Professional Statement of Manolis Kellis

Since the completion the human genome sequence, increasingly vast genomic datasets have created enormous opportunities for answering fundamental questions about life and evolution, with extraordinary implications for medicine and human health. My work aims to seize these unique opportunities and to develop the computational tools necessary for understanding genomic information systematically. On one hand, I develop computational approaches that exploit the genome-wide nature of biological datasets to derive common rules and genomic signatures which I use to annotate functional DNA elements, reconstruct their regulatory circuits, and understand their evolutionary principles. On the other hand, I work closely with experimental scientists with diverse expertise and working across a range of model organisms to validate computational predictions, obtain new datasets, and think deeply about foundational biological questions. The resulting interdisciplinary research program allowed me to gain fundamental new knowledge into the function, regulation and evolution of genomes by taking advantage of emerging technologies and data types, inventing new computational models that exploit the unique nature of biological problems, and ensuring that computational results benefit from expert experimental validation to confirm new insights.

My research spans four major areas, each central to our understanding of biological systems:

(1) Genome interpretation: To recognize the molecular basis of human biology and disease, we need a comprehensive annotation of all protein-coding genes, non-coding RNAs and regulatory control elements encoded in the human genome. While most of these genomic elements lack distinguishing sequence features for their *de novo* recognition in a single genome, comparison of multiple related species provides a new evolutionary lens for genome interpretation. Exploiting genome-wide comparative genomics datasets, my group recognized specific patterns of evolutionary change, or **evolutionary signatures** associated with each class of functional elements, and dictated by the specific constraints unique to each type of function. We used these signatures in the human, fly, and yeast genomes to recognize protein-coding genes and exons, RNA genes and structures, microRNAs and their targets, and several types of regulatory elements, revealing thousands of new coding exons, millions of regulatory motifs, and many novel biological insights.

(2) Gene regulation: Beyond a comprehensive encyclopedia of genomic elements we need to understand how these elements are put together in cellular circuits and to infer predictive models of cellular response, which can serve as foundations for modern medicine, rational drug design, and personalized treatments. Cellular regulatory networks that control gene expression are dictated by the binding of DNA and RNA regulators to sequence motifs typically only 6-8 nucleotides long. These **regulatory motifs** have previously remained elusive to systematic discovery due to their short lengths and varying distances at which they can act. However, by evaluating their genome-wide conservation patterns, their positional and strand biases, and their association with functional genomic datasets, my work pioneered new methods for their systematic discovery, the characterization of their functional roles, and the inference of the **regulatory networks** they define. We also developed experimental methods that can test tens of thousands of engineered regulatory regions in a single experiment, and used these to **dissect thousands of human distal regulatory regions** to uncover the role of regulatory motifs and individual nucleotide variants in gene regulation.

(3) Epigenomics: Beyond the primary sequence of the genome, a wide variety of post-translational modifications in the tails of histone proteins (around which DNA is wrapped into chromatin) play key roles in genome function, cellular differentiation, and human disease. With several dozen such chromatin marks already described, a wide array of chromatin regulators that can read and write them, and an astronomical number of mark combinations, discovering biologically meaningful sets of jointly acting marks and characterizing their specific function are daunting new challenges in genomics. My group pioneered new methods for addressing these challenges, using genome-wide maps of many histone marks to discover recurrent and spatially-coherent combinations of marks, or **chromatin states**. We also developed data integration methods to systematically characterize

chromatin state function, revealing diverse classes of enhancers, promoters, and insulators, which we used to discover new functional elements, to study **chromatin dynamics** across cell types and development, and to reveal motifs and regulators governing epigenetic changes in development, differentiation and disease.

(4) Genome evolution: Lastly, a unique and defining feature of biological systems is their ability to adapt to changing environments and evolve new functions and pathways, while constantly subjected to the forces of mutation and natural selection. Understanding the mechanisms associated with evolutionary innovation can teach us how to better interpret evolutionary signatures and comparative genomics evidence, yield new insights about the function of diverse genomic elements from their shared ancestry, and in the longer term reveal insights on how to engineer more evolvable and noise-tolerant systems. With the availability of many complete genomes, my group took a new approach to studying **gene evolution at the whole-genome level**, which resulted in numerous new insights that wouldn't be visible at the single-gene level of traditional phylogenetics. This revealed a unique genomic signature for **whole-genome duplication** which I demonstrated in yeast and fish, and has since been used to resolve the ancestry of many species, including our early vertebrate ancestors. We also developed **phylogenomic** methods for resolving gene phylogenies across complete genomes jointly, enabling much more accurate inference methods, by recognizing two distinct evolutionary forces, one acting at the gene level, and one at the genome level. Lastly, using population genomics, we were able to demonstrate that in addition to the 5% of the human genome that shows evidence of evolutionary constraint in mammals, an additional **4% is constrained within the human lineage** outside conserved regions, suggesting abundant evolutionary innovation.

Despite the challenges stemming from a broad research spectrum, the synergy between these four areas and the ability to work closely with experimental collaborators on both human cells and multiple model organisms have enabled discoveries that would otherwise not be possible, resulting in many new insights into genome biology and gene regulation. To name a few, our results revealed: (a) the first evidence of widespread translational readthrough in neuronal and regulatory proteins in the fruitfly and the human genome; (b) the first report of anti-sense microRNAs, and demonstration of their ability to function as regulatory switches and induce body-plan transformations; (c) the first demonstration that nearly one third of all human protein-coding exons encode overlapping functional elements including developmental enhancers that drive tissue-specific expression; (d) the first report of a large new class of long intergenic non-coding RNAs that play diverse roles in gene regulation and disease using chromatin and evolutionary signatures; (e) the first report that disease variants distant from genes are strongly enriched for specific chromatin states indicative of enhancer regulatory function; (f) the first global functional regulatory network in an animal genome with the ability to predict gene expression levels for several thousand genes. In each case, these results begun with our computational predictions, and were validated in experimental collaborations.

Specific research contributions:

I developed three types of genomic signatures that enable the systematic annotation of diverse classes of functional elements and the study of their regulatory principles.

(1) Evolutionary signatures. The first type of signature uses comparative genomics datasets to systematically annotate conserved functional regions across a complete genome. Early comparative genomics studies, including my own, demonstrated that functional elements are preferentially conserved across related species, enabling their detection in regions of increased nucleotide conservation. Evolutionary signatures go a step beyond this, and look specifically within alignments of multiple related species for **distinct patterns of nucleotide change characteristic of distinct types of functional elements**. It turns out that each class of functional elements evolves in specific and distinct ways, dictated by the specific level at which natural selection acts to maintain their function. For protein-coding genes, selective forces act at the amino-acid level, and changes in the primary

DNA sequence are tolerated as long as they preserve the amino-acid translation, leading to specific evolutionary signatures of biased codon substitution frequencies and increased reading-frame conservation. For RNA genes, selective forces act at the structural level and mutations are tolerated as long as they preserve the complementarity of paired nucleotides and thus the fold of the RNA structure. For microRNAs, selective forces are distinct from those on other RNA genes, and include precise sequence, pairing, and structural constraints, leading again to specific signatures. Regulatory motif instances are characterized by an increased phylogenetic conservation compared to control motifs of similar composition, while local movements and even strand changes are frequently tolerated. Lastly, regulatory motifs are defined by increased genome-wide conservation observed simultaneously across all their instances.

My group encoded these evolutionary signatures computationally and used them to **systematically annotate the human, fly, and yeast genomes** using phylogenies of 29 mammals, 12 flies, and 17 fungi. Our methods resulted in the discovery of (a) thousands of previously-undetected protein-coding exons, (b) new types of RNA secondary structures involved in localization, degradation, editing and regulation, (c) new microRNA genes and families leading to increased combinatorial regulation, (d) complete dictionaries of regulatory motifs, and (e) regulatory motif instances predicting targets for each regulator, leading to initial regulatory networks for both fly and human.

Our approach also revealed numerous **new insights about the regulation of animal genomes**, which would have been inaccessible using traditional methods. First, we found abundant translational readthrough in animal genomes, which incorporates new protein domains with important biological, evolutionary, and regulatory implications. While only five cases were known across all eukaryotes, we found 300 candidates in fly, of which numerous were validated experimentally. Second, we found a new class of anti-sense microRNAs encoded on the opposite strand of existing RNA genes and likely functioning as regulatory switches between different developmental programs. To validate our predictions, we over-expressed in fruitfly embryos one anti-sense microRNA encoded within the Hox developmental cluster, leading to a homeotic transformation of a haltere into a wing, as expected given the predicted Hox targets of the anti-sense miRNA. Third, we found that nearly one third of all human protein-coding genes contain regions of overlapping selective pressures that encode additional functional elements embedded within the protein-coding signal, including RNA structures, translation initiation and splicing signals, as well as experimentally confirmed developmental enhancers that drive tissue-specific expression of the host genes.

(2) Chromatin signatures. The second type of genomic signature is encoded in the epigenetic modifications of the nucleosomes around which DNA is wrapped. These post-translational modifications of specific amino-acid residues in the tails of histone proteins that make up nucleosomes have become the subject of intense scrutiny in recent years. Technological advances have enabled genome-wide maps of all genomic regions associated with each of many chromatin marks in numerous cell types. By mining these datasets, my group showed that specific **biologically-relevant and spatially-coherent chromatin states** can be discovered without any prior knowledge. We built a multivariate hidden Markov model to discover these states based solely on mark co-occurrence patterns across complete genomes, and found they are highly predictive of diverse classes of functions, including promoters, enhancers, transcribed regions, repressed and repetitive regions, and diverse classes of active intergenic regions. We used these states to provide **a new layer of annotation of the human genome**, which dramatically expanded on our understanding of the non-coding portion that had previously remained largely elusive.

We also used these chromatin signatures to gain many new insights into the biology of animal genomes. First, by combining the chromatin signature of promoter regions adjacent to transcribed regions with our evolutionary signatures of protein-coding selection, we helped recognize a new class of genes in the human genome which we call **long intergenic non-coding RNAs**, or lincRNAs,

that play diverse roles in gene regulation and cancer. Second, we used our chromatin state annotations to **revisit disease-associated variants** that were previously distant from genes but were strongly enriched for specific chromatin states indicative of enhancer regulatory function. Third, we found that distinct promoter chromatin states correspond to distinct functions of associated genes, and different enhancer states are associated with varying levels of expression for downstream genes more than 10kb away, suggesting that much more information about the function and utilization of genes is encoded at the chromatin level than previously suspected. Lastly, we recognized specific chromatin signatures associated with transcription end sites, suggesting a previously unexpected role for chromatin in transcription termination.

(3) Activity signatures. A third type of genomic signatures takes advantage of the increasing availability of functional genomic datasets across many conditions such as cell types, developmental stages or perturbations. Under the assumption that functionally interconnected genomic elements would show correlated changes in their activity across many conditions, we defined vectors of activity for diverse types of functional elements and used correlations between these vectors to reveal functional relationships between them. In the context of the human ENCODE project, we used high correlation between the enhancer activity of candidate regulatory regions and the expression level of neighboring genes in multiple cell types to **recognize functional targets of distal enhancers** acting at more than 100kb away. We also defined **activator and repressor signatures** for regulators based on three-way correlations between the expression or repression of a regulator, the enrichment or depletion of its motifs in cell-type specific chromatin states, and the active or repressed nature of the chromatin states. Lastly, we defined activity vectors measuring sharp depletions in chromatin signal associated with physical regulator binding and nucleosome displacement, which enabled us to recognize cell-type specific regulator binding events correlated with the predicted activity of activators and repressors in their target enhancers, thus validating our regulatory predictions.

We also applied these activity signatures in the developmental time-course of the fruitfly *Drosophila melanogaster* towards global predictive models of gene regulation in close collaboration with the modENCODE consortium. First, we recognized **functional targets of transcriptional regulators** based on their correlated gene expression and chromatin modification profiles across 30 developmental time-points. We used the resulting functional networks that we derived to **annotate predicted functions for genes previously lacking a functional annotation**, which we validated based on tissue-specific expression in precursor tissues for the predicted activity. We also used these functional networks to reveal stage-specific and cell-type specific regulators of gene expression changes and chromatin changes, based again on their correlated activity. Lastly, we used the functional network to study our ability to **predict the expression levels of target genes** based on the expression levels of their regulators. We found that indeed our functional networks have significantly higher predictive power than physical networks either predicted through conserved regulatory motif instances, or observed through binding of transcription factors. We were able to predict the expression levels of nearly 2000 genes across the developmental time-course, which we validated by measuring expression for both transcription factors and their targets in several cell lines and showing that our predictions indeed hold. In contrast, genes that were unpredictable by our methods showed significantly reduced reproducibility in their expression patterns, suggesting less precise regulatory control.

Going forward. These genomic signatures have proved extremely powerful in helping to interpret the human genome, and deriving predictive models of gene regulation. In contrast to previous modeling activities that have focused on small pathways, our work considers the combined effects of all motifs and regulators on all target genes, and observing responses at the systems level. To match the scale of our predictions, we are currently developing experimental assays that can test the action

of tens of thousands of enhancers in multiple cell lines simultaneously encoded using custom high-density microarrays with barcodes, and probed using next-generation sequencing. This can revolutionize our ability to evaluate genome-wide predictions, and serve as the basis of next-generation regulatory models for predicting the function of large numbers of motif combinations, and their specific arrangements to understand potential motif 'grammars'.

In an important new direction, we have begun exploring the interplay between **human polymorphisms, comparative genomics, and our regulatory predictions**. First, we extended our phylogenomics methods for gene duplication and loss to incorporate population models of allele coalescence within species, bringing closer together two previously-separate fields, and enabling studies of many loci and many individuals from densely-sampled species. Second, we studied correlations between selective pressures inferred from 29 mammals and variants inferred from 1000 human genomes, revealing that the major and minor alleles of human single-nucleotide polymorphisms (SNPs) match pairs of nucleotides frequently exchanged across the mammalian phylogeny, thus revealing that selective pressures still guide recent human evolution. Lastly, we used our genomic signatures to provide **mechanistic hypotheses for variants associated with human disease**, revealing both relevant cell types and motif instances affected by the SNPs and associated with predicted regulators, thus leading to specific mechanistic hypotheses for disease phenotypes.

In the coming decade, rapidly advancing sequencing technologies will generate thousands of vertebrate and mammalian genomes, tens of thousands of human genomes, and countless functional and disease datasets. Such datasets can revolutionize medicine, diagnosis and intervention, and our methods can help provide the computational tools necessary for enabling such transformations.

Educational contributions. As the next generation of scientists that will take on these questions will need to combine computational sophistication and strong biological intuition, I have strived to enhance our computational biology curriculum. I **developed three new courses** at the interface of computer science and biology: (1) an extension to the introductory algorithms course, (2) an independent introductory new course, and (3) an advanced course on personal genomics, all open at the undergraduate level, and the last two also for graduate students. Aiming these courses to computer science students allowed me to **tackle foundational computational concepts** that underlie the most pervasive techniques in genomics today, both to help students understand existing algorithms, and to enable them to design the next generation of algorithms. I have also **integrated the use of large-scale biological datasets** in the courses I am offering, to enable both graduate and undergraduate students to get the excitement and intuition that come with working directly with real data. I also developed an **integrated final project experience** including peer-review, midcourse reports, and extensive mentoring by postdocs and senior students. This often led to advanced final projects, resulting in several publications that grew out of final projects, as well as masters or PhD theses from work that started in the classroom. All materials are public at compbio.mit.edu/teaching.html. I also began work on **a book for my computational course**, titled "Computational Biology: Genomes, Networks, Evolution" aimed at advanced undergraduate quantitative students and seeking to cover theoretically-advanced concepts in computational biology in a simple and practical exposition, combined with biological insights that have been gained by these foundational developments. The book is at 300 pages and I'm in final stages of negotiation with Cambridge University Press.

Service contributions. I have served on the program and organizing committees of several conferences, including RECOMB, ISMB, CSHL Biology of Genomes, CSHL Genome Informatics, and RECOMB Regulatory Genomics and Systems Biology. I have also served on numerous panels advising the NIH, NHGRI, NIAID, and the NSF on future directions for computational biology.

For the past seven years, I have been the head organizer of a joint conference on Regulatory and Systems Genomics, bringing together 500+ attendees, with three partner journals, 50+ submitted papers, 300+ abstracts, and 90+ talks, all freely available at compbio.mit.edu/recombsat.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
School of Engineering Faculty Personnel Record

Date: March 11, 2013

Name: Manolis Kellis
Department: Electrical Engineering and Computer Science

1. Date of Birth: March, 1977

2. Citizenship: US

3. Education:

School	Degree	Date
MIT	BS	1999
Xerox Palo Alto Research Center	MIT VI-A	1997-1999
MIT	MEng	1999
MIT	PhD	2003
Cold Spring Harbor Laboratory, Genetics	N/A	2003
Broad Institute of MIT and Harvard	Postdoc	2004

4. Title of Thesis for Most Advanced Degree:

   Computational Comparative Genomics: Genes, Regulation, Evolution.

5. Principal Fields of Interest:

   Computational Biology: understanding complete genomes, their biological networks, and their evolutionary mechanisms using novel algorithms and machine learning techniques.

6. Name and Rank of Other Department Faculty in the Same Field:

   Robert Berwick, Professor
   David K. Gifford, Professor
   Tommi Jaakkola, Professor
   Bruce Tidor, Professor
   Collin Stultz, Associate Professor

7. Name and Rank of Faculty in Other Departments in the Same Field:

   David Bartel, Professor (Biology)
   Bonnie Berger, Professor (Math)
   George Church, Professor (Health Sciences and Technology)
   Edward DeLong, Professor (Biological Engineering)
   Eric Lander, Professor (Biology)
   Richard Young, Professor (Biology)
   Chris Burge, Professor (Biology)
   Ernest Fraenkel, Associate Professor (Biological Engineering)
   Eric Alm, Assistant Professor (Bioloical Engineering)

Aviv Regev, Associate Professor (Biology)

8. Non-MIT Experience (including military service):

Employer	Position	Beginning	Ending
Xerox Palo Alto Research Center	Summer Intern	May 1997	August 1997
Xerox Palo Alto Research Center	Summer Intern	May 1998	August 1998
Xerox Palo Alto Research Center	Summer Intern	May 1999	July 1999
Galeries Lafayette IT Strategy	Independent Consultant	July 1999	August 1999

9. History of MIT Appointments:

Rank	Beginning	Ending
Postdoctoral Fellow, Broad Institute	June 2003	June 2004
Associate Member, Broad Institute of MIT and Harvard	Sept. 2004	present
Assistant Professor, EECS	Sept. 2004	June 2008
Associate Professor without Tenure, EECS	July 2008	June 2011
Associate Professor with Tenure, EECS	July 2011	present

10. Consulting Record:

Firm	Beginning	Ending
Galeries Lafayette Information Technology Strategy	July 1999	Aug. 1999
Interim Advisory Board Member, National Human Genome Research Institute (NHGRI), NIH	Jan. 2002	Jan. 2002
Advisory Member, Planning meeting for Project ENCODE for the Human Genome, NIH	June 2002	June 2002
Advisory Member, Saccharomyces Genome Database (SGD) planning meeting in the yeast community	Aug. 2002	Aug. 2002
Novartis Pharmaceuticals	Nov. 2003	Nov. 2003
Advisory Member, NHGRI council meeting for planning mammalian comparative sequencing, NIH	2004	2004
Advisory Member, NHGRI/NIAID selection and planning of sequencing for Pathogen and Disease Vectors	2006	2006
Advisory Member, NHGRI planning meeting for activities in Gene Regulation	2009	2009
Advisory Member, NHGRI planning meeting for activities in computational data analysis and informatics	2010	2010
Advisory Member, NHGRI planning meeting for vision statement for future activities of the institute	2010	2010
Novartis Institute for Biomedical Research (NIBR)	2011	2012
DNAnexus Computational Advisory Board member	2012	2012

11. Department and Institute Committees, Other Assigned Duties:

Activity	Beginning	Ending
Graduate Admissions (EECS)	Sept. 2004	present
Research Qualifying Examinations (12 students)	Sept. 2004	present
Graduate admissions committee	Nov. 2004	present
Graduate Counselor (EECS)	Sept. 2005	present

Undergraduate Counselor (EECS)	Sept. 2005	present
Broad Scientific Council Meeting (Broad Institute)	Jan. 2007	Jan. 2007

12. Professional service:

Activity	Beginning	Ending
Interim Advisory Board Member, National Human Genome Research Institute (NHGRI), NIH	2002	
Advisory Member, Planning meeting for Project ENCODE for the Human Genome, NIH	2002	
Advisory Member, Saccharomyces Genome Database (SGD) planning meeting for yeast community	2002	
Program Committee Member, RECOMB Satellite Meeting on Gene Regulation (RECOMB 2004)	Mar. 2004	Dec. 2004
Advisory Member, NHGRI council meeting for planning mammalian comparative sequencing, NIH	2004	
Referee for Nature, Nature Genetics, PLoS Biology, Genome Research, Genome Biology	May 2004	present
Program Committee Member, Int. Conf. on Research in Computational Molecular Biology (RECOMB 2006)	Sept. 2005	May 2006
Program Committee Member, Regulatory Genomics, Intelligent Systems for Molecular Biology (ISMB 2007)	Nov. 2005	Aug. 2006
Program Committee Co-Chair, Computational Genomics, Cold Spring Harbor Laboratory (CSHL) Meeting on the Biology of Genomes	Nov. 2005	May 2006
Program Committee Member, Int. Conf. on Research in Computational Molecular Biology (RECOMB 2007)	Sept. 2006	present
Program Committee Co-Chair, Comparative Genomics, Intelligent Systems for Molecular Biology (ISMB 2006)	Nov. 2006	present
Advisory Member, NHGRI/NIAID selection and planning of sequencing, Pathogens/Disease Vectors	2006	
Member, Editorial Board, Genome Research	Aug. 2007	present
Co-Chair for Comparative Genomics, CSHL Meeting on Genome Informatics (CSHL Genome Informatics 2007)	Oct. 2007	Oct. 2007
Conference Chair, RECOMB Satellite Meeting on Regulatory Genomics (MIT / Broad Institute Oct 11-13) (250 people attended, 83 abstracts, 9 keynotes, 27 talks) Conference site: http://compbio.mit.edu/recombsat/2007/	May 2007	Oct. 2007
Conference Chair, Joint RECOMB Satellite Meeting on Regulatory Genomics and Systems Biology (MIT Oct 29-Nov 2) (508 people attended, 253 abstracts, 17 keynotes, 92 talks) Conference website: http://compbio.mit.edu/recombsat/2008/	May 2008	Nov. 2008
Conference Chair, Joint RECOMB Satellite Meeting on Regulatory Genomics and Systems Biology (MIT Dec 2-6) (441 people attended, 292 abstracts, 15 keynotes, 89 talks) Conference website: http://compbio.mit.edu/recombsat/	May 2009	Dec. 2009
Advisory Member, NHGRI planning meeting for activities in Gene Regulation	2009	

GCAT study section, Ad-hoc reviewer	2009	2011
Chair, Intelligent Systems for Molecular Biology (ISMB) Special Session on Computational Epigenomics	Jun. 2010	Jun. 2010
Advisory Member, NHGRI planning, activities in computational data analysis and informatics	2010	
Advisory Member, NHGRI planning meeting for vision statement for future activities of the institute	2010	
Register of External Experts of Hellenic Quality Assurance Agency for Higher Education (HQAA)	2010	
VectorBase External Scientific Advisory Board (SAB)	2010	
Register of External Experts of Hellenic Quality Assurance Agency for Higher Education (HQAA)	Feb. 2010	Present
VectorBase External Scientific Advisory Board (SAB)	Feb. 2010	Present
FlyBase External Scientific Advisory Board (SAB)	Mar. 2010	Present
Conference Co-Chair, Joint RECOMB Satellite Meeting on Reverse Engineering of Biological Systems, Systems Biology, and Regulatory Genomics (Columbia Nov 16-20) Conference website: http://recombsat.c2b2.columbia.edu/	May 2010	Dec. 2010
Aristeia Excellence Grants application reviewer for Greece	2011	Present
Conference Co-Chair, RECOMB Regulatory Genomics, Systems Biology, and DREAM6 Barcelona, Spain, 340 attendees, 200 abstracts, 60 talks, 12 keynotes		2011
FlyBase External Scientific Advisory Board (SAB)	2011	
BDMA study section (Bioinformatics Databases Mathematics Algorithms), Ad-hoc Reviewer	2012	
Conference Co-Chair, RECOMB Regulatory and Systems Genomics, and DREAM7 To be held in San Francisco, CA	2012	
ENCODE U01 computational applications review panel, member	2012	
GCAT study section (Genomics, Computational Biology and Technology), Rotating Member	2012	2018
Conference Chair, RECOMB Regulatory and Systems Genomics, and DREAM8 To be held in U. Toronto, Canada	2013	

For the past seven years, I have been head organizer of the RECOMB Satellite Conference on Regulatory Genomics. Since 2008, the meeting has been held jointly with RECOMB Systems Biology, and the DREAM Dialogue on Reverse-Engineering And Modeling of Biological Systems. The meeting brought together computational and experimental scientists in the areas of regulatory genomics and epigenomics, systems biology and network science, and applications to understanding biological systems and human disease.
The conference has been tremendously successful, growing from only 50-150 attendees in previous years to a major international meeting, attracting more than 500 attendees. We receive both paper and abstract submissions, and partner with several partner journals to provide a journal venue for accepted papers in Genome Research, PLoS Computational Biology, and the Journal of Computational Biology.
Abstracts, papers, and videos from previous years are linked at
http://compbio.mit.edu/meetings.html

13. Awards Received:

Awards Received Personally:	Date

First prize in South France in country-wide math competition	1993
French Baccalauréat with the Congratulations of the Jury, the highest distinction in France	1995
First prize, Intercollegiate Tau Beta Pi Engineering Design Competition	1997
MassGrant Award for academic excellence	1998
Chorafas Foundation Award	1999
Paris Kanellakis Fellowship	2000
Sprowls award, best Ph.D. thesis in Computer Science, MIT	2003
Museum of Science, One of three young scientists representing the Next Generation in Biotechnology	2004
Distinguished Alumnus (1964) Career Development Chair	2005
Genome Technology, Principal Investigators of the Future	2006
Technology Review magazine, Top Young Innovators, TR35	2006
Service appreciation award, CSAIL Student Workshop, Dinner Speaker	2007
Karl Van Tassel (1925) Career Development Chair	2007
NSF CAREER Award	2007
Alfred P. Sloan Foundation Award	2008
Ruth and Joel Spira Teaching Award, MIT School of Engineering, EECS	2009
Presidential Early Career Award in Science and Engineering (PECASE)	2010
Niki Award for Science and Technology, Kokkalis Foundation and AIT	2011
Martin Prize for Basic Research, MGH, best paper of 2011 (Ernst *et al* 2011 Nature paper)	2012

Awards Received by Students and Postdocs:

William A. Martin Thesis Award for Best M.S. thesis is Computer Science: Matthew Rasmussen, Ph.D. student	2006
Charles and Jennifer Johnson Thesis Award for outstanding M.Eng. thesis in Computer Science: Michael F. Lin, M. Eng. Student	2006
Charles and Jennifer Johnson Thesis Award for outstanding M.Eng. thesis in Computer Science: Joshua Grochow, M. Eng. Student	2007
Cold Spring Harbor The Biology of Genomes Best Poster Presentation Award Loyal Goff, Postdoctoral Fellow	2008
Computer Research Association (CRA) Finalist for Outstanding Undergraduate Researcher Award: Tom Morgan, UROP student	2009
Anna Pogosyants award for outstanding Undergraduate Research: Tom Morgan, UROP student	2010
Siebel Scholarship for academic excellence and demonstrated leadership Angela Yen, M.Eng. student	2010
National Science Foundation Computing Innovation (CI) Fellows Program Sushmita Roy, Postdoctoral Fellow	2010

14. Current Organization Membership:

Organization	Offices Held
International Society for Computational Biology, ISCB	*None*
Sigma Xi	*None*
Eta Kappa Nu (HKN)	*None*
Tau Beta Pi (TBP)	Chair, 1999

15. Patents and Patent Applications Pending: None

16. Professional Registration:
Registered Professional Engineer, Greece.

17. Major New Products, Processes, Designs, or Systems:
None

1. Teaching Materials Developed
   a. Redesigned the syllabus for 18.417, titled "Computational Molecular Biology," in Fall 2000, expanding genome analysis and comparative genomics, and developing several new lectures and updated material, as a guest co-lecturer for the course with Bonnie Berger.
   b. Developed 6.096, titled "Algorithms for Computational Biology," a new 6-unit extension to the 12-unit Introduction to Algorithms course (6.046), complementing the traditional algorithms curriculum with a practical component focused on Computational Biology, including one additional lecture each week, expanding on the material for the week and its applications to computational biology.
   c. Developed 6.047/6.878/HST.507, titled "Computational Biology: Genomes, Networks, Evolution," a new course for Computational Biology especially tailored for computational students, introducing fundamental algorithms and machine learning techniques, and applying them to central problems in computational biology. The course combine a rigorous in-depth foundational component that covers the algorithms and machine learning technique central to the field, and also a frontiers component that exposes students to recent developments and current directions of the field, which covers both recent landmark papers, and also includes guest lectures from faculty and active researchers in the field.
   d. All materials for all courses, including syllabi, lectures, recitations, problem sets, labs, and large-scale datasets are freely available through MIT open courseware (OCW). In addition to the lecture slides, in recent years we have made available scribe notes for each lecture, as well as audio recordings. All lectures are currently being videotaped for Fall 2010, and all video recordings will similarly become freely available. All materials can be accessed from compbio.mit.edu/teaching.html.
   e. Developed a term-long integrated mentoring coursework to help students begin their independent research career as part of their final projects in 6.047/6.878. Bringing together ~20 postdocs and senior students in computational biology as mentors for the class, meeting with students regularly throughout the term. With their help the students forming teams and brainstorm ideas based on their own background and interests, as shared with their peers and mentors in a one-page profile. Through the term, they:
      - submit an NIH-style fellowship/grant proposal with their specific aims, milestones, datasets, algorithms, and deliverables
      - review peer proposals and receive anonymous peer reviews and a summary statement generated by the course staff
      - respond to peer reviews in bullet points and revise their proposal accordingly
      - submit a mid-course progress report with an outline of their final report and progress on aims
      - submit a conference-style write-up of their independent research
      - present their project orally in front of their peers.

      This aims at introduce students to active research, and help them acquire several real-world skills that are not typically taught in a classroom. Several projects have been published in journals and conferences, been used as Masters and PhD theses, and led to fruitful long-term and ongoing collaborations.
   f. In addition to the teaching materials, I have developed and made available, I have also strived to share all materials from the conferences I have organized, including videos of all talks, the full conference proceedings, all published papers, and all submitted abstracts for talks and posters. All are freely available at: http://compbio.mit.edu/recombsat/.
   g. I have developed new courses at MIT at the interface of computer science and biology. These are taught in the Computer Science Department but attract students from Computer Science, Math, Physics, Biological Engineering, Computational Biology within MIT, and

also students outside MIT from Harvard, Boston University, Tufts, the MIT/Harvard Health Science and Technology program , the Harvard School of Public Health, and Harvard Medical School, as well as non-student attendees including staff members from the Broad Institute, Whitehead Institute, and Novartis. The resulting materials have been widely distributed freely on the web through http://compbio.mit.edu/teaching.html and MIT Open CourseWare (OCW).

h. Computational Biology: Genomes, Networks, Evolution (6.047/6.878/HST.507). I developed an introductory computational biology course covering the algorithmic and machine learning foundations of computational biology, combining theory with practice. We cover both foundational topics in computational biology, and current research frontiers, and study fundamental techniques, recent advances in the field, and work directly with large-scale biological datasets. Topics include: Genomes: Biological sequence analysis, hidden Markov models, gene finding, comparative genomics, RNA structure, sequence alignment, hashing; Networks: Gene expression, clustering/classification, EM/Gibbs sampling, motifs, Bayesian networks, microRNAs, regulatory genomics, epigenomics; Evolution: Gene/species trees, phylogenomics, coalescent, genome duplication, population genomics, human ancestry, recent selection. Aiming this course to computer science students has allowed me to tackle foundational computational concepts that underlie the most pervasive techniques in genomics today, both to help students understand existing algorithms, and to enable them to design the next generation of algorithms. I have also integrated the use of large-scale biological datasets in the courses I am offering, to enable both graduate and undergraduate students to get the excitement and intuition that come with working directly with real data.

i. Independent research lab in Computational Biology (part of 6.047/6.878). I also developed an integrated final project experience that includes extensive mentoring by postdocs and senior graduate students, enabling students to conduct their independent research and also gain experience in research-enabling skills such as proposals and peer-review. As part of their final projects, the students (1) write an NIH-style research proposal, (2) review peer proposals in a study section format, (3) incorporate reviewer feedback and provide a revised proposal, (4) provide a midcourse progress update report, (5) prepare a written report as a journal-style scientific paper, and (6) present their results orally in a conference setting in front of their peers. This has led to several advanced final projects through the years, resulting in several conference and journal publications, masters theses and PhD theses, including a best thesis award, all from work that started in the classroom.

j. Computational Personal Genomics (6.881). I recently developed a new course focused on the computational challenges associated with complete genomes, including genotype phasing and haplotype reconstruction, variant imputation and ancestry painting, predicting likely causal variants, identifying quantitative trait loci for gene expression and chromatin, measuring recent human selection, using network information to decipher weak contributions from interacting loci, and deciphering complex traits. Students read and discuss seminal papers, and wrote their own software for analyzing primary large-scale datasets in weekly labs. We also discussed limitations, challenges, and ethical and social implications.

All educational materials for these courses are available at http://compbio.mit.edu/teaching.html, including the audio and video of all the lectures, all slides, and a 300-page working draft of a textbook for the first course.

2. Educational Contributions:

a. Developing a book for computational biology, titled “Computational Biology: Genomes, Networks, Evolution,” jointly with James Galagan (Broad Institute), with contract offers from MIT press and Cambridge University Press. The book is tailored to computer scientists aiming to learn the algorithmic and machine learning foundation of genomics. It covers the topics taught in the identically-titled course that I have developed (6.047).
b. Developing a book titled “Genomics,” with Amy Caudy (Princeton), Rick Myers (Stanford/Hudson Alpha), Jan Witkowski (Cold Spring Harbor Lab). The book is the successor of “Recombinant DNA” originally authored by James Watson, and used as a reference for more than 30 years. This complements my computational biology book, as it covers the experimental and biological foundations of genomics, which can be useful as a reference for biology students aiming to learn more about the analysis techniques of genomics, and also computer scientists aiming to gain a more in-depth understanding of the underlying biological foundations of genomics.
c. Serving as external expert on the Hellenic Quality Assurance Agency for Higher Education (HQAA), an independent authority with a mandate to organize the evaluation of individual academic units including both Departments and Universities as a whole in Greece.

1. Books
    1. None

2. Papers in Refereed Journals.
    1. Galagan, Calvo, Borkovich, Selker, Read, Jaffe, FitzHugh, Ma, Smirnov, Purcell, Rehman, Elkins, Engels, Wang, Nielsen, Butler, Endrizzi, Qui, Ianakiev, Bell-Pedersen, Nelson, Werner-Washburne, Selitrennikoff, Kinsey, Braun, Zelter, Schulte, Kothe, Jedd, Mewes, Staben, Marcotte, Greenberg, Roy, Foley, Naylor, Stange-Thomann, Barrett, Gnerre, Kamal, Kellis (as Kamvysselis), Mauceli, Bielke, Rudd, Frishman, Krystofova, Rasmussen, Metzenberg, Perkins, Kroken, Cogoni, Macino, Catcheside, Li, Pratt, Osmani, DeSouza, Glass, Orbach, Berglund, Voelker, Yarden, Plamann, Seiler, Dunlap, Radford, Aramayo, Natvig, Alex, Mannhaupt, Ebbole, Freitag, Paulsen, Sachs, Lander, Nusbaum, Birren, “The Genome Sequence of the Filamentous Fungus Neurospora crassa,” Nature, v. 422, p. 859-868, Apr 24, 2003. (PDF)
    2. Kellis, Patterson, Endrizzi, Birren, Lander, “Sequencing and Comparison of Yeast Species to Identify Genes and Regulatory Motifs,” Nature, v. 423 p. 241-254. May 15, 2003. (PDF)
    Cited in 1254 publications and featured in numerous reviews, magazines and newspapers.
    3. Chiang, Moses, Kellis, Lander, Eisen, “Phylogenetically and Spatially Conserved Word Pairs Associated with Gene-Expression Changes in Yeasts,” Genome Biology, 4(7):R43, 19 pages, June 26, 2003. (PDF)
    4. Moses, Chiang, Kellis, Lander, Eisen. “Position Specific Variation in the Rate of Evolution in Transcription Factor Binding Sites,” BMC Evolutionary Biology, 3:19, 13 pages, Aug 28, 2003. (PDF)
    5. Kellis, Patterson, Birren, Berger, Lander, “Methods in comparative genomics: genome correspondence, gene identification, regulatory motif discovery,” Journal of Computational Biology, v. 11, Numbers 2-3, p. 319-55, 2004. (PDF)
    6. Kellis, Birren, Lander, “Proof and Evolutionary Analysis of Ancient Genome Duplication in Yeast,” Nature, 428 pp. 617-624, Apr 8, 2004. (PDF)
    Cited in 690 publications and featured in numerous reviews, magazines and newspapers.

7. Kellis, "The changing face of genomics," Genome Biology, 5(5):324, 3 pages, Apr 30, 2004. (PDF)

8. Harbison, Gordon, Lee, J.Rinaldi, Macisaac, Danford, Hannett, Tagne, Reynolds, Yoo, Jennings, Zeitlinger, Kellis, Rolfe, Takusagawa, Gifford, Fraenkel, Young, "Transcri-ptional regulatory code of a eukaryotic genome," Nature, 431 pp. 99-104, Sep 2, 2004. (PDF)
Cited in 1234 publications.

9. Jaillon, Aury, Brunet, Petit, Stange-Thomann, Mauceli, Bouneau, Fischer, Ozouf-Costaz, Bernot, Nicaud, Jaffe, Fisher, Lutfalla, Dossat, Segurens, Dasilva, Salanoubat, Levy, Boudet, Castellano, Anthouard, Jubin, Castelli, Katinka, Vacherie, Biemont, Skalli, Cattolico, Poulain, Berardinis, Cruaud, Duprat, Brottier, Coutanceau, Gouzy, Parra, Lardier, Chapple, McKernan, McEwan, Bosak, Kellis, Volff, Guigo, Zody, Mesirov, Lindblad-Toh, Birren, Nusbaum, Kahn, Robinson-Rechavi, Laudet, Schachter, Quetier, Saurin, Scarpelli, Wincke, Lander, Weissenbach, Crollius, "Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype," Nature, 431 pp. 946-957, Oct 21, 2004. (PDF)
Cited in 714 publications.

10. Bernstein, Kellis, "Large-scale discovery and validation of functional elements in the human genome," Genome Biology 6(3):312, 2 pages, Mar 1, 2005. (PDF)

11. Xie, Lu, Kulbokas, Golub, Mootha, Lindblad-Toh, Lander, Kellis, "Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals," Nature, 434 p. 338-345, Mar 17, 2005. ** (PDF)
Cited in 1270 publications.

12. Lindblad-Toh, Wade, Mikkelsen, Karlsson, Jaffe, Kamal, Clamp, Chang, Kulbokas, Zody, Mauceli, Xie, Breen, Wayne, Ostrander, Ponting, Galibert, Smith, DeJong, Kirkness, Alvarez, Biagi, Brockman, Butler, Chin, Cook, Cuff, Daly, DeCaprio, Gnerre, Grabherr, Kellis, Kleber, Bardeleben, Goodstadt, Heger, Hitte, Kim, Koepfli, Parker, Pollinger, Searle, Sutter, Thomas, Webber, Baldwin, Abebe, Abouelleil, Aftuck, Ait-Zahra, Aldredge, Allen, et al, Zimmer, Lander, "Genome sequence, comparative analysis and haplotype structure of the domestic dog," Nature, 438 p.803-819, Dec 8, 2005. (PDF)

13. Grochow, Kellis, "Network Motif Discovery Using Subgraph Enumeration and Symmetry-Breaking". Lecture Notes in Computer Science (4453) 92-106, 2007. (PDF)

14. Zeitlinger, Zinzen, Stark, Kellis, Zhang, Young, Levine, "Genome-wide binding of *Drosophila* developmental factors shows integration of diverse patterning processes," Genes & Development, 21(4):385-90, Feb 15, 2007. (PDF)

15. Brennecke, Arabin, Stark, Dus, Kellis, Sachidanandam, Hannon, "Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*," Cell, 128(6): 1089-103, Mar 23, 2007. (PDF)
Cited in 405 publications.

16. Xie, Mikkelsen, Gnirke, Lindblad-Toh, Kellis, Lander, "Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites," PNAS, 104:7145-50, Apr 24, 2007. ** (PDF)

** Outgrowth of Supervised Student Research

17. Lin, Carlson, Crosby, Matthews, Yu, Park, et al, Celniker, Gelbart, Kellis. "Revisiting the protein-coding gene catalog of Drosophila melanogaster using 12 fly genomes," Genome Research 17:1823-36, Nov 7, 2007. ** (PDF)
18. Ruby, Stark, Johnston, Kellis, Bartel, Lai. "Evolution, biogenesis, expression, and target predictions of a expanded set of Drosophila microRNAs," Genome Research 17:1850-64, Nov 7, 2007. (PDF)
19. Stark, Kheradpour, Parts, Brennecke, Hodges, Hannon, Kellis. "Systematic discovery and characterization of fly microRNAs using 12 Drosophila genomes," Genome Research 17:1865-79, Nov 7, 2007. ** (PDF)
20. Kherapdour, Stark, Roy, Kellis. "Reliable prediction of regulator targets using 12 Drosophila genomes," Genome Research 17:1919-31, Nov 7, 2007. ** (PDF)
21. Rasmussen, Kellis. "Accurate gene-tree reconstruction by learning gene- and species-specific substitution rates across multiple complete genomes," Genome Research 17:1932-42, Nov 7, 2007. ** (PDF)
22. Clark, Eisen, Smith, Bergman, Oliver, Markow, Kaufman, Kellis, Gelbart; Drosophila 12 Genomes Consortium. "Evolution of genes and genomes in the Drosophila Phylogeny," Nature 450:203-218, Nov 8, 2007. (PDF)
    Cited in 267 publications and featured in numerous reviews, magazines and newspapers.
23. Stark, Lin, Kheradpour, Pedersen, Parts, Carlson, Crosby, Rasmussen, Roy, Deoras, Ruby, Brennecke, FlyBase curators, Berkeley Drosophila Genome Project, Hodges, Hinrichs, Caspi, Paten, Park, Han, Maeder, Polansky, Robson, Aerts, vanHelden, Hassan, Gilbert, Eastman, Rice, Weir, Hahn, Park, Dewey, Pachter, Kent, Haussler, Lai, Bartel, Hannon, Kaufman, Eisen, Clark, Smith, Celniker, Gelbart, Kellis, "Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures," Nature, 450:219-232, Nov 8, 2007. ** (PDF)
    Cited in 207 publications and featured in numerous reviews, magazines and newspapers.
24. Zeitlinger, Stark, Kellis, Hong, Nechaev, Adelman, Levine, Young. "RNA Polymerase Stalling at Developmental Control Genes in the Drosophila Embryo," Nature Genetics 39:1512-6, Nov 11, 2007. (PDF)
25. Clamp, Fry, Kamal, Xie, Cuff, Lin, Kellis, Lindblad-Toh, Lander. "Distinguishing protein-coding and noncoding genes in the human genome," PNAS 49:19428-19433, Dec 4, 2007. (PDF)
26. Stark, Bushati, Jan, Kheradpour, Hodges, Brennecke, Bartel, Cohen, Kellis. "A single Hox locus in Drosophila produces functional microRNAs from opposite DNA strands," Genes and Development 22:8-13, Jan 1, 2008. ** (PDF)
27. Presser, Elowitz, Kellis, Kishony. "Evolutionary dynamics of the Saccharomyces cerevisiae protein interaction network after duplication," PNAS 105(3):950-4, Jan 22, 2008. ** (PDF)
28. Lin, Deoras, Rasmussen, Kellis. "Performance and scalability of discriminative metrics for gene identification in 12 Drosophila genomes," PLoS Computational Biology 4(4):e1000067, 14 pages, Apr 18, 2008. ** (PDF)

** Outgrowth of Supervised Student Research

29. Murchison, Kheradpour, Sachidanandam, Smith, Hodges, Xuan, Kellis, Gruetzner, Stark, Hannon. "Conservation of small RNA pathways in platypus," Genome Research 18:995-1004, May 7, 2008. (PDF)

30. Warren, Hillier, Marshall Graves, Birney, Ponting, Gruetzner, Belov, Miller, Clarke, Chinwalla, Yang, Murchison, et al. Stark, Kheradpour, Kellis, Flicek, Chen, Webber, Hardison, Nelson, Hallsworth-Pepin, Delehaunty, Markovic, Minx, Feng, Kremitzki, Mitreva, Glasscock, Wohldmann, Thiru, Nhan, Pohl, Smith, Hou, Renfree, Mardis, Wilson. "Genome analysis of the platypus reveals unique signatures of evolution," Nature 453:175-183, May 8, 2008. (PDF)

31. Czech, Malone, Zhou, Stark, Schlingeheyde, Dus, Perrimon, Kellis, Wohlschlegel, Sachidanandam, Hannon, Brennecke. "An endogenous small interfering RNA pathway in Drosophila," Nature 453:729-31, Jun 5, 2008. (PDF)

32. Guttman, Amit, Garber, French, Lin, Feldser, Hurate, Mikkelsen, Zuk, Jacks, Hacohen, Bernstein, Kellis, Regev, Rinn, Lander. "Chromatin signature reveals over a thousand highly conserved, large non-coding RNAs in mammals," Nature 458:223-7, March 12, 2009. ** (PDF)
Cited in 209 publications.

33. Heintzman, Hon, Hawkins, Kheradpour, Stark, Stuart, Harp, Ye, Ching, Ching, Antosiewicz, Liu, Zhang, Green, Stewart, Crawford, Kellis, Ren. "Histone modifications at human enhancers reflect global cell-type-specific gene expression," Nature 459:108-12, May 7, 2009. ** (PDF)

34. Butler, Rasmussen, Lin, Santos, Sakthikumar, Munro, Rheinbay, Grabherr, Agrafioti, Arnaud, Bates, Berman, Brown, Brunke, Constanzo, Fitzpatrick, Forche, deGroot, Harris, Hoyer, Hube, Klis, Kodira, Lennard, Logue, Martin, Neiman, Nikolaou, Quail, Quinn, Reedy, Schmitzberger, Sherlock, Shah, Silverstein, Skrypek, Soll, Staggs, Stumpf, Sudbery, Thyagarajan, Zeng, Berriman, Heitman, Lorenz, Gow, Birren, Kellis*, Cuomo. "Evolution of pathogenicity and sexual reproduction in eight Candida genomes," Nature 459:657-62, Jun 4, 2009. ** (PDF)
*=co-corresponding author

35. Pruitt, Harrow, Harte, Wallin, Diekhans, Maglott, Searle, Farrell, Loveland, Ruef, Hart, Suner, Landrum, Aken, Ayling, Baertsch, Fernandez-Banet, Cherry, Curwen, DiCuccio , Kellis, Lee, Lin, Schuster, Shkeda, Amid, Brown, Dukhanina, Frankish, Hart, Maidak, Mudge, Murphy, Murphy, Rajan, Rajput, Riddick, Snow, Steward, Webb, Weber, Wilming, Wu, Birney, Haussler, Hubbard, Ostell, Durbin, Lipman. "The consensus coding sequence (CCDS) project – identifying a common protein-coding gene set for the human and mouse genomes." Genome Research 19:1316-23, Jun 4, 2009. (PDF)

36. Celniker, Dillon, Gerstein, Gunsalus, Henikoff, Karpen, Kellis, Lai, Lieb, MacAlpine, Micklem, Piano, Snyder, Stein, White, Waterston; modENCODE Consortium. "Unlocking the secrets of the genome." Nature 459:927-30. Jun 18, 2009. (PDF)

37. Syed, Stultz, Kellis, Indyk, Guttag, "Motif Discovery in Physiological Datasets: A Methodology for Inferring Predictive Elements," ACM Transactions on Knowledge Discovery from Data 4:1 Article 2, 23 pages, January 2010. ** (PDF)

** Outgrowth of Supervised Student Research

38. Murchison, Tovar, Hsu, Bender, Kheradpour, Rebbeck, Obendorf, Conlan, Bahlo, Blizzard, Pyecroft, Kreiss, Kellis, Stark, Harkins, Marshall Graves, Woods, Hannon, Papenfuss. "The Tasmanian devil transcriptome reveals Schwann cell origins of a clonally transmissible cancer," Science 327(5961):84-7. Jan 1, 2010. (PDF)

39. Nègre, Brown, Shah, Kheradpour, Morrison, Henikoff, Feng, Ahmad, Russell, White, Stein, Henikoff, Kellis, White. "A comprehensive map of insulator elements for the Drosophila genome," PLoS Genetics 6(1):e1000814, 13 pages. Jan 15, 2010. ** (PDF)

40. Kellis, Rinn, "Sequences to systems," Genome Biology 11(5):303, 3 pages, May 25, 2010. (PDF)

41. Rasmussen, Kellis, "A Bayesian approach for fast and accurate gene tree reconstruction," Molecular Biology and Evolution, Mol Biol Evol. 28:273–290, Sep 7 2011, Advance Access on July 25, 2010, DOI 10.1093/molbev/msq189. ** (PDF)

42. Ernst, Kellis, "Discovery and characterization of chromatin states for systematic annotation of the human genome," Nature Biotechnology 28(8):817-25. July 25, 2010. ** (PDF)

43. Mueller, Kohajda, Findeiss, Stadler, Washietl, Kellis, von Bergen, Kalkhof. "Optimization of parameters for coverage of low molecular weight proteins," Annals of Bioanalytical Chemistry, 398(7-8):2867-81, Aug 28, 2010.** (PDF)

44. Bernstein, Stamatoyannopoulos, Costello, Ren, Milosavljevic, Meissner, Kellis, Marra, Beaudet, Ecker, Farnham, Hirst, Lander, Mikkelsen, Thomson, "The NIH Roadmap Epigenomics Mapping Consortium," Nature Biotechnology, 28(10):1045-8, Oct 2010. (PDF)

45. The modENCODE Consortium, Roy, Ernst, Kharchenko, Kheradpour, Nègre, Eaton, Landolin, Bristow, Ma, Lin, Washietl, Arshinoff, Ay, Meyer, Robine, Washington, Di Stefano, Berezikov, Brown, Brown, Candeias, Carlson, Carr, Jungreis, Marbach, Sealfon, Tolstorukov, Alekseyenko, Artieri, Boley, Booth, Brooks, Dai, Davis, Duff, Feng, Gorchakov, Gu, Henikoff, Kapranov, Li, Li, MacAlpine, Malone, Minoda, Nordman, Okamura, Perry, Powell, Riddle, Sakai, Samsonova, Sandler, Schwartz, Sher, Spokony, Sturgill, van Baren, Will, Wan, Yang, Yu, Feingold, Good, Guyer, Lowdon, Ahmad, Andrews, Berger, Bickel, Brenner, Brent, Cherbas, Elgin, Gingeras, Grossman, Hoskins, Kaufman, Kent, Kuroda, Orr-Weaver, Perrimon, Pirrotta, Posakony, Ren, Russell, Cherbas, Graveley, Lewis, Micklem, Oliver, Park, Celniker, Henikoff, Karpen, Lai, MacAlpine, Stein, White, Kellis, "Identification of functional elements and regulatory circuits in Drosophila by large-scale data integration," Science, 330(6012):1787-97, December 24, 2010. ** (PDF)

46. Kharchenko, Alekseyenko, Schwartz, Minoda, Riddle, Ernst, Sabo, Larschan, Gorchakov, Gu, Linder-Basso, Plachetka, Shanower, Tolstorukov, Bishop, Canfield, Sandstrom, Thurman, Stamatoyannopoulos, Kellis, Elgin, Kuroda, Pirotta, Karpen, "Comprehensive analysis of the Drosophila melanogaster chromatin landscape differentiates among chromosomes, genes, and regulatory elements," Nature 471:480-485, Dec 22, 2010, doi:10.1038/nature09725. ** (PDF)

47. Hubisz, Lin, Kellis, Siepel, "Error and Error Mitigation in Low-Coverage Genome Assemblies". PLoS ONE 6(2):e17034, 13 pages, Feb 14, 2011. (PDF)

48. Ay, Kellis, Kahveciy, "SubMAP: Aligning metabolic pathways with subnetwork mappings," Journal of Computational Biology 18(3):219-35, Mar 1, 2011.** (PDF)

49. Nègre, Brown, Ma, Bristow, Miller, Kheradpour, Loriaux, Sealfon, Li, Ishii, Spokony, Chen, Hwang, Wagner, Auburn, Domanus, Shah, Morrison, Zieba, Suchy, Senderowicz, Victorsen, Bild, Grundstad, Hanley, Mannervik, Venken, Bellen, White, Russell, Grossman, Ren, Posakony, Kellis*, White, "A Cis-Regulatory Map of the Drosophila Genome," Nature 471:527-531, March 23, 2011. **(PDF)
*=co-corresponding author

50. Ernst, Kheradpour, Mikkelsen, Shoresh, Ward, Epstein, Zhang, Wang, Issner, Coyne, Ku, Durham, Kellis*, Bernstein, "Systematic determination and analysis of chromatin state dynamics in nine human cell types," Nature 473(7345):43-9, Mar 23, 2011. ** (PDF)
*=sole corresponding author

51. Rhind, Chen, Yassour, Thompson, Haas, Habib, Wapinski, Roy, Lin, Heiman, Young, Furuya, Guo, Pidoux, Chen, Robbertse, Goldberg, Aoki, Bayne, Berlin, Desjardins, Dobbs, Dukaj, Fan, Fitzgerald, French, Gujja, Hansen, Keifenheim, Levin, Mosher, Muller, Pfiffner, Priest, Russ, Smialowska, Swoboda, Sykes, Vaughn, Vengrova, Yoder, Zeng, Allshire, Baulcombe, Birren, Brown, Ekwall, Kellis, Leatherwood, Levin, Margalit, Martienssen, Nieduszynski, Spatafora, Friedman, Dalgaard, Baumann, Niki, Regev, Nusbaum. "Comparative Functional Genomics of the Fission Yeasts". Science 332(6032):930-6, Apr 21, 2011. (PDF)

52. Magklara, Colquitt, Clowney, Yen, Allen, Markenscoff-Papadimitriou, Kheradpour, Mountoufaris, Carey, Barnea, Kellis, Lomvardas, "An Epigenetic Signature for Monoallelic Olfactory Receptor Expression," Cell 145:555-70, May 13, 2011. ** (PDF)

53. Hung, Wang, Lin, Koegel, Kotake, Grant, Horlings, Shah, Argani, Wang, Kim, van de Vijver, Sukumar, Whitfield, Kellis, Xiong, Wong, Chang, "Extensive and choreographed transcription of noncoding RNAs within cell cycle promoters," Nature Genetics 43:616-7, Jun 5, 2011. (PDF)

54. The ENCODE Project Consortium, "ENCODE 101: A user's guide to the ENCyclopedia of DNA Elements," PLoS Biology 9(4):e1001046, Apr 19, 2011. (PDF)

55. Lin, Jungreis, Kellis, "PhyloCSF: a comparative genomics method to distinguish protein-coding and non-coding regions," Bioinformatics, 27(13):i275-82, Jul 1, 2011. ** (PDF)

56. Lowe, Kellis, Siepel, Raney, Clamp, Salama, Kingsley, Lindblad-Toh, Haussler, "Three different periods of regulatory innovation during vertebrate evolution," Science 333:1019-24, Aug 19 2011. (PDF)

57. Wu, Rasmussen, Kellis, "Evolution at the sub-gene level: gene architecture rearrangements in the Drosophila phylogeny," Mol Biol Evol. Sep 7, 2011. ** (PDF)

58. Wu, Kumar, Mishra, Cheng, Keller Capone, Marie Dorman, Chen, Dogan, Morrissey, Weiss, Blobel, Ernst, Kellis, Taylor, Chiaromonte, Zhang, Miller, Hardison. "The epigenetic landscape of erythroid differentiation after GATA1 restoration," Genome Research 21:1659-71, October 2011 (PDF)

59. Lin, Washietl, Kheradpour, Parker, Pedersen, Kellis, "Locating protein-coding sequences under purifying selection for additional, overlapping functions in 29 mammalian genomes," Genome Research, 21:1916-28, Oct 12, 2011. ** (PDF)

** Outgrowth of Supervised Student Research

60. Parker, Moltke, Roth, Washietl, Wen, Kellis, Breaker, Pedersen, "New families of human regulatory RNA structures identified by comparative analysis of vertebrate genomes," Genome Research 21(11):1929-43, Nov 2011. (PDF)

61. Jungreis, Lin, Chan, Kellis, "Evidence of Abundant Stop Codon Readthrough in Drosophila and other Metazoa," Genome Research 21(12):2096-113, Dec 2011. ** (PDF)

62. Lindblad-Toh, Garber, Zuk, Lin, Parker, Washietl, Kheradpour, Ernst, Jordan, Mauceli, Ward, Lowe, Holloway, Clamp, Gnerre, Alfoldi, Beal, Chang, Clawson, Di Palma, Fitzgerald, Hubis, Jaffe, Jungreis, Kostka, Lara, Martins, Massington, Moitke, Raney, Rasmussen, Vilella, Wen, Zody, Broad Institute Sequencing Platform, Worley, Gibbs, Baylor College of Medicine Human Genome Sequencing Center, Warren, Wilson, Washington University Genome Sequencing Center Sequencing Platform, Birney, Margulies, Herrero, Green, Haussler, Siepel, Goldman, Pollard, Pedersen, Lander, Kellis, "A high-resolution map of evolutionary constraint in the human genome based on 29 placental mammals," Nature 478(7370):476-82, Oct 12, 2011. (PDF)**

63. Ward, Kellis. "HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants." Nucleic Acids Research, Nov 7, 2011. (PDF)**

64. Ram, Goren, Amit, Shoresh, Yosef, Ernst, Kellis, Gymrek, Issner, Coyne, Durham, Zhang, Donaghey, Epstein, Regev, Bernstein. "Combinatorial Patterning of Chromatin Regulators Uncovered by Genome-wide Location Analysis in Human Cells". Cell 147(7):1628-39, Dec 23, 2011. (PDF)**

65. Rasmussen, Kellis. "Unified modeling of gene duplication, loss, and coalescence using a locus tree". Genome Research Jan 23, 2012. (PDF)**

66. Washietl, Hofacker, Stadler, Kellis. "RNA folding with soft constraints: reconciliation of probing data and thermodynamic secondary structure prediction". Nucleic Acids Research, Jan 28, 2012. (PDF)**

67. Melnikov, Murugan, Zhang, Tesileanu, Wang, Rogov, Feizi, Gnirke, Callan, Kinney, Kellis, Lander, Mikkelsen. "Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay". Nature Biotechnology 30(3):271-7, Feb 26, 2012. (PDF)**

68. Ernst, Kellis. "ChromHMM: automating chromatin-state discovery and characterization". Nature Methods 9:215-6, Feb 28, 2012. (PDF)**

69. Marbach, Roy, Ay, Meyer, Candeias, Kahveci, Bristow, Kellis. "Predictive regulatory models in Drosophila melanogaster by integrative inference of transcriptional networks". Genome Research, 16 pages, Mar 28, 2012. (PDF)**

70. Wiggs, Yaspan, Hauser, Kang, Allingham, Olson, Abdrabou, Fan, Wang, Brodeur, Budenz, Caprioli, Crenshaw, Crooks, Delbono, Doheny, Friedman, Gaasterland, Gaasterland, Laurie, Lee, Lichter, Loomis, Liu, Medeiros, McCarty, Mirel, Moroi, Musch, Realini, Rozsa, Schuman, Scott, Singh, Stein, Trager, Vanveldhuisen, Vollrath, Wollstein, Yoneyama, Zhang, Weinreb, Ernst, Kellis, Masuda, Zack, Richards, Pericak-Vance, Pasquale, Haines. "Common variants at 9p21 and 8q22 are associated with increased susceptibility to optic nerve degeneration in glaucoma". PLoS Genetics 8(4):e1002654, 12 pages, April 26, 2012. (PDF)**

71. Bansal, Alm, Kellis, “Efficient Algorithms for the Reconciliation Problem with Gene Duplication, Horizontal Transfer, and Loss”. Bioinformatics 28:i283-i291 and ISMB, Jun 15, 2012. (PDF)**

72. Marbach, Costello, Kuffner, Vega, Prill, Camacho, Allison, DREAM5 Consortium, Kellis, Collins, Stolovitzky. “Wisdom of crowds for robust gene network inference”. Nature Methods 2012, doi:10.1038/nmeth.2016, 15 July 2012. (PDF)**

73. Wu, Rasmussen, Bansal, Kellis. “TreeFix: Statistically Informed Gene Tree Error Correction using Species Trees”. Systematic Biology, Sep 4, 2012. (PDF)**

74. Spivakov, Akhtar, Kheradpour, Beal, Girardot, Koscielny, Herrero, Kellis, Furlong, Birne. “Analysis of variation at transcription factor binding sites in Drosophila and humans”. Genome Biology 13:R49, Sep 5, 2012. (PDF)**

75. ENCODE Project Consortium. “An integrated encyclopedia of DNA elements in the human genome”. Nature 489:57-74. Sep 6, 2012. (PDF)**

76. Harrow, Frankish, Gonzalez, Tapanari, Diekhans, Kokocinski, Aken, Barrell, Zadissa, Searle, Barnes, Bignell, Boychenko, Hunt, Kay, Mukherjee, Rajan, Despacio-Reyes, Saunders, Steward, Harte, Lin, Howald, Tanzer, Derrien, Chrast, Walters, Balasubramanian, Pei, Tress, Rodriguez, Ezkurdia, van Baren, Brent, Haussler, Kellis, Valencia, Reymond, Gerstein, Guigo, Hubbard. “GENCODE: The reference human genome annotation for The ENCODE Project”. Genome Research 22:1760-74, Sep 2012. (PDF)**

77. Landt, Marinov, Kundaje, Kheradpour, Pauli, Batzoglou, Bernstein, Bickel, Brown, Cayting, Chen, Desalvo, Epstein, Fisher-Aylor, Euskirchen, Gerstein, Gertz, Hartemink, Hoffman, Iyer, Jung, Karmakar, Kellis, Kharchenko, Li, Liu, Liu, Ma, Milosavljevic, Myers, Park, Pazin, Perry, Raha, Reddy, Rozowsky, Shoresh, Sidow, Slattery, Stamatoyannopoulos, Tolstorukov, White, Xi, Farnham, Lieb, Wold, Snyder. “ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia”. Genome Research 22(9):1813-31, Sep 2012. (PDF)**

78. Ward, Kellis. “Evidence of Abundant Purifying Selection in Humans for Recently Acquired Regulatory Functions”. Science 337:1675-8, Sep 5 2012. (PDF)**

79. Sealfon, Gire, Ellis, Calderwood, Qadri, Hensley, Kellis, Ryan, Larocque, Harris, Sabeti. “High depth, whole-genome sequencing of cholera isolates from Haiti and the Dominican Republic”. BMC Genomics 13:468, Sep 11, 2012. (PDF)**

80. Washietl, Will, Hendrix, Goff, Rinn, Berger, Kellis. “Computational analysis of noncoding RNAs”. Wiley Reviews RNA, Sep 18, 2012. (PDF)**

81. Meuleman, Peric-Hupkes, Kind, Beaudry, Pagie, Kellis, Reinders, Wessels, van Steensel. “Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence”. Genome Research, Nov 2, 2012. (PDF)**

82. Ward, Kellis. “Interpreting noncoding genetic variation in complex traits and human disease”. Nature Biotechnology 30:1095-1106, Nov 2012. (PDF)**

83. Jin, Ernst, Tiedemann, Xu, Sureshchandra, Kellis, Dalton, Liu, Choi, Robertson. “Linking DNA Methyltransferases to Epigenetic Marks and Nucleosome Structure Genome-wide in Human Tumor Cells”. Cell Reports 2:1411-24, Nov 29, 2012. (PDF)**

84. Talkowski, Maussion, Crapper, Rosenfeld, Blumenthal, Hanscom, Chiang, Lindgren, Pereira, Ruderfer, Diallo, Lopez, Turecki, Chen, Gigek, Harris, Lip, An, Biagioli,

Macdonald, Lin, Haggarty, Sklar, Purcell, Kellis, Schwartz, Shaffer, Natowicz, Shen, Morton, Gusella, Ernst. “Disruption of a large intergenic noncoding RNA in subjects with neurodevelopmental disabilities”. American Journal of Human Genetics 91:1128-34, Dec 7, 2012. (PDF)**

85. Hoffman*, Ernst*, Wilder, Kundaje, Harris, Libbrecht, Giardine, Ellenbogen, Bilmes, Birney, Hardison, Dunham, Kellis*, Noble*. “Integrative annotation of chromatin elements from ENCODE data”. Nucleic Acids Research, Dec 5 2012. (PDF)**
86. Grote, Wittler, Hendrix, Koch, Wahrisch, Beisaw, Macura, Blass, Kellis, Werber, Herrmann. “The Tissue-Specific lncRNA Fendrr Is an Essential Regulator of Heart and Body Wall Development in the Mouse”. Dev Cell 24:206-14, Jan 28, 2013. doi: 10.1016/j.devcel.2012.12.012. (PDF)**
87. Sun, Goff, Trapnell, Alexander, Lo, Hacisuleyman, Sauvageau, Tazon-Vega, Kelley, Hendrickson, Yuan, Kellis, Lodish, Rinn. “Long noncoding RNAs regulate adipogenesis”. PNAS 110:3387-92, Feb 26, 2013. (PDF)**
88. Rajagopal, Xie, Li, Wagner, Wang, Stamatoyannopoulos, Ernst, Kellis, Ren. “RFECS: A Random-Forest Based Algorithm for Enhancer Identification from Chromatin State”. PLoS Computational Biology, Mar 14, 2013 doi/10.1371. (PDF)**
89. Roy, Wapinski, Pfiffner, French, Socha, Konieczka, Habib, Kellis, Thompson, Regev, “Arboretum: reconstruction and analysis of the evolutionary history of condition-specific transcriptional modules,” Genome Research, 12 pages, in press.
90. Ernst, Kellis, “Interplay between chromatin state, regulator binding, and regulatory motifs in six human cell types”, Genome Research, 12 pages, in press.
91. Kheradpour, Ernst, Melnikov, Rogov, Wang, Zhang, Alston, Mikkelsen, Kellis, “Systematic dissection of regulatory motifs in 2,000 predicted human enhancers using a massively parallel reporter assay”. Genome Research, 12 pages, in press.

Electronic copies of all papers are available from: http://compbio.mit.edu/papers.html or by clicking the PDF links on the electronic version of this document.

3. Proceedings of Refereed Conferences
    1. Amenta, Bern, Kellis (as Kamvysselis), “Crust: A new Voronoi-Based Surface Reconstruction Algorithm,” ACM SIGGRAPH, v. 32, p. 415-421, Jul 19, 1998. (PDF) *Cited in 686 publications.*
    2. Kellis, Patterson, Birren, Berger, Lander, “Whole-Genome Comparative Annotation and Motif Discovery in Multiple Yeast Species,” ACM RECOMB, p. 157-166, Apr 13, 2003. (PDF)
    3. Chiang, Moses, Kellis (as Kamvysselis), Lander, Eisen. “Phylogenetically and Spatially Conserved Word Pairs Associated with Gene-Expression Changes in Yeasts,” ACM RECOMB, p. 84-93, Apr 13, 2003. (PDF)
    4. Grochow, Kellis, “Network motif discovery using motif enumeration and symmetry conditions,” ACM RECOMB, p. 92-106, Apr 21, 2007. (PDF)
    5. Meyer, Marbach, Roy, Kellis, “Information-Theoretic Inference of Regulatory Networks Using Backward Elimination,” BioComp10, 12 pages, July 13, 2010. (PDF)
    6. Lin, Jungreis, Kellis, “PhyloCSF: a comparative genomics method to distinguish protein-coding and non-coding regions,” ISMB 2011, 12 pages, in press. ** (PDF)

7. Bansal, Alm, Kellis, “Efficient Algorithms for the Reconciliation Problem with Gene Duplication, Horizontal Transfer, and Loss”. ISMB 2012, 9 pages, July 2012.

4. Other Major Publications.
   1. Kellis (as Kamvysselis), Marina, “Imagina: Sketch-based Image Retrieval using Cognitive Abstraction,” Masters Thesis, MIT Libraries, 157 pages, May 20, 1999. (PDF)
   2. Kellis (as Kamvysselis), “Computational Comparative Genomics: Genes, Regulation, Evolution,” Ph.D. Thesis, MIT Libraries, 100 pages, May 25, 2003. (PDF)
   3. Kellis, “Gene finding using multiple related species: a classification approach,” chapter in Encyclopedia of Genetics, Genomics, Proteomics, John Wiley & Sons, 7 pages, July 15, 2005. (PDF)
   4. Organ, Rasmussen, Baldwin, Kellis, Edwards, “A Phylogenomic Approach to the Evolutionary Dynamics of Gene Duplication in Birds,” chapter in Evolution after Gene Duplication, Wiley-Blackwell, 17 pages, October 13, 2010. ISBN: 978-0-470-59382-0 (PDF)
   5. Kellis, Califano, Stolovitsky, eds., “RECOMB Systems Biology, Regulatory Genomics, and DREAM 2009 Special Issue,” Journal of Computational Biology (16):2, 133-406, January 29, 2009.
   6. Bar-Joseph, Califano, Kellis, Stolovitsky, eds., “RECOMB Systems Biology, Regulatory Genomics, and DREAM 2010 Special Issue,” Journal of Computational Biology 18(2):131, Feb 2011. (PDF)

   Electronic copies of all papers in review are available at: http://compbio.mit.edu/papers.html

5. Internal Memoranda and Progress Reports.
   1. Kellis (as Kamvysselis), Nielsen, “WebBot: Constraint Model for a Web Robot,” World Wide Web Consortium, Aug 1996. (PDF)
   2. Kellis (as Kamvysselis), “EciMorph: Curve Morphing in Extended Gaussian Space,” MIT Machine Vision Group, Dec 1997. (PDF)
   3. Kellis (as Kamvysselis), Gajos, Blum, Vassef, “3DMorph: Polygon Model Morphing,” MIT Lab for Computer Science, Dec 1997. (PDF)
   4. Kellis (as Kamvysselis), Lueck, Rohrs, “RoboLogo: Programming Environment for Interactive Robots,” MIT Microprocessor Lab, Dec 1998. (PDF)
   5. Kellis (as Kamvysselis), Marina, Vassef, Carter, “Mood: Music Classification using Patterns of Attentional State,” MIT AI Lab, May 1998. (PDF)
   6. Kellis (as Kamvysselis), Mehrotra, Warshawsky, “Invest: Applications of AI to Stock Market Prediction,” AI in Practice Project, May 1998. (PDF)
   7. Kellis (as Kamvysselis), “Handwritten character recognition using wavelets,” MIT AI Lab, Dec 1999. (PDF)

   Electronic copies of all internal reports are available at: http://compbio.mit.edu/papers.html

6. Invited Lectures

   Jun. 2002, “Yeast Comparative Genomics,” Whitehead Institute First Symposium on Genomics, Cambridge, MA.

Aug. 2002, “Yeast Comparative Genomics,” Harvard Institute of Proteomics Joint Colloquium, Cambridge, MA.

Mar. 2003, “Yeast Comparative Genomics,” Harvard Center for Genomics Research, Lectures on Genomics, Cambridge, MA.

Mar. 2004, “Computational Biology: Challenges and Opportunity,” Tufts Medical School Faculty Lectures on Computational Biology, Boston, MA.

Apr. 2004, “Extracting functional information from genome comparisons,” Human Genome Discovery and Gene Functional Analysis meeting, San Francisco, CA.

May 2004, “Regulatory motif discovery using comparative genomics,” Harvard Medical School Conference on Analysis of Genomic Data, Cambridge MA.

Jun. 2004, “Comparative genomics in *Saccharomyces cerevisiae*,” Gordon Conference Invited Plenary Lecture on Phylogenomics, Plymouth NH.

Dec. 2004, “Revisiting the yeast genome,” From Genome Sequencing to Biological Systems, ORFeome Meeting, Dana Farber.

May 2005, “Regulatory motif discovery using comparative genomics,” First Proteomics Meeting of the Hellenic Proteomics Society, Athens Greece.

Jun. 2005, “Comparative genomics: Genes, Networks, Evolution,” Bertinoro Meeting on Comparative Genomics, Invited Lecture, Italy.

Jun. 2005, “Computational Comparative Genomics,” Advanced Course in Functional Genomics, Invited Keynote Lecture, Sanger Center.

Jun. 2005, “Comparative genomics: Genes, Networks, Evolution,” Intelligent Systems for Molecular Biology, Invited Lecture, ISCB Student Council.

Aug. 2005, “Comparative Genomics in Yeast,” International Meeting on Yeast Genetics Keynote Lecture, Bratislava, Slovakia.

Oct. 2005, “Computational Comparative Genomics,” Princeton University, Quantitative and Computational Biology series, Lewis-Singler Institute.

Nov. 2005, “Computational genomics,” Tufts University Computer Science colloquium, Invited Lecture, Medford MA.

Dec. 2005, “Regulatory genomics,” RECOMB Satellite meeting on Regulatory Genomics, Invited Lecture, San Diego, CA.

Mar. 2006, “Regulatory Networks in the fly,” Systems Biology: Global Regulation of Gene Expression, Invited Lecture, Cold Spring Harbor, NY.

May 2006, “Fly Comparative Genomics,” Cold Spring Harbor Laboratory (CSHL), The Biology of Genomes, Invited Lecture, CSHL, NY.

Nov. 2006, “Discovering the human gene set using comparative genomics,” Sanger Institute, BioSapiens workshop, Invited Lecture, Wellcome Trust.

Nov. 2006, “Interpreting the human genome,” The Institute for Advanced Study, Simons Center for Systems Biology, Invited Lecture, Princeton, NJ.

Oct. 2006, “Interpreting the human genome,” Stanford University, Frontiers in Biology lecture series, Invited Lecture, Palo Alto, CA.

Nov. 2006, “Interpreting the human genome,” Algorithmic Biology, Cal(IT)2, UC San Diego, Invited Keynote Lecture, San Diego, CA.

Mar. 2007, “Biological signal discovery in 12 *Drosophila* genomes,” Drosophila 2007 meeting, Invited Lecture, Philadelphia, PA.

Mar. 2007, “Pre- and post-transcriptional regulatory networks in 12 Drosophila genomes,” Keystone Meeting on Systems Biology and Regulatory Networks, Steamboat Spring, CO.

May 2007, “Inferring the cis-regulatory code of Drosophila melanogaster,” modENCODE planning meeting, Fox Hollow, Woodbury, NY.

May 2007, “Comparative Genomics of 12 Drosophila species,” CSHL Genome Biology, Cold Spring Harbor, NY.

Jul. 2007, “Phylogenomics in mammalian, fly, and fungal genomes,” ISMB 2007 SIG, Invited Lecture, Vienna, Austria.

Sep. 2007, “Interpreting genomes using comparative genomics,” MIT Club of Greece celebrates the TR35, lecture given remotely by video conference from Boston to Athens.

Sep. 2007, “Computational methods for studying the human genome,” CSAIL student workshop invited lecture.

Oct. 2007, “Systematic discovery of protein-coding genes using evolutionary signatures,” BioCuration meeting, Invited Keynote Lecture, San Jose, CA.

Oct. 2007, “Phylogenomics of mammalian, fly, and fungal gnomes,” Harvard OEB, Invited Lecture, Cambridge MA.

Oct. 2007, “Computational Comparative Genomics,” Columbia C2B2, Invited Lecture, New York, NY.

Oct. 2007, “Systems Biology of gene regulation,” Linnaeus Tricentennial Workshop, Uppsala, Sweden.

Nov. 2007, “Machine learning for Phylogenomics,” CSHL Genome Informatics, Invited Lecture, Cold Spring Harbor, NY.

Nov. 2007, “Comparative genomics in flies and mammals,” Genome Biology, Broad Institute Retreat, Invited Lecture, Boston MA.

Dec. 2007, “Algorithms for regulatory network discovery using comparative genomics,” RECOMB Satellite on Systems Biology, Invited Keynote Lecture, San Diego, CA.

Dec. 2007, “Machine Learning and Probabilistic Methods in Computational Genomics,” NIPS 2007, MLCB workshop invited lecture, Vancouver, Canada.

Feb. 2008, “Interpreting the Human genome,” IEEE Engineering in Medicine and Biology Society, Cambridge MA.

Feb. 2008, “Regulatory Genomics of Drosophila and Mammalian Genomes,” Stanford University, Carnegie Institution, Palo Alto, CA.

Feb. 2008, “Comparative genomics in flies and mammals,” UC Berkeley, Statistics and Genomics, Berkeley, CA.

Feb. 2008, “Regulatory Genomics of Drosophila and Mammalian Genomes,” Columbia University, Center for Computational Biology and Bioinformatics (C2B2), New York, NY.

Mar. 2008, “Computational Challenges in Genomics,” University of New Mexico Annual Student Workshop Keynote lecture, Albuquerque, NM.

Mar. 2008, “Regulatory Genomics of Drosophila melanogaster,” University of New Mexico Biology Symposium, Albuquerque, NM.

Mar. 2008, “Genomics and Medicine,” Massachusetts General Hospital, Neurosurgery, Invited lecture, Boston, MA.

Apr. 2008, “Regulatory genomics of 12 *Drosophila* species,” Drosophila 2008, Invited Plenary Lecture, San Diego, CA.

Apr. 2008, “Interpreting the Human Genome,” National Technical University of Athens, Annual visit to MIT, Cambridge MA.

May 2008, “Comparative Genomics of Drosophila and Mammalian Species,” CSHL The Biology of Genomes, Cold Spring Harbor, NY.

May 2008, “Regulatory logic in *Drosophila* development,” Howard Hughes Medical Institute, Janelia Farms, The Logic of Gene Regulation, Invited Lecture, Ashburn, VA.

Jun. 2008, “Comparative genomics of 29 mammals,” ENCODE/modENCODE meeting ENCyclopedia Of DNA Elements, Invited Lecture, Rockville, MD.

Jun. 2008, “Regulatory circuit inference and dynamics in Drosophila melanogaster,” Aegean Conference on Networks, Pathways, and Systems, Invited Lecture, Crete, Greece.

Jun. 2008, “Regulatory genomics of Drosophila melanogaster,” Annual EMBO meeting on Drosophila Development, Invited Speaker, Crete, Greece.

Jul. 2008, “Interpreting the Human Genome,” National Technical University of Athens, Invited Lecture, Athens, Greece.

Jul. 2008, “Regulatory Genomics of Drosophila and Mammalian species,” ISMB 2008, Post-ENCODE workshop, Invited Speaker, Toronto, Canada.

Jul. 2008, “Computational Challenges in Genomics,” National Science Foundation (NSF) Emerging Models and Technologies (EMT) workshop, Invited Speaker, Princeton NJ.

Oct. 2008, “Mathematical Challenges in Computational Biology,” MIT Undergraduate Math Association, Cambridge MA.

Dec. 2008, “Computational Interpretation of the Human genome,” Innovations 2009 Invited Keynote Speaker, Al Ain, United Arab Emirates.

Dec. 2008, “Regulatory Genomics of Drosophila melanogaster,” Genome Institute of Singapore Invited Lecture, Singapore, Singapore.

Jan. 2009, “Computational Genomics of mammalian species,” Indian Institute of Technolgoy (IIT) Kanpur Invited Lecture, Kanpur, India.

Feb. 2009, “Computational Genomics and Epigenomics in Drosophila and Human,” Lawrence Berkeley National Labs, Berkeley, CA.

Feb. 2009, “Bayesian Gene-tree Reconstruction and Learning in Phylogenomics,” Stochastic Models of Sequence Evolution, Mathematical Biosciences Institute, Columbus, OH.

Mar. 2009, “Computational Genomics and Epigenomics of the Human Genome,” University of Chicago, Invited Lecture, Chicago, IL.

Mar. 2009, “Integrative Analysis of Drosophila epigenomic and genomic functional datasets,” modENCODE workshop invited lecture, Drosophila 2009, Chicago, IL.

Mar. 2009, “Challenges in Comparative Genomics,” Center for Algorithmic and Systems Biology Invited Lecture, University of California in San Diego, La Jolla, CA.

Mar. 2009, “Computational Epigenomics in Drosophila and Human,” Center for Molecular Genetics, UC San Diego, La Jolla, CA.

Mar. 2009, “Foundations and Frontiers in Computational Biology,” RECOMB Satellite on BioInformatics Education Invited Keynote Lecture, La Jolla, CA.

Mar. 2009, “Integrative analysis of chromatin, expression, and motif datasets,” ENCODE/modENCODE meeting invited lecture, Bethesda, DC.

Apr. 2009, “Computational methods for interpreting the human genome,” National Technological University of Athens, student visit, Invited lecture, Cambridge, MA.

May 2009, “Computational and comparative genomics and epigenomics in the human,” HST.500, Guest Lecture, Cambridge, MA.

Jun. 2009, “Regulatory Genomics and Epigenomics in Drosophila and Human,” Boston University Bioinformatics Symposium, Invited speaker, Boston, MA.

Jun. 2009, “Defining tissue- and stage-specific gene expression,” Mathematical Challenges in Developmental Biology conference, Mathematical Biosciences Institute, Invited speaker, Columbus, OH.

Jun. 2009, “Comparative genomics and phylogenomics of fly and fungal species,” Microsoft Research, Invited speaker, Cambridge, MA

Jul. 2009, “Bayesian reconstruction of thousands of gene trees in the phylogenomics setting,” Invited speaker, Quest for Orthologs, Hinxton UK.

Jul. 2009, “Chromatin and evolutionary signatures in mammalian genomes,” Jackson Laboratory 50th Annual course on Mammalian Genomics, Invited keynote, Bar Harbor, MN.

Sep. 2009, “Challenges and opportunities in computational genomics,” Dutch Math/CS students visit; CSBi students; EECS undergrad students; Invited lectures, Cambridge MA.

Sep. 2009, “Integrative analysis of large-scale epigenomics datasets,” NIH Roadmap on Epigenomics planning meeting, Houston, TX.

Sep. 2009, “Inference and dynamics of regulatory networks in *Drosophila melanogaster*,” 7th Annual CIFAR Genetic Networks meeting, Invited speaker, Princeton, NJ.

Oct. 2009, “Large-scale genomic and epigenomic analyses and human health,” Hellenic Biomedical Association First Annual Meeting, Invited speaker, Brookline, MA.

Oct. 2009, “Computational Regulatory Genomics and Epigenomics in human and fly,” CSHL Genome Informatics, Invited speaker, Cold Spring Harbor, NY.

Oct. 2009, “Regulatory Genomics and Epigenomics in the human genome,” Keck Institute Annual Research Conference, Invited Keynote, Houston, TX.

Nov. 2009, “Integrative analysis of genomic and epigenomic datasets,” NIH Epigenome Roadmap steering committee meeting, Invited Speaker, Washington, DC.

Jan. 2010, “Computational comparative genomics and epigenomics for interpreting the human genome,” EECS and Human Health seminar invited speaker, MIT, Cambridge MA.

Jan. 2010, “Efficient learning of chromatin mark combinations in many cell types across the whole genome,” Joint Genome Institute workshop, invited speaker, Walnut Creet, CA

Feb. 2010, “Interpreting the human genome using genomic, epigenomic, and evolutionary signatures,” U. of Washington Genome Sciences Seminars, invited speaker, Seattle, WA.

Feb. 2010, “Regulatory genomics and epigenomics in fly and human,” UCSD Frontiers in Bioinformatics and Systems Biology Colloquium, Invited Speaker, San Diego, CA.

Feb. 2010, “Regulatory genomics and epigenomics in fly and human,” Stanford Bio-X program, invited seminar speaker, Palo Alto, CA.

Feb. 2010, “Interpreting the human genome using genomic, epigenomic and evolutionary signatures,” Center for Biomolecular Science & Engineering, UC Santa Cruz, Santa Cruz, CA.

Feb. 2010, “Integrative analysis of *Drosophila* functional genomics datasets,” Stowers Institute for Medical Research, Stowers Lecture Series invited speaker, Kansas City, MO.

Feb. 2010, “Interpreting the human genome using genomic, epigenomic, and evolutionary signatures,” Washington University, Saint Louis, MI.

Mar. 2010, “Robust metrics for assessing reproducibility, resolution, and coverage in fly and worm modENCODE”, mod/ENCODE planning meeting, Washington, DC.

Mar. 2010, “Integrative analysis of Drosophila modENCODE datasets”, mod/ENCODE planning meeting, Washington, DC.

Mar. 2010, “Comparative analysis of 29 mammals for systematic annotation of the human genome”, mod/ENCODE planning meeting, Washington, DC.

Mar. 2010, “Interpreting the human genome using comparative, chromatin, and sequence signatures,” Biomedical sciences symposium, Athens Academy, Athens, Greece.

Mar. 2010, “Challenges and opportunities in computational biology and human genomics,” National Technical University of Athens (NTUA) student visit, Cambridge, MA.

Mar. 2010, “Regulatory Genomics and Epigenomics in fly and human,” CSHL Systems Biology: Global Regulation of Gene Expression, Invited Speaker, Cold Spring Harbor, NY.

Apr. 2010, “Phylogenomics and Epigenomics of human and fly,” New England Biostatistics Symposium invited speaker, Cambridge MA.

Apr. 2010, “Integrative analysis of genomic and epigenomic datasets in the fly,” Annual *Drosophila* Conference, modENCODE workshop, invited speaker, Cambridge MA.

Apr. 2010, “Interpreting the human genome using evolutionary and chromatin signatures,” NHGRI planning meeting on informatics and analysis, invited speaker, Washington, DC.

May 2010, “Comparative genomics and epigenomics in the human genome,” Evolutionary biology seminar series, invited speaker, Stanford University, Stanford, CA.

May 2010, “Integrative analysis of epigenomics datasets across multiple human cell types,” NIH Roadmap on Epigenomics steering committee meeting, San Diego, CA.

May 2010, “Bayesian gene tree reconstruction and reconciliation,” Bertinoro Computational Biology 2010, invited speaker, Bertinoro, Italy.

Jun. 2010, “Computational Regulatory Genomics and Epigenomics in human and fly,” Hebrew University in Jerusalem, Invited Lecture, Jerusalem, Israel.

Weizmann Institute, Invited Lecture, Tel Aviv, Israel.
Safra Bioinformatics Program Distinguished Lecture, Tel Aviv University, Israel.

Jul. 2010, “Dynamics of chromatin states in multiple human cell lines,” ISMB Epigenomics Special Session, Boston MA.

Jul. 2010, “Regulatory genomics and epigenomics in fly and human,” ISMB Regulatory Genomics Special Interest Group, invited keynote speaker, Boston MA.

Jul. 2010, “Integrative analysis of the Drosophila modENCODE (model organism ENCyclopedia of DNA Elements) project,” EMBO meeting on the molecular and developmental biology of Drosophila, invitation only meeting, Kolymbari, Crete, Greece.

Jul. 2010, “Epigenomics and disease: New tools for interpreting molecular phenotypes,” Scholars in Clinical Science, Harvard Medical School, invited lecture, Brookline, MA.

Sep. 2010, “Transcriptional regulatory networks in Drosophila modENCODE,” EMBO meeting invited speaker, Barcelona, Spain.

Sep. 2010, “Evolutionary, Chromatin, and Activity Signatures for Interpreting the Human Genome,” Keynote speaker, Swiss Institute of Bioinformatics PhD Network, Bern, Switzerland.

Sep. 2010, “Inferring regulatory codes from large-scale data integration in human and fly,” US National Academy of Science / Chinese Academy of Science ‘Frontiers of Science’ meeting, invited speaker, Irvine, CA.

Oct. 2010, “Interpreting the non-coding human genome using chromatin and regulator dynamics in multiple cell types,” IEEE Computer and Engineering in Medicine and Biology Societies, MIT biological engineering and biomedical engineering student group (BE-BMES), and GBC/ACM, Cambridge, MA.

Oct. 2010, “Regulatory network inference in human and fly using chromatin states and activity profiles,” Systems Biology: Molecular Networks, invited speaker, IRCM, Montreal, Canada.

Oct. 2010, “Disease epigenomics: Interpreting disease-associated non-coding variants”, Medical Genome Sequencing: Understanding the Genomes of Disease, invited speaker, CRG Barcelona, Spain.

Nov. 2010, “Integration of genomics and epigenomics datasets in health and disease”, NIH Epigenomics Roadmap Planning meeting, invited speaker, Bethesda, DC.

Nov. 2010, “Comparative Genomics: Biological Signal Discovery and Genome Evolution”, Guest Lecture, BU bioinformatics.

Dec. 2010, “Epigenomics: Chromatin state discovery and implications for disease”, Guest Lecture, BU Bioinformatics, Boston, MA.

Dec. 2010, “Evolutionary, Chromatin, and Activity Signatures for interpreting the human genome”, Annual Invited Lecture for student symposium, Harvard School of Public Health, Boston, MA.

Dec. 2010, “Genomic signatures for interpreting disease variants”, Computational Aspects of Biological Information 2010 Symposium, Microsoft Research, Cambirdge MA.

Feb. 2011, “Epigenomics of multiple human cell types and implications for disease datasets”, Quantitative Biology and Bioinformatics in Modern Medicine, Dublin, Ireland.

Feb. 2011, “Interpreting human disease with regulatory genomics and epigenomics”, Vertex Pharmaceuticals, Cambridge MA.

March 2011, “Integrative genomics and epigenomics”, Computer Science department, University of Athens, Invited speaker, Athens, Greece.

March 2011, “Integrative analysis of functional genomics datasets”, Computational and Systems Biology Initiative, Invited speaker, Cambridge MA.

March 2011, “Integrative analysis of the Drosophila modENCODE project”, International Drosophila Research conference, Invited speaker, San Diego CA.

Apr. 2011, “Computational Challenges and Opportunities in Genomics,” MIT Learning and Information Decision Systems, Invited seminar speaker, Cambridge MA.

Apr. 2011, “Epigenomics of human disease,” Tongji University Symposium, Invited speaker, Shanghai China.

Apr. 2011, “High-throughput determination of gene regulatory relationships and gene function prediction,” Cold Spring Harbor Lab Asia, High-Throughput Biology Meeting, Invited speaker, Suzhou, China.

May 2011, “Systems Biology of Developmental networks by Large-scale data integration in Drosophila modENCODE,” Systems Biology meeting, Invited keynote speaker, New York Academy of Science.

May. 2011, “Comparative genomics of 29 mammals,” ENCODE consortium project meeting, Invited speaker, Washington DC.

June 2011, “Regulatory network prediction by large-scale data integration in ENCODE and modENCODE”, Computational Systems Biology meeting, Invited speaker, Zurich, Switzerland.

June 2011, “Epigenomics of human health and disease”, European Molecular Biology Laboratory (EMBL), Invited speaker, Heidelberg, Germany.

June 2011, “Personal genomics and epigenomics of health”, Federal Drug Administration visit day, CSAIL, Invited speaker, Cambridge MA.

June 2011, “Computational challenges in personal genomics”, Computer Science retreat, Chatham MA.

July 2011, “Medical and personal genomics and epigenomics”, Dana Farber Cancer Institute retreat, Invited keynote speaker, Cambridge MA.

June 2011, “Personal and medical genomics in the age of complete genomes”, Qatar Innovation visit day, CSAIL, Invited speaker, Cambridge MA.

August 2011, “Regulatory Genomics and Epigenomics of human health and disease”, Invited special seminar speaker, Novartis, Cambridge MA.

September 2011, “Computational epigenomics of multiple human cell types and disease”, Epigenomics of Common Disease, Invited speaker, Hinxton, UK.

September 2011, “Personal genomics and epigenomics of human health and disease”, John Hopkins Computational Genomics symposium, Invited keynote speaker, Baltimore MD.

October 2011, “Thousands of loci contribute to complex human disease”, American Society of Human Genetics (ASHG) annual meeting, Montreal, Canada.

October 2011, “Interpreting human disease with large-scale functional genomics integration”, Invited keynote speaker, RECOMB Regulatory Genomics and Systems Biology meeting.

October 2011, “Comparative genomics across mammals and within the human population”, Cold Spring Harbor Laboratory Genome Informatics, Invited speaker, Cold Spring Harbor, NY.

November 2011, “Integrating comparative, regulatory, and epigenomics datasets for interpreting disease variants”, Invited keynote speaker, Salk Institute, La Jolla CA.

November 2011, “Evidence of purifying selection within the human population for an additional 4% of the human genome”, Cold Spring Harbor Laboratory Genome Informatics, Invited speaker, Cold Spring Harbor, NY.

December 2011, “Computational regulatory genomics and epigenomics”, Bioinformatics and Data Analysis, Invited Keynote speaker, Montreal, Canada.

December 2011, “Computational regulatory genomics and epigenomics”, Dana Farber Center for Cancer Computational Biology, Invited seminar speaker, Montreal, Canada.

December 2011, “Systems disease epigenomics reveal 1000s of functional variants associated with complex traits”, Dana Farber Center for Functional Cancer Epigenetics, Invited speaker, Cambridge MA.

January 2012, “Integrative analysis of Epigenomics Datasets”, National Institutes of Environmental Health Sciences, Epigenomics and Disease Symposium, Invited speaker.

February 2012, “Systems Epigenomics reveals thousands of functional SNPs associated with complex disease”, Keystone Meeting on Complex Traits, Invited speaker, Breckenridge, CO.

February 2012, “Comparative genomics of mammals and humans: selective signatures and evolutionary change”, Charles Darwin Symposium, Invited speaker, Harvard University, Cambridge MA.

March 2012, “Functional population genomics: GWAS interpretation and recent human selection”, Genomic Disorders meeting, Invited speaker, Hinxton, UK.

March 2012, “Selective turnover of active regulatory regions in the human lineage”, CHS meeting on Systems Biology: Global Regulation of Gene Expression, Selected speaker, Cold Spring Harbor, NY.

March 2012, “Regulatory genomics and epigenomics of multiple cell lines and complex disease”, Boston University Systems Biology seminar series, Invited speaker, Boston MA.

April 2012, “Interpreting the human genome in health and disease”, National Technical University of Athens, MIT visit day, Invited speaker, Cambridge MA.

April 2012, “Systems Epigenomics of Complex Disease”, Systems and Multiscale Biology, BioIT World Conference and Expo, Invited speaker, Boston MA.

April 2012, “Interpreting complex genotype-phenotype associations”, Genotype-Expression Analysis meeting (GTEx), Broad Institute, Invited speaker and participant, Cambridge MA

April 2012, “Epigenomic Views of Complex Disease Associations Reveal thousands of Regulatory SNPs”, Dana Farber Center for Cancer Computational Biology, Invited speaker, Boston MA.

April 2012, “Systems Epigenomics of Complex Disease”, Network Science Course, Northeastern Center for Complex Networks, Invited speaker, Boston MA.

May 2012, “Prioritizing genes and loci for regulatory variants associated with human disease”, NIH-Pharma-Academia Target Validation planning meeting, Invited speaker and participant, Cambridge MA.

May 2012, “Thousands of regulatory variants contribute to complex disease in Alzheimer's and type 1 diabetes”, Biology of Genomes Meeting, Selected speaker, Cold Spring Harbor, NY.

May 2012, “Chromatin and transcription factor binding dynamics and interplay”, Epigenomics Roadmap Steering Committee Meeting, Invited speaker, Bethesda, DC.

May 2012, “Prioritizing genes and loci for regulatory variants associated with human disease”, NIH-Pharma-Academia Target Validation planning meeting, Invited speaker and participant, Cambridge MA.

May 2012, “Thousands of regulatory variants contribute to complex disease in Alzheimer's and type 1 diabetes”, Biology of Genomes Meeting, Selected speaker, Cold Spring Harbor, NY.

May 2012, “Chromatin and transcription factor binding dynamics and interplay”, Epigenomics Roadmap Steering Committee Meeting, Invited speaker, Bethesda, DC.

June 2012, “Integrative analysis of Drosophila and human epigenomics and regulatory genomics datasets”, modENCODE symposium, Invited speaker, Washington, DC.

June 2012, “Systems Epigenomics of complex disease reveals thousands of regulatory variants”, Pathways Networks and Systems Medicine, Invited speaker, Rhodes, Greece.

June 2012, “Chromatin state and motifs in TF binding dynamics and disease”, Epigenetics meets Systems Biology (EpigeneSys), Invited speaker, Weizmann Institute, Rehovot, Israel.

June 2012, “Genomic and Epigenomic signatures for interpreting complex disease”, Israel Bioinformatics Meeting, Keynote speaker, Hebrew University, Jerusalem, Israel.

July 2012, “Interpreting human disease associations using comparative genomic and epigenomic signatures”, ISMB 2012, Highlights talk, Los Angeles, CA.

September 2012, “Genotype-methylation-disease in 750 Alzheimer’s patients and controls”, Nature Genetics/Wellcome Trust meeting on Genomics of Common Disease, Potomac, DC.

November 2012, “Selective pressures in 29 mammals and within the human population”, American Society of Human Genetics (ASHG), Invited Session speaker, San Francisco, CA.

November 2012, “Systems genomics approaches for interpreting human disease”, Stanford University, Invited seminar speaker, Palo Alto, CA.

November 2012, “Systems genomics approaches for interpreting human disease”, UC Berkeley, Invited seminar speaker, Berkeley, CA.

December 2012, “Regulator and chromatin dynamics reveal interplay in target specification”, Max Planck Epigenetics, Invited keynote speaker, Freiburg, Germany.

Theses Supervised by Manolis Kellis

	Total	Completed	In Progress
Bachelor's	1	1	0
Master's	2	2	0
MEng	6	3	3
Engineer's	0	0	0
Doctoral			
As Supervisor	10	3	7
As Reader	11	6	5

Bachelor's Theses

Parts, Leopold "Comparative RNA gene identification in the fly genome," June 2006.
Led to co-authored papers in Genome Research and Nature.

Master's Theses

Rasmussen, Matthew "Probabilistic Framework for Genome-wide Phylogeny and Orthology Determination," June 2006.
Received William A. Martin Thesis Award for M.S. thesis in the field of Computer Science at MIT.
Led to first-authored paper in Genome Research.

Deoras, Ameya "Gene Identification using Phylogenetic Metrics with Conditional Random Fields," June 2007.
Led to co-authored paper in PLoS Computational Biology.

MEng Theses

Lin, Michael F. "Comparative Gene Identification," June 2006.
Received Charles and Jennifer Johnson Thesis Award for outstanding M.Eng. thesis in the field of Computer Science at MIT.
Led to first-authored papers in Genome Research and PLoS Computational Biology.

Grochow, Joshua. "On the Structure and Evolution of Protein Interaction Networks," August 2006.
Received Charles and Jennifer Johnson Thesis Award for outstanding M.Eng. thesis in the field of Computer Science.
Led to first-authored paper in RECOMB 2007.

Fujiwara, Guilherme Issao. "De novo discovery of evolutionary signatures," June 2008.

Yen, Angela. "Epigenomics of olfaction and disease," June 2011.
Led to co-authored paper in review in Cell.

Brown, Wes. "Phylogenomics of human populations and bacterial species," June 2011.

Ayuso, Anna. "Learning two-dimensional embryo gene expression primitives using a semi-supervised learning approach," June 2011.

Holmes, Benjamin, “Interplay of RNA and chromatin in human gene regulation,” expected August 2013.

Goel, Rushil, “Network inference in high-throughput datasets”, expected May 2016.

Engineer’s Theses

None

Doctoral Theses, Supervisor

Rasmussen, Matthew, “Methods and Analysis of Genome-scale Gene Family Evolution Across Multiple Species,” August 2010.
Led to first-author paper in MBE and three co-authored papers in Nature.

Kheradpour, Pouya, “Regulatory network reconstruction in the human genome for transcription factors and microRNAs,” expected May 2011.
Led to first-author paper in Genome Research, four co-authored papers in Nature, and additional co-authored papers in Genome Research, Science, PLoS Genetics, G&D.

Lin, Michael F. “Probabilistic models for functional genome analysis,” expected May 2011.
Led to first-author papers in Genome Research, four co-authored papers in Nature, two in Genome Research, and Science.

Candeias, Rogerio, “Dynamics and structure of regulatory networks in fly,” expected May 2011.
Led to co-authored papers in review in Nature, Genome Research, and Genome Biology.

Jessica Wu, “Phylogenomic modeling of beyond protein-coding genes,” expected June 2012.
Led to first-author paper in submission to PLoS Computational Biology.

Sealfon, Rachel, “Machine Learning Methods for Regulatory Region Identification in Drosophila Genomes,” expected May 2013.
Led to co-authored papers in Nature and *Science.*

Altshuler, Robert, “Integrative analysis of functional and population genomic data,” expected June 2013.

Frogner, Charlie, “Translating genomic information into spatial developmental constraints,” jointly with Tommy Poggio, expected May 2014.

Yen, Angela, “Role of chromatin modifications in differentiation and disease”.

Doctoral Theses, Reader

Breva, Luis, “DNA binding economies,” EECS, completed August 2007 (Supervised by T. Jaakkola)

Weimann, Oren, “Accelerating Dynamic Programming,” EECS, completed June 2009 (Supervised by E. Demaine)

Missiuro, Patrycja, “Predicting Genetic Interactions in *C. elegans* using Machine Learning,” EECS, completed September 2009 (Supervised by T. Jaakkola)

David, Lawrence, “Novel Phylogenetic Approaches to Problems in Microbial Genomics,” CSBi, completed August 2010 (Supervised by E. Alm)

Shapiro, Jesse, “Genomic signatures of sex, selection and speciation in the microbial world,” Civil/Environ. Eng., completed August 2010, (Supervised by E. Alm)

Tsankov, Alexander, “Evolution of Nucleosome Positioning and Gene Regulation in Yeasts: a Genomic and Computational Approach,” EECS, completed August 2010 (Supervised by A. Regev)

Koche, Richard, “Sequence Determinants of an Epigenetic Memory Element in Mammalian Development,” HST, expected June 2011 (Supervised by B. Bernstein)

Abhoukhalil, Anton, “Computational analysis of cis regulatory codes in *Drosophila,*”Aero Astro, expected June 2012 (Supervised by M. Bulyk)

Timberlake, Sonia, “Selection in the environment and phenotype in the lab: Quantifying contributions of gene regulation to organismal fitness and adaptive change,” CSBi, expected June 2012 (Supervised by E. Alm)

Sun, James X, “Population genomics estimates of human mutation rates using satellite repeats,” HST, expected June 2013 (Supervised by D. Reich)

Barnett, John, “Drosophila development gene expression modeling using Dirichlet processes,” EECS, on leave (Supervised by T. Jaakkola)

Postdoctoral Associates and Fellows Supervised by Manolis Kellis

Current Postdocs

Name	Dates of Appointment	PhD Granting Institution	Current Position
Goff, Loyal	Oct. 2008 - present	Rutgers University	Postdoc Fellow
Washietl, Stefan	Aug. 2009 - present	Vienna University	Postdoc Fellow
Marbach, Daniel	Sep. 2009 - present	EPFL Lauzanne	Postdoc Fellow
Ward, Luke	June 2010 - present	Columbia University	Postdoc Assoc.
Eaton, Matt	June 2011 - present	Duke University	Postdoc Assoc.
Quon, Gerald	Feb. 2012 - present	U. Toronto	Postdoc Assoc.
Wang, Jianrong	Sep. 2012 - present	Georgia Tech.	Postdoc Assoc.
Meuleman, Wouter	Sep. 2012 - present	NKI Netherlands	Postdoc Assoc.
Pfenning, Andreas	Sep. 2012 - present	Duke University	Postdoc Assoc.
Jaillon, Olivier	Oct. 2012 - present	Genoscope, France	Research Scientist
Waterhouse, Robert	Feb. 2013 - present	U. Geneva	Marie Curie Fellow
Cowper-Sal.lari Richard	Feb. 2013 - present	Dartmouth College	Postdoc Assoc.
Kim, Ah-Ram	Mar. 2013 - present	U. Chicago	Postdoc Assoc.
Novoa, Eva Maria	Starting May 2013	IRB Barcelona	Marie Curie Fellow

Previous Postdocs

Name	Current Title	Current Employer
Xie, Xiaohui	Assoc. Prof.	UC. Irvine
Stark, Alexander	Asst. Prof.	MPI, Vienna
Diallo, Abdoulaye	Asst. Prof.	UQAM, Montreal
Ossowsky, Stephan	Asst. Prof	CRG, Barcelona
Bristow, Chris	Institute Res. Sci.	MD Anderson Cancer Center
DiStefano, Luisa	Asst. Prof.	CNRS
Roy, Sushmita	Asst. Prof.	UWisc. Madison
Ernst, Jason	Asst. Prof.	UC. Los Angeles
Bansal, Mukul	Asst. Prof.	U. Conn (offer accepted)
Hendrix, David	Asst. Prof.	Oregon State (offer accepted)
Rasmussen, Matt	Asst. Prof.	U. Souther California (offer not yet accepted)
Kundaje, Anshul	Asst. Prof.	Stanford (offer not yet accepted)

Manuscripts in preparation:

1. Washietl, Kellis*, Garber*, “Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals”. Science, in submission.
2. Rouskin, McKeon, Washietl, Manolis Kellis, Weissman. “Genome-wide RNA probing reveals mRNA structures are far less common in vivo than in vitro”, Nature, in submission.
3. De Jager, Srivastava, Eaton, Keenan, Ernst, McCabe, Tang, Chibnik, Brodeur, Gabriel, Chai, Younkin, Younkin, Zou, Szyf, Epstein, Schneider, Yu, Ertekin-Taner, Bernstein, Meissner, Kellis, Bennett, “Alzheimer’s disease pathology is associated with early alterations in DNA methylation of the human cortex”, Nature Genetics, in submission.
4. Boyle, Araya, Cheng, Xie, Kheradpour, Kundaje, Cayting, Niu, Yan, Kawli, Ma, Cheng, Wang, Slattery, Spokony, Gardner, Kasper, Janette, Rozowsky, Lixia, Jiang, Brdlik, Feingold, Good, Pazin, Reinke6, Waterston, Gerstein, White, Kellis, Snyder. “Comparative analysis of regulatory information and circuits across diverse species”. Nature, in submission.
5. Ho, Liu, Jung, Alver, Lee, Ikegami, Sohn, Minoda, Tolstorukov, Appert, Parker, Gu, Kundaje, Riddle, Bishop, Egelhofer, Hu, Alekseyenko, Rechtsteiner, Asker, Belsky, Bowman, Chen, Chen, Day, Dong, Dose, Duan, Epstein, Ercan, Feingold, Garrigues, Gehlenborg, Good, Haseley, He, Herrmann, Hoffman, Jeffers, Kharchenko, Kolasinska-Zwierz, Kotwaliwale, Kumar, Langley, Larschan, Latorre, Libbrecht, Lin, Park, Pazin, Pham, Plachetka, Qin, Schwartz, Shoresh, Stempor, Vielle, Wang, Whittle, Xue, Kingston, Kim, Bernstein, Dernburg, Pirrotta, Kuroda, Noble, Tullius, Kellis, MacAlpine, Strome, Elgin, Ahringer, Liu, Karpen, Lieb, Park, “Comparative analysis of metazoan chromatin architecture”, Nature, in review.
6. Bansal, Wu, Alm, Kellis, “Reliable and Accurate Gene Tree Reconstruction for Deciphering Microbial Evolution”, in preparation.
7. Hendrix, Goff, Hart, Kellis, “Genomic evidence for role of small RNAs in the recruiting and reprogramming of chromatin marks in human embryonic stem cells,” 6 pages, in preparation.
8. Lin, Kellis, “Bayesian inference of lineage-specific selection on synonymous sites using phylogenetic codon models,” 8 pages, in preparation. **
9. Sarkar, Ward, Kellis, “Genome-wide enrichments for regulatory regions across thousands of unlinked disease-associated variants”, 8 pages, in preparation. **
10. Altshuler, Kellis, “Allelic activity in regulator binding, chromatin state, and gene expression associated with human polymorphisms”, 8 pages, in preparation. **
11. Feizi*, Quon*, Mendoza, Medard, Kellis, “Spectral integration of regulatory networks in human, fly, worm reveals conservation of regulatory pathways.”, in preparation.
12. Quon*, Feizi*, Vartak, Kellis, “Disease gene associations across model organisms and human disease”, in preparation.
13. Wu, Mendoza, Bansal, Kellis, “Phylogenetic identification and functional validation of orthologous genes across human, mouse, fly, worm, yeast”, in preparation.

14. Kheradpour, Kellis, “Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments”, in preparation.
15. Kundaje, Kumar, Ernst, Kellis, “Joint chromatin state analysis across fly, worm, human”, in preparation.
16. Ernst, Kellis, “Imputation of chromatin datasets in densely phenotyped human epigenomes”, in preparation.
17. Wang, Kellis, “Prediction of gene expression patterns from short-range and long-range chromatin elements”, in preparation.
18. Eaton, Bernstein, Bennett, De Jager, Kellis, “Genotype and methylation variants associated with Alzheimer’s disease”, in preparation.

Books in preparation:

19. Kellis, “Computational Biology: Genomes, Networks, Evolution (Book),” Cambridge University Press, in preparation.

Teaching Evaluations of Manolis Kellis

Term	Course #	Course Title	Role	Course Type	# Students Registered	# Survey Responses	Instructor's Evaluation	Course Evaluation	Scale
IAP05	6.092 (new)	Bioinformatics and proteomics: an engineering-based approach	Lectures, co-in charge (with M. Ramoni and G. Alterovitz)	Lecture	7	7	6.4	6.0	7.0
ST05	6.096 (new)	Algorithms for Computational Biology	Lectures, in charge	Lecture	25	8	4.5	4.5	7.0
FT05	6.095 / 6.895 (new)	Computational Biology: Genomes, Networks, Evolution	Lectures, co-in charge (with P. Indyk)	Lecture	21	12	6.3	6.1	7.0
ST06	6.046	Introduction to Algorithms	Lectures, co-in charge (with R. Rivest)	Lecture	84	30	5.6	5.8	7.0
FT06	6.085 / 6.895	Computational Biology: Genomes, Networks, Evolution	Lectures, co-in charge (with P. Indyk)	Lecture	25	14	6.5	5.8	7.0
ST07	6.046	Introduction to Algorithms	Lectures, co-in charge (with S. Devadas)	Lecture	118	65	5.5	5.3	7.0
FT07	6.047 / 6.878	Computational Biology: Genomes, Networks, Evolution	Lectures, co-in charge (with J. Galagan)	Lecture	33	16	6.0	5.6	7.0
ST08	6.046	Introduction to Algorithms	Lectures, co-in charge (with P. Indyk)	Lecture	80	44	5.8	5.6	7.0
FT08	6.047 / 6.878	Computational Biology: Genomes, Networks, Evolution	Lectures, co-in charge (with J. Galagan)	Lecture	23	13	6.4	6.2	7.0
ST09	New 6.046	Design and Analysis of Algorithms	Lecture, co-in charge (with M. van Dijk)	Lecture	72	45	5.9	5.4	7.0
FT09	6.047/ 6.878/ HST.507	Computational Biology: Genomes, Networks, Evolution	Lectures, co-in charge (with J. Galagan)	Lecture	24	12	5.8	5.4	7.0
ST10	Junior Faculty Teaching Leave								
FT10	6.047/ 6.878/ HST.507	Computational Biology: Genomes, Networks, Evolution	Lectures, in charge	Lecture	36	16	5.5	5.1	7.0

ST11	6.006	Introduction to Algorithms	Lectures, co-in charge (with E. Demaine and P. Indyk)	Lecture	163	55	5.2	3.9	7.0
FT11	6.047/ 6.878/ HST.507	Computational Biology: Genomes, Networks, Evolution	Lectures, in charge	Lecture	26	15	5.9	6.0	7.0
ST12	6.881	Computational Personal Genomics (New)	Lectures, in charge	Lecture	8	5	6.6	6.6	7.0
FT12	6.047/ 6.878/ HST.507	Computational Biology: Genomes, Networks, Evolution	Lectures, in charge	Lecture	32	17	5.9	5.7	7.0