Cascadia Advanced Genomic Technologies Abstract Book

June 19-20, 2025, Vancouver, BC, Canada

Conference Program

Day 1 - June 19th

Time	Event
14:00-14:30	**Registration and Poster Setup**
14:40-14:45	**Welcome and Introduction**

Session I: Emerging Omics Technologies

14:45–15:15	*Saturation deletion scanning of the human genome.* **Jonas Koeppel**
15:15–15:30	*Single cell Strand-seq: a unique tool for genome analysis.* **Peter Lansdorp**
15:30–15:45	*Cuffed CRISPR guide RNA for microRNA activity-dependent genome editing.* **Arman Adel**
15:45–16:00	Break

Session II: Computational Approaches I

16:00–16:30	*Multi-scale deep phenotyping and statistical derivatization of mammalian developmental enhancers.* **Jean-Benoit Lalanne**
16:30–16:45	*Dissecting the cooperative, context-dependent gene regulatory syntax in human development.* **Selin Jessa**
16:45–17:00	*iQcell2.0: A GNN with Interpretable Message Passing for Single Cell Transcriptomics.*

	Tiam Heydari
17:00–17:15	**Break**

Session III: Future Frontiers

17:15–17:30	*Virtual cells in virtual human development: an international effort.* **Nika Shakiba**
17:30–17:45	*Reconstitution of mammalian life cycle in vitro.* **Nobu Hamazaki**
17:45–18:30	**Flash Talks**
18:30–20:30	**Dinner and Poster Presentations**

Day 2 - June 20th

Time	**Event**
8:00-9:00	**Breakfast**
9:00-9:40	**Keynote Speaker: Calin Plesa**

Session I: SynBio Parts

9:40 - 10:10	*Capsule-based single cell genome sequencing.* **Sanjay Srivatsan**
10:10–10:25	*Diversified, miniaturized and ancestral parts for mammalian genome engineering and molecular recording.* **Troy McDiarmid**
10:25–10:40	*Understanding upstream dominance in synthetic gene circuits in human cells.* **Ross Jones**
10:40–10:55	**Break**

Session II: Lineage Tracing

10:55–11:25	*Technologies for "look-back-in-time" biology.* **Nozomu Yachie**
11:25–11:40	*Massive-scale in vivo lineage recording of cancer metastasis.* **Kamen Simeonov**

Session III: Computational Approaches II

11:40–11:55	*Trustworthy machine learning for understanding genome regulation* **Maxwell Libbrecht**
11:55–12:10	*Generalizable design of human cell type-specific expression via deep learning models of genomic accessibility* **Sebastian Castillo-Hair**
12:10-12:25	**Closing Remarks**
12:25-13:25	**Lunch**

Sponsors


Chan
Zuckerberg
Initiative


BROTMAN BATY
INSTITUTE


GSAT


SBME
UBC
SCHOOL OF
BIOMEDICAL
ENGINEERING


Genome
BritishColumbia

How to Reach the Venue

The conference will be held at the Gordon B Shrum Building, located at 6088 University Boulevard, UBC, Vancouver, BC Canada V6T 1Z3. Here are the recommended transportation options to reach the venue:

By Public Transit

- Bus: The UBC campus is well-served by multiple bus routes. The most convenient buses are the #99 B-Line, #25, #33, #41, and #49, all of which stop at UBC Exchange, a short walk to the Gordon B Shrum Building.
- SkyTrain: While there is no direct SkyTrain service to UBC, you can take the Canada Line to Broadway-City Hall Station and transfer to the #99 B-Line bus which goes directly to UBC.

By Car

- From Downtown Vancouver: Take W 4th, W 10th or SW Marine Drive west to UBC.
- Parking: Available at the Health Sciences Parkade, which is a 5-minute walk from the venue, or at the North Parkade, which is a 10-minute walk.

By Bicycle

• Bike Racks: Available around the AMS Student Nest. UBC is bike-friendly with multiple bike lanes leading directly to the campus.

By Air

• Nearest Airport: Vancouver International Airport (YVR).

• Transport to Venue: From YVR, you can take the Canada Line from the airport to Broadway-City Hall Station and then transfer to the #99 B-Line bus to UBC. Alternatively, a taxi or rideshare from the airport will take about 30 minutes.

Oral Presentation Abstracts

Saturation deletion scanning of the human genome

Jonas Koeppel[1,2], Samantha Sgrizzi[2], Peixi Chen[1,2], Faaiz Quaisar[2], Jay Shendure[1,2,3], Sudarshan Pinglay[1,2]

1 Department of Genome Sciences, University of Washington, Seattle, WA, USA 2 Seattle Hub for Synthetic Biology, Seattle, WA, USA 3 Howard Hughes Medical Institute, Seattle, WA, USA

Understanding the constraints on genome structure beyond protein-coding regions is limited due to a lack of effective genome engineering tools capable of systematically interrogating the extensive non-coding genome. Consequently, we do not know which regions of the genome are truly dispensable. The Class I CRISPR system, CRISPR-Cas3, represents a promising solution to this challenge. CRISPR-Cas3 uses crisprRNAs (crRNAs) to target specific genomic loci and initiate unidirectional deletions of variable lengths ranging from hundreds to hundreds of thousands of base pairs. However, its use for high throughput mammalian genome engineering has been constrained by the difficulty in precisely mapping the variable deletion breakpoints without whole genome sequencing. To address this gap, we developed Shred-seq, a method for highly multiplexed large-scale deletion screens of the human genome. To achieve this, we randomly integrated barcoded synthetic target sites and used CRISPR-Cas3 to initiate deletions. The barcodes allowed rapid mapping of both integration location and deletion boundaries via T7 in vitro transcription, a method that is compatible with single-cell sequencing readouts. In a single experiment, we genotyped 140 Mb of deletions across 1700 unique variants distributed across all chromosomes. We characterized the length distribution of Cas3-mediated deletions across variable genomic contexts and identified characteristic repair junction patterns. We further improved the system by unbiased screening of crRNAs and engineering of guide RNA stem loops. Finally, we map how

genomic features affect the selection pressure acting on thousands of deletions to create an essentiality map of the human genome. Such a map will enable us to identify previously unknown functional sequence elements, better predict the impact of structural variants, and lay the groundwork towards a minimal human genome.

Single cell Strand-seq: a unique tool for genome analysis

Lansdorp, Peter

Terry Fox Laboratory, BC Cancer Agency, Vancouver, BC V5Z 1L3, Canada and Department of Medical Genetics, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada

Sequencing the human genome came with the promise of refined risk assessment for heritable diseases, adverse drug reactions, and other applications of “precision medicine”. Genome wide association studies and the linking of thousands of genetic alterations to specific heritable disorders have delivered on many of these promises. Subsequent “whole genome” sequencing studies have provided further advances. However, over 50% of rare disease patients remain undiagnosed after genomic sequencing and current “whole genome” sequencing studies struggle to delineate structural variations in human genomes. Importantly, most studies also fail to fully delineate or “phase” the parental, haploid genomes in diploid genomes. Global efforts to generate “telomere to telomere (T2T)” genomes are addressing this challenge, but clinical application of these approaches face multiple challenges including costs. Recent studies have shown that long read sequence data, combined with limited short read data obtained using single cell Strand-seq can be used to assign the parental origin to genetic variants. By resolving parental haplotypes as well as structural variations in human genomes, this approach is likely to replace many targeted and “whole genome” sequencing approaches. Resolution of parental haplotypes or “diploid sequencing” represents a novel baseline for precision medicine. Key to this approach is the Strand-seq method. In this presentation applications and technical aspects of the Strand-seq method will be reviewed and discussed

Cuffed CRISPR guide RNA for microRNA activity-dependent genome editing

Adel, Arman; Baatartsogt, Nemekhbayar; Shuto, Yutaro; Nakagawa, Ryoya; Ono, Hiroki; Chopra, Sanchit; Hayakawa, Morisada; Kashiwakura, Yuji; Suleimenova, Alima; Robinson, Meghan; Stepan, Tabea; Mori, Hideto; Kiyota, Brett; Flannigan, Ryan; Hoodless, Pamela; Hiroyuki, Aburatani; Saito, Hirohide; Nureki, Osamu; Ohmori, Tsukasa; Yachie, Nozomu

1. School of Biomedical Engineering, The University of British Columbia, Vancouver, BC, Canada 2. Department of Biochemistry, School of Medicine, Jichi Medical University, Shimotsuke, Tochigi, Japan 3. Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo, Japan 4. Department of Life Science Frontiers, Center for iPS Cell Research and Application, Kyoto University, Kyoto,

Sakyo-Ku, Japan 5. Vancouver Prostate Centre, Vancouver, British Columbia 6. Terry Fox Laboratory, BC Cancer, Vancouver, BC, Canada 7. Department of Urologic Sciences, The University of British Columbia, Vancouver, BC, Canada 8. Department of Urology, Weill Cornell Medicine, New York, NY, USA 9. Department of Medical Genetics, The University of British Columbia, Vancouver, BC, Canada 10. Cell and Developmental Biology Program, The University of British Columbia, Vancouver, BC, Canada 11. Premium Research Institute for Human Metaverse Medicine, Osaka University, Suita, Osaka, Japan 12. Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan 13. Institute for Quantitative Biosciences, The University of Tokyo, Tokyo, Japan 14. Center for Gene Therapy Research, Jichi Medical University, Shimotsuke, Tochigi, Japan 15. Graduate School of Media and Governance, Keio University 16. Japan Science and Technology Agency, Core Research for Evolutional Science and Technology, Kawaguchi-shi, Saitama, Japan 17. The McMillan Multiscale Human Program, the Canadian Institute for Advanced Research, Toronto, Canada

Cells in multicellular eukaryotic systems are diverse, and their molecular profiles can characterize their functions. While CRISPR-Cas9 genome editing has been widely utilized in the broad field of biology to modulate cells and interrogate gene function, there is no versatile and scalable method to edit the genome of a cell in response to endogenously expressed cellular signals. MicroRNAs (miRNAs) are widely conserved short non-coding RNAs involved in development and homeostasis as well as disease progression and disorder in eukaryotes. Here, we report a cuffed CRISPR guide RNA (cgRNA) that confers genome editing efficiently in response to the catalytic activity of a specified miRNA within a cell. By permutation of the commonly used CRISPR guide RNA sequence and concatenation of a miRNA target sequence, cgRNA forms a warped guide RNA structure that is inactive until cleaved by the corresponding miRNA. Structural analyses revealed how cgRNA inactivates the CRISPR genome editing pathways in the absence of the miRNA at multiple steps. The utilization of cgRNA enabled efficient miRNA activity-dependent genome editing in human and mouse cell lines, genome-wide screening of miRNA activity in mouse embryonic stem cells, and mouse tissue-specific genome editing. cgRNA would facilitate the development of cell state-specific genome editing and the mapping and understanding of miRNA activity landscape during regulation of multicellular systems.

Multi-scale deep phenotyping and statistical derivatization of mammalian developmental enhancers

Lalanne, Jean-Benoit; Li, Tony; Kajiwara, Emma; Jain, Shruti; Li, Xiaoyi; Marin, Beth; Regalado, Samuel; Daza, Riza; Shendure, Jay

Department of biochemistry and molecular medicine, Université de Montréal; Department of Genome Sciences, University of Washington

Gene expression control in space and time in metazoans is collectively mediated by non-coding regions called cis-regulatory elements (CREs). Despite its fundamental and practical importance, key features of the transcriptional 'regulatory code', the sequence-to-function relationship for CREs, remain empirically uncharacterized: What length of DNA is sufficient to capture complete regulatory activity? How much synergy exists between transcription factor binding sites? Can the flexibility of the regulatory grammar be statistically assessed? Focusing on select cell-type specific CREs as models, we functionally profiled diverse classes of variant libraries using massively parallel reporter assays to provide a comprehensive, multi-scale, and quantitative view of regulatory sequence-to-function maps. Further, leveraging CRE 'derivatization' and model-driven mutagenesis, we provide a roadmap for the engineering of compact, highly active CREs.

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

Jessa, Selin; Liu, Betty; Kim, Samuel; Ng, Yan Ting; Higashino, Soon il; Marinov, Georgi K.; Ben-David, Eyal; Kundaje, Anshul; Farh, Kyle; Greenleaf, William

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

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, the syntax of binding sites—their composition, orientation, and spacing—contributes to cell type-specific regulation, but we lack a systematic understanding of how this syntax mediates combinatorial activity of TFs.

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, generally corresponding to recognition motifs for TF families. 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 that are highly abundant in the genome yet inhibit accessibility. We show that different classes of motifs have preferential distributions relative to genomic features and inferred nucleosome positions. Using our trained models together with an in silico experimentation framework, we systematically dissected motif cooperativity across spacings and orientations, and discovered 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.

iQcell2.0: A GNN with Interpretable Message passing for single cell transcriptomics

*Tiam Heydari - Umar Ali - Divy Raval - Peter Zandstra**

UBC

Understanding how dynamic gene regulatory networks (GRNs) control cell fate decisions remains a fundamental challenge in biology. Current computational approaches often lack the causal and temporal structure of regulation in scalable manner, limiting their ability to predict cellular responses to perturbations. Here, we introduce iQcell2, a biologically informed graph neural network framework that learns biology-constrained regulatory interactions from single-cell RNA sequencing data and simulates dynamic cell trajectories under genetic perturbations. iQcell2 flexibly accommodates diverse biological priors into its loss function, allowing adaptation across different applications and datasets.

Each gene is modeled as a node in a directed graph, with activator and repressor influences combined via a differentiable BioLogic core that encodes structured

inductive biases derived from systems biology into message passing mechanism. To capture causal delays between regulator activity and transcriptional outcomes, we introduce Delayed Attention, a mechanism that learns optimal pseudo-temporal offsets for regulatory interactions. Training integrates biological priors, pseudotime-derived hierarchy constraints, and entropy-annealed attention regularization.

We demonstrate that BioLogic accurately recovers synthetic regulatory programs across both continuous and discrete expression spaces, establishing it as a universal regulatory network learner. The BioLogic core exhibits robustness under noise and generalizes to predict the outcomes of unseen perturbations, including knockouts and overexpression events. Once trained, iQcell2 enables forward simulation of cell trajectories, identification of combinatorial regulatory targets, and reconstruction of dynamic developmental landscapes in human and mouse in vitro systems. Together, iQcell2 offers a flexible, interpretable, and biologically grounded platform for dynamic GRN inference, perturbation modeling, and hypothesis generation from single-cell data.

Virtual cells in virtual human development: and international effort

Shakiba, Nika

The School of Biomedical Engineering, The University of British Columbia, Vancouver, Canada

Premium Research Institute for Human Metaverse Medicine (PRIMe), The University of Osaka, Osaka, Japan

This talk introduces the international Virtual Human Development (VHD) consortium. It explores the VHD mission of simulating human embryonic development, from zygote to fetus, and its vision for achieving this by creating experiment-to-theory cycles. Virtual twin models are permeating biomedicine, offering in silico windows into our tissues and serving as a tool for precision medicine. However, models to date do not capture the process by which our bodies are built, leaving underlying mysteries with exponential rewards for regenerative medicine. The dynamic nature of the embryo over space and time makes this a complex challenge. To meet this challenge, VHD uniquely bridges empirical and theoretical experts. The birth of VHD is marked by key advancements in both experimental and theoretical technologies. As single-cell omics, synthetic biology, and in vitro models of embryo and organ development have come of age, we can now have a high-dimensional, multi-scale and dynamic view of development. With the advent of artificial intelligence and multi-scale modeling, the computational boom in biology provides a means to unravel these rich empirical datasets. The VHD simulation platform

will not only unlock the mysteries of our origins, but will move us towards the era of rational cellular engineering.

Reconstitution of mammalian life cycle in vitro

Nobuhiko Hamazaki

University of Washington, Departments of Obstetrics & Gynecology, Departments of Genome Sciences, Institute for Stem Cell & Regenerative Medicine, Brotman Baty Institute, Seattle Hub for Synthetic Biology

Mammalian development remains a mysterious process, largely because it takes place within the mother's uterus. Historically, removing parts of this process from the uterus and studying them in vitro has led to groundbreaking discoveries. For instance, in vitro fertilization and the subsequent culture of embryos to the blastocyst stage have allowed us to observe, manipulate, and study post-fertilization development in unprecedented detail. In this context, we introduce two in vitro model systems: directly induced oocyte-like cells (DIOLs), which mimic mouse oocyte development, and the human advanced gastruloid model, which recapitulates human post-gastrulation development. These systems offer robust and scalable platforms for exploring the mechanisms governing mammalian life cycles and their associated failures, such as infertility and developmental diseases.

Capsule-based single cell genome sequencing

Dustin B. Mullaney[1†], Samantha R. Sgrizzi[1†], David Mai[2,3†], Ian Campbell[2†], Yuqi Huang[2,3], Andrius Šinkūnas[4], D. Lucas Kerr[2,3,5], Valentino E. Browning[5], Helen E. Eisenach[6,7], Jeremiah Sims[6,7], Eva K. Nichols[5], Christopher P. Lapointe[2], Yasuhiro Amimura[2], Kelley Harris[5], Rapolas Žilionis[4,8], Sanjay R. Srivatsan[2,3,5,6,9]

1. Molecular and Cellular Biology Program, University of Washington, Seattle, WA, USA 2. Basic Sciences Division, Fred Hutchinson Cancer Center, Seattle, WA, USA 3. Computational Biology Program, Fred Hutchinson Cancer Center, Seattle, WA, USA 4. Atrandi Biosciences, Vilnius, Lithuania 5. Department of Genome Sciences, University of Washington, Seattle, WA, USA 6. Institute for Protein Design, University of Washington, Seattle, WA, USA 7. Department of Biochemistry, University of Washington, Seattle, WA, USA 8. Institute of Biotechnology, Life Sciences Center, Vilnius University, Vil

Single-cell genomics methods have unveiled the heterogeneity present in seemingly homogenous populations of cells, however, these techniques require meticulous optimization. How exactly does one handle and manipulate the biological contents from a single cell? Here, we introduce and characterize a novel semi-permeable capsule (SPC), capable of isolating single cells and their contents while facilitating biomolecular exchange based on size-selectivity. These capsules maintain stability under diverse

physical and chemical conditions and allow selective diffusion of biomolecules, effectively retaining larger biomolecules including genomic DNA, and cellular complexes, while permitting the exchange of smaller molecules, including primers and enzymes. We demonstrate the utility of SPCs for single cell assays by performing the simultaneous culture of over 500,000 cellular colonies, demonstrating efficient and unbiased nucleic acid amplification, and performing combinatorial indexing-based single-cell whole genome sequencing (sc-WGS). Notably, SPC-based sc-WGS facilitates uniform genome coverage and minimal cross-contamination allowing for the detection of genomic variants with high sensitivity and specificity. Leveraging these properties, we conducted a proof-of-concept lineage tracing experiment using cells harboring the hypermutator polymerase ε allele (POLE P286R). Sequencing of 1000 single cell genomes at low depth facilitated the capture of lineage marks deposited throughout the genome during each cell division and the subsequent reconstruction of cellular genealogies. Capsule-based sc-WGS expands the single-cell genomics toolkit and will facilitate the investigation of somatic variants, resolved to single cells at scale.

Diversified, miniaturized and ancestral parts for mammalian genome engineering and molecular recording

McDiarmid, Troy A.[1,2,†,]; Taylor, Megan L.[1,2,†]; Chen, Wei[1,2]; Chardon, Florence M.[1,2]; Choi, Junhong[1,2,3]; Liao, Hanna[1,2]; Li, Xiaoyi[1,2]; Kim, Haedong[1,2]; Lalanne, Jean-Benoît[1]; Li, Tony[1]; Nathans, Jenny F.[1,2]; Martin, Beth K.[1,2]; Knuth, Jordan[2]; Coradini, Alessandro L.V.[2]; Gray, Jesse M.[2]; Pinglay, Sudarshan[1,2,4]; and Shendure, Jay[1,2,4,5,6]*

1. Department of Genome Sciences, University of Washington, Seattle, WA, USA 2. Seattle Hub for Synthetic Biology, Seattle, WA, USA 3. Developmental Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA 4. Brotman Baty Institute for Precision Medicine, Seattle, WA, USA 5. Howard Hughes Medical Institute, Seattle, WA, USA 6. Allen Discovery Center for Cell Lineage Tracing, Seattle, WA, USA

† These authors contributed equally to this work

** Correspondence to T.A.M. (troym13@uw.edu) or J.S. (shendure@uw.edu)*

As the synthetic biology and genome engineering fields mature and converge, there is a clear need for a “parts list” of components that are diversified with respect to both functional activity (to facilitate design) and primary sequence (to facilitate assembly). Here we designed libraries composed of extant, ancestral, mutagenized or miniaturized variants of Pol III promoters or guide RNA (gRNA) scaffolds and quantified their ability to mediate precise edits to the mammalian genome via multiplex prime editing. We identified thousands of parts that reproducibly drive a range of editing activities in

human and mouse stem cells and cancer cell lines, including hundreds exhibiting similar or greater activity than the sequences used in conventional genome engineering constructs. We further conducted saturation mutagenesis screens of canonical Pol III promoters (U6p, 7SKp, H1p) and the prime editing guide RNA (pegRNA) scaffold, which identified tolerated variants that can be superimposed on baseline parts to further enhance sequence diversity. While characterizing thousands of orthologous promoters from hundreds of extant or ancestral genomes, we incidentally mapped the functional landscape of mammalian Pol III promoter evolution. Finally, to showcase the usefulness of these parts, we designed a “ten key” molecular recording array that lacks repetitive subsequences in order to facilitate its one-step assembly in yeast. Upon delivering this 15.8 kb tandem array of promoters and guides to mammalian cells, individual pegRNAs exhibited balanced activities as predicted by the activity of component parts, despite their relocation to a single locus. Looking forward, we anticipate that the diversified parts and variant effect maps reported here can be leveraged for the design, assembly and deployment of synthetic loci encoding arrays of gRNAs exhibiting predictable, differentiated levels of activity, which will be useful for multiplex perturbation, advanced biological recorders and complex genetic circuits.

Understanding upstream dominance in synthetic gene circuits in human cells

Jones, Ross D.; Yun, Jiyoung; Murtaza, Ali; Chan, Matthew; Johnstone, Christopher P.; Love, Kasey S.; Galloway, Kate E.; Zandstra, Peter W.

RDJ 1, 2; JY 1, 2; AM 1, 2, 3; MC 4; CPJ 5; KSL 6; KEG 5; PWZ 1, 2, 7

1: School of Biomedical Engineering, University of British Columbia, Vancouver, BC, Canada 2: Michael Smith Labs, University of British Columbia, Vancouver, BC, Canada 3: Synthetic Biology Institute, Rice University, Houston, TX, USA 4: Microbiology & Oceanography, University of British Columbia, Vancouver, BC, Canada 5: Department of Chemical Engineering, MIT, Cambridge, MA, USA 6: Department of Biological Engineering, MIT, Cambridge, MA, USA 7: Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada

Context-dependence in gene expression muddies our ability to predictably design robust synthetic genetic circuits in cells. A significant and poorly-understood

context-dependency is how the expression of one gene affects that of neighboring genes in the genome, especially in mammalian cells. In collaboration with the Galloway Lab at MIT, we are studying this problem through the lens of genomic “syntax”, the relative position and orientation of synthetic genes, and its effects on gene expression. For example, we have found that when integrating two genes into human cells in a tandem orientation: {Promoter1:Protein1_PolyA1 // Promoter2:Protein2_PolyA2}, expression of a given gene will be much stronger when in the upstream position than the downstream position. This pattern of “upstream dominance” holds for different integration modalities (PiggyBac transposon, lentivirus, and CRISPR knockin), different promoters (both constitutive and induced), and different human cell types (e.g. HEK-293 and pluripotent stem cells [PSCs]). To further understand the mechanistic basis for upstream dominance, we engineered human PSCs with an array of two-gene constructs where we tested how combinations of different promoters, poly-A signals, gene orientations, and chromatin insulators affect gene expression. Our results suggest that several factors including DNA supercoiling, DNA looping, and silencing by the human silencing hub (HUSH) complex likely contribute to upstream dominance. Further improvements in our understanding of upstream dominance in human cells and the development of methods to predict and counteract its effects will enable much more precise control of complex gene circuits integrated into human cells. Our focus on engineering PSCs, which have vast potential to differentiate into diverse cells and tissues, will have significant implications for advancing regenerative medicine

Technologies for "look-back-in-time" biology

Yachie, Nozomu

UBC SBME

Biological studies often require the destruction of samples to observe molecular and cellular details, hindering the understanding of dynamic processes. To overcome this issue and observe the dynamics of living systems in high resolution, my group is developing two approaches: (1) a video camera system and (2) a “time-reversing”

system to observe dynamic molecular and cellular events. In the first system, which we call “DNA event recording,” molecular and cellular events within a multicellular organism are progressively stored in synthetic “DNA tapes” embedded in the cells’ genomes. We are currently working toward obtaining the first high-resolution whole-body developmental map of the mouse. In the second system, called “retrospective clone isolation,” cells can be isolated and analyzed after identifying their future states. Cells in a population are first tagged with unique, short DNA barcodes and propagated. A subpopulation is then subjected to a specific assay. Identifying the barcode of a clone of interest after an assay, the same clone is isolated in a barcode-dependent manner from the initial population before it expresses the target phenotype. I will share our current progress in these ideas.

Massive-scale in vivo lineage recording of cancer metastasis

Simeonov, Kamen; Clark, Megan; Lalanne, Jean-Benoît; Shendure, Jay; Lengner, Christopher

Fred Hutchinson Cancer Center; University of Pennsylvania; University of Montreal; University of Washington; University of Pennsylvania

The underpinnings of cancer metastasis remain poorly understood, in large part due to a lack of tools for probing their emergence at high resolution. We previously created macsGESTALT, an inducible CRISPR-based lineage recorder with single-cell capture of both transcriptional and phylogenetic information. Applying macsGESTALT to a model of metastatic cancer, we uncovered in vivo emergence of highly metastatic cell states driven by distinct hybrid-EMT (epithelial-to-mesenchymal transition) and S100 activation programs. To more systematically profile drivers of metastasis and seeding topologies, we have since developed Ouroboros, an improved lineage recorder with optimized editing rates, higher information capacity, and near-deterministic barcode recovery. By developing a protocol for high-throughput tissue harvesting and dissociation combined with sample multiplexing, we applied Ouroboros on a massive scale: harvesting ~633,000 single cancer cells with paired lineage, transcriptional, and sample hash information across 543 individually-dissected primary tumor and metastatic lesions derived from 52 mice. Via phylogenetic analysis of ~32 million recovered CRISPR target sites, we have reconstructed ~170,000 in vivo labeled cancer subclones with diverse

transcriptional states and metastatic behaviors. We believe these data will demonstrate the potential of genomic recording to generate a deeper understanding of cancer dissemination and heterogeneity, cell state heritability and plasticity, and clonal dynamics.

Trustworthy machine learning for understanding genome regulation

Maxwell W Libbrecht

Simon Fraser University, School of Computing Science

Genome-wide association studies (GWAS) have identified tens of thousands of genetic variants associated with human diseases and traits, but the vast majority of these associations are not backed by a hypothesized mechanism. Understanding such disease-associated variants is hampered by the incompleteness of annotations functional elements in the genome. To improve our annotation of genomic elements, large-scale projects like ENCODE, CEEHRC and IHEC have recently engaged in epigenome mapping. These projects are enabled by high-throughput sequencing techniques for genome-scale measurement of biochemical activity of chromatin in cellular samples. These datasets quantify various facets of gene regulation such as genome-wide measurements of transcription factor binding or histone modifications using ChIP-seq, measurements of open chromatin using DNase-seq or ATAC-seq, RNA transcription using RNA-seq, and others. My group aims to improve our understanding and annotating of the genome using machine learning methods. I will present our recent work aiming to comprehensively annotate genomic functional elements and statistically validate genomic predictions.

Generalizable design of human cell type-specific expression via deep learning models of genomic accessibility

Castillo-Hair, Sebastian; Yin, Christopher; VandenBosch, Leah; Cherry, Timothy; Meuleman, Wouter; Seelig, Georg

Department of Electrical & Computer Engineering, University of Washington, Seattle, WA; Department of Electrical & Computer Engineering, University of Washington, Seattle, WA; Center for Developmental Biology and Regenerative Medicine, Seattle Children's Research Institute, Seattle, WA; Center for Developmental

Biology and Regenerative Medicine, Seattle Children's Research Institute, Seattle, WA; Altius Institute for Biomedical Sciences, Seattle, WA; Department of Electrical & Computer Engineering, University of Washington, Seattle, WA

Enhancer sequences regulate eukaryotic gene expression with spatio-temporal and cell type specificity. Engineering enhancers that restrict expression to specific cell types and tissues could improve our ability to design more specific gene therapies with reduced side effects. However, enhancer identification and design remain limited by our understanding of their regulatory grammar. Here, we develop and experimentally validate deep learning-designed cell type-specific enhancers. We trained deep learning models on genomic accessibility data, based on the rationale that active enhancers reside in genomic regions with an open, accessible chromatin state. These data are available for hundreds of human cell types and tissues—including single-cell resolution datasets— and thus offers a vast resource for enhancer design. Using data of >3 million DNase-hypersensitive sites across 733 cell types and tissues from the ENCODE and Roadmap Epigenetics projects, we trained deep learning models to predict cell type-specific accessibility from sequence. We then applied gradient descent-based sequence optimization and generative neural networks to design de novo sequences predicted to maximize cell type-specific accessibility. Notably, synthetic sequences generated via deep learning exhibited a more complex regulatory grammar with a higher density of putative regulatory elements than genomic sequences. To validate our designs, we tested a library of 9,000 synthetic enhancers in a panel of 10 human cell lines—including HepG2 (liver), K562 (lymphoid), SJCRH30 (muscle), WERI-Rb1 (retina), and MCF7 (breast)—as well as in vivo in mouse retinas. In most cases, synthetic sequences displayed significantly higher enhancer activity and specificity in their target cells compared to putative genomic enhancers. Our results demonstrate that functional enhancers can be designed directly from models of genomic accessibility, highlighting the potential of deep learning-driven sequence design for synthetic biology.

Poster Presentation Abstracts

---

1. TARDIS: Enabling Large-Scale Genetic Library Screening in an Animal Model

Zachary C. Stevenson[1], Megan Moerdyk-Schauwecker[2], Patrick C. Phillips[2]

1: University of Washington, 2: University of Oregon

Synthetic biology has traditionally focused on single-cell chassis—bacteria, yeast, and mammalian cell cultures—due to the scalability of transgenesis techniques that enable high-throughput design-build-test-learn cycles. While animal models could expand the synthetic biology toolkit by incorporating complex features like multicellularity and development, their utility has been limited by the inability to test large genetic libraries. In the model nematode Caenorhabditis elegans, transgenesis is typically performed by microinjection, which results in very few transgenic progeny. To address this bottleneck, we developed Transgenic Arrays Resulting in Diversity of Integrated Sequences (TARDIS), a system that enhances transgenesis in C. elegans by approximately 1,000-fold. This breakthrough enables, for the first time, large-scale barcoding and genetic library screening in an animal model. TARDIS leverages C. elegans' unique genetic tool: customizable, heritable extrachromosomal DNA elements known as arrays. These arrays serve as in vivo genetic libraries, propagating across generations to distribute genetic diversity within a population. Effectively making the worm genetically 'bigger on the inside,' the TARDIS workflow splits transgenesis into two phases: (1) generating an in vivo TARDIS library array from an injected in vitro library and (2) integrating individual library elements into a genomically engineered safe-harbor "landing pad" via controlled homology-directed repair (HDR). By delaying HDR until Cas9 expression is induced via heat shock, we achieve simultaneous integration of thousands of unique sequences into the landing pad across a synchronized population. We validated TARDIS with two applications: a high-diversity barcode library and a tissue-specific promoter library. Custom landing pads were designed to accept these

elements at multiple genomic loci, ensuring stable transgene expression. By coupling scalable array generation with targeted integration, TARDIS unlocks the potential for high-throughput genetic screening in animal systems, bridging the gap between single-cell and multicellular synthetic biology.

2. Culture Wars: Empirically Determining the Best Approach for Plasmid Library Amplification

Mateyko, Nicholas; de Boer, Carl

Genome Science and Technology Graduate Program, University of British Columbia; School of Biomedical Engineering, University of British Columbia

DNA libraries are essential for many biological experiments: MPRAs, CRISPR screens, deep mutational scans, and directed evolution, among many others. They are typically constructed in plasmids and amplified in E. coli, but uneven amplification due to differences in clone growth rates can skew the representation of library elements, ultimately leading to suboptimal experimental outcomes. It is thought that the culture method used to amplify plasmid libraries can impact the uniformity of amplification but, to our knowledge, no one has directly compared library culture methods. To determine the impact of culture method on plasmid library amplification, we amplified two dissimilar plasmid libraries with five culture methods: liquid, semisolid agarose, cell spreader-spread plates with high or low colony density, and bead-spread plates with high colony density. We then sequenced the resulting plasmid libraries to determine library uniformity after amplification. We found that no method amplified libraries more uniformly than liquid culture, the simplest method. With high (~100X) transformation coverage, culture method had a minimal impact on library uniformity and the original library composition was always maintained. These results will allow the many users of plasmid libraries to make an informed choice of culture method when creating new plasmid libraries or amplifying existing ones.

3. RAPID-DASH: Single-Day Assembly of Guide RNA Arrays for Multiplexed CRISPR-Cas9 Applications

Salaudeen, Asfar Lathif; Mateyko, Nicholas; de Boer, Carl G.

Genome Science and Technology Graduate Program, University of British Columbia, Vancouver, BC, Canada; Genome Science and Technology Graduate Program, University of British Columbia, Vancouver, BC, Canada; School of Biomedical Engineering, University of British Columbia, Vancouver, BC, Canada

Guide RNA (gRNA) arrays can enable targeting multiple genomic loci simultaneously using CRISPR-Cas9. In this study, we present a streamlined and efficient method to rapidly construct gRNA arrays with up to 10 gRNA units in a single day. We demonstrate that gRNA arrays maintain robust functional activity across all positions, and can incorporate libraries of gRNAs, combining scalability and multiplexing. Our approach will streamline combinatorial perturbation research by enabling the economical and rapid construction, testing, and iteration of gRNA arrays.

4. A Combined Synthetic Biology and Mathematical Modeling Approach for the Large-Scale Supply of Human Pluripotent Stem Cells

Shahdoost, Ali; Leibovich, Nava; Barbaric, Ivana; Yachie, Nozomu; Goyal, Sidhartha; Shakiba, Nika

The University of British Columbia, University of Toronto, University of Sheffield, The University of British Columbia, University of Toronto, The University of British Columbia

The unlimited potential of human pluripotent stem cells (hPSCs) for self-renewal and differentiation into all body cell types positions them as a crucial resource for regenerative medicine cell therapies. To meet the escalating demand for these therapies, bioprocessing pipelines must scale up to produce trillions of hPSCs. However, prolonged cultivation of hPSCs often leads to the emergence of "variants" with genetic and epigenetic abnormalities. These variants not only display a cancer-like growth pattern that dominates cultures but also reduce differentiation efficiency and are unsafe for clinical applications. Current methods for identifying these variants are inadequate; standard assays like karyotyping only detect a narrow range of known genetic changes, leaving many harmful mutations undetected. Additionally, once detected, there is no effective method to eliminate these variants from cultures, often necessitating the disposal of affected hPSC batches.

In response to these challenges, this project introduces a novel bioengineering strategy that utilizes a CRISPR-based DNA barcoding technology combined with computational modeling. This innovative approach identifies variant hPSCs by analyzing growth rates, which act as a phenotypic fingerprint, thus bypassing traditional genotypic assays. The central goal of this initiative is to develop and implement a technology capable not only of tracking but also eliminating the problematic variants from cell cultures. This technology promises to be the first standardized method for detecting and removing problematic hPSC variants, potentially revolutionizing hPSC bioprocesses across the biotechnology industry. This advancement is especially significant as cell therapies emerge as a new class of treatments in Canada and globally. By enabling a large-scale, reliable supply of hPSCs, this approach could replace all current methods, which have significant limitations, and it is poised to be implemented in bioreactors and suspension cell cultures, enhancing safety and efficiency in cell therapy manufacturing.

5. Deciphering enhancer regulation with synthetic DNA: from molecular mechanism to disease modelling

Emmanuel Cazottes; Carl de Boer

School of Biomedical Engineering, University of British Columbia, Vancouver, BC, Canada

The vast majority (91%) of trait-associated genetic variants occur in non-coding regions and are enriched in enhancers. However, our understanding of how these enhancers communicate with their target genes remains severely limited, hampering our ability to interpret the clinical significance of enhancer variants. This challenge persists largely due to technological constraints - current methods cannot efficiently analyze the estimated 650,000 enhancers in the human genome with sufficient precision.

To address this critical knowledge gap, we are developing an innovative platform that combines high-throughput screening of over 100,000 enhancer-gene pairs with advanced AI modeling. This integrated approach will systematically identify the sequence features that govern enhancer activity and their communication with target genes.

Our research will generate a comprehensive enhancer-gene interaction atlas, revealing the fundamental sequence elements that control gene regulation. These insights will improve our ability to interpret disease-causing variants in enhancer regions and their downstream effects on gene expression.

6. Mapping Beta Cell Heterogeneity: Cellular Insights from hPSC-Derived Differentiation

Movasat, Hourieh; Shakiba, Nika*,**; Tahamtani, Yaser***; Lynn, Francis C.*,***,****

* School of Biomedical Engineering, University of British Columbia, Vancouver, BC ** Premium Research Institute for Human Metaverse Medicine (PRIMe), Osaka University, Osaka, Japan *** BC Children's Hospital Research institute, Vancouver, BC **** Department of Surgery, University of British Columbia, Vancouver, BC*

Beta cell death is a hallmark of type 1 diabetes (T1D), a condition in which the immune system attacks insulin-producing beta cells, leading to lifelong dependence on exogenous insulin. While islet transplantation offers a promising therapeutic approach, the scarcity of suitable donors remains a major barrier to widespread implementation. Human pluripotent stem cells (hPSCs) present an alternative and virtually unlimited source for generating beta cells. However, significant cell loss during differentiation poses a challenge, reducing yield and cost-effectiveness. We hypothesize that a subset of hPSCs and their progeny possess an intrinsic advantage in beta cell differentiation potential. To test this hypothesis, we employ DNA barcoding technology to track individual cell fates during two critical stages of differentiation: hPSCs to pancreatic progenitors (PPs) and PPs to functional beta cells. By analyzing differentiation outcomes at a clonal level, we aim to identify high-potential clones that exhibit superior efficiency, viability, and functional characteristics. Our objectives are threefold: (1) to pinpoint the subpopulations of hPSCs and PPs that exhibit the highest likelihood of successfully differentiating into mature beta cells, (2) to characterize the molecular signatures that define these high-performing clones, and (3) to elucidate the role of cellular heterogeneity in determining differentiation efficiency, yield, and purity. Understanding the intrinsic factors governing differentiation potential will enable the development of strategies to enhance beta cell production, improving the feasibility of

hPSC-derived beta cell therapies. By addressing the challenges of heterogeneity and optimizing differentiation protocols, our research aims to bridge a critical gap in regenerative medicine for diabetes. This work can potentially revolutionize beta cell replacement therapy, offering a scalable, donor-independent solution for T1D patients and advancing the broader field of stem cell-based disease modeling and treatment.

7. AlphaFold-Multimer accurately captures interactions and dynamics of intrinsically disordered protein regions

Omidi, Alireza; Harder Møller, Mads; Malhis, Nawar; Bui, Jennifer M.; Gsponer, Jörg

Michael Smith Laboratories, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

Interactions mediated by intrinsically disordered protein regions (IDRs) pose formidable challenges in structural characterization. IDRs are highly versatile, capable of adopting diverse structures and engagement modes. Motivated by recent strides in protein structure prediction, we embarked on exploring the extent to which AlphaFold-Multimer can faithfully reproduce the intricacies of interactions involving IDRs. To this end, we gathered multiple datasets covering the versatile spectrum of IDR binding modes and used them to probe AlphaFold-Multimer's prediction of IDR interactions and their dynamics. Our analyses revealed that AlphaFold-Multimer is not only capable of predicting various types of bound IDR structures with high success rate, but that distinguishing true interactions from decoys, and unreliable predictions from accurate ones is achievable by appropriate use of AlphaFold-Multimer's intrinsic scores. We found that the quality of predictions drops for more heterogeneous, fuzzy interaction types, most likely due to lower interface hydrophobicity and higher coil content. Notably though, certain AlphaFold-Multimer scores, such as the Predicted Aligned Error and residue-ipTM, are highly correlated with structural heterogeneity of the bound IDR, enabling clear distinctions between predictions of fuzzy and more homogeneous binding modes. Finally, our benchmarking revealed that predictions of IDR interactions can also be successful when using full-length proteins, but not as accurate as with

cognate IDRs. To facilitate identification of the cognate IDR of a given partner, we established “minD,” which pinpoints potential interaction sites in a full-length protein. Our study demonstrates that AlphaFold-Multimer can correctly identify interacting IDRs and predict their mode of engagement with a given partner.

8. A computational method to facilitate single-cell analysis at scale

Kiyota, Brett; Yachie, Nozomu

School of Biomedical Engineering, University of British Columbia, Vancouver, BC, Canada; Premium Research Institute for Human Metaverse Medicine, Osaka University, Suita, Osaka, Japan; Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan

The continued advancement of single-cell technologies has significantly accelerated the rate at which single-cell information is being collected. While there is promise surrounding the insights that can be gained from characterizing biological systems at increased resolution and sensitivity, it also marks a shift towards large-scale analysis which challenges many of the current practices of single-cell analysis. To address such imposing algorithmic demands, we are developing a method to ease the computational burden by creating information-rich subsamples that capture and integrate information across many single-cell RNA sequencing datasets. This method, referred to as MILK (Multi-resolution Integration of Large-scale single-cell Kernel information), is a non-parametric, distance-based approach that recursively stratifies similar cells into groups using a lenient percentile-based threshold (e.g., 0.1st percentile). Provided gene expression or latent information as input, it hierarchically organizes the population of cells at high-resolution in linear-time. This information can be useful for various downstream applications, including the creation of “meta”-cells, clustering, and exploring potential single-cell trajectories along hierarchical relationships. We are currently applying this method to the Chan-Zuckerberg CELLxGENE census with the goal of investigating whether foundational models can reliably represent information effectively spanning the complete spectrum of observed single-cell RNA sequencing information. In particular, we are applying MILK to latent representations of over 40 million transcriptomic profiles derived from scVI, Geneformer and scGPT models to

facilitate a comprehensive benchmarking analysis, potentially leading to a more holistic understanding of the value these models provide to the field of biology. By focusing efforts on exploring the massive amounts of data we have, this work hopes to conduct large-scale inferences to identify emergent biological properties using economic computing resources.

9. Spatiotemporal microRNA activity profiling during mouse embryogenesis

Chopra, Sanchit[1]; Yachie, Nozomu[1-6]

1. School of Biomedical Engineering, The University of British Columbia, Vancouver, BC, Canada 2. Premium Research Institute for Human Metaverse Medicine, Osaka University, Suita, Osaka, Japan 3. Center for Gene Therapy Research, Jichi Medical University, Shimotsuke, Tochigi, Japan 4. Graduate School of Media and Governance, Keio University 5. Japan Science and Technology Agency, Core Research for Evolutional Science and Technology, Kawaguchi-shi, Saitama, Japan 6. The McMillan Multiscale Human Program, the Canadian Institute for Advanced Research, Toronto, Canada

Mammalian development is a highly intricate and tightly regulated process governed by molecular programs encoded within the genome. It originates from a single fertilized egg that undergoes successive divisions and differentiation events to give rise to diverse, specialized cell types. As development progresses, cells selectively activate gene expression programs in response to intrinsic cues and extrinsic signaling, enabling context-specific functions. Elucidating the regulatory mechanisms underlying these processes is essential, as perturbations can result in congenital abnormalities and developmental disorders. Over the past three decades, researchers have identified a distinct class of small, non-protein-coding genes known as microRNAs, which play critical roles in regulating gene expression during development. Around 1,900 microRNAs have been identified in humans and 1,200 in mice. Each miRNA has the capacity to regulate numerous target transcripts, often through repression or fine-tuning of mRNA stability and translation. Many miRNAs are indispensable for normal development, and mutations in these genes have been linked to a range of developmental disorders Despite their significance, current technologies do not allow for a comprehensive and spatially resolved characterization of miRNA activity across developing tissues. To address this, we have developed a genome engineering platform

that enables simultaneous readout of microRNA activity across tissues. In this project, we will construct the first spatial atlas of miRNA activity during mouse development. The first tier involves engineering reporter mouse lines and applying whole-body 3D tissue clearing for volumetric imaging of miRNA activity. The second tier integrates spatial transcriptomics to spatially co-map miRNA activity with transcriptome-wide gene expression in tissue sections. Together, these complementary approaches will generate a powerful resource for decoding the spatial logic of developmental gene regulation and provide insights into how microRNA dysregulation contributes to disease, laying the foundation for future diagnostic and therapeutic strategies.

10. Variant effects depend on polygenic background: experimental, clinical, and evolutionary implications

Chapel, Madison; Dennis, Jessica; de Boer, Carl

MC - Bioinformatics Program, UBC JD - Bioinformatics Program, UBC. Department of Medical Genetics, UBC. BC Children's Hospital Research Institute. CD - School of Biomedical Engineering, UBC.

Both rare and common genetic variants contribute to human disease, and emerging evidence suggests that they combine additively to influence disease liability. However, due to the non-linear relationship between disease liability and disease prevalence, risk variants have more severe phenotypic consequences in high-risk polygenic backgrounds. Consequently, selecting individuals with appropriate polygenic backgrounds may improve variant characterization experiments, as disease-relevant phenotypes may be masked or revealed in different genomic contexts. Simultaneously, selection has limited power to eliminate risk variants because they have minimal impact in low polygenic risk backgrounds, yet is more powerful with higher polygenicity because more individuals achieve genetic and phenotypic extremes. This dependence on polygenic background means that selection acting on an allele should be modeled as a distribution that differs across populations, time, environments, and individuals rather than as a single value . Overall, considering polygenic backgrounds will be critical when investigating the origins of complex traits .

11. Variant Position Modulates Functional Impact in Massively Parallel Reporter Assays

Aninta, Sambina Islam; Tewhey, Ryan; G de Boer, Carl

School of Biomedical Engineering, University of British Columbia, Vancouver, BC, Canada; The Jackson Laboratory, Bar Harbor, ME, USA; School of Biomedical Engineering, University of British Columbia, Vancouver, BC, Canada

Genome-wide association assays (GWAS) have identified hundreds of thousands of disease-associated loci; however, the causal variants for the vast majority of these trait-associated loci are yet to be identified and functionally validated. To establish a mechanistic link between disease-causing variants in cis-regulatory elements (CREs) and their effects on gene regulatory pathways, massively parallel reporter assays (MPRAs) are performed (Tewhey et al., 2016). The measured effects of alternate and reference alleles in these experiments are used to prioritize causal loci (Mouri et al., 2022).

In a typical MPRA reporter, the variant of interest is positioned at the center before being introduced into the target cell, either via a lentiviral vector or episomal expression (Tewhey et al., 2016). However, it remains unclear whether the center is the optimal position for the variant within the reporter construct and how variant placement influences the prioritization of causal variants. In this study, we explored this question both experimentally and through predictive modelling, analyzing how the variant's effect changes depending on its position within the reporter.

Our results, from both experimental and predictive analyses, reveal that the effect of a single nucleotide mutation (SNP) on expression varies based on its position within the reporter. Specifically, when the variants are placed further apart, the correlation between their SNP effects decreases. Additionally, variants positioned closer to the promoter in the reporter vector exhibit stronger effects.

Our findings highlight the importance of considering variant positioning in MPRA experiments, emphasizing the need for further studies to determine the optimal approach for testing variant effects.

12. GAME: Genomic API for Model Evaluation

Luthra, Ishika; Priyadarshi, Satyam; Dooley, Damion; Guo, Rui; Sheng, Yilun; Tu, Xinming; Lal, Avantika; Klie, Adam; Penzar, Dima; Vorontsov, Ilya; Armand, Ethan; Drusinsky, Shiron; Floren, Alexander; Aerts, Stein; White, Michael A.; Pinello, Luca; Alasoo, Kaur; Zeitlinger, Julia; Zhou, Jian; Seelig, Georg; Tewhey, Ryan; Gosai, Sager; Koo, Peter; Agarwal, Vikram; Pollard, Katherine S.; Libbrecht, Maxwell; Kulakovskiy, Ivan; Carter, Hannah; Mostafavi, Sara; Hsiao, Will; de Boer, Carl G.

School of Biomedical Engineering, University of British Columbia, Vancouver, BC, Canada; Centre for Infectious Disease Genomics and One Health, Simon Fraser University, Burnaby, BC, Canada; Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA; Biology Research | AI Development, gRED Computational Sciences, Genentech, South San Francisco, CA, USA; Department of Medicine, Division of Genomics and Precision Medicine, University of California San Diego, La Jolla, CA, USA; Vavilov Institute of General Genetics, Russian Academy of Sciences, 119991 Moscow, Russia; Integrative Biology Laboratory, Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA; School of Computing Science, Simon Fraser University, Burnaby, British Columbia V51 1S6, Canada; University of California, San Francisco, San Francisco, CA 94143, USA; mRNA Center of Excellence, Sanofi, Waltham, MA 02451, USA; Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA; VIB Center for AI & Computational Biology, VIB-KU Leuven Center for Brain and Disease Research & KU Leuven Department of Human Genetics, Leuven, Belgium; Department of Genetics, Washington University in St. Louis, St. Louis, MO, 63110, USA; Molecular Pathology Unit, Massachusetts General Research Institute, Charlestown, MA, USA; Institute of Computer Science, University of Tartu, Tartu, Estonia; Stowers Institute for Medical Research, Kansas City, MO 64110, USA; Lyda Hill Department of Bioinformatics, University of Texas Southwestern Medical Center, Dallas, TX, USA; The Jackson Laboratory, Bar Harbor, ME, USA; Broad Institute of MIT and Harvard, Cambridge, MA, USA; Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada

The rapid expansion of genomics datasets and the application of machine learning has produced genomics models with ever-expanding capabilities. However, benchmarking these models on practical applications has been challenging because individual projects evaluate their models in ad hoc ways, and there is substantial heterogeneity of both model architectures and benchmarking tasks. We aim to define a system for large-scale, community-led standardized model benchmarking on user-defined evaluation tasks. We borrow concepts from the Application Programming Interface (API) paradigm to allow for seamless communication between pre-trained models and benchmarking tasks, ensuring consistent evaluation protocols. Because all models and benchmarks are inherently compatible in this framework, continual addition of new models and new benchmarks is easy. Containerization will enhance reproducibility and facilitate the deployment of models and benchmarks across computing platforms. We

provide examples of benchmarks and models implementing this framework, and anticipate that the community will contribute their own, leading to an ever-expanding and evolving set of models and evaluation tasks. This resource will accelerate genomics research by illuminating the best models for a given task, motivating novel functional genomic benchmarks, and providing a more nuanced understanding of model abilities.

13. High-content cell lineage tracing of mouse embryogenesis

Nanami Masuyama[1,2,3]; Yusuke Kijima[1]; Hideto Mori[4]; Hirofumi Nishizono[5]; Mikiko Negishi[4]; Madina Kagieva[1]; Shoma Matsumoto[6]; Eiichi Okamura[6]; Arman Adel[1]; Rina Sakata[1]; Hiromi Toyoshima[7]; Kentaro Kakoi[7]; Soh Ishiguro[1]; Yoko Tanimoto[8]; Osamu Masui[9]; Motoaki Seki[7]; Keiji Nishida[10]; Haruhiko Koseki[9]; Hiroyuki Aburatani[7]; Masatsugu Ema[6]; Seiya Mizuno[8]; Nozomu Yachie[1,3,4,7]

1 School of Biomedical Engineering, Faculty of Applied Science and Faculty of Medicine, The University of British Columbia. 2 Institute for Advanced Biosciences, Keio University. 3 Graduate School of Media and Governance, Keio University. 4 Premium Research Institute for Human Metaverse Medicine (WPI-PRIMe), Osaka University. 5 Medical Research Institute, Kanazawa Medical University. 6 Research Center for Animal Life Science, Shiga University of Medical Science. 7 Synthetic Biology Division, Research Center for Advanced Science and Technology, The University of Tokyo. 8 Laboratory Animal Resource Center, University of Tsukuba. 9 Laboratory for Developmental Genetics, RIKEN Center for Integrative Medical Sciences. 10 Graduate School of Science, Technology and Innovation, Kobe University.

Mammalian development begins with a single fertilized egg, which proliferates to form complex body structures. Several methods have been developed to trace cell lineages in metazoans, utilizing chromosome-embedded DNA barcodes that are mutated by CRISPR-Cas9 genome editing and inherited across generations. These mutation patterns enable the reconstruction of developmental lineages, similar to phylogenetic analysis. However, previous methods have limited resolution due to short DNA barcodes, a restricted mutation pattern (mainly deletions), and the toxic effects of Cas9. Here, we present Barclock, a high-content lineage tracing method developed over the past few years. Barclock leverages CRISPR base editing for massively parallel recording of cell lineage information across hundreds of gRNA target sequences, paired with scRNA-seq for reading the recorded data. We envision obtaining the high-content map of mammalian developmental cell lineage for the first time.

14. Interrogating the Role of Slc20a2 in Murine Tibia Development using Visium HD Spatial Transcriptomics

Minzlaff, Jade; Walczak, Philip A.; Scatena, Marta; Giachelli, Cecilia M.

Department of Bioengineering; School of Dentistry, Department of Oral Health Sciences, Department of Dentistry; Department of Bioengineering; Department of Bioengineering

Phosphate is critical for mineralization of bones and teeth, as well as composing cell membranes and nucleic acids, but mechanisms of phosphate transport and signaling are still largely unknown. Phosphate homeostasis in the body is regulated in part by the mammalian type-III sodium-dependent phosphate transporters PiT-1 and PiT-2, encoded by Slc20a1 and Slc20a2, respectively, the only known phosphate transporters expressed in bone cells. Previous work from our lab has shown that unlike Slc20a1, Slc20a2 knockout (KO) in mice disrupts bone mineralization and reduces osteoblast number and function, leading to impaired bone length, density, and volume, suggesting that Slc20a2 could alter osteogenesis-related signaling pathways. Herein, we investigate the effects of Slc20a2 on bone in situ utilizing spatial transcriptomics (ST).

ST is a molecular profiling technique that enables mapping of mRNA molecules to their spatial position in a tissue cross-section, allowing interrogation of molecular pathways. 10x Genomics' Visium HD is the first ST platform with a continuous lawn of barcoded capture areas with a 2 μm resolution. Due to its required demineralization and tissue heterogeneity, bone has been underrepresented in transcriptomics studies. Use of CytAssist and optimized demineralization with EDTA enabled successful ST on bone. Visium HD was used to compare demineralized left tibia cross-sections of female 25 day old WT and Slc20a2-KO mice (n=4), generated from C57Bl/6J Slc20a2<tm1a (EUCOMM)Wtsi>/Ieg (Slc20a2) mice, under University of Washington IACUC #201600170.

Qiagen Ingenuity Pathway Analysis was used to compare the averaged differential gene expression in the bone remodeling region of the WT and KO mice and showed significant differential expression of biological pathways related to bone development and maintenance, in alignment with the observed phenotype. These studies confirm an important gene regulatory role of Slc20a2 in bone development and will provide new hypotheses to be tested in future studies.

15. A multi-kingdom genetic barcoding system for precise clone isolation

*Soh Ishiguro1,14; Kana Ishida2,14; Rina C. Sakata1,14; Minori Ichiraku3; Ren Takimoto1; Rina Yogo1; Yusuke Kijima1; Hideto Mori4; Mamoru Tanaka5; Samuel King1; Shoko Tarumoto6; Taro Tsujimura6; Omar Bashth1; Nanami Masuyama1,7,8; Arman Adel1; Hiromi Toyoshima5; Motoaki Seki5; Ju Hee Oh9; Anne-Sophie Archambault9; Keiji Nishida10,11; Akihiko Kondo9-12; Satoru Kuhara13; Hiroyuki Aburatani5; Ramon I. Klein Geltink9; Takuya Yamamoto3,6; Nika Shakiba1,4; Yasuhiro Takashima3; Nozomu Yachie1,4,5**

1. School of Biomedical Engineering, Faculty of Applied Science and Faculty of Medicine, The University of British Columbia, Vancouver, Canada 2. Spiber Inc., Tsuruoka, Japan 3. Center for iPS Cell Research and Application, Kyoto University, Kyoto, Japan. 4. Premium Research Institute for Human Metaverse Medicine (WPI-PRIMe), Osaka University, Suita, Osaka, Japan 5. Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan 6. Institute for the Advanced Study of Human Biology (WPI-ASHBi), Kyoto University, Kyoto, Japan 7. Institute for Advanced Biosciences, Keio University, Tsuruoka, Japan 8. Systems Biology Program, Graduate School of Media and Governance, Keio University, Fujisawa, Japan 9. BC Children's Hospital Research Institute, Department of Pathology and Laboratory Medicine, The University of British Columbia, Vancouver, Canada 10. Engineering Biology Research Center, Kobe University, Kobe, Japan 11. Graduate School of Science, Technology and Innovation, Kobe University, Kobe, Japan 12. Department of Chemical Science and Engineering, Graduate School of Engineering, Kobe University, Kobe, Japan 13. Graduate School of Bioresource and Bioenvironmental Sciences, Faculty of Agriculture, Kyushu University, Fukuoka, Japan 14. These authors contributed equally

Cell-tagging strategies with DNA barcodes have enabled the analysis of clone size dynamics and clone-restricted transcriptomic landscapes in heterogeneous populations. However, isolating a target clone that displays a specific phenotype from a complex population remains challenging. Here, we present a multi-kingdom genetic barcoding system, CloneSelect, which enables a target cell clone to be triggered to express a reporter gene for isolation through barcode-specific CRISPR base editing. In CloneSelect, cells are first stably tagged with DNA barcodes and propagated so that their subpopulation can be subjected to a given experiment. A clone that shows a

phenotype or genotype of interest at a given time can then be isolated from the initial or subsequent cell pools stored during the experiment using CRISPR base editing. CloneSelect is scalable and compatible with scRNA-seq. We demonstrate CloneSelect's versatility in human embryonic kidney (HEK) 293T cells, mouse embryonic stem cells (mESCs), human pluripotent stem cells (hPSCs), yeast cells, and bacterial cells.

16. A rapid retrospective cell clone isolation method

Ye, Chengyu; Yachie, Nozomu

School of Biomedical Engineering, Faculty of Applied Science and Faculty of Medicine, The University of British Columbia, Vancouver, Canada.

In today's biology, it is challenging to interrogate how a cell's prior molecular state primes its next state, as molecular analysis often requires the destruction of the cell. A recently emerging technological concept, called "retrospective clone isolation," has begun to address this challenge. In this approach, cells are individually labeled with molecular barcodes, such as unique DNA sequences. The barcoded cell population is then divided, with one portion subjected to a defined assay and the other stored. Once a clone exhibits a phenotype of interest, it can be retrieved from the stored population in a barcode-dependent manner, allowing the investigation of the molecular profiles that led to the phenotype. However, current methods rely on CRISPR genome editing, which is limited by response time and cytotoxicity. Here, we propose an RNA-based system, RapidSelect, which uses an RNA deaminase enzyme to overcome the limitations of current methods. In RapidSelect, cells encoding a specific RNA sequence are labeled by a rapid activation of a fluorescent reporter expression. This is achieved through the interaction between a barcode-targeting trigger RNA and a target reporter RNA that encodes an impaired fluorescent gene. The genetic code is repaired by deamination (A→I substitution) at the interacting double-stranded RNA region, restoring fluorescence. We will present our recent progress in the development of this new technology.

17. Scaling up massive parallel reporter assays with bulk quantitative density-based cell sorting

Hum, JJ; de Boer, Carl

University of British Columbia

Variations in regulatory DNA sequence affect gene expression and phenotype. Massive Parallel Reporter Assays (MPRAs) are a key tool for understanding regulatory DNA grammar, enabling simultaneous testing of how thousands of cis-regulatory elements (CREs) affect expression of a reporter gene. Many sequence configurations must be tested to capture the vast combinatorial interaction space of DNA binding proteins and develop an accurate predictive model of sequence to gene expression, necessitating an improvement in reporter assay scale.

I aim to improve MPRAs by developing a bulk quantitative density-based cell sorting (DBCS) method. In DBCS, cells express a cell-surface reporter protein that can be labeled with dense nanoparticles (DNPs). Labelled cell density increases proportional to reporter expression level, enabling quantitative sorting by isopycnic density gradient centrifugation. Cells that express more reporter protein will capture more DNPs and sediment at higher density. After fractionating the gradient, cells in each fraction are sequenced to link CREs to expression level. Bulk sorting time is independent of cell quantity, significantly increasing throughput compared to sort-seq MPRA where sorting time increases with cell quantity. Cell-to-cell variation can also be captured with DBCS, unlike bulk RNA-seq MPRA.

I am currently developing DBCS using HEK293T and K562 cells with endogenous, variably-expressed surface protein CD59 as a reporter. CD59 is labeled with a fluorescent antibody, which is then labeled by protein A-conjugated DNPs. By correlating fluorescent and density-based readouts, I will optimize parameters for high-resolution separation. I will then validate DBCS using a CRISPR base editing screen to mutate the CD59 coding region, as these edits will have predictable effects on gene expression. Finally, DBCS will be evaluated by comparison to bulk RNA-seq MPRA using a reporter

plasmid that contains a minimal promoter, cell surface marker, barcode sequence, and an enhancer library chosen for a wide range of expected expression.

18. Investigation of Local Effects of Transposable Elements in Preimplantation Embryos

Chang, Kai-Wei; Bashth, Omar; Dorri, Yeganeh; Shakiba, Nika;

School of Biomedical Engineering, UBC; School of Biomedical Engineering, UBC; School of Biomedical Engineering, UBC; School of Biomedical Engineering, UBC;

Transposable elements (TEs) are repeating DNA fragments originating from ancient viral invasion that have been domesticated in the human genome. While TEs are heavily epigenetically repressed to prevent deleterious transposition that may threaten the integrity of the somatic cell genome, transient TE activity in preimplantation blastocysts has been linked to the establishment of totipotency and cell fate determination. Recent studies further demonstrate the differential contributions of TE subfamilies towards embryonic lineages, such as inner cell mass and trophectoderm. To date, TE activity has been explored only at subfamily resolution, owing to the fact that TE subfamily members are indistinguishable from one another by their sequence and only differ in their genomic integration site. Therefore, whether TE activity is involved in loci-specific modulation of cell fate during human embryonic development remains an open question. Here, we propose a DNA barcoding approach to assess individual TE activity within stem cell-derived blastocyst-like structures, so-called “blastoids”. Multiple barcode designs will be evaluated based on their multi-loci knock-in efficiency, their detectability at the transcriptional level, and their impact on blastoid formation. By uniquely barcoding the repeated TE sequences within subfamilies of interest, we aim to investigate the activities of each TE to attribute their individual roles in driving cell fate in early human embryonic development.

19. Human gene regulatory network inference through a custom Peter-Clark algorithm

Yao, Herbert; Zhang, Jian; Khalilitousi, Sam; Kiyota, Brett; Yachie, Nozomu

School of Biomedical Engineering, Faculty of Applied Science and Faculty of Medicine, The University of British Columbia, Vancouver, BC V6T 1Z3, Canada

A human gene regulatory network (GRN) illustrates transcriptional gene regulatory cascades within cells. Deciphering the cascades of causal regulator genes and their target genes through biological measurements to understand disease progression and development has become a major focus in biology today. While many existing GRN inference algorithms leverage large-scale single-cell RNA sequencing data using regression models, they suffer from high false-positive rates, due to the intricacy of transcription regulatory network causing many gene pairs to numerically correlate even if they are not TF-gene pairs. In contrast, causal discovery algorithms, particularly the Peter-Clark (PC) algorithm, effectively eliminate false causal relationships. However, PC tends to over-prune true regulator–target pairs and has high computational costs (high false negative rate). To address these challenges, we propose two key modifications to the PC algorithm: (1) reducing uninformative statistical tests through graph theory to avoid accidental removal of true TF-gene edges, and (2) introducing biased causal discovery to enable a divide-and-conquer approach for large datasets by leveraging local sample agreement. Furthermore, to improve computational efficiency, we implemented our custom PC algorithm, termed PCC, in C programming, achieving a runtime reduction of over two orders of magnitude compared to the current de facto standard implementation of PC in Python. These improvements enable robust human GRN inference through the PC concept, potentially resulting in a more robust understanding of gene regulation in broader biology research.

20. Retrospective cell clone isolation using protein barcodes

Takimoto, Ren; Yogo, Rina; Mori, Hideto; Yachie, Nozomu

School of Biomedical Engineering, The University of British Columbia; School of Biomedical Engineering, The University of British Columbia; Premium Research Institute for Human Metaverse Medicine (PRIMe), Osaka University. ; School of Biomedical Engineering, The University of British Columbia., Premium Research Institute for Human Metaverse Medicine (PRIMe), Osaka University., Research Center for Advanced Science and Technology, The University of Tokyo.

Retrospective clone isolation is an emerging approach to interrogate dynamics and fates heterogeneous cell populations. In this concept, cells are initially tagged with molecular barcodes, propagated, and then split into subpopulations. One of the subpopulations undergoes a given assay, while the other is preserved. Upon identifying the barcode for a clone of interest showing a phenotype, the same clones can be isolated from the preserved population in a barcode-specific manner. This framework enables the identification of molecular factors that prime the cell clone's fate to derive the identified phenotype in the population. We recently established a high-performance method, CloneSelect, using CRISPR base editing. This method, however, has limitations in sensitivity and clone labeling speed, mainly due to the CRISPR reagent delivery and the time required for genome editing. To overcome these issues, we are working on the development of a programmable DNA-binding protein-based method, LightningSelect. In this system, we use either transcription activator-like effectors (TALEs) or zinc fingers (ZnFs) projected onto the cell surface to barcode cells. A barcoded clone of target can then be labeled by a fluorophore-conjugated double-stranded DNA, with no time lag related to intracellular processing. We plan to utilize LightningSelect and investigate various untapped biological phenomena, including the fate-priming of stem cells in human body development using advanced gastruloid models.

21. Massively parallel survey of CRISPR base editing activities

Yin, Liu; Nozomu Yachie

1 School of Biomedical Engineering, Faculty of Applied Science and Faculty of Medicine, The University of British Columbia, Vancouver, BC V6T 1Z3, Canada; 2 Premium Research Institute for Human Metaverse Medicine, Osaka University, Osaka, 565-0871, Japan; 3 Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo 153-8904, Japan

CRISPR base editing is a genome editing technology that enables precise gene editing without introducing double-stranded DNA breaks. Cytosine base editors (CBEs) are one type of the base editors which can directly convert cytosine to thymine at target genomic sites. While most current CBEs use cytidine deaminases derived from humans

or mice, these cytidine deaminases are found across a wide range of species, and the editing potential of many remains unexplored. To address this, we propose a high-throughput screening system in yeast cells using barcode fusion genetics (BFG) to evaluate the editing activity of hundreds of novel base editors incorporating cytidine deaminases from diverse organisms. This approach will enable us to systematically characterize the editing activities of these newly engineered base editors.

22. Modeling functional evolutionary trajectories of protein-coding sequences by high dimensional genotype fitness landscape

Romero, Sofia; Yachie, Nozomu

UBC

Sequence homology-based analysis enables the extrapolation of functional annotations between homologs, prediction of protein structures, as well as the reconstruction of the protein evolutionary trajectories. While the sequence-based phylogeny estimation captures evolutionary paths, the neutral theory of molecular evolution suggests that sequence evolution does not always impact the functional evolution of genes since most DNA changes are not adaptive or subjects to natural selection. Besides functionally beneficial mutations that incrementally evolve protein function, mutations that are initially neutral can either become beneficial through subsequent mutations or develop into constraints. In contrast, certain mutations that initially yield beneficial effects in evolution may later be masked by subsequent mutations, rendering them seemingly neutral for biological observation. Our current view of gene evolution might underestimate such “evolutionary buffering mutations” and “evolutionary masked mutations.” Therefore, the understanding of evolution and evolvability of proteins remains elusive in this regard. Here, we plan to measure high-dimensional fitness landscapes of yeast and human histone subunits in yeast cells and associate them with their sequence evolutionary trajectories. We will establish a system to provide single, double, and higher-order synthetic mutations to a pair of yeast, human and ancestral histone subunits and measure their fitness. Aside from the measurement of the

evolutionary landscape, we also envision such a high- throughput genotype-fitness landscape would provide a wealth of resources for functional and structural prediction as well as protein engineering.

23. Massive parallel screening to increase genome editing outcomes in human stem cells

Kagieva, Madina (1); Ishiguro, Soh (1); Mori, Hideto (2); Masuyama, Nanami (3); Yachie, Nozomu (1, 2, 4).

1. School of Biomedical Engineering (SBME), the University of British Columbia, Vancouver, BC V6T 1Z3, Canada 2. Premium Research Institute for Human Metaverse Medicine (PRIMe), Osaka University, Osaka repair genes + 565-0871, Japan 3. Institute for Advanced Biosciences, Keio University, Tsuruoka 97-0035, Japan 4. Research Center for Advanced Science and Technology (RCAST), the University of Tokyo, Tokyo 153-8904, Japan

With the potential to differentiate into any cell type in the body and the ability for infinite self-renewal, human pluripotent stem cells (hPSCs) hold great promise for applications in cell replacement therapy, tissue engineering, and modelling human genetic diseases. The majority of the human disease mutations are single nucleotide variants (SNVs), which have led the genome editing field to develop CRISPR-Cas9 base editors to enable SNV corrections. However, the efficacy and precision of CRISPR base editors remain to be improved, especially in hPSCs. To address this, I am developing a system to screen base editor variants developed so far and identify ones with high efficiencies in hPSCs using a virus-like-particle (VLP) system. VLPs are self-assembling protein structures that mimic the architecture of viruses while encapsulating cargo molecules produced by host cells. Although they retain the ability to transduce mammalian cells, VLPs are non-infectious and incapable of replication, as they lack viral genetic material. Compared to conventional delivery systems, VLPs offer several advantages: they can accommodate larger cargo sizes, efficiently package ribonucleoproteins (RNPs) to reduce off-target effects and can be pseudotyped with specific glycoproteins to direct cargo selectively to target cells. In parallel, I am performing an assay to elucidate if certain markers like p53 and UGI make human stem cells more susceptible to editing.

With this research, I envision establishing a functional toolkit to edit human pluripotent stem cells more efficiently. This will push the field forward to unleash the full potential of hPSCs for cell therapy and disease modelling application.

24. An environment-dependent contactome map for a eukaryotic cell

Sheykhkarimli, Dayag; Liu, Betty B.; Laval, Florent; Spirohn, Kerstin; Kishore, Nishka; Knapp, Jennifer J.; Rayhan, Ashyad; Škalič, Miha; Li, Roujia; Horecka, Ira; Kuang, Da; Mee, Miles; Dugied, Guillaume; Colabella, Claudia; Cote, Atina G.; Gebbia, Marinella; Nguyen, Maria; Hao, Tong; Lambourne, Luke; Leutert, Mario; Li, Siyang; Verby, Marta; Pogoutse, Oxana; Torres, Nikko; Ho, Brandon; Brown, Grant W.; Rost, Hannes; Petrovič, Uroš; Villén, Judit; Bader, Gary; Calderwood, Michael A.; Hill, David E.; Vidal, Marc; Petsalaki, Evangelia; Yachie, Nozomu; Kim, Dae-Kyum; Roth, Frederick P.

*1Donnelly Centre, University of Toronto, Toronto, Ontario, Canada; 2Department of Molecular Genetics and Computer Science, University of Toronto, Toronto, Ontario, Canada; 3Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada; 4Unité de Génétique Moléculaire des Virus à ARN, Département de Virologie, Institut Pasteur, Paris, France; 5UMR3569, Centre National de la Recherche Scientifique, Paris, France; 6Université Paris Cité, Sorbonne Paris Cité, Paris, France; 7Computational Science Laboratory, University Pompeu Fabra, Barcelona, Spain; 8Istituto Zooprofilattico Sperimentale dell'Umbria e delle Marche "Togo Rosati", Perugia, Italy; 9Center for Cancer Systems Biology (CCSB), Dana-Farber Cancer Institute, Boston, MA, USA; 10Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA, USA; 11Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA, USA; 12TERRA Teaching and Research Centre, University of Liège, Gembloux, Belgium; 13Laboratory of Viral Interactomes, GIGA Institute, University of Liège, Liège, Belgium; 14Laboratory of Molecular and Cellular Epigenetics, GIGA Institute, University of Liège, Liège, Belgium; 15Department of Genome Sciences, University of Washington, Seattle, WA, USA; 16Department of Biochemistry, University of Toronto, Toronto, Ontario, Canada; 17Department of Biology, Biotechnical Faculty, University of Ljubljana, Ljubljana, Slovenia; 18Department of Molecular and Biomedical Sciences, Jozef Stefan Institute, Ljubljana, Slovenia; 19European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Hinxton, UK; 20School of Biomedical Engineering, University of British Columbia, Vancouver, BC, Canada; 21Research Center for Advanced Science and Technology, University of Tokyo, Tokyo, Japan. *To whom correspondence should be addressed*

Dynamic protein interactions maintain cellular homeostasis in response to perturbations. However, dependence of protein interaction on environmental change has not been explored at proteome scale. Here we investigated proteome-scale dynamics of the set of direct protein-protein contacts – the contactome – in the model eukaryote S. cerevisiae. We generated a 'pilot' atlas by measuring over a million protein pairs in 31 environments, yielding four proteome-scale maps (in baseline, DNA damage, carbon starvation, and oxidative stress environments) evaluating ~18M pairwise

combinations of ~5K yeast proteins. Orthogonal assays showed our interactions to have quality on par with interactions observed at least five times in the literature. The resulting validated atlas (the Environmental Yeast Reference Interactome or EYRI) contains a union of 1.5K high-quality interactions, of which we estimate ~30% to be dynamic across environments. Transcriptomic and phosphoproteomic measurements, obtained under matching environments, enabled systematic discovery of interaction-regulatory models. Thus, EYRI represents a multimodal proteome scale resource for exploring environment-dependent regulation of eukaryotic protein-protein contacts.

25. Massively parallel gene assembly in Semi-Permeable Capsules

Sgrizzi, Samantha ; Mai, David ; Pathak, Nandini ; Srivatsan, Sanjay

Molecular and Cellular Biology Program, University of Washington, Seattle, WA, USA; Basic Sciences Division, Fred Hutchinson Cancer Center, Seattle, WA, USA; Computational Biology Program, Fred Hutchinson Cancer Center, Seattle, WA, USA; Department of Genome Sciences, University of Washington, Seattle, WA, USA; Institute for Protein Design, University of Washington, Seattle, WA, USA; Brotman Baty Institute for Precision Medicine, Seattle, WA, USA.

Affordable synthesis of large DNA libraries remains a major obstacle for exploring sequence-to-function landscapes. Existing array-based DNA assembly platforms are limited by scale, while emulsion-based technologies lack the flexibility needed to perform sequential reactions (e.g. assembly, amplification, and error correction). Here we introduce CapSynth, a novel DNA assembly workflow that builds off the DropSynth technology. Central to the DropSynth method is the ability to isolate and concentrate DNA fragments specifically needed for assembling each gene within oil-based emulsions, increasing scalability and limiting bias. However, these emulsion-based systems inherently lack permeability, limiting them to single-step reactions and necessitating compartment disruption before downstream applications. CapSynth utilizes Semi-Permeable Capsules (SPCs) to perform programmed enzymatic DNA assembly within capsules. SPCs are solvent-based capsules that can facilitate the selective exchange of biomolecules based on size. Microarray-synthesized oligonucleotides are partitioned first onto a solid substrate and then encapsulated

within SPCs. Subsequent enzymatic assembly within SPCs generates full-length DNA that can be used for a variety of downstream applications. As an initial proof of concept, using a 60-member oligo pool, we assembled 5 distinct open reading frames of ~800bp in a single run, achieving highly uniform coverage compared to a matched in solution reaction. We are currently working to scale this platform to assemble thousands of genes simultaneously. In the future, we anticipate that the generation of assembled genes within capsules will enable sequential enzymatic reactions, in vitro expression, and screening applications while affording DNA assembly at scale.

26. Automated Cell Type Re-Annotation for Enhanced Reproducibility and Meta-analysis in Single-Cell Brain Studies

Schwartz, Rachel; Pavlidis, Paul

Bioinformatics graduate program, Department of Psychiatry; Michael Smith Laboratories, Department of Psychiatry

The rapid expansion of single-cell RNA sequencing (scRNAseq) data enables meta-analysis of gene expression, but inconsistent cell-type annotations hinder such efforts. Research in our lab uses large amounts of transcriptomic data for meta-analyses, managed through Gemma (gemma.msl.ubc.ca). To enable reuse of scRNAseq data, we developed a pipeline for automated annotation of public single-cell data, focusing on human and mouse neocortex and hippocampus using a unified cell type taxonomy with minimal need for manual curation. Preliminary results include benchmarking of 14 human and 13 mouse experiments across 480 and 240 variable parameters, respectively. Results of this benchmarking informed a Nextflow pipeline leveraging community-hosted deep learning models of single-cell atlases to automatically classify cell types. These parameters achieved high performance across diverse test datasets, with marginal sample-level F1 scores of 0.95 and 0.77 in human and mouse, respectively. Harmonized cell-type annotations generated by this pipeline enable statistical replication and re-analysis of gene expression patterns within public

data. Initial efforts focus on single-cell studies of neuropsychiatric disease and/or disease models in the human and mouse forebrain.

27. Exploration of Large-Scale Differential Expression Datasets with Co-Differential Expression

Patadia, Neera; Pavlidis, Paul

Graduate Program in Bioinformatics; UBC Department of Psychiatry

In genomics, we are interested in understanding how gene expression changes in response to biological conditions. This is often done through differential expression (DE) analysis, which compares gene expression between a baseline and a condition of interest. While individual DE studies provide insights into condition-specific changes, it remains unclear how DE patterns relate across different conditions. By aggregating DE data from multiple studies, we can identify broader patterns of gene regulation. To address this, the Pavlidis lab developed Gemma, a database that compiles thousands of DE analyses from diverse biological contexts. Using this data, I examined trends correlated in DE measures (such as Log2 fold changes) across all conditions in a “co-differential expression” analysis. This approach is inspired by co-expression analysis, which clusters genes based on expression correlations across samples. Co-differential expression on the other hand focuses on how genes respond similarly across conditions, and potentially overcome confounds that may arise in co expression analyses due to factors like tissue composition or sample sex. To interpret these patterns, I will compare co-differential expression results with co-expression data from GTEx tissue expression data. While I expect some overlap in gene correlation patterns, I aim to uncover biologically meaningful patterns unique to DE data. Furthermore, in my analyses I examine both human and mouse DE data, and compare trends in gene DE between orthologous genes. Understanding relationships between conditions through co-differential expression may reveal underlying molecular mechanisms that shape gene regulation across diverse contexts.

28. Creating a novel whole-animal synthetic biology platform using *C. elegans*

Speed, Haley N.; Moerdyk-Schauwecker, Megan J., Estevez, Annette O., Stevenson, Zach C., Phillips, Patrick C.

Institute of Ecology and Evolution, University of Oregon; Institute of Ecology and Evolution, University of Oregon; Institute of Ecology and Evolution, University of Oregon; Department of Genome Sciences, University of Washington; Institute of Ecology and Evolution, University of Oregon

Over the last 25 years, the field of synthetic biology has successfully translated fundamental principles from engineering towards building novel biological systems. A central feature of this work has been the remapping of modular molecular regulatory components, such as transcription factors and promoter sequences, into the base components of synthetic activation and deactivation switches. The vast majority of synthetic biology has been carried out using single-cell systems—first in bacteria, then in yeast, and now, importantly, within mammalian cell lines. The focus on single cells is natural, as one of the limiting factors in rigorously testing and implementing synthetic circuit design has been transforming the cells themselves with the novel genetic components needed to generate and test synthetic circuits at scale. However, single-cell systems lack critical features that are central to many important biological and medical applications. Importantly, humans are complex multicellular organisms that undergo developmental differentiation, whose genomes generate tissue and organ specific genetic function, and whose cells communicate over distance using neurological and chemical means—all within the context of dynamic interactions with environmental inputs including nutrition, potential toxins, administered drugs, and other medical treatments. Applying approaches from synthetic biology to these critical areas requires the development of a general-use, whole-animal research platform. Success in establishing this platform holds the promise of establishing an entirely new branch of synthetic biology that can be used to study fundamental biology, as well as creating the potential to assist in the development of novel therapeutics and health interventions. We present our current progress in combining novel genomic engineering approaches developed in our lab with existing artificial gene circuit design from single cell systems

to create a new whole-animal synthetic biology platform utilizing the power of the Caenorhabditis elegans nematode model system.

29. Comparative Single-Nucleus Transcriptomic Analysis of Gene Co-Expression in Microglia Across Alzheimer's and Healthy Brains

Elazzabi, Nairuz; Pavlidis, Paul

Michael Smith Laboratories, University of British Columbia

Recent studies suggest a prominent role of neuroinflammation and the effect of innate and adaptive immune activation in Alzheimer's disease (AD) progression. Advancements in single-nucleus RNA-sequencing have enabled the comprehensive profiling of cell-type-specific transcriptomic alterations within the postmortem frontal cortex of individuals at varying stages of AD pathology. These studies have shown that AD pathogenesis involves the complex interplay of virtually every major brain cell type, identifying gene expression signatures that delineate the transition from healthy aging to cognitive decline. In this study, I compile and analyze 14 publicly available single-nucleus RNA-seq datasets, encompassing postmortem brain samples from 3,659 donors across diverse AD pathological and clinical stages, as well as healthy controls and a unique cohort of fresh cortical biopsies. I focus on microglia, key mediators of neuroinflammation, to construct microglia-specific gene co-expression networks and perform comparative analyses between AD and healthy brains. This approach aims to identify reproducible patterns of differential coordinated expression between transcriptional regulators and their gene partners across major AD studies, offering insights into the molecular mechanisms underlying microglial dysfunction in AD.

30. Development of a High-Throughput Platform for Transcription Factor Characterization

William Cheney, Omar Tariq, Carl de Boer

School of Biomedical Engineering,, University of British Columbia

Transcription factors coordinate gene expression. They function by binding to specific DNA sequences and regulating transcription. A comprehensive understanding of gene regulation and potential effects of mutation is reliant on our ability to identify sequences a transcription factor will bind. Notably, transcription factor binding motifs, encapsulating the optimal binding sites, are unknown for approximately 60% of all eukaryotic transcription factors. Current techniques for identifying motifs are labour-intensive and have limited scalability. Here, we augment our previously described Gigantic Parallel Reporter Assay (GPRA), potentially enabling the characterization of binding motifs for a diverse library of transcription factors. A GPRA leverages randomized synthetic promoters to control expression of a fluorescent reporter. By adding an exogenous library of transcription factors into the GPRA, we facilitate parallel characterization of their respective binding motifs. We have implemented a pilot GPRA experiment encompassing 96 human transcription factors. Transcription factors were chosen to analyze the robustness of our approach, including well-studied transcription factors as well as those with hitherto unexplored DNA binding domains. The TF-GPRA generated a dataset of over 3 million sequences with corresponding expression levels. These data will be used to train a machine learning model that will identify transcription factor binding motifs.