Computational Biology and Bioinformatics

Rotation Guide 2019-2020

Julien Berro .................................................................2
Cynthia Brandt .............................................................2
Kei-Hoi Cheung .........................................................3
Chris Cotsapas ............................................................3
Mark Gerstein .............................................................3
Antonio Giraldez .........................................................4
Monika Jadi .................................................................5
Samah Fodeh Jarad .....................................................6
William Jorgensen ......................................................7
Naftali Kaminski .......................................................8
Steven Kleinstein .......................................................8
Smita Krishnaswamy ..................................................9
Morgan Levine ..........................................................10
Zachary Levine ........................................................11
Bluma Lesch .............................................................11
Haifan Lin .................................................................12
Elias Lolis .................................................................13
Jun Lu .........................................................................15
Robert McDougal .....................................................16
James Noonan ..........................................................16
Corey O'Hern ............................................................18
Lajos Pusztai .............................................................21
Anna Marie Pyle .........................................................21
Hemant D. Tagare .....................................................22
Andrew Taylor ..........................................................22
David van Dijk ..........................................................23
Anita Wang ...............................................................24
Hongyu Zhao ............................................................24
Steven Zucker ...........................................................25

Note: This is not a complete listing of all the faculty who are currently involved with the CBB program
The Berro lab is interested in unraveling the molecular mechanisms of force production and force sensing in vivo. The lab currently focuses on understanding the biochemical and biophysical regulations of clathrin-mediated endocytosis. Specifically, the lab aims to understand how the actin cytoskeleton generates forces to deform flat plasma membranes into small vesicles and, conversely, how the endocytic machinery senses membrane tension and adapts to it.

We are a collaborative cross-disciplinary team that combines experimental approaches (e.g., cell biology, quantitative microscopy) and computational tools (e.g., image analysis, mathematical modeling, machine learning). Depending on your interests, we can tailor the rotation project to be experimental, computational, or a mix of both. By joining the lab you will be exposed to all aspects of the cross-disciplinary approaches we develop and use. Below are examples of rotation projects.

Computational rotation projects:
- Development of a new data alignment algorithm using machine learning
- Simulation of actin filament assembly during clathrin-mediated endocytosis using Monte-Carlo simulations
- Mathematical modeling of subunits residence times at the tip of a growing filament using theory and Monte-Carlo simulations

Please do not hesitate to contact me to discuss your research interests and possible projects.

The Brandt lab focuses on the creative use of informatics tools on clinical and electronic health record data to inform health services research, and on the development and use of informatics systems for clinical and research studies. Areas of research are broad, with current projects focusing on women veteran’s health, and pain and complementary and integrative health.
Kei-Hoi Cheung
kei.cheung@yale.edu | 203-737-5783

Lab location: 464 Congress Avenue
Lab meeting: No specific schedule; typically meets with students on Thursdays or Fridays

Rotations preferred in the spring

Prof. Cheung’s research interests include the use of new database technologies (e.g., NoSQL), ontologies and data standards (e.g., FAIR data) to enable semantic data integration in the systems biology context. In addition, Prof. Cheung is keen on combining semantic technologies with natural language processing (NLP) to facilitate clinical text mining and machine learning. His research has spawned a range of applications across multiple domains including emergency medicine, systems vaccinology, genomics medicine, and Veteran healthcare research. Dr. Cheung’s research collaboration spans Yale, VA, and a number of consortia/communities.

Chris Cotsapas
chris.cotsapas@yale.edu | 203-737-2896

Lab location: 353H, 300 George Street
Lab meeting: Mondays 3:00 p.m., 353M, 300 George Street

Rotations available anytime.

Available rotation projects:

1. Computational annotation of autoimmune disease loci. We have identified a set of enhancers that drive risk for human autoimmune diseases. Many of these diseases show marked differences in incidence and prevalence between men and women, but we do not know why. To explore this connection, we would like to know if sex steroid hormone receptors (the estrogen and androgen receptors, in particular) bind to these enhancers in relevant immune cells. This project would suit CBB students wanting to explore methods applications to real-world problems, Immunology students interested in learning how computational biology can be applied to problems in immunology, and MCGD students interested in human complex disease genetics.

2. Expression of sex steroid hormone receptors in immune cell populations. Immune function and disease vary greatly between men and women, but we have little understanding how sex hormones affect immune cell function. We are curious about whether the receptors for estrogen and testosterone – both of which are transcription factors – are expressed in immune cells, and if these can be induced through stimulus. Your project will be to use PCR and single-cell RNA sequencing to determine if the receptors are expressed and if there are functional differences between the cells that do and don’t express them. This project would suit students in Immunology interested in human immunology, and MCGD students interested in signaling and hormone biology.

Mark Gerstein
mark.gerstein@yale.edu | 203-432-6105

Lab location: Bass 432
Lab meeting: Weekly at different times

Rotations available any time.
Rotation projects on Computational Biology & Biomedical Data Science related to topics at http://info.gersteinlab.org/General_Information_for_Potential_Graduate_Students

For relevant papers see http://papers.gersteinlab.org/

Antonio Giraldez
antonio.giraldez@yale.edu | www.giraldezlab.org

Stem Cell Center (associate member)

Lab location: SHM I room 147, 333 Cedar Street.
Lab meeting: Friday, 2:00 p.m. SHM I room 143.

Rotations available anytime. Please email Antonio if you are interested.

The problem: The main goal of my laboratory is to understand the cellular and molecular signals that initiate embryonic development to uncover universal principles that direct the development of a new life. A universal step in all animals is the maternal to zygotic transition, whereby the transcriptionally silent egg activates the new zygotic program and removes the old maternal program. This central step in animal development can be considered the beginning of life from a transcriptional standpoint, whereby subsequent developmental decisions will depend on the correct activation of the zygotic program and regulation of the previous maternal program.

The questions: Our laboratory aims to understand how does the vertebrate embryo activate the silent zygotic genome to jump start development, how does it regulate the previous developmental program and how do the building blocks in the genome, coding and non-coding elements, orchestrate these processes?

The approaches: We have a multidisciplinary infrastructure (wet/dry) that allows us to combine genomics, embryology, biochemistry and computational biology to leverage the powerful genetics in zebrafish to understand vertebrate development.

The potential projects:

Project 1: How is the genome activated. We have recently identified nanog, oct4 and soxB1 as three factors required to activate the zygotic genome. In this project, we aim to gain mechanistic insights on how these factors establish a competent genome. We will use biochemical approaches (pull down and mass spectrometry) to identify the factors that are recruited by nanog, oct4 and soxB1, and genetic tools (crispr-cas9 loss of function, Chip-seq) to investigate their function in genome activation and development.

Project 2: Computational/experimental analysis of translation regulation. Using ribosome footprinting, we are for the first time able to study how translation is regulated genome wide. This computational project will investigate how translation efficiency changes across development, with the goal of defining co-regulated transcripts and identifying common sequence and structural elements that mediate regulation.

Project 3: Computational/experimental analysis of the RNA regulatory elements in the transcriptome. We are combining machine learning with experimental approaches that use high-throughput sequencing tools, to identify RNA regulatory elements in vivo. This project will investigate how the regulatory elements in the transcriptome changes across development and how it influences mRNA turn over and translation, with the goal of defining novel structural elements that regulate gene expression.

Project 4: Uncovering the function of micropeptides in development. Using ribosome footprinting, we have identified a large number of long-noncoding RNAs that encode micropeptides that are conserved and translated during embryogenesis. Using crispr/cas9 mediated mutagenesis and immunoprecipitation in vivo, we will
investigate the function of these micropeptides in development, with the goal of identifying new signaling molecules.

**Project 5:** Genetic screening to define the factors that initiate vertebrate development. We have identified a set of chromatin remodelers, RNA binding proteins, transcription factors and genes of unknown function that are strongly translated in the early embryo. Using novel approaches developed in our lab, we are beginning a genetic screen to identify their function in vivo.

If you are potentially interested in rotating in our lab and need more information, please feel free to send Antonio an e-mail to arrange for a meeting.

**Selected Publications:**

- Jean-Denis Beaudoin, Eva Maria Novoa, Charles E Vejnar, Valeria Yartseva, Carter Takacs, Manolis Kellis, **Antonio J Giraldez.** mRNA structure dynamics identifies the embryonic RNA regulate. *Nature Structure Molecular Biology* 2018
- Moreno-Mateos MA, Fernandez JP, Rouet R, Vejnar CE, Lane MA, Mis E, Khokha MK, Doudna JA, **Giraldez AJ.** CRISPR- Cpf1 mediates efficient homology-directed repair and temperature-controlled genome editing. *Nature Commun.* 2017
- Yartseva V, Takacs CM, Vejnar CE, Lee MT, **Giraldez AJ.** RESA identifies mRNA-regulatory sequences at high resolution. *Nature Methods.* 2017
- Mishima Y, ... Giraldez AJ. Zebrafish miR-1 and miR-133 shape muscle gene expression and regulate sarcomeric actin organization. *Genes & Development.* 2009

**Monika Jadi**

monika.jadi@gmail.com | www.jadilab.org

**Lab location:** Room 32B, Suite 901, 300 George Street

**Lab meetings:** Thursdays @1:30 PM

**Rotations available any time**

Our lab is interested in discovering fundamental computational principles that are critical for flexible information processing in the brain. This is crucial for understanding brain function and uncovering the information processing abnormalities that are associated with various brain disorders. We use the framework of the canonical
computational unit in the sensory cortex: a 1x1x2mm column of cortical tissue.

The following are some of the research topics in the lab:

- **Computations in the columnar circuit:** We investigate the computational role of the structural diversity of inhibition in circuit function. We use both computer simulations and analytical treatment of computational models of the columnar circuit.

- **Information flow in neural circuits:** Newer techniques are allowing recording of dense neural datasets that are annotated in a variety of ways, such as putative cell types, laminar location and behavioral state. Our group investigates the dynamics of information flow through the columnar neural circuit as a function of behavior. Our computational approach involves methods to model the interactions and correlated activity of populations of simultaneously recorded neurons.

- **Temporal dynamics of population codes:** Population codes are neural representations of information at the level of groups of neurons (e.g. sparse coding). Our lab investigates the temporal dynamics of population coding of sensory information. We use computational approaches that include but are not limited to machine learning, to reveal the temporal dynamics of sensory information coding in a cell-type and layer specific manner.

**Recent Publications:**


**Samah Fodeh Jarad**

samah.fodeh@yale.edu | 203-737-5806

**Lab location:** Room 518, Suite 501, 300 George Street

**Lab meeting:** Fridays

*Rotations available anytime.*
The Data Mining lab led by Dr. Fodeh seeks to develop innovative methodologies to address challenging problems in healthcare and biomedical informatics. The current focus of the group is analysis of exceedingly large data sets; especially those that arise in the application areas of text mining, information retrieval and extraction, machine learning and deep learning. The emphasis of the lab is developing and applying sound algorithms that utilize and combine multiple data modalities for central tasks such as prediction and clustering of high-dimensional big data. Uncovering the latent low-dimensional structure of big data that preserves sparsity and non-negativity is essential to enhance representation, interpretation, and by extension success of algorithms. With the above goals in mind, the lab has recently been exploring difficulties associated with phenotyping different types of headaches and associated pharmacological and non-pharmacological therapies, mining social media and healthcare data for suicide and opioid overdose prediction, mining patient-generated data to characterize types of communication between patients and healthcare providers, and exploiting PubMed articles for gene molecular function prediction. The Data Mining lab has released the findings of multiple prior publications addressing those problems. The team meets every Friday and motivated students are always welcome to join the lab.

William Jorgensen  
william.jorgensen@yale.edu | 203-432-6278 | http://www.JorgensenResearch.com  
Lab location: CRB 324  
Rotations available fall and spring.

Research Description
Organic, medicinal, and computational chemistry including simulations of organic and enzymatic reactions, computer-aided drug design, and synthesis and development of therapeutic agents targeting infectious, inflammatory, and hyperproliferative diseases.

Computer-Aided Drug Discovery. Our approach features focused synthetic organic chemistry driven by state-of-the-art molecular design. The computations center on modeling protein-inhibitor complexes including docking for virtual high-throughput screening, growing of combinatorial libraries inside binding sites with BOMB, and lead-optimization guided by Monte Carlo free-energy simulations. Synthesis and optimization of the most promising leads are performed in our laboratory, and biological testing and crystallography are pursued with collaborators. The approach has allowed efficient discovery of extraordinarily potent anti-HIV, anti-inflammatory, and anti-cancer agents. Current protein targets include HIV-1 reverse transcriptase, CXCR4, human and Plasmodium MIF, and DNA methyl transferases.

Modeling of Organic and Enzymatic Reactions. The aims include elucidation of reaction mechanisms, medium effects on reaction rates, and effects of site-specific mutations on enzymatic reactions. A QM/MM approach is taken; the energetics of the reacting systems are described quantum mechanically with ab initio, DFT, or advanced semiempirical QM methods such as our PDDG/PM3 procedure. The environment including solvent molecules are represented using molecular mechanics and the sampling is normally performed with Monte Carlo statistical mechanics. Our group is also recognized as a leader in the development of force fields for water, organic and biomolecular systems and in the development of comprehensive software for molecular modeling (BOSS and MCPRO).

Reviews:  
Naftali Kaminski  
naftali.kaminski@yale.edu | 203-737-4612 | https://medicine.yale.edu/lab/kaminski/  
**Lab location:** Suite S441D, TAC  

Dr. Kaminski’s team main ambition is to uncover the mechanisms, and thus have a significant impact on the management and diagnosis, of advanced lung diseases with a specific focus on idiopathic Pulmonary Fibrosis (IPF), a chronic progressive interstitial lung disease that is currently incurable. To study these mechanisms Dr. Kaminski’s team applies systems biology approaches that incorporate a combination of traditional molecular biology methods, high-throughput genomic and epigenomic technologies, advanced bioinformatics approaches and studies in humans and animal models of disease. These studies have led to shifts in the perception of pulmonary fibrosis, the realization that aberrant activation of developmental pathways is at the core of lung fibrosis, the discovery of the role of microRNAs and large non-coding RNA in IPF, the identification and validation of novel prognostic biomarkers in the bloodstream, and discovery of novel therapeutic targets that are currently being developed. Dr. Kaminski’s team is deeply involved in the Human Cell Atlas and is leading the IPF Cell Atlas and the Normal Lung Aging Atlas.

To obtain more information about lab meetings, timing for rotations and other activities please email Dr. Kaminski. For more information, go to Dr. Kaminski’s lab website or follow Dr. Kaminski on twitter @kaminskimed.

You can view his publications here:  
https://www.ncbi.nlm.nih.gov/pubmed/?term=kaminski+n%5BAuthor%5D+NOT+kaminski+ne%5BAuthor%5D+NOT+kaminski+nj%5BAuthor%5D+NOT+kaminski+na%5BAuthor%5D

Steven Kleinstein  
steven.kleinstein@yale.edu | 203-785-6685 | https://medicine.yale.edu/lab/kleinstein  
**Lab location:** Suite 505, 300 George Street  
**Lab meeting:** Wednesdays, 2:30-4pm, Room 354, 300 George Street  

*Rotations available anytime. Please email if you are interested.*

Prof. Kleinstein is a computational immunologist with a combination of “big data” analysis and immunology domain expertise. His work seeks to understand how individual variations in immune status and function produce heterogeneity in infection and vaccination responses.

Research in the Kleinstein Lab leverages recent advances in immune profiling methods to characterize diverse states of the human immune system (in health and disease, and following infection and vaccination), such as B/T cell receptor repertoire sequencing (AIRR-seq), genome-wide transcriptional and cytokine profiling and high-dimensional cytometry. Novel computational methods are developed to analyze and integrate these data.

Examples of computational methods for large-scale genetic network modeling developed by the Kleinstein Lab include: (1) QuSAGE, which quantifies pathway activity from high-throughput transcriptional profiling data while accounting for gene-gene correlations, (2) LogMiNeR, which leverages prior knowledge networks to improve model interpretability in the analysis of high-throughput transcriptional profiling data. (3) SPEC, which predicts the specific cellular source (e.g., B cells, T cells, etc.) of a gene expression signature using data from total PBMCs, and (4) TIDAL, which integrates genome-wide expression kinetics and time-dependent promoter analysis to reconstruct transcriptional regulatory networks.

Prof. Kleinstein has particular expertise in the management and analysis of high-throughput Adaptive Immune Receptor Repertoire sequencing (AIRR-Seq) data, with a focus on B cell Immunoglobulin (Ig) receptor (BCR)
repertoire analysis. This includes the development of widely used computational analysis methods. These methods are seamlessly integrated and currently made available to the wider scientific community through the Immcantation framework (http://immcantation.org). Application of these computational methods have led to important biological insights in a wide range of systems, including: infection (Lyme, Salmonella, West Nile virus, HIV), vaccination (influenza), allergy (allergic rhinitis, atopic asthma) and autoimmune disease (Multiple Sclerosis (MS), Myasthenia Gravis (MG), Celiac).

For a more detailed description of research in the Kleinstein Lab, including publications and software, please see: https://medicine.yale.edu/lab/kleinstein

Smita Krishnaswamy
smita.krishnaswamy@yale.edu | 203-785-7833 | http://medicine.yale.edu/lab/krishnaswamy

Lab location: SHM I-336A
Rotations available anytime.

Research summary: The Krishnaswamy Laboratory is jointly affiliated with the Department of Genetics in YSM and the Department of Computer Science in SEAS. Our group is also affiliated with the Program in Applied Math, the Yale Institute for Network Science, Yale Cancer Center, and the Yale Center for Biomedical Data Science. The focus of our group is developing machine learning and applied mathematical techniques for extracting structure and patterns in high-dimensional and high-throughput biomedical data. Our recent projects include algorithms for: data denoising, data generation, deep learning-based manifold alignment, data visualization, and developing software for single cell analysis. The projects in the lab tend to be flexible, seeking computationally novel solutions to new problems that are motivated by the explosion of new biomedical measurement technologies and collected patient data. As such we work with data from many disciplines, including genomics, electronic health records, and neuro-imaging. The lab maintains active collaborations with research groups in the Departments of Immunology, Neuroscience, Genetics, Neurology, Radiation Oncology, and Endocrinology as well as outside institutions like Weill Cornell and the Salk Institute. Our group publishes in top-tier computational venues (such as NeurIPS and ICML) and biomedical journals (such as Cell and Nature family journals).

Recent publications include:

There are a wide variety of projects available for rotation students ranging from mathematical and computational
development to applications to domain specific areas. Specific projects for rotation students could include but are not limited to:

1. Project for learning the topology of gene regulatory networks from multimodal data such as scRNA-seq and scATAC-seq data.
2. Learning stochastic dynamics (such as the evolution of populations through differentiation) from single-cell data using neural networks.
3. Application of our manifold-learning techniques to novel datatypes such electronic health record data or gut microbiome data.
4. Embedding, classifying and learning features (using scattering transforms or deep neural networks) in naturally graph-structured data such as fMRI data or neural connectivity data.
5. Analyzing immune responses and deriving response signatures to a wide variety of infectious diseases (west nile virus, lyme disease, dengue) as well as autoimmune conditions (multiple sclerosis, lupus, type 1 diabetes).
6. Analysis of immune cells in response to immunotherapy and cell expansion protocols.

Morgan Levine  
morgan.levine@yale.edu | 203-785-4562  
Lab location: 300 George, Suite 505  
Lab meeting: Mondays at 2:00 p.m.

Do you know what the biggest risk factor for breast cancer is? How about heart disease, lung cancer, or even Alzheimer’s disease? The truth is that the answer for all of these is the same—aging. An 80-year-old has an almost 60 fold increase in his/her risk of developing cancer than a 20-year-old and similar statistics exist for heart disease, diabetes, Alzheimer’s disease, and even many infectious diseases. In the Levine lab we combine computational and molecular studies to discover what happens to the organismal system as it ages and how these changes are implicated in pathogenesis of most major chronic diseases. We have a major focus on quantifying the biological aging process by modeling age-related changes in the epigenetic landscape. Further, evolution has already shown us that biological aging is malleable, as evidenced by the profound diversity in lifespans of different species. Thus, we aim to elucidate mechanisms that regulate the pace of aging and identifying potential targets to slow or ameliorate it.

Potential Rotation Projects:

1. Developing epigenetic (DNA methylation) signatures of cellular senescence and comparing/contrasting with alterations involved in neoplastic transformation.
2. Computational modeling of the DNA methylation landscape with aging in multiple mouse tissues. In moving forward, this project will also model changing epigenetic dynamics in response to interventions (e.g. heterochronic parabiosis, dietary restriction, senolytic therapies, etc.).
3. Proteomic and epigenomic networks from brain, blood, and CSF that are associated with pathogenesis of Alzheimer’s disease in humans—particularly focused on involvement of ApoE4.

A major goal of mine is to create a multidisciplinary research environment in the lab, with the idea that this will lead to more creative and impactful science. We are focused on big questions, rather than employing specific techniques. As a result, students and postdocs in my lab have come from a variety of backgrounds—computation/math, chemistry, genetics, engineering, and cell biology. I am also very passionate about career development and work to help my mentees develop fulfilling independent and interdisciplinary research careers. I
believe it is important for students to have ample opportunity to present their work at scientific conferences. I also work to help my trainees form strong professional networks both at Yale and outside of the institution. In addition to weekly lab meetings, I also strive to meet with my students-one-on-one to provide both research and career advice. Finally, as a woman in STEM, I recognize the potential career challenges that many woman and other underrepresented minorities face, and thus make a point to help support my trainees with whatever obstacles they may face when building their careers in science (either in academia or the private sector).

Zachary Levine
zachary.levine@yale.edu | http://www.zachlevine.com/

Lab location: Room 320, Bass Center, 266 Whitney Avenue
Lab meeting: Wednesdays at 2pm in Bass 326

Rotations available anytime. Please email if you are interested.

The primary goal of the Levine Research Group at the Yale School of Medicine (LRG@YSM) is to understand and intervene in protein folding diseases using biophysical insights. In particular, we strive to understand the underlying physical and thermodynamic mechanisms that lead to age-related disease in a diverse set of biological environments.

We focus predominantly on intrinsically disordered proteins (IDPs) and regions (IDRs), which tend to interconvert between three-dimensional structures. These behaviors enable IDPs to carry out diverse physiological roles, however it can also lead to the aggregation of soluble and insoluble protein plaques (amyloids) over the lifecourse, which is a notable hallmark in certain cancers and in over twenty-one degenerative diseases such as Alzheimer's Disease, Parkinson's Disease, and Type II Diabetes. To better understand these behaviors, LRG@YSM utilizes a combination of molecular dynamics simulations (direct MD, REMD, Umbrella Sampling, Thermodynamic Integration, and Free-Energy Perturbation) and experiments (solution NMR, single-molecule FRET, and optical spectroscopy) that combine physical, chemical, biological, computational, and clinical knowledge in order to manipulate IDPs in cancer, human aging, and in traditional amyloid diseases

Bluma Lesch
bluma.lesch@yale.edu | 203-737-1074 | lesclab.org

Lab location: SHM I-141A
Lab meeting: Mondays, 10:00 a.m., SHM I 143

Rotations available anytime.

Regulation and evolution of chromatin state
We are interested in how chromatin regulates the genome, with special emphasis on how chromatin evolution impacts evolution of gene expression and phenotype. This is a huge, complex, and understudied problem, and we are tackling it at multiple levels. In vitro, we use molecular and genetic tools to manipulate histone modification states in cell culture and test the transcriptional consequences. In vivo, we use mouse models to understand
the developmental roles of chromatin perturbations. And in silico, we compare functional genomics data (ChIP-seq and RNA-seq) from multiple mammalian species to understand the rules governing evolution of chromatin state and how it relates to evolution of gene expression.

Germ cells and gametes are a particularly important place to look at chromatin evolution because (1) an intact germ line is absolutely required for fitness, and (2) chromatin-mediated regulatory states in germ cells have the potential to be inherited. We are especially (but not exclusively) interested in the evolutionary dynamics and developmental role of poising, a chromatin state defined by the simultaneous presence of activating and repressive histone modifications and thought to ‘poise’ developmental regulatory genes for expression, in male mammalian germ cells.

We are looking for students enthusiastic about gene regulation, evolution, and/or reproduction. I especially encourage students interested in combining experimental and computational approaches to rotate. No previous computational experience is required.

Some current projects include:

- **Testing the role of divergent chromatin states in humanized mouse cells.** We have identified hundreds of loci at which chromatin state differs between mammalian species. Does chromatin state diverge randomly, or do species differences in chromatin drive biologically important differences in gene expression and function? To begin to answer this question, we “swapped” human regulatory sequence into the equivalent locus in mouse embryonic stem cells, and we are testing the transcriptional and functional consequences.

- **Direct manipulation of histone modification state using dCas9 fusion proteins.** We are developing a molecular toolkit in which various histone modifier domains are fused to a catalytically dead Cas9 protein. These will allow us to add or remove specific histone modifications at specific loci, and experimentally test hypotheses about the importance of differences in chromatin state.

- **Defining the dynamics of enhancer evolution in the mammalian germ line.** We are comparing ChIP-seq data from several mammalian species to define rules for enhancer evolution in the germ line.

- **Deep evolution of poising in animals.** Our previous work examined evolution of poising in mammals. However, the evolutionary origins of poising are still a mystery. This project will involve collecting germ cells from deeply branching animal species – such as the sea anemone Nematostella – for analysis of genome-wide chromatin state, in order to determine if poising extends to the root of the animal phylogeny.

- **Transgenerational epigenetic inheritance.** We recently found that knocking out the chromatin regulator Kdm6a in the male germ line results in increased cancer rates in genetically wild type offspring. We will now look more closely at germ cells and early embryos to define the molecular mechanism underlying this effect.

---

**Haifan Lin**

haifan.lin@yale.edu | 203-785-6239 (Office) | 203-785-6215 (Lab)

**Lab location:** 10 Amistad Street, Room 237

**Lab meeting:** Mondays from 9:00-11:00 a.m. in room 237A, Amistad Building.

Please contact madeline.riccio@yale.edu if you plan to attend.
We study molecular mechanisms underlying the self-renewing division of stem cells. Currently, we focus on small RNA-mediated epigenetic programming and post-transcriptional regulation that are required for the self-renewal of germline and embryonic stem cells.

Meanwhile, we are exploring the clinical implications of our findings. Stem cells are characterized by their abilities to self-renew and to produce numerous differentiated daughter cells. These two special properties enable stem cells to play a central role in generating and maintaining most tissues in higher organisms. Over-proliferation of stem cells can cause cancer, whereas under-proliferation of stem cells leads to tissue dystrophy, anemia, immuno-deficiency, and infertility. Drosophila and the mouse represent two powerful systems for studying stem cells because they allow easy access to combined genetic, cell biological, and molecular analyses. We use Drosophila and the mouse as complementary models to investigate novel gene regulation mechanisms underlying stem cell division and relevant developmental processes. Previously, we identified germline stem cells in the Drosophila ovary and revealed their self-renewing asymmetric division. We showed that the asymmetric division of these stem cells is controlled by both niche signaling and intracellular mechanisms. Using systematic genetic screens, we then identified key genes involved in both niche signaling and intracellular regulation of stem cell division. Among them, the argonaute/piwi genes represent the only known family of genes required for stem cell self-renewal in both animal and plant kingdoms. We and others also independently discovered a complex class of small-noncoding RNAs that interact with Piwi proteins that we named piRNAs (for Piwi-interacting RNAs). Currently, our research is focused on epigenetic programming and post-transcriptional regulation of germline stem cell self-renewal mediated by the Piwi/Argonaute proteins and piRNAs. We also work on how a class of post-transcriptional regulators called Pumilio proteins function in embryonic and adult stem cells. In addition, we explore the role of these mechanisms in human oncogenesis.

Elias Lolis  
elias.lolis@yale.edu | 203-785-6233

Lab location: SHM B345  
Lab meeting: Mondays, 9:00 – 10:00 a.m., SHM B322

Rotations available any time.

Structural Biology of Chemokines and GPCR Chemokine Receptors
Chemokines are pro-inflammatory proteins that function during an immune response through activation of specific G-protein coupled receptors (GPCRs). Some of the ~45 human chemokines are active during embryonic development, others have homeostatic function in adults, but most function during infection or other cellular stresses to recruit specific immune cells where they are needed. When these receptors are dysregulated, they induce autoimmune diseases and cancer metastasis among other diseases. Projects include the structural biology of chemokines, chemokine receptors complexed to small molecule antagonists, chemokine antagonists, and other binding partners.

Genetics and Therapeutics of Chemokine Receptor-Mediated Cancer
Recently, we identified a chemokine receptor as a potential target for an aggressive T-cell lymphoma. We used Crispr/Cas to make an inducible Cre/lox chemokine receptor knockout in a mouse model where 50% of the mice get lymphoma in 6 months. We are in the process of determining whether (1) this receptor is involved in lymphomagenesis by knocking out the chemokine receptor gene soon after mice are born and examining how many get lymphoma in 6 months (and characterizing the lymphoma properties, e.g., the size of the tumor) and (2) knocking out the gene after the tumors appear, in a separate experiment, to observe what happens to the tumors.
(size of tumors, etc. In addition, we are using high throughput screening (HTPS) to identify small molecule antagonists of the receptor using its chemokine in phage display screening to identify a mutant variant that functions as antagonist for fusing to IgG and testing its effect on the tumors, and creating a bispecific antibody for specificity against this lymphoma. Any therapeutic from this project could also be tested against autoimmune diseases. We are also interested in the structural biology of the receptor and with other complexes.

This project is in collaboration with Francine Foss, MD (Oncology) who treats patients with this lymphoma and Demetrios Braddock, MD/PhD (Pathology) who diagnoses these patients.

**Macrophage migration inhibitory factor and inhibition of its function**

Human macrophage migration inhibitory factor (MIF) is a pro-inflammatory cytokine involved in immunity and, when dysregulated, induces inflammation and cancer, among other diseases. Interestingly, its three-dimensional structure is similar to microbial enzymes that have isomerase/tautomerase activity. Many parasites also express homologues of this homotrimeric protein, presumably to interfere with the immune response. We have determined the human apo structure as well complexes with inhibitors that function as competitive, non-competitive, covalent, and allosteric and are used as reagents to probe its inflammatory and oncogenic activities. Molecular dynamics studies identified a second allosteric site that regulates some of the protein’s biological activities. This second allosteric site needs further characterization. We are also interested in determining structures of complexes with MIF binding proteins including its receptor, CD74.

**Potential rotation projects:**

1. Characterize constitutive active mutants of human chemokine receptor using a genetically modified S. cerevisiae as well as mammalian cells.
2. Screen and characterize chemokine receptors for antagonists and inverse agonists.
3. Crystallize the intracellular domain of CD44, the signaling receptor for MIF (NEW PROJECT).
4. Characterization of allosteric site of MIF.
5. Characterize the mechanism of action of a NCI-screened inhibitor that inhibits MIF only on promyelocytic leukemia.

**Publications**

My laboratory focuses on noncoding and epigenetic mechanisms that are underlying cellular identity, cellular response, cancer and anti-cancer immunity. We utilize the amazing blood-forming system, or hematopoiesis, as a model system. In a normal adult human being, ~100 to 200 billion new blood cells are generated every day to replace similar numbers of existing blood cells. These mature blood cells originate from hematopoietic stem cells and exhibit vastly different forms, shapes and functions, regulating processes such as innate and adaptive immune responses, oxygen transport and coagulation. We are addressing a number of intriguing questions in this system. What regulates the cross-talk between cancer cells and hematopoietic cells? What regulates the competition between clones of hematopoietic stem cells? What regulates the transformation of normal hematopoietic stem cell into malignant cells? How is the speed of hematopoietic regeneration regulated? What controls the subcellular organelle shape and function in blood cells?

The following rotation projects are available.

1. **The epigenetic and signaling control of anti-cancer immunity**
   Immune cells are capable of eliminating cancer cells, but they are often not fully functional against cancer cells, or even effectively assist cancer cells during cancer progression. We study the epigenetic and signaling controls that regulate immune cell’s function toward cancer, with a focus on genetic alterations in hematopoietic stem cells that accumulate through aging. This project will involve in vitro and in vivo experiments that test and elucidate mechanisms of epigenetic enzymes, signaling pathways and drugs in anti-cancer immunity.

2. **Mapping and visualizing of functional noncoding elements in mammalian genomes**
   The noncoding genome has been recognized to play critical functions in mammalian cells, including functional elements such as noncoding RNAs, enhancers and transposable elements. This project is aimed at understanding both the function and 3D organization of noncoding sequences in the genome. We will utilize a new technology, named Molecular Chipper, that our lab developed to generate CRISPR libraries. We will further make use of these libraries to identify novel noncoding genetic elements in stem cells and cancer cells and visualize their localization in living cells.

3. **Decoding the RNA licensing codes in cancer**
   RNA binding proteins (RBPs) function through binding to coding and noncoding RNAs. There are thousands of such RBPs in the genome that each has its own specific set of targets RNAs. For each RBP, there must exist sequence and structural codes in the RNAs that permit their binding to the RBP, a process we refer to as RNA licensing. These licensing codes are poorly understood in general and could be regulated in disease conditions such as cancer. This project involves both computational and experimental approaches, and participation by rotation students with either computational or experimental background is highly welcome.

4. **Molecular control of organelle shape**
   The morphologies of subcellular organelles often differ between cell types, but what controls of the shapes of such organelles are often enigmatic. This rotation project focuses on the unique morphology of neutrophil nuclei to investigate the principles and mechanisms that govern the shape of subcellular organelles. Unlike most other cell types, human neutrophil condenses and fragments its nucleus during
differentiation, which is a process recently evolved. Investigation will involve testing the importance of adopting special nuclear shapes, and using live cell imaging coupled.

Robert McDougal
robert.mcdougal@yale.edu | 203-737-4828
Lab location: Room 526, Suite 501, 300 George St.
Rotations available any time.

My lab uses computational approaches, including biophysically detailed simulations and machine learning to gain insight into brain function.

Potential project areas include:

1. Investigate the effects of ischemic stroke and its aftermath - design models, run simulations, test hypotheses to gain insights on how to maximize neuron survival following a stroke.
2. Alzheimer’s data mining - help identify promising treatments by building tools to mine the literature to look for treatments showing promise in a wide range of model organisms.
3. Neuroscience data sharing - build NLP tools to identify experimentally derived neuron properties or computational models, automatically predict metadata annotations, apply the FAIR data sharing principles to maximize utility, extract and visualize structured data from model code, and more.
4. Advance computational neuroscience infrastructure - help enable others to use computational techniques to study the brain by designing simulator algorithms, interfaces, or standards. My lab is especially interested in building tools to explore (and then exploring) the interaction between intracellular chemical processes and electrophysiology.
5. Study of the role of morphology in the regulation of neuronal activity.

James Noonan
james.noonan@yale.edu (contact me by e-mail to arrange a visit) | 203-737-1922 | noonanlab.org
Lab location: SHM I-142C
Lab meeting: Fridays, 10:00 a.m., SHM I-143
Rotations available any time.

Research Overview:
My current research program is focused on addressing two distinct, but complementary, questions. I started my laboratory in 2007 to decipher the role of gene regulatory changes in the evolution of uniquely human traits. Our initial efforts targeted a class of elements that I and others first characterized over a decade ago: Human Accelerated Regions (HARs). These elements are highly conserved across species but show many human-specific changes, suggesting they encode uniquely human functions of potentially large effect. We have since expanded the scope of our work to globally identify human-specific regulatory innovations using experimental methods. Our current research is aimed at understanding the phenotypes these regulatory changes specify using humanized mouse models, massively parallel genetic screens, and cellular models of primate neurodevelopment.

More recently, we have expanded our interests further to identify gene regulatory mechanisms underlying a
human neurodevelopmental phenotype of enormous public health significance: autism. Our current research is focused on characterizing regulatory networks disrupted in autism using mouse and cellular models of brain development. This work is highly synergistic with our studies of human brain evolution, drawing on the same intellectual and experimental resources we have established over the last ten years. Ultimately, we anticipate that our parallel research interests will converge to reveal uniquely human molecular, cellular and developmental processes in the brain that are perturbed in autism.

**Current Research Topics:**

**Linking uniquely human regulatory changes to human traits using genetic and experimental models.**

Humans possess unique biological features that distinguish us from all other species on this planet. My laboratory is pursuing the long-standing hypothesis that changes in gene regulation underlie the evolution of many uniquely human phenotypes. Our prior studies have provided insight into the landscape of uniquely human developmental regulatory functions in the genome, yielding a rich collection of loci for experimental studies. In our current work, we are moving beyond the discovery of novel human regulatory functions towards understanding the molecular, cellular and developmental traits they specify.

To accomplish this, we are developing two essential tools. The first is the means to overcome the species barrier and employ the power of experimental genetics to study uniquely human genomic features in model organisms. The second is the ability to access and compare developmental processes in humans and other great apes. Current projects underway in the lab include:

**Humanized mouse models.** Building on our prior work, we are using genome editing to generate humanized mouse models for HARs and other regulatory elements that encode human-specific functions during development. These models will allow us to link uniquely human genetic changes to the cellular and morphological phenotypes they produce. We have established several models and shown that each humanized regulatory element maintains its human-specific function in the mouse genomic context. We are using these models to identify changes in gene expression and regulation throughout development. We hypothesize that each humanized allele results in a "humanization" of developmental gene expression patterns and regulatory functions.

To identify specific humanized cell types, we are carrying out unbiased single-cell transcriptome analyses. Guided by these studies, we will then identify changes in developmental processes during organogenesis.

**Comparative studies of primate neurodevelopment using induced pluripotent stem cells.** The inability to compare molecular and cellular events across great ape neurodevelopment is a fundamental obstacle to understanding how the human brain evolved. To meet this challenge, we have assembled a panel of induced pluripotent stem cells from several great ape and primate lineages. This resource enables us to model great ape neurodevelopment and to globally identify molecular and cellular changes in humans. Our panel has two important features: we include chimpanzee and orangutan, which will allow us to identify changes specific to the human lineage; and we include multiple individuals from each species, providing robustness and statistical power for our comparisons.

We are generating both monolayer neural stem cells (NSCs) and human cortical organoids from human, chimpanzee and orangutan iPSCs, using rhesus and marmoset iPSCs as additional outgroup species. We will use these models to identify human-specific evolutionary changes in gene expression and regulation, neuronal progenitor proliferation, and differentiation as well as the genetic changes that drive them. We will then model human-specific sequence changes by using genome editing to generate "humanized" chimpanzee iPSCs and organoids.

**Global screens to characterize uniquely human regulatory functions.** We are using CRISPR-Cas9 knockout screens in human and primate neural stem cells and neurons to disrupt thousands of enhancers showing human-specific changes in activity. In NSCs, this has allowed us to precisely measure the contribution of each enhancer - and the substitutions within them - to human progenitor proliferation. Going forward, we will combine this approach with single-cell transcriptome surveys to identify enhancer mutations with large effects on their target genes.

We are also using comparative massively parallel reporter assays (MPRAs) to measure the effects of thousands of
uniquely human genetic changes on enhancer function. In this approach, the activities of thousands of short human regulatory sequences with human-specific sequence changes are measured and compared to the activities of their chimpanzee orthologs. MPRA is enabling us to discover novel human-specific regulatory functions in an unbiased manner and hone in on the molecular mechanisms that generated them.

**Identifying neurodevelopmental regulatory networks perturbed in autism spectrum disorder**
The biological foundations of autism — the specific cell types in the brain that are affected, the neurodevelopmental events that are perturbed, and the critical stages when pathology begins — are unknown. However, whole exome and whole genome studies have identified genes that contribute to autism risk. Many of these genes encode transcriptional regulators that converge in gene co-expression networks in mid-fetal human cortex. Loss of function mutations in these genes may cause haploinsufficiency, producing widespread expression changes across regulatory networks that result in altered neurodevelopment.

One of the most promising autism risk genes identified in whole-exome surveys is the chromatin remodeler CHD8. We recently identified the regulatory targets of CHD8 in developing mouse and human brain. We found that other autism risk genes are strongly overrepresented among CHD8 binding targets, and that shRNA-induced knockdown of CHD8 expression in cultured human neural stem cells (hNSCs) results in dysregulation of CHD8-bound autism risk genes. Our results suggest that loss of CHD8 disrupts regulatory networks in the prenatal human brain, thereby contributing to autism pathology.

Given the cellular heterogeneity of the developing cortex, there is likely to be substantial cell-type specific variation in the effects of loss of function mutations in CHD8 or other regulatory genes associated with autism. Using single cell RNA sequencing in a Chd8 knockout mouse model, we are characterizing the transcriptional effects of Chd8 haploinsufficiency within and across cellular compartments throughout cortical development. In parallel, we are mapping cell-type specific binding targets of CHD8 as well as other autism-associated chromatin modifiers that CHD8 regulates in the cortex, and in human neural progenitors and derived neuronal subtypes. We are also using CRISPR-mediated gene and enhancer knockouts in these cellular models to identify critical upstream regulators of CHD8 and other autism risk genes, and to measure the effects of autism risk gene haploinsufficiency on global gene expression. By integrating these approaches, we will elucidate cell-type specific CHD8 regulatory networks and establish how those networks are disrupted due to loss of CHD8.

Corey O'Hern
corey.ohern@yale.edu | 203-432-4258

**Lab location:** ML 203  
**Lab meeting:** Every Monday at 3:00 p.m., Mason Lab Room 200

*Rotations available any time.*

1 - **Computational studies of protein structure, dynamics and design**
Although proteins form the basis of almost all biophysical processes in living organisms, computational approaches still cannot fully explain and predict their structure and dynamics. For instance, it is difficult to predict even local structural changes caused by single amino-acid mutations to protein cores with high accuracy. Current computational methods tend to focus on knowledge-based scoring functions, but these have difficulty distinguishing between experimentally-derived protein structures and protein decoys that do not occur in nature. We take an alternative approach, using simple physical models to describe amino acid interactions and predict hydrophobic sidechain placement and packing fractions in protein cores. We seek to extend these models to describe protein structural fluctuations, predict the response to amino acid mutations, and detect protein decoys.
While proteins tend to adopt a single, global conformation, fluctuations about a mean structure are crucial to proper function, and thus modeling these fluctuations is important for understanding allostery and the effects of amino acid mutations. We seek to understand folded proteins as an ensemble of related conformations using an all-atom, hard-sphere-like model for proteins, where amino acids with backbone connectivity interact via repulsive atomic potentials and satisfy experimentally-determined atomic distance constraints from NMR spectroscopy data. Similar models have been used in our group to model fluctuations in intrinsically disordered proteins. These studies will help illuminate which regions of proteins are the most flexible, and which regions are the most rigid. We can also use these models to predict structural changes in response to single amino acid mutations to cores. In addition, understanding the underlying ensemble of conformational states of a given protein will allow us to differentiate between naturally occurring and disease-associated mutations.

Understanding a protein’s native state fluctuations will also aid efforts to differentiate between accurate computational predictions and inaccurate “decoy” structures in folding and design applications. When researchers try to computationally fold a sequence into a well-defined structure or try to design a sequence that fits a given structural motif, current methods generate many low-energy decoy predictions that deviate from natural proteins. Physics-based models can shed light on these issues by providing fundamental bounds on what relevant criteria define natural proteins. These projects involve developing physical metrics for evaluating the accuracy of experimental and computational protein structures, as well as using simulation techniques such as molecular dynamics simulations and Monte Carlo sampling to understand the physical principles that govern protein structure and dynamics.

Publications:

2 - Modeling active polymers with applications to chromatin organization
Many biopolymers are acted on by nonequilibrium, active processes that consume energy to induce motion or remodel the polymer. For example, a number of active processes are used to convert the information stored in DNA into proteins. These active processes drive the polymers out of equilibrium, leading to strong fluctuations and deviations from the results of equilibrium polymers. There have been several recent theoretical and computational studies to understand the structure and dynamics of active polymers. We have developed a new model of an active polymer, where a particle with an attractive interaction to the polymer moves along the polymer and applying a constant tangential force. The dynamics of this model as characterized by the mean square displacements of individual monomers are consistent with the active Rouse model. Furthermore, we have shown that upon increasing the activity, there is a dramatic drop in the radius of gyration of the polymers. The collapse of the polymer can be induced by increasing the number of active pullers and increasing the magnitude of the active force. In future studies, we will test the hypothesis that this model can be used to explain the globular segregation of chromosomes in the nuclear envelope. In future applications, this model can be adapted to more closely model loop extrusion of the chromatin fiber and the effect of local volume exclusion of nascent transcripts via transcription. This project is carried out in collaboration with Professors Megan King in the Department of Cell Biology and Simon Mochrie in the Departments of Physics and Applied Physics.

3 - Modeling of cells in tissues and tumors
Clinical pathologists typically employ qualitative visual analyses to identify different cell types and diagnose disease in tissue samples. The emergence of in vivo imaging techniques has advanced our ability to visualize and model cell spatiotemporal dynamics in three spatial dimensions (3D) during tissue and tumor development. The goal of this project is to use physics-based approaches and computational methods to quantitatively describe and model the melanoma tumor microenvironment. The project is a collaboration between the O’Hern and Bosenberg research laboratories. Prof. Bosenberg has significant expertise in murine models of melanoma and state-of-the-art in vivo imaging of melanoma tumors in 3D using two-photon microscopy. There have been very few studies that have reported 3D in vivo imaging of cells in tumors, and this data will allow us to develop a robust, experimentally validated, computational model of the tumor microenvironment. Prof. O’Hern has significant expertise in computer simulations of jamming and crowding in biological systems including proteins, cells, and tissues. We will use novel image analysis techniques, molecular dynamics simulations, and cell shape and topology classifications to investigate melanoma tumor formation, answering several key, open questions in the field of cancer biology: (1) How do cancer and immune cells move in dense tissues and tumors? (2) Can we determine whether immune cells have been activated to fight tumor formation by studying their time-dependent shape and mechanical properties? and (3) How does the spatial organization of the tumor microenvironment influence whether melanoma tumors will proliferate? To address these questions, we will first develop an image-processing and analysis pipeline to identify cell surfaces and centers from in vivo two-photon microscopy of melanoma tumors in mice. Second, we will generalize the deformable particle model (DPM) that we developed to model bubbles and emulsions in 2D to 3D to simulate cell packing in tissues and tumors. Third, we will identify the key factors that determine whether a tumor will form or regress in mice that have been injected with large populations of cancer cells, and develop an experimentally validated computational model that can simulate the response to cancer cell invasion.

Publications:

Lajos Pusztai
lajos.pusztai@yale.edu | 203-737-8309

Lab location: Room 786, 300 George St. (wet lab); Suite 120, 300 George St. (computational biology group)
Lab meeting: Thursdays from 3-5pm in Room 130, 300 George Street

Rotations available any time.

The Breast Cancer Translational Research Lab includes a wet lab and a computation biology team. Our research covers a broad range of areas including (i) characterizing the immune microenvironment in breast cancer tissues, (ii) develop drugs against metabolic enzymes in cancer as new therapies, (iii) develop a novel antibody drug conjugate that targets an aberrantly expressed receptor on cancer cells, and (iv) study how germ line functional polymorphisms interact with somatic mutations to cause malignant transformation.

Recent publications can be accessed at: https://medicine.yale.edu/lab/pusztai/publications/

Anna Marie Pyle
anna.pyle@yale.edu | 203-432-5633

Lab location: KBT 826
Lab meeting: Fridays, 10:00 a.m., 1202 KBT

Rotations available any time.

Research Summary
The Pyle Laboratory studies RNA structure and RNA recognition by protein enzymes. We use a combination of experimental biochemistry and crystallography to study the architectural features of large RNA molecules, such as self-splicing introns and other noncoding RNAs. This is accompanied by complementary work on RNA-dependent ATPase enzymes that bind and remodel RNA structures, with an emphasis on proteins that are involved in viral replication and host innate immune response. Our studies involve a combination of solution biochemistry, enzymology, crystallography, and cell-based functional approaches. In parallel, we develop new computational methods for solving, analyzing and predicting RNA structures.

We have rotation projects on the following topics:
1. Innate immune receptor structure and mechanical function
2. Small molecule activators and inhibitors of innate immune receptors.
4. Self-splicing group II introns; structure, folding and catalytic mechanism
5. Functional architecture of lincRNA molecules.
6. Computational approaches for modeling and predicting RNA tertiary structure
### Hemant D. Tagare

**Email:** hemant.tagare@yale.edu  |  **Phone:** 203-737-4271  |  **Website:** [http://noodle.med.yale.edu/hdtag/profile.html](http://noodle.med.yale.edu/hdtag/profile.html)

**Lab location:** TAC 309C

*Rotations available. Please contact me.*

### Computational Structural Biology via Cryogenic Electron Microscopy

The Tagare lab is interested in developing the mathematical theory and computational algorithms for understanding three-dimensional structures of biological macromolecules from cryogenic electron microscopy (cryo-EM) data. Cryo-EM is a relatively new and promising tool for structural biology. Its inventors (Jaques Dubochet, Joachim Frank, and Richard Henderson) won the Nobel Prize in 2017. Cryo-EM's success is intimately related to the success of several underlying algorithms.

We are interested in posing new computational problems and developing new algorithms for cryo-EM. We have several projects underway to develop methods for understanding structural changes in heterogeneous proteins, for understanding local resolution and validation of structures, and for developing fast reconstruction algorithms.

Please contact me if you are interested in a rotation in my lab. Students with a computational and mathematical background who are interested applying computation to real-world problems might find cryo-EM problems interesting.

**Sample cryo-EM publications:**


[Image: Heterogeneity in the RNA-dependent RNA-polymerase of the influenza virus (from [1] above). This is a result from a new algorithm which can directly reconstruct principal components of a heterogeneous protein from cryo-EM images.]

### Andrew Taylor

**Email:** richard.taylor@yale.edu  |  **Phone:** 203-909-3012

**Lab location:** Room 524, Suite 501, 300 George Street

**Lab meeting:** Tuesday and Thursday

*Rotations available anytime.*
The Taylor Lab focuses on applying data science to various aspects of clinical care (emergency medicine focus). Current areas of research include:

**Machine learning/Deep learning for predictive analytics** – Emergency medicine is a unique and exciting field for the application of predictive analytics. Providers must make numerous decisions (admission/discharge; ordering tests, medications, etc.) in a chaotic environment within a compressed timeframe that can lead to a variety of cognitive errors. Our lab is focused on augmenting this decision process and lessening the cognitive burden of providers through integration of machine learning tools into clinical workflows. To accomplish this task, we use a variety of methods including deep learning.

**Data Mining/Unsupervised Learning** – Adoption of EHRs has led to an explosion of secondary data available for research. We use a variety of data science tools to mine EHR emergency medicine data, find novel relationships, and gain better insight into care processes. Our current research is focused on finding low-dimensional representations of ED encounters and using cluster analysis for phenotype discovery.

**Discovery of optimal pathways of care through the use of decision analysis** – Our work in this area is primarily focused on establishing appropriate testing thresholds and cost-effective clinical pathways for emergency conditions including: aortic dissection, renal colic, trauma, and head injury.

**EHR-driven, outcomes-based research** – Current work in this area focuses on causal analysis of difficult to randomize interventions in emergency research using observational EHR data. For example, we are interested in examining the effect of point-of-care ultrasound on mortality and other patient-centered outcomes.

---

**David van Dijk**

david.vandijk@yale.edu | 203-785-2241 | www.vandijklab.org

**Lab location:** Yale Cardiovascular Research Center Office Room 770K, 300 George Street

**Lab meeting:** Fridays, 1:00pm, 300 George Street. Visit for more details.

*Rotations available any time.*

Big data and machine learning are revolutionizing biomedical discovery. Algorithms allow us to find patterns in data that are hidden to the human eye and enable building of predictive models of disease, disease progression, and treatment outcomes. In the van Dijk lab we leverage state-of-the-art machine learning and AI to provide new insights into large-scale genomic, biomedical imaging, gut microbiome, and patient health record data.

**Ongoing areas of research include:**

**Deep learning on biomedical imaging:** Hospitals are generating large amounts of imaging data that are used for diagnostic purposes. Traditionally, most of the analysis is done manually by physicians. We want to use machine learning to automatically analyze, diagnose, and provide insights into such biomedical imaging data using both unsupervised (without labels) and supervised (with labels, such as diagnosis) approaches.

**Predicting disease and therapy outcome from high-throughput antibody reactivity assays:** In collaboration with the Ring lab (Yale Immunobiology) we are building machine learning models to predict disease progression and therapy outcome for a large number of cancer patients, patients with autoimmune disease, and healthy individuals. The data comes from a new high-throughput assay that the Ring lab has developed in which reactivity of thousands of antibodies can be measured for thousands of samples at a time. The models that we will build could ultimately be used for personalized medicine and discovery of new therapeutic targets.
**Predicting microbe-host interaction:** Together with the Palm lab (Yale Immunobiology) we are analyzing data from gut microbiome host interaction in order to predict which microbes and in particular which proteins, genes, and pathways are involved in the microbe-host interactions. The Palm lab uses a novel high-throughput assay to measure these interactions and is generating large amounts of data that require machine learning for discovery and prediction of new microbe-host interactions that can be used for therapeutic purposes.

**Unsupervised learning on single-cell data:** Recent advances in single-cell sequencing (e.g. 10x genomics) allow us to measure transcriptomes (as well as epigenomes) in thousands of cells at a time resulting in large, but noisy, data. Together with our collaborators in Genetics and Immunobiology we are generating big single-cell datasets on cancer, immunology, infectious disease, embryonic development, and much more. We are developing new and applying existing unsupervised machine learning algorithms to find structure, visualize, provide biological insights, and generally make sense of this data.

**Selected Publications:**

**Anita Wang**
zuoheng.wang@yale.edu | 203-737-2672

**Lab location:** 60 College St, Yale School of Public Health
**Lab meeting:** Tuesdays, 1-3pm

*Rotations available any time*

Dr. Wang’s research focuses on the development of statistical methods and computational algorithms to better identify genes contributing to complex human diseases such as cancer, pulmonary diseases, and psychiatric disorders.

**Hongyu Zhao**

hongyu.zhao@yale.edu | 203-785-3613

**Lab location:** 300 George Street, Suite 503
**Lab meeting:** Mondays and Fridays, 9:00 - 10:30 a.m., 300 George Street, Suite 503

*Rotations available any time.*
Available rotation projects:
1. Interpretation of personal genomes through integrated analysis of GWAS data, phenotype information, and other data sources of information.
2. Analysis of large biobank data, e.g. UK Biobank, with hundreds of thousands of individuals and thousands of traits to delineate the genetic architecture of complex diseases.
4. Disease risk prediction.
5. Analysis of cancer genomics data to identify cancer subtypes, biomarkers for treatment options and responses, and combination therapies. Focus on lung and breast cancers.
6. Joint analysis of transcriptome (gene expression data), genome (genetic variations such SNPs and CNVs), and imaging data to identify genetic basis of neurodevelopment and psychiatric disorders.
7. Analysis of electronic medical record and wearable device data for precision health and precision medicine.

Recent Publications:

Steven Zucker
steven.zucker@yale.edu | 203-432-6434

Lab location: AKW 407, 51 Prospect Street
Lab meeting: Nothing is regularly scheduled because of the small lab size
Rotations are only available in the fall

Our group studies computational vision and computational neuroscience. We have developed a class of ecologically-relevant stimuli that exercise more of the visual system than classical laboratory stimuli, yet are more analyzable than 'natural' images. We work with experimental groups at UCSF (visual cortex) and Duke (retina), and are developing machine learning techniques to derive a functional manifold of neural activity based on these stimuli. We also work on the perception of shape and color -- and their interactions -- in primates. Using mathematical ideas from differential geometry and topology, we discovered novel invariants connecting image patterns to surface patterns generically.