Stowers Institute

Stowers.org

STOWERS RESEARCH CONFERENCES:

Development, Regulatory Genomics & AI

#SRCKC26

PROGRAM

Table of Contents

Speaker instructions

Speakers **must upload their presentations to the "Presenter" folder that was shared with them in advance of the meeting OR at the podium** at the designated times listed below, prior to their scheduled sessions.

Speaker file upload times:

Wednesday, April 22
1:00-2:00PM

Thursday, April 23
9:00–9:30 AM
1:30–2:00 PM

Friday, April 24
9:00–9:30 AM
1:30–2:00 PM

Speakers **must get equipped with a microphone at the podium** 15 minutes prior to their scheduled sessions.

Conference Program

All talks will be presented in the Stowers Auditorium
Food and beverages offered are complementary, please enjoy!

25 min presentation + 10 min Q&A
10 min presentation + 5 min Q&A
45 min presentation + 15 min Q&A

WEDNESDAY, APRIL 22

12:30 PM TO 2:00 PM	**CONFERENCE CHECK-IN & POSTER SET-UP** *Light snacks and refreshments are available outside of the Auditorium*

SESSION 1: TRANSCRIPTIONAL MECHANISMS

2:00 PM	**OPENING REMARKS**
2:05 PM TO 2:40 PM	**WILLIAM GREENLEAF & BENJAMIN DOUGHTY, STANFORD UNIVERSITY** *"Single-molecule states link transcription factor binding and expression in human cells"*
2:45 PM TO 3:00 PM	**XIN HE, UNIVERSITY OF CHICAGO** *"Combining natural genetic variations with single molecule footprinting to learn about the rules of chromatin regulation"*
3:05 PM TO 3:40 PM	**SUSHMITA ROY & SPENCER HALBERG, UNIVERSITY OF WISCONSIN-MADISON** *"Unsupervised learning to decipher short and long-range gene regulatory networks of cellular state"*
3:40 TO 4:00 PM	**BREAK** – *refreshments available outside of the Auditorium*
4:00 PM TO 4:35 PM	**LACRA BINTU & ELI COSTA, STANFORD UNIVERSITY** *"Bifunctional transcriptional effector domains control gene expression pulses in an occupancy-dependent manner"*
4:40 PM TO 4:55 PM	**SHAOXUN LIU, COLUMBIA UNIVERSITY** *"Pervasive multi-specificity of bZIP transcription factors revealed by biophysically interpretable machine learning"*
5:00 PM TO 7:30 PM	**WELCOME PARTY at LINDA HALL LIBRARY** *A shuttle is available to take you to the Linda Hall Library.* *Tours of their Rare Books Collection will take place every 20 minutes.* *Small bites, desserts, and refreshments available*

THURSDAY, APRIL 23

9:00 AM TO 9:45 AM	*Coffee and light breakfast options are available outside of the Auditorium*

SESSION 2: 3D ORGANIZATION AND EVOLUTION

9:45 AM	**OPENING ANNOUNCEMENTS**
9:50 AM TO 10:25 AM	**JAMES DAVIES & HANGPENG LI, UNIVERSITY OF OXFORD** *"Defining genome structure at sub-nucleosome resolution"*
10:30 AM TO 10:45 AM	**FRANCOIS SPITZ, UNIVERSITY OF CHICAGO** *"Evolution of tissue-specific gene-regulatory programs after vertebrate whole genome duplications"*
10:50 AM TO 11:05 AM	**ARMAAN MEHTA, STOWERS INSTITUTE FOR MEDICAL RESEARCH** *"Gene Structure Alone Predicts Most Eukaryotic Homology"*
11:05 AM TO 11:25 AM	**BREAK** – *refreshments available outside of the Auditorium*

11:25 AM TO 12:00 PM	**JUDITH ZAUGG & KRISTY OU, UNIVERSITY OF BASEL** *"Understanding Hematological Malignancies Through the Lens of Gene Regulation and Spatial Profiling"*
12:05 PM TO 12:20 PM	**JEAN-BENOIT LALANNE, UNIVERSITY OF MONTREAL** *"Multi-scale dissection, compaction and derivatization of mammalian developmental enhancers"*
12:30 PM TO 2:00 PM	**LUNCH BREAK** – *join us in the Stowers Library for lunch.* *Tours of the Institute will be offered from 1:15-1:45pm. Meet your tour leaders outside of the library.*

SESSION 3: REGULATORY NETWORKS AND DISEASE

2:00 PM TO 2:35 PM	**CHRISTINA LESLIE & YANG YANG, MEMORIAL SLOAN KETTERING CANCER CENTER** *"AI models of gene regulation with single-cell and 3D genomics"*
2:40 PM TO 2:55 PM	**TATJANA SAUKA-SPENGLER, STOWERS INSTITUTE FOR MEDICAL RESEARCH** *"From states to decisions: decoding cranial neural crest fate with dynamic regulatory models"*
3:00 PM TO 3:15 PM	**REBECCA MCAVOY, UNIVERSITY OF MICHIGAN** *"Small-effect variants interacting epistatically explain HST2 expression divergence between Saccharomyces cerevisiae and Saccharomyces paradoxus"*
3:15 PM TO 3:35 PM	**BREAK** – *refreshments available outside of the Auditorium*
3:40 PM TO 4:40 PM	**KEYNOTE LECTURE: BING REN AND GUOJIE ZHONG, NEW YORK GENOME CENTER, COLUMBIA UNIVERSITY** *"Exploring the role of noncoding DNA in human health and disease through single-cell atlases and AI"*
4:45 PM TO 6:15 PM	**POSTER SESSION 1** proceed down the staircase to the Stowers Gallery **ODD # POSTERS** 4:45 PM - 5:30PM **EVEN # POSTERS** 5:30 PM - 6:15 PM Light snacks and refreshments are available at the base of the stairs
6:30 PM	**DINNER** – *join us in the Stowers Library for dinner*

FRIDAY, APRIL 24

9:00 AM TO 9:45 AM	*Coffee and light breakfast options are available outside of the Auditorium*

SESSION 4: DEEP LEARNING THE CIS-REGULATORY CODE

9:45 AM	**OPENING ANNOUNCEMENTS**
9:50 AM TO 10:25 AM	**GEORG SEELIG & SEBASTIAN CASTILLO-HAIR, UNIVERSITY OF WASHINGTON** *"Decoding and engineering cell state-specific regulation with AI and synthetic biology"*
10:30 AM TO 10:45 AM	**MICHAEL WHITE, WASHINGTON UNIVERSITY IN ST. LOUIS** *"Deep mutagenesis and deep learning to understand epistatic interactions among transcription factor binding sites in a model promoter"*
10:50 AM TO 11:05 AM	**MELANIE WEILERT, STOWERS INSTITUTE FOR MEDICAL RESEARCH** *"Widespread low-affinity motifs enhance chromatin accessibility and regulatory potential"*
11:05 AM TO 11:25 AM	**BREAK** – *refreshments available outside of the Auditorium*

11:25 AM TO 12:00 PM	**PETER KOO & YIJIE KANG, COLD SPRING HARBOR LABORATORY** *"Dissecting the cis-regulatory landscape with deep learning"*
12:05 PM TO 12:20 PM	**SHAUN MAHONY, PENN STATE UNIVERSITY** *"Interpreting deep learning genomics models via concept attribution"*
12:30 PM TO 2:00 PM	**LUNCH BREAK** – *join us in the Stowers Library for lunch.* *Tours of the Institute will be offered from 1:15-1:45pm. Meet your tour leaders outside of the library.*
2:00 PM TO 3:30 PM	**POSTER SESSION 2** proceed down the staircase to the Stowers Gallery **EVEN # POSTERS** 2:00 PM – 2:45 PM **ODD # POSTERS** 2:45 PM – 3:30 PM Light snacks and refreshments are available at the base of the stairs

SESSION 5: FROM CIS-REGULATORY CODE TO COMPLEX TISSUES

3:30 PM TO 4:05 PM	**ANSHUL KUNDAJE & SELIN JESSA, STANFORD UNIVERSITY** *"Deep learning context-specific cis-regulatory syntax and genetic variation influencing human development"*
4:10 PM TO 4:25 PM	**TONY LI, UNIVERSITY OF WASHINGTON** *"Functional plasticity and tunability in the evolution of mammalian cis-regulatory elements"*
4:25 PM TO 4:45 PM	**BREAK** – *refreshments available outside of the Auditorium*
4:45 PM TO 5:00 PM	**FAHAD KAMULEGEYA, STOWERS INSTITUTE FOR MEDICAL RESEARCH** *"Deep learning-guided dissection and manipulation of neuronal subtype-specific enhancers at base-pair precision"*
5:05 PM TO 5:40 PM	**HONGKUI ZENG & REMI MATHIEU, ALLEN INSTITUTE** *"Developmental Origins of Brain Cell Type Diversity"*
5:45 PM	CLOSING REMARKS
6:00 PM TO 8:00 PM	**CLOSING RECEPTION** – *join us in the Stowers Library for a reception.* *Small bites, desserts, and refreshments available.*

Poster Sessions

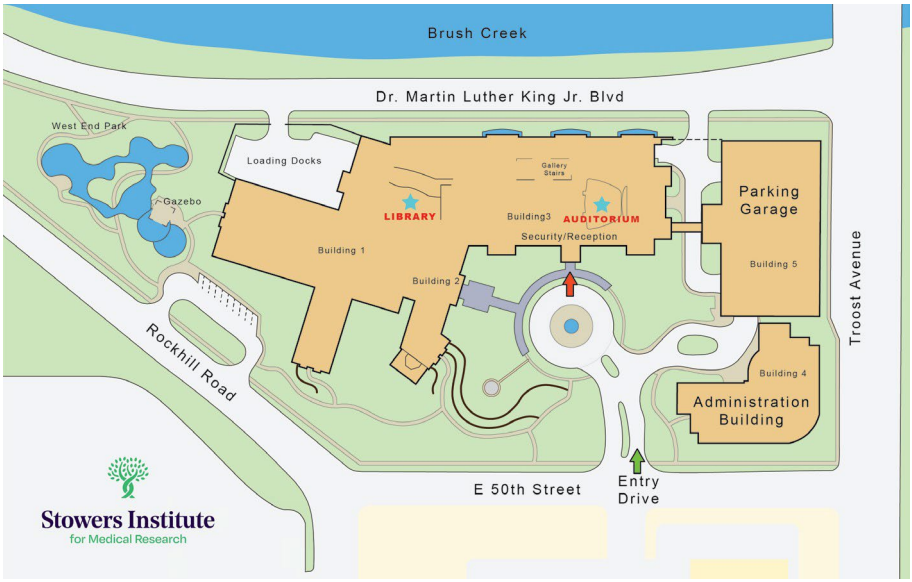
Poster Session

Poster Session 1: Thursday, April 23 | 4:45 PM – 6:15 PM
Odd numbered posters: 4:45 PM – 5:30 PM
Even numbered posters: 5:30 PM – 6:15 PM

Poster Session 2: Friday, April 24 | 2:00 PM – 3:30 PM
Even numbered posters: 2:00 PM – 2:45 PM
Odd numbered posters: 2:45 PM – 3:30 PM

Author	Title	Poster #
Simon Bourdareau Stowers Institute for Medical Research	ChIP-SMF: Scalable single-molecule profiling to reveal protein co-binding dynamics on DNA	1
Udit K. Chakraborty California Institute of Technology	Promoter-State Switching Underlies Dosage Compensation of the Developmental Regulator Snail	2
Trevor Christensen & Yash Vinaykumar Mundewadi Cold Spring Harbor Laboratory	Iterative Design of Training Datasets for Generalizable Sequence-to-Function Models	3
Rose Coyne Stowers Institute for Medical Research	Temporal regulation of a spatial patterning factor during neuronal identity specification in Drosophila	4
Zixuan Ding Stowers Institute for Medical Research	Integration of Promoter-Enhancer Interaction and Gene Regulatory Network during zebrafish gastrulation	5
Belén Gaete Humada Kansas State University	Functional Characterization of Argonaute Syndromes Variants in Caenorhabditis elegans	6
Hani Zakaria Girgis Texas A&M University-Kingsville	EnhancerDetector: An Intelligent Interpretable Framework for Identifying Enhancers Across Cell Types and Species	7
Taslima Haque University of Michigan	Evolutionary analyses of gene expression variation in Saccharomyces cerevisiae	8
Haining Jiang Stowers Institute for Medical Research	Deep learning models reveal separable sequence rules for chromatin accessibility and enhancer activation in Drosophila	9
Jean-Benoit Lalanne University of Montreal	Multi-scale dissection, compaction and derivatization of mammalian developmental enhancers	10
Mitchell R. Lewis University of Utah	ScRNASeq and Atomic Sketch Integration Reveal Ethanol-Induced Transcriptional Changes in Drosophila Cholinergic Neurons	11
Pablo Mantilla Puccetti Cold Spring Harbor Laboratory	Dissecting the cis-regulatory code beyond motif syntax	12

Campus Map



All t**alks** will be presented in the Auditorium.
**Food and beverages <u>are not</u> allowed inside of the auditorium.*

Breaks will be in the hallway outside of the auditorium.

Lunch will be in the Stowers library.

Poster Session will be in the Gallery, down the staircase outside of the auditorium.

Conference Room 114 is reserved as a **workspace**.
** Food and beverages <u>are allowed</u> inside Conference Room 114.*

Shuttle Schedule

A shuttle will run between the Kansas City Marriott – Country Club Plaza Hotel (4445 Main St, Kansas City, MO 64111) and the Stowers Institute for Medical Research (1000 E. 50th St, Kansas City, MO 64110) at the times listed below:

Wednesday, April 22:

12:30 PM & 1:30 PM	Pick up at Marriott Hotel, drop off at Stowers
5:00 PM	Pick up at Stowers, drop off at Linda Hall Library for welcome party
7:30PM	Pick up at Linda Hall Library, drop off at Marriott Hotel and Stowers

Thursday, April 23:

8:30 AM & 9:00 AM	Pick up at Marriott Hotel, drop off at Stowers
7:45 PM	Pick up at Stowers, drop off at Marriott Hotel

Friday, April 24:

8:30 AM & 9:00 AM	Pick up at Marriott Hotel, drop off at Stowers
7:45 pm & 8:15 PM	Pick up at Stowers, drop off at Marriott Hotel

SRC Social Media Guidelines

The Stowers Research Conference organizers **encourage the use of social media to share information and network with other attendees.**

We remind you to **remain courteous and respectful in your comments and posts.**

Avoid sharing speaker or poster presentation content that's not your own, copyrighted or trademarked, or material protected by other intellectual property rights.

Follow and tag **@stowersinstitute** on Instagram

Follow and tag **Stowers Institute for Medical Research** on LinkedIn

Follow and tag **@ScienceStowers** and **@Stowers_SRC** on X (Twitter)

Follow and tag **@stowersinstitute.bsky.social** on Bluesky

Use hashtag **#SRCKC26**

Invited Speakers and Trainees

Bing Ren and Guojie Zhong, New York Genome Center, Columbia University
Lab Website

Hongkui Zeng and Remi Mathieu, Allen Institute
Lab Website

Sushmita Roy and Spencer Halberg, University of Wisconsin-Madison
Lab Website

Judith Zaugg and Kristy Ou, University of Basel University
Lab Website

Georg Seelig and Sebastian Castillo-Hair, University of Washington
Lab Website

Christina Leslie and Yang Yang, Memorial Sloan Kettering Cancer Center
Lab Website

Anshul Kundaje and Selin Jessa, Stanford University
Lab Website

James Davies and Hangpeng Li, University of Oxford
Lab Website

Lacra Bintu and Eli Costa, Stanford University
Lab Website

Peter Koo and Yijie Kang, Cold Spring Harbor Laboratory
Lab Website

William Greenleaf and Benjamin Doughty, Stanford University
Lab Website

ChIP-SMF: Scalable single-molecule profiling to reveal protein co-binding dynamics on DNA

Simon Bourdareau1, Julia Zeitlinger1,2

1Stowers Institute for Medical Research, Kansas City, MO, 2University of Kansas Medical Center, Kansas City, KS

Protein-DNA interactions are a key aspect of nuclear processes such as gene transcription and chromatin organization. Current methods for mapping these interactions are limited by their bulk nature. These approaches average across cells and thus cannot capture the co-occurrence and dynamics of proteins on single DNA molecules. To address these limitations, we introduce ChIP-SMF, a method that combines single-molecule footprinting with chromatin immunoprecipitation. ChIP-SMF retains the scalability of a ChIP-seq bulk assay while embedding single-molecule information and providing high-coverage single-molecule footprints of genomic regions bound by a protein of interest. By locally increasing sequencing coverage, this method enables the detailed characterization of protein interactions and their co-binding dynamics on DNA. ChIP-SMF is particularly well suited for studying transcription factors and larger complexes such as the transcription machinery at promoters. To showcase the method, we performed ChIP-SMF on multiple components of the Pre-Initiation Complex (PIC) in Drosophila melanogaster Kc167 cells. We uncovered co-dependencies across promoter types and revealed the relationship between RNA Polymerase II pausing and the positioning of the +1 nucleosome, within a single dataset. We also developed flexible computational tools tailored for single-molecule resolution data to quantify and visualize the results. In summary, we developed an experimental and analytical framework to decode the combinatorial rules by which transcription factors and other proteins bind simultaneously on DNA.

Using deep learning sequence models to understand nucleosome and 3D genome organization in the early Drosophila embryo

Sam Campbell1, Charles McAnany1, Melanie Weilert1, Kaelan Brennan2, Julia Zeitlinger1,3

1Stowers Institute for Medical Research, 2Stanford Medicine, 3University of Kansas Medical Center

Transcriptional regulation is encoded in the DNA sequence, which specifies transcription factor binding, chromatin accessibility, enhancer activation and target gene expression. A poorly understood element is how DNA sequence encodes the three-dimensional (3D) genome organization and how these sequence features impact gene regulation. While individual DNA motifs have been implicated in shaping genome architecture, the broader, combinatorial sequence rules linking 3D genome organization and gene expression remain elusive. To address this gap, we trained a BPNet model (as part of the BPReveal package) on MNase-nucleosome data from the early Drosophila embryo from DNA sequence, extracted the learned motifs and analyzed the role of insulator motifs. Moreover, we trained Akita to predict the pairwise contacts of Micro-C data in the early Drosophila and conducted perturbation tests to understand the role of insulator motifs. We found that some motifs show a similar perturbation response in the nucleosome model and the Micro-C model. We discover a new potential regulatory syntax at insulators, nucleosome-phased syntax, in which insulator motif instances preferentially occur in periodic distances of 200bp. This raises the possibility that nucleosomes play an underappreciated role in the 3D chromatin organization. We are currently probing the models to understand how insulator motifs cooperate in nucleosome organization and TAD boundary formation, and the role that nucleosome-phased syntax plays in this process.

Promoter-State Switching Underlies Dosage Compensation of the Developmental Regulator Snail

Udit K. Chakraborty1, Leslie Dunipace1, Angelike Stathopoulos1

1Department of Biology and Biological Engineering, California Institute of Technology

Snail (Sna) is a conserved transcription factor (TF) responsible for epithelial-to-mesenchymal transition and is critical for gastrulation and later tissue development. Deviations in nuclear Sna levels, whether from environmental or genetic perturbation, can lead to improper tissue development and have been linked to cancer, necessitating a tight regulation of the sna gene. Dosage compensation is one mechanism by which alleles buffer transcriptional output when a partner allele is absent or exhibits altered expression due to environmental changes or mutations, including heterogeneous changes in key cis-regulatory modules. Sna represents an ideal system to study dosage compensation because of its autoregulatory feedback: together with partner TFs, Sna can either repress or activate its own expression. Using MS2-based live embryo imaging combined with mathematical inference, we find that a wild-type (wt) sna allele can either up- or down-regulate its expression when paired with a partner allele carrying enhancer mutations that drive either low or high expression. To dissect the underlying regulatory mechanism, we applied a two-state Hidden Markov model to characterize how the promoter toggles between ON and OFF states to modulate transcriptional output. We observe that the dosage compensation of sna, an autosomal gene, is accompanied by a change in bursting amplitude and frequency. Currently, we are investigating the mechanistic insights regarding Sna action that underlie the observed differences in these transcriptional parameters. These findings provide additional understanding into how tight regulation of key developmental genes is achieved and highlight the spatio-temporal contributions of multiple enhancers working together to maintain optimum nuclear protein levels.

Iterative Design of Training Datasets for Generalizable Sequence-to-Function Models

Trevor Christensen1,2, Yash Vinaykumar Mundewadi1,2, Peter Koo2

1co-first author, 2Cold Spring Harbor Laboratory

Predicting regulatory element activity from DNA sequence is a central challenge in biology. While modern sequence-to-function models achieve strong performance on held-out genomic sequences, they fail to generalize reliably to perturbations and synthetic sequences. This generalization gap reflects a fundamental limitation of current training data: genome-wide profiling assays sample only a narrow and biased slice of regulatory sequence space, leaving models without exposure to the genetic variation needed to learn transferable regulatory rules. Improving generalization therefore requires not just more data, but strategically selected data that better samples the distribution of sequences models will encounter at deployment. Here, we introduce S2F-LearningLoop, a modular platform for systematically exploring how training data composition and selection strategy affect sequence-to-function model performance. Using in-silico oracle models as experimental proxies, we compare a spectrum of approaches ranging from informed biological baselines that leverage prior knowledge of regulatory sequence features, to model-guided active learning strategies that use uncertainty and diversity criteria to identify maximally informative training sequences. Both the composition of candidate sequence pools and the criteria used to select from them are treated as separable design choices, enabling a structured exploration of how each contributes to model improvement. Performance is evaluated across test sets spanning genomic sequences, low-shift perturbations, high-shift synthetic variants, and purely random sequences, capturing the range of distribution shifts that arise in regulatory genomics applications. The goal of this platform is to identify which sequence generation and selection strategies most efficiently improve model generalization, ultimately informing the design of prospective experiments.

Temporal regulation of a spatial patterning factor during neuronal identity specification in Drosophila

Rose Coyne1, Kenzie Treese1, Cathleen Lake1, Raghuvanshi Rajesh2, Yen-Chung Chen2, M. Neşet Özel1

1Stowers Institute for Medical Research, Kansas City, MO, 64110,
2Department of Biology, New York University, New York, NY, 10003

During neurogenesis, numerous cell types must be produced and specified from a limited number of progenitors that are patterned across space and time by signaling molecules and transcription factors (TFs). Once specified, the cell type identities of >200 neurons in the Drosophila optic lobe are established and maintained by unique combinations of terminal selectors (tsTFs) in each cell type. However, how patterning in progenitors activates tsTF expression is not well understood. Two cell types, Dm2 and Mi15, share all tsTFs except for Vsx1 and Vsx2 expressed in Dm2 neurons. Previous work established that ectopic expression of either Vsx1 or Vsx2 was sufficient to convert Mi15 to Dm2 in morphology and neurotransmitter identity. Vsx1 is also an established spatial patterning factor in a restricted neuroepithelial domain which produces many Vsx1/2+ neurons. We show that these neurons are eliminated upon neuroepithelial knockdown of Vsx1. In contrast, Dm2 neurons are produced in all spatial domains and also express Vsx1/2, which was not affected by neuroepithelial knockdown of Vsx1. We hypothesized that Vsx1/2 expression in Dm2 is specified by temporal patterning instead. Indeed, we found that temporal TF BarH1 is responsible for Dm2 specific activation of Vsx1/2, which distinguishes them from Mi15 neurons that are born shortly after Dm2 during the Tailless temporal window . Using our single-cell multiome (RNA+ATAC-seq) dataset during neurogenesis, we identified the enhancers that control Vsx1/2 activation in Dm2, which are distinct from their spatially patterned enhancers. In addition to providing a direct connection between the upstream patterning in progenitors and downstream terminal identity, this work makes an important contribution to our understanding of how enhancers can be selectively activated via different patterning mechanisms to regulate the same genes and helps explain the ‘reuse’ of TFs across development for vastly different purposes.

Investigating Differential miRNA Strand Selection During Development

Heather Crawshaw1, Sumire Kurosu1, Jeff Medley1, Anna Zinovyeva1

1Kansas State University

Regulation of gene expression is a fundamental process that enables diverse cellular functions by controlling the magnitude, timing, and location of different gene activities. Proper gene regulation is essential for normal animal development, and disruptions in this process are associated with several cancers and developmental disorders. Among the key regulators of gene expression are microRNAs (miRNAs), which are a

class of small, non-coding RNAs that play a crucial role in post-transcriptional gene silencing by repressing the activity of target genes. The final step in miRNA biogenesis is the formation of a miRNA duplex. This

duplex contains two strands: a functional guide miRNA strand and a star miRNA strand which

is typically degraded. The guide miRNA is preferentially loaded by an Argonaute protein to form

a functional complex that represses miRNA specific target transcripts. Here, we investigate possible strand switching of certain miRNAs throughout various developmental stages of the model

organism Caenorhabditis elegans. In addition, we investigate whether terminal nucleotidyl transferases (TENTs), that add non-templated nucleotides to the 3' end of miRNA molecules, play a role in miRNA strand selection. To address this question, we first investigated the developmental profile of several TENTs and assayed what effects TENT gene mutations have on miRNA abundances and uridylation or adenylation of miRNA 3' ends. We are currently assessing the effects TENT gene mutations have on miRNA strand selection during C.

elegans development. Deepening our understanding of miRNA dynamics across development and how TENTs shape miRNA stability and biogenesis will shed light on the mechanisms that fine-tune miRNA-mediated gene regulation during animal development.

Integration of Promoter-Enhancer Interaction and Gene Regulatory Network during zebrafish gastrulation

Zixuan Ding1, Evgeny Akkuratov1, Yuri Iwamura1, Andrew Price1, Tatjana Sauka-Spengler1

1Stowers Institute for Medical Research

Zebrafish gastrulation is characterized by limited number of cell types, rapid differentiation, and extensive cell movements, providing an ideal model for studying how gene regulatory networks (GRNs) assemble in spatial context to drive cell fate decisions. Existing GRN frameworks perform well when inferring regulatory hierarchies of transcription factors (TFs), but lack the resolution necessary to identify the specific TF mediated promoter–enhancer (P–E) interactions. Here, we developed a framework to integrate in-vivo P–E interaction information into a deep-learning-inferred GRN during zebrafish gastrulation. Using scRNA-seq and scATAC-seq, we identified 16 cell states and four differentiation trajectories and demonstrated that the chromatin accessibility landscape can predict cell fates during gastrulation. Applying SCENIC+ to single-cell multiomic data, we inferred preliminary GRNs and then integrated base-pair–resolution chromatin conformation capture data obtained by Micro Capture-C (MCC) to systematically characterize the promoter–enhancer interaction landscape during gastrulation. We demonstrated that MCC data can be deconvolved using scATAC-seq data to provide cell type–specific P–E interactions and incorporate them into the GRN. This integrative framework enables the identification of key transcription factors driving cell fate and potential repressors inhibiting P–E interactions.

Functional Characterization of Argonaute Syndromes Variants in Caenorhabditis elegans

Belén Gaete Humada1, Rebecca Mitchell1, Amélie Piton2, Davor Lessel3, Hans-Jürgen Kreienkamp4, Victor Ambros5, Anna Zinovyeva1

1Division of Biology, Kansas State University, Manhattan, KS, 2Department of Translational Medicine and Neurogenetics, Institute of Genetics and Molecular and Cellular Biology, Strasbourg University, Strasbourg, France, 3Institute of Human Genetics, University of Regensburg, Regensburg, Germany, 4Institute of Human Genetics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany, 5Program of Molecular Medicine, UMass Chan Medical School, Worcester, MA

Gene regulation is essential for animal development. Argonaute (AGO) proteins, guided by microRNAs (miRNAs), play key roles in post-transcriptional regulation by silencing target genes. Recently, coding variants in human AGO1 and AGO2 genes have been identified as causative for rare developmental disorders termed Argonaute Syndromes (AS). The AS variants, primarily missense alleles, are associated with clinical manifestations that range from mild to severe. As the number of distinct AS variants increases, there is a need to rapidly characterize their effects on molecular AGO functions. Given that human and Caenorhabditis elegans miRNA-associated AGOs are highly conserved, modeling AS variants in C. elegans allows for their rapid in vivo functional characterization. We previously modeled four human AGO1 AS mutations in the C. elegans homolog alg-1, revealing allele-specific developmental and miRNA-related molecular phenotypes. In this study, we used CRISPR-Cas9 genome editing to engineer 14 additional alg-1 AS variants, further increasing our coverage of AS variations. Genetic analysis of alg-1 AS strains revealed varying degrees of developmental defects. While some variants cause only mild effects, a subset of alg-1 AS alleles disrupt development more strongly than loss of alg-1, exhibiting antimorphic phenotypes. Introducing wildtype alg-1 into alg-1 AS mutant strains partially restores normal development, consistent with dosage-dependent effects of ALG-1 function. Our current work aims to characterize the molecular effects of alg-1 AS variations in miRNA biogenesis and activity. Genetic and molecular analysis will enable functional classification of AS variants, contributing to our understanding of genotype-phenotype causalities and disease mechanism of AS.

Mapping and Functional Dissection of Enhancer-Gene Interactions Governing Primordial Germ Cell Development

Jingy i Gao1, Haiyuan Yu2, John Schimenti3

1Department of Biomedical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853., 2Weill
Institute for Cell and Molecular Biology, Cornell University, Ithaca, NY, USA, 3Department of Biomedical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853

Infertility is a common reproductive health problem, affecting ~15% of couples worldwide. Studies over the past decades have identified hundreds of causative genes, yet the genetic causes of most cases still remain unexplained. Due to the highly polygenic and heterogeneous nature of the disease, genome-wide association studies (GWAS) have had limited success. Risk variants identified so far mostly reside in non-coding regulatory regions, suggesting that defects in gene regulation may underlie many infertility cases. As germline stem cells, primordial germ cells (PGCs) are the progenitors of all germ cell lineages. During early embryogenesis, this crucial developmental phase governs the production of gametes with optimal quality and quantity.

Dysregulation in PGC specification, migration, or proliferation will result in germ cell development failure and compromise fertility. However, the regulatory landscape governing germline development remains poorly characterized.

Here, leveraging a high-efficiency, four–transcription–factor–based in vitro differentiation system that produces up to 80% Stella^{+} primordial germ cell-like cells (PGCLCs), we generated nascent RNA sequencing data to identify active cis-regulatory elements (CREs) across PGCLC differentiation. We then identified candidate PGC enhancers and established their linkage to target genes through Activity-By-Contact analyses. We further validated the enhancer activity of top candidates in the PGCLC model and designed a CRISPR knockout and a CRISPR interference screen to identify the essential genes and enhancers that safeguard PGC development. Together, these findings will provide a foundation for constructing a comprehensive enhancer–gene interaction map and for interpreting infertility-associated non-coding variants, thereby offering new insights into infertility at the level of genomic regulation.

RNA-binding protein HRPK-1 coordinates with miRNAs to regulate C. elegans development

Mika Ghosh1, Li Li1, Anna Zinovyeva1

1Kansas State University

MicroRNAs (miRNAs) are small, noncoding RNAs that regulate gene expression post-transcriptionally. Argonaute proteins load miRNAs, which guide the complex to target mRNAs through partially complementary sequences, leading to gene repression. RNA-binding proteins (RBPs) also regulate mRNA expression by influencing RNA processing, transport, stability, and translation and can do so both independently and in conjunction with miRNA pathways. To identify factors that affect miRNA function, previous work screened for protein interactors of the Argonaute protein ALG-1 in Caenorhabditis elegans using mass spectrometry and identified HRPK-1 as a physical interactor of ALG-1. HRPK-1 is the C. elegans homolog of the human RNA-binding protein hnRNPK. Previous studies demonstrated that HRPK-1 modulates miRNA activity during development, as loss of hrpk-1 enhances phenotypes associated with reduced activity of the let-7 and mir-35 miRNA families. However, the molecular mechanisms by which HRPK-1 regulates miRNA function remain unclear. To address this question, we performed a functional domain analysis by generating deletions or mutations in six distinct HRPK-1 domains or predicted signaling sequences. Our results indicate that the RNA-binding activity of all three KH domains is essential for proper miRNA function and normal development, whereas deletion or mutation of other HRPK-1 domains and signaling motifs produced variable effects on development and miRNA activity. In addition to domain analysis, we also generated a C. elegans model for one of the Au-Kline Syndrome (AKS)-associated clinical variants, hrpk-1(L68R). Although hrpk-1(L68R) animals do not exhibit developmental defects on their own, they enhance miRNA reduction-of-function phenotypes in mir-48 mir-241(nDf51) and lsy-6(ot150) mutant backgrounds, suggesting a functional impact of this variant on miRNA activity. Together, these findings define domains required for HRPK-1 function coordinating with miRNA gene regulation during C. elegans development.

EnhancerDetector: An Intelligent Interpretable Framework for Identifying Enhancers Across Cell Types and Species

Luis M. Solis1, Geyenna Sterling-Lentsch2, Marc S. Halfon2, Hani Z. Girgis1

1Texas A&M University-Kingsville, 2The State University of New York at Buffalo

Background: Accurately identifying transcriptional enhancers remains a major challenge in regulatory genomics. Existing tools often rely on chromatin features, extensive feature engineering, or complex thresholding strategies, limiting usability and cross-species applicability. Deep learning–based models show great potential for locating enhancers, but many lack broad generalization across species. Currently, large-scale sequencing project are taking place, resulting in generating thousands of newly assembled genomes that require regulatory annotation. To overcome this bottleneck, we developed a sequence-based framework that directly predicts enhancer activity from short DNA sequences while providing biologically meaningful explanations for its predictions. Our tool leverages common sequence properties of enhancers—such as multiple degenerate transcription factor binding sites—regardless of cell type or species.

Results: We present a convolutional neural network that learns enhancer-associated sequence patterns from 400 bp human enhancer data and generalizes strongly across species and experimental assays. The model achieved high precision and F1 on human test sets and outperformed or matched existing enhancer-prediction methods, including LS-GKM, Enhancer-FRL, DeepSEA, and Enformer-derived chromatin rules. Without modification, it identified mouse enhancers with high precision, and fine-tunned on as few as 20,000 enhancer sequences increasing its performance. Experimental validation in transgenic Drosophila demonstrated that five of six predicted enhancers drove reporter gene expression. To provide interpretability, we apply class activation maps to highlight sequence regions most influential in classification. Knockout, insertion, and context-disruption experiments revealed that these regions capture essential regulatory motifs and their surrounding sequence context.

Conclusion: This work introduces a robust, interpretable, and cross-species deep learning framework for enhancer prediction. The model performed strongly across independent datasets, reduced reliance on chromatin assays, and provided insight into enhancer architecture. Its ability to transfer across species with minimal data makes it well suited for enhancer discovery in newly sequenced genomes, offering a scalable approach for future regulatory genomics research.

Leveraging sequence models to infer TSS usage from standard RNA-seq data

Daniel Witoslawski1, Andrew Hsu2, Mingon Kang1, Mira Han1

1University of Nevada, Las Vegas, 2University of Nevada, Reno

Background: Standard RNA-seq lacks the resolution to precisely quantify Transcription Start Site (TSS) usage due to sparse and noisy coverage at the 5' end. While CAGE-seq provides high-resolution TSS maps, it comes with higher costs. Current computational approaches typically rely on split reads at first-exon junctions, failing to capture 5'-end variation within exons and overlooking internal or overlapping promoters.

Results: We extended the deep learning framework EPInformer to integrate promoter sequence features with RNA-seq coverage patterns to predict TSS usage. Using CAGE-seq as ground truth in K562 and GM12878 cell lines, we demonstrate that incorporating RNA-seq data significantly improves prediction over DNA-only models, increasing the Pearson R from ~0.65 to ~0.78. While RNA-seq data provides a substantial boost to sequence-based models, it offers more moderate gains when integrated with models already utilizing chromatin marks and Hi-C data. We identified specific motifs driving differential TSS usage between these cell lines and evaluated the regulatory contribution of distal enhancer sequences to TSS selection.

Conclusion: By coupling DNA sequence models with existing RNA-seq datasets, we achieve higher accuracy in TSS prediction. This approach increases the utility of standard transcriptomic data, enabling high-resolution regulatory analysis when specialized 5'-end sequencing is out of reach.

Evolutionary analyses of gene expression variation in Saccharomyces cerevisiae

Taslima Haque1, Patricia J. Wittkopp1,2

1Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, 2Department of Molecular, Cellular, and Developmental Biology, University of Michigan Ann Arbor

Gene expression variation is a major driver of phenotypic evolution, with both cis- and trans-regulatory variants contributing to changes in expression. Although the relative contribution of these two types of changes differs among species, it is generally observed that trans-regulatory variants make a larger contribution to intraspecies than interspecies variation. Despite the prevalence of trans-regulatory variation in gene expression within a species, we still know little about whether this variation is primarily neutral or adaptive. In this current study, we leveraged gene expression data from a large biparental cross between 26 genetically diverse natural isolates of Saccharomyces cerevisiae, which demonstrated extensive trans-regulatory variation underlying gene expression to investigate whether individual gene expression levels are more likely to be adaptive or neutral. We also examined their collective co-expression networks for adaptive expression variation. Using the genomic relatedness among these natural strains, we are building neutral models to test for excessive parental expression variation at both the level of individual genes and their co-expression modules. Additionally, we aim to investigate whether genes with different genetic architectures show similar or different degrees of adaptive expression variation. Overall, our investigation will enhance our understanding of the driving forces of gene expression evolution and their underlying genetic architectures.

Combining natural genetic variations with single molecule footprinting to learn about the rules of chromatin regulation

Kaixuan Luo1, Ayelen Lizarraga1, Xiaotong Sun1, Diana Vera Cruz1, Xin He1, Sebastian Pott1

1University of Chicago

Gene regulation is a highly dynamic process mediated through interactions of transcription factors (TFs), nucleosomes, and DNA methylation (DNAm). The key challenge of the field is to unravel the rules of these molecular interactions. Single-molecule footprinting (SMF) emerges as a promising technology to address this challenge. SMF combines DNA methyltransferase treatment and direct long-read sequencing to simultaneously capture chromatin accessibility, DNAm, and TF binding across intact 10-20 kb molecules. Given the complexity of gene regulation, however, it is still difficult to extract causal and mechanistic rules from SMF data. We reason that natural genetic variations in DNA sequences provide large-scale perturbations of gene regulatory processes, creating an excellent opportunity to learn about chromatin regulation. For example, a SNP disrupting the motif of a TF would provide information of the effect of this TF on the nearby chromatin state.

We generated SMF data in 30 human lymphoblastoid cell lines. Using these data, we detected ~6,000 cis-regulatory elements that show differential accessibility between different alleles of genetic variants, denoted as allele-specific fiber-seq inferred regulatory elements (AS-FIRE). Many of the variants underlying AS-FIRE disrupt motifs of TFs controlling chromatin accessibility such as CTCF. These variants often have broad effects on nucleosome positioning and phasing. We also detected ~2,000 TF footprints that show allelic difference, denoted as AS-footprints. Notably, half of the AS-footprints are missed by AS-FIREs, and these genetic variants often exhibit more local effects on nearby chromatin states. Our results thus suggest that different TFs have variable chromatin effects. We are currently characterizing the TFs underlying AS-FIREs and AS-footprints, and map additional genetic variants on gene regulation through molecular Quantitative trait loci analysis.

Our results demonstrate that combining natural genetic variations with SMF is a powerful strategy of learning about gene regulation.

Deep learning models reveal separable sequence rules for chromatin accessibility and enhancer activation in Drosophila

Haining Jiang1, Charles McAnany1, Julia Zeitlinger1,2

1Stowers Institute for Medical Research, Kansas City, MO 64110, USA, 2Department of Pathology & Laboratory Medicine, The University of Kansas Medical Center, Kansas City, KS 66160, USA

Identifying and predicting the activity of all enhancers across various cell types during development is a pivotal yet unresolved challenge. ATAC-seq stands out as a versatile assay that measures chromatin accessibility at single-cell resolution across diverse cell types. However, chromatin accessibility is not a direct measure of enhancer activity, and models of chromatin accessibility do not reliably predict enhancer activity levels derived from reporter assays such as STARR-seq performed in the same cell type. To dissect and directly compare these two assays, we trained BPNet models to accurately predict ATAC-seq data and STARR-seq data in Drosophila S2 cells from sequence alone. Each model learned and mapped motifs whose individual contributions correlate with the binding affinity of their corresponding transcription factors measured with ChIP-nexus, validating the quality of motifs derived from ATAC-seq and STARR-seq. However, the various transcription factor motifs exhibit different relative contributions towards chromatin accessibility and enhancer activation. While the chromatin accessibility model relies more on motifs bound by pioneer transcription factors, the enhancer activation model depends on a broader combination of transcription factors, including both pioneer and non-pioneer motifs, and more extensive motif-motif cooperativity. To validate the differences in sequence rules, we used a genetic algorithm to mutate enhancer sequences to various levels of enhancer activation while keeping the chromatin accessibility unchanged (and vice versa). We conclude that chromatin accessibility and enhancer activation are encoded by overlapping but distinct sequence rules, implying that accessibility and activation are two separable molecular processes that consecutively determine the activity of enhancers. This suggests that sequence rules derived from chromatin accessibility data could be supplemented with additional motif and syntax rules to better predict enhancer activity.

Deep learning-guided dissection and manipulation of neuronal subtype-specific enhancers at base-pair precision

Fahad Kamulegeya1, Rose Coyne1, Cathleen Lake1, Julia Zeitlinger1, M. Neset Özel 1

1Stowers Institute for Medical Research

The Drosophila visual system is an excellent model for investigating how diverse neural cell types are generated and form complex circuits. Distinct combinations of transcription factors, known as terminal selectors (tsTFs), establish and maintain cell-type identity in postmitotic neurons, but how cis-regulatory regions read out the various tsTF combinations in a cell-type-specific manner remains largely unexplored. To address this, we generated a single-cell multiomics atlas of developing optic lobes at four key stages, providing a high-resolution characterization of gene expression and chromatin accessibility for all its cell types. We then trained deep learning models on cell-type-specific accessibility profiles and inferred the underlying grammar of cis-regulatory sequences by interpreting these models. As a proof of concept, we present BPNet models of pseudobulk ATAC-seq profiles at various developmental time points for four cell types, Tm1, Tm2, Tm4, and Tm6. Interpretation of these models revealed canonical pioneering motifs like GAGA, architectural TF motifs such as CTCF, and promoter elements known to be associated with accessible regions. Interestingly, the models also effectively capture the binding motifs of key tsTFs in each cell type. To evaluate the utility of these models, we tested

model-derived mutations on a Tm1-specific enhancer of the tsTF gene Drgx, which we had previously shown to be critical for specifying Tm1 fate over Tm4. Upon interpretation of the enhancer, we identified binding motifs of Lola-I as key in activating this enhancer, and Aop (a Tm4 tsTF) as key in repressing this enhancer in Tm4. As predicted by the models, mutation of the Aop motif resulted in increased Tm4 reporter activity, while mutations in the Lola-I motif decreased Tm1 reporter activity. Our results demonstrate that accurate interpretations of BPNet models reveal cis-regulatory rules of tsTFs and provide a way to design targeted mutations to disrupt the regulatory circuitry.

Multi-scale dissection, compaction and derivatization of mammalian developmental enhancers

Jean-Benoit Lalanne1,2, Tony Li3, Emma A.N. Kajiwara3, Chau Huynh3, Tiffany V. Do3, Beth K. Martin3, Samuel G. Regalado3, Jay Shendure3,4

1University of Montreal, Department of Biochemistry, 2Courtois Institute for Biomedical Innovation, 3University of Washington, Department of Genome Sciences, 4HHMI

Gene expression during mammalian development is orchestrated by non-coding cis-regulatory DNA elements (CREs) such as distal enhancers. Despite their fundamental importance, many high-level properties of enhancer 'grammar' remain unresolved. How does the length of an autonomously active CRE constrain its activity? How robust are CREs to mutations or rearrangements of transcription factor binding sites (TFBSs)? And how much epistasis exists among these sites? As predictive models solely trained on endogenous CREs are unlikely to resolve these questions, we subjected several endogenous CREs to intensive sequence-level perturbation.

Specifically, we assayed >35,000 variants of 5 parietal endoderm enhancers, with each variant fragmenting, compacting, mutating, or reassembling the endogenous architecture. The variants were organized into four perturbation classes, designed to probe: (i) the functional sufficiency of sub-fragments via dense multi-size tiling, (ii) local epistasis via multi-hit saturation mutagenesis, (iii) activity-size tradeoffs via model-guided compaction, or (iv) functional resilience via sequence derivatization anchored on key TFBSs, including random deposition, reconstitution, and synthetic thripsis. This multi-scale dissection revealed rich phenomena. Sub-tiling uncovered sharp non-additivity between activity and fragment size, highlighting strongly synergistic TFBS clusters. Compaction showed that natural CREs lie far from the activity-size Pareto front, and that model-guided deletions can yield shorter yet stronger elements. Mutational scanning exposed a spectrum of robustness, from highly tolerant to extremely fragile, together with rare but consequential epistasis between individual TFBSs.

Finally, TFBS-anchored derivatization demonstrated that background sequence can influence activity on par with TFBS arrangement. Strikingly, a substantial fraction of synthetic thripsis variants exceeded the activity of their endogenous progenitors. Taken together, these results reveal both 'soft' and 'stiff' directions in regulatory sequence space, advancing a quantitative phenomenology of how enhancer sequences encode function and robustness.

ScRNASeq and Atomic Sketch Integration Reveal Ethanol-Induced Transcriptional Changes in Drosophila Cholinergic Neurons

Mitchell R. Lewis1, Madelyn Miles1, Aakriti Bhandari1, Adrian Rothenfluh1,2

1University of Utah, 2Huntsman Mental Health Institute

Alcohol use disorder is a debilitating disease affecting approximately 10% of Americans over the age of 12. Disrupted sleep is a common symptom and leading cause of relapse, as sleep disturbances may persist for months after abstinence from alcohol. To better understand the neurological changes causing these long-last effects, we use the vinegar fly, Drosophila Melanogaster, because of a high percentage of functional homologs of human disease related genes, the ability to develop tolerance, and similar alcohol-induced sleep changes.

Previous work in our laboratory found that cholinergic neurons are involved in long-lasting alcohol-induced sleep deficits. To determine the impacts of alcohol on neural gene expression, flies were exposed to either air or ethanol vapor and allowed to recover for 18 hours. Cholinergic neurons were isolated using INTACT pull-down and single cell RNA sequencing (scRNASeq) was performed. To identify alcohol-induced changes, atomic sketch integration was performed using Seurat and BPCells to link air and ethanol samples. We focused on the cells from the mushroom bodies (MBs), structures in the drosophila brain associated with learning, memory, sleep, and alcohol response. Cholinergic MB neuron clusters were isolated based on FasII expression, and differentially expressed genes were identified. Gene ontology returned terms involving stress response, metabolic processes, and protein folding. Several heat-shock proteins (Hsp26, Hsp23, Hsp27, Hsp68, Hsp68Bc, and Hsp70Ab had increased expression in the ethanol-exposed neurons. Heat-shock protein 26 (Hsp26) has previously been associated with alcohol response in flies. Interestingly, several immune-related proteins (IM1, IM2, IM3, and IM4) were also increased by ethanol exposure. While other studies have also observed increased IM2 expression from alcohol exposure, the role of these genes in alcohol response is unknown. This hypothesis-forming assay utilizes the power of scRNASeq to survey neuronal transcriptomes and identify pathways to further our understanding of alcohol response in flies and humans.

Functional plasticity and tunability in the evolution of mammalian cis-regulatory elements

Tony Li1,2, Jean-Benoit Lalanne1,3, Emma A.N. Kajiwara1,2, Shruti Jain1,2, Xiaoyi Li1, Tiffany V. Do1,2, Beth K. Martin1,2, Samuel G. Regalado1,2, Riza M. Daza1,2, Jay Shendure1,2,4,5,6

1Department of Genome Sciences, University of Washington, Seattle, WA, USA, 2Seattle Hub for Synthetic Biology, Seattle, WA, USA, 3Department of Biochemistry and Molecular Biomedicine & Courtois Institute for Biomedical Innovation, Université de Montréal, Montréal, QC, Canada, 4Brotman Baty Institute for Precision Medicine, Seattle, WA, USA, 5Howard Hughes Medical Institute, Seattle, WA, USA, 6Allen Discovery Center for Cell Lineage Tracing, Seattle, WA, USA

Cis-regulatory elements (CREs) such as enhancers play a central role in orchestrating mammalian development, yet how they gain, lose, maintain or change function over evolutionary timescales remains poorly understood. To address this gap, we densely mapped the functional evolution of five mouse developmental enhancers by testing orthologous sequences from 480 extant and ancestrally reconstructed mammalian genomes with massively parallel reporter assays (MPRAs). This phylogenetic dissection revealed diverse modes of evolution, from lineage-restricted activity to deep functional conservation despite extensive sequence divergence. To pinpoint causal changes, we developed a model-driven reconstitution strategy that uses deep learning-based predictions of chromatin accessibility to re-introduce a succession of mutations into ancestral orthologs, which revealed critical transcription factor binding site changes (TFBSs), pervasive epistasis, and super-additive mutational effects. When we extended this strategy to tune the activity of extant orthologs, we found that a handful of computationally prioritized mutations are sufficient to either enhance or ablate function. Together, these results shed light on how the plasticity of mammalian enhancers intersects with their evolution, and advances a framework for fine-turning the activity of cis-regulatory elements at nucleotide resolution.

Hierarchical Chromatin Structures Drive Coordinated Keratin Gene Expression Across Vertebrate Skin Development

Ya-Chen Liang1, Ping Wu1, Wen-Chien Jea1, Chih-Kuan Chen1, Tsz Yau Law1, Cheng-Ming Chuong1

1Department of Pathology, Keck School of Medicine, University of Southern California, Los Angeles, California, USA

Keratin genes are essential for the structural integrity, protection, and function of various organs, including skin, hair, nails, feathers, and scales, while also serving as signaling factors involved in cellular processes such as wound healing, tissue regeneration, and stress response across vertebrates. In humans and mice, keratins are organized into two major clusters, whereas birds and reptiles have four clusters. During organ development and regeneration, precise transcriptional control of keratin genes is critical for tissue stratification and region-specific characteristics, yet their transcriptional regulation remains poorly understood. Using multi-omic approaches, we investigated keratin gene clusters during skin development and specification in chicken, mouse, and human, uncovering a hierarchical regulatory model. Type-I and Type-II clusters form distinct chromatin configurations initially confined within a topologically associating domain (TAD) that progressively subdivides into two sub-TADs during avian embryonic skin development. Regional TAD and sub-TAD formation at Keratin Scaffold Regions (KSRs) establishes a foundational chromatin framework independent of active transcription, while promoters and enhancers within these structures are pre-occupied by transcription factors, poised for activation through micro-scale promoter-promoter, promoter-enhancer, and enhancer-enhancer loops. Disruption of CTCF expression impairs keratin expression and alters skin morphology. These findings reveal a novel three-dimensional chromatin regulatory mechanism governing keratin gene cluster organization and expression across species.

Pervasive multi-specificity of bZIP transcription factors revealed by biophysically interpretable machine learning

Shaoxun Liu1, Chase Ende1, Paul Nieuwerburgh1, Richard S. Mann2,3, Harmen J. Bussemaker1,3

1Department of Biological Sciences, Columbia University, New York, NY, USA, 2Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY, USA, 3Department of Systems Biology, Columbia University, New York, NY, USA

Basic leucine zipper (bZIP) transcription factors (TFs) are key regulators of diverse biological processes that bind DNA as homodimers or heterodimers. They are believed to be classifiable into subfamilies that bind either an odd-symmetry binding site (TGASTCA for the AP-1 subfamily) or an even-symmetry one (TGACGTCA for the CREB subfamily; TTGCGCAA for the C/EBP subfamily). However, we previously showed anecdotally that individual bZIPs can allosterically adapt to the DNA sequence and bind with either symmetry, and with relative preference that is an intrinsic property of the bZIP protein (Riley et al., eLife, 2015, PMID 26701911). To investigate how general this phenomenon is, we systematically analyzed 171 SELEX-seq datasets covering 67 bZIP homodimers from multiple species using ProBound, a biophysically interpretable machine-learning framework (Rube et al., Nature Biotechnol., 2022, PMID 35606422). We find that for a substantial number of bZIPs a single weight matrix does not suffice to fully capture their DNA binding behavior. Instead, it is necessary to quantify DNA sequence preference in terms of self-cooperativity between two half-sites that depends on the spacing between them.These models show greatly improved cross-platform predictive performance on ChIP-seq and protein binding microarray (PBM) data. Family-level analysis of these allosteric binding models revealed specific amino-acid residue positions that govern half-site specificity, which we validated experimentally.

Interpreting deep learning genomics models via concept attribution

Jianyu Yang1, Shaun Mahony1

1Penn State University

Interpreting genomics deep learning models remains challenging. Existing feature attribution methods are largely restricted to one-hot DNA inputs, and therefore cannot assess the influence of more general genomic features such as chromatin states or genomic repeats. Concept attribution methods offer an input-agnostic global interpretation framework, yet they have not been systematically applied to interpret neural network applications in genomics.

We present the first application of concept attribution to interpret genomics deep learning models by adapting the Testing with Concept Activation Vectors (TCAV) method. We improve upon the original TCAV method by incorporating a PCA-based decorrelation transformation to address correlated and redundant embedding features commonly observed in genomics deep learning models, resulting in the Testing with PCA-projected Concept Activation Vectors (TPCAV) method. We also introduce a strategy for extracting concept-specific input attribution maps. We evaluate our approach by interpreting influential biological concepts across a diverse set of genomics models spanning multiple input representations and prediction tasks.

We demonstrate that TPCAV provides more reliable DNA motif interpretation than TCAV and provides comparable motif interpretation to TF-MoDISco on one-hot coded DNA-based transcription factor binding prediction models. TPCAV also enables robust interpretive analysis of more general concepts such as repetitive elements and chromatin accessibility and generalizes to tokenized foundation models as well as models incorporating chromatin signals as inputs. We further show that TPCAV can identify representative transcription factor binding sites associated with specific concepts, motivating downstream investigation of distinct binding mechanisms. Overall, TPCAV provides a flexible and robust complement to existing model interpretation techniques.

Dissecting the cis-regulatory code beyond motif syntax

Pablo Mantilla Puccetti1,2, Peter K. Koo2

1School of Biological Sciences, Cold Spring Harbor Laboratory, New York, NY, USA, 2Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, New York, NY, USA

Background. Precise regulation of gene expression underlies development, cellular differentiation, and disease. The cis-regulatory code, which governs how transcription factors (TFs) read regulatory DNA, is characterized by recurrent sequence motifs that serve as TF recognition sites. These motifs form a combinatorial syntax in which their arrangements and spacings influence downstream regulatory processes such as nucleosome remodeling, co-factor recruitment, and ultimately gene expression during development and differentiation. Deep Neural Networks (DNNs) trained to map regulatory sequence to functional outcome serve as powerful virtual experimental platforms, having been shown to learn cis-regulatory syntax including motif distance, orientation, affinity, and flanking sequence effects. However, most virtual experiments remain motif-centric, ignoring regulatory information embedded in surrounding sequence context-a major limitation in our ability to design regulatory elements that recapitulate developmental gene expression programs using known motif grammars.

Results. To address this gap, we designed in-silico experiments to systematically quantify sequence context contributions to model predictions. Leveraging SEAM (Systematic Explanation of Attribution-based Mechanisms), we identify attribution signals robust to partial random mutagenesis to disentangle motif syntax from sequence context, and develop motif-context swap experiments to quantify their respective contributions. In fly, we find that motif syntax alone is insufficient to reproduce enhancer activity on developmental and housekeeping promoters. Instead, sequence context sets a strong baseline activity level while motif syntax further tunes activity within each context, demonstrating that both contribute to model predictions. In human promoters, we systematically characterize Hox gene and oncogene transcriptional start sites to understand context-dependent regulatory logic at disease-relevant developmental loci. Again, we find that motif syntax alone cannot describe the regulatory activity learned by the DNN.

Conclusion. This context-aware dissection of the cis-regulatory code moves us closer to a comprehensive understanding of how regulatory DNA encodes developmental gene expression programs, extending our understanding beyond motif syntax alone.

Collier coordinates shared neuronal identity across two distinct temporal cohorts in a birth-order-dependent manner

Zarion Marshall1, Ellie Heckscher1

1University of Chicago

During neural development, neural stem cells generate sequentially born progeny that assemble into neural circuits in highly reproducible patterns. In many systems, sibling neurons born at similar times are organized into temporal cohorts, groups of neurons that originate from a shared neural stem cell and often adopt related circuit roles based on birth order. Temporal cohorts are molecularly regulated by neural stem cell–derived patterning genes, including temporal and Hox transcription factors, which regulate both the number and identities of neurons produced. However, it remains unclear how sibling neurons from distinct temporal cohorts converge on a shared molecular identity. Here, we investigate how temporal cohorts accomplish shared identity in the Drosophila neuroblast 3-3 (NB3-3) lineage, which produces approximately 10–12 Even-skipped–expressing lateral interneurons (ELs). ELs are divided into early-born and late-born temporal cohorts that preferentially participate in mechanosensory and proprioceptive circuits, respectively. Despite their distinct birth times and circuit functions, both cohorts express the conserved neuronal identity transcription factor Even-skipped (Eve). We find that the COE (Collier/Olf/EBF)-family transcription factor Collier is dispensable for EL neurogenesis and survival but is selectively required for Eve expression in late-born ELs alone. Notably, Collier is broadly expressed in the NB3-3 lineage and transiently present in both early- and late-born ELs, indicating that its cohort-specific requirement cannot be explained by temporally restricted expression. Through enhancer analysis and neuronal rescue experiments, we show that Collier indirectly initiates Eve expression in late-born ELs. Together, these findings demonstrate that temporal cohorts generated from a single neural stem cell can converge on a shared neuronal identity through birth-order-dependent regulatory logic independent of canonical temporal patterning mechanisms.

PISA: a versatile interpretation tool for visualizing cis-regulatory rules in genomic data

Charles McAnany1, Melanie Weilert1, Grishma Mehta1, Fahad Kamulegeya1, Jennifer Gardner1, Jacob Schreiber2, Anshul Kundaje3, Julia Zeitlinger1

1Stowers Institute for Medical Research, 2UMass Chan Medical School, 3Stanford University

Sequence-to-function neural networks learn cis-regulatory sequence rules driving many types of genomic data. However, interpreting these models to relate the sequence rules to underlying biological processes remains challenging, especially for complex genomic readouts such as MNase-seq, which maps nucleosome occupancy but is confounded by experimental bias. We introduce pairwise influence by sequence attribution (PISA), an interpretation tool that combinatorially decodes which bases contributed to the readout at a specific genomic coordinate. PISA visualizes the effects of transcription factor motifs, detects undiscovered motifs with complex contribution patterns, and reveals experimental biases. By learning the bias for MNase-seq, PISA enables unprecedented nucleosome prediction models, allowing the de novo discovery of nucleosome-positioning motifs and their effects, as well as the design of sequences with altered nucleosome configurations.

Furthermore, we use our bias-corrected MNase-seq model to discover barriers that limit the effects of motifs and resemble chromatin domains identified by Micro-C. These results show that PISA is a versatile tool that expands our ability to extract novel cis-regulatory sequence rules from genomics data, paving the way towards deciphering the cis-regulatory code.

Small-effect variants interacting epistatically explain HST2 expression divergence between Saccharomyces cerevisiae and Saccharomyces paradoxus

Rebecca L. McAvoy1, Patricia J. Wittkopp1,2

1Department of Molecular, Cellular, and Developmental Biology, University of Michigan, 2Department of Ecology and Evolutionary Biology, University of Michigan

Gene expression is a fundamental biological process that is a vital link between genotype and phenotype, impacting critical processes including development, structure, and disease. Its regulation is complex with both cis-regulatory elements (e.g. promoters and enhancers) and trans-regulatory (e.g. transcription factors) factors determining gene expression. Models designed to predict gene expression solely from DNA sequence are gaining predictive power, but still fail to predict the impact of genetic variation in many cis-regulatory sequences. These models often focus on predicting transcription factor binding, although a growing body of work shows that changing transcription factor binding often does not correlate with changes in gene expression. Our work aims to identify causative nucleotide(s) impacting gene expression and understanding their relationship to transcription factor binding. Using genome-wide, allele-specific transcription factor binding and gene expression data for species-specific alleles from two closely related species of yeast, Saccharomyces cerevisiae and Saccharomyces paradoxus, in F1 hybrids, we identified orthologous genes with extensive transcription factor binding data and expression divergence to test. We swapped regions of the promoter between the S. cerevisiae and S. paradoxus alleles, used these promoters to drive expression of a fluorescent protein, and measured fluorescence as a proxy for promoter activity. Using this approach for one gene, HST2, revealed multiple nucleotide changes with small impacts on expression that interact epistatically and explain the majority of the expression divergence. Interestingly, our results show that the genetic changes in known transcription factor binding sites are not causing the expression divergence. Further testing is needed to tease apart whether the genetic variation is impacting transcription factor binding in other ways.

3’ Nucleotide Asymmetry Directs miRNA Strand Selection

Jeffrey C. Medley1, Sumire Kurosu Moriya1, Huiwu Ouyang2, Heather Crawshaw1, Sarah Y. Zhang1, Ganesh Panzade1,3, Will J. Sydzyik1, Joel T. Sydzyik1,4, Mira Bhandari1,5, Christopher M. Hammell2, Anna Zinovyeva1

1Division of Biology, Kansas State University. Manhattan, KS, 2Cold Spring Harbor Laboratory. Cold Spring Harbor, NY, 3Laboratory of Human Retrovirology and Immunoinformatics, Frederick National Laboratory for Cancer Research, Frederick, MD, 4University of Kansas School of Medicine, Kansas City, KS, 5Department of Molecular Genetics and Cell Biology, University of Chicago. Chicago, IL

Accurate microRNA (miRNA) strand selection is essential for defining the regulatory landscape of the miRNA-induced silencing complex (miRISC). While 5’ nucleotide identity and duplex thermodynamics have been proposed to bias strand choice, these features cannot fully explain in vivo strand preferences. Here, we uncover a conserved and previously unrecognized role for 3’ nucleotide asymmetry in directing miRNA strand selection in Caenorhabditis elegans and human cells. A favorable 3’ terminal nucleotide on the passenger strand promotes selective loading of the opposing guide strand into miRISC, revealing a cooperative interplay between 5’ and 3’ terminal asymmetries that ensures precise strand discrimination. We show that changes in 3’ nucleotide identity are associated with alternative miRNA strand preference throughout animal development, providing a basis for how strand asymmetry may be regulated in a developmental or tissual context. Collectively, these findings establish a unified, evolutionarily conserved mechanism for miRNA duplex sorting and expand the fundamental rules governing small RNA biogenesis.

Gene Structure Alone Predicts Most Eukaryotic Homology

Armaan Mehta1, Srini Turaga2, David Stern1

1Stowers Institute, HHMI, 2HHMI Janelia Research Campus

Traditional methods of detecting protein homology depend on sequence similarity, but this approach fails when sequences are highly divergent. In earlier studies, we and others have found that highly divergent genes can be identified as homologs from features of their gene structure alone, without any sequence information. Based on these findings, we have explored to what extent machine learning can identify homologous genes across eukaryotes using only information on gene structure. We analyzed several hundred thousand genes from 300 diverse organisms and have identified at least 25 specific patterns of exons, introns, exon phase, and UTR lengths that are highly predictive of gene homology. Furthermore, we show that new bidirectional state-space models can utilize gene structural information to correctly classify 75% of eukaryotic homologs. We trace patterns of gene structure evolution throughout the eukaryotic tree and identify core differences in how gene structures are conserved in different lineages. These models have many applications, including de-orphanization of rapidly evolving genes and improving gene annotations across eukaryotes.

Single-Cell RNA Sequencing of Human Adipose Tissue Identifies Distinct AKR1C2-High Transcriptional States in Depot- and Sex-Specific Adipocyte Progenitors

Ayush Shah1, Alan Ramalho2, Boryana Koseva1, Alex Rader1, Rebecca McLennan1, Andre Tchernof2, Elin Grundberg1

1Children's Mercy Research Institute, Kansas City, Missouri, 2Quebec Heart and Lung Institute, School of Nutrition, Laval University, Quebec, Canada

The AKR1C2 (Aldo-keto reductase family 1 member C2) is a key regulator of adipose tissue function, influencing androgen metabolism and, in turn, fat distribution. While its expression has been positively associated with abdominal obesity specifically in women, exploration of AKR1C2 expression patterns in adipose subpopulations across depots is lacking. To this end, we performed single-cell RNA sequencing on adipose tissue derived from adult individuals (11 Females and 5 Males) across different depots (subcutaneous and omental) undergoing bariatric surgery, profiling a total of 48,175 single cells. First, we mapped key cell populations in human adipose tissue, including immune, endothelial, mesothelial, and adipogenic progenitor cells (APCs), confirming that APCs have unique signatures based on their origin (depot). Next, we showed that AKR1C2 and its partner genes AKR1C1 and AKR1C3 are preferentially expressed in APCs specifically from subcutaneous adipose, with AKR1C1 frequently co-expressed with AKR1C2. When restricting to single cells from the subcutaneous depot, we noted significantly ($P<2E-16$) higher expression of AKR1C1, AKR1C2, and AKR1C3 in females versus males. These findings provide not only a detailed single-cell map of adipose tissue in adults with severe obesity but also highlight AKR1C-positive subpopulations as potentially important contributors to adipose remodeling and local hormone regulation in the context of sex-specific aspects of metabolic disease.

Machine-guided design of large-scale chromatin accessibility patterns in mammalian cells

Shreyash Sonthalia1, Zachary Amador1, Jacob Schreiber2, William Stafford Noble1, Sudarshan Pinglay1

1University of Washington, 2UMass Chan Medical School

Recent advances in DNA design have enabled the engineering of small regulatory sequences (<1 kb) with targeted activity, such as cell type-specific expression. However, extending these approaches to much longer sequences remains limited by both computational design algorithms and experimental methods for validation. Design algorithms typically rely on sequence-to-function model predictions whose computational cost scales with sequence length, while synthesis and genomic delivery of long DNA constructs remain expensive and technically challenging. The ability to design and validate much longer sequences (>10–100 kb) would expand the synthetic biology toolkit toward entire regulatory circuits composed of multiple interacting elements.

To address these challenges, we combined Ledidi, a fast gradient-based design algorithm, with sequence-to-function models and recent advances in bottom-up assembly and delivery of long DNA sequences to mammalian genomes. Using this framework, we designed synthetic DNA with specified regulatory properties and experimentally evaluated these designs in a cost-effective manner. Across in silico analyses, Ledidi more effectively optimized ~20 kb sequences initialized from genome-like distributions toward desired accessibility patterns than sequences initialized from simpler distributions. Attribution analyses suggested that the strongest designs enriched combinations of transcription factor motifs relevant to the tested cell types. When integrated into mammalian genomes and assayed by ATAC-seq, some designs showed concordance with predefined accessibility patterns, whereas others deviated from the intended outputs. These results demonstrate that efficient design algorithms coupled with long-payload delivery can program custom chromatin accessibility patterns in mammalian cells, while also revealing important failure modes in current sequence-to-function models.

Overall, this work establishes a framework for designing increasingly complex combinations of cis-regulatory elements and supports the feasibility of long-context synthetic regulatory DNA design.

Evolution of tissue-specific gene-regulatory programs after vertebrate whole genome duplications

Francois Spitz1

1Department of Human Genetics, University of Chicago

Evolutionary changes are usually gradual. However, in a few instances, animal evolution has been associated with dramatic changes in genome size, organization, and content. Whole-genome-duplications (WGD) provide organisms with a full complement of genes for sub- and neo-functionalization. In the vertebrate lineage, successive rounds of WGD and the chromosomal rearrangements that follow, as well as independent gene losses and cis-regulatory innovation have contributed to the re-wiring and diversification of ancestral cell-type-specific gene regulatory networks (GRN). To identify these changes, we are generating high-resolution single-cell transcriptomic and chromatin accessibility atlases in the lamprey (Petromyzon marinus) and the little skate (Leucoraja erinicea). We find several instances in the lamprey or skate of highly specific transcription factors (TF) added to the core TF complement of conserved vertebrate cell-types. These TFs are usually ohnologs of genes present in mammals but expressed in totally different cell types. These examples provide unique opportunities to investigate how GRNs changed their wiring, their target genes, and their cis-regulatory grammar to accommodate these transcriptional factor gains (or losses). I will discuss a few examples of our findings and how they can shed light on how our ancestors formed new genomic blueprints for vertebrate organismal complexity.

Turnip crinkle virus remodels host chromatin to induce a dark-like physiological state in Arabidopsis thaliana

Alyssa M Stanfield1, Dana J Rademacher1, Akashata Dawane1, Jared P May1

1University of Missouri - Kansas City

Plant viruses extensively reprogram host gene expression to promote infection and disease, yet the chromatin-level mechanisms underlying these transcriptional changes remain poorly understood. Here we investigate how the positive-sense single-stranded RNA virus Turnip crinkle virus (TCV) epigenetically alters histone-mediated chromatin remodeling and gene expression in Arabidopsis thaliana. To determine how TCV shifts chromatin regulation, we used Cleavage Under Targets and Tagmentation (CUT&Tag) to map the active histone mark H3K4me3. Parallel RNA-seq analyses revealed that transcriptionally upregulated genes during infection were enriched for H3K4me3, particularly genes involved in the response to salicylic acid and systemic acquired resistance. In contrast, genes associated with the photosynthetic dark reaction showed reduced H3K4me3.

Notably, trimethylation was reduced at the light-regulated RuBisCO Small Subunit 2B (RBcS2B), a nuclear-encoded chloroplast gene whose expression decreases during darkness. To test whether infection induces a dark-like physiological state, uninfected plants were subjected to a 72-hr dark treatment. RT-qPCR analysis showed that RBcS2B transcript levels decreased and were comparable to those observed during TCV infection.

When plants infected with TCV for 7 days were subjected to a 72-hr dark treatment, viral accumulation increased, suggesting that dark conditions enhance viral fitness. TCV also led to increased formation of RuBisCo containing bodies indicating degradation is also occurring at the protein level. RNA-seq of 72-hr dark-treated plants revealed 28% and 17% overlap with 14 dpi TCV infected plants among down- and upregulated DEGs, respectively. Ongoing CUT&Tag experiments targeting additional histone modifications, including H3K9ac, will further define chromatin remodeling during infection. TCV induces a dark-like state through H3K4me3-mediated chromatin remodeling, reducing RBcS2B at both the transcript and protein level, along with RNA-seq revealing overlapping transcriptional responses with dark-treated plants. Defining the mechanisms by which TCV exploits this dark-like state to enhance viral fitness is the focus of ongoing work.

Inferelator-Multiome: Context-specific prior construction for robust gene regulatory inference

McKenzie Treese1, Hua Li1, Neşet Özel 1

1Stowers Institute for Medical Research

Accurate inference of gene regulatory networks (GRNs) remains a central challenge for understanding developmental regulation in multicellular organisms. Although several methods integrate single-cell transcriptomic and chromatin accessibility data to infer transcription factor (TF)–gene relationships, most GRN inference frameworks remain constrained by their reliance on ground-truth regulatory priors. In complex organisms, where regulatory interactions are highly context-specific and poorly annotated, available priors are often incomplete and potentially irrelevant to the tissue of interest. Here, we build upon the Inferelator 3.0 framework, where prior networks are used to estimate transcription factor activity (TFA) at the single-cell level, rather than to constrain inferred regulatory edges. This enables inference of interactions outside the provided prior and avoids using TF mRNA abundance as a proxy for TFA. However, the current Inferelator-Prior method relies primarily on motif enrichment in accessible enhancers, failing to fully leverage multiomic data.

We introduce Inferelator-Multiome, which can construct high-quality TF-gene priors from scRNA/ATAC-seq data. First, regulatory elements are linked to target genes using expression/accessibility correlations weighted by genomic distance. Then, for each accessible element, TF occupancy is estimated by correlating its accessibility with expression of TFs whose motifs are present within, providing an in silico approximation of TF binding. These information are combined to generate signed TF–gene edges, filtered to retain highest-confidence interactions. By integrating motif information with coordinated variation in TF expression and chromatin accessibility, Inferelator-Multiome moves beyond motif-restricted priors and captures context-specific regulatory evidence.

Benchmarking on Drosophila optic lobe data demonstrates that Inferelator-Multiome dramatically improves network performance compared to Inferelator-Prior and improves recovery of known regulatory interactions. Importantly, networks inferred using Inferelator-Multiome also show improved predictive power on gene expression changes upon perturbation of specific TFs. Together, these results highlight that prior quality is critical for inferring accurate GRN models with greater biological explanatory power.

Widespread low-affinity motifs enhance chromatin accessibility and regulatory potential

Melanie Weilert1, Kaelan J. Brennan1, Khyati Dalal1,2, Sabrina Krueger1, Haining Jiang1, Rosa Martinez-Corral3, Julia Zeitlinger1,2

1Stowers Institute for Medical Research, Kansas City, MO, 64110, USA, 2Department of Pathology & Laboratory Medicine, The University of Kansas Medical Center, Kansas City, KS, 66160, USA, 3CRG (Barcelona Collaboratorium for Modelling and Predictive Biology), Dr. Aiguader 88, Barcelona 08003, Spain

Low-affinity transcription factor (TF) motifs are an important element of the cis-regulatory code, yet they are notoriously difficult to map and mechanistically incompletely understood, limiting our ability to interpret non-coding variation in development, evolution, and disease. Here we investigate their role in pioneering and leverage sequence-to-profile models of chromatin accessibility in mouse embryonic stem cells to reliably map and interpret low-affinity motifs across the genome. We find that low-affinity motifs have outsized effects by cooperating with nearby motifs through intra-nucleosomal soft syntax. By modeling nucleosome-mediated cooperativity with a kinetic model, we discover and validate that pioneer cooperativity makes a motif operate at higher pioneering ranges across changing TF concentrations, thereby raising the regulatory potential. These results show that low-affinity motifs can be accurately mapped, shape the properties of developmental enhancers and likely play a widespread role in fine-tuning enhancers during evolution.

Deep mutagenesis and deep learning to understand epistatic interactions among transcription factor binding sites in a model promoter

David Granas1, Daniel Lyon1, Ryan Friedman1,2, James Shepherdson1, Daniel Murphy3, Joseph Corbo3, Michael White1

1Department of Genetics, Washington University in St. Louis, 2Department of Genome Sciences, University of Washington, 3Department of Pathology & Immunology, Washington University in St. Louis

Genetic variation in regulatory DNA has a substantial impact on health and disease. Variants can affect regulatory DNA function by altering transcription factor (TF) binding sites, but predicting the impact of such variants on transcription is challenging because the function of a TF binding site is often highly sensitive to the local context. We sought to learn the epistatic interactions between TF binding sites and local context by coupling massively parallel reporter assay (MPRA)-based deep mutagenesis with interpretable deep learning of the Rho promoter, a long-time paradigmatic cell-type-specific regulatory element. Two high-affinity TF binding sites matching the consensus motif were largely intolerant of sequence variants, while a conserved, non-canonical TF binding site was more tolerant of affinity-altering variants. A 3 or 4 bp spacing between two highly conserved binding sites for the TFs CRX and NRL was critical for strong promoter activity, but beyond 4 bp there was little effect of increasing these distance between these binding sites. A convolutional neural network (CNN) trained on MPRA data from retina open chromatin regions was able to accurately predict the effects of variants in the Rho promoter, including variants that increased promoter activity. To discover the TF binding site interactions learned by the model, we are performing systematic in silico sequence perturbations and comparing the results to data collected from extensive motif rearrangements of Rho promoter binding sites. Our results demonstrate how TF binding sites within a single promoter vary in their sensitivity to local context and affinity altering variants.

Postnatal conversion of methylcytosine to hydroxymethylcytosine reconfigures the human neuronal epigenome

Heng Xu1, Jo-Fan Chien2, Alexey Kozlenkov3,4, Ramu Vadukapuram3,4, Junhao Li5, Yu Wei6, Andrew J. Dwork7, Chunyu Liu6, Stella Dracheva3,4, Eran A. Mukamel5

1Bioinformatics and Systems Biology, University of California San Diego, La Jolla, CA 92037, US, 2Department of Physics, University of California San Diego, La Jolla, CA 92037, US., 3Research & Development and VISN2, James
J. Peters VA Medical Center, Bronx, NY, 10468, US., 4Friedman Brain Institute and Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY 10029, US., 5Department of Cognitive Science, University of California San Diego, La Jolla, CA 92037, US., 6Department of Psychiatry, SUNY Upstate Medical University, Syracuse, NY 13244, US., 7Department of Pathology and Cell Biology, Columbia University, NY 10027, US.

Background: Healthy brain development requires a coordinated process of postnatal cellular maturation throughout the first two decades of life that transforms neuronal morphology, connectivity, physiology, and gene expression. DNA methylation is a major component of this process, but the distinct developmental roles of 5-methylcytosine (mC) and 5-hydroxymethylcytosine (hmC) remain poorly understood because most assays do not distinguish these marks. In neurons, which are highly enriched for hmC, resolving these modifications may reveal how epigenomic state is linked to transcriptional and cis-regulatory programs during development.

Results: We profiled mC and hmC in excitatory and inhibitory neurons from human prefrontal cortex across 103 donors spanning 38 days to 77 years of age using bisulfite and oxidative-bisulfite sequencing, together with transcriptomic and histone modification data. We found widespread postnatal conversion of mCG to hmCG, with up to half of all CG dinucleotides undergoing this transition over the first decade of life, indicating extensive remodeling of the neuronal epigenome during maturation. hmC accumulated asymmetrically on the sense strand of actively transcribed genes and increased in a linear, clock-like manner across the lifespan, supporting a close relationship between transcription and hmCG. At dynamic cis-regulatory elements, changes in hmCG were associated with coordinated remodeling of active and repressive histone marks in a cell-type-specific manner, linking hydroxymethylation to developmental regulatory state transitions. We also identified sex-specific differences in X-linked DNA methylation driven primarily by hmCG rather than mCG.

Conclusion: Our results identify 5hmC as a major feature of postnatal neuronal epigenome remodeling and implicate it in the regulation of transcriptional and cis-regulatory programs across human brain development. These findings position hmC as an important layer of the regulatory code underlying neuronal maturation and stable cell identity.

Reinforcement learning enables de novo design of tissue-specific enhancers through motif-level regulatory grammars

Liwen Yao1, Liangqi Xie1

1Microbial Sciences in Health, Cleveland Clinic Research

Designing functional enhancers from first principles requires understanding the regulatory grammar through which transcription factor binding sites encode tissue-specific activity. Current generative approaches operate in nucleotide space, where the design units are misaligned with the motif-level logic that the transcriptional machinery reads. Whether this representation limits design capacity has not been systematically tested.

Here we develop a reinforcement learning framework that designs tissue-specific enhancers by composing explicit motif grammars. Using Group Relative Policy Optimization (GRPO) guided by a VISTA-trained Enformer oracle, we compare two strategies operating at different levels of abstraction. De novo nucleotide-level generation produces functional enhancers under single-tissue optimization (69.9% active in STARR-seq for heart), but fails comprehensively under contrastive training designed to enforce tissue selectivity: on-target activity collapses, canonical motifs are replaced by non-canonical signatures, and tissue specificity inverts—revealing reward hacking enabled by the unconstrained nucleotide action space. We therefore introduce motif embedding, in which a policy network learns to select, position, and orient transcription factor binding sites from an interpretable motif library derived from Enformer attribution analysis. This representation enables what nucleotide space cannot: contrastive optimization achieves up to 68.5% on-target activity with up to 69-fold tissue selectivity in STARR-seq across seven embryonic tissues. The learned grammars recover known developmental programs—MEF2/GATA/TBX5 for heart, SOX2/RFX/ATOH1 for neural tissues—without explicit supervision of TF–tissue relationships. Ablation experiments confirm that activity depends on combinatorial grammar rather than motif vocabulary alone. Transgenic mouse assays validate that designed 200-bp enhancer cassettes drive tissue-specific expression in vivo, achieving comparable function to native enhancers one-tenth their size.

These results establish that successful enhancer design requires representational alignment between design units and transcription factor motifs, and objective alignment between optimization and regulatory function.

Unlocking cis-regulatory landscapes across 500 million years of evolution and disease mechanisms

Tássia Mangetti Gonçalves , Casey L. Stewart2 , Samantha Baxley2, Jason Xu3 , Kevin Boyer1, Bijesh George , Daofeng Li1 , Chengran Yang4, Harrison W. Gabel5, Xianhua Piao6, Carlos Cruchaga4, Yang E. Li7, Ting Wang , Oshri Avraham2, Guoyan Zhao1

1Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110, USA., 2Department of Cellular Biology, University of Georgia, Athens, GA, 30602, USA., 3Missouri University of Science & Technology, Rolla, MO 65409, USA., 4Department of Psychiatry, Washington University School of Medicine, St. Louis, MO 63110, USA., 5Department of Neuroscience, Washington University School of Medicine, St. Louis, MO 63110, USA., 6Division of Neonatology, Department of Pediatrics, University of California, San Francisco, CA, USA., 7Department of Neurosurgery, Washington University School of Medicine, St. Louis, MO 63110, USA.

Background: Genomic DNA encodes regulatory information that determines where, when, and to what extent genes are expressed. Theoretically, we should be able to identify these transcriptional “instructions” by examining genomic DNA sequence alone, yet this has remained challenging.

Results: Here we present the Vertebrate Regulatory MOdule Detector (VRMOD), a method that accurately predicts gene regulatory sequences using only the query genomic sequences. We applied VRMOD to 309 Ensembl genomes, generating a compendium of high-resolution, genome-position-fixed cis-regulatory modules without parameter tuning. We performed extensive computational evaluation and experimental validation of VRMOD predictions. Notably, VRMOD predicted three sub-enhancers within the human hs52 enhancer at the FTO locus from the VISTA database, including one missed by existing methods. Using a chicken embryo system and 3D tissue imaging, we showed that each sub-enhancer exhibits restricted spatiotemporal activity within specific subsets of tissues where the full enhancer is active. We further demonstrated VRMOD’s utility for identifying evolutionarily non-conserved enhancers, annotating regulatory sequences in non-model organisms, and identifying candidate disease-causal variants. By integrating 12 genomic and epigenomic datasets with 10 neurological disease-associated single-nucleotide polymorphisms (SNPs) and expression quantitative trait loci (eQTLs), VRMOD enables systematic prioritization of candidate disease-causal variants located in non-coding regions.

Applications of VRMOD across diverse biological systems have generated testable hypotheses regarding transcription factors, transcription factor binding site disruptions, and cell-type-specific cis-regulatory mechanisms in neurodegenerative diseases and chronic pain. The VRMOD CRM UCSC Genome Browser public hub (https://genome.ucsc.edu/s/gzhao/VRMOD_hg38) and WashU Epigenome Browser enable users to explore VRMOD-predicted CRMs, supporting experimental evidence, candidate neurological-disease-causal variants, and additional regulatory annotation.

Conclusion: Collectively, VRMOD provides a universal coordinate reference system for regulatory sequences across 309 vertebrate genomes, and a universal framework for elucidating regulatory mechanisms underlying disease and evolution across 309 vertebrate genomes.

Investigate enhancer activity associated with metabolism-related traits across mammals using TACIT

Junjie Ma1, Andrew Bellesis1, Maddie Kellogg1, Wynn Meyer3 and Irene Kaplow1,2

1Department of Biological Sciences, Carnegie Mellon University, 2Department of Ray and Stephanie Lane Computational Biology, Carnegie Mellon University, 3Department of Biological Sciences, Lehigh University

Multiple dietary specializations, including herbivory, carnivory, and omnivory, have evolved multiple times during mammalian evolutionary history, enabling mammals to inhabit diverse ecological niches. Some studies have identified morphological adaptations in organs such as teeth and intestines associated with these phenotypes, but the molecular mechanisms underlying the dietary evolution remain less understood. A recent study identified protein-coding regions associated with dietary phenotype evolution and found only weak associations; the genes with identified associations tended to be highly expressed in liver, suggesting that there may also be a role of liver transcriptional regulatory element changes in dietary phenotype evolution. We have generated data using ATAC-seq, an assay for transcriptional regulatory element activity, from the livers of mammals with a diversity of diets and integrated it with publicly available data. We plan to use this data to train machine learning models to predict liver transcriptional regulatory element differences between species then make predictions of liver transcriptional regulatory element activity at orthologous regions in hundreds of mammals using the Vertebrate Genomes Project Cactus alignment. We will then associate these predictions with various dietary phenotypes, including carnivory and percentages of macronutrients such as fiber in the diet. We will then use publicly available liver RNA-seq and Hi-C data, which measures three-dimensional DNA interactions, to identify genes likely to be regulated by these transcriptional regulatory elements and test these regulatory relationships using CRISPR knock-out experiments. We anticipate that our findings will reveal transcriptional regulatory mechanisms involved in the evolution of dietary phenotypes and that some of them will provide insights into human disorders associated with meat consumption, such as type 2 diabetes and non-alcoholic fatty liver disease.

Dissecting the cooperative, context-dependent gene regulatory syntax in human development

Selin Jessa1, Betty Liu1, Samuel Kim1, Yan Ting Ng2, Soon il Higashino1, Georgi K. Marinov1, Eyal Ben-David2, Kyle Farh2, Anshul Kundaje1, William Greenleaf1

1Department of Genetics, Stanford University
2Illumina AI Lab for Genome Interpretation

Transcription factors (TFs) establish cell identity during development by binding noncoding regulatory DNA in a sequence-specific manner, often promoting local chromatin accessibility, and regulating gene expression. Yet, the map of TFs that bind the genome in each cell type remains incomplete. Furthermore, we lack an understanding of how the syntax of binding sites—their composition, orientation, and spacing—contributes to combinatorial TF activity and cell type-specific regulation.

Here, we profiled chromatin accessibility and gene expression in 800k cells from 12 human fetal organs between post-conception weeks 10-23, annotating 203 cell types. For each cell type, we trained a deep convolutional neural network model (ChromBPNet) to predict basepair-resolution chromatin accessibility profiles in open chromatin regions from local DNA sequence. Systematic interpretation of each model identified recurrent sequence patterns which influence chromatin accessibility, corresponding to recognition motifs for TFs. Through this de novo motif discovery, we identify 508 unique motifs across 189 models, including known ubiquitous and cell type-specific motifs, putative novel motifs, as well as motifs which are highly abundant in the genome yet inhibit accessibility. Using our trained models together with an in silico experimentation framework, we systematically dissected motif cooperativity across spacings and orientations, and discovered dozens of composite motifs with “hard” syntactic rules requiring precise motif organization, or “soft” rules allowing flexible motif arrangements.

Altogether, we show that deep learning models trained on epigenetic features can reveal novel rules of TF binding site grammar and cis-regulatory element

organization. Our work delineates how motif syntax governs cell type-specific chromatin accessibility and provides a foundational resource for decoding cis-regulatory logic and interpreting genetic variation during human development.

Dissecting the cooperative, context-dependent gene regulatory syntax in human development

Selin Jessa1, Betty Liu1, Samuel Kim1, Yan Ting Ng2, Soon il Higashino1, Georgi K. Marinov1, Eyal Ben-David2, Kyle Farh2, Anshul Kundaje1, William Greenleaf1

1Department of Genetics, Stanford University
2Illumina AI Lab for Genome Interpretation

Transcription factors (TFs) establish cell identity during development by binding noncoding regulatory DNA in a sequence-specific manner, often promoting local chromatin accessibility, and regulating gene expression. Yet, the map of TFs that bind the genome in each cell type remains incomplete. Furthermore, we lack an understanding of how the syntax of binding sites—their composition, orientation, and spacing—contributes to combinatorial TF activity and cell type-specific regulation.

Here, we profiled chromatin accessibility and gene expression in 800k cells from 12 human fetal organs between post-conception weeks 10-23, annotating 203 cell types. For each cell type, we trained a deep convolutional neural network model (ChromBPNet) to predict basepair-resolution chromatin accessibility profiles in open chromatin regions from local DNA sequence. Systematic interpretation of each model identified recurrent sequence patterns which influence chromatin accessibility, corresponding to recognition motifs for TFs. Through this de novo motif discovery, we identify 508 unique motifs across 189 models, including known ubiquitous and cell type-specific motifs, putative novel motifs, as well as motifs which are highly abundant in the genome yet inhibit accessibility. Using our trained models together with an in silico experimentation framework, we systematically dissected motif cooperativity across spacings and orientations, and discovered dozens of composite motifs with “hard” syntactic rules requiring precise motif organization, or “soft” rules allowing flexible motif arrangements.

Altogether, we show that deep learning models trained on epigenetic features can reveal novel rules of TF binding site grammar and cis-regulatory element

organization. Our work delineates how motif syntax governs cell type-specific chromatin accessibility and provides a foundational resource for decoding cis-regulatory logic and interpreting genetic variation during human development.

Programming human cell type-specific gene expression via AI-designed enhancers

Sebastian M. Castillo-Hair1, Christopher H. Yin1, Leah VandenBosch2, Timothy Cherry2, Wouter Meuleman3,4, Georg Seelig1,4

1Department of Electrical & Computer Engineering, University of Washington, Seattle, WA, 2Center for Developmental Biology and Regenerative Medicine, Seattle Children's Research Institute, Seattle, WA, USA, 3Altius Institute for Biomedical Sciences, Seattle, WA, 4Paul G. Allen School of Computer Science & Engineering, University of Washington, Seattle, WA

Across tissues, developmental stages, and disease conditions, cells adopt distinct states characterized by unique molecular profiles. The ability to write DNA-encoded programs that sense and modulate cell states hold transformative potential for biotechnology, with applications that include smart gene therapies with targeted activity and guided stem cell differentiation for regenerative medicine. Yet, an incomplete understanding of how core cellular processes such as gene expression are regulated across states limits their rational design. To overcome these challenges, we developed Artificial Intelligence (AI) models to design synthetic DNA elements to selectively target gene expression across cell states.

Trained on a large corpus of chromatin accessibility data, our models enabled us to generate an atlas of tens of thousands of synthetic regulators with specificity towards hundreds of human cell types, tissues, and in vitro differentiated states. Experimental testing of thousands of designs in a representative subset of ten human cell lines and in mouse retinas demonstrated their function as specific regulators of gene expression, not only in the case of one-versus-all objectives but also when targeting two or three cell types. An explainable AI analysis of synthetic sequences allowed us to identify "grammar" features learned from natural sequences and amplified by our models, such as combinations of transcription factor binding sites. Our work resulted in the largest resource of synthetic regulators of cell type-specific expression to date, paving the way for programmable, state-aware genetic programs that can be leveraged for biotechnology and therapeutics.

Context-dependent role of Snail repressor in early Drosophila development

Kimberly Escobar Alvarado1, 2, Melanie Weilert1, Julia Zeitlinger1, 2

1Stowers Institute for Medical Research, Kansas City, MO, 2University of Kansas Medical Center, Kansas City, KS

Transcriptional repressors bind sequence-specific motifs within enhancers and counteract the effect of transcriptional activators, but the molecular mechanisms underlying this repression remain poorly understood. Similarly, how the arrangement of repressive motifs within enhancers influences repression is unclear. To address these questions, we used the deep learning model BPNet to investigate the role of the repressor Snail during dorsoventral patterning in the early Drosophila embryo. By training BPNet to predict ChIP-nexus data from sequence alone, we found that many Snail motifs prevent the binding of other transcription factors, though the mechanism remains unresolved. ATAC-seq and H3K27ac ChIP-seq experiments done in mutant embryos with ubiquitous Snail expression revealed that Snail binding reduces chromatin accessibility and H3K27 acetylation in a sequence context-dependent manner.

Differential modeling of H3K27ac data in the ubiquitous Snail-expressing mutant vs. a control mutant where Snail targets are derepressed, further supports Snail's role in repressing this active enhancer mark. To elucidate these context-dependent effects, we are now investigating whether Snail influences nucleosome stability, given the high nucleosome occupancy observed at Snail-repressed enhancers.

3D Genome-informed Gene Regulation Modeling Links Noncoding Diabetes Variants and Enhancers to Target Genes in Pancreatic Differentiation

Nan Zhang[1, #], Yang Yang[2, #], Jiaxin Li[2,3, #], Wilfred Wong[2,3, #], Xianming Wang[1, #], Qianzi Li[2,3], Nan Hu[1], Christopher McGinnis[4], Ann Le[4], Ansuman Satpathy[4], Kushal Dey[2], Christina S. Leslie[2,*], Danwei Huangfu[1,*]

[1]Developmental Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA, [2]Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA, [3]Tri-institutional Training Program in Computational Biology and Medicine, New York, NY 10065, USA, [4]Department of Pathology, Stanford University, Stanford, CA, USA; Department of Bioengineering, Stanford University, Stanford, CA, USA; Department of Immunology, Stanford University, Stanford, CA, USA; Parker Institute for Cancer Immunotherapy, San Francisco, CA, USA; Stanford Cancer Institute, Stanford University, Stanford, CA, USA; Gladstone-UCSF Institute of Genomic Immunology, San Francisco, CA, USA. #indicate equal contribution; *corresponding authors

Enhancer-gene interactions are essential for gene regulation. Genome-wide association studies have uncovered large numbers of disease-associated genetic variants, most of which reside in the non-coding genome and are presumed to impact gene expression. As non-coding variants may act through disrupting enhancers, linking enhancers to target genes is important for elucidating disease mechanisms. However, linking distal enhancers to target genes remains a challenge. 3D genome organization and multi-omics provide complementary information for identifying regulatory interactions. We developed GraphReg+, a novel extension of the GraphReg algorithm, which utilizes graph attention networks to perform 3D contact-informed gene regulation inference. GraphReg requires CAGE data, which quantifies gene expression at transcription start sites and is unavailable in our setting. GraphReg+ leverages two sub-graphs to establish an integrated graph, gaining generalizability to broader gene expression profiling techniques. The first sub-graph captures 3D interactions between genomic loci to enable information exchange between regulatory elements. The second sub-graph connects genes with genomic loci containing the transcription start sites, enabling gene-level model supervision. Moreover, GraphReg+ combines chromatin accessibility with DNA sequence features to augment feature representations of genomic loci. We apply GraphReg+ to single-cell mult-omics and HiCAR data at key stages in guided differentiation of human pluripotent stem cells to late endocrine cells, to infer enhancer-gene interactions during pancreatic differentiation and systematically link diabetes risk variants to potential target genes. Through feature attribution analysis, we identified distal enhancers of important pancreatic developmental genes, several of which were supported by existing studies. We inferred target genes of 1220 fine-mapped diabetes risk variants, including 527 variants with associations at a distance over 500Kb. Our method enables identifying candidate regulatory interactions to prioritize for experimental validation, facilitates a more comprehensive understanding of enhancer rewiring in pancreatic differentiation, and advances functional interpretation of disease risk variants.

How data quality affects the interpretability of sequence-to-function models

Minal Khatri1, Julia Zeitlinger1
1Stowers Institute for Medical Research, Kansas City, MO, USA

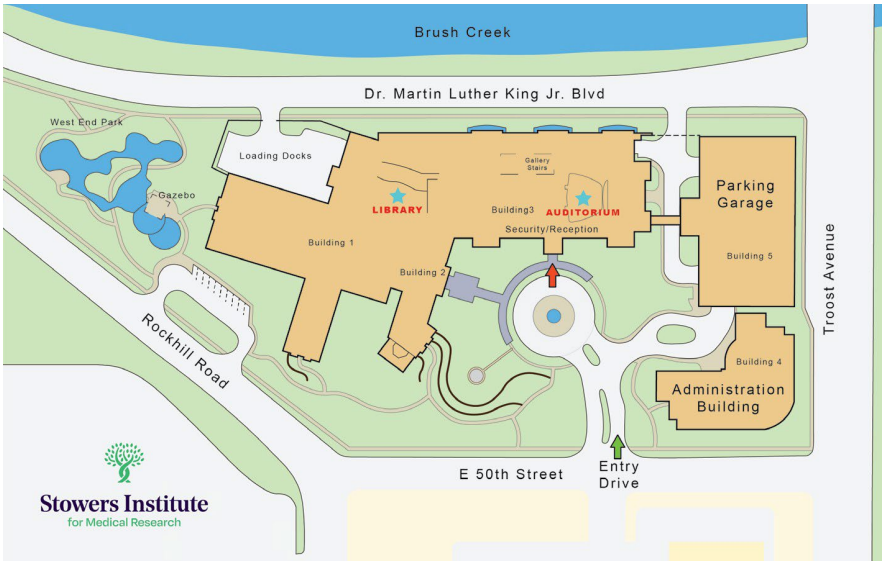
Sequence-to-function models such as BPNet (BPReveal) predict genomic assay signal from DNA sequence, and interpretation of these models enables the identification of the underlying transcription factor binding motifs across the genome. The accuracy of the regulatory patterns learned by model is important for predicting genetic variants, yet it is unknown how low coverage settings such as single-cell or pseudo-bulk data affect the quality of the model interpretation. Here, we systematically evaluate the effect of sequencing depth on both model performance and interpretability by training BPNet models on down-sampled ChIP-nexus and ATAC-seq datasets across a range of read depths. Interpretability is assessed using motif discovery and motif instance recovery. We find that predictive performance decreases only slightly with sequencing depth, and this in part because signal strength diversity is maintained in low depth data. However, the model interpretation is more severely affected by sequencing depth, with weaker motif instances being preferentially lost. These results suggest that predictive performance alone does not fully capture changes in learned regulatory features.

Attendee Resources

Location: Stowers Institute for Medical Research – 1000 E. 50th Street, Kansas City, MO 64110
General Phone Number: 816-926-4000
Driving directions from Kansas City International Airport:
(approximately 30 minutes under normal driving conditions)

1. Follow the signs on International Circle to the airport exit.
2. Take Cookingham Drive to I-29 South (right ramp), heading toward Kansas City.
3. Continue to follow I-29 South as it merges into I-29 South 71 (merge in left lane), then I-35South (merge in left lane again).
4. Cross the Paseo Bridge and get in the lane for I-70 South 71 (Exit 3)
5. Merge right and follow sign to South 71 Highway (Exit 2M).
6. Take the Emmanuel Cleaver II Boulevard exit and turn right onto Emmanuel Cleaver II Boulevard.
7. Follow Cleaver II to Troost Avenue and turn left on Troost.
8. Turn right on 50th Street and take another immediate right at the Stowers Institute's entrance (1000 E. 50th Street, Kansas City, MO 64110).

Campus Map



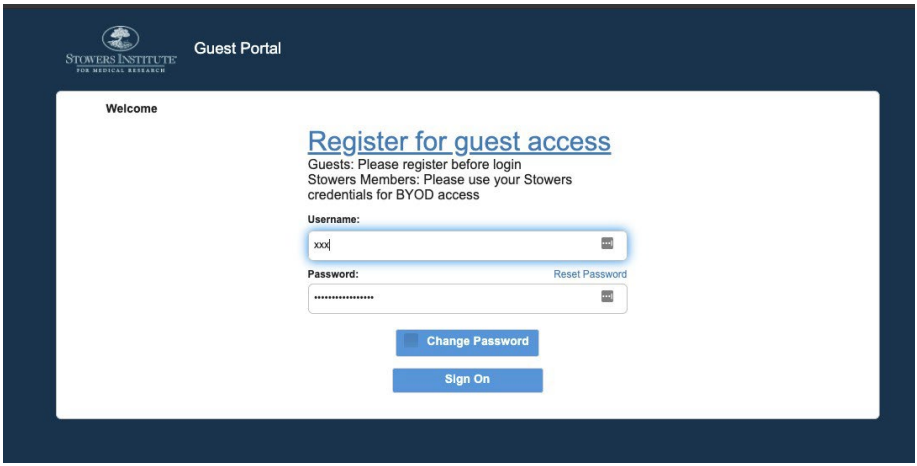
Parking: Visitor parking is available in the parking structure located between the Administration Building (to your right as you enter the campus) and the Research Building (to your left). Please park on the 5th floor of the parking garage and take the elevator down to the 1st floor to enter the Research Building.

Restrooms: Located by the two large planters on the 1st floor of the Research Building. All-gender restrooms are available at the base of the stairs on the B1 level.

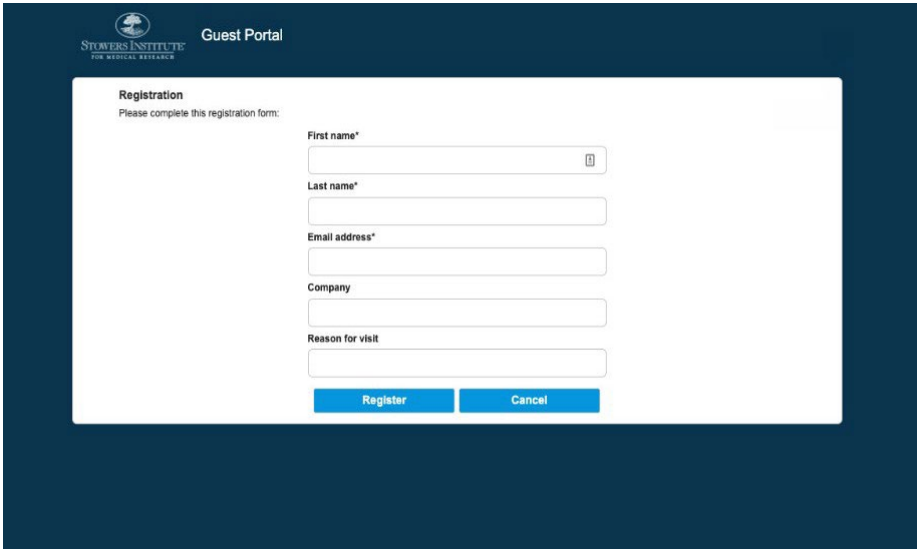
Mother's Room: At the base of the stairs on the B1 level, located next to the all-gender restrooms. An entry key will be provided by security to guests that request to use it.

Tobacco Free Campus: Tobacco of any kind is prohibited in all the Institute facilities and on the Institute grounds. The Institute does not offer any designated smoking areas.

Wireless Access: Stowers Guests may connect to the **stowers_guest SSID**
When connected, you will see a captive portal. Click on the link at the top of the page to register for guest access.



- You will be prompted to enter their information. Name and email are the only required fields.
- After entering the required information click Register.



- You will see a page like this with sign on information. You can use the userid and password to connect other devices without going through the registration, by entering the userid and password in the first page of the captive portal.

- Click the "Sign On" button.

After clicking sign on, you will be prompted to accept the "Acceptable use Policy". After clicking accept you should now have internet access.

If you have any questions, please contact the Stowers Help Desk at 816-926-4150.

Kansas City Fun: The Kansas City area offers a wealth of cultural, educational, and entertaining opportunities to explore, many of which are inexpensive or free of charge. The following is a brief listing of local attractions, event calendars, and resources.

Useful Resources
Information about Kansas City and the surrounding area can be found at kcmo.gov or at visitkc.com. Two publications to help you discover things to do in the area are: "Insider's Guide to Kansas City" by Katie van Luchene and "Day Trips from Kansas City" by Shifra Stein. Both can be found in local bookstores.

Linda Hall Library
5109 Cherry Street, 816.363.4600
The Linda Hall Library in Kansas City is one of the world's foremost independent science research libraries. Founded in 1946 through a philanthropic bequest, it houses vast collections spanning science, engineering, and technology, serving researchers, students, and the public. Known for its international research programs and collaborations, the Library promotes lifelong learning with innovative programming, digital access, and educational resources. The Library also maintains its grounds as a public urban arboretum, offering both scholarly and community engagement opportunities.
www.lindahall.org

Nelson-Atkins Museum of Art
4525 Oak Street, 816-751-1278
Opened in 1933, the Nelson-Atkins Museum has more than 50 galleries and several period rooms. The museum's outstanding feature is its collection of Asian art. The collection of Chinese landscape paintings is one of the finest in the West, and the museum's holdings of Chinese ceramics and decorative arts are also noteworthy. Besides European paintings from the Renaissance on, the museum also has notable collections of ancient Egyptian sculpture, Japanese porcelains and lacquer, and English pottery. The E.F. Pierson Sculpture Garden was dedicated

in 1972, and the Henry Moore Sculpture Garden opened in 1989. Admission is free.
nelson-atkins.org

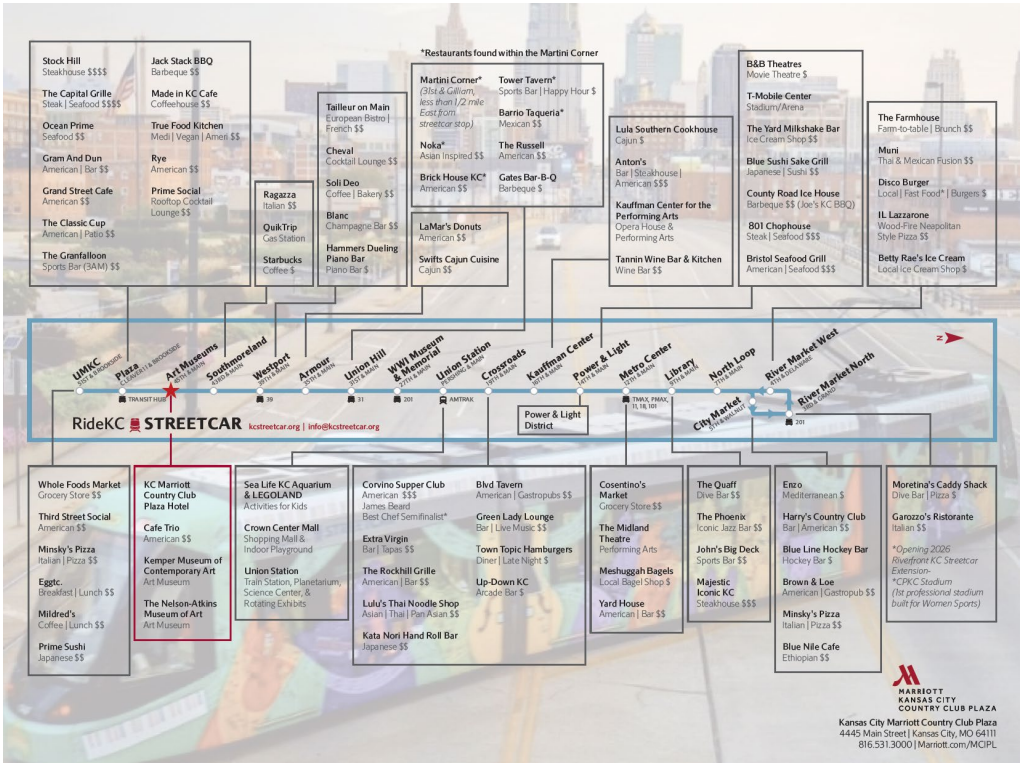
Kemper Museum of Contemporary Art
4420 Warwick, 816-753-5784
Founded in 1994, the Kemper Museum of Contemporary Art presents modern and contemporary art of the highest quality and significance. It collects, preserves, documents, interprets, and exhibits a growing permanent collection; develops and presents special exhibitions; and offers a variety of educational programs. Admission is always free, and the Museum serves a diverse and inclusive public population.
kemperart.org

Union Station
30 West Pershing Road, 816-460-2020
This fully restored 1914 landmark is Kansas City's most prominent destination for entertainment and cultural activities. The Station is home to a permanent rail exhibit with vintage rail cars, an interactive science center, a vibrant Theater District featuring giant screen movies and live theater, fine restaurants, unique shops, spaces for meetings and events and much more. Of course, you can still catch the train at Union Station, once again among Amtrak's busiest stops.
unionstation.org

Loose Park
Intersection: Wornall Road and 51st Street, 816-784-5300
Loose Park is one of Kansas City's most beautiful parks. The park is home to a lake, a walking path, a shelter house, Civil War markers, tennis courts, a wading pool, picnic areas and the famous Rose Garden. The Rose Garden is popular for all types of outdoor special events including theatrical performances and wedding ceremonies.

Westport
Westport is one of Kansas City's premier destinations for dining, shopping, site seeing and is the heart of the city's nightlife. Located in the midtown, Westport is just north of the Country Club Plaza and a few miles south of downtown Kansas City. Historically, Westport was built along the Santa Fe Trail as an outfitting center for wagon trains heading west. Today the area is filled with renovated and new buildings housing trendy shops, restaurants, and nightspots.

RideKC Streetcar Map



SRC Meeting Policies

The Stowers Research Conference (SRC) series supports an environment for the exchange of scientific ideas that is grounded in dignity and respect for all program participants. SRC believes that a diverse, inclusive and collegial community culture promotes scientific creativity and progress. The conference code of conduct outlined here has been adapted from the policies outlined by the Society for Developmental Biology, sponsor of the first SRC meeting.

No Harassment Policy

Program participants are expected to conduct themselves in a professional manner and to treat each other with dignity and respect. This expectation applies to the organizers, event attendees, volunteers, employees, consultants, vendors, and others while on Stowers premises, while representing SRC elsewhere, and while attending events organized by SRC.

SRC will not tolerate any discrimination in the form of sexual harassment and other forms of harassment. Program participants shall not engage in any conduct that could reasonably be construed as unlawful harassment against an individual. Program participants shall not make unwelcome sexual advances, make requests for sexual favors, or engage in other verbal or physical conduct of a sexual or offensive nature.

No Violence Policy

SRC does not tolerate any type of violence or threats of such violence while on Stowers premises, while representing SRC elsewhere, and while attending events organized by SRC. SRC prohibits acts or threats of violence by or against any program participants. In addition, SRC does not permit the possession or the concealed or open carrying of weapons anywhere on the Stowers premises.

Reporting

If a program participant discovers any conduct which they believe violates this policy or is otherwise detrimental to the organization, they are asked to promptly report it to the organizers.

Recording, Photography, and Session Etiquette

While in sessions, please mute all cell phones and other electronic devices. Photography or the electronic capture of scientific sessions and posters is not permitted without the expressed consent of the presenting author(s). Respect other individuals' and organizations' intellectual property and confidential information.

Photo Release

SRC has an official photographer for the meeting. Photos taken at the meeting may be used in future SRC publications, on the SRC website, or in other materials.

By registering for the meeting, you agree to allow SRC to use your photo in and SRC related publication or website.

Smoking Policy

The use of tobacco, in any form, or the use of any smoking device is prohibited in all Stowers facilities and on the Stowers campus. Stowers facilities are defined as common work areas, auditoriums, classrooms, conference and meeting rooms, private offices, elevators, hallways, food service facilities, employee lounges, stairs, restrooms, vehicles, the parking garage, and all facilities owned by the Stowers Group of Companies.

Stowers Institute

Stowers.org

Creative, innovative, fearless science

At the Stowers Institute, we aim to attract scientists who are visionary risk-takers, who seek to solve the most challenging and intractable problems in biology.

Meet the Scientists at the Stowers Institute for Medical Research
stowers.org/scientists