Yale Computational Biology and Bioinformatics Rotation Guide 2020-2021

Julien Berro	
Cynthia Brandt	3
Kei-Hoi Cheung	4
Chris Cotsapas	4
Richard Flavell	4
Joel Gelernter	5
Mark Gerstein	6
Antonio Giraldez	7
Jeffrey Gruen	8
George Hauser	
Monika Jadi	
Samah Fodeh Jarad	
William Jorgensen	14
Amy Justice	
Naftali Kaminski	
Steven Kleinstein	
Smita Krishnaswamy	16
Bluma Lesch	17
Morgan Levine	
Zachary Levine	
Haifan Lin	
Elias Lolis	
Jun Lu	
Robert McDougal	
Kathryn Miller-Jensen	
John Murray	
James Noonan	
Corey O'Hern	
Lajos Pusztai	
Anna Marie Pyle	
Hemant D. Tagare	
Andrew Taylor	
Jeffrey P. Townsend	
Serena Tucci	

David van Dijk	
Anita Wang	
Jack (Kai) Zhang	
Hongyu Zhao	
Steven Zucker	

Note: This is not a complete listing of all the faculty who are currently involved with the CBB program.

Julien Berro

julien.berro@yale.edu | http://campuspress.yale.edu/berrolab/ | @BerroLab

Lab location: Bass 230 (Science Hill) and ISTC 214C (West Campus) Lab meeting: Fridays, 9:00 a.m., ISTC 201

Journal Club: Mondays, 9:00 a.m., ISTC 262A

Rotations available anytime.

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 clathrinmediated 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 crossdisciplinary approaches we develop and use.

The Berro lab strongly believes that diversity – of ideas, identities, perspectives and backgrounds – is key to scientific creativity and productivity. Therefore, we recruit lab members with a broad range of backgrounds and professional goals.

Please do not hesitate to contact us to discuss your research interests and arrange a visit of the lab.

Below are a small sample of possible rotation projects (we will design your rotation project together according to your interests)

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

Cynthia Brandt

cynthia.brandt@yale.edu | 203-737-5762

Lab location: Room 524, Suite 501, 300 George Street Lab meeting: Mondays: 8:00-12:00 p.m.; Tuesdays, Wednesdays, Fridays: 8:00-4:00 p.m.;

Rotations available in the spring term.

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 (Zoom)

Rotations available anytime.

Our lab combines experimental and analytical techniques to uncover the pathogenic mechanisms underlying autoimmune and neurological diseases and identify outcome predictors. Many of these diseases occur at different rates in men and women; we are therefore also interested in sex-specific effects in disease. Due to restrictions in lab occupancy due to SARS-CoV-2 we are only offering analytic rotations in 2020 but will be offering experimental projects in 2021.

The following rotation projects are now available:

- 1. Establish if age at puberty is causal for multiple sclerosis risk. Earlier puberty is associated with higher disease risk, but we do not know if this is causal. In this project you will use causal inference methods to assess if pubertal timing and associated parameters such as childhood BMI affect disease risk. Some programming or mathematical background is an advantage.
- 2. New methods for correlated trait association. We often measure multiple cellular parameters in the same individuals, and no single parameter captures the underlying biology. Your project will be to test if we can capture the underlying, unmeasured phenotype using principal component analysis and other decomposition techniques, then map genetic determinants of these factors.

Richard Flavell

richard.flavell@yale.edu |

Lab location: TAC S-569 Lab meeting: Every Wednesday, 9:30 a.m., TAC S-647

Rotations available any time.

Rotation projects are available in the following areas:

Regulation of Immunity at mucosal barriers

The role of the nervous system The role of microbial metabolites

Jarret A, Jackson R, Duizer C, Healy ME, Zhao J, Rone JM, Bielecki P, Sefik E, Roulis M, Rice T, Sivanathan KN, Zhou T, Solis AG, Honcharova-Biletska H, Vélez K, Hartner S, Low JS, Qu R, de Zoete MR, Palm NW, Ring AM, Weber A, Moor AE, Kluger Y, Nowarski R, Flavell RA. Enteric Nervous System-Derived IL-18 Orchestrates Mucosal Barrier Immunity. Cell. (2020) Jan 9;180(1):50-63.e12. doi:10.1016/j.cell.2019.12.016. PubMed PMID: 31923399.

Immunometabolism

Metabolic checkpoints in immune cell differentiation

 Bailis W, Shyer JA, Zhao J, Canaveras JCG, Al Khazal FJ, Qu R, Steach HR, Bielecki P, Khan O, Jackson R, Kluger Y, Maher LJ 3rd, Rabinowitz J, Craft J, Flavell RA. Distinct modes of mitochondrial metabolism uncouple T cell differentiation and function. <u>Nature</u>. (2019) Jun 19. PMID: 31217581

Modeling the human immune system in mice

• Das R, Strowig T, Verma R, Koduru S, Hafemann A, Hopf S, Kocoglu MH, Borsotti C, Zhang L, Branagan A, Eynon E, Manz MG, Flavell RA, Dhodapkar MV. <u>Microenvironment-dependent growth</u> of preneoplastic and malignant plasma cells in humanized mice. <u>Nat Med.</u> Nov (2016). PMC5101153.

Joel Gelernter

joel.gelernter@yale.edu | 203-932-5711, ext 3590

Lab location: VA Building 2, Rm 7-121 Lab meeting: Friday mornings, 8:45 a.m. (at VA)

Journal club: Tuesdays, 2:30 p.m.

Rotations available any time.

Available rotation projects: Negotiable

Psychiatric traits including substance use disorders, and other behaviors, are strongly influenced by genetics, and are genetically complex. We study genetics of psychiatric illness (focusing on substance dependence traits, posttraumatic stress disorder, depression, and anxiety) and related traits (such as personality and neuroimaging), issues relevant for complex trait genetics, and population genetics. We do this by means of study of genetic polymorphism, sequence, and copy number variation, in the context of exome and genome sequencing, association, genomewide association studies (GWAS), post-GWAS analyses, and studies of gene-environment interaction. We have a local (>15,000 subject) deeply-phenotyped GWAS dataset for substance dependence traits, with whole exome sequence (WES) data on 3000. We have projects underway with several large consortia, including the Psychiatric Genomics Consortium (PGC) (where Dr Gelernter shares leadership of the Substance Use Disorders group) and the Million Veterans Program (MVP), which now has genotype and phenotype data available from >400,000 participants, with >825,000 recruited to date. Our main publications based on MVP data have addressed PTSD, depression, anxiety, problematic alcohol use, and other traits. The value of the MVP dataset continues to increase as more subjects are recruited, and epigenetic and whole genome sequence data become available. We study populations in the US and Thailand; one ongoing project involved obtaining WES and GWAS data for 2000 extremephenotype Thai subjects to identify genes influencing risk for methamphetamine dependence. Finally, we are now initiating a VA focused biobank project with the aim of recruiting 50,000 additional subjects with primary psychiatric diagnoses. These resources provide rich opportunities to motivated trainees for study of psychiatric phenotypes, genetic association, and deep statistical analyses in a highly collegial environment.

Recent representative publications:

- Gelernter J, Sun N, Polimanti R, Pietrzak R, Levey DF, Bryois J, Lu Q, Hu Y, Li B, Radhakrishnan K, Aslan M, Cheung KH, Li Y, Rajeevan N, Sayward F, Harrington K, Chen Q, Cho K, Pyarajan S, Sullivan PF, Quaden R, Shi Y, Hunter-Zinck H, Gaziano JM, Concato J, Zhao H, Stein MB; Department of Veterans Affairs Cooperative Studies Program (#575B) and Million Veteran Program. Genome-wide association study of post-traumatic stress disorder reexperiencing symptoms in >165,000 US veterans. Nat Neurosci. 2019 Sep;22(9):1394-1401. doi: 10.1038/s41593-019-0447-7. Epub 2019 Jul 29. PubMed PMID: 31358989.
- Montalvo-Ortiz JL, Cheng Z, Kranzler HR, Zhang H, Gelernter J. Genomewide Study of Epigenetic Biomarkers of Opioid Dependence in European- American Women. Sci Rep. 2019 Mar 15;9(1):4660. doi: 10.1038/s41598-019-41110-7. PubMed PMID: 30874594; PubMed Central PMCID: PMC6420601.
- Zhou H, Sealock J, Sanchez-Roige S, Clarke TK, Levey D, Cheng Z, Li B, Polimanti R, Kember RL, Vickers Smith R, Thygesen JH, Morgan MY, Atkinson SR, Thursz MR, Nyegaard M, Mattheisen M, Børglum AD, Johnson EC, the VA Million Veteran Program, Justice AC, Palmer AA, McQuillin A, Davis LK, Edenberg HJ, Agrawal A, Kranzler HR, Gelernter J: Genome-wide Meta-Analysis of Problematic Alcohol Use in 435,563 Individuals Yields Insights Into Biology and Relationships With Other Traits. Nat Neurosci 2020 May 25. doi: 10.1038/s41593-020-0643-5. Online ahead of print.
- Kranzler HR, Zhou H, Kember RL, Vickers Smith R, Justice AC, Damrauer S, Tsao PS, Klarin D, Baras A, Reid J, Overton J, Rader DJ, Cheng Z, Tate JP, Becker WC, Concato J, Xu K, Polimanti R, Zhao H, Gelernter J. Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. Nat Commun. 2019 Apr 2;10(1):1499. doi: 10.1038/s41467-019-09480-8. PubMed PMID: 30940813; PubMed Central PMCID: PMC6445072. Correction: Nat Commun. 2019 May 17;10(1):2275. doi: 10.1038/s41467-019-10254-5.
- Polimanti R, Walters RK, Johnson EC, McClintick JN, Adkins AE, Adkins DE, Bacanu SA, Bierut LJ, Bigdeli TB, Brown S, Bucholz KK, Copeland WE, Costello EJ, Degenhardt L, Farrer LA, Foroud TM, Fox L, Goate AM, Grucza R, Hack LM, Hancock DB, Hartz SM, Heath AC, Hewitt JK, Hopfer CJ, Johnson EO, Kendler KS, Kranzler HR, Krauter K, Lai D, Madden PAF, Martin NG, Maes HH, Nelson EC, Peterson RE, Porjesz B, Riley BP, Saccone N, Stallings M, Wall TL, Webb BT, Wetherill L; Psychiatric Genomics Consortium Substance Use Disorders Workgroup, Edenberg HJ, Agrawal A, Gelernter J. Leveraging genome-wide data to investigate differences between opioid use vs. opioid dependence in 41,176 individuals from the Psychiatric Genomics Consortium. Mol Psychiatry. 2020 Feb 26. doi: 10.1038/s41380-020-0677-9. [Epub ahead of print] PubMed PMID: 32099098.
- Levey DF, Stein MB, Wendt FR, Pathak G, Zhou H, Aslan N, Quaden R, Harrington KM, Sanacora G, MacIntosh AM, Concato J, Polimanti R, Gelernter J, on behalf of the Million Veteran Program: GWAS of Depression Phenotypes in the Million Veteran Program and Meta-analysis in More than 1.2 Million Participants Yields 178 Independent Risk Loci. MedRxiv doi: https://doi.org/10.1101/2020.05.18.20100685

Mark Gerstein

mark.gerstein@yale.edu | 203-432-6105 | http://www.gersteinlab.org/

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 foot printing, 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. <u>mRNA structure dynamics identifies the embryonic RNA regulate</u>. Nature Structure Molecular Biology 2018
- Moreno-Mateos MA, Fernandez JP, Rouet R, Vejnar CE, Lane MA, Mis E, Khokha MK, Doudna JA, Giraldez AJ.<u>CRISPR- Cpf1 mediates efficient homology-directed repair and temperature-controlled genome editing.</u>Nature Commun. 2017
- Yartseva V, Takacs CM, Vejnar CE, Lee MT, Giraldez AJ.<u>RESA identifies mRNA-regulatory</u> sequences at high resolution.**Nature Methods.** 2017
- Bazzini AA, Del Viso F, Moreno-Mateos MA, Johnstone TG, Vejnar CE, Qin Y, Yao J, Khokha MK, Giraldez AJ. Codon identity regulates mRNA stability and translation efficiency during the maternalto-zygotic transition.EMBO J. 2016
- Bazzini AA^{‡#}, Johnstone TG[#], Giraldez AJ[‡]. Identification of smallORFs in animals using ribosome footprinting and evolutionary conservation. **EMBO J.** 2014.
- Lee MT[#], Bonneau AR[#], ... Giraldez AJ[‡]. Nanog, SoxB1 and Pou5f1/Oct4 regulate widespread zygotic gene activation during the maternal-to-zygotic transition. **Nature**, 2013.
- Stahlhut C, ... Giraldez AJ[‡]. miR-1/206 regulate angiogenesis by modulating Vegf-A expression. Development, 2012.
- Bazzini AA, Lee MT, Giraldez AJ[‡]. Ribosome Profiling Shows That miR-430 Reduces Translation Before Causing mRNA Decay in Zebrafish. Science 2012.
- Cifuentes D, Xue H, Giraldez AJ[‡]. A novel miRNA processing pathway independent of Dicer requires Argonaute2. Science. 2010.
- Mishima Y, ... Giraldez AJ. Zebrafish miR-1 and miR-133 shape muscle gene expression and regulate sarcomeric actin organization. Genes & Development. 2009

Jeffrey Gruen

jeffrey.gruen@yale.edu | 203-727-2202 | http://medicine.yale.edu/lab/gruen/index.aspx

Lab location: Yale Child Health Research Center (YCHRC), 464 Congress Avenue Lab meeting: Mondays, 9:30 a.m., YCHRC Conference Room, (Zoom meetings)

Rotations available anytime.

We use a multidisciplinary approach to study the **genetics of learning disabilities** including, reading disability (**dyslexia**) and attention deficit hyperactivity disorder

(ADHD), and the genes that enable language development called the human "lexinome". These are complex traits, meaning they are caused by a combination of genetic and environmental factors, but the genetic





components are quite large, accounting for 40% to 80% of the variance. Disorders of language, reading, and speech are very common. For example, **dyslexia is present in 10 to 20% of school children and is the most common cause of learning disability.**

(From: Truong et al, 2019)

We have three main areas of investigation:

1. Gene Discovery for Learning Disabilities

We deploy various methods for identifying common and rare variants including linkage analysis, genome wide association studies (GWAS), polygenic risk scores, whole exome sequencing, and whole genome sequencing. We study a range of subjects from extant and on-going studies including multigenerational families with learning disabilities, a case-control study of dyslexia in African Americans and Hispanic Americans (GRaD Study), Philadelphia Neurocognition Cohort (PNC), New Haven Lexinome Study (NHLP), the ABCD study, and several others large data sets.

2. Gene Function



⁽From: Li et al, 2018)

We approach gene function from two directions. Using a "top-down" approach (above left) we combine functional MRI imaging in awake subjects with genotyping/sequencing in **imaging-genetic** studies to determine the contribution of genetic variants to functional language and attention networks in children with learning disabilities and control subjects. We also use a "bottom-up" approach (above right) where we recapitulate genetic variants in **human embryonic stem cells** using **CRISPR**, differentiating them along neuronal cell lineages, and comparing changes in growth, expression, and electrophysiologic properties, as well as epigenetic changes.

3. Evolution of Language



(From: DeMille et al, 2018)

Languages evolve rapidly due to an interaction between sociocultural interactions and underlying phonological processes that are modified by population-genetics. By correlating the frequency of genetic variants with linguistic features (e.g., consonants and vowels), we are studying genetic factors that have influenced changes in over 50 languages distributed throughout the world, both at the population level and in the evolution of language (see DeMille et al, 2018).

Diversity and Inclusion

Our group is composed of individuals from diverse scientific and personal backgrounds who bring their talents, willingness to strive, and work ethic to all of our professional and social activities. We are committed to treating each other with respect, dignity, fairness, caring, equality, to help build and maintain each other's self-esteem, and to support each other in our collective achievements.

Mentoring and Commitment to Graduate Student Development

Our goal is to mentor students to develop their scientific and laboratory skills as well as other vital career skills required to succeed in diverse careers. We support students in my lab pursuing career training activities (e.g. teaching, writing) outside of lab. The average time for PhD in our lab is just over 5 years. PhD trainees go on to distinguished careers in academia, biotech, and pharma.

Current Lab Members

Andrew Adams: Graduate Student, Dept. of Genetics, "Investigating the influence of rare and common variation on reading disability"

Steven Paniagua: Graduate Student, Dept. of Genetics, "Modeling genes associated with reading and dyslexia in human embryonic stem cells"

Mellissa DeMille, PhD: Associate Research Scientist, "Evolutionary genetics of language" Nhu Truong, PhD: Associate Research Scientist, "Genome-wide analyses of reading endophenotypes"

Previous Graduate Students (degree, years in lab, present position)

Natalie Powers¹, PhD, 2008-2014, Post-Doctoral Fellow, Jackson Labs, Bar Harbor, ME **John Eicher**², PhD, 2009-2014, Post-Doctoral Fellow, NHLBI Cardiovascular, Epidemiology and Human Genetics Branch; Senior Scientist, Merck Pharmaceuticals, Cambridge, MA

¹ Recipient of the 2013 CW Cotterman Award from the American Society of Human Genetics ² NIH Ruth L. Kirschstein National Research Service Award

Recent Student Publications (Gruen lab students in bold)

- Truong DT, Adams AK, Paniagua S, Frijters JC, Boada R, Hill DE, Lovett MW, Mahone EM, Willcutt EG, Wolf M, DeFries JC, Gialluisi A, Francks C, Fisher SE, Olson RK, Pennington BF, Smith SD, Bosson-Heenan J, Gruen JR; Pediatric, Imaging, Neurocognition, and Genetics Consortium. Multivariate genome-wide association study of rapid automatised naming and rapid alternating stimulus in Hispanic American and African-American youth. *J Med Genet*. 2019 Apr 17.
- DeMille MMC, Tang K, Mehta CM, Geissler C, Malins JG, **Powers NR**, **Bowen BM**, **Adams AK**, Truong DT, Frijters JC, Gruen JR. Worldwide distribution of the *DCDC2* READ1 regulatory element and its relationship with phoneme variation across languages. *Proc Natl Acad Sci* U S A. 2018 May 8;115(19):4951-4956.
- Adams AK, Smith SD, Truong DT, Willcutt EG, Olson RK, DeFries JC, Pennington BF, Gruen JR. Enrichment of putatively damaging rare variants in the DYX2 locus and the reading-related genes CCDC136 and FLNC. *Human Genetics*. 2017; 136(11-12):1395-1405.
- **Powers NR**, **Eicher JD**, Miller LL, Kong Y, Smith SD, Pennington BF, Willcutt EG, Olson RK, Ring SM, Gruen JR. The regulatory element READ1 epistatically influences reading and language, with both deleterious and protective alleles. *J Med Genet*. 2016 Mar;53(3):163-71.
- Truong DT, Shriberg LD, Smith SD, Chapman KL, Scheer-Cohen AR, DeMille MM, Adams AK, Nato AQ, Wijsman EM, Eicher JD, Gruen JR. Multipoint genome-wide linkage scan for nonword repetition in a multigenerational family further supports chromosome 13q as a locus for verbal trait disorders. *Human Genetics*. 2016; 135(12):1329-1341.
- Eicher JD, Montgomery AM, Akshoomoff N, Amaral DG, Bloss CS, Libiger O, Schork NJ, Darst BF, Casey BJ, Chang L, Ernst T, Frazier J, Kaufmann WE, Keating B, Kenet T, Kennedy D, Mostofsky S, Murray SS, Sowell ER, Bartsch H, Kuperman JM, Brown TT, Hagler DJ Jr, Dale AM, Jernigan TL, Gruen JR; Pediatric Imaging Neurocognition Genetics study. Dyslexia and language impairment associated genetic markers influence cortical thickness and white matter in typically developing children. *Brain Imaging Behavior*. 2016 Mar;10(1):272-82.

Complete List of Published Work in MyBibliography:

http://www.ncbi.nlm.nih.gov/sites/myncbi/jeffrey.gruen.1/bibliography/40115979/public/?sort=date&direction=descending

George Hauser

ronald.hauser@yale.edu

Lab location: Room 216B, Building 35A, VA Lab meeting: Mondays at 9:00 a.m.

Rotations available anytime. Please email if interested.

Dr. Hauser's lab focuses on the development of innovative informatics methods to improve clinical decisions and enable large scale research. Methods employed in past projects include machine learning, graphical methods, data standardization, and web services. His lab operates in a consultative capacity, which leads to work with many groups in a variety of healthcare fields. For example, the lab works closely with the VA's national HIV program to provide better care for Veterans with HIV and with the Million Veterans Program, one of the world's largest programs on genetics and health. Dr. Hauser has a strong project management background, which allows students to publish one or more papers during a typical rotation. Due to the close integration of the student and the team, the lab can host only a single student at a time.

Monika Jadi

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

Lab location: Room 32B, Suite 901, 300 George Street Lab meetings: Thursdays at 1:30 p.m.

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 reveal the temporal dynamics of sensory information coding in a cell-type and layer specific manner.

Recent Publications:

- Veit, J., Hakim, R., Jadi, M.P., Sejnowski, T.J., Adesnik, H. (2017). Cortical gamma band synchronization through somatostatin neurons. Nature Neuroscience, In press
- Nandy, A.S., Mitchell, J.F., Jadi, M.P., Reynolds, J.H., (2016). Neurons in macaque Area V4 are tuned for complex spatio- temporal patterns. Neuron, 91(1-11)
- Coppola, J.J., Ward, N.J., Jadi, M.P., Disney, A. (2016). Modulatory compartments in cortex and the local regulation of cholinergic tone. J. Physiology, 110(1-2)
- Jadi, M.P., Behrens, M.M. & Sejnowski, T. J., (2015). Abnormal Gamma Oscillations in N-Methyl-D-Aspartate Receptor Hypofunction Models of Schizophrenia. Biological Psychiatry, 79(9).
- Jadi, M.P., & Sejnowski, T. J., (2014). Cortical oscillations arise from contextual interactions that regulate sparse coding. Proceedings of the National Academy of Sciences, 111(18).
- Jadi, M.P., & Sejnowski, T. J., (2014). Regulating cortical oscillations in an inhibition- stabilized network. Proceedings of the IEEE, 102(5).
- Jadi, M.P., Behabadi, B., Polsky, A., Schiller, J. & Mel, B. W., (2014). An augmented 2- layer model captures nonlinear analog spatial integration effects in pyramidal neuron dendrites. Proceedings of the IEEE, 102(5).
- Jadi, M.*, Polsky, A.*, Schiller, J. & Mel, B. W., (2012). Location-dependent effects of inhibition on local spiking in pyramidal neuron dendrites. PLoS Comput Biol, 8(6), e1002550. *=equal contribution

Samah Fodeh Jarad

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

Lab location: Room 518, Suite 501, 300 George Street **Lab meeting:** 2nd Friday of every month in the fall and the spring semesters

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 nonnegativity 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 the following research problems:

COVID-19 Research:

• Utilize social media data to estimate social distancing. Our recent work has shown that Twitter encapsulates invaluable information that allows for the estimation of different facets of SD in spatiotemporal settings14 including: purpose, implementation, social disruption, adaptation, and emotional response to SD as presented on the project's dashboard. https://samahfodehlab.github.io/ The plan is to extend this work and build predictive model of COVID-19 cases and deaths by either adjusting the compartmental SEIR transmission dynamics models or build hybrid predictive models based on statistics and machine learning.



• Develop multi-modal prediction models based on deep learning approaches that combine structured clinical variables, unstructured clinical text, and imaging data to enhance the prognosis of four COVID-19-related outcomes (levels of care) including: no admission, admission to hospital, intubation and mortality.

Phenotyping different types of headaches: The goal of this project is to characterize the different types of headaches and the associated pharmacological and non-pharmacological therapies. Many patients report headache symptoms and ways of treating those symptoms, as documented in their clinical notes, but do not receive an International Classification of Diseases (ICD) diagnostic code for their headaches. We utilize

natural language processing (NLP) and machine learning (ML) approaches to conduct a "deeper dive" into unstructured clinical notes to explore the entirety of the documented symptoms and treatments of headaches.

Linking social media and healthcare data: Despite their potential benefits, social media contents have been studied in isolation of the electronic medical records (EHR) of patients, making it difficult to fully understand important behavioral outcomes including fatal/nonfatal opioid overdose and suicide. Sharing social media data for research purposes has been shown to be possible when patients consented to share their social media activity when approached at a Yale addiction recovery clinic. The goal of this project is to compose a combined database on patients by collecting their social media content and link it to their clinical information from the EHR.

Characterizing communication between patients and healthcare providers: Effective and rapid Patient-Provider Communication (PPC) is key to improve population health management and patient-centered outcomes. Timely communication makes it easier to closely monitor patients' adherence and swiftly react to changes in health. In this project, using machine learning and natural language processing, we will automate the process of extracting patterns of communications such as symptoms, adverse events, medications, emotions, and expressions of empathy.

Exploiting Medline articles for gene molecular function prediction: The goal is to develop predictive models to automatically assign molecular functions to genes using the biomedical literature. We have shown in previous work that PubMed abstracts can be used to classify genes based on functionality using a multi-label classification approach. Our method performed based compared to existing models. We aim to enhance our classification model.

Motivated students are always welcome to join the lab at any time of the year.

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 stateof-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:

- a) Efficient Drug Lead Discovery and Optimization. Jorgensen, W. L. Acc. Chem. Res. 2009, 42, 724-733.
- b) Advances in QM/MM Simulations for Organic and Enzymatic Reactions. Acevedo, O.; Jorgensen, W. L. Acc. Chem. Res. 2010, 43, 142-151.
- c) Computer-Aided Discovery of Anti-HIV Agents. Jorgensen, W. L. *Bioorg. Med. Chem.* 2016, 24, 4768-4788.

Amy Justice

amy.justice2@va.gov | 203-932-5711 x3541

Lab location: 950 Campbell Avenue, Ste Building 35a **Lab meeting:** Fridays at noon in the 2nd floor conference room of Building 35a

Rotations available in the spring term; please contact Angela Consorte

The Justice lab combines national VA electronic medical record data, death index data, CMS data, biomarker data including genetic and omic data and other related databases to create large cohorts for in-depth longitudinal epidemiologic study. Clinical domains of interest include HIV, HCV, cancer, multi-morbidity, poly substance use, polypharmacy, and life expectancy. Analytic methods include conventional epidemiology and causal methods to artificial intelligence and machine learning. Our lab collaborates with a number of Department of Energy National Laboratories and has access to their expertise and high-powered computing. There are opportunities for extended visits to one or more of these labs as well.

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

Steven Kleinstein

steven.kleinstein@yale.edu | 203-785-6685 | https://medicine.yale.edu/lab/kleinstein | @skleinstein

Lab location: Suite 505, 300 George Street

Lab meeting: Wednesdays, 2:30-4:00 p.m., Zoom (email if interested in attending)

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 (COVID-19, 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: <u>https://medicine.yale.edu/lab/kleinstein</u>

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:

- Lindenbaum, O., Stanley, J., Wolf, G., & Krishnaswamy, S. (2018). Geometry Based Data Generation. In Advances in Neural Information Processing Systems (pp. 1407-1418).
- Amodio, M., & Krishnaswamy, S. (2018). Magan: Aligning biological manifolds. In Proceedings of the 35th Annual International Conference on Machine Learning.
- Van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A. J., ... & Krishnaswamy, S.* & Pe'er, D.* (2018). Recovering gene interactions from single-cell data using data diffusion. Cell, 174(3), 716-729. (*equal contribution, co- corresponding)
- Amodio, M., & Krishnaswamy, S. (2019). TraVeLGAN: Image-to-image Translation by Transformation Vector Learning. arXiv preprint arXiv:1902.09631 in Proceedings of the Conference on Computer Vision and Pattern Recognition, 2019.
- Amodio, M., Van Dijk, D., Srinivasan, K., Chen, W. S., Mohsen, H., Moon, K. R., ... & Krishnaswamy, S. (2019). Exploring single-cell data with deep multitasking neural networks. BioRxiv, 237065. to appear in Nature Methods 2019.

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.

Bluma Lesch

bluma.lesch@yale.edu | 203-737-1074 | leschlab.org

Lab location: SHM I-141A

Lab meeting: Mondays at 2:00 p.m., SHM I-143 (virtual until further notice; contact Bluma for zoom link)

Rotations available anytime.

Regulation and evolution of chromatin in development

How does chromatin evolution impact evolution of gene expression and phenotype? We are tackling this huge, complex, and understudied problem 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 how chromatin state evolves and how it relates to evolution of gene expression.

Germ cells and gametes are critically important to understanding 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 a chromatin state called poising, defined by the simultaneous presence of activating and repressive



histone modifications and thought to 'poise' developmental regulatory genes for expression.

We are looking for experimental and computational students enthusiastic about gene regulation, evolution, and/or reproduction and excited by 'big questions' in biology.

Some ongoing projects:

- *Testing the significance of evolutionary divergence in chromatin state* in vitro. We have identified hundreds of loci at which chromatin state in the *in vivo* germ line differs between mammalian species (Lesch et al, *Nat Genet* 2016). We are using mouse cell lines containing "swapped" regulatory sequence from other species as a tractable system to test the biological significance of these changes.
- **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, allowing us to add or remove specific histone modifications at specific loci.
- *Molecular mechanisms of epigenetic inheritance.* We recently found that knocking out a specific chromatin regulator in the male germ line results in increased cancer rates in genetically wild type offspring (Lesch et al, *eLife* 2019). We are now looking more closely *in vivo* to define the molecular mechanism underlying this effect. At the molecular level, we are exploring the role of H3K4 monomethylation in preservation of the poised state across cell divisions (Bae & Lesch, *Front Cell Dev Biol* 2020).
- **Defining the dynamics of enhancer evolution in the mammalian germ line.** We are using our rich dataset of published and unpublished functional genomics data in germ cells from multiple mammalian species to analyze patterns of enhancer evolution in the germ line.

Key publications:

- S. Bae and B.J. Lesch. H3K4me1 predicts transcription state and poising at promoters. *Front Cell Dev Biol.* 8:289 (2020).
- B.J. Lesch, Z. Tothova, E.A. Morgan, Z. Liao, R.T. Bronson, B.L. Ebert, D.C. Page. Intergenerational epigenetic inheritance of cancer susceptibility in mammals. *eLife*. 8 (2019).
- B.J. Lesch, S.J. Silber, J.R. McCarrey, D.C. Page. Parallel evolution of male germline epigenetic poising and somatic development in animals. *Nat Genet.* 48, 888 (2016).
- B.J. Lesch, G.A. Dokshin, R.A. Young, J.R. McCarrey, D.C. Page. A set of genes critical to development is epigenetically poised in mouse germ cells from fetal stages through completion of meiosis. *PNAS* 110, 16061 (2013).

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.
- 4. Training multi-omics predictors of biological age and disease risk in humans.

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 | 203-785-3478 | http://www.zachlevine.com/

Lab location: Room 320, Bass Center, 266 Whitney Avenue Lab meeting: Tuesdays at 11:00 a.m. in Bass Center/Zoom

Rotations available anytime. Please email if you are interested.

Computational and Experimental Biophysics of Amyloid Diseases



A central dogma in protein folding is that structure dictates function, however more than a third of the human proteome lacks a stable threedimensional structure. Instead, disordered proteins adopt transient structures in a wide variety of environments, making them versatile workhorses in human biology. Despite this versatility, over 22 major diseases are routinely associated with the aggregation of disordered proteins such as Alzheimer's Disease, Type II Diabetes, and Parkinson's Disease, which are all known as amyloid disorders. In addition, tumorigenesis and premature cellular aging can also occur when disordered proteins are unregulated over the life course, highlighting their central role in most degenerative diseases.

The goal of my lab, broadly, is to deduce how heterogeneous amyloids cause degenerative diseases through a combination of biophysical modeling and fluorescence microscopy techniques. This involves combining computational and experimental observations at both the single-molecule and macroscopic level. Students will have unique opportunities to learn about both molecular dynamics simulations and single-molecule FRET (smFRET) experiments, which can be easily combined to reveal the heterogeneities of protein disorder and function. Currently, our lab boasts the largest number of dedicated computing resources on campus and houses a dedicated smFRET system that is one-of-a-kind at Yale. This enables us to study protein dynamics in-silico, in-solution, in-vitro, and in tissues, which yields a comprehensive understanding of human disease from the ground up

Rotation projects in our lab include (but are not limited to):

- 1. Investigate how mutations in p53 affect their ability to fold, function, and bind to DNA. Aggregation of mutated tumor suppressor proteins, such as p53, are a common hallmark in many cancers, however little is known about their pathological aggregation propensities.
- 2. Model and experimentally test how the preparation of lab-derived amyloids in-solution compares to patient-derived protein structures. This would answer an important question of whether protein aggregates in the lab "look" like real pathological aggregates, or if these structures are contrived.
- **3.** Deduce the solubility of individual protein oligomers and aggregates. This is easy to do in large protein ensembles but challenging for individual protein oligomers that interconvert between physiological and pathological states.
- 4. Study the intersection between protein phase separation (LLPS) and the formation of pathogenic granules in cells.
- 5. Quantify how soluble amyloids stress or senesce cells over the life course. This would help reveal protein-based determinants of human aging and longevity.

I also value diversity in our lab, and strongly encourage researchers from unique backgrounds to get in touch with us. No prior experience in molecular modeling or fluorescence techniques is required to rotate with us. Instead, we emphasize the importance of learning and synergizing different methods together to paint a comprehensive picture of human biology and disease. Students who join the lab will have multiple opportunities to present at conferences, workshops, and seminars. Similarly, each member will have access to faculty in both basic research departments (e.g. MB&B) and in the School of Medicine.

Haifan Lin

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

Lab location: Room 237, 10 Amistad Street 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.

Rotations are available anytime.

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. Overproliferation 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 peeded. When these receptors are dwaregulated, they induce outcimmune cells where they are peeded.

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

- Pantouris G, Bucala R, Lolis EJ. Structural Plasticity in the C-Terminal Region of Macrophage Migration Inhibitory Factor-2 Is Associated with an Induced Fit Mechanism for a Selective Inhibitor. Biochemistry. 2018 Jul 3;57(26):3599-3605.
- Pantouris G, Ho J, Shah D, Syed MA, Leng L, Bhandari V, Bucala R, Batista VS, Loria JP, Lolis EJ. Nanosecond Dynamics Regulate the MIF-Induced Activity of CD74. Angew Chem Int Ed Engl. 2018 Jun 11;57(24):7116-7119.
- Identification of an Arg-Leu-Arg tripeptide that contributes to the binding interface between the cytokine MIF and the chemokine receptor CXCR4. Lacy M, Kontos C, Brandhofer M, Hille K, Gröning S, Sinitski D, Bourilhon P, Rosenberg E, Krammer C, Thavayogarajah T, Pantouris G, Bakou M, Weber C, Lolis E, Bernhagen J, Kapurniotu A. Sci Rep. 2018 Mar 26;8(1):5171.
- Gil-Yarom N, Radomir L, Sever L, Kramer MP, Lewinsky H, Bornstein C, Blecher-Gonen R, Barnett-Itzhaki Z, Mirkin V, Friedlander G, Shvidel L, Herishanu Y, Lolis EJ, Becker-Herman S, Amit I, Shachar I. CD74 is a novel transcription regulator. Proc Natl Acad Sci U S A. 2017 Jan 17;114(3):562-567

- Rajasekaran D., et al. Macrophage Migration Inhibitory Factor-CXCR4 Receptor Interactions: EVIDENCE FOR PARTIAL
- ALLOSTERIC AGONISM IN COMPARISON WITH CXCL12 CHEMOKINE. J Biol Chem. 2016 Jul 22;291(30):15881-95.
- Pantouris G, Syed MA, Fan C, Rajasekaran D, Cho TY, Rosenberg EM Jr, Bucala R, Bhandari V, Lolis EJ. An Analysis of MIF Structural Features that Control Functional Activation of CD74. Chem Biol. 2015 Sep 17;22(9):1197-205.

Jun Lu

jun.lu@yale.edu | 203-737-3426

Lab location: AMISB 237C

Lab meeting: You are welcome to participate in our lab meetings. Fridays, 1:00-3:30 p.m. via Zoom, Please email Jun Lu if you are interested in attending.

Rotations are available any time. If you are interested in seeing the lab, we can show you either online or in person (depending on the COVID situation, with safety measures in mind).

Overview:

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?

Rotation Projects:

1. Molecular control of organelle shape (experimental and/or computational)

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 with computational analysis to classify 3D nuclear shapes. A related computational project aims to use deep learning to classify blood cells and blood diseases and understand the relationship between organelle shape and genetic mutations in disease.

2. The epigenetic control of anti-cancer immunity (experimental and/or computational)

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 pathogen and cancer, with a focus on genetic alterations in hematopoietic stem cells that accumulate through aging. A related project aims to computationally identify drugs that improve anti-cancer immune responses utilizing large genomic datasets.

3. Mapping and visualizing of functional noncoding elements in mammalian genomes (experimental) 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.

Robert McDougal

robert.mcdougal@yale.edu | 203-737-4828

Lab location: Room 526, Suite 501, 300 George Street

Lab meetings: Most Wednesdays at 1:00 p.m.; email to confirm.

Rotations available any time.

The McDougal lab develops and applies informatics and simulation approaches to gain insight into biomedical phenomena, especially multiscale phenomena involving the brain. We are especially interested in deriving information from neuroscience artifacts (models and papers) and in improving the practice, rigor, and capabilities of simulation studies.

Potential project areas include:

1. Applying machine learning to neuron simulation

Whether or not a neuron generates action potentials or "spikes" is a complex response to patterns of input from other neurons. Multiple excitatory inputs -- inputs that bring the neuron closer to a firing "threshold" -- within a short period of time might trigger a spike, but perhaps surprisingly sustained inhibition -- holding the cell farther below threshold than when at rest -- can also lead to a spike. For decades, scientists have used calculus-based simulations to predict how neurons respond to input, but this requires a significant amount of computer processing power. This project seeks to use the power of machine learning and large datasets to develop an alternative prediction approach.

2. Biochemistry simulation

We develop simulation techniques to gain insight into multiscale biological function. From the perspective of physics, chemical activity within a cell is governed primarily by two fundamental actions: diffusion and reaction. If concentrations are high enough, the chemical processes can be modeled using deterministic methods, but low conentrations introduce stochastic effects that when combined with other nonlinear phenomena within the cell can introduce new behaviors not seen in the deterministic case. For example, a calcium gated potassium channel may on average keep a neuron just below threshold for firing, but stochastic effects might allow the cell to fire anyways. How is the brain able to robustly function despite this variation? Join us as we learn more.

3. Biomedical literature mining

Biomedical data is being generated at a rapid rate, but to me most effective at guiding future research, this data must be easily discoverable and understood in context. We are developing text-mining techniques to semi-automatically monitor and curate the biomedical literature as well as databases for analyzing, visualizing, and disseminating the derived data. We are especially interested in trends within the Alzheimer's, computational neuroscience, and covid-19 literature.

4. Ischemia modeling

Stroke is one of the leading causes of death in the United States. We are studying ischemic stroke, the most common kind, where blood flow to a region of the brain is cut off due to a blocked blood vessel. Even after blood flow is restored, the brain continues to lose neurons as dead and dying cells create a toxic environment for other neurons. We are building a computational model to study this dangerous chain reaction, with the hope of gaining insights that could ultimately suggest ways to interrupt the process and minimize brain loss.

Kathryn Miller-Jensen

kathryn.miller-jensen@yale.edu | 203-432-4265

Lab location: MEC 303

Lab meeting: Time still TBD for fall semester. Contact Kathryn if interested.

Rotations available any time.

Research Interests

Our lab uses quantitative, systems-level approaches to study intracellular and extracellular signaling networks regulating immune cell functions. We are particularly interested in how non-genetic variability between cells impacts immune responses. We use imaging and microfluidic-based approaches for multiplexed single-cell measurements of signaling, transcription, and secretion, combined with traditional cell-population assays, and we use a range of computational approaches to analyze data. Current research projects in the lab focus on the regulation of macrophage activation heterogeneity in response to pathogenic and resolving stimuli at the signaling, transcription, and epigenetic level. We also have a project that focuses on studying cell-to-cell communication networks between immune cells and cancer cells in the melanoma tumor microenvironments.

Rotation opportunities include both experimental and computational projects to explore a range of questions. **Some examples include:**

- 1. How do macrophages choose between competing environmental cues? We are using several approaches to determine the specificity and durability of combinations of stimuli that activate distinct functional cells states, both in vitro and in vivo.
- 2. What are the changes in immune cell networks that are associated with growing melanoma tumors versus tumors that respond to cancer immunotherapy? We are using machine learning methods to build cell-cell communication networks from single-cell RNA-sequencing data and then testing these experimentally.

Mentoring and Career Training

My group offers an interdisciplinary environment with training in experiments, computation and engineering. In addition to mentoring students to develop as scientists, I encourage them to develop other skills required for them succeed in the diverse careers that interest them. In the past, students in my lab have pursued a range of career training activities outside of lab, including teaching, scientific outreach, and entrepreneurial pursuits.

John Murray

john.murray@yale.edu | 203-737-2382 | <u>http://murraylab.yale.edu/</u> Lab location: Suite 6E, 40 Temple Street

Rotations available any time.

Research in our computational lab focuses on investigating the dynamics and function of neural circuits, and their dysfunction in psychiatric and neurological disorders, through computational modeling, theoretical, and data-analytic approaches to systems neuroscience, in close collaboration with experimentalists.

Rotation projects can include:

- 1. Analysis of human neuroimaging data (fMRI, DWI, MEG), in relation to computational models (see: Demirtas et al., 2019, *Neuron*)
- 2. Analysis of gene expression data from the human brain and their whole-brain topographic patterns, and integration with computational modeling (see: Burt et al., 2018, *Nature Neuroscience*).
- 3. Analysis of neuronal spike-train recordings collected from awake behaving animals, in relation to models.

- 4. Computational modeling of neural circuit dynamics, from biophysically detailed microcircuits to large-scale whole-brain networks.
- 5. Applications of machine learning in analysis of complex neural datasets, and as models of cognitive computations in the brain (e.g. trained artificial recurrent neural networks)

Relevant publications:

- Burt JB, et al. (2018) Hierarchy of transcriptomic specialization across human cortex captured by structural neuroimaging topography. *Nature Neuroscience* 21:1251
- Demirtas M, et al. (2019) Hierarchical heterogeneity across human cortex shapes large-scale neural dynamics. *Neuron* 101:1181
- Lam NH, et al. (2017) Effects of altered excitation-inhibition balance on decision making in a cortical circuit model. *bioRxiv* doi:10.1101/100347
- Murray JD, et al. (2017) Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proceedings of the National Academy of Sciences* 114:394

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: Mondays 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 knowledgebased 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 lowenergy 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.



Caption: An example of a protein structure prediction submitted to CASP Commons for the Sar-CoV-2 protein nsp4. On a scale of 0 to 1, the average decoy deception score for this prediction was 0.49.

Publications:

- Q. Zhou, C. S. O'Hern, and L. Regan, "The power of hard-sphere models: Explaining side-chain dihedral angle distributions of Thr and Val," Biophys. J. 102 (2012) 2345.
- Q. Zhou, C. S. O'Hern, and L. Regan, ``Revisiting the Ramachandran plot from a new angle," Protein Science 20 (2011) 1166.
- Q. Zhou, D. Caballero, C. S. O'Hern, and L. Regan, "New insights into the interdependence between amino acid stereochemistry and protein structure," Biophys. J. 105 (2013) 2403.
- Caballero, J. Maatta, A. Q. Zhou, M. Sammalkorpi, C. S. O'Hern, and L. Regan, ``The intrinsic α-helical and β-sheet preferences: A computational case study of Alanine," Protein Science 23 (2014) 970.
- Q. Zhou, C. S. O'Hern, and L. Regan, ``Predicting the side-chain dihedral angle distributions of nonpolar, aromatic, and polar amino acids using hard-sphere models," Proteins: Structure, Function, and Bioinformatics 82 (2014) 2574.
- D. Caballero, W. W. Smith, C. S. O'Hern, and L. Regan, ``Equilibrium transitions between sidechain conformations in leucine and isoleucine," Proteins: Structure, Function, and Bioinformatics 83 (2015) 1488.
- J. C. Gaines, W. W. Smith, L. Regan, and C. S. O'Hern, ``Random close packing in protein cores," Phys. Rev. E 93 (2016) 032415.
- Virrueta, C. S. O'Hern, and L. Regan, ``Understanding the physical basis for the side chain conformational preferences of Met," Proteins: Structure, Function, and Bioinformatics 84 (2016) 900.
- D. Caballero, A. Virrueta, C. S. O'Hern, and L. Regan, ``Steric interactions alone determine side chain conformations in protein cores," Protein Eng. Des. Sel. 29 (2016) 367.
- J. C. Gaines, A. Virrueta, D. A. Buch, S. J. Fleishman, C. S. O'Hern, and L. Regan, "Collective repacking reveals that the structures of protein cores are uniquely specified by steric repulsive interactions," Protein Eng. Des. Sel. 30 (2017) 387.

- J. C. Gaines, A. H. Clark, L. Regan, and C. S. O'Hern, "Packing of Protein Cores," J. Phys.: Condensed Matter 29 (2017) 293001.
- J. C. Gaines, S. Acebes, A. Virrueta, M. Butler, L. Regan, and C. S. O'Hern, "Comparing side chain packing in soluble proteins, protein-protein interfaces, and transmembrane proteins," Proteins: Structure, Function, and Bioinformatics 86 (2018) 581.
- J. D. Treado, Z. Mei, L. Regan, and C. S. O'Hern, "Void distributions reveal structural link between jammed packings and protein cores," Phys. Rev. E 99 (2019) 022416.

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.



Caption: Top: A polymer acted on by an active pulling agent (green) remains in a relatively extended state for a Peclet number, $Pe \approx 0.1$ (top right), and collapses for $Pe\approx 100$ (top left). Bottom: The distribution of the radius of gyration R_g of polymers of 200 monomers (with diameter b) acted upon by increasingly strong tangent forces. For increasing Pe, the width of the distribution narrows and the mean shifts to smaller values of R_g .

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.



Caption: The upper experimental image shows a slice from breast cancer tissue. The tumor cells are stained, and the adipocytes are unstained. The lower image shows a snapshot from numerical simulations that model cancer cells (black circles) invading adipocytes shaded by the value of the cell's shape factor.

Publications:

- Boromand, A. Signoriello, F. Ye, C. S. O'Hern, and M. D. Shattuck, "Jamming of deformable polygons," Phys. Rev. Lett. 121 (2018) 248003.
- Boromand, A. Signoriello, J. Lowensohn, C. S. Orellana, E. R. Weeks, F. Ye, M. D. Shattuck, and C. S. O'Hern, "The role of deformability in determining the structural and mechanical properties of jammed packings of bubbles and emulsions," Soft Matter (2019); doi: 10.1039/c9sm00775j

Lajos Pusztai

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

Lab location: Room 133, Suite 120, 300 George Street (computational biology group) Lab meeting: Thursdays from 3:00-5:00 p.m. in Room 130, 300 George Street

Rotations available any time.

The Pusztai Lab focuses on breast cancer translational research. We work with human breast cancer tissues and preclinical laboratory cell models of breast to discover new therapeutic targets and understand the biology of this disease. The group includes wet bench researchers, computational biologists, and clinicians in various levels of training.

The four major current projects include:

- 1. Characterizing and studying the immune microenvironment of breast cancer and use this information to design new clinical trials.
- 2. Develop a new class of anti-cancer agents that target cancer-dominant metabolic enzymes.
- **3.** Optimize our lead compound to generate a novel antibody drug conjugate that targets the aberrantly expressed GABRP cell surface receptor subunit in breast cancer.
- 4. Study the complimentary functional interactions between the large number of proteins altering germline single nucleotide polymorphisms, that we all carry, and acquired somatic mutations.

Students can find recent publications of the lab at: <u>https://scholar.google.com/citations?hl=en&user=lHogWesAAAAJ&view_op=list_works&sortby=pubdate</u>

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.
- 3. Hepatitis C replication mechanisms: role of genomic RNA structure and viral proteins.
- 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

hemant.tagare@yale.edu | 203-737-4271 | 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:

- Tagare H. D., Kucukelbir A., Sigworth F. J., Wang H., Rao M., "Directly reconstructing principal components of heterogeneous particles from cryo-EM images", J. Struct. Biol., 191(2):245-62, 2015.
- Dvornek N. C., Sigworth F. J., Tagare H. D., "SubspaceEM: A fast maximum-a-posteriori algorithm for cryo-EM single particle reconstruction", J. Struct. Biol., 190(2):200-14, 2015.
- Kucukelbir A., Sigworth F. J., Tagare H. D., ``Quantifying the local resolution of cryo-EM density maps'', Nature Methods, 11(1):63-5, 2013.



Heterogeneity in the RNA-dependent RNApolymerase 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

richard.taylor@yale.edu | 203-909-3012

Lab location: Room 524, Suite 501, 300 George Street Lab meeting: Tuesdays and Thursdays

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 of 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.

Jeffrey P. Townsend

jeffrey.townsend@yale.edu | 203-737-7042 Lab location: Room 222, Suite 200, 135 College Street Lab meeting: Thursdays at 10:15 a.m.

Rotations available anytime.

Professor Townsend's research interests include cancer genetics and evolution, bioinformatics, mathematical modeling, and gene expression analysis. Research topics include the estimation of selection coefficients on mutations within genes and pathways that underly the somatic evolution of cancer, outwitting the evolution of drug resistance, and global sensitivity and uncertainty analyses of dynamic infectious disease models. Professor Townsend also performs research on the evolution of gene expression using pathogenic and non-pathogenic fungi, including the model filamentous fungus Neurospora crassa.

Potential projects include:

- 1. Analyze the genetic architecture of cancers; predict the fitness landscape of diverse cancer tumors in response to targeted drug therapies; improve treatment.
- 2. Perform mathematical modeling and probabilistic uncertainty analysis for infectious disease or systems biology.
- 3. Develop powerful new computational, bioinformatic, or probabilistic approaches to the characterization of the molecular evolution of genes, their phylogenetic informativeness, or the level of selection on individual DNA sites in a gene or genome or amino acid sites in a protein or proteome.
- 4. Discover a gene differentially expressed during fruiting body development of species of Neurospora and Fusarium; knock it out; phenotype the resulting strain; identify the function of the gene

Serena Tucci

serena.tucci@yale.edu | https://campuspress.yale.edu/stucci/ Lab location: Room 216, 10 Sachem Street Lab meeting: Tuesdays at 12:00 p.m.

Rotations available anytime.

Dr. Tucci's research addresses fundamental questions in human evolution and population history using DNA from present-day and ancient humans. Her interdisciplinary approach combines expertise from anthropology, population genetics, and computational biology, to reconstruct past demographic events and disentangle the genetic basis of human adaptation. By integrating field work, laboratory work and cutting-edge computational methods, her work sheds light on mechanisms of evolutionary change, and on the genetic legacy that extinct humans - such as Neandertals and the enigmatic Denisovans - left in the genomes of human populations in Island Southeast Asia and Oceania.

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, 1:00 p.m., 300 George Street / Zoom. Visit <u>www.vandijklab.org</u> 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. 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:

COVID-19: In recent months, the Van Dijk lab has become one of the main computational labs at Yale working on SARS-CoV-2 and COVID-19 (<u>https://medicine.yale.edu/news-article/24911/</u>). We have been working on a wide range of projects, including single-cell analysis of airway epithelium (<u>https://www.biorxiv.org/content/10.1101/2020.05.06.081695v1</u>), electronic health records (<u>https://www.medrxiv.org/content/10.1101/2020.05.07.20094573v2</u>, <u>https://onlinelibrary.wiley.com/doi/full/10.1002/emp2.12145</u>), and immunoprofiling of COVID patients (<u>https://arxiv.org/abs/2006.12971</u>) using scRNA-seq and large-scale cytokine panels. We have collaborations with all the top virologists, clinicians, and immunologists at Yale and have access to the best possible data for studying COVID-19.

Single-cell disease state prediction: 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, and neurodegeneration in order to train our algorithms to detect and understand the differences between healthy and diseased cell types and cell states. Recently, we have developed several such algorithms using Graph Neural Networks (<u>https://arxiv.org/abs/2006.12971</u>, https://arxiv.org/abs/2002.07128).

Neuroscience (wide-field calcium imaging): Together with the Cardin Lab we are developing algorithms to discover patterns in wide-field calcium imaging data. We are using these algorithms to correlate behavior of mice to their cortical activation patterns in order to understand how behavior is encoded in the cortex and to predict behavior from cortical activation.

Machine learning / Deep learning: In addition to biomedical applications of machine learning, we also contribute to the machine learning community by developing new algorithms that we publish at computer science conferences such as ICML, NeurIPS, etc. We work on topics such as: Deep Learning, Graph Neural Networks, Natural Language Processing, Physics Machine Learning, and much more! See arxiv for some of our CS manuscripts: <u>https://arxiv.org/search/cs?searchtype=author&query=van+Dijk%2C+D</u>

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.

Gut Microbiome / Molecule discovery: Together with the Palm Lab at Immunobiology we are developing algorithms to understand the complex interactions between the host and its microbiome. In addition, we are using Machine Learning to discover new compounds, from microbial genes, that can be used as therapeutics.

Natural Language Processing: We are excited about the developments in deep learning and natural language processing and we are contributing to this field by developing algorithms that can do machine translation (<u>https://arxiv.org/abs/2006.11578</u>). Ultimately we would like to apply these methods to biomedical literature.

We publish in top biomedical journals such as Cell, Nature Biotech, etc. -- e.g.: <u>Van Dijk, David, et al.</u> <u>"Recovering gene interactions from single-cell data using data diffusion." *Cell* 174.3 (2018): 716-729. Feel free to contact us at: david.vandijk@yale.edu -- We are always excited to have new students in the lab and we have lots of data for you to play with!</u>



Anita Wang

zuoheng.wang@yale.edu | 203-737-2672

Lab location: 60 College Street, Yale School of Public Health **Lab meeting:** Tuesdays, 1:00-3:00 p.m.

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.

Jack (Kai) Zhang

jack.zhang@yale.edu | 203-436-4902 Lab location: Bass 322

Rotations available any time

We work on the molecular mechanisms of cellular cargo transport and ciliary motility by a combination of a variety of techniques, spanning biophysics, biochemistry, molecular biology, and computational techniques and mathematical modelling. Particularly, we are interested in dyneins, and dynein-related protein complexes for their fascinating roles in the intracellular and intraflagellar transports, cell division, cell polarity, organelle positioning, cell motility, neurodevelopment, neurodegeneration, viral infection, extracellular environment sensation, and their elaborate mechanisms of regulation. The Zhang lab also develops methods on cryoelectron microscopy and tomography, aiming to visualize highly dynamic protein complexes as vast interaction networks in cells at the atomic level to finally allow a profound understanding of high-level cellular activities from their structural basis.

In the past, we have revealed the mechanisms of how cytoplasmic dynein-1 is autoinhibited and activated by its cofactors in unprecedented details (*Science*, 347 (6229): 1441-1446, Cover; *Cell*, 169: 1303–1314). Recently, by developing our own methods, we determined the cryo-EM/ET structures of several fundamentally important components of cilia and revealed the mechanisms of beating regulation in atomic details (~10-fold resolution improvement, *to be published*). One of our methods (JSB, 193:1-12) on cryoEM/ET has been cited for >1000 times within a couple of years and many new methods are being developed to solve much more challenging problems in the cryo-EM/ET field.



Rotation opportunities:

- 1. Viral transport: to over-express co-factors and/or adaptors that involve in dynein-mediated viral transport;
- 2. Disease mutations: to over-express disease-related mutant dyneins;
- 3. Dynein motility: to test how the co-factors, adaptors and cargos affect dynein motility using singlemolecule imaging;
- 4. The ciliary beating mechanism: to work on the structures of outer dynein arms (ODA) and their interaction with microtubule doublets (MTD) and docking complexes (DC);
- 5. Environmental signal sensing by cilia: to test our hypotheses on how cilia sense and response to their environment;
- 6. Cryo-EM data processing: to learn and practice 3D reconstruction on basic and challenging cryo-EM/ET datasets;
- 7. Structural analysis: analyze the many highly complicated structures determined by the Zhang lab and their interaction networks in atomic details to reveal their roles in ciliary motility;
- 8. Computation: write simple programs and/or scripts to improve the 2D and 3D classification, alignment, refinement, special particle recognition, cryo-EM image contrast enhancement, cryo-EM image quality diagnosis, weak signal detection, CTF refinement, and structural geometry optimization etc. (whichever is suitable).

NOTE: students learn directly from the PI and postdocs; the PI deeply works together with all his postdocs and students!

Hongyu Zhao

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

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

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.
- **3.** Single cell analysis.
- 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:

- Q. Lu, R. L. Powles, S. Abdallah, D. Ou, Q. Wang, Y. Hu, Y. Lu, W. Liu, B. Li, S. Mukherjee, P. K. Crane, H. Zhao (2017) Systematic tissue-specific functional annotation of the human genome highlights immune-related DNA elements for late onset Alzheimer's disease. PLOS Genetics, 13: e1006933.
- Q. Lu, B. Li, D. Ou, M. Erlendsdottir, R. L. Powles, T. Jiang, Y. Hu, D. Chang, C. Jin, W. Dai, Q. He, Z. Liu, S. Mukherjee, P. K. Crane, H. Zhao (2017) A powerful approach to estimating annotationstratified genetic covariance using GWAS summary statistics. American Journal of Human Genetics, 101: 939-964.
- Y. Hu, Q. Lu, R. Powles, X. Yao, C. Yang, F. Fang, X. Xu, H. Zhao (2017) Leveraging functional annotations in genetic risk prediction for human complex diseases. PLOS Computational Biology, 13: e1005589.
- Y. Hu, M. Li, Q. Lu, H. Weng, J. Wang, S. M. Zekavat, Z. Yu, B. Li, J. Gu, S. Muchnik, Y. Shi, B. W. Kunkle, S.
- Mukherjee, P. Natarajan, A. Naj, A. Kuzma, Y. Zhao, P. K. Crane, Alzheimer's Disease Genetics Consortium, H. Lu, H. Zhao (2019) A statistical framework for cross-tissue transcriptome-wide association analysis. Nature Genetics, 51: 568576.
- D. Tang, S. Park, H. Zhao (2020) NITUMID: Nonnegative Matrix Factorization-based Immune-TUmor MIcroenvironment Deconvolution. Bioinformatics, 36: 1344-1350.
- S. Park, H. Xu, H. Zhao (2020) Integrating multidimensional data for clustering analysis with applications to cancer patient data. Journal of American Statistical Association, in press.
- K-Y Lee, T. Liu, B. Li, H. Zhao (2020) Learning causal networks via additive faithfulness. Journal of Machine Learning Research, in press.

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.