. *Full House: the spread of excellence from Plato to Darwin*. (Three Rivers Press: 1997).

Carroll, S.B. Chance and necessity: the evolution of morphological complexity and diversity. *Nature* **409**, 1102-9(2001).3. Buss, L.W. *The evolution of individuality*. (Princeton University Press: Princeton, 1987).

Buss, L.W. *The evolution of individuality*. (Princeton University Press: Princeton, 1987).

Andras Fiser, Ph.D.
Professor, Department of Systems and Computational Biology
Professor, Department of Biochemistry
Price Center, Room 453A
andras.fiser@einstein.yu.edu, (718) 678-1068

Computational protein structure modeling and design

The main interest of our group is to study the evolution of protein structures from a modular perspective and to develop theoretical methods based on these observations to model and design protein structures and functions. We are developing methods to design new molecular shapes, either by redesigning existing proteins into biologically more viable shapes or to explore thermodynamically possible molecular architectures. In terms of modeling we are developing techniques in hybrid modeling, specifically, using NMR chemical shift information for de novo prediction of protein structures.

T-cell costimulatory proteins

We are interested to understand how co-stimulation is enacted on a molecular level in the immunological synapse. We are exploring the molecular classification, recognition, regulation of Immunoglobulin Superfamily proteins playing role in T-cell recognition, and modulating the immune response. We are developing techniques that combine molecular design with docking to establish cognate partners within this essential set of proteins. This work is conducted within the Protein Structure Initiative.

Glioblastoma multiforme stem cells

Despite the advances in surgical, radiation and chemotherapy treatments of human glioblastoma (GBM) the prognosis remains poor with fewer than 3% of glioma patients alive 5-years after diagnosis. Molecular targeted therapies hold the promise of providing new anticancer treatments that are more effective than traditional therapies. In a four-way collaboration with brain surgeons, molecular biologists and a high throughput sequencing facility, we are exploring molecular signatures of GBM stem cells from the perspective of gene expression, RNA editing, RNA methylation.

References:

Khafizov K, Madrid-Aliste C, Almo SC, Fiser A. Trends in structural coverage of the protein universe and the impact of the Protein Structure Initiative. *Proc Natl Acad Sci U S A* (2014) **111**(10) : 3733-8

Yap EH, Rosche T, Almo S, Fiser A. Functional clustering of immunoglobulin superfamily proteins with protein-protein interaction information calibrated hidden markov model

sequence profiles. *J Mol Biol* (2014) **426**(4) : 945-61

Menon V, Vallat BK, Dybas JM, Fiser A. Modeling proteins using a super-secondary structure library and NMR chemical shift information. *Structure* (2013) **21**(6) : 891-9

Rubinstein R, Ramagopal UA, Nathenson SG, Almo SC, Fiser A. Functional classification of immune regulatory proteins. *Structure* (2013) **21**(5) : 766-76

Pujato M, MacCarthy T, Fiser A, Bergman A. The underlying molecular and network level mechanisms in the evolution of robustness in gene regulatory networks. *PLoS Comput Biol* (2013) **9**(1) : e1002865

Fajardo JE, Fiser A. Protein structure based prediction of catalytic residues.
BMC Bioinformatics (2013) **14**, 63

Libusha Kelly, Ph.D.
Assistant Professor, Department of Systems and Computational Biology
Price Center, Room 553B
Libusha.kelly@einstein.yu.edu; (718) 678-1148

We study how ecosystem-level genetic variability influences adaptive responses to dynamic environments.

Probing the influence of genetic variation in the human genome and microbiome on drug response

My lab is at the forefront of microbial ecology, studying the interplay between diversity and functional capacity in natural microbial communities. We focus on: 1) the capacity for microbes to metabolize excreted drugs in human populations, and 2) the influence of mobile element pools on microbial ecosystem functional capacity. Microbial genomes harbor tremendous diversity at the gene level even within closely related taxonomic groups. Microbes exchange DNA, with each other and with viruses, and can also take up DNA from the environment, leading to variability in the functional capacity of individual cells. Microbial ecosystems, therefore, are a social network of interacting and mobile genes with the capacity for tremendous functional plasticity. Microbes in the gut carry enzymes with the potential to metabolize excreted drugs, some of which cause adverse drug responses in patients. We study the abundance and phylogenetic distribution of microbial enzymes in human guts to predict the capacity of patients to metabolize drugs. Our focus is on the question: what forces enable genetic mobility, or information flow, in microbial ecosystems between distantly related bacteria, and how do these forces contribute to the evolution of community functions? The goals of the lab are to develop a pharmacokinetics of the human microbiome by incorporating the many enzymes with the potential to interact with excreted drugs and to predict gene mobility and spread to enable targeted manipulation of the metabolic capacity of microbial communities in diverse environments.

Tracking and predicting the flow of genetic information in a model microbial/viral ecosystem.

A key challenge in biology is determining the mechanisms that make biological systems stable to perturbation. We find that genomic flexibility enables both microbes and the viruses that infect them to adapt to variable environmental conditions in marine ecosystems. For example, viruses that infect marine cyanobacteria (cyanophages) often carry genes with orthologs in their cyanobacterial hosts, and the frequency of these genes can vary with habitat. Numerous shared phage/host genes differed in relative frequency between environments, including genes related to phosphorous acquisition, photorespiration, photosynthesis and the pentose phosphate pathway, possibly reflecting environmental selection for these genes in cyanophage genomes. We found a strong emergent signal related to phosphorous availability; a higher fraction of viral genomes from

relatively low-phosphorus environments–the Sargasso and Mediterranean Sea–contained host-like phosphorus assimilation genes compared with those from the N. Pacific Gyre. These genes are known to be upregulated when the host is phosphorous starved, a response mediated by pho box motifs in phage genomes that bind a host regulatory protein. Eleven cyanomyoviruses have predicted pho boxes upstream of the phosphate-acquisition genes *pstS* and *phoA*; eight of these have a conserved cyanophage-specific gene (*PhCOG173*) between the pho box and *pstS*. *PhCOG173* is also found upstream of other shared phage/host genes, suggesting a unique regulatory role during the infection cycle. Pho boxes are also found upstream of high light-inducible (*hli*) genes in cyanophages, suggesting that this motif may have a broader role than regulating phosphorous-stress responses in infected hosts or that these hlis are involved in the phosphorous-stress response. Future work will continue to explore how genetic exchange between bacteria and viruses in additional marine and human-associated microbial ecosystems reveals information about environmental conditions and influences the ability of bacteria to adapt to different environments.

Selected References:

† Indicates that authors contributed equally to the publication

Xenobiotic metabolism and the human microbiome

Structure and Inhibition of GI Microbiome Targets that Alleviate Cancer Drug Toxicity Bret D. Wallace†, Adam B. Roberts†, Rebecca Pollet, Madhu Kumar Venkatesh, Leah Guthrie, Sara O'Neil, James Ingle, Sara J. Robinson, Makani Dollinger, Esteban Figueroa, Sarah R. McShane, Jian Jin, Stephen V. Frye, William Zamboni, Sridhar Mani, **Libusha Kelly**, and Matthew R. Redinbo‡ Submitted, *Chemistry and Biology*

Smith VH, Rubinstein RJ, Park S, **Kelly L**, Klepac-Ceraj V. Microbiology and Ecology Are Vitally Important to Premedical Curricula. *Evol Med Public Health*. 2015 Jul 21. pii: eov014. [Epub ahead of print]

Cyanophage and cyanobacterial genomics and metagenomics

Kelly L, Ding H, Huang KH, Osburne MS, Chisholm SW. Genetic diversity in cultured and wild marine cyanomyoviruses reveals phosphorus stress as a strong selective agent. *ISME J. 2013 May 9. doi: 10.1038/ismej.2013.58. [Epub ahead of print]*

Kelly L†, Huang KH†, Ding H, Chisholm SW. ProPortal: A resource for integrated systems biology of *Prochlorococcus* and its phage. *Nucleic Acids Res. 2012 Jan;40(Database issue):D632-40.*

Malmstrom RR†, Rodrigue S†, Huang KH**, Kelly L**, Kern SE, Thompson A, Roggensack S, Berube PM, Henn MR, Chisholm SW. Ecology of uncultured *Prochlorococcus* clades

revealed through single-cell genomics and biogeographic analysis. *ISME J*. 2012 Aug 16. *[Epub ahead of print].*

Thompson LR†, Zeng Q†, **Kelly L**, Huang KH, Singer AU, Stubbe J, Chisholm SW. Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proc Natl Acad Sci U S A*. 2011 Sep 27;108(39):E757-64.

Li B†, Sher D†, **Kelly L**, Shi Y, Huang K, Knerr PJ, Joewono I, Rusch D, Chisholm SW, van der Donk WA. Catalytic promiscuity in the biosynthesis of cyclic peptide secondary metabolites in planktonic marine cyanobacteria. *Proc Natl Acad Sci* U S A. 2010 Jun 8;107(23):10430-5.

Computational analysis of human genetic variation

Kelly L†, Fukushima H†, Karchin R, Gow JM, Chinn LW, Pieper U, Segal MR, Kroetz DL, Sali A. Functional hotspots revealed by mutational, evolutionary, and structural characterization of ABC transporters. *Protein Sci. 2010 Nov;19(11):2110-21.*

Kelly L, Karchin R, Sali A. Protein interactions and disease phenotypes in the ABC transporter superfamily. *Pac Symp Biocomput*. 2007:51-63.

Karchin R, Diekhans M, **Kelly L**, Thomas DJ, Pieper U, Eswar N, Haussler D, Sali A. LS-SNP: large-scale annotation of coding non-synonymous SNPs based on multiple information sources. *Bioinformatics*. 2005 Jun 15;21(12):2814-20.

Karchin R, **Kelly L**, Sali A. Improving functional annotation of non-synonomous SNPs with information theory. *Pac Symp Biocomput. 2005:397-408.*

Olivier M, Aggarwal A, Allen J, Almendras AA, Bajorek ES, Beasley EM, Brady SD, Bushard JM, Bustos VI, Chu A, Chung TR, De Witte A, Denys ME, Dominguez R, Fang NY, Foster BD, Freudenberg RW, Hadley D, Hamilton LR, Jeffrey TJ, **Kelly L**, Lazzeroni L, Levy MR, Lewis SC, Liu X, Lopez FJ, Louie B, Marquis JP, Martinez RA, Matsuura MK, Misherghi NS, Norton JA, Olshen A, Perkins SM, Perou AJ, Piercy C, Piercy M, Qin F, Reif T, Sheppard K, Shokoohi V, Smick GA, Sun WL, Stewart EA, Fernando J, Tejeda, Tran NM, Trejo T, Vo NT, Yan SC, Zierten DL, Zhao S, Sachidanandam R, Trask BJ, Myers RM, Cox DR. A High-Resolution Radiation Hybrid Map of the Human Genome Draft Sequence. *Science. 2001 Feb 16;291(5507):1298-302.*

Computational sequence and structure analysis of membrane proteins

Kelly L, Pieper U, Eswar N, Hays FA, Li M, Roe-Zurz Z, Kroetz DL, Giacomini KM, Stroud RM, Sali A. A survey of integral α–helical membrane proteins. *J Struct Funct Genomics. 2009 Dec;10(4):269-80.*

Li M, Hays FA, Roe-Zurz Z, Vuong L, **Kelly L**, Ho CM, Robbins RM, Pieper U, O'Connell JD 3rd, Miercke LJ, Giacomini KM, Sali A, Stroud RM. Selecting optimum eukaryotic integral membrane proteins for structure determination by rapid expression and solubilization screening. *J Mol Biol.* 2009 Jan 23;385(3):820-30.

Hays FA, Roe-Zurz Z, Li M, **Kelly L**, Gruswitz F, Sali A, Stroud RM. Ratiocinative screen of eukaryotic integral membrane protein expression and solubilization for structure determination. *J Struct Funct Genomics.* 2009 Mar;10(1):9-16.

Jessica Mar, Ph.D.
Assistant Professor, Department of Systems and Computational Biology
Assistant Professor, Department of Epidemiology and Population Health
Price Center, Room 253A
jessica.mar@einstein.yu.edu, (718) 678-1251

Understanding the impact of variability in biological systems and disease

Characterizing cellular phenotypes based on distinct gene expression profiles has become a standard part of understanding biological function. Typically, we identify subsets of genes with differential expression levels that on average distinguish one phenotypic group from another. While studying genes on the basis of absolute expression is important to understanding regulation, we are only just beginning to recognize that variability in gene expression is an insightful regulatory parameter too. Work from our lab, as well as others, has shown that variability gives us an additional window into regulatory control of the transcriptome. We are interested in understanding variability in the context of transcriptional regulation of human stem cells, and its impact on disease.

Modeling single cell gene expression networks

Cell populations are inherently heterogeneous, and even for isogenic cells, we know that a gene's expression level exists along a distribution. With single cell expression profiling techniques now readily available, we are in a position to characterize expression heterogeneity across the genome and its role in regulation of the cell population. We are building single cell gene expression networks that incorporate measures of heterogeneity. Our objective is to identify cellular states that deviate from average single cell behavior.

Investigating tissue specificity of cancer-causing mutations

Over the past 10 years an increasing number of mutated genes have been associated with familial predisposition to cancer. Interestingly for more than half of these genes their involvement in cancer is restricted to only a few cancer types (e.g. BRCA1 mutations in breast and ovarian cancers). Even more interestingly some of these genes are expressed in all cell types, and perhaps we would expect to see them causing many more different types of cancer but they don't. Working with the RIKEN FANTOM 5 Consortium, our group is examining how these mutations are tolerated in most cell types but not in others by considering the network of genes expressed in different cell types and how that determines whether they are susceptible or resistant.

Selected Publications

Nayler S, Gatei M, Kozlov S, Gatti R, Mar JC**,** Wells CA, Lavin M, Wolvetang E. (2012). An induced pluripotent stem cell model for ataxia telangiectasia. *Stem Cells Translational Medicine.* Jun;1(7):523-535.

Gewurz BE, Towfic F, Mar JC, Shinners NP, Takasaki K, Zhao B, Cahir-McFarland ED, Quackenbush J, Kieff E. (2012). Genome-wide siRNA screen for mediators of NF-kappa B activation. *PNAS.* Feb 14;109(7):2467-72.

Mar JC**,** Matigian NA, Mackay-Sim A, Miotto A, Mellick G, Sue C, Silburn P, McGrath J, Quackenbush J, Wells CA. (2011). Variance of gene expression identifies altered network constraints in neurological disease. *PLoS Genetics.* Aug;7(8):e1002207.

Mar JC, Wells CA, Quackenbush J. (2011). Defining an Informativeness Metric for Clustering Gene Expression Data. *Bioinformatics.* Apr 15;27(8):1094-100.

Mar JC, Matigian NA, Quackenbush J, Wells CA. (2011). attract: A method for identifying core pathways that define cell phenotypes. *PLoS One.* 6(10):e25445.

Gewurz BE, Mar JC, Padi M, Zhao B, Takasaki K, Cahir-Mcfarland E, Quackenbush J, Kieff E. (2011). Canonical NF-kB activation is essential for Epstein-Barr Virus Latent Membrane Protein 1 TES2/CTAR2 gene regulation. *Journal of Virology.* Jul;85(13):6764-73.

Zhao B, Mar JC, Maruo S, Lee S, Gewurz B, Johannson E, Holton K, Rubio R, Takada K, Quackenbush J, Kieff E. (2010). Epstein Barr virus nuclear antigen 3C regulated genes in lymphoblastoid cell lines. *PNAS.* Jan 4;108(1):337-42.

Mar JC, Quackenbush J. (2009). Decomposition of Gene Expression State Space Trajectories. *PLoS Computational Biology.* Dec;5(12):e1000626.

FANTOM Consortium, Suzuki H, Forrest AR, van Nimwegen E, Daub CO, et al Mar JC. (2009). The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nature Genetics.* May 41(5):553-62.

Mar JC, Kimura Y, Irvine K, Schroder K, Hume D, Quackenbush J. (2009). Normalization Strategies for High-throughput quantitative RT-PCR Data. *BMC Bioinformatics.* 10:110.
Mar JC, Rubio R. and Quackenbush J. (2006). Inferring steady state single-cell gene expression distributions from analysis of mesoscopic samples. *Genome Biology.* 7(12):R119.

Ian Willis, Ph.D.
Professor, Department of Systems and Computational Biology
Professor, Department of Biochemistry
Forchheimer Building, Room 316
ian.willis@einstein.yu.edu; (718) 430-2839

Signaling pathways and transcriptional regulation in growth control and metabolism

Our laboratory is conducting basic research on the mechanisms of eukaryotic transcriptional regulation in response to nutrients and environmental and cellular stress. We are especially interested in defining the signaling pathways and the mechanisms that regulate transcription of ribosomal components and transfer RNAs since these processes are critically important for controlling cell growth. Deregulation of cell growth control is widely recognized as a key event in cell transformation and tumorigenesis and is relevant to a broad range of human diseases. In addition, as the synthesis of new protein synthetic machinery constitutes ~85% of nuclear gene transcription in growing cell populations, the tight coordinate control of this process, which involves all three nuclear RNA polymerases, is considered to be critical for metabolic economy. Our research programs span genetics, molecular biology, biochemistry and structural biology and utilize budding yeast, mammalian cells and mice as model experimental systems. Much of our current focus is on Maf1, a structurally and functionally novel protein that integrates the outputs of diverse signaling pathways and regulates transcription by all three nuclear RNA polymerases. The mechanisms of Maf1-dependent repression, the biological consequences of deleting Maf1 in the mouse and studies on novel downstream regulators and targets in the TOR signaling pathway are central areas of investigation in the lab. The conservation of Maf1 along with the signaling pathways that regulate Maf1 function enables the reciprocal translation of knowledge between yeast and mammalian systems and facilitates the discovery of new biology.

Genetic arrays, gene networks and functional genomics

Synthetic genetic array analysis and other systematic genome-wide genetic approaches such as synthetic dosage lethality and suppression are being conducted by robotic pinning of high density arrays of yeast strains. This technology enables the mapping of genetic interaction networks, defines the function of genes and establishes functional relationships between biochemical pathways. These genetic array-based approaches are being used to interrogate a range of biological processes including transcriptional regulation as described above. The robot also serves as a resource to other researchers at Einstein and elsewhere who are working in yeast or in mammalian systems on genes that have homologs in yeast. The integration of genetic interaction data with other large scale datasets such as DNA microarray, RNA and ChIP-sequencing and protein-protein interaction data is used to inform testable hypotheses of the systems level behavior of genes and their products.

Selected References:

Moir, R.D., Lee, J., Haeusler, R.A., Desai, N., Engelke, D.R. and Willis, I.M. (2006) Protein kinase A regulates RNA polymerase III transcription through the nuclear localization of Maf1. *Proc Natl Acad Sci U S A.* **103**:15044-15049.

Lee J, Moir RD, Willis IM. (2009) Regulation of RNA polymerase III transcription involves SCH9-dependent and SCH9-independent branches of the target of rapamycin (TOR) pathway. *J Biol Chem*. 284:12604-8.

Johnson, A.A., Zhang, C., Fromm, J., Willis, I.M. and Johnson, D.L. (2007) Mammalian Maf1 is a negative regulator of transcription by all three nuclear RNA polymerases. *Mol Cell.* 26:367-79.

Willis, I.M., Chua, G., Tong, A.H., Brost, R.L., Hughes, T.R., Boone, C., Moir, R.D. (2008) Genetic interactions of *MAF1* identify a role for Med20 in transcriptional repression of ribosomal protein genes. *PLoS Genet*. 4, e1000112.

Short, M.K., Hallett, J.P., Tar, K., Dange, T., Schmidt, M., Moir, R.D., Willis, I.M. and Jubinsky, P.T. (2012) The Yeast Magmas ortholog Pam16 has an essential function in fermentative growth that involves sphingolipid metabolism. *PLoS One* **7**, e39428.

Lee, J., Moir, R.D., McIntosh, K.B. and Willis, I.M. (2012) TOR Signalling Regulates Ribosome and tRNA Synthesis via LAMMER and GSK-3 Family Kinases. *Mol. Cell*. 45, 836-843.

Moir, R.D., Lee, J., and Willis, I.M. (2012) Recovery of RNA polymerase III Transcription from the Glycerol-Repressed State: Revisiting the Role of Protein Kinase CK2 in Maf1 Phosphoregulation *J. Biol. Chem.* 287.

Moir, R.D., Gross, D.A., Silver, D.L. and Willis, I.M. (2012) *SCS3* and *YFT2* Link Transcription of Phospholipid Biosynthetic Genes to ER Stress and the UPR. *PLoS Genet*. 8.

Moir, R.D. and Willis, I.M. (2013) Regulation of Pol III Transcription by Nutrient and Stress Signaling Pathways. *Biochim. Biophys. Acta.* 1829, 361-375.

Yinghao Wu, Ph.D.
Assistant Professor, Department of Systems and Computational Biology
Price Center, Room 553A
Yinghao.wu@einstein.yu.edu; (718) 678-1232

The aggregation of membrane receptors during cell adhesion initiates the elaborate networks of signaling pathways. The complexities of the networks originate from the spatial-temporal interactions of their numerous cellular components. By integrating computational analysis with experimental measurements, our lab is focusing on developing a multi-scale modeling framework to understand the molecular mechanisms of protein interactions underlying the physics of cell adhesion, as well as their biological significance.

Method Development:
The development of a multi-scale modeling framework could lead to an integrative understanding of how extracellular signals regulate cell adhesion and downstream signaling pathways in various biological systems. By designing different simulation scenarios on **molecular level**, **sub-cellular level**, **systems level** and **multi-cellular level**, the framework could serve as a guide to reveal the molecular mechanism of specific disease-related problems.

Biological Applications:
Cadherin/Wnt Signaling The epithelial-mesenchymal transition (EMT), characterized by repression of cell adhesion, is the hallmark of both normal embryonic development and cancer metastasis. Wnt is one of the most important signaling pathways triggering EMT. The key players in Wnt signaling is β-catenin, which is involved in both intercellular adhesion and gene regulation. The binding of β-catenin to the cytoplasmic domain of E-cadherin results in the stabilization of adherens junctions. On the other hand, its association with the T-cell factor/lymphoid enhancer factor (TCF/LEF) DNA binding proteins changes the transcription of target genes, initializing the canonical Wnt pathway. The fate of β-catenin in adhesion and signaling is further regulated by Wnt activation and its downstream phosphorylations. As the functions of β-catenin have been studied separately in cadherin-based junction formation and in Wnt signaling pathway, relatively little has been done to connect these two systems. Our goal for this project is to quantitatively interrogate the interplay between cadherin-mediated junction formation and canonical Wnt signaling pathway by asking the direct question: How can competition of β-catenin between these two systems serve to integrate cell adhesion with gene expression?

Integrin Signaling During cell migration, large macromolecular assemblies form at focal adhesions to transmit mechanical force and regulatory signals across cell membranes. Integrins serve as the mechanical linkages to the extracellular matrix (ECM). Their clustering based on ligand binding provides a biochemical signaling hub to direct numerous signaling and adapter proteins such as talin and focal adhesion kinase (FAK). We study the molecular mechanism of integrin clustering and its impact on mechano-chemical coupling by multi-scale modeling. Our studies can be directly compared with

cellular imaging experiments, and will give insights into the dynamic coupling of integrin clustering with downstream signaling events, for instance, the recruitment of FAK.

T-cell Signaling T cells play a pivotal role in cell-mediated immunity. The spatiotemporal patterning between T-cells and antigen-presenting cells (APCs) leads to the maturation of the immunological synapse (IS). This process is highly correlated to T-cell activation. Although size of membrane receptors was suggested to drive synaptic patterning, detailed structural information has not been used to study such sub-cellular process. Combining knowledge from molecular and cellular levels, we are using multi-scale studies to understand why specific patterns can be formed on T cell surfaces and how they are related to the intracellular signaling.

Publications:

J. Chen, Z. Xie and Y. Wu , "Computational Modeling of the Interplay between Cadherin-Mediated Cell Adhesion and Wnt Signaling Pathway". *PLoS ONE* (2014) 9(6): e100702. doi:10.1371/journal.pone.0100702.

J. Chen, Z. Xie and Y. Wu , "A multiscale model for simulating binding kinetics of proteins with flexible linkers". *Proteins: Structure, Function, and Bioinformatics* (2014) doi: 10.1002/prot.24614.

Z. Xie, J. Chen and Y. Wu , "A coarse-grained model for the simulations of biomolecular interactions in cellular environments". *Journal of Chemical Physics* (2014) 140: 054112.

Y. Wu, B. Honig and A. Ben-Shaul , "Theory and simulations of adhesion receptor dimerization on membrane surfaces". *Biophysical Journal* (2013) 104: 1221-1229.

Y. Wu, J. Vendome, L. Shapiro, A. Ben-Shaul and B. Honig, "Transforming Binding Affinities from Three Dimension to Two With Application to Cadherin Clustering". *Nature* (2011) 475: 510-513.

O. Harrison, X. Jin, S. Hong, F. Bahna1, G. Ahlsen, J. Brasch, Y. Wu, J. Vendome, K. Felsovalyi, etc. "The extracellular architecture of adherens junctions revealed by crystal structures of type I cadherins". *Structure* (2011) 19:244-256.

Y. Wu, X. Jin, O. Harrison, L. Shapiro, B. Honig and A. Ben-Shaul, "Cooperativity between trans and cis interactions in cadherin-mediated junction formation". *P.N.A.S.* (2010) 107 (41) 17592-17597.

Y. Wu, AD. Dousis, M. Chen, J. Li, and J. Ma, "OPUS-Dom: Applying the Folding-Based Method VECFOLD to Determine Protein Domain Boundaries". *J. Mol. Biol.* (2009) 385, 1314-1329.

M. Chen, AD. Dousis, Y. Wu, and J. Ma, "Predicting protein folding cores by empirical potential functions". *Archives of Biochemistry and Biophysics* (2009) 483, 16-22.

Y. Wu, M. Lu, M. Chen, J. Li, and J. Ma, "OPUS-Ca: A knowledge-based potential function requiring only C-alpha Positions". *Protein Science* (2007) 16, 1449-1463.

M. Chen, C.J. Wilson, Y. Wu, P. Wittung-Stafshede, and J. Ma, "Correlation between Protein Stability Cores and Protein Folding Kinetics: A Case Study on Pseudomonas aeruginosa Apo-Azurin". *Structure* (2006) 14, 1-10.

Y. Wu, X. Tian, M. Lu, M. Chen, Q. Wang and J. Ma, "Folding of Small Helical Proteins Assisted by Small-Angle X-Ray Scattering Profiles". *Structure* (2005) 13, 1587-1597.

Y. Yuan, Y. Wu and J. Zi, "Heat Capacities of Globular Proteins". J. of Phys.: *Condensed*

Matter, (2005) 17, 469-473.
Y. Wu, M. Chen, M. Lu, Q. Wang and J. Ma, "Determining Protein Topology from Skeletons of Secondary Structures". *J. Mol. Biol.* (2005) 350, 571-586.
Y. Wu and J. Ma, "Refinement of F-actin Model against Fiber Diffraction Data by Long-Range Normal Modes". *Biophysical J.* (2004) 86, 116-124.
Y. Wu and J. Ma, "Normal-mode-based Refinement of an F-actin Model Against Fiber Diffraction Data". *Fiber Diffraction Review,* (2004) 12, 25-28.
D. Ming, Y. Kong, Y. Wu and J. Ma, "Simulation of F-actin filaments of Several Microns". *Biophysical J.* (2003) 85, 27-35.
Y. Kong, D. Ming, Y. Wu, J. Stoops, H. Zhou and J. Ma, "Conformational Flexibility of Pyruvate Dehydrogenase Complexes: A Computational Analysis by QEDM". *J. Mol. Biol.* (2003) 330, 129-135.
D. Ming, Y. Kong, Y. Wu and J. Ma, "Substructure Synthesis Method for Simulating Large Molecular Complexes". *P.N.A.S.* (2003) 100, 104-109.
Y. Wu and J. Zi, "Universal Behavior of Localization of Residue Fluctuations in Globular Proteins". *Physical Reviews E.* (2003) 67, 041909-041912.