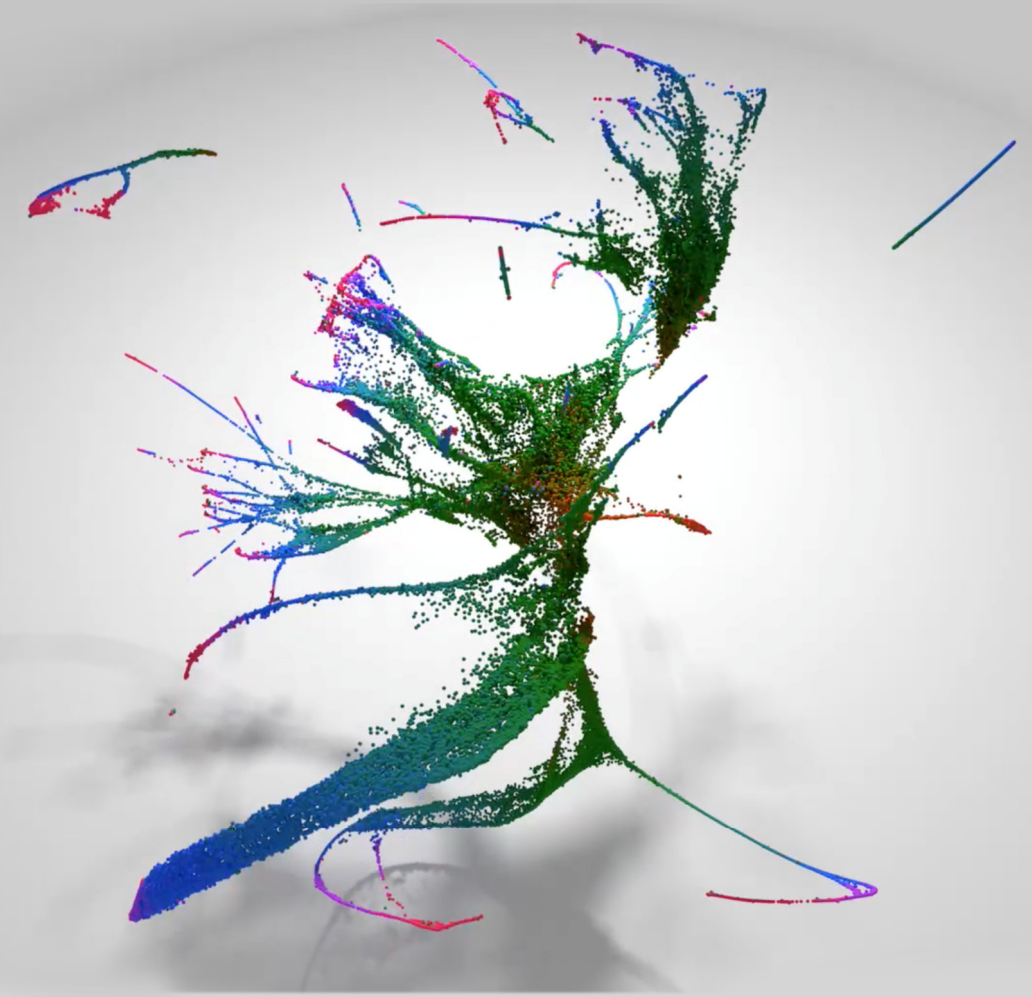
Abstracts of papers presented
at the 2019 meeting on

SINGLE CELL ANALYSES

November 13–November 16, 2019



CSH Cold Spring Harbor Laboratory
MEETINGS & COURSES PROGRAM

Abstracts of papers presented
at the 2019 meeting on

SINGLE CELL ANALYSES

November 13–November 16, 2019

Arranged by

Nancy Allbritton, *University of North Carolina*
Scott Fraser, *University of Southern California*
Junhyong Kim, *University of Pennsylvania*

This meeting was funded in part by **Bio-Techne; Cell Microsystems; SeqGeq; Takara Bio;** and **10x Genomics.**

Contributions from the following companies provide core support for the Cold Spring Harbor meetings program.

Corporate Benefactors

Estée Lauder Companies
Regeneron
Thermo Fisher Scientific

Corporate Sponsors

Agilent Technologies
Bayer
Bristol-Myers Squibb Company
Calico Labs
Celgene
Genentech, Inc.
Merck
New England BioLabs
Pfizer

Corporate Partners

Alexandria Real Estate
Enzo Life Sciences
Gilead Sciences
Lundbeck
Novartis Institutes for Biomedical Research
Sanofi

Cover: 3D UMAP plot of 80,000 *C. elegans* cells during embryo development.
Image credit: Cole Trapnell, University of Washington.

SINGLE CELL ANALYSES

Wednesday, November 13 – Saturday, November 16, 2019

Wednesday	7:30 pm	Welcome Remarks
		Keynote Speaker
Thursday	9:00 am	**1** Spatial Single Cell Biology
Thursday	2:00 pm	**2** Epigenome, Informatics, and Applications
Thursday	4:30 pm	**3** Flash Talks
Thursday	7:30 pm	**4** Poster Session & Wine and Cheese Party
Friday	9:00 am	**5** Photonics and Imaging for Single Cell Biology
Friday	1:30 pm	**6** Single Cells in Development
Friday	5:00 pm	Keynote Speaker
Friday	6:00 pm	Banquet
Saturday	9:00 am	**7** Physical / Chemical Single Cell Biology

Workshops

Cell Microsystems: Thursday, following morning session (abstract T-1)
SeqGeq: Friday, following morning session (abstract T-2)

Mealtimes at Blackford Hall are as follows:

Breakfast 7:30 am-9:00 am
Lunch 11:30 am-1:30 pm
Dinner 5:30 pm-7:00 pm

Bar is open from 5:00 pm until late

Cold Spring Harbor Laboratory is committed to maintaining a safe and respectful environment for all meeting attendees, and does not permit or tolerate discrimination or harassment in any form. By participating in this meeting, you agree to abide by the Code of Conduct, which is available both online and at the back of this book.

Abstracts are the responsibility of the author(s) and publication of an abstract does not imply endorsement by Cold Spring Harbor Laboratory of the studies reported in the abstract.

These abstracts should not be cited in bibliographies. Material herein should be treated as personal communications and should be cited as such only with the consent of the author(s).

Please note that photography or video/audio recording of oral presentations or individual posters is strictly prohibited except with the advance permission of the author(s), the organizers, and Cold Spring Harbor Laboratory.

Any discussion via social media platforms of material presented at this meeting requires explicit permission from the presenting author(s).

Printed on 100% recycled paper.

PROGRAM

WEDNESDAY, November 13—7:30 PM

Welcome and Introduction

Junhyong Kim
University of Pennsylvania

Scott Fraser
University of Southern California

KEYNOTE SPEAKER

Xiaoliang Sunney Xie
Peking University & Harvard University

THURSDAY, November 14—9:00 AM

SESSION 1 SPATIAL SINGLE CELL BIOLOGY

Chairperson: **Leeat Keren,** Stanford University, California

seqFISH+—in situ transcriptome profiling
Long Cai.
Presenter affiliation: Caltech, Pasadena, California. 1

Shared and specific long-range projections across transcriptionally-defined neuronal subtypes revealed by *in situ* barcode sequencing
Xiaoyin Chen, Yu-Chi Sun, Anthony M. Zador.
Presenter affiliation: Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. 2

In situ transcriptome accessibility sequencing (INSTA-seq) reveals in vivo mRNA-protein interactions
Daniel Furth, Victor Hatini, Je H. Lee.
Presenter affiliation: Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. 3

THURSDAY, November 14—2:00 PM

SESSION 2 EPIGENOME, INFORMATICS, AND APPLICATIONS

Chairperson: **Junhyong Kim,** University of Pennsylvania, Philadelphia

THURSDAY, November 14—4:30 PM

SESSION 3 FLASH TALKS

Chairperson: **Junhyong Kim,** University of Pennsylvania, Philadelphia

THURSDAY, November 14—7:30 PM

SESSION 4 POSTER SESSION

Wine & Cheese Party
Sponsored by Thermo Fisher Scientific

FRIDAY, November 15—9:00 AM

SESSION 5 PHOTONICS AND IMAGING FOR SINGLE CELL BIOLOGY

Chairperson: **Hernan Garcia,** University of California, Berkeley

SESSION 6 SINGLE CELLS IN DEVELOPMENT

Chairperson: **Erik van Nimwegen,** University of Basel, Switzerland

Reconstruction of an invariant lineage in chordate embryogenesis
Tengjiao Zhang, Yichi Xu, Teng Fei, Kaoru Imai, Tianwei Yu, Yutaka Satou, Zhirong Bao, Weiyang Shi.
Presenter affiliation: Sloan Kettering Institute, New York, New York. 118

Single-cell transcriptome analysis of CD34+ stem cell-derived myeloid cells infected with human cytomegalovirus
Laura Hertel, Melissa Galinato, Kristen Shimoda, Alexis Aguiar, Fiona Hennig, Dario Boffelli, Michael McVoy.
Presenter affiliation: Children's Hospital Oakland Research Institute, Oakland, California. 119

Inferring gene regulatory landscapes from single-cell data
Erik van Nimwegen.
Presenter affiliation: University of Basel, Basel, Switzerland. 120

FRIDAY, November 15—5:00 PM

KEYNOTE SPEAKER

James Eberwine
University of Pennsylvania

FRIDAY, November 15

BANQUET

Cocktails 6:00 PM Dinner 6:30 PM

SESSION 7 PHYSICAL / CHEMICAL SINGLE CELL BIOLOGY

Chairperson: **Scott Fraser,** University of Southern California, Los Angeles

seqFISH+: IN SITU TRANSCRIPTOME PROFILING

Long Cai

Caltech, Biology and Biological Engineering, Pasadena, CA

Imaging the transcriptome in situ with high accuracy has been a major challenge in single cell biology, particularly hindered by the limits of optical resolution and the density of transcripts in single cells. We developed seqFISH+, that can image the mRNAs for 10,000 genes in single cells with high accuracy and sub-diffraction-limit resolution, in the mouse brain cortex, subventricular zone, and the olfactory bulb, using a standard confocal microscope. The transcriptome level profiling of seqFISH+ allows unbiased identification of cell classes and their spatial organization in tissues. In addition, seqFISH+ reveals subcellular mRNA localization patterns in cells and ligand-receptor pairs across neighboring cells. This technology demonstrates the ability to generate spatial cell atlases and to perform discovery-driven studies of biological processes in situ.

SHARED AND SPECIFIC LONG-RANGE PROJECTIONS ACROSS TRANSCRIPTIONALLY-DEFINED NEURONAL SUBTYPES REVEALED BY *IN SITU* BARCODE SEQUENCING

Xiaoyin Chen, Yu-Chi Sun, Anthony M Zador

Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Understanding how neuronal circuits function requires resolving the interactions among myriad neurons with diverse characteristics, such as anatomy, gene expression, and activity. Deciphering the relationship among neuronal types requires high throughput and cellular resolution, which is difficult to achieve using conventional anatomical methods. To address these challenges, we have developed BARseq, a high-throughput technique for mapping long-range neuronal projections at cellular resolution using RNA barcode sequencing. Combining BARseq with Cre-line labeling, in situ hybridization, or in situ sequencing of endogenous genes further allows simultaneous measurement of neuronal projections and gene expression at high spatial resolution and high throughput. We used BARseq to determine the projections of 2288 neurons in mouse auditory cortex to 11 brain areas. BARseq confirmed the laminar organization of the three major classes (IT, PT and CT) of projection neuron. In depth analysis uncovered both combinatorial projection patterns specific to transcriptionally-defined subtypes of IT neuron and projection patterns shared across subtypes. BARseq thus uncovers a complex organization of projections across transcriptionally-defined subtypes of IT neurons that differs from the distinct separation among the major classes. By combining projection patterns and gene expression, BARseq bridges anatomical and transcriptomic approaches at single-neuron resolution with high throughput, and thereby can potentially uncover the organizing principles underlying the structure and formation of neural circuits.

IN SITU TRANSCRIPTOME ACCESSIBILITY SEQUENCING (INSTA-seq) REVEALS IN VIVO mRNA-PROTEIN INTERACTIONS

Daniel Furth[1], Victor Hatini[2], Je H Lee[1]

[1]Cold Spring Harbor Laboratory, Functional Genomics, Cold Spring Harbor, NY, [2]Tufts School of Medicine, Sackler School of Biomedical Science, Boston, MA

Subcellular RNA localization regulates spatially polarized cellular processes, but unbiased investigation of its control *in vivo* remains challenging. Current hybridization-based methods cannot differentiate small regulatory variants, while *in situ* sequencing is limited by short reads. We solved these problems using a bidirectional sequencing chemistry to efficiently image transcript-specific barcode *in situ*, which are then extracted and assembled into longer reads using NGS. In the *Drosophila* retina, genes regulating eye development and cytoskeletal organization were enriched compared to methods using extracted RNA. We therefore named our method In Situ Transcriptome Accessibility sequencing (INSTA-seq). Sequencing reads terminated near 3' UTR *cis*-motifs (e.g. *Zip48C*, *stau*), revealing RNA-protein interactions. Additionally, *Act5C* polyadenylation isoforms retaining zipcode motifs were selectively localized to the optical stalk, consistent with their biology. Our platform provides a powerful way to visualize any RNA variants or protein interactions *in situ* to study their regulation in animal development.

BAG-seq: COPOLYMERIZATION OF NUCLEIC ACIDS INTO BALLS OF ACRYLAMIDE GEL (BAG) FOR SINGLE-CELL DNA/RNA SEQUENCING

Siran Li[1], Jude Kendall[1], Sarah Park[1], Zihua Wang[1], Joan Alexander[1], Andrea Moffitt[1], Nissim Ranade[1], Cassidy Danyko[1], Jordan Gegenhuber[1], Jessica Tollkuhn[1], Brian D Robinson[2], Herbert Lepor[3], Stephan Fischer[1], Jesse Gillis[1], Eric Brouzes[4], Alex Krasnitz[1], Dan Levy[1], Michael Wigler[1]

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, [2]Weill Medical College of Cornell University, Department of Pathology and Laboratory Medicine, New York, NY, [3]New York University Langone Medical Center, Department of Urology, New York, NY, [4]Stony Brook University, Department of Biomedical Engineering, Stony Brook, NY

Single cell analyses are increasingly employed for understanding the patterns of gene expression and genomic variation in complex populations of cells and tissues. Many droplet-based technologies have emerged as high-throughput ways to study DNA or RNA of single cells. However, these methods often lack the breadth of coverage. Droplet merging and breakage give rise to cross-contamination. Moreover, some droplet-based methods suffer from inefficient use of samples. In addition, these technologies were mostly designed to do only "one thing".

Here, we introduce a novel high-throughput method by using 5'-Acrydite oligonucleotides to copolymerize single-cell DNA or RNA into balls of acrylamide gel (BAG). We transform liquid droplets into solid porous BAGs with cell components covalently linked within. Combining this step with split-and-pool techniques for creating barcodes yields a method with advantages in cost and scalability, depth of coverage, ease of operation, minimal cross-contamination and efficient use of samples. We successfully perform DNA copy number profiling on mixtures of cell lines, nuclei from frozen prostate tumors, and cancer biopsy washes. As applied to RNA, we demonstrate that the BAG-seq method exceeds the number of genes and unique templates captured per cell, and has fewer barcode collisions compared to other high-throughput methods. We apply this method to individual nuclei from the mouse brain. In addition, by using varietal tags to achieve sequence error correction, we show extremely low levels of cross-contamination by tracking source-specific SNVs. The method is very flexible and readily modified, and we will talk about its great adaptability and its reusability.

GIOTTO, A PIPELINE FOR INTEGRATIVE ANALYSIS AND VISUALIZATION OF SINGLE-CELL SPATIAL TRANSCRIPTOMIC DATA

Ruben Dries[1,4], Qian Zhu[1,4], Chee-Huat Linus Eng[2], Arpan Sarkar[3], Feng Bao[1], Rani George[1], Nico Pierson[2], Long Cai[2], Guo-Cheng Yuan[1]

[1]Dana Farber Cancer Institute, Pediatric Oncology, Boston, MA, [2]California Institute of Technology, Biology and Biological Engineering, Pasadena, CA, [3]Harvard University, Molecular and Cellular Biology, Cambridge, MA, [4]Equal contribution

The rapid development of novel spatial transcriptomics technologies has provided new opportunities to investigate the interactions between cells and their native microenvironment. However, effective use of such technologies requires the development of innovative computational algorithms and pipelines. Here we present Giotto, a comprehensive, flexible, robust, and open-source pipeline for spatial transcriptomic data analysis and visualization. The data analysis module implements a wide range of algorithms ranging from basic tasks such as data pre-processing to innovative approaches for cell-cell interaction characterization. The data visualization module provides a user-friendly workspace that allows users to interactively visualize, explore and compare multiple layers of information. These two modules can be used iteratively for refined analysis and hypothesis development. We illustrate the functionalities of Giotto by using the recently published seqFISH+ dataset for mouse brain. Our analysis highlights the utility of Giotto for characterizing tissue spatial organization as well as for the interactive exploration of multi- layer information in spatial transcriptomic and imaging data. We find that single-cell resolution spatial information is essential for the investigation of ligand-receptor mediated cell-cell interactions. Giotto is generally applicable and can be easily integrated with external software packages for multi-omic data integration.

The first module (Giotto Analyzer) provides instructions about the different steps in analyzing spatial single-cell data, whereas the second module (Giotto Viewer) provides a fast and interactive viewer of spatial single-cell data and additional annotations (such as stainings, morphology and subcellular transcript localization information).

Giotto Analyzer requires as minimal input a gene-by-cell count matrix and the spatial coordinates for the centroid position of each cell. At the basic level, Giotto Analyzer can be used to perform common steps similar to single-cell RNAseq analysis, such as pre-processing, feature selection, dimension reduction and unsupervised clustering; however, the main strength comes from its ability to integrate gene expression and spatial information in order to gain insights into the structural and functional organization of a tissue. The Giotto Viewer module is designed both to interactively explore the outputs of Giotto Analyzer and to visualize additional information such as cell morphology and transcript locations. Giotto Viewer allows users to easily explore the data in both physical and expression space and identify relationships between different data modalities.

Giotto is available at http://spatial.rc.fas.harvard.edu/ with links to repository websites and tutorials. An online demo of the Giotto Viewer can be found at http://spatial.rc.fas.harvard.edu/giotto-viewer/.

SINGLE CELL TRANSCRIPTOME AND CHROMATIN CONFORMATION ANALYSIS REVEAL PANCREATIC HORMONE LINEAGE DECISIONS

Eliza Duvall[1], Cecil Benitez[2], Martin Enge[3], Songjoon Baek[1], Nathan Sheffield[4], Steve Quake[3], Seung Kim[2], Efsun Arda[1]

[1]Laboratory of Receptor Biology and Gene Expression, NCI, National Institutes of Health, Bethesda, MD, [2]Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA, [3]Department of Bioengineering and Applied Physics, Stanford University, Stanford, CA, [4]Center for Public Health Genomics, University of Virginia, Charlottesville, VA

All forms of diabetes result from loss or impairment of pancreatic beta cell function. Recent work demonstrating the feasibility of replacing beta cells with insulin-producing cells generated from renewable sources such as embryonic stem cells, has sparked great excitement. However, the molecular and cellular processes leading to beta cell differentiation are not well understood. During embryonic development, activation of a transcription factor, Neurog3, in select pancreatic duct cells specifies the endocrine progenitors, which will then differentiate into hormone producing cells, such as beta, alpha, and delta cells. Despite recent progress, molecular features of endocrine progenitors remain elusive, and how subsequent hormone cell fates are determined is largely unknown.

In this study, we used lineage-traced Neurog3 cells labeled with eGFP from mouse embryos to perform single cell transcriptome analysis in order to elucidate gene expression changes during differentiation. By establishing pseudotime trajectories and using differential gene expression analysis, we identified unique combinations of transcription factors for each of the beta, alpha, and delta lineages. We found that these transcription factors turn on and off at precise moments to guide endocrine cell differentiation. In addition, we performed ATAC-seq to investigate Neurog3-dependent chromatin conformation changes in endocrine progenitor cells. We discovered genomic regions that undergo substantial transformation during development as well as enriched motifs in differentiation-stage specific open regions. Further, we applied genomic footprinting analysis to identify transcription factor activity in open chromatin regions and found evidence of specific transcription factor footprints linked to their associated motifs. Taken together, we found 17 significant transcription factors common in both analyses that are enriched during the transition from progenitor to specific endocrine lineage. Our results demonstrate the feasibility of using a combined scRNA-seq and ATAC-seq analysis to gain insights into gene regulatory networks that define cell lineages. We anticipate our findings from these and similar efforts will reveal gene regulatory networks governing pancreatic endocrine development, and facilitate the efforts toward stem cell based therapies and tissue regeneration.

UNRAVELING THE TUMOR IMMUNE MICROENVIRONMENT USING MULTIPLEXED ION BEAM IMAGING

Leeat Keren

Stanford University, Pathology, Stanford, CA

The immune system is critical in modulating cancer progression, but knowledge of immune composition, phenotype, and interactions with tumor is limited. Tumors are spatially organized ecosystems that are comprised of distinct cell types, each of which can assume a variety of phenotypes defined by co-expression of multiple proteins. Thus, to interrogate solid tumor biology and response to treatment, it is necessary to gauge the expression of a multitude of proteins with single-cell or even subcellular resolution while preserving spatial information. In this talk I will present Multiplexed Ion Beam Imaging by Time of Flight (MIBI-TOF), a method in which human tissue sections are stained with a mix of antibodies conjugated to metals and their expression levels are read out by secondary ionization mass spectrometry.The result is a multi-dimensional image, depicting sub-cellular expression of 40 distinct proteins in situ. I will present results in which we applied MIBI-TOF to study the immune tumor microenvironment (iTME) in triple-negative breast cancer (TNBC) and uncovered several layers of organization. We showed that in TNBC there are several archetypical iTME organizations, linking molecular expression patterns, cell composition and histology, and these are predictive of patient survival. I will present algorithmic advances in analyzing multiplexed spatial datasets and how these are applied to study the iTME across a variety of malignancies.

REVERSE ENGINEERING OF THE GLOBAL NEURAL CREST GENE REGULATORY NETWORK IN VIVO

Ruth M Williams, Ivan Candido-Ferreira, Emmanouela Repapi, Tatjana Sauka-Spengler

University of Oxford, Radcliffe Department of Medicine, Oxford, United Kingdom

Precise control of developmental processes is encoded in the genome in the form of gene regulatory networks (GRNs). Such multi-factorial systems are difficult to decode in vertebrates owing to their complex gene hierarchies and dynamic molecular interactions. We present a genome-wide *in vivo* reconstruction of the GRN underlying development of the multipotent neural crest (NC) embryonic cell population. By coupling NC-specific epigenomic and transcriptional profiling at population and single-cell levels with genome/epigenome engineering *in vivo*, we identify multiple regulatory layers governing NC ontogeny, including NC-specific enhancers and super-enhancers, novel *trans*-factors and *cis*-signatures allowing to reverse engineer the NC-GRN at unprecedented resolution. Furthermore, identification and dissection of divergent upstream combinatorial regulatory codes has afforded new insights into opposing gene circuits that define canonical and neural NC fates early during NC ontogeny. Our integrated approach, based on the global epigenomic and transcriptional analyses of specific embryonic cell types at the single-cell and population level allows dissection of cell-type-specific regulatory circuits in vivo, with broad implications for GRN discovery and investigation.

OPTIMIZATION OF A WHOLE GENOME AMPLIFICATION METHOD FOR SINGLE *PLASMODIUM FALCIPARUM* PARASITE SEQUENCING

Shiwei Liu[1], Adam Huckaby[1], Michael J McConnell[2], Jennifer Guler[1,3]

[1]University of Virginia, Department of Biology, Charlottesville, VA, [2]University of Virginia Health System, Department of Biochemistry and Molecular Genetics, Charlottesville, VA, [3]University of Virginia Health System, Division of Infectious Diseases and International Health, Charlottesville, VA

Malaria is caused by protozoan *Plasmodium* parasites. Antimalarial resistance is one of the greatest threats to malarial control. Genetic causes of antimalarial resistance (i.e. SNPs or copy number variations) are commonly detected by DNA sequencing of pooled populations of parasites. This approach lacks the sensitivity to reveal heterogeneity, particularly in clinical samples that may harbor minor genotypes from multiple infections. Single cell genomics can detect these genetic variants, but it requires a whole genome amplification (WGA) step to generate enough material for sequencing. WGA methods often introduce bias, which prevents identification of copy number variants in particular. This bias is exacerbated when studying *Plasmodium falciparum* due to its small genome (23Mb), high AT content (80.6%), and contaminating DNA from host's blood. WGA using multiple annealing and looping-based amplification cycles (MALBAC) has been reported to exhibit excellent reproducibility and higher accuracy for copy number variation detection over other WGA methods. Here, we evaluated and optimized MALBAC for single laboratory-derived *P. falciparum* genomes. We show that our optimized MALBAC can amplify early stage parasite genomes (1n, ~30fg) from single infected erythrocytes. The optimization of MALBAC reduces the proportion of contaminating DNA (from 91% to an average of 9%) and improves the breadth of genome coverage (from 25% to an average of 61%). Moreover, the amplification pattern of optimized MALBAC is consistent across single parasite samples. This finding suggests that computational modeling and normalization can be used to remove amplification bias and ultimately, conduct copy number analysis. Finally, we tested this approach on parasite DNA from clinical patients. Despite high levels of human DNA contamination, clinical material is suitable for sequencing (contamination proportion: ~63%, breadth of coverage: ~51%). Standard MALBAC preferentially amplifies GC-rich sequences and exhibits picogram level sensitivity; our optimized MALBAC efficiently amplifies low-GC sequences with femtogram level sensitivity, allowing selective amplification of parasite DNA over contaminating DNA. The reproducible nature of MALBAC will facilitate the study of copy number variation at a single parasite level for the first time, thus improving tools to study the mechanisms of genetic adaptation in the deadly parasite.

MAPPING VECTOR FIELD OF SINGLE CELLS

Xiaojie Qiu[1,2], Yan Zhang[3,4], Dian Yang[1,2], Shayan Hosseinzadeh[5], Li Wang[6], Ruoshi Yuan[7], Song Xu[8], Yian Ma[9,10], Joseph Replogle[1,2], Spyros Darmanis[5], Jianhua Xing[3,4,11,12], Jonathan S Weissman[1,2]

[1]UCSF, Cellular Molecular Pharmacology, San Francisco, CA, [2]Howard Hughes Medical Institute, UCSF, San Francisco, CA, [3]University of Pittsburgh, Department of Computational and Systems Biology, Pittsburgh, PA, [4]University of Pittsburgh, Joint CMU-Pitt Ph.D. Program in Computational Biology, Pittsburgh, PA, [5]Chan Zuckerberg Biohub, Chan Zuckerberg Biohub, San Francisco, CA, [6]University of Texas at Arlington, Department of Mathematics, Arlington, TX, [7]Harvard Medical School, Department of Systems Biology, Boston, MA, [8]Stanford University, Center for Biomedical Informatics Research, Department of Medicine, Stanford, CA, [9]University of California, Berkeley, Department of Electrical Engineering and Computer Sciences, Berkeley, CA, [10]University of California, San Francisco, Medical Scientist Training Program, San Francisco, CA, [11]University of Pittsburgh, UPMC-Hillman Cancer Center, Pittsburgh, PA, [12]University of Pittsburgh, Department of Physics, Pittsburgh, PA

Understanding how gene expression in single cells progresses over time is vital for revealing the mechanisms governing cell fate transitions. RNA velocity, which infers immediate changes in gene expression by comparing levels of new (unspliced) versus mature (spliced) transcripts (La Manno et al. 2018), represents an important advance to these efforts. A key question remaining is whether it is possible to predict the most probable cell state backward or forward over arbitrary time-scales. To this end, we introduce an inclusive model (termed Dynamo) capable of predicting cell states over extended time periods, that incorporates promoter state switching, transcription, splicing, translation and RNA/protein degradation by taking advantage of scRNA-seq and the co-assay of transcriptome and proteome. We also implement scSLAM-seq by extending SLAM-seq to plate-based scRNA-seq (Hendriks et al. 2018; Erhard et al. 2019; Cao, Zhou, et al. 2019) and augment the model by explicitly incorporating the metabolic labelling of nascent RNA. We show that through careful design of labelling experiments and an efficient mathematical framework, the entire kinetic behavior of a cell from this model can be robustly and accurately inferred. Aided by the improved framework, we show that it is possible to reconstruct the transcriptomic vector field from sparse and noisy vector samples generated by single cell experiments. The reconstructed vector field further enables global mapping of potential landscapes that reflect the relative stability of a given cell state, and the minimal transition time and most probable paths between any cell state in the state space. This work thus foreshadows the possibility of predicting long-term trajectories of cells during a dynamic process instead of short time velocity estimates. Our methods are implemented as an open-source tool, dynamo (https://github.com/aristoteleo/dynamo-release).

FRIEND OR FOE: MISSING DATA IMPUTATION IN SINGLE CELL ANALYSIS

Jungmin Han, Vipul Periwal

NIH/NIDDK, Laboratory of Biological Modeling, Bethesda, MD

With advances in single cell techniques, collecting a large quantity of data has become more accessible and efficient. However, due to experimental limitations, some missing data is still inevitable. One of the long-standing conundrums in the field is how to account for these missing data. A common solution for the missing data is to impute them using the information gathered from the measured data. In this study, we present a case where even a high performance imputation method utilizing neural networks may not be an appropriate data pre-processor.

Fowlkes et al. (*Cell*, 2008) published a set of gene expressions measured from 6078 *Drosophila* blastoderm during 6 different time cohorts that spanned the 50 minutes prior to the onset of gastrulation. Out of 95 genes and 4 proteins, only 27 of them had complete temporal information from all the cells, while the rest of them were measured only in a subset of cells. The missing data constituted about 37% of the whole data set. To impute the missing data, we trained and tested neural networks with one hidden layer on the complete 27 genes as predictors and the genes that were measured in only in subsets of cells as targets. Once we confirmed that the neural networks were well trained, the system was used to predict the missing gene expressions.

We then used a statistical inference method to infer two dynamic gene networks from two different sets of the gene data. One set had all the gene expression levels, including the imputed missing data, and the other set was only for those 27 genes with complete temporal information. We investigated which of the inferred networks is better at describing the dynamics of the 27 genes with the complete temporal profile. As a result, the gene dynamics predicted from the gene network inferred from the data set with the imputed data had a larger deviation from the true gene dynamics, compared to the prediction made from the gene network with the 27 genes. This suggests that, although missing data imputation is a commonly used data pre-processor, it may not be advantageous for inferring a gene network to forecast the dynamics of gene evolution.

UNCOVERING T CELL DYNAMICS OVER THE COURSE OF CHIMERIC ANTIGEN RECEPTOR (CAR) T CELL THERAPY

Vanessa D Jonsson[1], Rachel Ng[1], Natalie Dullerud[2], Robyn Wong[1], Christine E Brown[1]

[1]City of Hope Beckman Research Institute, Hematology, Duarte, CA, [2]University of Southern California, Mathematics, Los Angeles, CA

At present, a major clinical effort aims to assess how immunotherapies can modulate the strength, timing, and evolution of an individual patient's anticancer immune response. Tumor immune cell composition – specifically T cell quantity, functionality and T cell receptor (TCR) diversity – plays a major role in response to immunotherapy. However, recent studies have shown that reactivity of tumor infiltrating lymphocytes (TILs) with respect to tumor antigens is rare and variable; therefore, the characterization of T cell dynamics in peripheral compartments is critical for deeper understanding of T cell mediated immunity during immunotherapy. Here, we collected and performed combined analysis of longitudinal, single cell RNA and TCR sequencing of tumor, peripheral blood (PB) and cerebrospinal fluid (CSF) samples in a patient with recurrent multifocal glioblastoma that received chimeric antigen receptor (CAR) T cells targeting the tumor-associated antigen interleukin-13 receptor alpha 2 (IL13Rα2). This patient's disease course – a remarkable regression of all tumors for 7.5 months followed by a recurrence – provided the opportunity to characterize the longitudinal dynamics of T cell mediated immunity during response and recurrence in a clinical setting.
Single cell analysis from pre-treated and recurrent tumors revealed significant T cell functional and clonal changes; these evolved from a large, cytotoxic and clonally expanded T cell population to a sparse T cell population with evenly distributed TCR repertoire– suggesting an active, pre-existing endogenous T cell response that disappeared in the recurrent setting. To further explore the dynamics of T cell response, we characterized T cell clonality over blood and cerebrospinal fluid samples collected longitudinally during the treatment course (20 timepoints over 228 days, 16931 clonotypes). The application of both density and shape-based clustering on timeseries data revealed three equivalence classes of prominent TCR signals that emerged in the CSF: these patterns consisted of contracting, persistent and transient T cell dynamics. Transient T cell dynamics were characterized by high TCR frequency for no more than one or two timepoints and were composed mostly of CAR T and endogenous T cells from peripheral blood; these identified both the transience of CAR T cell infusions in the CSF, and a prominent recruitment of endogenous T cells from the PB into the CSF. Strikingly, the identified contracting T cell subset, was not only enriched for a diverse TCR repertoire, but also decreased concomitantly with tumor volume, suggesting the existence of an immuno-editing T cell population that combined with CAR T cell therapy to produce an unprecedented complete response.

FROM SINGLE-CELL PERTURBATION TO SUBCELLULAR ANALYSIS USING FLUIDIC FORCE MICROSCOPY

Julia A Vorholt

ETH Zurich, Department of Biology, Zurich, Switzerland

Single-cell biology provides unique insights into intracellular processes, cell-cell communication, and cell responses to exogenous stimuli, otherwise inaccessible to traditional bulk studies. While micromanipulation or cell sorting followed by cell lysis is already used for subsequent molecular examinations, approaches to directly extract the content of living cells remain a challenging but promising alternative to achieving non-destructive sampling and cell-context preservation. We developed a novel strategy to selectively perturb, isolate, or analyse single cells within physiological environments using fluidic force microscopy (FluidFM). The technology combines atomic force microscopy with microfluidics via microchanneled cantilevers with nano-sized apertures. Pressure control through the microchannel enables liquid release and aspiration with femtolitre resolution, whereas force spectroscopy allows for non-destructive insertion of the probe pyramidal tip inside a cell with real-time monitoring of cell indentation and membrane perforation. We demonstrate the quantitative extraction from single cells with spatiotemporal control. Molecular analyses of the withdrawn extracts permit the detection of enzyme activities, transcripts and metabolites. The established approaches provide unprecedented means to quantitatively inject or extract soluble molecules into/from single cells while preserving cell viability and physiological context.

CITE-seq IDENTIFIES UNIQUE MYELOID DERIVED SUPPRESSOR CELLS (MDSCS) POPULATION SIGNATURES IN GLIOBLASTOMA PATIENTS THAT DO NOT RESPOND TO IMMUNE CHECKPOINT INHIBITION

Tyler J Alban[1,4], Matthew Grabowski[1,3], Balint Otvos[1,3], Pat Rayman[2], C. Marcela Diaz-Montero[2], Manmeet S Ahluwalia[1,3,4], Justin D Lathia[1,3,4]

[1]Cleveland Clinic, CVMS, Cleveland, OH, [2]Cleveland Clinic, Inflammation and Immunity, Cleveland, OH, [3]Cleveland Clinic, Rose Ella Burkhardt Brain Tumor Center, Cleveland, OH, [4]Case Western, Comprehensive Cancer Center, Cleveland, OH

Glioblastoma (GBM) creates an immunosuppressive environment that presents a challenge to efficacy of immunotherapeutic approaches aimed at increasing T cell activation by immune checkpoint inhibition. Results from the CheckMate-143 trial demonstrated responses in 8% of patients treated with nivolumab, underscoring the need for further insight into the mechanisms and markers of immune suppression and response. Given a limited set of biomarkers predictive of immunotherapy response in GBM, we explored the changes in immune cell populations in GBM patients' pre and post-treatment in order to help predict response. We utilized traditional and newly developed approaches, including mass cytometry time-of-flight (CyTOF) and 10X Genomics simultaneous cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq). We analyzed patient's samples in a randomized, phase 2 study of nivolumab and bevacizumab at GBM first recurrence (NCT03452579). Greater than 16 patients were clinically identified as responders or non-responders at 8 weeks after therapy initiation with 8 responders and 8 non-responders utilized in the first round of CITE-seq. Peripheral blood samples of patients were analyzed pre and post treatment using CITE-seq, which identified differential gene expression in myeloid derived suppressor cells (MDSCs) of responders and non-responders. Additionally, preliminary data shows that responders had increased IL7R-positive T cells post-treatment, which was not observed in non-responders. These results are currently being validated by flow cytometry and an additional cohort of CITE-seq is being performed on a control cohort of bevacizumab only treated patients. Taken together, differences in immune-phenotypes that were specific to responders and non-responders were observed, and characterization of these immune populations may be helpful in identifying GBM patients likely to benefit from immunotherapy.

SINGLE CELL RNA-seq REVEALS DEVELOPMENTAL PLASTICITY WITH COEXISTING ONCOGENIC STATES AND IMMUNE EVASION PROGRAMS IN ETP T-ALL

Praveen Anand*[1,2,3], Amy Guillaumet-Adkins*[1,3], Huiyoung Yun*[1,3], Yotam Drier*[2,4,5], Anna Rogers[1], Madhu M Ouseph[6], Monica Nair, Sayalee Potdar[1], Jake A Kloeber[1], Tushara Vijaykumar, Randi Isenhart[1], Julia Frede, Guangwu Guo[2,3], Marian Harris[7], Lewis Silverman[1,8], Andrew A Lane[2,3,9], Daniel DeAngelo[2,9], Jon C Aster[6,9], Bradley E Bernstein[3,4,9], Jens G Lohr[2,3,9], Birgit Knoechel[1,3,8,9] *Authors contributed equally in the abstract body.

[1]Dana-Farber Cancer Institute, Department of Pediatric Oncology, Boston, MA, [2]Dana-Farber Cancer Institute, Department of Medical Oncology, Boston, MA, [3]Broad Institute of MIT and Harvard, Boston, MA, [4]Massachusetts General Hospital, Department of Pathology, Boston, MA, [5]The Hebrew University of Jerusalem, Faculty of Medicine, Jerusalem, Israel, [6]Brigham and Women's Hospital, Department of Pathology, Boston, MA, [7]Boston Children's Hospital, Department of Pathology, Boston, MA, [8]Boston Children's Hospital, Division of Hematology/Oncology, Boston, MA, [9]Harvard Medical School, Boston, MA

Early T-cell precursor acute lymphoblastic leukemia (ETP T-ALL) is a distinct subtype of T-ALL characterized by higher rates of relapse and induction failure. Although large-scale genetic sequencing studies have identified frequently mutated oncogenes and gene fusions, the role of functional heterogeneity and microenvironment towards treatment failure remain unknown. We performed full-length single-cell RNA-sequencing of 5,077 malignant and normal immune cells from bone marrow or blood from five patients with relapsed/refractory ETP T-ALL (based on immunophenotyping, all with NOTCH1 activating mutations), before and after targeted therapy against NOTCH1. These patients were enrolled on a phase I trial with the γ-secretase inhibitor (GSI) BMS-906024 (NCT01363817). The transcriptome analyses revealed a deranged developmental hierarchy characterized by co-expression of multiple progenitor programs in malignant cells implying ineffectual commitment to either lymphoid or myeloid lineage. The leukemic cells subjected to GSI treatment revealed down regulation of Notch signaling but simultaneously demonstrated high PI3K activity. This population of high PI3K activity preexisted along with high Notch activity cells in the untreated population and thus could explain the poor response of patients towards GSI treatment and emergence of resistance. Analysis of the immune microenvironment revealed an oligoclonal T-cell population in ETP T-ALL compared to normal donor T-cells. CD8+ T-cells from ETP patients expressed markers of T-cell exhaustion (PDCD1, TIGIT, LAG3, HAVCR2). Analyses of expression levels of the respective ligands on leukemic blasts and the predicted interaction with their receptors on endogenous CD8 T-cells demonstrated the highest interaction score between HAVCR2 and its ligand LGALS9. LGALS9 was universally expressed in all leukemic cells, which was confirmed by flow cytometry staining in leukemic blasts and IHC staining in bone marrow of 8 patients with ETP T-ALL and 7 patients with T-ALL. We confirmed LGALS9 mediated exhaustion *in vitro* using activated CD8+ T-cell and DND-41 cell line. This demonstrates a possible role for HAVCR2-LGALS9 interaction in causing CD8+ T-cell dysfunction in ETP T-ALL patients, which may provide a novel therapeutic strategy.

INCORPORATION OF SPATIAL MAPPING BY A MULTIPLEX *IN SITU* HYBRIDIZATION TECHNOLOGY INTO SINGLE CELL RNA SEQUENCING WORKFLOWS

Jyoti Phatak, Han Lu, Li Wang, Hailing Zong, Morgane Rouault, Xiao-Jun Ma, Courtney M Anderson

Advanced Cell Diagnostics/Bio-Techne, R&D, Newark, CA

Complex and highly heterogenous tissues such as the brain are comprised of multiple cell types and states with exquisite spatial organization. Single-cell RNA sequencing (scRNA-seq) is now being widely used as a universal tool for classifying and characterizing known and novel cell populations within these heterogenous tissues, ushering in a new era of single cell biology. However, the use of scRNA-seq presents some limitations due to the use of dissociated cells which results in the loss of spatial context of the cell populations being analyzed. Incorporating a multiplexed spatial approach that can interrogate gene expression with single cell resolution in the tissue context is a powerful addition to the scRNA-seq workflow. In this study, we used the RNAscope Multiplex Fluorescent and RNAscope HiPlex *in situ* hybridization (ISH) assays to confirm and spatially map the diverse striatal neurons that have been previously identified by scRNA-seq in the mouse brain (Gokce et al, Cell Rep, 16(4):1126-1137, 2016). We confirmed the gene signatures of two discrete D1 and D2 subtypes of medium spiny neurons (MSN): *Drd1a/Foxp1, Drd1a/Pcdh8, Drd2/Htr7*, and *Drd2/Synpr*. The heterogenous MSN subpopulations were marked by a transcriptional gradient, which we could spatially resolve with RNA ISH. Numerous striatal non-neuronal cell populations identified by scRNA-seq, including vascular cells, immune cells, and oligodendrocytes, were also confirmed with the multiplex ISH assay. Finally, the spatial relationship between the D1 and D2 MSN subtypes identified by Gokce et al. was visualized using the RNAscope HiPlex assay, which allows for detection of up to 12 RNA targets simultaneously in intact tissues. In conclusion, we have demonstrated the utility of two multiplexed RNAscope ISH assays for the confirmation and spatial mapping of scRNA-seq transcriptomic results in the highly complex and heterogenous mouse striatum at the single cell level. Incorporating spatial mapping by the RNAscope technology into single cell transcriptomic workflows complements scRNA-seq results and provides additional biological insights into the cellular organization and functional states of diverse cell types in healthy and disease tissues.

SENSITIVITY OF GENE FUSION DETECTION FROM HIGH-THROUGHPUT SMART-seq CHEMISTRY ON THE ICELL8 CX SYSTEM

Rachel Fish[1], Michael J Young[1], Lyndon Liu[1], Sherry Wei[1], Karthikeyan Swaminathan[1], Magnolia Bostick[1], Nidhanjali Bansal[2], Andrew Farmer[1]

[1]Takara Bio USA, R&D, Mountain View, CA, [2]Takara Bio USA, Marketing, Mountain View, CA

The use of next-generation sequencing for transcriptome analysis in clinical and applied spaces requires accurate, parallel processing of large numbers of samples and the availability of chemistries that enable robust library preparation from the desired targets. The method chosen for library preparation – for example, end-capture versus full-length transcript capture – will yield different quality data to aid in answering specific questions about gene expression. Full-length capture, which provides more uniform coverage of the transcript, enables more thorough examination of gene fusions, SNP detection, and alternative splicing, over other methods that capture only the 3' or 5' end of the transcript. The application of our full-length SMART-seq chemistry on the ICELL8® cx system provides a high-throughput solution to obtaining richer data on single-cell transcriptomics. We identified fusions in a solid tumor cell line as well as in a leukemia cell line using this application that could not be identified with a 3'DE approach, with fusions discovered in more than three times as many single cells. This method can also be used to obtain a full understanding of the immune response in single immune cells by combining cell type identification with paired T-cell receptor and B-cell receptor information – all obtained using different analysis of a single whole transcriptome assay. This sensitive SMART-Seq method for automated, full-length RNA-seq offers benefits in increased gene body coverage, enabling improved detection of fusions, SNPs, and splice variants—applications that will aid in the advancement of biomarker identification and the development of novel therapeutics.

HIGHLY ACCURATE SMALL-RNA SEQUENCING OF SINGLE CELLS (REALseq-SC)

Sergio Barberan-Soler, Quincy Okobi, Van Lau

RealSeq Biosciences, Sequencing, Santa Cruz, CA

Single cell analyses of mRNA have allowed the identification of crucial differences between cells that were otherwise considered identical. These findings have shown that there is intrinsic “noise” in the regulation of gene expression that plays an important role in determining cell fates. Unfortunately, there is currently a lack of information about the cell-to-cell variability of levels of small RNAs, including microRNAs. Indeed, there is no commercially available library preparation kit for small RNAs that can profile single cells. We are currently developing our library preparation technology, RealSeq®-AC, to be able to quantify small RNAs from single cells. RealSeq®-AC uses a scheme involving a single combo-adapter and circularization that accurately quantifies over 75% of all miRNAs detected, compared to ~35% from the best competitor kit. To adapt this technology for single-cell sequencing we use three separate strategies to dramatically reduce the presence of adapter-dimers in the library. Our protocol performs all steps from cell lysis to final purification of amplified libraries in a single tube. Preliminary data shows that this technology allows the reliable quantification of miRNAs and other small RNAs from MCF7 single-cells. RealSeq-Single Cells allows the accurate quantification of small RNAs from single cells.

SINGLE-CELL RNA SEQUENCING IDENTIFIES FIBRO-INFLAMMATORY PERIVASCULAR STROMAL CELLS CONTROLLING ADIPOSE TISSUE INFLAMMATION IN OBESITY

Spencer D Barnes[1], Bo Shan[2], Mengle Shao[2], Qianbin Zhang[2], Chelsea Hepler[2], Gervaise H Henry[3], Vivian A Paschoal[2], Lavanya Vishvanath[2], Yu A An[2], Venkat S Malladi[1], Douglas W Strand[3], Dayoung Oh[2], Rana K Gupta[2]

[1]University of Texas Southwestern Medical Center, Lyda Hill Department of Bioinformatics, Dallas, TX, [2]University of Texas Southwestern Medical Center, Touchstone Diabetes Center, Department of Internal Medicine, Dallas, TX, [3]University of Texas Southwestern Medical Center, Department of Urology, Dallas, TX

White adipose tissue (WAT) remodeling is dictated by coordinated interactions between adipocytes and resident stromal-vascular cells; however, the functional heterogeneity of adipose stromal cells has remained unresolved. We recently combined single-cell RNA-sequencing and FACS to identify and isolate functionally distinct subpopulations of PDGFRbeta+ stromal cells within visceral WAT of adult mice. LY6C- CD9- PDGFRbeta+ cells represent highly adipogenic visceral adipocyte precursor cells (‘APCs’), whereas LY6C+ PDGFRbeta+ cells represent fibro inflammatory progenitors (‘FIPs’). FIPs lack adipogenic capacity, display pro-fibrogenic/pro-inflammatory phenotypes, and can exert an anti-adipogenic effect on APCs. We now present further analysis of these subpopulations identified by single-cell techniques and demonstrate that FIPs activate pro-inflammatory signaling cascades shortly after the onset of high-fat diet feeding of mice and control the accumulation of pro-inflammator macrophages in WAT. The activation of FIPs is mediated by the downregulation of ZFP423, identified here as a transcriptional co-regulator of NFkB. Biochemical analysis of ZFP423-protein complexes and ChIP-seq analysis reveal that ZFP423 suppresses the DNA-binding capacity of the p65 subunit of NFkB by inducing a co-regulator switch. Doxycycline-inducible expression of Zfp423 in PDGFRbeta+ cells suppresses inflammatory signaling in FIPs and attenuates macrophage accumulation within visceral WAT of obese mice. Conversely, inducible inactivation of Zfp423 in PDGFRbeta+ cells increases FIP activity, exacerbates adipose macrophage accrual, and promotes WAT dysfunction in obese mice. These studies unveil the functional heterogeneity of the adipose tissue stromal and implicate perivascular PDGFRbeta+ cells as sentinels and gatekeepers of adipose tissue inflammation in obesity.

THE NIH COMMON FUND, HUMAN BIOMOLECULAR ATLAS PROGRAM (HUBMAP): BUILDING A FRAMEWORK FOR MAPPING THE HUMAN BODY

Tyler K Best, Dena C Procaccini, Richard Conroy

National Institutes of Health, Office of Strategic Coordination, Rockville, MD

Knowing how tissue organization influences a cell's molecular state, interactions, and history is critical for enhancing our awareness of variation in organ function across the lifespan and health-disease continuum. Despite vastly improved imaging and omics technologies as well as many important foundational discoveries, our understanding of how tissues are organized is restricted to a very limited number of microscopic structures. Better insights into the principles governing organization-function relationships will potentially give researchers a better sense of the significance of inter-individual variability, changes across the lifespan, tissue engineering, and the emergence of disease at the biomolecular level. However, integrating imaging and omics analysis to comprehensively profile biomolecular distribution and morphology of tissues in a high throughput manner and placing this information into 3D tissue maps amenable to modelling and molecular perturbation has yet to be fully realized.

The goal for HuBMAP is to catalyze the development of a comprehensive atlas of cellular organization in human tissues that will elucidate the principles of organization-function by:
♣ Accelerating development of tools for constructing comprehensive spatial tissue maps and integrating data types,
♣ Building and generating tissue maps from validated high-content, high-throughput imaging and omics assays,
♣ Coordinating and collaborating with other funding agencies, programs and the biomedical research community,
♣ Rapidly making data findable, accessible, interoperable, and reusable in standardized formats.

HuBMAP has been designed as a collaborative consortium, having a culture of openness and sharing using team science-based approaches. HuBMAP is committed to sharing protocols, analytical pipelines, quickly making data available and FAIR: Findable, Accessible, Interoperable, and Re-usable.

HuBMAP is organized into five integrated focus areas:
1. Transformative Technology Development
2. Rapid Technology Integration
3. Tissue Mapping Centers
4. The HIVE (HuBMAP Integration, Visualization & Engagement)
5. Demonstration Projects

PROTOCELL: AUTOMATED IDENTIFICATION OF CELL TYPES IN SINGLE-CELL RNA-seq DATA USING PROTOTYPICAL NEURAL NETWORKS

Ash Blibaum, Alexander Dobin

Cold Spring Harbor Laboratory, Genomics, Cold Spring Harbor, NY

Single-cell RNA-seq technologies have become invaluable for studying the molecular biology of individual cells. One of the most crucial components of scRNA-seq data analysis is the classification of each cell into its characteristic type. Typically, this is done via unsupervised clustering followed by manual curation of marker genes, which are used to assign cell type based on existing literature and prior knowledge of the researcher. However, this approach is time consuming, has poor reproducibility due to a lack of canonical markers for many cell types, and is influenced by unwanted sources of variation in clustering such as batch effects. There now exist several tools for automated identification of cell types via machine learning models which address these limitations; however, these methods fail in realistic scenarios with limited data for some cell types, such as is common in studies of immune response. In addition, these models have to be pre-trained on all cell types of interest, requiring a researcher to re-train the model if they wish to accommodate new cell types, necessitating non-trivial amounts of computational resources and machine learning expertise.

Here we present ProtoCell, a model that resolves the aforementioned limitations using the *prototypical network* model from the domain of few-shot deep learning. (Snell et. al, NIPS 2017) The model learns a non-linear mapping of the full gene expression vector into a low-dimensional embedding space using a neural network, and takes the *prototype* for each cell type to be the centroid of its vectors in the embedding space. Classification is then performed for a query cell by computing its embedding and finding the nearest euclidean distance to a class prototype. ProtoCell has equivalent or better performance to existing models for the prediction of cell types with a large amount of training data, while significantly outperforming existing models for cell types with few training examples. ProtoCell retains good prediction accuracy (MCC>80%) on cell types with as few as a single training cell, enabling the prediction of rare cell types. Furthermore, ProtoCell can confidently predict novel cell types by thresholding the distance to the nearest prototype centroid, achieving AUROC>0.94 in our tests, significantly better than existing approaches. Since ProtoCell only learns the embedding space during pre-training and dynamically constructs a classifier for provided cell type examples without updating its parameters, it eliminates the need for computationally expensive model retraining to predict new cell types. ProtoCell is fast and easy to use, interpretable, and enables accurate cell type discovery for challenging single-cell datasets. ProtoCell is open source and available at https://ashblib.github.io/protocell/.

MERCURY: HIGH QUALITY VISUALIZATION AND REPRODUCIBLE EXPLORATORY ANALYSIS OF scRNA-seq DATA

Ash Blibaum, Alexander Dobin

Cold Spring Harbor Laboratory, Genomics, Cold Spring Harbor, NY

Single cell RNA-seq (scRNA-seq) has recently become an extremely popular methodology for the analysis of biological systems, providing transcriptome profiles at single cell resolution. However, scRNA-seq data is challenging to analyse due to its high dimensionality and sparsity, making visual and supervised analysis of the data crucially important. This typically involves visualization of cell clustering, manual editing and definition of cell clusters/groups, and selection of cell groups for differential expression analysis. Though hundreds of tools exist for statistical analysis of scRNA-seq data, very few interactive processing and visualization solutions are available, and all have limitations in both their functionality and workflow. Here we present Mercury, a tool for high quality visualization and reproducible exploratory analysis of scRNA-seq data, that addresses the shortcomings of current visualization software and unifies the scRNA-seq analysis pipelines.

Mercury provides flexible and aesthetically pleasing 2D and 3D visualization of scRNA-seq data that's fully integrated with the popular Seurat and Scanpy statistical analysis libraries. It uses Unreal Engine 4, a modern game development engine, for ultra-high quality, performant rendering and a sophisticated user experience. Mercury's close integration with analysis tools allows the user to define a sequence of data processing steps, change their parameters, reprocess the data using Seurat or Scanpy as a backend and update the visualization, all in an intuitive user interface and without leaving the app. Mercury speeds up data exploration by providing a streamlined user experience and workflow designed specifically for the core analysis tasks of cell cluster creation/labeling and differential gene expression analysis. In addition to core workflow improvements, Mercury provides important analysis functions beyond most existing tools, including high quality 3D visualization, integration of multiple datasets, and automated cell type identification. Mercury also enhances reproducibility by tracking user actions, which can allow another researcher, or the user themselves, to replay the steps they took in their analysis. Mercury is fully open source and uses well documented file formats for interoperability between analysis pipelines. Mercury closes the gap in the current scRNA-seq analysis pipeline, providing a sophisticated, usable, and open source tool for reproducible visual scRNA-seq analysis. Mercury is available at https://dobinlab.org/mercury.

UNPRECEDENTED SENSITIVITY WITH SMART-seq SINGLE CELL TECHNOLOGY

Nathalie Bolduc[1], Tommy Duong[1], Magnolia Bostick[1], Nidhanjali Bansal[2], Andrew Farmer[1]

[1]Takara Bio USA, Inc., Research and Development, Mountain View, CA, [2]Takara Bio USA, Inc., Marketing, Mountain View, CA

Since the emergence of next-generation sequencing (NGS), the importance and demand for single-cell analysis have risen rapidly. Extracting meaningful biological information from the small amount of mRNA present in each cell requires an RNA-seq preparation method with exceptional sensitivity and reproducibility. To date, the SMART-Seq v4 chemistry has been the most sensitive commercial single-cell RNA-seq method, in part due to its incomparable capability to retrieve information from full-length mRNA and not just the 3' end. However, there is still room for improvement for extremely challenging samples such as cells with very low RNA content or nuclei. To address this need, we have further modified our core technology to create a new chemistry with higher sensitivity, SMART-Seq Single Cell, that outperforms all current commercial and non-commercial full-length methods, particularly with as little as 2 pg of total RNA. When validating with a B lymphocyte cell line or peripheral blood mononuclear cells from a healthy donor, we were able to detect 50-60% more genes with the new chemistry compared to current methods. The improvement in sensitivity was associated with a clear reduction of the dropout rate as well as increase in reproducibility. In addition, the new SMART-Seq Single Cell chemistry generates a high yield of cDNA, which is extremely useful when dealing with difficult cells such as clinical samples that tend to carry very low RNA content.

IDENTIFICATION OF TRANSCRIPTIONAL PROFILES OF THE MS4A GENE CLUSTER AND THEIR ASSOCIATION WITH ALZHEIMER'S DISEASE

Logan Brase, Jorge del-Aguila, Bruno Benitez, Celeste Karch, Carlos Cruchaga, Oscar Harari

Washington University in St. Louis, Psychiatry, St. Louis, MO

Alzheimer's disease (AD) is a complex and heterogenous trait in which multiple molecular pathways are disrupted in different cell-types culminating in disease. Genetic variants in TREM2, a gene expressed in microglia, have been associated with high-risk with AD, and similarly soluble TREM2 (sTREM2) levels in the cerebrospinal fluid (CSF) have been associated with AD. In a previous Genome-wide association study (GWAS) of CSF sTREM2 levels, a genetic variant near the MS4A gene cluster accounted for 6 percent of the protein's variance (rs1582763 p= $1.15\times10-15$). This finding suggests that TREM2 is involved in AD risk in the general population, beyond TREM2 risk-variant carriers, as the minor allele frequency of rs1582763 > 0.30. rs1582763 is also a cis-eQTL for both MS4A4A and MS4A6A genes in bulk RNA-seq of blood. Identifying the transcriptional effects of this variants in myeloid cells, and their overall impact in the brain, will help improve our understanding of AD biology. We generated unsorted single-nuclei RNA-seq from human brain tissue from the Knight-ADRC brain bank. We collected the parietal lobe for 21 AD brains (13 are carriers of the rs1582763 minor allele) and 12 neuropath-free (9 carriers) for a total of 33 brains. Using the 10X Chromium 3' chemistry v3 we generated high quality data for 284,109 nuclei to identify the transcriptomic effects of this variant at a cell specific resolution. We identified 10,889 microglial nuclei to analyze microglial subtypes, activation states, co-expression networks and eQTLs. We are applying several methods for analyzing these nuceli: DEsingle, which is specifically developed to analyze single-cell RNA-seq and alternative methodologies to identify differentially expressed genes, including linear models to test for gene expression modeled as totalized pseudo-counts; and mixed models to account for subject of origin. This effort is designed to test over 13,000 genes to help uncover the role of the MS4A cluster and its connection with TREM2 levels in microglia in human brains.

LATTICE LIGHT-SHEET MICROSCOPY MULTI-DIMENSIONAL ANALYSES (LAMDA) OF T-CELL RECEPTOR DYNAMICS PREDICT T-CELL SIGNALING STATES

Jillian Rosenberg*[1], Guoshuai Cao*[2], Fernanda Borja-Prieto[2], Jun Huang[2]

[1]University of Chicago, Committee on Cancer Biology, Chicago, IL,
[2]University of Chicago, Pritzker School of Molecular Engineering, Chicago, IL

*Authors contributing equally

The methods of collecting, analyzing, and understanding the dynamics of cell surface receptors with single-cell imaging currently lag behind those for other high-dimensional analyses, such as single-cell omics assays. Here we report the development of lattice light-sheet microscopy multi-dimensional analyses (LaMDA), a pipeline that combines high spatiotemporal-resolution four-dimensional lattice light-sheet microscopy, machine learning, and diffusion maps to analyze T-cell receptor (TCR) dynamics and predict T-cell signaling states without the need for complex biochemical measurements. LaMDA images thousands of TCR microclusters on the surface of live primary cells to collect high-dimensional dynamic data for machine learning, which extracts key dynamic features to build predictive diffusion maps. LaMDA spatiotemporally reveals global changes of TCRs across the 3D cell surface, accurately differentiates stimulated cells from unstimulated cells, precisely predicts attenuated T-cell signaling after CD4 and CD28 receptor blockades, and reliably discriminates between structurally similar TCR ligands. We anticipate broad usage of this approach for other receptors and cells, as well as for guiding the design and development of future immunotherapies for cancer, infection, and autoimmunity.

CONSTRUCTING A COMPREHENSIVE TRANSCRIPTOMIC ATLAS OF DEVELOPING FOLLICULAR EPITHELIAL CELLS IN ADULT *DROSOPHILA* OVARY

Deeptiman Chatterjee[1], Allison M Jevitt[2], Gengqiang Xie[2], Xian-Feng Wang[1], Yi-Chun Huang[1], Taylor Otwell[2], Wu-Min Deng[1,2]

[1]Tulane University, Department of Biochemistry and Molecular Biology, New Orleans, LA, [2]Florida State University, Department of Biological Science, Tallahassee, FL

Oogenesis, the biological process for the egg to develop and mature, is a highly regulated process that involves coordinated action between germline and supporting somatic cells. To gain more insight into how this complex process is regulated, we generated a single-cell transcriptome atlas of the adult *Drosophila* ovary and surrounding tissues. This dataset allows us to track the developmental trajectory of the supporting somatic cells known as the follicular epithelial cells. As follicle cells develop through different stages of oogenesis, clear single-cell transcriptional programs are distinguishable temporally and spatially as the cells undergo differentiation, cell-cycle switching, morphogenesis and termination. This comprehensive resource also provides new markers for different cell types in the ovary and surrounding tissues, lending to the study of inter-cellular and inter-tissue signaling which coordinates oogenesis and ovulation in *Drosophila*. Finally, in this study, we have provided a framework to filter potentially contaminating cells by expected gene markers, identify known cell types and expected transcription states in developing tissues, and also perform in-depth clustering and pseudotemporal analyses of the continually evolving transcriptome during specific stages of development.

SIMULTANEOUS HIGH-THROUGHPUT SINGLE-CELL SEQUENCING OF DNA METHYLATION, CHROMATIN ACCESSIBILITY AND mRNA FROM THE SAME CELL ENABLES MARKER-FREE GENERATION OF CELL TYPE-SPECIFIC EPIGENETIC LANDSCAPES

Alex J Chialastri[1,2], Chatarin Wangsanuwat[1,2], Monte J Radeke[3], Siddharth S Dey[1,2]

[1]University of California, Santa Barbara, Chemical Engineering, Santa Barbara, CA, [2]University of California, Santa Barbara, Center for Bioengineering, Santa Barbara, CA, [3]University of California, Santa Barbara, Neuroscience Research Institute, Santa Barbara, CA

The regulatory role of epigenetic marks and features on transcription for many rare cell types is unknown. This is because isolation of these cell types at high purity either require creating complex transgenic animals or the use of cell type-specific antibodies that are frequently unavailable. Current methods allow the parallel detection of some epigenetic marks and features together with the transcriptome from the same cell, but these techniques are limited in throughput. Here, we describe scMAT-seq (single-cell **M**ethylation, chromatin **A**ccessibility, and **T**ranscriptome sequencing), which allows for early pooling of single cells prior to amplification and does not require physical separation of nucleic acids decreasing the risk of dropouts and drastically increasing the throughput compared to existing single-cell multiomics methods. We validate scMAT-seq by applying it to human embryonic stem cells and *in vivo* rat retinas to quantify cell type-specific epigenetic landscapes.

DOUBLETDECON: DECONVOLUTING DOUBLETS FROM SINGLE-CELL RNA-SEQUENCING DATA

Erica A DePasquale[1,2], Daniel J Schnell[1,3], Pieter-Jan Van Camp[1,2], Inigo Valiente-Alandi[3], Burns C Blaxall[3,4], H. Leighton Grimes[4,5,6], Harinder Singh[7], Nathan Salomonis[1,2,4]

[1]Cincinnati Children's Hospital, Biomedical Informatics, Cincinnati, OH, [2]University of Cincinnati, Biomedical Informatics, Cincinnati, OH, [3]Cincinnati Children's Hospital, Heart Institute and Center for Translational Fibrosis Research, Cincinnati, OH, [4]University of Cincinnati, Pediatrics, Cincinnati, OH, [5]Cincinnati Children's Hospital, Immunobiology and Center for Systems Immunology, Cincinnati, OH, [6]Cincinnati Children's Hospital, Experimental Hematology and Cancer Biology, Cincinnati, OH, [7]University of Pittsburgh, Center for Systems Immunology, Pittsburgh, PA

Methods for single-cell RNA sequencing (scRNA-Seq) have greatly advanced in recent years. While droplet- and well-based methods have increased the capture frequency of cells for scRNA-Seq, these technologies readily produce technical artifacts, such as doublet cell captures. Doublets occurring between distinct cell types can appear as hybrid scRNA-Seq profiles, but do not have distinct transcriptomes from individual cell states. We introduce DoubletDecon, an approach that detects doublets with a combination of deconvolution analyses and the identification of unique cell-state gene expression. We demonstrate the ability of DoubletDecon to identify synthetic, mixed-species, genetic, and cell-hashing cell doublets from scRNA-Seq datasets of varying cellular complexity with a high sensitivity relative to alternative approaches. Importantly, this algorithm prevents the prediction of valid mixed-lineage and transitional cell states as doublets by considering their unique gene expression. DoubletDecon has an easy-to-use graphical user-interface and is compatible with diverse species and unsupervised population detection algorithms (https://github.com/EDePasquale/DoubletDecon).

INVESTIGATING CENTRAL DOGMA IN THE CONTEXT OF NUCLEAR ARCHITECTURE AND CELL CYCLE AT SINGLE-CELL RESOLUTION

Shivnarayan Dhuppar, Aprotim Mazumder

Tata Institute of Fundamental Research, Centre for Interdisciplinary Sciences, Hyderabad, India

Nuclear architecture is the way the genome is arranged within the nucleus inside a cell. Of late, it has emerged as one of the major factors regulating gene expression in mammalian cells. This has led to what is now called the Topological Model of Gene Regulation. According to this model the timely expression and regulation of different genes in a cell is a consequence of specific arrangement of genes with respect to different nuclear compartments such as heterochromatin, interchromatin compartment or nuclear lamina [1]. The model, as seminal as it is interesting, still lacks a direct evidence the procurement of which relies on the ability to interrogate nuclear architecture and gene expression at the same time in the same 3-dimensionally intact cells. Here we present a method that we standardized to investigate central dogma of molecular biology in the context of nuclear architecture and the cell cycle. This was achieved by combining DNA FISH† with smFISH‡ and immunofluorescence in an elegant, simple and non-invasive way that preserves the 3D structure of cells. It was further combined with image-based cell cycle staging of cells [2] to quantify cell cycle-dependent changes in nuclear architecture and hence gene expression. We chose Cyclin A gene in this case to study how nuclear architecture can be correlated with the known cell cycle dependence of its expression.

___

† FISH: Fluorescence in situ Hybridization
‡ smFISH: Single molecule RNA Fluorescence in situ Hybridization

---

1. Cremer, T., & Cremer, C. (2001). Chromosome Territories, Nuclear Architecture and Gene Regulation in Mammalian Cells. Nature Reviews Genetics, 2(4), 292–301. doi:10.1038/35066075

2. Dhuppar, S., & Mazumder, A. (2018). Measuring cell cycle-dependent DNA damage responses and p53 regulation on a cell-by-cell basis from image analysis. Cell Cycle, 15384101.2018.1482136. doi:10.1080/15384101.2018.1482136

MOLECULAR DISSECTION OF CELLULAR HETEROGENEITY IN PANCREATIC ADENOCARCINOMA BY TRANSCRIPTOMIC PROFILING

Pierluigi Di Chiaro[1], Chiara Balestrieri[1], Stefania Brandini[2], Giuseppe R. Diaferia[2], Gioacchino Natoli[1,2]

[1]Humanitas University, Department of Biomedical Sciences, Milan, Italy, [2]European Institute of Oncology (IEO), Department of Experimental Oncology, Milan, Italy

BACKGROUND: Pancreatic ductal adenocarcinoma (PDAC) is a lethal disease with an extremely aggressive behavior due to both the advanced stage of the disease at diagnosis and the peculiar biology of this tumor. PDAC is characterized by nests of poorly differentiated cells (high-grade areas) coexisting with well- or moderately- differentiated glandular structures (low-grade areas) embedded in a highly fibrotic stroma that hinders canonical single cell analyses. This histological intratumor variability reflects distinct underlying gene-regulatory networks and transcriptional outputs and may contribute to the clinical properties of this tumor, including resistance to the therapeutic treatments. Therefore we aimed at obtaining a mechanistic understanding of cellular heterogeneity in human PDACs.
METHODS: Low- and high-grade tumor areas from individual human PDAC samples of a cohort of patients were morphologically identified by immunohistochemistry and immunofluorescence. Transcriptional profiling was assessed by RNA-seq of small groups of morphologically homogeneous cells captured by Laser Capture Microdissection (LCM). Multiplexed single molecule fluorescence in situ hybridization (smFISH) was performed for the detection and quantification of a selected panel of spatial distributed mRNA transcripts in their native environment with a single cell resolution.
RESULTS: We generated a robust protocol for maintaining RNA integrity during laser capture microdissection (LCM) of human cryo-sectioned tumor tissues and for library preparation of RNA sequencing. Our preliminary RNA-seq analysis showed separate transcriptional programs between low- and high-grade tumor areas.
To further investigate pancreatic cancer cell heterogeneity, we set up ad hoc multiplexed smFISH protocol showing high efficiency and accuracy in the detection of single molecules of mRNA targets in complex tissues like human PDAC.
CONCLUSIONS:
Our analysis provides an experimental approach to study transcriptional differences between low- and high-grade tumor areas in human PDAC. We expect that the integration of RNA-seq coupled with LCM with multiplexed smFISH will provide a most comprehensive view of transcriptome of heterogeneous tumor cell areas in a large cohort of PDAC patients.

STARSOLO: SINGLE-CELL RNA-seq DATA ANALYSIS BEYOND GENE EXPRESSION.

Ash Blibaum, Alexander Dobin

Cold Spring Harbor Laboratory, Genomics, Cold Spring Harbor, NY

Rapid improvements of single-cell RNA-seq (scRNA-seq) technologies are paving the way towards comprehensive characterization of molecular states of all human cells. Here we present **STARsolo[1], a comprehensive ultra-fast turnkey solution for analyzing scRNA-seq data,** built directly into our RNA-seq aligner STAR. Currently, the most commonly used software for mapping and quantifying the 10X Genomics scRNA-seq data is Cell Ranger. Even though Cell Ranger's read alignment step uses STAR and thus is fast, its cell barcode demultiplexing and UMI counting take a long time and require significant computing resources. The STARsolo output is designed to be a drop-in replacement for the CellRanger gene quantification output and produces nearly identical gene counts in the same format. To increase computational efficiency, cell barcode demultiplexing, UMI collapsing, mapping and quantification are integrated into a single code and are performed simultaneously. This avoids input/output bottlenecks and results in **massive (~10 fold) boost in processing speed** and a significant reduction in analysis complexity. Unlike Cell Ranger, STARsolo is compatible with many established and emerging scRNA-seq protocols, such as inDrop, Microwell-seq, sci-RNA-seq, SPLiT-seq, SMART-seq. We have compared STARsolo with other popular tools, such as UMI-tools[2], Alevin[3], kallisto/bustools[4], and will discuss their advantages and drawbacks.

In addition to gene expression quantification, **STARsolo is capable of detecting pre-mRNA, alternative and novel splicing, novel transcription start and termination sites, unannotated transcript/genes, RNA editing, chimeric (fusion) junctions**. Detection of these transcriptomic features may enable discovery of novel transcriptional and post-transcriptional regulation in rare cell types. To this purpose, we re-analyzed multiple publicly available massive scRNA-seq datasets containing millions of mouse and human cells. It is a common preconception that 3' cloning bias in scRNA-seq data prevents studying alternative splicing. We found that despite this bias, 20-40% of the reads are spliced and it is possible to observe alternative splicing exclusive to certain cell types. We will also show multiple examples of cell-population-dependent regulation of RNA polyadenylation and RNA-editing in the 3' UTR of the transcripts.

[1] https://github.com/alexdobin/STAR/blob/master/docs/STARsolo.md
[2] Smith, Heger, Sudbery: Genome Research 2017
[3] Srivastava, Malik, Smith, Sudbery, Patro: Genome Biology 2019
[4] Melsted, Booeshaghi, Gao, Beltrame, Lu, Hjorleifsson, Gehring, Pachter: bioRxiv 2019

TUMOR PROMOTING MACROPHAGES PREVAIL IN THE MALIGNANT ASCITES OF ADVANCED GASTRIC CANCER

Hye Hyeon Eum[1,2], Areum Jo[1], Hae Ock Lee[1,3], Woong-Yang Park[1,3]

[1]Samsung Medical Center, Samsung Genome Institute, Seoul, South Korea, [2]Sungkyunkwan University School of Medicine, Department of Molecular Cell Biology, Suwon, South Korea, [3]Sungkyunkwan University, Department of Health Sciences and Technology, Samsung Advanced Institute for Health Sciences & Technology, Seoul, South Korea

Malignant ascites frequently accompanies the end-stage disease of advanced gastric cancer, resulting from peritoneal carcinomatosis. Affected patients show an abysmal survival and treatment at this stage is focused on palliation of symptoms. To define the cellular characteristics of malignant ascites and search for therapeutic strategies, we applied single cell RNA sequencing to cancer cells and tumor associated macrophages (TAM), present in large numbers in malignant ascites.

We captured 180 cells, including cancer cells, macrophages and mesothelial cells. Analysis of ligand-receptor expression suggests highly activation of cancer cells and tumor promoting interactions between macrophages and cancer cells. The most active pathways between cancer cell and macrophage were TAM recruitment and anti-inflammatory signaling. By constructing the reference transcriptomes for inflammatory M1 and non-inflammatory M2 macrophages in the single cell level, we evaluated inflammatory or non-inflammatory characteristics of the TAM. Macrophages recovered from the ascites were polarized towards non-inflammatory M2 type. The highly polarized M2 phenotype was unique for the macrophages recovered from the malignant ascites of gastric cancer compared to other cancer types. Single cell transcriptome-deriven M2-specific signature has a strong value for prognosis prediction.

These results indicate that tumor-promoting M2 characteristics are concocted by the associated tumor cells. These data present important implications in the treatment strategies of advanced gastric cancer that combinatorial treatment targeting cancer cells and macrophages may have a reciprocal synergistic effect.

SINGLE CELL EXPLORER, COLLABORATION-DRIVEN TOOLS TO LEVERAGE LARGE-SCALE SINGLE CELL RNA-seq DATA

Di Feng[1], Charles E Whitehurst[1], Dechao Shan[1], Jon D Hill[1], Yue G Yong [2]

[1]Boehringer Ingelheim Pharmaceuticals, Inc, Computational Biology, Ridgefield, CT, [2]Camp4 Therapeutics Corp, Data Science, Cambridge, MA

Single Cell Explorer is a Python-based web server application we developed to enable computational and experimental scientists to iteratively and collaboratively annotate cell expression phenotypes within a user-friendly and visually appealing platform. These annotations can be modified and shared by multiple users to allow easy collaboration between computational scientists and experimental biologists. Data processing and analytic workflows can be integrated into the system using Jupyter notebooks. The application enables powerful yet accessible features such as the identification of differential gene expression patterns for user-defined cell populations and convenient annotation of cell types using marker genes or differential gene expression patterns. Users are able to produce plots without needing Python or R coding skills. As such, by making single cell RNA-seq data sharing and querying more user-friendly, the software promotes deeper understanding and innovation by research teams applying single cell transcriptomic approaches. The tool 1) serve as a database with data portal function for core genomic labs to share large-scale single cell RNA-seq data, 2) support for all types of model organisms, 3) flexible data visualization tools including freehand selection, labeling, annotation, and marker gene curation 4) Integration with python notebook API for reanalysis by Computational Biologists.

HOW WELL DEFINED ARE REFERENCE CELL TYPES IN THE BRAIN?

Stephan Fischer, Jesse Gillis

Cold Spring Harbor Lab, Quantitative Biology, Cold Spring Harbor, NY

The Brain Initiative Cell Census Network (BICCN) is an NIH-funded consortium aiming at discovering and characterizing all cell types of neurons. In the past two year, the BICCN Mini-Atlas generated several large molecular datasets specifically focused on the primary motor cortex (mainly in mouse). Analyses of single-cell expression data yielded two levels of clustering. The “subclass” level represents larger groups of cell types that are functionally or developmentally related (e.g. somatostatin-expressiong inhibitory neurons, intra-cortical projecting excitatory neurons), it contains approximately 20 populations of cells. The “cell type” level represents potential cell types (e.g. Chandelier cells, long projecting interneurons), it contains up to 100 populations of cells for the largest dataset. We assess the quality and replicability of clusters using two approaches. Internally, we apply MetaNeighbor, a statistical framework quantifying replicability using leave-one-dataset-out cross-validation, to identify consistent clusters across 9 BICCN datasets totaling more than 500k cells. Externally, we extract meta-analytical gene signatures from replicating BICCN clusters and find additional evidence for these clusters by applying naive classifiers to 6 previously published datasets. We show that the subclass level is well defined overall, with high internal replicability scores and high external classification scores across all cortical regions, but with some exceptions for clusters of inhibitory neurons. At the cell type level, we find various levels of evidence for internal replicability, with approximately half of the clusters showing good evidence of replicability. We show that the clusters with the highest evidence of internal replicability can be readily identified in external datasets, even in developing populations of neurons. We find that clusters with lower internal replicability within the BICCN cannot be externally validated from datasets outside the BICCN, suggesting caution in using provided cell-type clusters purely as gold standards.

SINGLE CELL RNA SEQUENCING REVEALS DEVELOPMENTAL ORIGINS OF MOUSE OVIDUCT EPITHELIAL CELL HETEROGENEITY

Matthew J Ford[1], Alain S Pacis[2], Helen Maunsell[1], Keerthana Harwalkar[1], Yu C Wang[2], Dunarel Badescu[2], Katie Teng[1], Nobuko Yamanaka[1], Ioannis Ragoussi[2], Yojiro Yamanaka[1]

[1]McGill University, Goodman Cancer Research Centre, Montreal, Canada, [2]McGill University, Genome Quebec Innovation Centre, Montreal, Canada

The oviduct also known as the fallopian tube in humans, is a conduit connecting the ovary to the uterus derived from the distal region of the Müllerian duct. It is lined by a highly folded epithelium of multi-ciliated and secretory cells which provides an environment to support sperm transport, fertilization and preimplantation embryonic development. The morphology of the epithelial lining and proportions of multi-ciliated to secretory cells transitions along the length of the oviduct. However it is unknown if the secretions and function of multi-ciliated cells remain constant throughout this transition.

Using a *Pax2-GFP BAC* transgenic mouse line, which faithfully reports *Pax2* expression, we have identified the absence of *Pax2* expression in the distal region of the oviduct in adult mice and throughout the development of the Müllerian duct. To investigate the regional heterogeneity further we conducted two single cell RNA sequencing experiments using epithelial cells isolated prior to epithelial differentiation at postnatal day 4 and epithelial cells from adult mice. Five epithelial cell types were identified in adult mice and located to different regions of the oviduct. Differential expression of secretory proteins implicated in sperm activation and embryonic development, in addition to two distinct multi-ciliated populations, highlights the regional specific functions of the oviduct during reproduction. The expression profiles of undifferentiated epithelial cells at postnatal day 4 confirmed proximal-distal specification occurred before differentiation. *Wt1* expression was identified to be specifically expressed in distal epithelial cells complimentary to *Pax2* expression throughout Müllerian duct development.

In this study we have used single cell RNA sequencing to identify and characterize oviduct epithelial cells and spatially mapped these subtypes to different locations in the oviduct. Regional specification occurred early in Müllerian duct development suggesting distinct developmental origins of distal and proximal epithelial cells. The results from this study have important implications for the understanding of oviduct development and homeostasis.

COMPARATIVE SINGLE-CELL TRANSCRIPTOMIC ANALYSIS OF THE HEMATOPOIETIC SYSTEM BETWEEN HUMAN AND MOUSE

Shouguo Gao, Zhijie Wu, Xingmin Feng, Vivian Dong, Sachiko Kajigaya, Neal S Young

NHLBI, National Institutes of Health, Hematology Branch, Bethesda, MD

The cross-species comparison of hematopoietic hierarchy is still not thoroughly studied. We constructed a single-cell transcriptomic atlas of Hematopoietic Stem and Progenitor Cells (HSPCs) in human and mouse, from a total number of 32,805 single cells. By known marker genes, we grouped human cells as hematopoietic stem cell (HSC), multilymphoid progenitor, granulocyte-monocyte progenitor (GMP), ProB cell, and megakaryocytic-erythroid progenitor (MEP); and mouse cells as long-term HSC, lymphoid multipotent progenitors, and multipotent progenitor (MPP), GMP, MEP and common myeloid progenitors. After alignment with CCA in Seurat 2.0 (https://satijalab.org/seurat/), the cells of mouse and human were well mixed and separated by the same cell type categories. The cells were grouped into 17 subpopulations and cluster-specific genes were species -conserved and shared same functional themes.

After calculating an average of expression of cells in each population of human and mouse, the clustering dendrogram indicated that cell types were highly conserved between human and mouse. We used Monocle to examine the differentiation trajectory of hematopoiesis in human and mouse. Graphically, an intuitive representation of HSPC differentiation emerged. In both mouse and human, three branches (Erythroid/megakaryocytic, Myeloid, and Lymphoid) derived directly from HSC and LTHSC. Single-cell transcriptomes of human cells were compared with those of mouse cells using scmap. Most human MEP cells (85%) mapped to mouse MEP cell types based on transcriptional similarity, suggesting species conservation of functional organization. Further, 45% and 24% human HSC cells were mapped to mouse LTHSC and MPP cell types, respectively. Our analysis confirms evolutionary conservation in hematopoietic systems between mouse and man.

DIFFERENTIAL 3'UTR USAGE ALONG THE OLFACTORY BULB NEUROGENIC LINEAGE

Nikhil Oommen George[1,3], Manuel Goepferich[1,3], Georgios Kalamakis[1,3], Ulrike Mueller[4], Claudia Pitzer[5], Jeroen Krijgsveld[2], Isabel Fariñas[7], Raúl Méndez[8], Simon Anders[6], Wolfgang Huber[9], Ana Martin Villalba[1,3]

[1]DKFZ, Molecular Neurobiology, Heidelberg, Germany, [2]DKFZ, Proteomics of Stem Cells and Cancer, Heidelberg, Germany, [3]Heidelberg University, Biosciences, Heidelberg, Germany, [4]Heidelberg University, IPMB, Heidelberg, Germany, [5]Heidelberg University, INBC, Heidelberg, Germany, [6]Heidelberg University, ZMBH, Heidelberg, Germany, [7]University of Valencia, Cell Biology, Valencia, Spain, [8]Institute for Research in Biomedicine (IRB), Molecular Medicine, Barcelona, Spain, [9]EMBL, Genome Biology Heidelberg, Germany

A small population of neural stem cells (NSCs) contributes to tissue homeostasis and repair in the adult brain. Single-cell transcriptome analysis has revealed heterogeneity within the NSC population, that transitions between different activation states. Molecular mechanisms leading to this transition are not clearly understood. However, data suggests translational control could play an important role.
Our single cell RNA-seq (scRNA-seq) data analysis has revealed that alternative polyadenylation (APA) is sufficient to separate NSCs and their progeny in silico by defining 3'UTR-wise distances of single cells. A pseudo temporal ordering of the NSC activation states based on their 3'UTR information shows a similar lineage progression as with their transcriptome profiles. scRNA-seq analysis of active and BMP4 induced quiescent NSCs in vitro also revealed significant changes in the 3'UTR usage. Thus, changes in the activation state of NSCs in vitro consequentially altered 3'UTR landscape. These changes in the 3'UTR usage could provide an additional layer of regulation for translational control via miRNA biding sites and other regulatory elements.
Analysis of polyadenylation factors with respect to 3'UTR usage identified amyloid-precursor like protein 1 (APLP1) as a potential upstream regulator. A small set of regulated targets with gene ontology (GO) categories such as neurogenesis and differentiation appear to be downstream to APLP1. Our data shows that APLP1 along with CPEB4 is up-regulated in the dormant subpopulation of quiescent NSCs. FACS sorting the NSCs lineage ($GLAST^{+}Ter119^{-}O4^{-}CD45^{-}$) showed reduced cell numbers in the $APLP1^{-/-}$ mice as compared to their WT counterparts. Further, differences in the 3'UTR landscape were observed between $APLP1^{-/-}$ and WT mice. These genes showed a strong association to neuronal disorders like autism. Preliminary behavior studies show decreased social novelty in the $APLP1^{-/-}$ mice. In light of these data, we hypothesize that APLP1 could play a role in maintaining quiescence via alternative polyadenylation and thereby regulate neurogenesis in the SVZ.

SPATIOTEMPORAL DYNAMICS OF CALVARIAL SUTURE DEVELOPMENT REVEALED BY INTEGRATED ANALYSIS OF SINGLE-CELL AND WHOLE TRANSCRIPTOME DATA

Ana S Gonzalez-Reiche*[1,2], Xianxiao Zhou*[1,2], Greg Holmes*[1], Na Lu*[1], Joshua Rivera[1], Divya Kriti[1,2], Robert Sebra[1,2], Anthony A Williams[1], Michael J Donovan[3], S. Steven Potter[4], Dalila Pinto[1,2,5], Ethylin Wang Jabs[1,6], Bin Zhang[1,2], Harm van Bakel[1,2]

[1]Icahn School of Medicine at Mount Sinai, Genetics and Genomic Sciences, New York, NY, [2]Icahn School of Medicine at Mount Sinai, Icahn Institute for Data Science and Genomic Technology, New York, NY, [3]Icahn School of Medicine at Mount Sinai, Department of Pathology, New York, NY, [4]Cincinnati Children's Medical Center, Division of Developmental Biology, Cincinnati, OH, [5]Icahn School of Medicine at Mount Sinai, Department of Psychiatry, and The Mindich Child Health and Development Institute, New York, NY, [6]Icahn School of Medicine at Mount Sinai, Department of Cell, Developmental and Regenerative Biology and Department of Pediatrics, New York, NY *authors contributed equally.

Craniofacial sutures are the growth sites of the developing skull. Mutations in a variety of genes can result in calvarial pathologies, such as craniosynostosis, characterized by premature fusion of the suture bones. Understanding the organization of molecular and cellular processes within sutures during development requires the definition of their architecture at the cellular level. To address this, we generated single-cell RNA expression profiles of the murine frontal suture. Analysis of over 7,000 cells from the frontal suture at embryonic days (E)16.5 and E18.5 revealed the presence of at least 14 cell type clusters characterized by unique gene expression signatures, including nine distinct subpopulations of mesenchymal and osteoblast cells, which exhibit distinct spatial localization within the suture. An inferred differentiation trajectory of these nine subpopulations suggests that mesenchymal cells differentiate into osteoblasts, as well as into other cell populations within the suture. To further characterize transcriptional changes during frontal suture development in normal and pathological conditions, we generated bulk RNA-seq profiles. Transcriptome analysis of 60 frontal suture subregions of suture mesenchyme and osteogenic fronts at E16.5 and E18.5 of WT mice and two mutant models of craniosynostosis with altered frontal suture growth, *Twist1*$^{+/-}$ and *Fgfr2*$^{+/S252W}$, revealed novel gene expression signatures that could be mapped to the mesenchymal cell populations identified by single-cell RNA-seq. These signatures included stem cell markers and genes associated with connective tissue, vascularization, and ribogenesis. By integrating cell type-specific signatures with gene co-expression networks inferred from the bulk RNA-seq data, we identified a novel mesenchymal gene expression module of potential regulators of osteogenesis. Altogether, our integrative approach provides a comprehensive framework for studying the transcriptional dynamics of suturogenesis and other developmental processes.

NETWORK CONTROL OF PHENOTYPIC PLASTICITY IN SMALL CELL LUNG CANCER

Sarah M Groves[1], David Wooten[1,2], Ken Lau[3], Carlos Lopez[1], Vito Quaranta[1]

[1]Vanderbilt University, Biochemistry, Nashville, TN, [2]Penn State University, Physics, State College, PA, [3]Vanderbilt University, Cell and Developmental Biology, Nashville, TN

Phenotypic plasticity plays a major role in the ability of a cancer to acquire resistance after chemotherapy. Understanding the structure and dynamics of a gene regulatory network (GRN) that regulates cell phenotype can uncover the mechanisms behind this plasticity exhibited in heterogeneous tumors. For example, small cell lung cancer (SCLC) is an aggressive tumor type that often relapses and metastasizes early, but the chemotherapeutic standard of care for SCLC has not changed in almost half a century. During much of this time, SCLC has been treated as a homogenous disease, resulting in a high rate of relapse after initial response to therapy. In order to address inevitable acquired resistance after treatment, more recent research has investigated the role that intratumoral, transcriptional heterogeneity may play (Rudin et al., 2019). Our work has revealed at least four transcriptional subtypes of SCLC cells defined by a GRN, including a novel neuroendocrine variant (NEv2) distinguished by gene co-expression modules and drug response. However, the ability of individual cells to change their phenotype in response to drug has not been adequately characterized. By combining an RNA velocity analysis of single cell RNA-seq data, Markov chain modeling, and statistical mechanics theory, we have been able to understand and quantify movement through the phenotypic landscape, and well as make predictions of perturbations that may induce or prohibit those transitions. Our analysis of human and mouse tumors suggests that SCLC phenotypes are highly plastic, both stochastically and in response to drug. This characterization of the highly plastic transitions between core SCLC phenotypes and predictions of perturbation strategies may lead to new methods of personalized treatment.

MAPPING ADIPOGENESIS IN HUMANS USING SINGLE-CELL AND SINGLE-NUCLEI RNA-SEQUENCING

Anushka Gupta[1], Aaron M Streets[2,3]

[1]UC Berkeley-UC San Francisco Graduate Program in Bioengineering, Berkeley Graduate Division, Berkeley, CA, [2]UC Berkeley, Bioengineering, Berkeley, CA, [3]Chan Zuckerberg, Biohub, San Francisco, CA

There are at least two known kinds of fat: the white adipose tissue (WAT) which participates in energy storage, and brown and the related beige adipose tissue (BAT), which specializes in energy expenditure. Together, WAT and BAT maintain energy balance in almost all animals, dysregulation of which is implicated in the pathology of metabolic disorders like obesity. Despite its importance, heterogeneous cell populations that reside in human WAT and BAT and their developmental trajectories are poorly understood. A key challenge to this undertaking is the variation of adipose tissue composition, not only from one individual to another but also between distinct anatomical locations in the body. To tackle these challenges, we use an *in vitro* adipogenesis model involving WAT and BAT progenitors isolated from the neck region of a single individual, which can be differentiated into lipid-laden cells, demonstrating molecular and functional characteristics of human white and brown adipocytes respectively. We apply high throughput microfluidics techniques with barcode enabled multiplexing to profile the transcriptome of single WAT and BAT progenitors undergoing adipogenesis at multiple time points, from the precursor state to the mature adipocyte state, using single-cell and single-nuclei RNA-sequencing. Our goal is to generate a high-resolution scaffold of the WAT and BAT differentiation landscape in order to characterize BAT and WAT adipogenic signaling, identify new cell populations and their functions in energy homeostasis, and provide a reference for studies of primary tissue.

VGATE: AN R PACKAGE TO AUTOMATICALLY DEFINE CELL TYPES/STATES AND CELL-OF-ORIGIN FOR SINGLE CELL RNASEQ DATA

Guangchun Han[1], Yuanxin Wang[1], Ruiping Wang[1], Shaojun Zhang[1], Jaffer Ajani[2], Michael Green[3], Andrew Futreal[1], Linghua Wang[1]

[1]U.T. MD Anderson Cancer Center, Department of Genomic Medicine, Houston, TX, [2]U.T. MD Anderson Cancer Center, Department of Gastrointestinal Medical Oncology, Houston, TX, [3]U.T. MD Anderson Cancer Center, Department of Lymphoma/Myeloma, Houston, TX

Automatically identification of cell types and cellular states and cell-of-origin is the first and also an essential step for large-scale single cell RNAseq (scRNAseq) data, particularly for cancer research. Traditionally, cell types are identified by using canonical gene markers. But gene dropout events post a great challenge for this approach, especially for cell types with gene markers expressed at a low level, such as CD4 gene from CD4 positive T cells. In addition, cell type determination is generally performed by manual inspection of marker gene expression from feature plots, a process which is labor-intensive and heavily biased. Moreover, for cancer research, cell-of-origin analysis helps in understanding the initiation of cancer and intratumoral heterogeneity and may favor early detection of malignancies. Currently, there isn't a tool to help identify the origin of cells from compartments other than the immune system. Cell-of-origin analysis is also paramount to understand the biology of developmental disease. We have developed vGATE, an R package to automatically identify cell types and cell-of-origin for scRNAseq data. vGATE combines marker- and clustering-based strategies to iteratively boost the cell type identification process. This process has been validated to be able to rescue ~20% of cells for each cell type with high accuracy. Cell-of-origin is determined by computing the transcriptomic similarity between cells of interests and our curated large scale scRNAseq database (~1M cells). vGATE takes either raw or processed cell-gene matrix and the processing is fast. Cell type assignment is fully automatic and the outputs are in various format for downstream integrative analysis. Together, vGATE is a useful tool that can be incorporated into existing scRNAseq bioinformatics pipelines.

COMBINED SINGLE-CELL QUANTIFICATION OF GENOME AND TRANSCRIPTOME IN BREAST CANCERS

Kyung Yeon Han[1], Taeseob Lee[2], Jeongmin Woo[2], Woong-Yang Park[1,2], Jinho Kim[1], Dong Hyun Park[3]

[1]Samsung Medical Center, Samsung Genome Institute, Seoul, South Korea, [2]Sungkyunkwan University of Medicine, Samsung Advanced Institute for Health Sciences and Technology, Seoul, South Korea, [3]GENINUS Inc, Research center, Seoul, South Korea

Breast cancers display highly variable patterns of genetic diversity and clonal architecture, which is associated with inter- and intra-tumor heterogeneity. The heterogeneity in breast cancers is known to be significant challenge in cancer diagnosis and treatment design. Luminal breast cancer is the most heterogeneous, molecularly and clinically, and frequently occurring breast cancer subtype in the population. In order to understand the relationship between genome and transcriptome of luminal breast cancers in single cell level, the genome and the transcriptome sequencing data from the same single cells, using the SIDR-seq, from breast cancer patients were generated. By comparing copy number alteration patterns with transcriptome profiles, we quantitatively estimated the contribution of copy number alteration in generating phenotypic heterogeneities of breast cancers. Estimation of transcriptomic variability created by copy number variation revealed significant associations far exceeding random expectation. Our observation was further supported by analyzing high-throughput scRNA-seq data generated using droplet-mediated RNA-seq platform in parallel with scSIDR-seq. The analysis of gene expressions across tumor cells suggested that interaction between tumor and tumor microenvironment is important factor to shape transcriptomic heterogeneity of tumor cells.

METHODS AND STATISTICS FOR DIFFERENTIAL CO-EXPRESSION IN SINGLE CELL RNA SEQUENCING DATA

Benjamin D Harris[1,2], Jesse Gillis[1]

[1]Cold Spring Harbor Laboratory, Genomics, Cold Spring Harbor, NY,
[2]Cold Spring Harbor Laboratory, Watson School of Biological Sciences, Cold Spring Harbor, NY

Co-expression or co-variation of expression profiles between genes across conditions are used to build co-expression networks. Because shared expression patterns are thought to reflect shared functionality, co-expression networks are typically used to describe the functional relationships between genes. Single-cell RNA-sequencing presents us with the capability to study co-expression in a cell-type specific way. Comparing cell-type specific networks, such as using differential co-expression analysis, identifies regulatory relationships between genes unique to individual cell-types. However, study-specific characteristics of datasets and the sparseness of single-cell data make accurate analysis impossible in single datasets.

We mitigate these problems and produce high performing cell-type specific co-expression networks using meta-analysis across over 350,000 cells from 8 publicly available datasets from the Brain Initiative Cell Census Network and Allen Institute for Brain Science that profile the mouse cortex. We compute differential co-expression between networks built from Glutamatergic and GABAergic neurons, medial (MGE) and caudal ganglionic eminence (CGE) derived GABAergic neurons and Pvalb and Sst GABAergic neurons to study the statistical power of differential co-expression at different levels of cell type classification/abundance. The different abundances of cells across these three levels of classification depict the decrease in power as cell classification becomes more specific. Similar to previous studies in bulk RNA-sequencing tissue specific networks, our results show that both differential expression and average expression are highly predictive of network rewiring. In conclusion, we highlight the importance of a well powered meta-analysis in building and comparing cell type specific co-expression networks from Single-cell RNA-sequencing data.

SINGLE-CELL RNA-seq REVEALS NAÏVE B CELLS ASSOCIATED WITH BETTER PROGNOSIS OF HEPATOCELLULAR CARCINOMA

Jian He[1], Yingxin Lin[2], Xianbin Su[1], Jean Yang[2], Ze-guang Han[1]

[1]Shanghai Jiao Tong University, Shanghai Center for Systems Biomedicine, Shanghai, China, [2]The University of Sydney, School of Mathematics and Statistics, Sydney, Australia

Hepatocellular Carcinoma (HCC) is a type of malignant solid tumor, causing high morbidity and mortality around the world and the major portion of HCC patients is from China. Cancer immunotherapies have shown some clinical responses in treating some types of cancer but did not shown significant efficiency in HCC treatment. This in part due to the impact of immune cells in the tumor microenvironment. It is commonly believed that HCC is a heterogeneous solid tumor and the microenvironment of HCC plays an important role in tumorgenesis and development. Currently, the residents of the microenvironment of HCC is not well-defined and clarification, especially the immune cells, which we believe that paly pivotal roles in tumorgenesis and development. To depict the landscape of the composition, lineage and functional states of the immune cells in HCC, we performed single-cell RNA sequencing on Diethylnitrosamine (DEN)-induced mouse HCC model. We observed heterogeneity within the immune and hepatocytes both in the precancerous condition of tumorigenesis and cancerous condition of HCC. In this study we found that the disease-associated changes appeared early in pathological progression and were highly cell-type specific. Specific subsets of T and B cells preferentially enriched in HCC, and we identified signature genes for each subset. Additionally, we mapped this group of specific cells to the human TCGA database. We found a cluster of naïve B cells characterized by high expression of CD38 associated with better prognosis of human HCC. Our study demonstrates signaling interaction map based on receptor-ligand bonding on the single-cell level could broaden our comprehending of cellular networks in varies status. Our finding provides a new approach for patient stratification and will help further understand the functional states, dynamics and signaling interaction of B cells in hepatocellular carcinoma, and may provide a novel insight and therapeutics for the HCC.

A CELLULAR ATLAS OF HUMAN PROSTATE DISEASE

Gervaise H Henry[1], Alicia Malewska[1], Venkat S Malladi[2], Jeffrey C Gahan[1], Jeff C Reese[3], Douglas W Strand[1]

[1]UT Southwestern Medical Center, Department of Urology, Dallas, TX, [2]UT Southwestern Medical Center, Department of Bioinformatics, Dallas, TX, [3]Southwest Transplant Aliance, Dallas, TX

Single-cell RNA-sequencing was used to define the cellular and molecular landscape of human prostatic disease states. Primary tissue from normal organ donors (donor, Pd); patients undergoing simple prostatectomy for benign prostate hyperplasia (BPH, Pb); and radical prostatectomy or targeted biopsy for localized prostate cancers (Pc) were freshly digested for single-cell RNA-sequencing. These data were aggregated to distinguish the cellular composition of normal versus diseased prostate. These single-cell data were also used to deconvolute the cellular composition of TCGA Prostate Adenocarcenoma bulk expression data. This powerful approach allows us to better clarify the role of specific cell types in clinical outcomes and genetic features.

THE GEAR PORTAL - GENE EXPRESSION ANALYSIS RESOURCE

Joshua Orvis[1], Brian Gottfried[1], Kevin Rose[1], Jayaram Kancherla[2], Yang Song[1], Beatrice Milon[1], Dustin Olley[1], Hector Corrada Bravo[2], Anup Mahurkar[1], Ronna Hertzano[1,3,4]

[1]University of Maryland School of Medicine, Institute for Genome Sciences, Baltimore, MD, [2]University of Maryland College Park, Department of Computer Science, College Park, MD, [3]University of Maryland School of Medicine, Otorhinolaryngology Head and Neck Surgery, Baltimore, MD, [4]University of Maryland School of Medicine, Anatomy and Neurobiology, Baltimore, MD

The gEAR portal (gene Expression Analysis Resource, umgear.org) is an online tool for multi-omic and multi-species data visualization, sharing, and analysis. Originally designed for auditory and vestibular researchers, the gEAR portal has now been expanded for general use. The gEAR is unique in its ability to allow users to upload, view and analyze their own data in the context of previously published datasets, as well as confidentially share their data with collaborators prior to publication. It is also unique in combining not only multiple species but multiple data types including bulk RNA-seq, sorted cell RNA-seq, single cell RNA-seq (scRNA-seq) and epigenomics in a one page, user-friendly, browsable format. Individual expression datasets can be displayed in a variety of ways alongside each other, including interactive bar, line or violin plots, and colorized anatomical images.

Most recently, scRNA-seq has matured as a commonly used technique for measuring gene expression across tissues. To provide researchers access to scRNA-seq data regardless of their programming knowledge, we have integrated a scRNA-seq workbench into the gEAR. The gEAR workbench provides access to both the raw data of scRNA-seq datasets and saved expert analyses where cell types have already been assigned – giving researchers rapid insight into gene expression of their cell type of interest. After the users upload their data they can immediately import their datasets into profiles alongside other public ones of interest, searching genes and comparing their expression across experiment. Using the workbench, they filter for variable genes, perform PCA, tSNE, clustering, marker gene identification, label clusters based on marker genes, and compare genes or clusters. All steps are performed using the interface and stored on the system. These analyses can be made public or stored privately for registered users.

A goal of the gEAR is to become a primary method of data access for expression studies in publication. A user can upload their dataset privately pre-publication, perform analyses, and then shift the dataset to be publicly available upon manuscript submission, including a gEAR permalink within the manuscript. Reproducibility of analysis is also bolstered, as dataset authors can store their analysis from the workbench along with their dataset. This allows others to load it, create a copy, and further interrogate their own copy. The main gEAR site currently has over 500 registered users and over 300 expression datasets. It is developed primarily in Python3, uses a mix HDF5 and relational storage, currently runs completely on an individual cloud node and is accessible at https://umgear.org

HIPPO PATHWAY DELETION IN ADULT RESTING CARDIAC FIBROBLASTS INITIATES A CELL STATE TRANSITION WITH SPONTANEOUS AND SELF-SUSTAINING FIBROSIS

Yang Xiao*[1], Matthew C Hill*[2], Lele Li[1], Vaibhav Deshmukh[1], Thomas J Martin[1], James F Martin[1,2,3,4]

[1]Baylor College of Medicine, Molecular Physiology and Biophysics, Houston, TX, [2]Baylor College of Medicine, Program in Developmental Biology, Houston, TX, [3]Baylor College of Medicine, Cardiovascular Research Institute, Houston, TX, [4]Texas Heart Institute, Cardiomyocyte Renewal Lab, Houston, TX

Cardiac fibroblasts (CFs) respond to injury by transitioning through multiple cell states, including resting CFs, activated CFs, and myofibroblasts. We report here that Hippo signaling cell-autonomously regulates CF fate transitions and proliferation, and non-cell-autonomously regulates both myeloid and CF activation in the heart. Conditional deletion of Hippo pathway kinases, Lats1 and Lats2, in uninjured CFs initiated a self-perpetuating fibrotic response in the adult heart that was lethally exacerbated by myocardial infarction (MI). Single cell transcriptomics showed that uninjured Lats1/2 mutant CFs spontaneously transitioned to a myofibroblast cell state. Through gene regulatory network reconstruction, we found that Hippo-deficient myofibroblasts deployed a network of transcriptional regulators of endoplasmic reticulum (ER) stress, and the unfolded protein response (UPR) consistent with elevated secretory activity. Moreover, we observed an expansion of myeloid cell heterogeneity in uninjured Lats1/2 CKO hearts with a striking similarity to cells recovered from infarcted control hearts. Integrated genome-wide analysis of Yap chromatin occupancy revealed that Yap directly activates myofibroblast cell identity genes, the proto-oncogene Myc, and an array of genes encoding pro-inflammatory factors through enhancer-promoter looping. Thus, our data indicate that Lats1/2 maintain the resting CF cell state through restricting the Yap-induced injury response.

*authors contributed equally

SINGLE CELL TRANSCRIPTOME PROFILING IN COLORECTAL CANCER REVEALS TUMOR HETEROGENEITY AND INTRINSIC TUMOR CELL SIGNATURES ASSOCIATED WITH THE CONSENSUS MOLECULAR SUBTYPE

Yourae Hong[1,2], Hae-Ock Lee[1,3], Yong Beom Cho[4], Woong-Yang Park*[1,2,3]

[1]Samsung Medical Center, Samsung Genome Institute, Seoul, South Korea, [2]Sungkyunkwan University, Department of Health Sciences and Technology, Samsung Advanced Institute for Health Sciences &Technology, Seoul, South Korea, [3]Sungkyunkwan University School of Medicine, Department of Molecular Cell Biology, Suwon, South Korea, [4]Samsung Medical Center, Sungkyunkwan University School of Medicine, Department of Surgery, Seoul, South Korea

* Correspondence

To understand characteristics of colorectal cancer, subtype classification using molecular characteristics are applied. Consensus molecular subtype (CMS) classifies four colorectal cancer subtypes demonstrating immune, canonical, metabolic, and mesenchymal signatures from bulk whole transcriptome data. The classification represents the molecular characteristics of tumor tissues including tumor cells and their microenvironmental components. With single cell transcriptome profiling, these mixed phenotypes can be explicitly characterized. Here, we performed single-cell RNA sequencing for 23 colorectal cancer and 10 patient-matched distant normal tissues using 10X chromium system. At single-cell resolution, we discovered that the molecular characteristics in tissue can be derived tumor cell or microenvironment. Cancer cells revealed intrinsic transcriptional signatures, reminiscent of absorptive or secretory cell lineages, matching the CMS2 and CMS3 subtypes. On the other hand, tumor microenvironment cells mostly determine the CMS1 and CMS4 subtypes. These data define the intra-tumor heterogeneity of colorectal cancer that may lead to a diverse epithelial functions and environmental adaptations that can affect treatment resistance and progression.

GENETIC VARIATION REGULATORY NETWORKS FROM SINGLE CELL RNA SEQUENCING

NM Prashant[1], Hongyu Liu[1], Pavlos Bousounis[1], Justin Sein[1], Liam Spurr[2,3,1], Anelia Horvath[1]

[1]George Washington University, Biochemistry and Molecular Medicine, Washington, DC, [2]Dana Farber Cancer Institute, Department of Medical Oncology, Boston, MA, [3]The Broad Institute of MIT and Harvard, Boston, MA

Single cell RNA sequencing (scRNA-seq) emerges as a powerful platform to study allele-specific regulatory relationships, and, respectively, to identify functional genetic variants. We present an scRNA-seq–focused approach for assessment of variant allele frequency (VAF_{RNA}) at single nucleotide positions and its correlation to other genomic features including gene expression, splicing and RNA-editing. We apply our approach on over 300,000 heterozygous SNV positions on scRNA-seq data (10x Genomics Chromium) from ~24,000 human adipose-derived mesenchymal stem cells (ADSCs) obtained from three donors. To ensure robust VAF_{RNA} estimation we correct for allele-mapping bias through SNV-aware alignment, and for PCR-duplicates using unique molecular identifiers (UMI). We call variants (GATK) on the aggregated alignments (all cells from a donor) and determine heterozygous positions based on bi-allelic signal; we then estimate VAF_{RNA} in the individual cells' alignments at three levels of coverage - 10, 5 and 3 unique sequencing reads. In parallel, we estimate: gene expression and genetic heterogeneity (FeatureCount, Seurat), proportion of alternatively spliced junctions (LeafCutter) and level of RNA-editing in known RNA-editing loci. We next apply Quantitative Loci Traits (QTL)-based analyses to search for pair-wise correlations between VAF_{RNA}, gene expression, splicing patterns and RNA-editing. Our analysis shows wide-spread and strong correlation in cis (adjacent positioning on the same chromosome) between genetic alleles and both gene expression and splicing. While trans-correlations were fewer, they were enriched in known molecular interactions. Finally, we observe wide-spread random monoallelic expression (from only one of the chromosomes at a time) for genes with low and intermediate expression levels, whereas highly expressed genes were frequently transcribed simultaneously from the two chromosomes. Our results show that VAF_{RNA} from scRNA-seq data is an informative estimation for genetic variation, and can be implemented in QTL-based strategies for studying genetic regulatory networks.

INCREASED T-CELL EXHAUSTION IN ANTIGEN-SPECIFIC CD8 T-CELLS FROM EARLY-ONSET TYPE 1 DIABETIC SUBJECTS AFTER TEPLIZUMAB TREATMENT

Alex Hu[1], Hai Nguyen[2], Hannah DeBerg[1], Elisavet Serti[3], Gerald Nepom[2,3], Eddie James[2], Peter Linsley[1]

[1]Benaroya Research Institute, Systems Immunology, Seattle, WA, [2]Benaroya Research Institute, Seattle, WA, [3]Immune Tolerance Network, Bethesda, MD

In patients with recent onset type 1 diabetes (T1D), induction of a global exhausted CD8 T-cell phenotype has been shown to be associated with preserved c-peptide after treatment with teplizumab, a monoclonal anti-CD3 antibody. However, it is not known whether this phenotype is manifested by pancreatic islet specific cells. We plate-sorted single tetramer positive T cells by flow cytometry and performed single-cell RNA sequencing to characterize the phenotypes and T-cell receptor sequences of 455 CD8 T-cells reactive to either Epstein-Barr virus (EBV) or pancreatic islet antigens from 9 patients with T1D sampled before and two months after a 2-week course of teplizumab treatment in the AbATE clinical trial. A higher proportion of EBV-reactive cells than islet-reactive cells shared T-cell receptor sequences, suggesting more clonal expansion in response to in-vivo EBV antigens. Dimensionality-reduction and clustering on RNA sequencing data performed by Monocle across all cells identified multiple T cell states (or subsets) exhibiting distinct phenotypes. Phenotypic differences were driven in part by markers of T-cell exhaustion and memory. EBV-reactive cells had a higher proportion of cell subsets with markers of T-cell exhaustion and antigen exposure than islet-reactive cells. However, teplizumab-treated samples contained a higher proportion of a cell subset characterized by higher expression of T-cell exhaustion markers and ribosomal genes and a lower proportion of a cell subset characterized by higher expression of genes regulated by NF-kB in response to TNF. The changes in cell subset proportions after treatment occurred in both the EBV-reactive and islet-reactive cells. The increased abundance of the exhausted cell subset after treatment with teplizumab suggests that teplizumab promotes the exhaustion of antigen-specific CD8 T-cells.

MASSIVELY PARALLEL, TIME-RESOLVED SINGLE-CELL RNA SEQUENCING WITH scNT-seq

Qi Qiu*[1,2], Peng Hu*[1,2], Kiya Govek[1,3], Pablo Gonzalez-Camara[1,3], Hao Wu#[1,2]

[1]University of Pennsylvania, Department of Genetics, Philadelphia, PA,
[2]University of Pennsylvania, Penn Epigenetics Institute, Philadelphia, PA,
[3]University of Pennsylvania, Institute of Biomedical Informatics, Philadelphia, PA

*These authors contributed equally to this work
#Correspondence should be addressed to H.W (haowu2@pennmedicine.upenn.edu).

Standard high-throughput single-cell RNA sequencing offers snapshots of transcriptomic diversity but obscures temporal dynamics of RNA biogenesis and decay. Here we present **s**ingle-**c**ell **n**ascent transcript **t**agging sequencing (scNT-seq), a method for massively parallel analysis of nascent and pre-existing RNAs from the same cell. The method integrates metabolic RNA labeling with droplet microfluidics to enable high-throughput chemical conversion of 4-thiouridine to cytidine analogs, thereby marking nascent transcripts with T-to-C mutations. The scNT-seq analysis readily uncovers neuronal subtype-specific nascent transcriptome, gene regulatory networks and RNA velocity trajectories in response to neuronal activity. Pulse-chase scNT-seq enables transcriptome-wide measurements of RNA stability in rare two-cell-embryo-like (2C-like) cells in mouse embryonic stem cell (mESC) cultures. Analysis of mESC metaplastic transition by scNT-seq reveals RNA dynamics associated with and a critical role of DNA methylcytosine dioxygenases in regulating the transition between the primed state and 2C-like state. Time-resolved single-cell transcriptomic analysis thus opens new lines of inquiry regarding cell-type specific RNA regulatory mechanisms.

SINGLE-CELL RNA SEQUENCING IDENTIFIES THE COMMON AND UNIQUE CELL POPULATION BETWEEN IN AN *IN VITRO* AND *IN VIVO* CULTURED CONDITION

Sadahiro Iwabuchi, Shinichi Hashimoto

Kanazawa University, Graduate School of Medical Sciences, Kanazawa, Japan

It is well known that cancer development is influenced by differential microenvironmental conditions such as hypoxia, immune infiltration and aging. In this study, cell lines were used to examine cell changes in different environments *in vivo* and *in vitro*. It intuitively expects that the gene expression profiling in a single-cell may different between these different culture conditions because of the optimized environment for cells in the *in vitro*. However, it is still unknown how the gene expression pattern in single-cell level changes from the *in vitro* cultured cell to the *in vivo* transplanted cell. Here, we used our developed Nx1-seq (next generation 1-cell sequencing) method to compare 2 different environment-cultured mouse tumor cells. The *in vitro* cultured tumor cells were collected and transplanted subcutaneously into a mouse, and tumor cells were dissected and isolated into single cells after housing. At the same time, the *in vitro* cultured cancer cells were also isolated. Isolated cells from the *in vivo* culture were cultured again in the *in vitro* condition. The obtained results were that 1) transplanted mouse hepatocellular carcinoma cell holds some cancer-relate genes (*Hmox1*, *Ube2c*, *Cks2*) that original *in vitro* cultured cell had, 2) unique gene slightly was decreased (*Fabp5*) or induced (*Ndrg1*) during the tumor growth in the *in vivo* transplanted culture, 3) all of these genes were correlated to overall survival of Kaplan-Meier estimate in the liver cancer referred from The Cancer Genome Atlas, 4) the subsequent *in vitro* cultured cells expressed *Ndrg1*, meaning that the cancer cells kept the *in vivo* condition.Therefore, cell lines *in vivo* are more likely to reflect biological responses than *in vitro* culture conditions.

HOMOGENEOUS, SINGLE-CELL LEVEL PROTEIN AND PROTEIN COMPLEX DETECTION

Csaba Jeney, Tobias Gross, Peter Koltay

Albert-Ludwigs-Universität Freiburg, IMTEK, Freiburg i. Br., Germany

'Emulsion coupling' technology was developed to detect proteins and proteins complexes using cellular material below of a single cell amount. The assay is based on compartmentalized linkage of two antibody DNA labels in a bi-component, homogeneous immune-reaction using lysed cellular material. The assay is capable to detect any antibody detectable analytes with extremely high sensitivity, as a single protein target molecule is detectable per compartment, which detection is limited only by the antibody binding equilibrium. The formed DNA dimers can then be amplified using conventional digital PCR (dPCR) technology. The assay has two different ways for readout: using real-time digital PCR, as a single-step assay; and next-generation sequencing, the latter provides the individual enumeration of protein complexes.
Detection of HER2 - human epidermal growth factor receptor 2 (ERBB2) was carried out by both readout and shows the possibility of analysis of very small amount of material (under 1/1000th cells, MCF7 cells). The assay recognizes 8 attomoles of HER2 – at binding conditions –, which means that 20 MCF7 cells in 2 µl are sufficient as a starting material. However, due to the required high dilution (> 25000 times), the detection sensitivity of the assay is in the range of hundreds of yoctomols, which corresponds ~ 200 HER2 molecules in the final assay. This sensitivity can be exploited my reducing the volume of the sample. This low-volume single-cell detection was carried out by a single-cell printer, which shows similar analytical behavior. The same detection conditions were read by NGS, as well, which allow the reconstruction of the molecular content of each droplet. The self-dimerization and raft-formation of the HER2 molecules were demonstrated.
The ‘emulsion coupling’ assay is a promising, new tool for protein and interacting proteins analysis, and albeit its projected high parallelism need to be demonstrated, its extreme sensitivity, versatility and the possibility to enumerate individual protein complexes makes it an ideal research tool for single-cell analysis.

COMBINATORIAL SINGLE CELL ANALYSIS TO LINK GENOTYPE TO PHENOTYPE IN COLORECTAL CANCER

Joshua Thomas*, Michelle Dourado, Jun Zou, Shan Lu, Anneleen Daemen, Felipe de Sousa e Melo, Radia M Johnson*

Genentech Inc, Research & Early Development, South San Francisco, CA

Recent development in single cell transcriptomics has highlighted the critical role of intercellular heterogeneity in phenotype variability in both health and disease (Gawad et al., 2016). More recently, combinatorial single cell genomics has emerged as a powerful approach to integrate different layers of biological insight. In particular, methods that provide joint DNA/RNA sequencing allow us to deconvolute the effect of genetic alterations on gene expression profile and are ideally suited to study cancer evolution and progression. More recently, technological advancements have enabled the assessment of the transcriptome and chromatin accessibility jointly in the same cells (Cao et al., 2018). Here, we have adapted the sci-CAR method to jointly label the RNA and DNA cellular fractions for RNA and DNA sequencing from the same cells. For single cell whole genome amplification (WGA), we have combined single-cell combinatorial indexing with multiple displacement amplification (MDA). A major limitation of MDA for WGA is the amplification errors and bias that occurs during library preparation. To limit our detection of artifactual single-cell variants, we have used the SCcaller (Dong et al., 2017) to more accurately identify single-nucleotide variants (SNVs) from the single cells by correcting for local allelic amplification bias in SNV calling. Using this adapted protocol, we have been able to sequence the DNA portion of LS174T colon cells to call SNVs as well as the RNA portion of these cells to correlate these variants with transcriptomic changes. We will present our protocol, the robustness of our method, and results from multiple colorectal cell lines. We anticipate this methodology to be broadly applicable to many cancers and could provide important insights on the mechanisms that underlie the evolution of clonal lineages of cells in cancer.

References:

Cao, J., Cusanovich, D.A., Ramani, V., Aghamirzaie, D., Pliner, H.A., Hill, A.J., Daza, R.M., McFaline-Figueroa, J.L., Packer, J.S., Christiansen, L., et al. (2018). Joint profiling of chromatin accessibility and gene expression in thousands of single cells. Science 361, 1380-1385.
Dong, X., Zhang, L., Milholland, B., Lee, M., Maslov, A.Y., Wang, T., and Vijg, J. (2017). Accurate identification of single-nucleotide variants in whole-genome-amplified single cells. Nat Methods 14, 491-493.
Gawad, C., Koh, W., and Quake, S.R. (2016). Single-cell genome sequencing: current state of the science. Nat Rev Genet 17, 175-188.

QUANTIFYING BREAST CANCER HETEROGENEITY AND CLONAL FITNESS USING SINGLE CELL TIMESERIES POPULATION DYNAMICS

Farhia Kabeer*[1,2], Sohrab Salehi*[1,3], Emma Laks[1], Kieran Campbell[1], Mirela Andronescu[3], Ciara O'Flanagan[1], Jazmine Brimhall[1], Justina Biele[1], Jecy Wang[1], Danial Lai[1], Andrew McPherson[1], Teresa Algara[1], Hakwoo Lee[1], Alexandre Bouchard-Cote[3], Samuel Aparicio[1,2], Sohrab Shah[1,2,4]

[1]British Columbia Cancer Research Centre, Department of Molecular Oncology, Vancouver, Canada, [2]University of British Columbia, Department of Pathology and Laboratory medicine, Vancouver, Canada, [3]University of British Columbia, Department of Bioinformatics, Vancouver, Canada, [4]Memorial Sloan Kettering Cancer Centre, Department of Epidemiology and Biostatistics, New York, NY

*These authors contributed equally to this work

Cancer is an ecosystem of genetically diverse evolving clones, which emerge in time and space. The assessment of clonal dynamics in cancer and the changes that occur in heterogenous tumor population is fundamental to understand clinically important behaviour such as drug resistance and metastasis. The stochastic nature of mutation accumulation and subsequent selection dictates that the tumour growth process is inherently random. The identification of expanding or suppressing clones is potentially discernible through serial sampling of the populations over time which underpins the concept of clonal fitness, a key quantitative measure. However, the cancer field has lacked timeseries measurements to observe cancer evolution in real time by defining clonal population at the single cell copy number space. Our goal is to understand clonal fitness and develop experimental and mathematical framework for predicting the likely trajectories of clones in patients. By single cell whole genome sequencing and mathematical approaches to quantifying dynamics and fitness, we are mapping the relationships between clones and their fitness attributes and matching them with their resultant phenotypes. We present a method that enables a compelling measure of clonal fitness by directed evolution experiments and serial breast cancer patient derived xenografts permitting a real-time observation of cancer evolution.

We have adopted a new group of mathematical models based on Wright-Fisher diffusion approximations to estimate fitness coefficients of clones and create their temporal growth trajectories. Significantly, we establish the capacity of the model to predict clonal competition dynamics and engineered mixtures of clones with differential fitness coefficients, re-injected and serially passaged. We found a clone showing re-producible high fitness co- efficient in the growing tumor.

Quantitative reasoning about the future trajectory, dominance, and depletion of sub populations in a tumour will be invaluable in predicting patient response to treatment, including precision medicine.

A HIERARCHICAL RANDOM FOREST APPROACH FOR CELL TYPE PROJECTIONS ACROSS SINGLE-CELL RNAseq DATASETS

Yasin Kaymaz[1], Ming Tao[1], Nathan Lawless[2], Timothy Sackton[1]

[1]Harvard University, Informatics Group, Cambridge, MA, [2]Boehringer Ingelheim Pharma GmbH & Co KG, Computational Biology, Biberach an der Riss, Germany

The emergence of single-cell RNA sequencing (scRNAseq) has led to an explosion in novel methods to study biological variation among individual cells and to classify cells into biologically meaningful categories. These approaches have revealed novel cell types and previously unknown relationships between cell types and diseases. The power of scRNAseq has motivated an exponential increase in the scope of scRNAseq studies, including large scale cell atlas projects. Being able to utilize the rich information from such projects provides new dimensions to smaller-scale individual studies. Therefore, integration and accurate information transfer between existing cell atlases and newly generated, targeted data is a critical step for proper interpretation of biological features. While methods exist for cell type assignment across experiments, current approaches have limitations. Existing algorithms work well when the reference training data is composed of a few well-represented cell types, and when the query data contains few or no novel types and a good representation of known cell types. However, an ideal classification should be able to handle many candidate cell classes and not rely on a minimum input threshold of query data as some single-cell protocols produce low-throughput data in which rare cell types are represented with only a few cells.
Here, we present a new cell type projection tool based on hierarchical random forests that overcomes these limitations by using *a priori* information about cell type relationships for improved classification accuracy. We named our tool as **'HieRFIT'**, which stands for "Hierarchical Random Forest for Information Transfer". This novel classification algorithm takes as input a hierarchical tree structure representing the class relationships, along with the reference data. We use an ensemble approach combining multiple random forest models, organized in a hierarchical decision tree structure. We show that our hierarchical classification approach improves accuracy and reduces incorrect predictions. We use a scoring scheme that adjusts probability distributions for candidate class labels and resolves uncertainties while avoiding the assignment of cells to incorrect types by labeling cells at internal nodes of the hierarchy when necessary. Using HieRFIT, we re-analyzed publicly available scRNAseq datasets showing its effectiveness in cell type cross-projections with inter/intra-species examples. HieRFIT is implemented as an R package, freely available through GitHub (https://github.com/yasinkaymaz/HieRFIT).

COMPUTATIONAL PIPELINE TO INFER THE GENE REGULATORY NETWORK DURING CELL DIFFERENTIATION FROM scRNA-seq ANALYSIS

Kyung Dae Ko, Stefania Dell'Orso, Aster Juan, Vittorio Sartorelli

NIH, NIAMS, Bethesda, MD

Bulk transcriptomes have provided considerable insights and fostered the discovery and characterization of GRN(Gene Regulatory Network)s. However, bulk transcriptomes provide population-based averaged measurements which blur cell heterogeneity and developmental dynamics of asynchronous cell populations. Single-cell transcriptome technologies (scRNA-seq) capture cell heterogeneity and thus are useful for the discovery of cell populations, identification of cell mutants, and quantification of subpopulations.
Leveraging on the ability of generating thousands of individual measurements from scRNA-seq, lots of methods have been developed to capture spatial or temporal information from cell populations. For example, clustering and dimensionality reduction algorithms and pseudo-timing algorithms are implemented to identify unique transition paths among different cell states or by predicting future states of gene expression based on measurements of un-spliced and spliced transcripts. Using spatial or temporal information from clustering and trajectories, many algorithms have been successfully employed to infer GRNs.
However, their accuracy depends on the size of the network, methods of normalization, and the availability of annotated datasets. Here, we present an integrated pipeline for GRNs inference which uses clustering, temporal, and biological signatures extracted directly from scRNA-seq datasets. To evaluate its predictive power, we apply it to public datasets related to differentiate human pluripotent stem cells, embryonic stem cells, and mouse embryonic stem cells. The pipeline correctly identifies signaling pathways activated and dismantled at specific stages of cell differentiation and reveals the composition of gene hubs underlying discrete GRNs in each sample.

CELL-TYPE-SPECIFIC GENOMICS REVEALS HISTONE MODIFICATION DYNAMICS IN MAMMALIAN MEIOSIS

Gabriel Lam, Kevin Brick, Gang Cheng, Florencia Pratto, Dan Camerini-Otero

NIH, NIDDK, GBB, Bethesda, MD

Meiosis is the specialized cell division during which parental genomes recombine to create genotypically unique gametes. Despite its importance, mammalian meiosis cannot be studied *in vitro*, greatly limiting mechanistic studies. *In vivo*, meiocytes progress asynchronously through meiosis and therefore the study of specific stages of meiosis is a challenge. Here, we describe a method for isolating pure sub-populations of nuclei that allows for detailed study of meiotic sub-stages. Interrogating the H3K4me3 landscape revealed dynamic chromatin transitions between sub-stages of meiotic prophase I, both at sites of genetic recombination and at gene promoters. We also leveraged this method to perform the first comprehensive, genome-wide survey of histone marks in meiotic prophase, revealing a heretofore unappreciated complexity of the epigenetic landscape at meiotic recombination hotspots. Ultimately, this study presents a straightforward, scalable framework for interrogating the complexities of mammalian meiosis.

OXIDATIVE STRESS-INDUCED KLF4 ACTIVATES INFLAMMATORY RESPONSE THROUGH IL17RA AND ITS DOWNSTREAM TARGETS IN RETINAL PIGMENT EPITHELIAL CELLS

David W Li

Sun Yat-sen University, ZHongshan Ophthalmic Center, Guangzhou, China

Age-related macular degeneration (AMD) is a leading cause of irreversible blindness worldwide. Oxidative stress (OS), inflammation and genetics are considered the key pathogenic factors contribute to AMD development. Recent evidence shows that the pro-inflammatory interleukin 17 (IL17) signaling is activated in AMD patients and promotes disease pathogenesis. However, the interplay between OS and IL17 signaling, and the regulatory mechanism of IL17 pathway are largely unknown. OS-induced retinal pigment epithelial cell (RPE) damage causes both the initial pathogenesis of AMD and secondary degeneration of rods and cones. Healthy RPE is essential for ocular immune privilege, however, damaged RPE cells can activate inflammatory response. In the present study, we identified IL17RA, the principle receptor of IL17 signaling, is one of the most upregulated inflammatory genes in human RPE cells upon OS exposure. The prominent increase of IL17RA was also observed in RPE and retina of an AMD-like mouse model. Suppression of IL17RA in RPE cells prevented OS-induced RPE cell apoptosis and reduced the inflammatory response in both RPE and macrophages. Furthermore, we found that transcription factor KLF4 directly activates IL17RA expression, therefore, promotes the production of IL1β and IL8 in an IL17RA-dependent manner. In addition, the mRNA level of KLF4 isoform 2 was positively correlated with that of IL17RA in AMD patients. Together, our study demonstrates an unrevealed relationship between IL17RA and OS, and a new regulatory mechanism of IL17RA by KLF4 in RPE cells. These findings suggest that inhibition of IL17RA as a new potential therapeutic target for AMD through RPE protection and inflammatory suppression upon OS exposure.

ELIMINATING SPURIOUS CORRELATIONS FROM REGULARIZED SINGLE CELL RNA-seq DATA

Ruoyu Zhang, Gurinder S Atwal, Wei Keat Lim

Regeneron Pharmaceuticals Inc., Tarrytown, NY

Advances in single cell RNA-seq (scRNA-seq) technology and analyses have recently driven the development of data preprocessing methods, such as transcript abundance normalization and imputation, to address the numerous sources of substantial technical variability inherent in this technology. While these methods have been demonstrated to be effective in recovering individual gene expression, the suitability of these methods to the inference of gene-gene association and subsequent gene networks reconstruction have not been systemically investigated. In this study, we benchmarked five representative scRNA-seq normalization/imputation methods on human cell atlas bone marrow data in respect to their impact on inferred gene-gene associations. The benchmarked methods include global scaling normalization, regularized negative binomial regression-based normalization (NBR), and three widely used imputation methods MAGIC, DCA and SAVER.

We report widespread and significant inflation of gene expression correlations across the genome for all benchmarked preprocessing approaches, with the median Spearman's rank correlation coefficient as high as 0.839 (NBR). Furthermore, there was little overlap in the highly correlated gene pairs across the methods. These results reveal the extent to which regularization of single-cell transcriptomic can induce spurious correlations, resulting in erroneous inferences of molecular pathways and networks. We propose and implement a model-agnostic correction method that can effectively eliminate correlation artifacts whilst still effectively imputing individual gene expression. Agreement of the top correlated gene pairs among the methods, as well as their protein-protein interaction (PPI) enrichment, were significantly improved with the most dramatic change seen in NBR (PPI enrichment of the top 1000 gene pairs changes from <2% to >90%). Gene-gene association derived from the corrected data were further used to reconstruct gene co-expression networks and successfully reconstructed several known immune cell modules. Network topology analyses retrieved canonical biomarker genes for several known cell types (e.g. NKG7 for natural killer cells, CD79B for B cells), further demonstrating the need for correlation correction of regularized single cell RNA-seq data.

RISC: ROBUST INTEGRATION OF SINGLE-CELL RNA-seq DATASETS WITH DIFFERENT EXTENTS OF CELL CLUSTER OVERLAP

Yang Liu[1], Tao Wang[1,2], Deyou Zheng[1,3,4]

[1]Albert Einstein College of Medicine, Department of Genetics, Bronx, NY, [2]Albert Einstein College of Medicine, Department of Epidemiology and Population Health, Bronx, NY, [3]Albert Einstein College of Medicine, Department of Neurology, Bronx, NY, [4]Albert Einstein College of Medicine, Department of Neuroscience, Bronx, NY

Single cell RNA-seq (scRNA-seq) has remarkably advanced our understanding of cellular heterogeneity and dynamics in tissue development, diseases, and cancers. Integrated data analysis can often uncover molecular and cellular links among individual datasets and thus provide new biological insights, such as developmental relationship. Due to differences in experimental platforms and biological sample batches, the integration of multiple scRNA-seq datasets is challenging. To address this, we developed a novel computational method for robust integration of scRNA-seq (RISC) datasets using principal component regression (PCR). Because of the natural compatibility of eigenvectors between PCR model and dimension reduction, RISC can accurately integrate scRNA-seq datasets and avoid over-integration. Compared to existing software, RISC shows particular improvement in integrating datasets that contain cells of the same types (more accurately clusters) but at distinct functional states. To demonstrate the value of RISC in finding small groups of cells common between otherwise heterogenous datasets, we applied it to scRNA-seq datasets of normal and malignant cells and successfully identified small clusters of cells in healthy kidney tissues that may be related to the origin of renal tumors.

PROBING APOPTOSIS SIGNALING PROTEINS IN SINGLE LIVING CELLS ENABLES PRECISE EFFICACY EVALUATION OF ANTI-CANCER DRUGS

Yanrong Wen, Jia Liu, Zhen Liu

Nanjing University, School of Chemistry and Chemical Engineering, Nanjing, China

In vitro efficacy evaluation is critical for anti-cancer drug development and precision medicine. Conventional methods, such as MTT assay and clonogenic assay, which mainly rely on cell viability and large cell populations, suffer from apparent disadvantages, including signaling pathway-nonspecific, tedious, time-consuming, and so on. Herein, we present a single-cell approach called single cell plasmonic immunosandwich assay (scPISA) for precise anti-cancer efficacy evaluation. It allowed for facilely probing individual signaling proteins and protein-protein complexes in single living cells. Two key signaling proteins were proposed as apoptosis indicators, while three parameters reflecting the speed and strength of apoptosis were proposed as criteria for quantitative evaluation. Only a limited number (ca. 60) of cells was required for one whole set of assays. The evaluation was verified to be consistent among the indicators and parameters. Metformin, a potential anti-cancer drug, was then evaluated using this approach. Interestingly, metformin alone was found to be a less effective anti-cancer drug but its combination with actinomycin D dramatically improved the overall efficacy. This approach exhibited several unique strengths, including comprehensive and self-consistent evaluation, signaling pathway-specific, requiring a limited number of cells, simple procedure and speed. It holds great promise for personalized drug screening and precision medicine.

WHEN *FRANKENSTEIN* MEETS *VAN HELSING*: A WORKFLOW FOR snRNA/ATAC-seq FROM THE SAME NUCLEI PREP FOR 10X GENOMICS PLATFORM

Luciano G Martelotto, Fernando J Rossello, Sean M Grimmond

University of Melbourne, Centre for Cancer Research, Victoria Comprehensive Cancer Centre, Pathology, Melbourne, Australia

One of the challenges in single-cell genomic studies using solid tissues, is generating a high-quality single-cell suspension that preserves rare or difficult-to-dissociate cell types, and is free of both RNA degradation and artifactual transcriptional responses (e.g. stress). Single nuclei RNA-Seq (snRNA-Seq) has emerged as a powerful methodology to interrogate the transcriptome of thousands of nuclei isolated from fresh, frozen/archival and post-mortem tissues. snRNA-Seq overcomes the technical challenge of isolating intact single cells from complex tissues, reduces biased recovery of easily dissociated cell types, and minimises aberrant gene expression during whole-cell dissociation. While transcriptomic profiling enables the identification of functionally distinct cell types within complex tissues, the inclusion of epigenetic information can provide a more complete picture of how these expression profiles are regulated and maintained. Additionally, single-nucleus assay of transposase-accessible chromatin followed by sequencing (snATAC-seq) is a new technology that measures chromatin accessibility within each individual nucleus, yielding new insights into epigenetic regulation at the cellular level. snATAC-Seq generates complementary information on gene control and can uniquely reveal enhancer regions and regulatory logic. Droplet microfluidic technologies applied to the above-mentioned methodologies (e.g. Chromium™, 10x Genomics) is enabling researchers to profile thousands of nuclei at low cost and high throughput. While some profiling transcriptomes and chromatin accessibility from the same cell is possible, this is not yet available for microfluidics. One solution is to apply both methods to the same sample preparation; however, this can be challenging for small biopsies. Here, we introduce the *'Van Helsing'* protocol, an adaptation and extension of the *'Frankenstein'* protocol – originally developed for isolating nuclei from fresh, post-mortem and frozen small tissue biopsies for snRNA-Seq on the 10x Genomics platform – that allows the application of snRNA-Seq and snATAC-Seq on the *same* nuclei preparation. We show the step-by-step workflow, troubleshooting, practical examples and data integration results.

SINGLE NUCLEI RNA-seq ANALYSIS OF WHITE MATTER FROM HUMAN VASCULAR DEMENTIA

Daniel N Mitroi[1], Andrew J Lund[2], Brittany N Dugger[3], Charles DeCarli[4], William Lowry[2], Thomas S Carmichael[1]

[1]UCLA, Neurology, Los Angeles, CA, [2]UCLA, Molecular, Cell, and Developmental Biology, Los Angeles, CA, [3]UCD, Alzheimer's Disease Center, Davis, CA, [4]UCD, Neurology, Davis, CA

Vascular dementia is one of the most common causes of dementia after Alzheimer's Disease, and co-exists with Alzheimer's Disease in most cases (termed "mixed dementia"). Vascular dementia occurs as ischemic lesions in the subcortical white matter of the brain. These lesions expand radially over time into initially normally appearing white matter. Unlike the other major white matter disease in humans, multiple sclerosis, there is no process of repair in the ischemic white matter lesions that lead to vascular dementia. The cellular biology of the white matter injury and disease progression are poorly defined, in part because animal models do not replicate all aspects of the human disease. Our goal is to identify the molecular events that occur within and adjacent to white matter lesions in human vascular dementia, so as to determine molecular mechanisms that lead to disease progression and limit repair in human. For this we selected samples based on MRI scans of demyelinated areas of vascular dementia patients, after death and frozen storage, specimens were further assessed for RNA integrity. Only the samples with a suitable RNA integrity were further processed for myelin staining and then targeted microdissection of demyelinated area and adjacent white matter for nuclei isolation. Control, normal white matter was taken from the cerebellum. We selectively separated endothelial nuclei and all the remaining nuclei from these chosen areas and sent them for single nuclei RNA-seq.

MULTI-RESOLUTION SINGLE-CELL STATE CHARACTERIZATION VIA JOINT ARCHETYPAL/NETWORK ANALYSIS

Shahin Mohammadi*[1,2], Jose Davila-Velderrain*[1,2], Manolis Kellis[1,2]

[1]MIT, Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, [2]Broad Institute of MIT and Harvard, Cambridge, MA

*authors contributed equally

Dissecting the cellular heterogeneity embedded in single-cell transcriptomic data is challenging. Although a large number of methods and approaches exist, robustly identifying underlying cell states and their associations is still a major challenge; given the nonexclusive and dynamic influence of multiple unknown sources of variability, the existence of state continuum at the time-scale of observation, and the inevitable snapshot nature of experiments. As a way to address some of these challenges, here we introduce ACTIONet, a comprehensive framework that combines archetypal analysis and network theory to provide a ready-to-use analytical approach for multiresolution single-cell state characterization. ACTIONet uses multilevel matrix decomposition and network reconstruction to simultaneously learn cell state patterns, quantify single-cell states, and reconstruct a reproducible structural representation of the transcriptional state space that is geometrically mapped to a color space. A color-enhanced quantitative view of cell states enables novel visualization, prediction, and annotation approaches. Using data from multiple tissues, organisms, and developmental conditions, we illustrate how ACTIONet facilitates the reconstruction and exploration of single-cell state landscapes.

TRANSCRIPTOMIC ANALYSIS OF THE CELL RELEASED CIRCULATING EXOSOMES REVEALS BIOMARKERS OF THE MULTIPLE SCLEROSIS ACTIVITY.

Marcin P Mycko[1], Lukasz Paukszto[2], Jan P Jastrzębski[2], Krzysztof W Selmaj[1]

[1]University of Warmia and Mazury, Collegium Medicum, Department of Neurology, Laboratory of Neuroimmunology, Olsztyn, Poland, [2]University of Warmia and Mazury, Department of Plant Physiology, Genetics and Biotechnology, Olsztyn, Poland

Exosomes are a membrane vesicles released from the endocytic compartment of live cells that play an important role in cell-to-cell communication. The major contents of the exosomes are short RNAs that can interfere with the function of the acceptor cells. To profile the circulating exosomes transcriptome during multiple sclerosis (MS), a major autoimmune condition of the central nervous system, we have isolated exosomes and exosomal RNA from serum and urine of the MS patients and control subjects. Subsequently we have generated exosomal RNA libraries and processed them for the next generation sequencing analysis. Count RNA expression matrix was examined by two statistical methods, DESeq2 and edgeR. Both of them were applied to obtain the stringent results of differentially-expressed transcripts (DETs). Candidate DETs were scanned by Gene ontology (GO) and KEGG pathway databases using g.Profiler software. We have found that both serum and urine exosomes are a reach source of the shortRNA (< 300 nt) in MS patients both during relapse and remission. The sequences have been grouped into the following categories: CDBox, HAcaBox, RefSeq, lincRNA, lincRNA, miRNA, other_ncRNA, rRNA, piRNA, rfam, scaRNA, tRNA and tRNA_like. Among the 50 most significant DETs in remission versus controls only 3 categories of RNA have been found (Refseq, 80%; tRNA, 10% and lincRNA, 10%). Whereas in the top 50 most significant DETs in relapse versus control additionally to Refseq, tRNA and linRNA we have found 12% of rfam transcripts and 12% of rRNA. Many of these DETs have been derived from transcription factors coding genes. Interestingly almost third of urine exs DETs between MS vs controls (n=39) overlapped with DETs in serum. The RNA sequences differentially expressed were validated with qPCR of 48 serum and urine samples from RRMS patients during relapse and 48 in remission. We have identified and validated four serum exosomal microRNA (miRNA) that were differentially expressed during relapse in comparison to remission. Interestingly, these miRNA have been also differentially secreted within the exosomes by PBMCs from RRMS patients and from controls. Thus circulating exosomes transcripts showed significant difference in the pattern of expression between MS both relapse and remission vs. controls. These findings suggest disturbance in cell-to-cell communication in MS.

INTEGRATIVE SINGLE-CELL ANALYSIS TO DISSECT HEART FAILURE BIOLOGY

Seitaro Nomura[1,2], Issei Komuro[2], Hiroyuki Aburatani[1]

[1]Research Center for Advanced Science and Technology, The University of Tokyo, Genome Science Laboratory, Tokyo, Japan, [2]The University of Tokyo Hospital, Department of Cardiovascular Medicine, Tokyo, Japan

Single-cell analysis has comprehensively elucidated cell classification, cell trajectory, spatial heterogeneity, gene regulatory network, and cell-cell communication, in development, physiology, and disease at the single-cell and molecular levels. The heart consists of various cell types, including cardiomyocytes, fibroblasts, endothelial cells, and immune cells, and is exposed to hemodynamic stress. The heart shows compensated hypertrophy in response to pathological overload, and persistent and excessive overload collapses cardiac homeostasis, leading to heart failure, a primary cause of death worldwide. We conducted integrative analysis of single-cell RNA-seq and multiplexed single-molecule RNA in situ hybridization of the heart from heart failure model mice and human heart failure patients to reveal the transcriptional signature and spatial heterogeneity involved in cardiac phenotypes in mice and human, reconstructed a cardiomyocyte trajectory during heart failure progression, and identified a bifurcation into adaptive and failing cardiomyocytes. We also generated the cell-cell communication map during heart failure progression and found that cardiac fibroblasts HtrA3 degraded TGF-β1 to prevent TGF-β signaling activation cycle between cardiac fibroblasts and cardiomyocytes. Single-cell network analysis revealed that cardiomyocytes from HtrA3 knockout mice with pressure overload showed DNA damage, p53, and TGF-β signaling activation. We further generated cardiomyocyte-specific p53 knockout mice to show that p53 is essential for the induction of failing cardiomyocytes. We also classified human cardiomyocytes into three pathological states, integrated with clinical phenotypes, and developed a molecular pathology analysis pipeline to quantitatively assess the extent of DNA damage response and predict clinical prognosis and treatment response in human heart failure. In addition, we have not only developed a in vivo perturb-seq pipeline to comprehensively analyze functions of multiple genes and elements in the heart, but also established a method to simultaneously trace both lineages (barcodes) and phenotypes (transcriptomes) at the organism-level in mice. Collectively, we demonstrate that integrative single-cell analysis can resolve heart failure biology.

INTEGRATIVE COMPUTATIONAL FRAMEWORK FOR LINKING CELL SURFACE PROTEINS TO DOWNSTREAM TRANSCRIPTIONAL PROGRAMS IN SINGLE CELLS

Hatice U Osmanbeyoglu

University of Pittsburgh, Biomedical Informatics, Pittsburgh, PA

Complex signaling and transcriptional programs control the development and functionality of specialized cell types. Recently multi-omics single-cell technologies have been developed to analyze transcriptional and chromatin states and coupling them with expression of cell surface proteins. However, to date they have not been used to link the signaling and transcriptional states of individual cells. Another challenge with such high dimensional noisy data sets is the extensive amount of missing data inherent in single-cell sequencing. In this study, we developed a novel computational framework to address these challenges, thereby enabling biologically meaningful and actionable inferences to be drawn from such datasets. More specifically, we presented a statistical method to link upstream cell surface signaling to downstream transcriptional response by exploiting a cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq) datasets with cis-regulatory information. CITE-seq allows to profile surface protein expression, and RNA measurements from the same cell by first staining with DNA-barcoded antibodies and then performing single cell RNA-seq. Formally, we used an algorithm called affinity regression to learn an interaction matrix between upstream surface proteins and downstream transcription factors (TFs) that explains target gene expression. The trained model can then predict the TF activity, given a cell's protein expression profile, or infer the surface protein activity, given a cell's gene expression profile. We applied our strategy to existing CITE-seq dataset of cord blood mononuclear cells (CBMCs) and peripheral blood mononuclear cells (PBMCs) and identified shared and cell type-specific roles for TF/surface proteins. For instance, we identified TF regulators such as GATA3 and RUNX2 for T-cells. Indeed, these TFs are known to be associated with activation of cytotoxic T cells and CD8 T-cell development. Importantly, we did not observe those differences at the gene expression level, which does not incorporate upstream surface-protein expression and downstream TF target-gene expression. Our framework provides a proof of principle for broader, larger-scale studies at the single cell level aiming to link upstream cell surface signaling to downstream transcriptional programs for diseases such as cancer, as well as other applications (e.g., developmental biology and immunology).

COUPLING FLUIDFM TECHNOLOGY WITH MULTI-OMIC ANALYSES FOR UNDERSTANDING THE EFFECT OF PI5P ON CHROMATIN

Aaron Pantoja, Sarah Kreuz, Wolfgang Fischle

King Abdullah University of Science and Technology, Bioscience, Thuwal, Saudi Arabia

Different studies have established correlative links between cellular condition, nuclear phospholipid levels and chromatin state. However, due to a lack of experimental tools for measuring and perturbing phospholipids in nuclear context our understanding of the biological role these cellular intermediate molecules might play in genome regulation is still scarce. In vitro studies have demonstrated interactions between various phospholipids and different chromatin components suggesting a putative physiological role in the regulation of chromatin structure and function.
In particular, phosphatidylinositol 5-phosphate (PI5P) has been described by others and us to interact with multiple epigenetic regulators like UHRF1, ING2, ATX1, SAP30, and TAF3. It is thought that PI5P exerts signaling functions regulating the activity of the nuclear factors and thereby potentially affecting the overall chromatin landscape. Although in vitro studies provide insights into how PI5P might regulate nuclear target proteins, comprehending the impact PI5P might have on chromatin structure and function, requires developing experimental tools for cellular studies. We will present our efforts in establishing complementary approaches to this end. On a global and ensemble level, we are setting up cell lines where the activity of nuclear enzymes metabolizing PI5P can be tightly controlled. This is paired with omics readout for pools of and single cells of different aspects of chromatin state (i.e. accessibility, methylation). On a single cell level, we are applying FluidFM (Cytosurge) technology to directly inject PI5P and other phospholipids into the nucleus of cells. After isolation and recovery of single cells, this is paired with multi-omics analysis for determining multiple chromatin parameters in parallel.

DECODING THE IDENTIFY AND FLUX OF CARDIAC CELLS IN INJURY AND HOMEOSTASIS AT SINGLE-CELL RESOLUTION

Ralph Patrick[1,2], Nona Farbehi[1,3], Munira Xaymardan[1,4], Aude Dorison[1], Kitty Lo[5], David Humphreys[1], Vaibhao Janbandhu[1], Joshua WK Ho[6], Robert Nordon[7], Richard P Harvey[1,2]

[1]Developmental and Stem Cell Biology Division, Victor Chang Cardiac Research Institute, Sydney, Australia, [2]St Vincent's Clinical School, UNSW, Sydney, Australia, [3]Garvan Weizmann Centre for Cellular Genomics, Garvan Institute of Medical Research, Sydney, Australia, [4]School of Dentistry, Faculty of Medicine and Health, University of Sydney, Sydney, Australia, [5]School of Mathematics and Statistics, University of Sydney, Sydney, Australia, [6]School of Biomedical Sciences, The University of Hong Kong, Hong Kong, Hong Kong, [7]Graduate School of Biomedical Engineering, UNSW, Sydney, Australia

Cardiovascular disease, which includes myocardial infarction (MI), is the leading cause of death worldwide. After MI, a billion cardiomyocytes (CM) may die and will not be replaced, with the infarcted area healed by replacement fibrosis. However, insights into cardiac repair mechanisms in mammals and lower animals has spurred interest in developing heart regeneration strategies for humans. Cardiac injury triggers a complex cascade of cellular and molecular events that control the injury response. Understanding cardiac sub-lineages and their response to injury, including the communication and signaling networks that control the injury response, will be critical for developing strategies to augment cardiac repair and achieve heart regeneration in humans.

For this study we have applied single-cell RNA-seq (scRNA) to investigate the impact of MI on adult mouse hearts and the pro-reparative effect of treating mice with platelet-derived growth factor (PDGF)-AB following injury. We have profiled over 75,000 cardiac single-cell transcriptomes from the total non-CM fraction, as well as enriched (Pdgfra-GFP^{+}) fibroblast lineage cells, from adult mouse hearts at 3- and 7-days post-sham or MI surgery, and additionally the non-CM fraction with PBS or PDGF-AB treatment. Clustering revealed ~30 cell sub-populations, comprising both known and novel cell types, which showed significant flux across the injury time-points. Within the fibroblast lineage we observed non-linear dynamics of activation, proliferation and differentiation to myofibroblasts. Analysis of cells from PDGF-AB treated mice revealed a cell type-specific response to treatment including an increased biosynthetic profile and a prominent shift in gene usage by activated fibroblasts and myofibroblasts. Together, these data provide an unparalleled, high-resolution map of the cellular heterogeneity within the interstitial compartments of the heart, their response to MI and PDGF-AB treatment following injury.

GLUER: INTEGRATIVE ANALYSIS OF MULTI-OMICS DATA AT SINGLE-CELL RESOLUTION

Tao Peng[1], Kai Tan[1,2]

[1]Children's Hospital of Philadelphia, Division of Oncology and Center for Childhood Cancer Research, Philadelphia, PA, [2]University of Pennsylvania, Department of Pediatrics, Philadelphia, PA

Single-cell RNA-Seq has been developed to identify cell types in heterogeneous tissues. Furthermore, recent advancement of genome-wide assays for measuring DNA methylation and chromatin accessibility in thousands to millions of single cells provides more information regarding cellular types/states. However, it is a challenging to integrate those different types of single-cell omics data. Here we develop a new algorithm for inteGrative anaLysis of mUlti-omics at single-cEll Resolution (GLUER). We tested GLUER using the datasets generated by SNARE-seq (Song Chen et al, Biorxiv 2019), which was taken as the ground truth. Our results demonstrate that GLUER has promising performance in terms of the accuracy of matching cells with different data modalities. We applied GLUER to analyze single-cell RNA-Seq and single-cell ATAC-Seq data of peripheral blood mononuclear cells. Our result show that GLUER can identify new cell clusters that both single cell RNA-Seq and single cell ATAC-Seq miss. GLUER provides a principled analytical framework for studying the heterogeneity of cell populations using multi-omics data.

Reference:
Chen, Song, Blue Lake, and Kun Zhang. "Linking transcriptome and chromatin accessibility in nanoliter droplets for single-cell sequencing." Biorxiv (2019): 692608.

CELL LINEAGE INFERENCE FROM DROSOPHILA BLASTODERM GENE EXPRESSION

Jung Min Han, Vipul Periwal

Laboratory of Biological Modeling, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD

Single cell RNA-sequencing (scRNA) data can be a starting point for investigating cell differentiation and development, i.e., how cell phenotypes and changes are related to gene expression. Estimating biological time from scRNA data is therefore an interesting modeling problem, and has been studied extensively. There are many algorithms that aim to achieve this goal, and each of them comes with pros and cons. In this study, we present comparisons of these algorithms applied to a controlled developmental data set, albeit one with a much smaller set of measured gene expression values with much better precision in the values than presently available in scRNA data. Fowlkes et al. (Cell, 2008) published FISH-based expression levels of a total of 99 genes in 6078 Drosophila blastoderm cells at six developmental stages that spanned 50 minutes prior to the onset of gastrulation. The data set has some missing values, which were imputed through a supervised learning method. Using the Python package NeuroLab, we trained and tested a set of neural networks to fill in the missing data. We then used pseudotime estimation algorithms to order the cells along a differentiation trajectory and assign pseudotime values to each cell, in order to determine where the cell is positioned along the trajectory. To evaluate the performance of the algorithms, the inferred cell lineage was compared to the true developmental stage of the cells.

SINGLE-CELL ANALYSIS IDENTIFIED A LEF-1 POSITIVE DRUG-RESISTANT SUBPOPULATION IN LUMINAL SUBTYPE OF BREAST CANCER CELLS.

Marta Prieto-Vila[1,2], Wataru Usuba[2,3], Ryou-u Takahashi[2,4], Iwao Shimomura[2], Takahiro Ochiya[1,2], Yusuke Yamamoto[2]

[1]Tokyo Medical University, Department of Molecular and Cellular Medicine, Tokyo, Japan, [2]National Cancer Center Research Institute, Division of Cellular Signaling, Tokyo, Japan, [3]St Marianna University School of Medicine, Department of Urology, Kanagawa, Japan, [4]Hiroshima University, Department of Cellular and Molecular Biology, Hiroshima, Japan

Breast cancer is the leading type of cancer and the second cause of cancer-related death in women worldwide. Despite the continuous progresses in breast cancer treatment, a subset of patients present drug resistance. Surviving cells lead to tumor recurrence and metastasis, which remains the main cause of cancer-related mortality. In addition, breast cancer cells are highly heterogeneous, hindering the identification of individual cells with the capacity to survive to anticancer treatment. To address this issue, we performed an extensive single-cell qPCR profiling to a luminal subtype of breast cancer cell line and several derivatives, including docetaxel-resistant cells. Here we show that the upregulation of EMT and stemness-related genes and the downregulation of cell cycle–related genes were observed in the drug-resistant cells, suggesting a cancer stem cell-like phenotype. This docetaxel-resistance gene profile, which included 27 genes highly correlated among them at single-cell level, was mainly regulated by *LEF1*. Furthermore, those genes were translated to protein and their expression was confirmed by immunostaining *in vitro* and tissue samples. The small molecule Quercetin, a well-known flavonoid, was able to target Lef1 and decrease the overall expression of docetaxel-resistance gene profile, sensitizing the cells to docetaxel again. Interestingly, a small number of cells in the parental population exhibited a gene-expression profile similar to that of the drug-resistant cells, indicating that the untreated parental cells already contained a rare subpopulation of stem-like cells with an inherent predisposition. Our data suggest that during chemotherapy, this population may be positively selected, leading to treatment failure.

DEFINING MULTISTEP CELL FATE DECISION PATHWAYS DURING PANCREATIC DEVELOPMENT AT SINGLE-CELL RESOLUTION

Xin-Xin Yu*[1,2], Wei-Lin Qiu*[1,3], Liu Yang[1,2], Yu Zhang[1,2], Mao-Yang He[1,3], Lin-Chen Li[1,2], Cheng-Ran Xu†[1]

[1]Peking University, College of Life Sciences, Peking-Tsinghua Center for Life Sciences, Beijing, China, [2]Peking University, Academy for Advanced Interdisciplinary Studies, Beijing, China, [3]Peking University, PKU-Tsinghua-NIBS Graduate Program, Beijing, China

The generation of terminally differentiated cell lineages during organogenesis requires multiple, coordinated cell fate choice steps. However, this process has not been clearly delineated, especially in complex solid organs such as the pancreas. Here, we performed single-cell RNA-sequencing in pancreatic cells sorted from multiple genetically modified reporter mouse strains at embryonic stages E9.5–E17.5. We deciphered the developmental trajectories and regulatory strategies of the exocrine and endocrine pancreatic lineages as well as intermediate progenitor populations along the developmental pathways. Notably, we discovered previously undefined programs representing the earliest events in islet α- and β- cell lineage allocation as well as the developmental pathway of the "first wave" of α-cell generation. Furthermore, we demonstrated that repressing ERK pathway activity is essential for inducing both α- and β-lineage differentiation. This study provides key insights into the regulatory mechanisms underlying cell fate choice and stepwise cell fate commitment and can be used as a resource to guide the induction of functional islet lineage cells from stem cells *in vitro*.

*These authors contributed equally to this work.
†Corresponding author. Tel: +86 10 6275 7119; E-mail: cxu@pku.edu.cn

SCCONSENSUS: AN APPROACH COMBINING SEMI-SUPERVISED AND UNSUPERVISED CLUSTERING FOR CELL TYPE ANNOTATION IN SINGLE-CELL RNA-seq DATA

Bobby Ranjan, Wenjie Sun, Jinyu Park, Nirmala A Rayan, Joanna Tan, Xingliang Liu, Shyam Prabhakar

Computational and Systems Biology, Genome Institute of Singapore, A*STAR, Singapore, Singapore

Although cell type annotation is a crucial step in scRNA-seq data analysis, the single-cell community has yet to agree upon a single methodology. Semi-supervised clustering methods like *RCA*[1] and *SingleR*[2] robustly detect major cell types by correlating reference bulk transcriptomic datasets of sorted cell types with single-cell data. However, these methods are limited by the resolution of their reference data sets and cannot annotate intermediate cell states and case-specific phenotypes. De-novo clustering methods, implemented in packages like *Seurat*[3] and *Scanpy*[4], overcome these limitations. Despite these advantages, they are more susceptible to batch and quality variation, occasionally producing heterogeneous cell clusters.

We utilise the merits and account for the demerits of both methodologies in *scConsensus*, a hybrid approach to obtain cell type annotation using a consensus of semi-supervised and unsupervised clustering. *scConsensus* uses the reference-based *RCA* to identify known cell types, simultaneously implementing de-novo clustering using *Seurat* on the same dataset. Both clustering results are then manually annotated using the strongest correlated reference cell type for *RCA* and marker gene expression for *Seurat*. In *RCA*, clusters that do not correlate with a unique reference, are case-specific or are expected to be subtypes of a reference cell type are annotated using differential expression of marker genes of their corresponding *Seurat* clusters. Conversely, heterogeneous *Seurat* clusters are reassigned using their best correlating *RCA* reference cell types. A consensus cluster annotation is thus obtained, where a unique cell type is assigned to each cell. *scConsensus* then looks for differentially expressed (DE) genes between every pair of cell types, and uses the union set of DE genes to cluster the cells using hierarchical clustering.

We evaluated the performance of *scConsensus* on 12 scRNA-seq (UMI and SmartSeq2) datasets. We show that *scConsensus* consistently outputs more homogeneous clusters compared to *Seurat* or *RCA*. Furthermore, we used 6 CITE-seq datasets to show that *scConsensus*' results correlate better with antibody-based cell type annotations. Thus, we conclude that *scConsensus* is a robust methodology to detect known cell types and characterise new cell types in scRNA-seq data.

[1] Li, H. et al. Nature Genet. 49.5(2017):708.
[2] Aran, D. et al. Nature Immunol. 20.2(2019):163.
[3] Satija, R. et al. Nature Biotechnol. 33.5(2015):495.
[4] Wolf, F. A. et al. Genome Biol. 19.1(2018):15.

SINGLE NUCLEI RNA SEQUENCING IN NON-DISEASED AND DISEASED HUMAN CARDIAC TISSUE

Daniel Reichart, Daniel Delaughter, Josh Gorham, Barbara McDonough, Jonathan Seidman, Christine Seidman

Harvard Medical School, Department of Genetics, Boston, MA

Cardiomyopathies are a clinically heterogeneous group of diseases which comprise structural and functional changes in the cardiac muscle. Dilated cardiomyopathy (DCM) and Hypertrophic/Hypertrophic-obstructive cardiomyopathy (HCM/HOCM) belong to primary cardiomyopathies, which may have genetic causes and are of specific interest, since 5-year survival rate can be still as low as 50%.
The myocardium contains different types of cells including cardiomyocytes, which are of particular interest due to their substantial contribution to cardiac work and mutational content in primary cardiomyopathies. Bulk RNAseq often masks the uniqueness of each cells, which was initially successfully hurdled by using Single-Cell RNAseq (scRNAseq) resulting in new insights into the heterogeneity and subtypes of the myocardium cell population.
However, reaching out for analyses of human cardiac tissues, scRNAseq is not sufficient due to its need of viable and intact cells. Furthermore, the most common used analyses systems (10X Genomics) have cell size restrictions of 25-30µm, which excludes cardiomyocytes from any scRNAseq experiment due to their cylindrical shape with 100µm length and 25µm width. By using the Single-Nuclei RNAseq approach, these just mentioned problems can be overcome with subsequent usage of frozen and postmortem cardiac tissue of DCM and HCM/HOCM patients and analyses of single cardiac cells including cardiomyocytes. With this technique we are aiming to establish a single cell/nuclei atlas of non-diseased and diseased cardiac tissue to receive further insights into the gene expression pattern of wildtype and end-stage heart disease.

DEVELOPMENTAL MAP OF OTIC MESENCHYME CELL POPULATIONS OF THE INNER EAR

Kevin Rose[1,2], Beatrice Milon[1,2], Likhitha Kolla[3], Michael Kelly[3], Matthew Kelley[3], Ronna Hertzano[1,2,4]

[1]University of Maryland School of Medicine, Institute for Genome Sciences, Baltimore, MD, [2]University of Maryland School of Medicine, Otorhinolaryngology Head and Neck Surgery, Baltimore, MD, [3]National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Laboratory of Cochlear Development, Baltimore, MD, [4]University of Maryland School of Medicine, Anatomy and Neurobiology, Balt, MD

Hearing impairment is a large and growing healthcare problem that negatively impacts quality of life by impairing communication. Hearing is dependent on the conversion of mechanical sound waves into electrochemical information, which is performed by specialized mechanoreceptors found within the cochlea called hair cells. Electrical information is then relayed from the cochlea to the brainstem and to the rest of the ascending auditory pathway via the spiral ganglion neurons of the auditory nerve. This complex connectivity enables the precise encoding and processing of sound stimuli. While many studies have investigated the roles of these critical cell types in hearing, the important roles of other cell types in the inner ear are not as fully characterized or appreciated. The role of one such cell type found throughout the cochlea- the otic mesenchyme cells- is not yet fully characterized. Mutations in otic mesenchymal cells cause profound deafness, clearly indicating an important functional role for these cells in the development of the inner ear. In this study, we aim to better understand the functional importance of otic mesenchymal cells of the cochlea within the inner ear developmental framework using single cell RNA sequencing (scRNA-seq). Cochleae from mice at embryonic days 10, 12, 14, and 16, and postnatal days 1, 2, and 7 were procured and analyzed by scRNA-seq. Mesenchymal cells were identified based on the expression of *Pou3f4* and the lack of expression of *Epcam*. Cells were then sub-clustered to reveal multiple mesenchyme cell populations, each with unique molecular markers. Pseudotime analysis was performed at all time-points to understand otic mesenchyme fate trajectories during cochlear maturation. Understanding this abundant cell type of the inner ear will give insights into key developmental programs during cochlear maturation.

ANALYTICAL METHODS TO IDENTIFY TUMOR HETEROGENEITY AND RARE SUBCLONES IN SINGLE CELL DNA SEQUENCING DATA FROM TARGETED PANELS

Sombeet K Sahu, Manimozhi Manivannan, Shu Wang, Dong Kim, Saurabh Gulati, Nianzhen Li, Adam Sciambi, Nigel Beard

Mission Bio, Bioinformatics, So. San Francisco, CA

With the advancements of single cell sequencing technologies it is now possible to interrogate thousands of cells in a single experiment. ScRNA-Seq has been available for several years but high-throughput single-cell DNA analysis is in its infancy. To address these challenges and enable the characterization of genetic diversity in cancer cell populations, we developed a novel approach to identify mutation signatures which define subclones present in a tumor population.

Here we present subclone identification method using data generated on the Tapestri platform and analyzed by Tapestri analytical workflow. The variant-cell matrix is then subjected to identification of subclones. Top variants defined the signature of each subclone are also identified. To validate our methodology, we used two different targeted sequencing panels on model systems with known truth mutations. Our pipeline shows the distinct clusters correlating with titration and cell line ratios. Cluster associated signature mutations were also identified. Our approach addresses key issues of identifying rare subpopulations of cells down to 0.1%, and transforms the ability to accurately characterize clonal heterogeneity in tumor samples. This high throughput method advances research efforts to improve patient stratification and therapy selection for various cancer indications.

A NEUTROPENIA-ASSOCIATED TRANSCRIPTION FACTOR MUTATION DIFFERENTIALLY IMPACTS TARGET GENES IN THE CELL-STATES TRAVERSED DURING GRANULOCYTE SPECIFICATION AND COMMITMENT

David E Muench[1,2], Kyle Ferchen[1,3], Giang Pham[4], Somchai Chutipongtanate[3,5], Pankaj Dwivedi[3], Stuart Hay[6], Kashish Chetal[6], Lisa R Trump-Durbin[7], Jayati Mookerjee-Basu[8], Kejian Zhang[9], Carolyn Lutzko[7], Kasiani C Myers[10], Kristopher L Nazor[11], Kenneth Greis[3], Dietmar J Kappes[8], Sing Sing Way[4], H. Leighton Grimes[1,7], Nathan Salomonis[6]

[1]Cincinnati Children's Hospital Medical Center, Division of Immunobiology and Center for Systems Immunology, Cincinnati, OH, [2]Cincinnati Children's Hospital Medical Center, Molecular and Developmental Biology Graduate Program, Cincinnati, OH, [3]University of Cincinnati, Department of Cancer Biology, Cincinnati, OH, [4]Cincinnati Children's Hospital Medical Center, Division of Infectious Diseases, Cincinnati, OH, [5]Mahidol University, Pediatric Translational Research Unit, Department of Pediatrics, Faculty of Medicine Ramathibodi Hospital, Bangkok, Thailand, [6]Cincinnati Children's Hospital Medical Center, Division of Biomedical Informatics, Cincinnati, OH, [7]Cincinnati Children's Hospital Medical Center, Division of Experimental Hematology and Cancer Biology, Cincinnati, OH, [8]Fox Chase Cancer Center, Philadelphia, PA, [9]Cincinnati Children's Hospital Medical Center, Division of Human Genetics, Cincinnati, OH, [10]Cincinnati Children's Hospital Medical Center, Division of Bone Marrow Transplantation and Immune Deficiency, Cincinnati, OH, [11]BioLegend, BioLegend, San Diego, CA

Advances in genetics and sequencing have lead to a deluge of disease-associated and disease-causing genetic alterations. Resolving causality between genetics and disease requires generating accurate models for molecular dissection; however, the rapid expansion of single-cell landscapes presents a major challenge to accurate comparisons between mutants and their wild type equivalents. Here, we generated the first mouse models of human severe congenital neutropenia (SCN) using patient-derived mutations in the Growth factor independent-1 (GFI1) transcription factor. To delineate the impact of SCN mutations, we first generated single-cell references for granulopoietic genomic states with linked epitopes (CITE-Seq), then developed a new computational approach (cellHarmony) to align mutant cells to their wild-type equivalent and derive differentially expressed genes. Surprisingly, the majority of differentially expressed GFI1-target genes are sequentially altered as cells traverse successive states. Using an integrative single-cell RNA- and ATAC-Seq approach, we find dysregulation of chromatin accessibility during early specification steps which propagate into commitment. Moreover, genes associated with commitment and innate immune functional programming were conversely the epigenome surrounding genes normally activated during commitment was impaired. These cell-state-specific insights facilitated genetic rescue of granulocytic specification but not post-commitment defects in the expression of innate-immune effectors, providing regulatory insights into granulocyte dysfunction.

ESTABLISHMENT OF HUMAN INDUCED PLURIPOTENT STEM CELL DERIVED DOPAMINERGIC NEURONS FOR IN VITRO MODELING OF PARKINSON'S DISEASE PATHOBIOLOGY.

Harry Samaroo[1], Evgeny Shlevkov[2], Joost Groot[1], Patrick Cullen[1], Ravi Challa[1], Kejie Li[1], Sandra J Engle[1], Andreas Weihofen[2], Warren Hirst[2], Birgit Obermeier[1], Tom Lanz[1], Robin Kleiman[1]

[1]Biogen Inc., Translational Biology, Cambridge, MA, [2]Biogen Inc., Neuromuscular and Movement Disorders Research, Cambridge, MA

Parkinson's disease (PD) is characterized by pathologic deposits of α-synuclein (α-SYN), and progressive degeneration of dopaminergic (DA) neurons that project from the substantia nigra (A9 DA neurons). A translational model enriched for this vulnerable population of DA neurons could be useful for studying disease relevant phenotypic endpoints. We characterized induced pluripotent stem cells (iPSCs) differentiated into DA neurons based on a modification of the protocol published by Kriks et al., 2011 as well as DA neurons available from a commercial source. We evaluated select disease relevant endpoints such as mitochondrial health and function in resulting DA neurons derived from wildtype or A53T α-SYN mutant iPSC. The efficiencies of both differentiation protocols were evaluated using immunocytochemistry for key markers and single cell RNAseq from triplicate samples. Approximately half of the cells generated with the modified Kriks differentiation protocol stained with neuronal markers, while DA neurons from a commercial source were >80% positive for the same markers. A subset of the neurons from both protocols also stained positive for tyrosine hydroxylase, an established marker of mature DA neurons and the rate limiting enzyme in the dopamine metabolic pathway. Single cell RNAseq profiling (10X Genomics) revealed significant heterogeneity of the cellular population. Approximately 9% of cells expressed mRNA for KCNJ6, which is a marker characteristic of A9 DA neurons. Similar profiling of commercially available DA neurons revealed higher purity of neuronal cell types, but heterogeneity was also observed with 8% of the overall cells expressing mRNA for KCNJ6. These data on human iPSC-derived models of DA neurons will serve as a benchmark for producing DA neuron culture preparations enriched in A9 DA neurons that will be needed to enable translational models of PD pathobiology.

TOOMANYCELLS IDENTIFIES RARE TARGETED-THERAPY-RESISTANT T-ALL CELLS.

Gregory W Schwartz[1], Yeqiao Zhou[1], Jelena Petrovic[1], Robert B Faryabi[1,2]

[1]University of Pennsylvania Perelman School of Medicine, Pathology and Laboratory Medicine, Philadelphia, PA, [2]University of Pennsylvania Perelman School of Medicine, Abramson Family Cancer Research Institute, Philadelphia, PA

Introduction: Observation of gain-of-function Notch mutations in ~50% of T-cell acute lymphoblastic leukemia (T-ALL) patients led to consideration of therapies such as gamma-secretase inhibitors (GSIs) to target oncogenic Notch. However, GSI treatment often leads to a transient response and drug resistance. Current lack of effective single-cell visualization algorithms to elucidate rare populations hinder our ability to study potential rare pre-resistant subclones of T-ALL.
Hypothesis: There exists a rare GSI-resistant subclone in drug-naïve T-ALL with a transcriptomic profile similar to resistant cells.
Methods: To understand the underlying mechanisms of GSI-resistance in T-ALL, we developed TooManyCells, a suite of new algorithms and tools for delineation and visualization of cell clade relationships. The TooManyCells novel visualization model is distinct from and outperforms dimensionality reduction and clustering methods in rare cell identification. To identify GSI-resistant cell clade relationships, we applied TooManyCells to genetically homogeneous drug-naïve T-ALL cells and as well as GSI-resistant cells selected under two distinct treatment regimens of long-term ascending-dose or sustained high-dose GSI.
Results: Rigorous controlled purity and rare population benchmarking showed that TooManyCells outperformed all other tested clustering and visualization algorithms in detecting rare cells. Applying TooManyCells to GSI-sensitive and -resistant T-ALL cells showed that ascending resistant cells were markedly more heterogeneous and were transcriptionally distinct from sustained resistant cells. Compared to ascending cells, expression of *MYC* and its known targets were significantly higher in sustained resistant cells, implying different evolutionary pathways to resistance. Importantly, TooManyCells identified a rare subclone of drug-naïve cells that resembled sustained GSI-resistant cells. These pre-resistant cells had *MYC* levels comparable to resistant cells and significantly higher than other GSI-sensitive cells. Independent single molecule RNA fluorescence in situ hybridization verified the presence of high *MYC*-expressing drug-naïve cells.
Conclusion: We have identified a putative pre-resistant T-ALL population using TooManyCells, a suite of new tools for clear visualization for cell clade relationships. TooManyCells revealed two distinct pathways for genetically similar T-ALL cells to acquire GSI resistance. Given that ascending and sustained regimens usually correspond to different phases of clinical trials, our results suggest the importance of dosage selection in developing targeted therapies.

APPLYING SINGLE CELL ANALYSIS TECHNIQUES TO CHARACTERIZE THE ROLE OF MAIT CELLS IN ULCERATIVE COLITIS

Siddhartha Sharma[1], Louis Gioia[2], Brian Abe[1], Marie Holt[1], Anne Costanzo[1], Lisa Kain[1], Andrew Su[2], Luc Teyton[1]

[1]The Scripps Research Institute, Department of Immunology and Microbiology, La Jolla, CA, [2]The Scripps Research Institute, Department of Integrative Structural and Computational Biology, La Jolla, CA

Ulcerative colitis (UC) is an increasingly prevalent form of inflammatory bowel disease, that is characterized by chronic T cell mediated inflammation of colonic mucosa. UC is likely caused by a confluence of immune dysregulation, genetics, gut microbiome dysbiosis, and environmental factors. A sub-population of CD8+ T cells, the mucosal associated invariant T cells (MAITs), may "link" microbiome dysbiosis, environmental factors and chronic colonic inflammation. MAITs are uniquely suited for this role as they recognize microbial derived vitamin B2 based antigens and are a significant population in homeostatic mucosa. MAITs are likely a heterogenous population and a subset of MAITs may contribute to the chronic inflammation in UC. To assess the role of MAIT cells in UC, we have analyzed MAITs isolated from matched colonic biopsies, both uninflamed and inflamed mucosa, and blood from a cohort of UC patients. Our approach combines MR1 tetramers, to specifically identify MAIT cells, and flow cytometry with single cell T cell receptor sequencing of both alpha and beta chains, and single cell qPCR of the same cells. Using this approach, we have confirmed a decrease in MAIT cell frequency in peripheral blood of UC patients relative to healthy donors. We have also found an increase in MAIT frequency in inflamed colonic mucosa of UC patients relative to uninflamed mucosa. Single cell gene expression profiling revealed tissue specific differences between blood MAIT cells and colonic MAIT cells as well as a trend for higher expression of activation and pro-inflammatory genes in MAITs isolated from inflamed mucosa. Most intriguingly, single cell TCR sequencing revealed the expansion of a subset of MAITs in inflamed colonic mucosa with a distinct TCRVβ and CDR3β sequence. This suggests that a sub-population of MAIT cells in inflamed colonic mucosa are reacting to unique microbial ligands. A new cohort will extend these findings, but our preliminary results support a role for MAIT cells in the pathology of UC.

SINGLE CELL TRANSCRIPTOME MAPPING REVEALS CELL TYPE-SPECIFIC EFFECTS ON GENE EXPRESSION BY ACUTE DELTA9-TETRAHYDROCANNABINOL IN HUMANS

Ying Hu[1], Mohini Ranganathan[2,3], Chang Shu[2,3], Xiaoyu Liang[2,3], Suhas Ganesh[2,3], Chunhua Yan[1], Xinyu Zhang[2,3], Bradley E Aouizerat[4,5], John H Krystal[2,3], Deepak C D'Souza[2,3], Ke Xu[2,3]

[1]National Cancer Institute, Center for Biomedical Information and Information Technology, Bethesda, MD, [2]Yale University, Department of Psychiatry, New Haven, CT, [3] Connecticut Veteran Healthcare System, West Haven, CT, [4]New York University, Bluestone Center for Clinical Research, New York City, NY, [5]New York University, Department of Oral and Maxillofacial Surgery, New York City, NY

Delta 9-tetrahydrocannabinol (THC), the principal psychoactive constituent of cannabis, is also known to modulate immune response in peripheral cells. However, little is known about how THC regulates gene expression in human immune cells in vivo. Combining a within-subject design with single cell transcriptome mapping, we report that administration of THC acutely alters gene expression in 15,973 human peripheral blood mononuclear cells. Controlled for high inter-individual transcriptomic variability, we identified 294 transcriptome-wide significant genes among eight cell types including 69 significant genes shared in at least two cell types. Broadly, the 294 genes are involved in immune response (e.g. *HLA-A, HLA-DQA1/A2, IGLC2, IGKC*), cytokine production (e.g. *CCL4, GNLY*), signal transduction (e.g. JUNB, FOS, GPR55), and cell proliferation and apoptosis (e.g. *GZMB, ID2, DDIT4*). In major cell types, we found distinct transcriptomic sub-clusters affected by THC. THC perturbed cell type-specific intracellular gene expression correlation. Gene set enrichment analysis further supports the findings of THC's effects on immune response and cell toxicity. Finally, we found that THC alters the correlation of cannabinoid receptor gene, *CNR2*, with other genes in B cells, in which *CNR2* showed the highest level of expression. This comprehensive cell-specific transcriptomic profiling provides important insights into THC's acute effects on immune function that may have important medical implications.

GENE EXPRESSION CORRELATES OF CORTICAL PROJECTIONS REVEALED BY *IN SITU* SEQUENCING

Yu-Chi Sun, Xiaoyin Chen, Anthony Zador

Cold Spring Harbor Laboratory, Neuroscience, Cold Spring Harbor, NY

In vertebrate brains, the combinatorial expression of genes, such as cell adhesion molecules, are required to establish the highly diverse neuronal projections in the vertebrate brain. Understanding the organization of neuronal circuits thus requires characterizing, among various neuronal properties, how combinatorial gene expressions directs neuronal connectivity. This requires high-throughput gene detection and projection mapping in the same neurons while retaining spatial information at cellular resolution. Here, we combine targeted *in situ* sequencing and BARseq, a high-throughput technique for mapping long-range axonal projections based on *in situ* sequencing of RNA barcodes, to correlate gene expression and projections at cellular resolution. We use this approach to test the hypothesis that combinatorial cadherin expressions specify neuronal connectivity in the mouse auditory cortex. Our method uncovers the degree to which combinatorial cadherin expressions can predict neuronal projection patterns and thus potentially illuminate the gene expression logic underlying neuronal connectivity.

ACCOUNTING FOR CELL TYPE COMPOSITION IN BULK TRANSCRIPTOME AND METHYLOME DATA BY USING scRNA-seq REFERENCE

Fumihiko Takeuchi

National Center for Global Health and Medicine (NCGM), Research Institute, Tokyo, Japan

Gene expression and CpG methylation in patients have been measured in bulk samples of blood, liver, fat, brain etc. Such bulk tissues are mixtures of multiple cell types. Disease association of a gene (or CpG site) might vary by cell type, in which case the average is observed at tissue level. Here, I account for cell type composition in transcriptome and methylome data and refine the associations by cell types.
Cell types and their gene expression profile were obtained by published single-cell RNA sequencing experiments of human and mouse. By using gene expression profile of cell types as reference, the cell type composition of each tissue sample was inferred. By including the inferred cell type composition in multiple regression analysis, association of the omics measurements with disease were refined. Published data on smoking and obesity were reanalyzed.

A SUITE OF TOOLS FOR WORKING WITH SINGLE-CELL ATAC-seq DATA

Ming Tang[1,2], Brandon Logeman[2], Catherine Dulac[1,2,3,4], Tim Sackton[1]

[1]Harvard University, Informatics, FAS, Cambridge, MA, [2]Harvard University, Molecular and Cellular Biology, Cambridge, MA, [3]Harvard University, Center for Brain Science, Cambridge, MA, [4]Howard Hughes Medical Institute, Molecular and Cellular Biology, Chevy Chase, MD

Assaying the chromatin accessibility of the entire genome at single-cell resolution, using single-cell ATAC-seq (scATAC-seq), is an increasingly popular and powerful technique. scATAC-seq has provided valuable information on how the genome is regulated during development, and given insights into the progression of cancer and other diseases. However, scATAC-seq datasets are typically large and sparse, and working with them can be challenging, especially in the absence of dedicated tools.
Here, we present scATACutils, an R package for scATACseq quality control and visualization. This R package provides functions to calculate a variety of quality control metrics for scATAC-seq data, including FRIP (fraction of reads in called peaks) score, nucleosome banding score and TSS enrichment score. In addition, we provide a series of visualization tools to efficiently plot relevant metrics from large, sparse, scATAC-seq datasets.
Downstream analysis of scATAC-seq data can be cumbersome, as typical workflows require processing each cluster separately to create a bigwig file and call peaks, following by merging, filtering, and scoring read counts per peak in each cluster. To facilitate these steps, we developed a Snakemake workflow to automate these tasks. Although MACS2 is routinely used for calling peaks for ChIP-seq data, it is not designed specifically for ATAC-seq data. We use Genrich, a program developed in our group to call peaks for ATAC-seq data. We show that Genrich is at least as accurate as MACS2, but substantially faster, facilitating analysis of complex scATAC-seq datasets. We also benchmark methods for finding differential accessible regions across potentially up to a million peaks and thousands of cells and show that presto (a fast implementation of Wilcoxon rank-sum test and auROC) is significantly more performant than Seurat. To demonstrate the extensibility of this workflow, we show how pyflow-scATAC can be used to carry out downstream motif analysis for each cluster, finding potential driver transcription factors. Finally, the workflow identifies potential gene targets of the distal ATAC-seq peaks and carries out gene set enrichment analysis for eachcluster. To facilitate the reproducibility of the workflow, we have made Docker containers available for the users.

The pyflow-scATACseq Snakemake workflow can be found at https://github.com/crazyhottommy/pyflow-scATACseq
The scATACutils R package can be found at https://github.com/crazyhottommy/scATACutils/

SINGLE CELL GENOMICS IDENTIFIES CELL-TYPE SPECIFIC MOLECULAR CHANGES IN PRECISION-MODELS OF GENETIC EPILEPSIES

Parul Varma[1,2], Vanesa Nieto-Estevez[1,2], Zane Lybrand[1,2], Drew M Thodeson[3], Sean C Goetsch[4], Jay W Schneider[4], Jenny Hsieh[1,2]

[1]The University of Texas at San Antonio, Department of Biology, San Antonio, TX, [2]The University of Texas at San Antonio, Brain Health Consortium, San Antonio, TX, [3]UT Southwestern Medical Center, Departments of Pediatrics and Neurology and Neurotherapeutics, Division of Child Neurology, Dallas, TX, [4]UT Southwestern Medical Center, Department of Internal Medicine and Hamon Center for Regenerative Science and Medicine, Dallas, TX

Introduction: Millions of people in the US suffer from epilepsy and approximately 75% of this epilepsy begins during childhood. About 1/3rd of these patients are refractory to medication and/or not candidates for other treatments. Advances in genomic technologies have identified more than 100 genes in humans that can cause epilepsy when mutated. Rodent studies on some of these genes have defined the pathology and mechanisms associated with various genetic epilepsies. However, despite being widely studied in past few years, the epilepsy field still lacks a specific disease-modifying treatment. Currently, a major obstacle to developing better treatment strategies which are not only preventive but also curative is the lack of direct assessment of cell-type specific molecular changes occurring during development in the human brain affected by an epilepsy associated gene mutation.

Methods: To elucidate the molecular mechanisms of genetic epilepsies during development, we have taken two new promising approaches. We have first generated "epilepsy-in-a-dish" model for two genes associated with epilepsy (*ARX and CHD2*). These are 3D models of human brain development derived from human induced pluripotent stem cells (iPSCs) and we are using cortical and subpallium spheroids (hCS and hSS, respectively) to model two different regions of the brain. Next, we are performing unbiased single cell RNA sequencing (scRNA-seq) and single cell ATAC (scATAC-seq) sequencing at different time points to identify cell type specific molecular changes associated with these epilepsies. We are using R packages Seurat and Cicero for scRNA-seq and scATAC-seq data analysis respectively.

Results: We have identified significant changes in the molecular profiles of various cell types in cortical and subpallium spheroids. For example, we have detected changes in gene expression profiles of different layers of cortical neurons. Specifically, in *ARX* mutant cortical spheroids, we have found an increase in the layer 1, layer 3 and layer 6 neuronal expression markers. Moreover, we have also identified changes in the expression profiles of progenitors, intermediate progenitors and astrocytes in the *ARX* mutant cortical spheroids. In *CHD2* mutant spheroids, along with the cell type specific gene expression changes, we have also identified changes in the genome wide chromatin accessibility.

Conclusion: Our results show that single cell genomics combined with the approach of 3D models of human brain development are excellent tools to uncover epilepsy driven cell type specific changes in gene expression and chromatin accessibility during development. These studies can eventually help in the identification of novel molecular targets for therapeutic intervention.

CELLULAR NOISE AND METABOLISM, AS A MATTER OF FAT

Andreas E Vasdekis[1], Gregory Stephanopoulos[2]

[1]University of Idaho, Physics, Moscow, ID, [2]Massachusetts Institute of Technology, Chemical Engineering, Cambridge, MA

No two individual cells "look" the same, even if they share the same genes and grow under identical conditions. This unexpected phenomenon, frequently termed as 'cellular noise', emerges in part due to the stochastic nature of molecular-level interactions within individual cells that occur during protein production. As such, a form of non-clonal heterogeneity in the copy number and types of proteins between cells arises.

In this presentation, we will discuss the statistical physics nature of cellular noise and our recent findings of how cellular noise propagates from gene regulatory to metabolic networks with a focus on lipid production and growth. To this end, we will first present the interferometric and fluorescent bioimaging methods [1] that we employ to dynamically track cellular metabolism of more than 1,000 single-cells on a microfluidic chip.

We will then proceed with our recent findings of how cellular noise affects lipid production and accumulation [2] and how we can use cellular noise to better understand metabolic dynamics [3]. We will conclude with our more recent results on the independent impacts of cellular noise on growth and lipid accumulation, as well as the underlying metabolic trade-offs and competition between these two metabolic objectives [4].

[1] Metabolic Engineering 27, 115 (2015).
[2] Scientific Reports 5, 17689 (2015).
[3] PLOS ONE 12, e0168889 (2017).
[4] Nature Communications 10, 848 (2019).

BENCHMARKING OF NORMALIZATION AND IMPUTATION WORKFLOWS ACROSS REAL AND SIMULATED scRNA-seq DATA.

Priyanka Vijay[1], Astrid Wachter[2], Jennifer Mollon[2], Sunantha Sethuraman[1], Zheng Zha[1], Jozsef Karman[3], Cyril Ramathal[1], James M Barnes[1], Michael Pickup[4], Xin Lu[1], Tyler Mansfield[1], Relja Popovic[1], Josue Samayoa[4], Albert Park[5], Janina S Ried[2], Justin W Davis[1]

[1]AbbVie Inc., 1 North Waukegan Rd., North Chicago, IL, [2]AbbVie Deutschland GmbH & Co. KG, Neuroscience Research, Ludwigshafen, Germany, [3]AbbVie Bioresearch Center, 100 Research Drive, Worcester, MA, [4]Abbvie Biotherapeutics, Redwood City, CA, [5]AbbVie Stemcentrx, 450 E. Jamie Court, South San Francisco, CA

Single cell RNA-seq technologies have revolutionized our ability to characterize healthy and disease tissue. Although there are many analysis methods for scRNA-seq data, we lack a clear consensus on which workflows are best suited for a given question, data collection platform, and scale of data. Recent comparative studies have addressed this gap to an extent, especially for trajectory inference and differential expression analysis. Prior to these analyses however, robust normalization workflows must be used to ensure the proper detection of signal from noise, and a comprehensive comparison of normalization methodologies across real and simulated data is needed. To that end, we benchmarked the performance of 13 normalization and 2 imputation methods and follow-up steps of the removal of unwanted variation, totaling 78 distinct workflows. The normalization methods tested include both methods originally developed for bulk RNA-seq (DESeq, Upper Quantile, Full Quantile, TMM, TPM, Sum scaling) as well as methods specifically developed for scRNA-seq (Seurat, SCnorm, Linnorm, SCRAN, ZINB-WaVE, SCTransform, Census). The imputation methods used were scImpute and ALRA. The testing was performed across 10 real datasets from plate-based, microfluidic, and droplet-based platforms ranging in size from 90 cells to 1.3 million cells, and 34 simulated datasets. The real datasets include public data (ex. the Human Cell Atlas Immune Cell dataset) as well as a novel dataset generated for this study consisting of different mixtures of 3 cancer cell lines in known proportions. This dataset provides a ground-truth for biological variability that can be leveraged in the comparison of normalization methods and subsequent clustering. The 34 simulated datasets include multiple group and multiple batch simulations and span a parameter space representative of real data. Performance was primarily evaluated using 8 different metrics defined in the Scone R package. These metrics quantify some key aspects of an ideal single cell workflow such as the preservation of cellular heterogeneity, the preservation of biological clusters, and the removal of batch structure. Our results provide the field with valuable guidelines on how to best choose a normalization method and perform denoising given a specific experimental design. This can not only help reduce the decision load imposed on single cell data analysts but can also provide a standard resource against which newly developed methods can be compared.

Disclosures: All authors are employees of AbbVie. The design, study conduct, and financial support for this research were provided by AbbVie. AbbVie participated in the interpretation of data, review, and approval of the publication.

IN SITU STRUCTURAL BIOLOGY – ENABLED BY FOCUSED ION BEAM CRYO-LAMELLAE PREPARATION AND ELECTRON TOMOGRAPHY

Tomas Vystavel[1], Radovan Spurny[1], Alexander Rigort[2]

[1]Thermo Fisher Scientific, Materials & Structural Analysis, Brno, Czech Republic, [2]Thermo Fisher Scientific, Materials & Structural Analysis, Planegg, Germany

Cryo techniques for electron microscopy allow to preserve cellular samples in a close-to-native, fully hydrated state, frozen in non-crystalline vitreous ice. Recent advances in cryo-focused ion beam (cryo-FIB) based sample preparation have enabled the preparation of site-specific thin sections (called cryo-lamellae) from such vitrified cells. Cryo-lamellae are essentially snapshots of functional cellular environments whose supramolecular architecture can be studied *in situ* at high resolution and in 3D by electron tomography. This enables structural investigations of biomolecules in the context of cellular structures.
To study the molecular machinery of cells, a comprehensive workflow is needed covering vitreous freezing, cryo-fluorescence microscopy, sample thinning by cryo-FIB, and high-resolution electron tomography. Cryo-lamella preparation typically begins with cryo-correlative microscopy. Here, coordinate information from fluorescently labeled structures is obtained. Afterwards, the sample along with its target coordinates, is transferred to the Thermo Scientific Aquilos cryo-FIB. This innovative and user-friendly microscope brings the cell biologist a preparation tool which allows producing optimal samples for electron tomography while maintaining cryogenic conditions. Utilizing the ion beam of the cryo-FIB microscope, frozen cellular specimens are thinned down to an appropriate thickness of 200 nm. This renders the cryo-lamellae transparent for tomography data acquisition. The thinned cellular sample is then transferred to a cryo-transmission electron microscope (cryo-TEM) where a tomographic image series is acquired as the lamella is tilted incrementally. Finally, the images are combined computationally to reconstruct a 3D tomogram of the cellular interior.

CROSS-SPECIES TRANSCRIPTOMIC AND EPIGENOMIC ANALYSIS REVEALS KEY REGULATORS OF INJURY RESPONSE AND NEURONAL REGENERATION IN VERTEBRATE RETINAS

Thanh Hoang[1], Jie Wang[2], Patrick Boyd[4], Fang Wang[2], Clayton Santiago[1], Guohua Wang[2], John Ash[6], Andrew J Fischer[5], David R Hyde[4], Jiang Qian[2], Seth Blackshaw[1,3]

[1]Johns Hopkins University School of Medicine, Solomon H. Snyder Department of Neuroscience, Baltimore, MD, [2]Johns Hopkins University School of Medicine, Department of Ophthalmology, Baltimore, MD, [3]Johns Hopkins University School of Medicine, Department of Neurology, Baltimore, MD, [4]University of Notre Dame, Department of Biological Sciences, Notre Dame, IN, [5]Ohio State University Wexner Medical Center, Department of Neuroscience, Columbus, OH, [6]University of Florida School of Medicine, Department of Ophthalmology, Gainesville, FL

Injury induces retinal Müller glia of cold-blooded, but not mammalian, vertebrates to generate neurons. To identify gene regulatory networks that control neuronal reprogramming in retinal glia, we comprehensively profiled injury-dependent changes in gene expression and chromatin conformation in Müller glia from zebrafish, chick and mice using bulk RNA and ATAC-Seq, as well as single-cell RNA-Seq. Cross-species integrative analysis of these data, together with functional validation, identified evolutionarily conserved and species-specific gene networks controlling glial quiescence, gliosis and neurogenesis. In zebrafish and chick, transition from quiescence to gliosis is a critical stage in acquisition of retinal regeneration, while in mice a dedicated network suppresses this transition and rapidly restores quiescence. Selective disruption of NFI family transcription factors, which maintain and restore quiescence, enables Müller glia to proliferate and robustly generate neurons in adult mice following retinal injury. These comprehensive resources and findings may facilitate the design of cell-based therapies aimed at restoring retinal neurons lost to degenerative disease.

MULTIPLEX SINGLE CELL PROFILING TO DISSECT THE TUMOR IMMUNE MICROENVIRONMENT IN HUMAN CANCERS

Linghua Wang[1], Guangchun Han[1], Ruiping Wang[1], Shaojun Zhang[1], Michael Wang[2], Michael Green[2], Andrew Futreal[1], Jaffer Ajani[3]

[1]U.T. MD Anderson Cancer Center, Genomic Medicine, Houston, TX,
[2]U.T. MD Anderson Cancer Center, Lymphoma & Myeloma, Houston, TX,
[3]U.T. MD Anderson Cancer Center, Gastrointestinal Medical Oncology, Houston, TX

Malignant cells have been the focus of cancer research for several decades. However, it is being increasingly recognized that there is an ongoing interplay between tumor cells and the tumor immune microenvironment (TIME). Tumor cells orchestrate and interact dynamically with the non-tumor components, and the crosstalk between tumor, stroma and immune system can promote tumor cell growth, migration, and sustain the hallmarks of cancer, and confer drug resistance. For example, numerous studies in other type of cancers have shown that the infiltrating myeloid cells such as tumor-associated macrophages, myeloid-derived suppressor cells, and the cancer-associated fibroblasts contribute to cancer progression and metastasis. Single cell RNA sequencing (scRNA-seq) is a robust tool to help dissect the TIME and study tumor-immune interaction. However, the TIME has not been studied systematically across different types of human cancers, neither longitudinally in the context of targeted or immunotherapy in cancer patients. Combining scRNA-seq data generated from different laboratories encounters unavoidable batch effects that significantly limits downstream analyses. We have leveraged our access to freshly collected/frozen tumor specimens from cancer patients with rich clinical information and have generated a huge amount of high-quality scRNA-seq data for multiple cohorts of patients across different cancer types. Importantly, scRNA-seq was performed using the same library preparation protocol and the libraries were sequenced in the same core with minimal batch effects. We have profiled different cell compartments (cell types and cellular states) of both tumor, stroma and immune cells of the TIME in unprecedented details, and compared their profiles across different cancer types, during disease progression or the treatment course. In addition, we have investigated cell-to-cell communication using the bioinformatics tool developed in our lab and built the cellular interaction networks of TIME. By integrating with the rich clinical information, we have made a number of novel discoveries that leads to identification of new diagnostic/predictive signatures and therapeutic targets, which will be demonstrated in details during this presentation. Together, the data we have generated can be a valuable resource for the community as well as the novel insights derived from this study.

CELL-OF-ORIGIN ANALYSIS OF METASTATIC GASTRIC CANCER UNCOVERS THE ORIGIN OF INHERENT INTRATUMORAL HETEROGENEITY AND A FUNDAMENTAL PROGNOSTIC SIGNATURE

Ruiping Wang[1], Shumei Song[2], Kazuto Harada[2], Guangchun Han[1], Melissa P Pizzi[2], Meina Zhao[2], Shaojun Zhang[1], Yuanxin Wang[1], Jeannelyn S Estrella[3], Sinchita Roy-Chowdhuri[3], Ahmed Adel Fouad Abdelhakeem[2], Guang Peng[4], George A Calin[5], Samir M Hanash [4], Alexander J Lazar [1,3,6], Andrew Futreal[1], Jaffer A Ajani*[2], Linghua Wang*[1]

[1]U.T. MD Anderson Cancer Center, Genomic Medicine, Houston, TX, [2]U.T. MD Anderson Cancer Center, Gastrointestinal Medical Oncology, Houston, TX, [3]U.T. MD Anderson Cancer Center, Pathology, Houston, TX, [4]U.T. MD Anderson Cancer Center, Clinical Cancer Prevention, Houston, TX, [5]U.T. MD Anderson Cancer Center, Experimental Therapeutics, Houston, TX, [6]U.T. MD Anderson Cancer Center, Translational Molecular Pathology, Houston, TX

Gastric adenocarcinoma (GAC) remains a common and lethal disease with a poor prognosis. Often diagnosed at an advanced stage, GAC is frequently resistant to therapy. A common site of metastases is the peritoneal cavity (peritoneal carcinomatosis; PC) and there is an unmet need for improved therapeutic options in advanced GAC patients. It is well recognized that intra-tumoral heterogeneity (ITH) is fundamental for GAC survival as it confers therapy resistance and is a major obstacle to improving patient outcome. However, the origins of ITH are poorly understood. Deeper understanding of the cellular/molecular basis of ITH could influence how GACs are treated. Here we performed single-cell RNA sequencing of peritoneal carcinomatosis (PC) from 20 patients with advanced gastric adenocarcinoma (GAC), constructed a transcriptome map of 45,048 PC cells, determined the cell-of-origin of each tumor cell, and incisively explored ITH of PC tumor cells at single-cell resolution. A key finding of this study is that diversity of cell-of-origin appears to mirror and may even dictate inherent ITH of PC tumor cells at multiple molecular levels. The links between cell-of-origin and ITH was illustrated for the first time at transcriptomic, genotypic, molecular, and phenotypic levels. Most intriguingly, the cell-of-origin-based analysis classified PC tumor cells into two cellular subtypes that were prognostic independent of histopathological features. Further analyses led us to discover a 12-gene signature that appears to be fundamental to GAC carcinogenesis/propagation as it was not only highly prognostic in GAC metastatic validation cohort but perform just as robustly in several large-scale localized GAC cohorts. Currently, there is no such signature in clinical use and this signature has a high potential to stratify patients for more effective therapies as this becomes available.

SCOPE: A NORMALIZATION AND COPY NUMBER ESTIMATION METHOD FOR SINGLE-CELL DNA SEQUENCING

Rujin Wang[1], Danyu Lin[1,2], Yuchao Jiang[1,2,3]

[1]Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, [2]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, [3]Department of Genetics, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC

INTRODUCTION: Whole-genome single-cell DNA sequencing (scDNA-seq) enables characterization of copy number profiles at the cellular level. This increases resolution and decreases ambiguity in assessing tumor heterogeneity and tracking cancer evolutionary history. ScDNA-seq data is, however, sparse, noisy, and highly variable even within a homogeneous cell population due to the biases and artifacts that are introduced during library preparation and sequencing procedure. Existing methods either build an optimized normal/reference set for normalization or adopt a cell specific normalization procedure for removing biases, followed by a post hoc procedure for ploidy estimation and adjustment. They cannot adequately remove such artifacts and, furthermore, they do not address the challenges and complexities caused by aberrant copy number changes or the complicating factor of tumor ploidy.
METHODS: SCOPE integrates both null and non-null genomic regions for unbiased estimation of both GC content bias and latent factors for cell- and region-specific background correction. The distinguishing features of SCOPE include: (i) utilization of cell specific Gini coefficients for quality control and for identification of normal cells, which are then used as negative control samples in a Poisson latent factor model for read depth normalization; (ii) modeling of GC content bias using an expectation-maximization algorithm embedded in the Poisson regression models to account for the different copy number states along the genome; and (iii) a cross-sample iterative segmentation procedure to identify breakpoints that are shared across cells with the same genetic background.
RESULTS: SCOPE is shown to more accurately estimate subclonal copy number aberrations and to have higher correlation with array-based copy number profiles of purified bulk samples from the same breast cancer patient. Ploidy estimates by SCOPE are highly concordant with those from previous reports based on single-cell flow sorting. We further show that the copy number profiles returned by SCOPE can also be recapitulated by whole-exome sequencing and single-cell RNA-sequencing. We finally demonstrate SCOPE on scDNA-seq data that was produced using the 10X Genomics single-cell CNV pipeline, showing that it can reliably recover proportions of the cancer cell spike-ins from a background of normal cells and that it can successfully reconstruct cancer subclonal structure across 10,000 breast cancer cells.

A PROBABILISTIC FRAMEWORK FOR CELLULAR LINEAGE RECONSTRUCTION USING SINGLE-CELL 5-HYDROXYMETHYLCYTOSINE SEQUENCING

Chatarin Wangsanuwat[1,2], Javier F Aldeguer[3], Nicolas C Rivron[3], Siddharth S Dey[1,2]

[1]Department of Chemical Engineering, University of California Santa Barbara, Santa Barbara, CA, [2]Center for Bioengineering, University of California Santa Barbara, Santa Barbara, CA, [3]Hubrecht Institute–KNAW and University Medical Center Utrecht, Utrecht, Netherlands

Lineage reconstruction is central to understanding tissue development and maintenance. While powerful tools to infer cellular relationships have been developed, these methods typically have a clonal resolution that prevents the reconstruction of lineage trees at an individual cell division resolution. Moreover, these methods require a transgene, which poses a significant barrier in the study of human tissues. To overcome these limitations, we report scPECLR, a probabilistic algorithm to endogenously infer lineage trees at a single cell-division resolution using 5-hydroxymethylcytosine. When applied to 8-cell preimplantation mouse embryos, scPECLR predicts the full lineage tree with greater than 95% accuracy. Further, scPECLR can accurately extract lineage information for a majority of cells when reconstructing larger trees. Finally, we show that scPECLR can also be used to map chromosome strand segregation patterns during cell division, thereby providing a strategy to test the “immortal strand” hypothesis in stem cell biology. Thus, scPECLR provides a generalized method to endogenously reconstruct lineage trees at an individual cell-division resolution.

TRACKING EGFR SIGNALING PATHWAY AND ITS CROSSTALK OF APOPTOSIS PATHWAY VIA MULTIPLEXED SINGLE CELL ANALYSIS

Yanrong Wen, Jia Liu, Zhen Liu*

Nanjing University, State Key Lab of Analytical Chemistry for Life Science, School of Chemistry and Chemical Engineering, Nanjing, China

Dissecting cellular signaling pathways requires the measurement of a large number of signaling proteins. However, conventional methods such as phosphoproteomics, protein–protein interaction studies and gene expression profiling average information over many cells and do not take into account cell-to-cell heterogeneity, which has been shown to play key parts in the function of signaling networks. Tools for simultaneously monitoring the activation dynamics of multiple components of signaling in the same living cell are still limited. Herein, we describe a multiplexed single cell analysis approach for the simultaneous detection and quantitation of multiple cytoplasmic and nucleus signaling proteins in the same living cell. This approach can provide ultrasensitive detection at the single-molecule level and allow for simultaneous and parallel multiplexing non-destructive cell analysis of specific cells. We applied this approach to characterize the modulation of key regulatory signaling proteins and the corresponding signaling nodes of proliferation and apoptosis in individual cells following acute EGFR inactivation. We found that differential attenuation rates of pro-survival and pro-apoptotic signals in EGFR-dependent cells contribute to cell death following EGFR inactivation. These findings implicate a transient imbalance in survival and apoptotic oncogenic outputs in the apoptotic response to EGFR inactivation. Our approach could utilize signaling protein quantitative data to predict therapeutic effect using only a limited number of cells. This approach can be a very powerful tool to reveal signaling dynamic pattern at the single cell level. It holds great promise for personalized drug screening and precision medicine.

TO CLUSTER OR NOT TO CLUSTER: ASSESSING CLUSTERABILITY IN SCRNA-seq DATA.

Elijah S Willie[1,2], Sara Mostafavi[1,2,3], Ellis Patrick[4]

[1]University of British Columbia, Statistics and Medical Genetics, Vancouver, Canada, [2]University of British Columbia, Centre for Molecular Medicine and Therapeutics, Vancouver, Canada, [3]Canadian Institute for Advanced Research, Child and Brain Development Program, Toronto, Canada, [4]University of Sydney, Mathematics and Statistics, Sydney, Australia

Single-cell RNA-sequencing (scRNA-seq) has allowed researchers over the last decade to profile cells and tissue types with great resolution thus increasing the number of cells catalogued. This has resulted in a pressing need for robust and scalable analysis pipelines. Existing pipelines include methods for preprocessing, which reduces technical and experimental errors; methods for statistical modelling; and unsupervised learning. Unsupervised learning is one of the most important parts of any single-cell analysis pipeline as it most often helps to identify naturally occurring clusters for a given dataset. Identifying these potential clusters is helpful when inferring novel cell types. However, the majority of clustering algorithms will sometimes find clusters in a given dataset regardless of if said clusters are artifacts of the algorithm used. As such the idea of clusterability can be defined as the presence of natural cluster structures within a given dataset and are not in fact artifacts of clustering algorithms. In this work, the problem of assessing clusterability is addressed by defining a statistically robust framework that formally tests this clusterability notion and provides a degree of evidence for the existence of clusters.

There are several parameters has a significant impact on the ability of methods to find meaningful clusters. This includes modelling and preprocessing of the counts as well as feature selection and dimension reduction. Theses parameters are carefully considered and incorporated into the proposed method and as such, this method consists of three stages. Firstly, the counts are modelled and highly informative features are selected using a log and multinomial model. Secondly, meaningful projections are computed by projecting the processed data using linear and nonlinear methods. Finally, the Dip test that assesses the multimodality of a distribution function is used on distributions pairwise dissimilarities between cells. Pairwise dissimilarities of projections are computed using Pearson Correlations, Euclidean and Manhattan distances.

Clusterability in the scRNA-seq domain is not well defined and to the best of our knowledge, this study is the first to address this. Methods for assessing clusterability should ideally be done before clustering is applied. This may help to save computation time or suggest more suitable types of analyses. Results on real data and simulation show low type one error and high power making this method a robust way to assess clusterability. By using a robust statistical framework, based on meaningful projections, and highly informative features, it is possible to obtain highly accurate prediction of clusterability. We expect this type of analysis to become a crucial part many scRNA-seq analysis pipeline.

RECOVERY OF GENETIC AND EPIGENETIC HETEROGENEITY USING SINGLE-CELL DNA AND METHYLOME SEQUENCING

Chi-Yun Wu[1], Nancy R Zhang[2]

[1]Perelman School of Medicine, Genomics and Computational Biology Graduate Program, Philadelphia, PA, [2]The Wharton School, Department of Statistics, Philadelphia, PA

Genetic and epigenetic heterogeneity within tumor samples reflect the molecular mechanisms underlying tumor evolutionary dynamics. Intratumor heterogeneity is commonly studied at the clonal level where a "clone" is usually defined by DNA mutations such as single nucleotide variations (SNVs) and copy number variations (CNVs). Recently developed single-cell DNA-sequencing (scDNA-seq) enables mutation detection at the single cell level. In addition to DNA mutations, epigenetic heterogeneity is also prevalent in tumor samples. Single-cell methylome sequencing enable methylation detection in individual cells. These two types of sequencing data are low-coverage and plagued by dropout events, and computational methods are required to address these issues. Currently, mutation detection methods for scDNA-seq and methylation profiling methods for single-cell methylome sequencing data are lacking in accuracy and scalability. Although the two types of data reflect different genomic features, they can be treated under the same statistical and computational framework.
In this study, we develop a statistical framework to denoise scDNA-seq and single-cell methylome sequencing data using an unsupervised neural network called autoencoder. For scDNA-seq imputation, we borrow the raw allele frequency of alternative alleles/CNV states across all detected SNV/CNV sites in all cells to impute the missing genotypes and denoise all positions. Results show that the autoencoder outperforms two existing methods in terms of SNV signal estimation. We found that the autoencoder was able to impute missing SNV genotypes in the single-cell targeted DNA sequencing data and under different simulated scenarios. The SNV signals estimated by the proposed method allows the recovery of underlying clonal structures. A similar strategy is employed to impute the methylation states for single-cell methylome data. Our results reveal the potential of low dimensional latent space methods in the recovery of genetic and epigenetic signals for scDNA-seq and single-cell methylome sequencing data, facilitating the characterization of intratumor genetic and epigenetic heterogeneity.

A COMPARISON OF HIGH-THROUGHPUT SINGLE-CELL RNAseq METHODS FOR PROFILING IMMUNE CELLS

Tracy M Yamawaki, Daniel Lu, Daniel Ellwanger, Hong Zhou, Oliver Homann, Songli Wang, Chi-Ming Li

Amgen Research, GAU, South San Francisco, CA

Advances in single-cell methods have made it possible to profile transcriptomes from thousands of individual cells in parallel, but these developments have led to a veritable mélange of methods and choosing an experimental method is far from straightforward. To characterize differences in common single-cell methods for the characterization of immune cells, we have assessed the performance of several commercially-available single-cell systems: the 10X Chromium, the ddSeq, the Dolomite Bio SC RNAseq system (DropSeq), and the iCell8 cx. Here, we present data examining the sensitivity and variation of mRNA detection as well as cell-recovery efficiency for a real-life test case: a mixture of four lymphocyte cell-lines. We also compare each method's ability to identify differentially-expressed genes between cell-types. We find variation in performance across methods as well as across cell-types. Our comparison of these single-cell systems can provide useful metrics to guide method selection for profiling more complex mixtures immune cells found *in-vivo*.

ADROIT: ACCURATE AND ROBUST INFORMATION-THEORETIC DECONVOLUTION FROM BULK TISSUE TRANSCRIPTOMES

Tao Yang[1], Michael Schaner[2], Michael LaCroix-Fralish[2], Lynn E Macdonald[1,2], Wen Fury[1], Gurinder S Atwal[1], Yu Bai[1]

[1]Regeneron Pharmaceuticals, Molecular Profiling & Bioinformatics, Tarrytown, NY, [2]Regeneron Pharmaceuticals, Neurosciences, Tarrytown, NY

RNA sequencing technology promises unprecedented opportunity in learning disease mechanisms and discovering new treatment targets. In controlled experiments, it's often of great importance to know the variability of cell composition under treatment interventions. While bulk tissue RNA-seq enables accurate quantification of transcriptome abundance with whole genome coverage, single cell RNA-seq has the power to reveal cell type composition and expression heterogeneity in different cells. To leverage the bulk tissue RNA-seq data when sequencing at the single cell level is not feasible, here we present AdRoit, a method to estimate the proportions of each cell type in the bulk RNA-seq data using independently acquired single cell data that become abundant in the public domain. Our method uniquely employs probabilistic modeling to overcome the significant cross-subject and cross-cell type variabilities in the single cell data, thus is more accurate and robust. In particular, the method exiles in detecting very rare population cell types (i.e., < 0.5%). Our systematic benchmarking evaluation shows superior sensitivity and specificity to other existing methods, even in neuronal cells where there exist many closely related subtypes. We further estimate the variance of the proportions existing in the reference single cell data, which can later be utilized to infer the confidence of the proportion estimates and do statistical comparison of proportions between different conditions.

NORMAL AND MALIGNANT GERMINAL CENTER B CELL MAPS REVEALED BY SINGLE-CELL RNA SEQUENCING

Xiaofei Ye[1], Lei Wang[2,3], Man Nie[1,4], Mirjam van der Burg[5], Guibo Li[2,3], Zhiming Li[4], Kui Wu[2,3], Qiang Pan-Hammarström[1,4]

[1]Karolinska Institutet, Department of Biosciences and Nutrition, Stockholm, Sweden, [2]BGI-Shenzhen, BGI-Shenzhen, Shenzhen, China, [3]China National GeneBank, BGI-Shenzhen, Shenzhen, China, [4]Department of Medical Oncology, Sun Yat-Sen University Cancer Center, Guangzhou, China, [5]Department of Pediatrics, Leiden University Medical Center, Leiden, Netherlands Antilles

Diffuse large B-cell lymphoma (DLBCL), follicular lymphoma (FL), and Burkitt's lymphoma (BL) are subtypes of non-Hodgkin lymphomas that are likely derived from transformed germinal center B cells. DLBCL is the major subtype with notable genetic heterogeneity. Although current immune-chemotherapy has significantly improved the survival of DLBCL patients, approximately 30% patients still show relapsed or refractory disease. Additionally, virus infections, such as hepatitis B virus (HBV) infections, may also contribute to a poor therapeutic response in DLBCL patients. Here, we performed deep single-cell RNA sequencing (scRNA-seq) on 102,934 cells from 23 samples, including tumor biopsies from 17 DLBCL (7 HBV-infected), 1 FL and 1 BL patients, 1 reactive lymphoid hyperplasia sample from a patient with HBV infection, as well as samples from normal donors obtained from peripheral blood, bone marrow and tonsil. Largescale copy number variations inferred from single-cell RNA-seq expression profiles allowed us to distinguish malignant B cells from other infiltrated non-malignant cells. A trajectory analysis has been performed with all B cells to provide a comprehensive developmental path from pre-pro-B cells to mature B cells as well as their associates to the malignant B cells. When focusing on DLBCLs, a HBV infection associated principal component was identified by dimension reduction analysis from malignant B cells. Several HBV infection related genes were unregulated in the HBV associated DLBCLs, including *TNF*, *FOS*, *EGR2*, and *JUN*. Additionally, different expression patterns of infiltrating T cells have also been observed. The HBV infected DLBCLs have higher proportions of CD8+ T cells with upregulations of inflammatory and apoptosis responses pathways. Furthermore, the percentage of CD4+ T cells was significantly reduced in HBV positive DLBCLs. Overall, this study has provided an overview of normal and malignant germinal center B cell compartments in humans. Moreover, this study sheds new lights on the mechanism of HBV infection and B cell lymphomagenesis.

RECONSTRUCTION OF TUMOR-IMMUNE MICROENVIRONMENT USING A SYNGENEIC CT26 MOUSE TUMOR TRANSPLANT MODEL

Jinhee Yeo[1,2], Yourae Hong[1,2], Hye-Rim Park[1], Dasom Jeong[1], Se-Jin Kee[1,3], Woong-Yang Park [1,2,3], Hae-Ock Lee*[1]

[1]Samsung Medical Center, Samsung Genome Institute, Seoul, South Korea, [2]Sungkyunkwan University, Department of Health Sciences and Technology, Seoul, South Korea, [3]Sungkyunkwan University school of medicine, Department of Medicine, Seoul, South Korea

With tumor progression, the tumor microenvironment(TME) is reconstructed with the recruitment, differentiation, and local expansion of immune and stromal cells. To model the formation of TME, we used a syngeneic CT26 mouse tumor transplant model and characterized the alterations of stromal and immune microenvironment along the tumor growth. We found that the tissue resident cell populations in the skin were replaced with newly recruited immune cells, and that those recruits differentiated from monocytes to the tumor-associated macrophages or naïve T cells to the regulatory T cells. We also utilized this model to track changes in cellular dynamics and gene expression phenotypes after the treatment of immune checkpoint inhibitors. Our results demonstrate that the mouse model well-recapitulated cellular dynamics in human cancer, and further suggest that this model can be exploited to investigate the effects and mechanisms of action for various immune modulators.

LIMITATIONS OF SINGLE CELL RNA-seq

Chunling Zhang[1], Rujia Dai[2], Chunyu Liu[2]

[1]Upstate Medical University, Department of Neuroscience & Physiology, Syracuse, NY, [2]Upstate Medical University, Department of Psychiatry, Syracuse, NY

Single cell RNA-seq (scRNA-seq) can provide high resolution of cell type classification. It offers a unique opportunity to investigate transcriptomes of tissues with high heterogeneity, and detect gene expression changes among biological processes or disease/normal states in specific cell types. There is a myth that scRNA-seq will/should replace bulk cell/tissue RNA-seq when affordable. Here we explored the technical aspects of scRNA-seq and identified major limitations of scRNA-seq. 1) Current scRNA-seq technologies have difficulties detecting low abundance transcripts. Quantification of intermediate abundance transcripts can also be challenging due to high technical variance and dropout rates. We assessed the characteristics of genes that can be properly measured by scRNA-seq. 2) Majority of current scRNA-seq methods are not able to detect full-length transcripts, therefore cannot investigate RNA editing related events. That leaves out a large field critical for transcriptome research. Although a few methods such as Quartz-seq, SUPeR-seq, Smart-seq, and MATQ-seq can sequence full-length transcripts, but the low sequencing depth restricts the ability of assaying rare splicing forms. 3) Most current scRNA-seq methods target polyA transcripts, missing many important non-coding transcripts. SUPeR-seq, MATQ-seq, and SPLiT-seq may detect some ncRNAs, but the transcript coverage limits the capacity for studying many ncRNAs given the nature of low abundance. 4) Biases in cell isolation, tagging and amplification during the lab procedures may systematically drop out some important genes. By analyzing existing scRNA-seq data, we identified blind spots of scRNA-seq methods and genes not covered in scRNA-seq. Some of limitations may be addressed by future technology development, while some of them may require a continuous use of bulk cell/tissue RNA-seq or sorted cell RNA-seq to fill the gap.

SIMULATING MULTIPLE FACETED VARIABILITY IN SINGLE CELL RNA SEQUENCING

Xiuwei Zhang[3], Chenling Xu[1], Nir Yosef[1,2,4]

[1]University of California Berkeley, Department of Electrical Engineering and Computer Sciences, Berkeley, CA, [2]Ragon Institute of Massachusetts General Hospital, MIT and Harvard, Cambridge, MA, [3]Georgia Institute of Technology, School of Computational Science & Engineering, Atlanta, GA, [4]Chan-Zuckerberg Biohub, San Francisco, CA

The abundance of new computational methods for processing and interpreting transcriptomes at a single cell level raises the need for *in silico* platforms for evaluation and validation. Here, we present SymSim, a simulator that explicitly models the processes that give rise to data observed in single cell RNA-Seq experiments. The components of the SymSim pipeline pertain to the three primary sources of variation in single cell RNA-Seq data: noise intrinsic to the process of transcription, extrinsic variation indicative of different cell states (both discrete and continuous), and technical variation due to low sensitivity and measurement noise and bias. We demonstrate how SymSim can be used for benchmarking methods for clustering, differential expression and trajectory inference, and for examining the effects of various parameters on their performance. We also show how SymSim can be used to evaluate the number of cells required to detect a rare population under various scenarios.

SINGLE NUCLEUS RNA-seq AND ATAC-seq SURVEY OF CELL DIVERSITY AND FUNCTIONAL MATURATION IN POSTNATAL MAMMALIAN HEARTS

Juanjuan Zhao[1,2], Peng Hu[1,3], Jian Liu[1,2], Benjamin J Wilkins[2,4], Katherine Lupino[1,2], Changya Chen[5], Shiping Zhang[6], Kai Tan[5,7], Hao Wu[1,3], Liming Pei[1,2,4]

[1]Children's Hospital of Philadelphia, Center for Mitochondrial and Epigenomic Medicine, Philadelphia, PA, [2]Children's Hospital of Philadelphia, Department of Pathology and Laboratory Medicine, Philadelphia, PA, [3]University of Pennsylvania, Department of Genetics, Penn Epigenetics Institute, Philadelphia, PA, [4]University of Pennsylvania, Department of Pathology and Laboratory Medicine, Perelman School of Medicine, Philadelphia, PA, [5]Children's Hospital of Philadelphia, Division of Oncology and Center for Childhood Cancer Research, Philadelphia, PA, [6]Children's Hospital of Philadelphia, Department of Biomedical and Health Informatics, Philadelphia, PA, [7]University of Pennsylvania, Department of Cell and Developmental Biology, Perelman School of Medicine, Philadelphia, PA

A fundamental challenge in understanding cardiac biology and disease is that the remarkable heterogeneity in cell type composition and functional states have not been well characterized at single-cell resolution in maturing and diseased mammalian hearts. Massively parallel single nucleus RNA sequencing (snRNA-seq) and transposase-accessible chromatin profiling (snATAC-seq) have emerged as powerful tools to address these questions by interrogating the transcriptome and chromatin accessibility of tens of thousands of nuclei. Here, we comprehensively investigated the cell type diversity and functional maturation of postnatal maturing mouse hearts in healthy and disease states using snRNA-Seq and snATAC-seq methods. By profiling of transcriptome and open chromatin from about 20,000 nuclei respectively, we identified major and rare cardiac cell types and revealed significant heterogeneity of cardiomyocytes, fibroblasts, and endothelial cells in postnatal developing hearts. When applied to a mouse model of pediatric mitochondrial cardiomyopathy, we uncovered profound cell type-specific modifications of the cardiac transcriptome and open chromatin at single-nucleus resolution, including changes of subtype composition, maturation states, and functional remodeling of each cell type. We further deciphered the cardiac cell type-specific gene regulatory network (GRN) and open chromatin characteristics of GDF15, a heart-derived hormone and clinically important diagnostic biomarker of heart disease. Together, our results provide new insights into heart biology and disease at the single-nucleus resolution.

LATTICE LIGHT SHEET MICROSCOPY – INNOVATIONS, APPLICATIONS AND FUTURE DIRECTIONS

Wesley R Legant[1,2]

[1]University of North Carolina - Chapel Hill and North Carolina State University, Joint Department of Biomedical Engineering, Chapel Hill and Raleigh, NC, [2]University of North Carolina - Chapel Hill, Pharmacology, Chapel Hill, NC

Living specimens are both animate and three-dimensional. Lattice Light Sheet Microscopy (LLSM) is a new light-sheet microscopy method that utilizes non-diffracting beams and is capable of imaging fast 3D dynamic processes in vivo at signal to noise levels approaching those obtained by total internal reflection fluorescence (TIRF) illumination. Using this technique, we demonstrate substantial advantages in speed, sensitivity and reduced phototoxicity compared to conventional point scanning and spinning disc confocal microscopes as well as light-sheet microscopes utilizing single Gaussian or Bessel beams. In addition to discussing recent biological applications, I will highlight the advantages of combining lattice light sheet with: 1) single molecule super-resolution imaging and novel fluorescent probes for high spatial resolution, 2) adaptive optics for deeper penetration depth, and 3) multiplexed microscopic imaging to balance trade offs in spatiotemporal resolution while longitudinally monitoring single cells.

NOVEL TECHNIQUES FOR PROBING SINGLE CELL ADHESION AND SINGLE CELL ISOLATION

Tamás Gerecsei[1,2], Rita Ungai-Salánki[2], Milán Sztilkovics[2], Beatrix Peter[1], Sándor Kurunczi[1], Inna Székács[1], Bálint Szabó[1,2,3], Robert Horvath[1]

[1]Hungarian Academy of Sciences, Institute for Technical Physics and Materials Science, Nanobiosensorics Group, Budapest, Hungary, [2]Eötvös Loránd University, Department of Biological Physics, Budapest, Hungary, [3]Cellsorter, Company for Innovations, Budapest, Hungary

Measuring the interaction of cells with different surfaces is one of the great challenges of the emerging science of single cell biology. Extracting trustworthy, quantitative information has proved to be a difficult task using traditional methods such as AFM tip functionaliaztion. In our laboratory we apply and develop novel methods to measure single cell adhesion. Using a powerful micropipette based method we are able to separate cells or microbeads of various sizes from planar surfaces by applying a force exerted by a controlled fluid flow. The adhesion spectrum of the beads is then determined using computer vision. Such adhesion spectra has been compared to measurements by FluidFM BOT (an automated AFM based technique with a nanofluidic channel incorporated in the cantilever) on a model system. We found that the corresponding histograms can be correlated, providing a method for micropipette-based adhesion measurements on microbeads or cells.The average time needed per bead is half a minute, demonstrating the exceptional high-throughput capacity of our system [1].
We also developed a computer vision-based robot applying a micropipette to recognize and gently isolate intact individual cells for subsequent analysis, e.g., DNA/RNA sequencing from a cell suspension [2]. Our new piezoelectric micropipette with a precision of <1 nanoliter improves the efficiency of single cell isolation to above 90 % [3].
Furthermore, we executed combined experiments using FluidFM and a label-free high spatial resolution optical biosensor to develop a novel method for the measurement of the adhesion kinetics of living single cells. These unique methods provide accurate data on the adhesion force and adhesion energy values of single living cells to provide real time kinetic curves of the adhesion process.

[1] Gerecsei, Tamas, et al. "Adhesion force measurements on functionalized microbeads: an in-depth comparison of computer-controlled micropipette and fluidic force microscopy." Journal of colloid and interface science (2019).

[2] R. Ungai-Salánki, T. Gerecsei, P. Fürjes, N. Orgovan, N. Sándor, E. Holczer, R. Horvath, B. Szabó: Automated single cell isolation from suspension with computer vision, Nature Scientific Reports 6 (2016)

[3] B. Francz, R. Ungai-Salánki, É. Sautner, R. Horvath, B. Szabó: Subnanoliter precision piezo pipette for single cell isolation and droplet printing, to be published

SIMULTANEOUS QUANTIFICATION OF PROTEIN-DNA CONTACTS AND TRANSCRIPTOMES IN SINGLE CELLS

Koos Rooijers[1], Corina M Markodimitraki[1], Franka J Rang[1], Sandra S de Vries[1], Alex Chialastri[2], Kim de Luca[1], Dylan Mooijman[1], Siddharth S Dey*[2], Jop Kind*[1]

[1]Hubrecht Institute, KNAW and University Medical Center, Utrecht, Netherlands, [2]University of California Santa Barbara, Department of Chemical Engineering and Center for Bioengineering, Santa Barbara, CA

*Co-corresponding authors

The epigenome plays a critical role in regulating gene expression in mammalian cells. However, our understanding of how cell-to-cell heterogeneity in the epigenome influences gene expression variability remains limited. To address this question, we have developed a novel method for simultaneous quantification of protein-DNA contacts by combining single-cell DNA adenine methyltransferase identification (DamID) with mRNA sequencing from the same cell without requiring physical separation of the nucleic acids (scDam&T-seq). This method enables us to directly correlate protein-DNA contacts to a particular gene expression program in a cell. By profiling lamina-associated domains (LADs) in human cells, we reveal different dependencies between genome-nuclear lamina (NL) association and gene expression in single cells. Specifically, we find that gene expression heterogeneity is dependent on NL positioning for genomic regions that infrequently contact the lamina. Surprisingly though, for genomic regions that frequently contact the NL, we find that gene expression is not sensitive to NL positioning. In addition, we introduce the E. coli methyltransferase, Dam, as an *in vivo* marker of chromatin accessibility in single cells and show that scDam&T-seq can be used to simultaneously map chromatin accessibility and the transcriptome in single cells. Further, we show that scDam&T-seq can be used to map the genome-wide binding patterns of chromatin modifying factors by quantifying the genome binding landscape of a polycomb-group protein, RING1B, and the associated transcriptome. Finally, using mouse embryonic stem cells, we show that this approach can be used as a general technology to identify cell types *in silico* using the transcriptome of single cells and simultaneously determining the underlying chromatin landscape of each cell type using the corresponding protein-DNA interaction maps. Thus, scDam&T-seq will enable the analysis of protein-mediated mechanisms that regulate cell-type-specific transcriptional programs in heterogeneous tissues.

SINGLE CELL ANALYSIS….A BILLION CELLS AT A TIME

Shana Kelley

University of Toronto, Toronto, Canada

The analysis of heterogeneous ensembles of rare cells requires single-cell resolution to allow phenotypic and genotypic information to be collected accurately. We developed a new approach for high-throughput cell sorting and profiling, Magnetic Ranking Cytometry, that uses the loading of individual cells with functionalized magnetic nanoparticles as a means to report on biomarker expression at the single cell level. This approach can be used to profile circulating tumor cells in blood and provides a high-information content liquid biopsy in a single measurement. It profiles both protein (*Nature Nanotechnology*, 2017) and nucleic acid (*Nature Chemistry*, 2018) analytes at the single cell level. Recently, we have used this approach to perform high-throughput, phenotypic CRISPR screens at the whole genome level (*Nature Biomedical Engineering*, 2019) and are now using this platform as a tool for therapeutic target discovery.

INTEGRATIVE SINGLE-CELL ANALYSIS RECONSTRUCTS THE QUANTITATIVE CELL STATE LANDSCAPE OF THE HUMAN DEVELOPING CORTEX.

Jose Davila-Velderrain[1,2], Shahin Mohammadi[1,2], Manolis Kellis[1,2]

[1]MIT, Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, [2]Broad Institute of MIT and Harvard, Cambridge, MA

Single-cell technologies provide a high-throughput approach to simultaneously survey the transcriptomic state distribution across populations of thousands of single-cells. This technical development, in turn, provides the opportunity to empirically characterize the observable cell state space associated with a developmental process. Here we study the cell state space of an intricate cellular process: embryonic neurogenesis within the human developing cortex. Cell state heterogeneity is particularly pronounced during neurogenesis, as different classes of neural progenitor cells dynamically balance proliferation and differentiation to generate migrating neurons. This process results in cellular heterogeneity within and in-between multipotent and differentiated cellular populations, producing a trace of transitory cell states. Using a novel computational framework to quantify hierarchical latent cell states, we analyze publicly available, as well as novel, single-cell transcriptomic data of human fetal cortices at the pick of neurodegeneration. Independent and joint analysis of cells at different postconceptional weeks allows the reconstruction of cell state ontologies underlying a quantitative cell state landscape. The topography of the landscape recovers a stereotypical, time-dependent arrangement of neuron progenitor cells (NPCs), intermediate progenitors (IPs), and neurons (Ns). The relative organization of subregions of the state space is consistent with the spatial localization of observed cell-types, known cellular events across the embryonic cortical zones, and with the orientation of neuronal subpopulations. Extensive quantitative mapping of single-cell observations with rich tissue-level bulk references facilitate the interpretation of and support the conclusions of the reconstructed cell state landscape.

IMAGING AND SEQUENCING PROTEIN-DNA INTERACTIONS IN SINGLE CELLS

Aaron Streets[1,2]

[1]University of California, Berkeley, Bioengineering, Berkeley, CA, [2]Chan-Zuckerberg Biohub, Investigator, San Francisco, CA

Genome regulation depends on carefully programmed protein-DNA interactions that maintain or alter gene expression states, often by influencing chromatin organization. Most studies of these interactions to date have relied on ensemble methods to identify the sequence position of protein binding. In many systems, such bulk methods cannot capture the dynamic single-cell nature of these interactions as they modulate cell states. DNA adenine methyltransferase identification (DamID), is a technique which records a protein's DNA-binding history by methylating adenine bases in its vicinity, then selectively amplifying these methylated regions. When coupled with high-throughput sequencing, DamID can achieve high enough sensitivity to enable mapping of protein-DNA interactions in single-cells. Here, we implement DamID in an integrated microfluidic platform for single-cell processing. This technique, called micro-DamID (μDamID) allows imaging of the spatial location of protein-DNA interaction sites in the nucleus of single cells prior to lysis and DamID processing using fluorescent proteins that bind to methyladenines. We apply this system to generate paired single-cell imaging and sequencing data from a human cell line, in which we map and validate interactions between DNA and nuclear lamina proteins, providing a measure of chromatin organization and broad gene regulation patterns. μDamID provides the unique ability to compare paired imaging and sequencing data for each cell and between cells, enabling the joint analysis of the nuclear localization, sequence identity, and variability of protein-DNA interactions.

THREE DIMENSIONAL NANOSCALE MOLECULAR IMAGING BY EXTREME ULTRAVIOLET LASER ABLATION MASS SPECTROMETRY

Carmen S Menoni, Ilya Kuznetsov, Jorge J Rocca

Colorado State University, Electrical & Computer Engineering, Fort Collins, CO

Analytical probes capable of mapping molecular composition in 3-D at the nanoscale will transform materials research, biology and medicine. Mass-spectral imaging (MSI) is one of the most powerful methods to visualize the spatial organization of multiple molecular components on solid samples. However, it is challenging for MSI to map molecular composition in 3-D with submicron resolution. We have demonstrated a new MSI method that combines extreme ultraviolet (EUV) laser ablation with mass spectrometry to obtain 3-D composition images with nanoscale resolution [1]. In EUV MSI, bright laser pulses from a compact 46.9-nm-wavelength laser [2] are focused into nanometer size spots to ablate craters a few nanometers deep on selected regions of the sample. Elemental and molecular ions in the laser-created plasma are extracted and identified by their mass-to-charge ratio (m/z) using a time-of-flight (TOF) mass spectrometer. Analysis of the spatially resolved mass spectra obtained as the sample is displaced with respect to the focused laser beam enables one to construct 3-D composition images with nanoscale resolution. In this talk, I will describe advances of EUV MSI that show the unique capabilities of the method to identify intact organic analyte ions, and to map molecular composition in 3-D of a single micro-organism. In addition, I will describe the implementation of post-ablation soft ionization that is aimed at selectively ionize the most abundant neutrals in the EUV laser-created plasma. Post-ablation ionization is implemented using the 9th harmonic (118 nm) of Yb:YAG. Focusing the vacuum ultraviolet (VUV) light onto the sample region allows for softer ionization of intact neutral molecules. Using selected organic solids, it will be shown that the combination with EUV MSI with post-ablation ionization provides increased imaging resolution, sensitivity and mass detection range. In combination, these results open up attractive opportunities to visualize composition in biological systems with with high spatial resolution and sensitivity.

[1] I. Kuznetsov et al, "Three dimensional nanoscale molecular imaging by extreme ultraviolet laser ablation mass spectrometry," Nature Communications, **6,** Article No. 6944 (2015); doi:10.1038/ncomms7944.
[2] S. Heinbuch et al, "Demonstration of a desk-top size high repetition rate soft x-ray laser," Opt. Express **13**, 4050-4055 (2005).

SINGLE-CELL TRANSCRIPTOMICS OF THE MOUSE KIDNEY REVEALS POTENTIAL CELLULAR TARGETS OF KIDNEY DISEASE

Katalin Susztak

University of Pennsylvania, Renal Electrolyte and Hypertension Division, Department of Medicine and Genetics, Philadelphia, PA

Our understanding of kidney disease pathogenesis is limited by an incomplete molecular characterization of the cell types responsible for the organ's multiple homeostatic functions. To help fill this knowledge gap, we characterized 57,979 cells from healthy mouse kidneys by using unbiased single-cell RNA sequencing. On the basis of gene expression patterns, we infer that inherited kidney diseases that arise from distinct genetic mutations but share the same phenotypic manifestation originate from the same differentiated cell type. We also found that the collecting duct in kidneys of adult mice generates a spectrum of cell types through a newly identified transitional cell. Computational cell trajectory analysis and in vivo lineage tracing revealed that intercalated cells and principal cells undergo transitions mediated by the Notch signaling pathway. In mouse and human kidney disease, these transitions were shifted toward a principal cell fate and were associated with metabolic acidosis.

THE ORIGIN AND INDUCTION PROCESSES OF HAIR FOLLICLE STEM CELLS

Ritsuko Morita[1], Noriko Sanzen[1], Tetsutaro Hayashi[2], Mana Umeda[2], Mika Yoshimura[2], Itoshi Nikaido[2], Takaya Abe[3], Hiroshi Kiyonari[3,4], Yasuhide Furuta[3,4], Hironobu Fujiwara[1]

[1]RIKEN Center for Biosystems Dynamics Research, Laboratory for Tissue Microenvironment,, Kobe, Japan, [2]RIKEN Center for Biosystems Dynamics Research, Laboratory for Bioinformatics Research, Kobe, Japan, [3]RIKEN Center for Biosystems Dynamics Research, Laboratory for Animal Resource Development, Kobe, Japan, [4]RIKEN Center for Biosystems Dynamics Research, Laboratory for Genetic Engineering, Kobe, Japan

Tissue stem cells (SCs) are generated from an embryonic progenitor population as a result of spatiotemporal cell proliferation and differentiation. Although many fundamental regulatory mechanisms of adult tissue SCs have been explored, it remains obscure how tissue SCs are induced in the right place at the right time during development, mainly due to the lack of specific markers that exclusively label prospective SCs. In this study, we investigate the molecular and cellular mechanisms underlying induction of hair follicle SCs with two marker-independent single cell methods; single-cell live imaging and single-cell transcriptomics. Retrospective tracking of the stem and differentiated cells in our live imaging system of developing hair follicles enable us to continuously track individual cell behaviours within tissues. Single-cell transcriptome analysis identified new cell types, which cannot be distinguished by known makers, their signature markers and transcriptional state changes in developing hair follicles. We then integrated the transcriptional profile, in vivo location and developmental lineages of each cell type within tissues. Integration of the knowledge from single-cell live imaging and single-cell transcriptomics demonstrated that i) different epidermal lineages are aligned in a concentric manner in the basal layer of the hair placode, ii) each 2D concentric zone in the placode is enlarged and telescopes out to form longitudinally aligned 3D cylindrical compartments in the hair follicle, as seen in the development of Drosophila imaginal disk, iii) prospective SC compartments are derived from the basal layer of the epidermal cells located at the peripheral zone of the placode, iv) SC precursors do not significantly contribute to the hair follicle morphogenesis, v) molecular properties of adult SCs are gradually induced during hair follicle morphogenesis. Taken together, our marker-independent single-cell omics analysis allows us to understand the origin and developmental trajectories of tissue SCs, leading to the new model of hair follicle development and SC induction with unprecedented resolution.

WHOLE-BODY SINGLE-CELL RNA SEQUENCING REVEALS COMPONENTS OF ELEMENTARY NEURAL CIRCUITS IN A SPONGE

Jacob M Musser[1], Klaske Schippers[1], Michael Nickel[1], Giulia Mizzon[1], Leonid Moroz[2], Detlev Arendt[1]

[1]European Molecular Biology Laboratory, Developmental Biology, Heidelberg, Germany, [2]University of Florida, Whitney Laboratory for Marine Bioscience, St. Augustine, FL

Sponges are the sister group to nearly all other animals, and lack a nervous system, musculature, and gut. However, genes encoding important neuronal proteins, including key synaptic proteins, have been found in sponge genomes. Using single-cell RNAseq, single-molecule FISH, and Focused Ion Beam SEM (FIB-SEM) we provide the first deep molecular and morphological characterization of cell types in *Spongilla lacustris*, freshwater demosponge. We identify many specialized cell types bearing functional and regulatory signatures similar to those of other animals. This includes contractile epithelial cells, which we demonstrate experimentally are responsive to nitric oxide signaling, phagocytes involved in innate immunity, and digestive cells that express a nearly complete set of postsynaptic genes. Remarkably, we also find immune cells expressing presynaptic genes and show via FIB-SEM that they send neurite-like extensions that directly contact and enwrap microvilli on postsynaptic digestive cells. This reveals new evidence linking neuronal and immune function, and suggests a neuro-immune system may have predated the origins of true nervous and immune systems.

COMPREHENSIVE SINGLE-CELL TRANSCRIPTOME LINEAGES OF A PROTO-VERTEBRATE

Chen Cao[1], Laurence A Lemaire*[1], Wei Wang[2], Peter H Yoon[3], Yoolim A Choi[3], Lance R Parsons[1], John C Matese[1], Wei Wang[1], Michael S Levine[1,3], Kai Chen[1]

[1]Princeton University, Lewis-Sigler Institute for Integrative Genomics, Princeton, NJ, [2]Stowers Institute for Medical Research, Stowers Institute for Medical Research, Kansas City, MO, [3]Princeton University, Department of Molecular Biology, Princeton, NJ

Ascidian embryos highlight the importance of cell lineages in animal development. As simple proto-vertebrates, they also provide insights into the evolutionary origins of cell types such as cranial placodes and neural crest cells. Here we have determined single-cell transcriptomes for more than 90,000 cells that span the entirety of development—from the onset of gastrulation to swimming tadpoles—in Ciona intestinalis. Owing to the small numbers of cells in ascidian embryos, this represents an average of over 12-fold coverage for every cell at every stage of development. We used single-cell transcriptome trajectories to construct virtual cell-lineage maps and provisional gene networks for 41 neural subtypes that comprise the larval nervous system. We summarize several applications of these datasets, including annotating the synaptome of swimming tadpoles and tracing the evolutionary origin of cell types such as the vertebrate telencephalon.

These authors contributed equally: Chen Cao, Laurence A. Lemaire.

SINGLE CELL ANALYSES REVEAL DEVELOPMENTAL BOTTLENECK AND TWO WAVES OF BLOOD PROGENITOR FORMATION FROM ARTERIAL ENDOTHELIUM

Qin Zhu[1], Peng Gao[2], Joanna Tober[3], Laura Bennett[3], Changya Chen[2], Yasin Uzun[2], Yan Li[3], Nancy A Speck[3], Kai Tan[2,4,5]

[1]University of Pennsylvania, Genomics and Computational Biology Graduate Group, Philadelphia, PA, [2]The Children's Hospital of Philadelphia, Division of Oncology and Center for Childhood Cancer Research, Philadelphia, PA, [3]University of Pennsylvania, Abramson Family Cancer Research Institute and Department of Cell and Developmental Biology, Philadelphia, PA, [4]University of Pennsylvania, Department of Pediatrics, Philadelphia, PA, [5]University of Pennsylvania, Department of Genetics, Philadelphia, PA

Hematopoietic stem and progenitor cells (HSPCs) in the bone marrow are derived from a small population of hemogenic endothelial (HE) cells located in the yolk sac and major caudal arteries of the mammalian embryo. HE cells undergo an endothelial to hematopoietic cell transition (EHT), giving rise to HSPCs that accumulate in intra-arterial clusters (IACs) before colonizing the fetal liver. To examine the molecular transitions between endothelial (E), HE, and IAC cells, and the heterogeneity of HSPCs within the IACs, we profiled ~40,000 cells from the caudal arteries (dorsal aorta, umbilical, vitelline) of embryonic day (E) 9.5 to 11.5 mouse embryos by single-cell RNA sequencing (scRNA-seq) and single-cell chromatin accessibility sequencing (scATAC-Seq). A continuous developmental trajectory leads from E to IAC cells, with identifiable intermediate stages between E and HE. The intermediate stage most proximal to HE, which we have termed pre-HE, is characterized by elevated expression of genes regulated by GATA and SOX transcription factors. Developmental bottlenecks separate E from pre-HE, and pre-HE from HE, with the efficiency of transit through the latter bottleneck regulated by RUNX1 dosage. Distinct developmental trajectories within IAC cells result in two populations of CD45+HSPCs; an initial wave of multi-lineage committed progenitors followed by precursors of hematopoietic stem cells (pre-HSCs).

RECONSTRUCTION OF AN INVARIANT LINEAGE IN CHORDATE EMBRYOGENESIS

Tengjiao Zhang*[1], Yichi Xu*[2], Teng Fei[3], Kaoru Imai[4], Tianwei Yu[3], Yutaka Satou[5], Zhirong Bao[2], Weiyang Shi[1,6]

[1]Tongji University, School of Life Sciences and Technology, Shanghai, China, [2]Sloan Kettering Institute, Developmental Biology Program, New York, NY, [3]Emory University, Department of Biostatistics and Bioinformatics, Atlanta, GA, [4]Osaka University, Department of Biological Sciences, Osaka, Japan, [5]Kyoto University, Department of Zoology, Kyoto, Japan, [6]Ocean University of China, College of Marine Life Sciences, Qingdao, China

In multicellular organisms, a simple fertilized egg develops along divergent embryonic lineages to produce distinct cell types. What governs these processes is central to the understanding of cell fate specification and stem cell engineering. Recent advances of single-cell technologies enabled systematic characterization of cell types as well as the molecular programs that define them. Here we examined the molecular programs of lineage specification in chordate embryogenesis. We reconstructed the entire embryonic cell lineage in the ascidian Ciona savignyi with single-cell transcriptome profiling up to the 110-cell stage and compared the molecular programs to those of the mouse. We showed that early Ciona gene expression is dominated by complex combinatorial expression patterns and rapid turnover, and further provided molecular insights on FGF-induced fate specification and asymmetric cell division. Between species, we showed a substantial drift of gene expression over chordate evolution, with similarity between homologous cell types largely limited to key regulators of cell fates. This study demonstrates a general approach to study embryonic development and gene network regulation across a broad spectrum of non-traditional model organisms at an unprecedented single cell resolution.

*Those authors contributed equally.

SINGLE-CELL TRANSCRIPTOME ANALYSIS OF CD34+ STEM CELL-DERIVED MYELOID CELLS INFECTED WITH HUMAN CYTOMEGALOVIRUS

Laura Hertel[1], Melissa Galinato[1], Kristen Shimoda[1], Alexis Aguiar[1], Fiona Hennig[2], Dario Boffelli[2], Michael McVoy[3]

[1]Children's Hospital Oakland Research Institute, Center for Immunobiology and Vaccine Development, Oakland, CA, [2]Children's Hospital Oakland Research Institute, Center for Genetics, Oakland, CA, [3]Virginia Commonwealth University, Department of Pediatrics, Richmond, VA

Myeloid cells are important sites of lytic and latent infection by human cytomegalovirus (CMV). We previously showed that only a small subset of myeloid cells differentiated from CD34+ hematopoietic stem cells is permissive to CMV replication, underscoring the heterogeneous nature of these populations. The exact identity of resistant and permissive cell types, and the cellular features characterizing the latter, however, could not be dissected using averaging transcriptional analysis tools such as microarrays and, hence, remained enigmatic. Here, we profile the transcriptomes of ~7000 individual cells at day 1 post-infection using the 10X genomics platform. We show that viral transcripts are detectable in the majority of the cells, suggesting that virion entry is unlikely to be the main target of cellular restriction mechanisms. We further show that viral replication occurs in a small but specific sub-group of cells transcriptionally related to, and likely derived from, a cluster of cells expressing markers of Colony Forming Unit – Granulocyte, Erythrocyte, Monocyte, Megakaryocyte (CFU-GEMM) oligopotent progenitors. Compared to the remainder of the population, CFU-GEMM cells are enriched in transcripts with functions in mitochondrial energy production, cell proliferation, RNA processing and protein synthesis, and express similar or higher levels of interferon-related genes. While expression levels of the former are maintained in infected cells, the latter are strongly down-regulated. We thus propose that the preferential infection of CFU-GEMM cells may be due to the presence of a pre-established pro-viral environment, requiring minimal optimization efforts from viral effectors, rather than to the absence of specific restriction factors. Together, these findings identify a potentially new population of myeloid cells permissive to CMV replication, and provide a possible rationale for their preferential infection.

INFERRING GENE REGULATORY LANDSCAPES FROM SINGLE-CELL DATA

Erik van Nimwegen

University of Basel, Basel, Switzerland

I will discuss a new method for inferring, from scRNA-seq data, the transcription regulatory interactions that guide single-cell gene expression trajectories. The method combines three new ideas: First, a scRNA-seq normalization method that rigorously deconvolves sampling noise from true variations in transcription rates. Second, a Bayesian method that infers the 'regulatory states' of each single cell by modeling measured transcription rates in terms of genome-wide computational predictions of transcription factor binding sites. And third, a maximum entropy approach that infers an effective epigenetic landscape that guides the distribution of single cells in the space of regulatory states.

Time permitting I will briefly mention studies of our lab on single-cell gene regulation in bacteria using a combination of microfluidics and time-lapse microscopy. These studies highlight how gene regulation is not only strongly coupled to fluctuations in the physiological state of cells but that, by propagation of noise through the regulatory network, expression noise and gene regulation are intimately entangled.

TRANSCRIPTIONAL CONTROL STRATEGIES FOR PATTERN FORMATION IN EMBRYONIC DEVELOPMENT

Nicholas C Lammers[1], Vahe Galstyan[2], Armando Reimer[1], Sean A Medin[3], Chris H Wiggins[4], Hernan G Garcia[1,3]

[1]UC Berkeley, Biophysics Graduate Group, Berkeley, CA, [2]Caltech, Biochemistry and Molecular Biophysics Option, Pasadena, CA, [3]UC Berkeley, Molecular & Cell Biology and Physics, Berkeley, CA, [4]Columbia University, Applied Physics and Applied Mathematics, New York, NY

During embryonic development, tightly choreographed patterns of gene expression—shallow gradients, sharp steps, narrow stripes—specify cell fates. Predicting developmental outcomes from these transcription factor patterns and from regulatory DNA sequence remains an open challenge in physical biology. Such predictive power necessitates a quantitative understanding of the mechanisms that facilitate the flow of information along the central dogma. We used stripe 2 of the even-skipped gene in Drosophila embryos as a case study in the dissection of the regulatory forces underpinning a key step along the developmental decision-making cascade: the generation of cytoplasmic mRNA patterns via the control of transcription in individual cells. Using live imaging and computational approaches, we found that the transcriptional burst frequency is modulated across the stripe to control the mean mRNA production rate. However, we discovered that bursting alone cannot quantitatively recapitulate the formation of the stripe, and that control of the window of time over which each nucleus transcribes even-skipped plays a critical role in stripe formation. Theoretical modeling revealed that these two regulatory strategies—the analogue control of the mean transcription rate by bursting and the binary control of the transcription time window—obey different kinds of regulatory logic, suggesting that the stripe is shaped by the interplay of two distinct underlying molecular processes. Thus, our work provides an example of how biological numeracy can be used as driver for discovery. Further, it provides a stark reminder of how reaching a predictive understanding of developmental decision-making will require a precise understanding of the regulation of gene expression timing in embryos.

MASSIVELY MULTIPLEXED CHEMICAL TRANSCRIPTOMICS AT SINGLE-CELL RESOLUTION

Sanjay Srivatsan*[1,2], José L McFaline-Figueroa*[1], Vijay Ramani*[3], Lauren M Saunders[1], Junyue Cao[1], Jonathan S Packer[1], Hannah A Pliner[1], Dana Jackson[1], Riza M Daza[1], Lena Christiansen[4], Fan Zhang[4], Frank Steemers[4], Jay Shendure#[1,5,6,7], Cole Trapnell#[1,5,7]

[1]University of Washington, Department of Genome Sciences, Seattle, WA, [2]University of Washington, Medical Scientist Training Program, Seattle, WA, [3]University of California San Francisco, Department of Biochemistry and Physics, San Francisco, CA, [4]Illumina Inc., San Diego, CA, [5]University of Washington, Allen Discovery Center for Cell Lineage Tracing, Seattle, WA, [6]University of Washington, Howard Hughes Medical Institute, Seattle, WA, [7]University of Washington, Brotman Baty Institute for Precision Medicine, Seattle, WA
*Authors contributed equally #Co-senior authors

High-throughput screens (HTS) are fundamental to the drug discovery pipeline allowing for the characterization of chemical libraries with ease. However, conventional HTS is generally limited to determining the effect of small molecules on a simple phenotypic outcome (e.g. proliferation or cellular morphology) or a highly-specific molecular readout (e.g. enzyme activity or expression of a reporter). Chemical transcriptomic screens allow for global measurements of changes to transcriptomes following small molecule exposure. A limitation of these bulk assays is that even cells of the same "type" can exhibit heterogeneity in their response. Here we combine sample multiplexing via nuclear "hashing" and the throughput of single-cell combinatorial indexing RNA-seq (sci-RNA-seq) to create an HTS compatible single-cell RNA-seq platform which we term "sci-chem".

We apply our sci-chem paradigm to comprehensively profile the transcriptional response of 3 cancer cell lines to a collection of 188 compounds each at 4 doses and in biological duplicate. Compounds assayed included small molecules targeting epigenetic and transcriptional components as well as diverse cellular processes including replication, protein phosphorylation and protein folding, amongst others. In total, we profiled close to 5,000 independent samples across 649,340 single-cell transcriptomes. Our analysis identified heterogeneous responses of cells to specific compounds and recovered commonalities in the transcriptional response of cells treated with compounds with similar mechanisms of action, including for those that do not directly alter transcription.

We observe that exposure to one of several HDAC inhibitors results in a concerted response characterized by an exit from the cell cycle and upregulation of genes involved with increasing acetyl-CoA pools. These results and additional validation experiments support the view that chromatin acts as an important reservoir of acetate in cancer cells.

SIMULTANEOUS CAPTURE OF DNA AND mRNA FROM SINGLE MAMMALIAN CELLS IN DROPLETS

Claire M Bell, Aniruddha M Kaushik, Pengfei Zhang, Weixiang Fang, Hongkai Ji, Tza-Huei Wang, Donald J Zack

Johns Hopkins University, Baltimore, MD

Combining genome and transcriptome analysis at the single-cell level allows the direct correlation between gene expression and genotype on a cell by cell basis. This capability can extend a range of single-cell applications, from associating gene expression changes with genome editing to diagnostics of tissue samples. We have developed a novel droplet-based method for the parallel characterization of genomic sequences and the transcriptome of thousands of single cells in a given experiment. The method is based upon the Drop-seq workflow; a microfluidic chip is used to co-encapsulate single Drop-seq beads suspended in a custom lysis buffer and single cells into droplets. The beads are coated with oligos containing bead-specific and transcript-specific barcodes as well as a poly-T sequence designed to capture mRNA upon cell lysis within the droplet. To capture genomic sequence information, we introduced an in-droplet PCR amplification step that uses oligonucleotides that incorporate poly-A tails into the DNA amplicons, enabling their capture on the same oligo-coated Drop-seq beads. After droplet breakage, both the mRNA and amplified DNA molecules are copied off of the beads, resulting in the incorporation of the unique cell-specific barcodes into both the cDNA and amplified DNA sequences. Two types of libraries, the standard cDNA library as well as a "DNA amplicon" library, are then processed and sequenced separately. The sequenced transcripts are bioinformatically paired with the DNA amplicon sequences, based on the shared bead-specific barcode. We demonstrated this method by performing in-droplet genotyping and transcriptomic profiling of a mixed population of edited and unedited human stem cells. Using CRISPR/Cas9 genome editing, we introduced a stop codon into either the *C-MYC* or *KLF4* gene. Single cells from the entire population of transfected cells- without clonal selection, expansion or genotyping- were captured in droplets. In-droplet PCR was performed with primers designed to amplify the Cas9-targeted regions. After droplet breakage, the cDNA and amplicon samples were processed, and the libraries constructed were sequenced. Based on the shared bead-barcodes between DNA amplicons and transcripts, each cell was assigned a genotype, allowing for direct comparison of differential gene expression between cells with divergent genotypes. This method could be a powerful approach for applications including comparing cellular heterogeneity in clinical settings, eliminating batch effects in single cell RNA-seq experiments, determining genetic threshold effects, and combining DNA-barcoding with single cell transcriptomics.

PERTURBATION PANEL PROFILING IDENTIFIES TRANSCRIPTION FACTORS THAT ENHANCE DIRECTED CHANGES OF CELL IDENTITY

Ian A Mellis[1,2], Hailey Edelstein[3], Rachel Truitt[3], Lauren Beck[1], Margaret C Dunagin[1], Ricardo A Linares[4], Parisha P Shah[4], Wenli Yang[3], Rajan Jain[4], Arjun Raj[1,2,5]

[1]University of Pennsylvania, Bioengineering, Philadelphia, PA, [2]University of Pennsylvania, Genomics & Computational Biology, Philadelphia, PA, [3]University of Pennsylvania, Institute for Regenerative Medicine, Philadelphia, PA, [4]University of Pennsylvania, Cell and Developmental Biology, Philadelphia, PA, [5]University of Pennsylvania, Genetics, Philadelphia, PA

It is extremely difficult to change the cellular identity of differentiated cells, both in terms of picking the combination of factors needed to direct cells toward a cell type of interest and in terms of achieving a high efficiency of conversion. These challenges are present when converting cells either directly to another cell type (transdifferentiation) or to pluripotency (reprogramming). For example, many protocols for transdifferentiation of human fibroblasts to cardiomyocytes convert only ~1% of cells, if any at all, and for some human cell types reprogramming to pluripotency is even less efficient. We aim to better understand and then engineer cell type-intrinsic patterns of gene regulation underlying the maintenance of cell identity for two human cell types of interest: fibroblasts and cardiomyocytes. To this end, we conducted transcriptome-wide expression profiling using RNAtag-seq on panels of hundreds of samples of drug-perturbed fibroblasts and cardiomyocytes. We found that the expression levels of genes encoding transcription factors known to enable cardiac transdifferentiation and to regulate cardiomyocyte development are more frequently up-regulated across a broad range of drug conditions in cardiomyocytes than other expressed transcription factor genes. Further, in fibroblasts, we discovered that several frequently perturbed transcription factors were barriers to reprogramming; knockdown of transcription factors often perturbed in fibroblasts increased the efficiency of Yamanaka factor-mediated reprogramming to pluripotency. Together, these results suggest that, when overexpressed, transcription factors that are often perturbed in a given cell type can drive other cells toward that cell type, and, inversely, when knocked down they can enable more efficient conversion away from that cell type. In summary, we hope to establish high “perturbability” as a useful property for experimentally identifying transcription factors for use in directed changes in cellular identity and important for cellular identity maintenance.

SINGLE-CELL PROTEOMICS/METABOLOMICS IN THE DEVELOPING EMBRYO AND THE BRAIN USING MASS SPECTROMETRY

Peter Nemes

University of Maryland, Department of Chemistry & Biochemistry, College Park, MD

Formation and maintenance of molecular differences between cells (cell heterogeneity) is critical to normal development and function, but how proteins and small molecules (metabolites) coordinate cell heterogeneity during embryonic patterning is little understood. The challenge has been mass spectrometry (MS) characterization of low signal intensities that are yielded by miniscule amounts of proteins and metabolites contained in single cells. While the classical approach to combine multiple cells certainly enhances detection sensitivity, cell pooling averages out signals across cells, losing potentially important cell-specific molecular events, such as those responsible for inducing differential cell fates and neurogenesis. In this presentation, we will discuss microprobe capillary electrophoresis electrospray ionization (CE-ESI) MS that our laboratory has developed to enable the detection of hundreds-to-thousands of proteins and hundreds of metabolite signals in identified single cells in the early South African clawed frog (*Xenopus laevis*), a powerful model in cell and developmental biology, and electrophysiologically identified single neurons in the mouse cortex. We will also provide vignettes on how we use these single-cell mass spectrometers to quantify previously unknown proteomic and metabolic cell heterogeneity in the 4-to-128-cell embryo, discover molecules that can alter the normal tissue fate of embryonic cells, and how these results can be leveraged to formulate new hypotheses regarding the functional significance of proteins or metabolites during early embryogenesis.

TECHNICAL ABSTRACTS
FOR WORKSHOPS

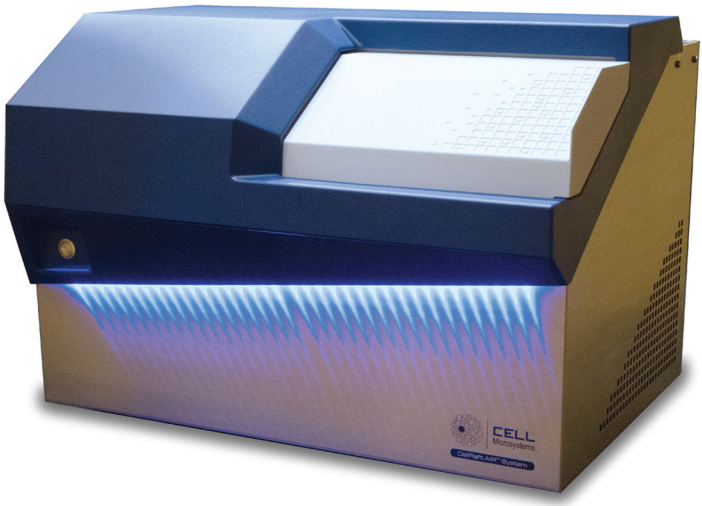
AUTOMATED IMAGING, SORTING AND ISOLATION OF SINGLE CELLS USING THE CELLRAFT AIR SYSTEM

Cell Microsystems, Inc. Research Triangle Park, North Carolina

The CellRaft AIR™ System is a bench-top instrument capable of imaging, sorting and isolating single cells for downstream molecular analysis. The system is based on Cell Microsystems' core CellRaft technology which employs a microwell array for culturing and phenotyping cells prior to isolation. Both genomic and transcriptomic workflows have been optimized and published using the system to analyze single cells. The CytoSort Array, Cell Microsytems' proprietary cell culture consumable, closely replicates standard *in vitro* conditions, rendering the array an ideal substrate for time-course phenotyping, drug sensitivity assays and other imaging-based evaluation prior to isolation of cells for single cell molecular analysis. The system has been shown to be effective in isolating single cells from cell lines, primary samples, non-adherent cells and stem cells such as iPSCs and human embryonic stem cells. Using the imaging capabilities of the CellRaft AIR System, unique and subtle phenotypes can be leveraged as sorting criteria. Also, each CellRaft is imaged prior to isolation, providing validation for single cell isolation (i.e. elimination of empties and multiplets) and confirmation of other phenotypes which can be linked to downstream sequencing data. At this workshop, users of the CellRaft AIR™ System will share data from their experience using the technology for single cell analysis and present complete workflows.

The CellRaft® AIR

Image, Sort, Isolate.



Expand your single cell analysis research with imaging-based sorting.

The CellRaft AIR System integrates single cell imaging and viable cell isolation, with no minimum input requirement – and works with virtually any cell type.

- **Phenotyping by imaging provides more information than RFUs alone**
- **Perform 'on-array' functional assays like co-culture or drug challenges**
- **Verify live single cells and eliminate empties, doublets, or dying cells**
- **Flexible for any downstream process – NGS, multi-omics, expansion culture**
- **Inform, QC, and cross-reference your genomic data analysis with cell images**

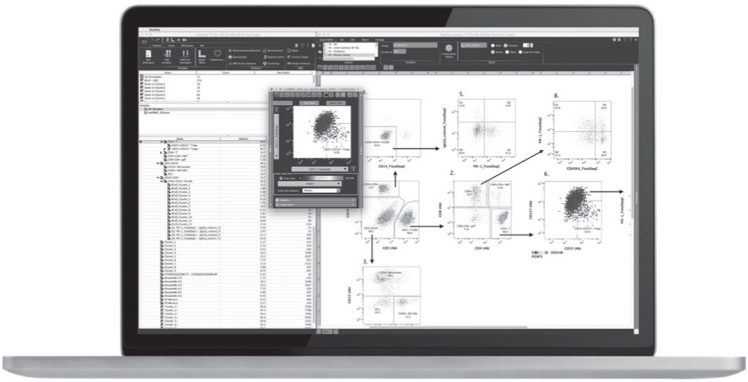
Learn more about the CellRaft AIR system and how it can accelerate your single cell analysis research at cellmicrosystems.com

CELL Microsystems®

SeqGeq™ v1.6 helps you extract insight from the individuality of single-cell gene expression. Each cell is a complex, unique individual, containing between 50,000 and 300,000 transcripts. SeqGeq [seek-geek] analyzes gene expression data, particularly from single-cell RNA sequencing. SeqGeq was built by and for scientists, and has powerful data reduction tools to help you discover, cluster, and subset populations and relationships. It also includes visualization tools—such as heat mapping and report editors—to help you communicate your results. With an open plugin architecture, SeqGeq provides the user access to workflows and cutting-edge tools like Seurat, Monocle, and ClassyDL (Deep Learning) through a friendly point-and-click interface.

SeqGeq v1.6

from the makers of FlowJo™



The world is waiting for your results. Unlock your scRNA seq data in minutes.

SeqGeq v1.6 [seek-geek] is a desktop bioinformatics platform that makes complex scRNA seq analysis accessible with an intuitive interface.

With SeqGeq, you control your analysis.
No more writing R scripts to visualize your data.
Easily share results for publication and collaboration.

flowjo.com

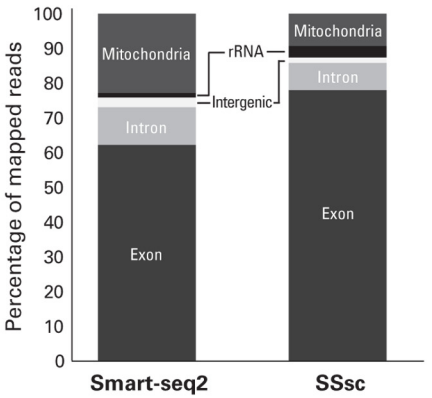
BD and FlowJo are one!
Supporting you and your research from design to discovery.

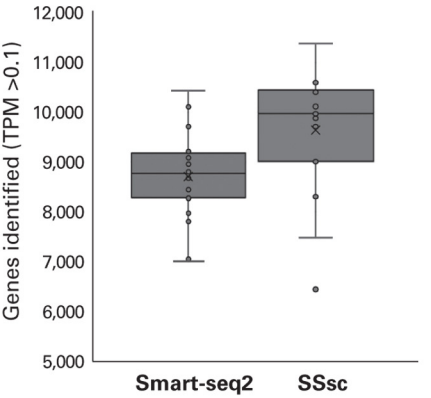
Our robust scRNA-seq technology just got better

Get data you didn't know you were missing

that's GOOD science!

Introducing the SMART-Seq® Single Cell Kit (SSsc), which features high sensitivity and reproducibility and outperforms all current full-length methods for single-cell and nuclei applications.





Higher sensitivity results in a greater number of useful reads and more genes detected than the popular full-length method.

To learn more about our new single-cell kit, visit takarabio.com/single-cell-seq

Takara Bio USA, Inc.
United States/Canada: +1.800.662.2566 • Asia Pacific: +1.650.919.7300 • Europe: +33.(0)1.3904.6880 • Japan: +81.(0)77.565.6999

TaKaRa

Clontech TaKaRa cellartis

takarabio.com

NOTES

NOTES

NOTES

NOTES

NOTES

NOTES

Participant List

Mr. Dominic Acri
IU School of Medicine/Stark Neuro Research Inst.
dacri@iu.edu

Mr. Faraz Ahmed
Cornell University
fahmed@cornell.edu

Mr. Tyler Alban
Cleveland Clinic
albant@ccf.org

Dr. Nancy Allbritton
University of North Carolina
nlallbri@email.unc.edu

Dr. Opal Allen
Huntsman Cancer Institute, University of Utah
opal.allen@hci.utah.edu

Ms. Abeer Almutairy
Camp4 Therapeutics Inc.
almutairy@camp4tx.com

Dr. Maneesha Aluru
Georgia Institute of Technology
maneesha.aluru@biology.gatech.edu

Dr. Lin An
CAMP4
an@camp4tx.com

Dr. Praveen Anand
Dana-Farber Cancer Institute
praveen_anand@dfci.harvard.edu

Dr. Courtney Anderson
Advanced Cell Diagnostics, A Bio-Techne Brand
courtney.anderson@bio-techne.com

Dr. Russell Bainer
Maze Therapeutics
rbainer@mazetx.com

Ms. Nidhanjali Bansal
Takara Bio USA, Inc
nidhanjali_bansal@takarabio.com

Dr. Sergio Barberan Soler
RealSeq Biosciences, Inc.
sbarberan@realseqbiosciences.com

Mr. Spencer Barnes
University of Texas Southwestern
spencer.barnes@utsouthwestern.edu

Dr. Thomas Bartlett
University College London
thomas.bartlett.10@ucl.ac.uk

Dr. Amir Bayegan
Celgene
a.h.bayegan@gmail.com

Ms. Claire Bell
Johns Hopkins University School of Medicine
cwenger3@jhmi.edu

Dr. Tyler Best
NIH
tyler.best@nih.gov

Ms. Ash Blibaum
Cold Spring Harbor Laboratory
blibaum@cshl.edu

Dr. Nathalie Bolduc
Takara Bio USA
nathalie_bolduc@takarabio.com

Mr. Daniel Borshagovski
University of Helsinki
daniel.borshagovski@gmail.com

Mr. Logan Brase
Washington University in St. Louis
brasel@wustl.edu

Prof. Long Cai
Caltech
longcai@gmail.com

Dr. Chen Cao
Princeton University
chencao@princeton.edu

Mr. Guoshuai Cao
University of Chicago
guoshuaicao@uchicago.edu

Dr. Yun-Juan Chang
Rutgers University
yc759@oarc.rutgers.edu

Mr. Deeptiman Chatterjee
Tulane University
dchatterjee@tulane.edu

Dr. Jian Chen
CHDI Management
jian.chen@chdifoundation.org

Dr. Xiaoyin Chen
Cold Spring Harbor Laboratory
xichen@cshl.edu

Mr. Alex Chialastri
University of California, Santa Barbara
alexchialastri@ucsb.edu

Dr. Tracy Chow
Celgene
thchow@celgene.com

Dr. Elizabeth Christie
Peter MacCallum Cancer Centre
liz.christie@petermac.org

Mr. Anthony Colombo
Cedars-Sinai Medical Center
anthony.colombo@cshs.org

Ms. Cassidy Danyko
Cold Spring Harbor Laboratory
cdanyko@cshl.edu

Dr. Jose Davila-Velderrain
MIT
jdavilav@mit.edu

Ms. Erica DePasquale
Cincinnati Children's Hospital Medical Center
careyea@mail.uc.edu

Dr. Siddharth Dey
University of California, Santa Barbara
sdey@ucsb.edu

Dr. Shivnarayan Dhuppar
Tata Institute of Fundamental Research
dhupparst@tifrh.res.in

Mr. Pierluigi Di Chiaro
Humanitas University
pierluigi.dichiaro@humanitasresearch.it

Dr. Alexander Dobin
Cold Spring Harbor Laboratory
dobin@cshl.edu

Dr. Pablo Domizi
Stanford University
domizi@stanford.edu

Ms. Eliza Duvall
National Institutes of Health
duvaller@nih.gov

Dr. James Eberwine
University of Pennsylvania
eberwine@upenn.edu

Dr. Hye Hyeon Eum
Samsung Medical Center
hheum1107@gmail.com

Dr. Di Feng
Boehringer Ingelheim Pharmaceutical, Inc
di.feng@boehringer-ingelheim.com

Dr. Stephan Fischer
Cold Spring Harbor Lab
fischer@cshl.edu

Dr. Matthew Ford
McGill University
matthew.ford@mail.mcgill.ca

Dr. Scott Fraser
University of Southern California
sfraser@provost.usc.edu

Dr. Akihiro Furuya
Okinawa Institute of Science and Technology
akihiro.furuya@oist.jp

Dr. Shouguo Gao
NHLBI/NIH
gaos2@nih.gov

Dr. Hernan Garcia
University of California Berkeley
hggarcia@berkeley.edu

Dr. Zhongqi Ge
MD Anderson Cancer Center
zge@mdanderson.org

Mr. Nikhil Oommen George
German Cancer Research Center (DKFZ)
n.george@dkfz-heidelberg.de

Dr. Joshy George
The Jackson Laboratory
joshy.george@jax.org

Mr. Tamas Gerecsei
Nanobiosensorics Laboratory
tamasger@yahoo.com

Dr. Francesca Giannese
Ospedale San Raffaele
giannese.francesca@hsr.it

Dr. Ana Gonzalez-Reiche
Icahn School of Medicine at Mount Sinai
anasilvia.gonzalez-reiche@mssm.edu

Mr. Chris Gordon
Stony Brook University
chris.gordon@stonybrook.edu

Dr. Ian Green
Nature Communications
ian.green@nature.com

Mr. Tobias Gross
University Freiburg
tobias.gross@imtek.uni-freiburg.de

Ms. Sarah Groves
Vanderbilt University
sarahmaddoxgroves@gmail.com

Dr. Nehaben Gujarati
Stony Brook School of Medicine
nehaben.gujarati@stonybrookmedicine.edu

Dr. Jennifer Guler
University of Virginia
jlg5fw@virginia.edu

Dr. Yiqing Guo
Stony Brook University
yiqing.guo@stonybrookmedicine.edu

Ms. Anushka Gupta
UC Berkeley-UCSF Bioengineering Graduate Program
anushka_gupta@berkeley.edu

Dr. Jungmin Han
National Institutes of Health
jungmin.han@nih.gov

Dr. Kyung Yeon Han
Samsung Medical Center
ky2.han@samsung.com

Dr. Guangchun Han
University of Texas MD Anderson Cancer Center
ghan1@mdanderson.org

Dr. Yuan Hao
Cold Spring Harbor Laboratory
yhao@cshl.edu

Mr. Benjamin Harris
Cold Spring Harbor Laboratory
bharris@cshl.edu

Prof. Shinichi Hashimoto
Kanazawa University
hashimoto@med.kanazawa-u.ac.jp

Dr. Sikander Hayat
Bayer US LLC
sikander.hayat@bayer.com

Dr. Jian He
Shanghai Jiao Tong University
hejianwan@hotmail.com

Mr. Gervaise Henry
UT Southwestern Medical Center - Dallas
gervaise.henry@utsouthwestern.edu

Dr. Laura Hertel
Children's Hospital Oakland Research Institute
lhertel@chori.org

Dr. Ronna Hertzano
University of Maryland School of Medicine
rhertzano@som.umaryland.edu

Mr. Samuel Higgins
Roche
samuel.higgins@roche.com

Mr. Matthew Hill
Baylor College of Medicine
matthew.hill@bcm.edu

Ms. Yourae Hong
Samsung Medical Center
yrhong1023@gmail.com

Dr. Anelia Horvath
The George Washington University
horvatha@email.gwu.edu

Dr. Alex Hu
Benaroya Research Institute
ahu@benaroyaresearch.org

Dr. Peng Hu
University of Pennsylvania
penghu@upenn.edu

Dr. Qing Huan
Chinese Academy of Sciences
qhuan@genetics.ac.cn

Mr. Salvatore Inganne
Stony Brook University
salvatore.inganne@stonybrook.edu

Dr. Sadahiro Iwabuchi
Kanazawa University
s_iwabuchi@staff.kanazawa-u.ac.jp

Dr. Csaba Jeney
Albert-Ludwigs-Universität Freiburg
csjeney@gmail.com

Dr. Gerrit John-Schuster
Springer Nature - BMC - Genome Medicine
gerrit.john-schuster@genomemedicine.com

Dr. Radia Johnson
Genentech
johnson.radia@gene.com

Mr. Taylor Jones
Vanderbilt University Medical Centr
tjonesster@gmail.com

Dr. Vanessa Jonsson
City of Hope National Medical Center
vjonsson@coh.org

Ms. Farhia Kabeer
University of British Columbia
fkabeer@bccrc.ca

Dr. Yasin Kaymaz
Harvard University
yasinkaymaz@fas.harvard.edu

Dr. Shana Kelley
University of Toronto
shana.kelley@utoronto.ca

Dr. Leeat Keren
Stanford University
lkeren@stanford.edu

Dr. Junhyong Kim
University of Pennsylvania
junhyong@sas.upenn.edu

Ms. Young Kim
Stony Brook University
yschonkim5@gmail.com

Mr. Kyung mo Kim
Stony Brook University
yschon@hanmail.net

Mr. Bill Kleinberg
Bio-Techne
bill.kleinberg@bio-techne.com

Dr. Kyung Dae Ko
NIH
kok3@nih.gov

Dr. Robert Kokoska
Army Research Office
robert.j.kokoska2.civ@mail.mil

Mr. Lukas Kremer
DKFZ German Cancer Research Center
L.Kremer@dkfz.de

Dr. Parveen Kumar
The Jackson Laboratory of Genomic Medicine
parveen.kumar@jax.org

Dr. Gabriel Lam
NIH
gabriel.lam@nih.gov

Dr. Dejan Lazarevic
Ospedale San Raffaele
lazarevic.dejan@hsr.it

Dr. Je Lee
Cold Spring Harbor Laboratory
jlee@cshl.edu

Dr. Wesley Legant
University of North Carolina at Chapel Hill
legantw@email.unc.edu

Mr. Wesley Lewis
Cold Spring Harbor Laboratory
lewiswesley66@gmail.com

Dr. David Li
Sun Yat-sen University
dwli1688@hotmail.com

Dr. Jian-Liang Li
NIEHS
jianliang.li@nih.gov

Dr. Siran Li
Cold Spring Harbor Laboratory
siranli@cshl.edu

Dr. Wei Keat Lim
Regeneron
weikeat.lim@regeneron.com

Ms. Ying Lin
University of California Los Angeles
yinglin1202@gmail.com

Ms. Shiwei Liu
University of Virginia
sl5xa@virginia.edu

Ms. Hongyu Liu
The George Washington University
hliu5259@gwu.edu

Dr. Yang Liu
Albert Einstein College of Medicine
yang.liu@einstein.yu.edu

Dr. Zhen Liu
Nanjing University
zhenliu@nju.edu.cn

Dr. Jia Liu
Nanjing University
liujia2019@nju.edu.cn

Ms. Fu Luo
Stony Brook University
fu.luo@stonybrook.edu

Dr. Xizeng Mao
MD Anderson Cancer Center
xmao1@mdanderson.org

Dr. Luciano Martelotto
Victorian Comprehensive Cancer Center
luciano.martelotto@unimelb.edu.au

Dr. Grinu Mathew
Cold Spring Harbor Laboratory
gmathew@cshl.edu

Dr. Jose McFaline
University of Washington
joselmc@uw.edu

Dr. David McKinnon
Stony Brook University
david.mckinnon@stonybrook.edu

Mr. Ian Mellis
University of Pennsylvania
iamellis@pennmedicine.upenn.edu

Dr. Melissa Mendez
GlaxoSmithKline
melissa.g.mendez@gsk.com

Dr. Carmen Menoni
Colorado State University
Carmen.Menoni@colostate.edu

Dr. Collin Merrill
University of Utah
collin.merrill@utah.edu

Dr. Bud Mishra
CSHL
mishra@nyu.edu

Dr. Apratim Mitra
NIH
apratim.mitra@nih.gov

Dr. Daniel Mitroi
UCLA
dmitroi@mednet.ucla.edu

Dr. Shahin Mohammadi
Massachusetts Institute of Technology
mohammadi@broadinstitute.org

Dr. Carme Mont
TakaraBio
carme_mont@takarabio.com

Mr. Austin Montgomery
University of Utah
u0873970@utah.edu

Dr. Ritsuko Morita
RIKEN Center for Biosystems Dynamics Research
ritsuko.morita@riken.jp

Dr. Jacob Musser
European Molecular Biology Laboratory
jacob.musser@embl.de

Prof. Marcin Mycko
University of Warmia and Mazury
marcin.mycko@uwm.edu.pl

Dr. Youngji Na
Columbia University
yn2318@cumc.columbia.edu

Ms. Sarah Nahle
Montreal Clinical Research Institute (IRCM)
sarah.nahle@ircm.qc.ca

Dr. Scott Neal
Upstate Medical University
neals@upstate.edu

Dr. Peter Nemes
University of Maryland, College Park
nemes@umd.edu

Ms. Rachel Ng
City of Hope National Medical Center
racng@coh.org

Dr. Tom Nguyen
Amgen Inc.
tnguye07@amgen.com

Dr. Xiaochun Ni
Celgene
xni@celgene.com

Mr. Prashant NM
The George Washington University
pnm27@gwu.edu

Dr. Seitaro Nomura
RCAST, The University of Tokyo
senomura-cib@umin.ac.jp

Dr. Hatice Osmanbeyoglu
University of Pittsburgh
osmanbeyogluhu@pitt.edu

Dr. Aaron Pantoja Angles
King Abdullah University of Science and Technology
aaron.pantojaangles@kaust.edu.sa

Ms. Sarah Park
CSHL
sapark@cshl.edu

Dr. Ralph Patrick
Victor Chang Cardiac Research Institute
r.patrick@victorchang.edu.au

Dr. Liming Pei
Children's Hospital of Philadelphia/UPenn
peil@email.chop.edu

Dr. Tao Peng
Children's Hospital of Philadelphia
pengt@email.chop.edu

Dr. Vipul Periwal
NIH/NIDDK
vipulp@mail.nih.gov

Dr. Stephanie Philtjens
IU School of Medicine/Stark Neuro Research Inst.
sphiltje@iu.edu

Dr. Sebastian Pott
University of Chicago
spott@uchicago.edu

Dr. Steve Potter
Cincinnati Children's Medical Center
steve.potter@cchmc.org

Dr. Jonathan Preall
Cold Spring Harbor Laboratory
jpreall@cshl.edu

Dr. Christina Pressl
The Rockefeller University
cpressl@rockefeller.edu

Dr. Marta Prieto-Vila
Tokyo Medical University
mprietov@tokyo-med.ac.jp

Dr. Joanna Przybyl
Stanford University
jprzybyl@stanford.edu

Dr. Xiaojie Qiu
UCSF
xiaojie.qiu@ucsf.edu

Mr. Wei-Lin Qiu
Peking University
qiuweilin@pku.edu.cn

Dr. John Quinn
FlowJo
john.flowjo@bd.com

Mr. Bobby Ranjan
Genome Institute of Singapore
Ranjan_Bobby@gis.a-star.edu.sg

Mr. Daniel Reichart
Harvard Medical School
Daniel_Reichart@hms.harvard.edu

Mr. Markus Roeder
TU Darmstadt
markus.roeder@bcs.tu-darmstadt.de

Mr. Kevin Rose
University of Maryland School Medicine
rosekp@umaryland.edu

Mr. Robert Ruzic
Cold Spring Harbor
robert.ruzic@stonybrook.edu

Mr. Sombeet Sahu
Mission Bio
sahu@missionbio.com

Dr. Nathan Salomonis
Cincinnati Children's Hospital Medical Center
nathan.salomonis@cchmc.org

Mr. Harry Samaroo
Biogen Inc.
Harry.Samaroo@biogen.com

Dr. Tatjana Sauka-Spengler
University of Oxford
tatjana.sauka-spengler@imm.ox.ac.uk

Dr. Gregory Schwartz
University of Pennsylvania
gsch@pennmedicine.upenn.edu

Mr. Tim Semple
Peter MacCallum Cancer Centre
timothy.semple@petermac.org

Dr. Sunantha Sethuraman
AbbVie
sunantha.sethuraman@abbvie.com

Mr. Siddhartha Sharma
The Scripps Research Institute
smsharma@scripps.edu

Dr. April Shu
Yale University
aprilchangshu@gmail.com

Dr. Remigijus Skirgaila
Thermo Fisher Scientific
remigijus.skirgaila@thermofisher.com

Prof. Simon Sprecher
University of Fribourg
simon.sprecher@gmail.com

Mr. Avi Srivastava
Stony Brook University
asrivastava@cs.stonybrook.edu

Dr. Aaron Streets
University of California, Berkeley
astreets@berkeley.edu

Ms. Yu-Chi Sun
Cold Spring Harbor Laboratory
ysun@cshl.edu

Dr. Katalin Susztak
University of Pennsylvania
ksusztak@pennmedicine.upenn.edu

Dr. Fumihiko Takeuchi
National Center for Global Health and Medicine
fumihiko@takeuchi.name

Dr. Ming Tang
Harvard University
tangming2005@gmail.com

Mr. Tomas Uzusienis
Thermo Fisher Scientific
tomas.uzusienis@thermofisher.com

Prof. Erik van Nimwegen
University of Basel
erik.vannimwegen@unibas.ch

Dr. Parul Varma
The University of Texas at San Antonio
parul.varma@utsa.edu

Prof. Andreas Vasdekis
University of Idaho
andreas.vasdekis@gmail.com

Mr. Jai Vekeria
Compass Therapeutics
jai.vekeria@compasstherapeutics.com

Dr. Priyanka Vijay
Abbvie
priyanka.vijay@abbvie.com

Prof. Julia Vorholt
ETH Zurich
jvorholt@ethz.ch

Dr. Pavle Vrljicak
Warwick University
P.J.Vrljicak@warwick.ac.uk

Dr. Tomas Vystavel
Thermo Fisher Scientific
tomas.vystavel@thermofisher.com

Dr. Li Wang
Icahn School of Medicine at Mount Sinai & Sema4
li.wang@mssm.edu

Dr. Siyuan Wang
Yale University
siyuan.wang@yale.edu

Dr. Ruiping Wang
The University of Texas MD Anderson Cancer Center
RWang12@mdanderson.org

Dr. Xujing Wang
National Institutes of Health
xujing.wang@nih.gov

Dr. Jie Wang
Johns Hopkins University
jwang240@jhmi.edu

Ms. Rujin Wang
University of North Carolina at Chapel Hill
rujin@email.unc.edu

Dr. Linghua Wang
UT MD Anderson Cancer Center
lwang22@mdanderson.org

Mr. Chatarin Wangsanuwat
University of California, Santa Barbara
cwangsanuwat@ucsb.edu

Prof. Hidemi Watanabe
Hokkaido University
watanabe@ist.hokudai.ac.jp

Ms. Yanrong Wen
Nanjing University
wenyanrong@foxmail.com

Mr. Elijah Willie
University of British Columbia
ewillie@cmmt.ubc.ca

Ms. Chi-Yun Wu
University of Pennsylvania
chiyunwu@pennmedicine.upenn.edu

Prof. Xiaoliang Xie
Peking University
sunneyxie@pku.edu.cn

Dr. Xiaojiang Xu
NIEHS, NIH
xuxiao@niehs.nih.gov

Dr. Yichi Xu
Memorial Sloan Ketter Cancer Center
xuy2@mskcc.org

Dr. Cheng-Ran Xu
Peking University
cxu@pku.edu.cn

Dr. Tracy Yamawaki
Amgen
tyamawak@amgen.com

Dr. Tao Yang
Regeneron Pharmaceuticals
tao.yang@regeneron.com

Dr. Na Yang
National Institute On Aging
na.yang2@nih.gov

Dr. Xiaofei Ye
Karolinska Institutet
xiaofei.ye@ki.se

Dr. Johannes Yeh
Cold Spring Harbor Laboratory
jyeh@cshl.edu

Ms. Jin-Hee Yeo
Sungkyunkwan University
ver.9402@gmail.com

Ms. Zhezhen Yu
Cold Spring Harbor Laboratory
zhezhen@cshl.edu

Ms. Tengjiao Zhang
Sloan Kettering Cancer Center
1153020@tongji.edu.cn

Dr. Chi Zhang
Indiana University School of Medicine
czhang87@iu.edu

Ms. Chunling Zhang
Upstate Medical University
zhangch@upstate.edu

Dr. Xiuwei Zhang
University of California, Berkeley
zhangxiuwei03@gmail.com

Dr. Hao Zhao
Ziopharm Oncology
hzhao@ziopharm.com

Dr. Yixin Zhao
Cold Spring Harbor Laboratory
yizhao@cshl.edu

Dr. Juanjuan Zhao
Children's Hospital of Philadelphia
zhaoj3@EMAIL.CHOP.EDU

Dr. Wei Zheng
Moderna
wei.zheng@modernatx.com

Mr. Qin Zhu
University of Pennsylvania
zhuqin@pennmedicine.upenn.edu

Dr. Qian Zhu
Dana Farber Cancer Institute
zqian@jimmy.harvard.edu

VISITOR INFORMATION

EMERGENCY	CSHL	BANBURY
Fire	**(3) 742-3300**	**(3) 692-4747**
Ambulance	**(3) 742-3300**	**(3) 692-4747**
Poison	**(3) 542-2323**	**(3) 542-2323**
Police	**(3) 911**	**(3) 549-8800**
Safety-Security	**Extension 8870**	

Emergency Room **Huntington Hospital** 270 Park Avenue, Huntington	**631-351-2000**
Dentists Dr. William Berg Dr. Robert Zeman	 **631-271-2310** **631-271-8090**
Doctor MediCenter 234 W. Jericho Tpke., Huntington Station	**631-423-5400**
Drugs - 24 hours, 7 days Rite-Aid 391 W. Main Street, Huntington	**631-549-9400**

GENERAL INFORMATION

Books, Gifts, Snacks, Clothing, Newspapers

BOOKSTORE 367-8837 (hours posted on door)
Located in Grace Auditorium, lower level.

Photocopiers, Journals, Periodicals, Books, Newspapers

Photocopying – Main Library
Hours: 8:00 a.m. – 9:00 p.m. Mon-Fri
10:00 a.m. – 6:00 p.m. Saturday
***Helpful tips* – Use PIN# 57290** to enter Library after hours.
See Library staff for photocopier code.

Computers, E-mail, Internet access

Grace Auditorium
Upper level: E-mail and printing in the business center area
STMP server address: mail.optonline.net
To access your E-mail, you must know the name of your home server.

Dining, Bar

Blackford Hall
Breakfast 7:30–9:00, Lunch 11:30–1:30, Dinner 5:30–7:00
Bar 5:00 p.m. until late (Cash Only)
Helpful tip - If there is a line at the upper dining area, try the lower dining room

Messages, Mail, Faxes, ATM

Message Board, Grace, lower level

Swimming, Tennis, Jogging, Hiking

June–Sept. Lifeguard on duty at the beach. 12:00 noon–6:00 p.m. Two tennis courts open daily.

Russell Fitness Center

Dolan Hall, east wing, lower level
PIN#: **Press 57290 (then enter #)**

Meetings & Courses Front Office

Hours during meetings: 8am – 7pm, until 9pm on arrival day
After hours – From guest house phones, dial x8870 for assistance

Pay Phones, House Phones

Grace, lower level; Cabin Complex; Blackford Hall; Dolan Hall, foyer

CSHL's Green Campus

Cold Spring Harbor Laboratory is pledged to operate in an environmentally responsible fashion wherever possible. In the past, we have removed underground oil tanks, remediated asbestos in historic buildings, and taken substantial measures to ensure the pristine quality of the waters of the harbor. Water used for irrigation comes from natural springs and wells on the property itself. Lawns, trees, and planting beds are managed organically whenever possible. And trees are planted to replace those felled for construction projects.

Two areas in which the Laboratory has focused recent efforts have been those of waste management and energy conservation. The Laboratory currently recycles most waste. Scrap metal, electronics, construction debris, batteries, fluorescent light bulbs, toner cartridges, and waste oil are all recycled. For general waste, the Laboratory uses a "single stream waste management" system, removing recyclable materials and sending the remaining combustible trash to a cogeneration plant where it is burned to provide electricity, an approach considered among the most energy efficient, while providing a high yield of recyclable materials.

Equal attention has been paid to energy conservation. Most lighting fixtures have been replaced with high efficiency fluorescent fixtures, and thousands of incandescent bulbs throughout campus have been replaced with compact fluorescents. The Laboratory has also embarked on a project that will replace all building management systems on campus, reducing heating and cooling costs by as much as twenty-five per cent.

Cold Spring Harbor Laboratory continues to explore new ways in which we can reduce our environmental footprint, including encouraging our visitors and employees to use reusable containers, conserve energy, and suggest areas in which the Laboratory's efforts can be improved. This book, for example, is printed on recycled paper.

1-800 Access Numbers

AT&T **9-1-800-321-0288**

Local Interest

Fish Hatchery	631-692-6758
Sagamore Hill	516-922-4788
Whaling Museum	631-367-3418
Heckscher Museum	631-351-3250
CSHL DNA Learning Center	x 5170

New York City

Helpful tip -

Take Syosset Taxi to Syosset Train Station
($9.00 per person, 15 minute ride), then catch Long Island Railroad to Penn Station (33rd Street & 7th Avenue).
Train ride about one hour.

TRANSPORTATION

Limo, Taxi

Syosset Limousine	516-364-9681
Executive Limo Service	631-696-8000
Super Shuttle	800-957-4533
Limos Long Island	833-545-4667,ext.3

To head west of CSHL - Syosset train station

Syosset Taxi	516-921-2141

To head east of CSHL - Huntington Village

Orange & White Taxi	631-271-3600

Trains

Long Island Rail Road	822-LIRR

Schedules available from the Meetings & Courses Front Office.

Amtrak	800-872-7245
MetroNorth	877-690-5114
New Jersey Transit	973-275-5555

Ferries

Bridgeport / Port Jefferson	631-473-0286
Orient Point/ New London	631-323-2525

Car Rentals

Avis	631-271-9300
Enterprise	631-424-8300
Hertz	631-427-6106

Airlines

American	800-433-7300
British Airways	800-247-9297
Delta	800-221-1212
Japan Airlines	800-525-3663
Jet Blue	800-538-2583
KLM	800-374-7747
Lufthansa	800-645-3880
Southwest Airlines	800-435-9792
United	800-241-6522
Virgin American	877-359-9792

Code of Conduct

Cold Spring Harbor Laboratory is dedicated to pursuing its twin missions of research and education in the biological sciences. The Laboratory is committed to fostering a working environment that encourages and supports unfettered scientific inquiry and the free and open exchange of ideas that are the hallmarks of academic freedom. The Laboratory aims to maintain a safe and respectful environment for all attendees of our meetings and courses and associated support staff, by providing an environment free from discrimination and harassment, in accordance with federal, state, and local laws.

Consistent with the Laboratory's missions, commitments and policies, the purpose of this Code is to set forth the Laboratory's expectations for the professional conduct of individuals participating in the Laboratory's meetings program, including organizers, session chairs, invited speakers, presenters, attendees and sponsors.

By attending a CSHL meeting, participants agree to:

1. Treat fellow meeting participants and CSHL staff with respect, civility and fairness, and without bias based on race, color, religion, sex, national origin, citizenship status, sexual orientation, gender identity or expression, age, disability, marital status, veteran status, genetic information, or any other criteria prohibited under applicable federal, state or local law.

2. Use all CSHL facilities, equipment, computers, supplies and resources responsibly and appropriately, as you would at your home institution.

Meeting participants agree to refrain from:

1. Discrimination in violation of Laboratory policy based on age, gender, race, ethnicity, national origin, religion, disability, or sexual orientation.

2. Behavior that is disrespectful of others, and unprofessional interpersonal behavior that interferes with the working and learning environment.

3. Unwanted physical contact with others, or threats of such contact.

4. Sexual harassment or harassment based on age, gender, race, ethnicity, national origin, religion, disability or sexual orientation.

5. Loss of civility that interferes with the working and learning environment (for example shouting, personal attacks or insults, throwing objects or other displays of temper).

6. Misappropriation of Laboratory property or excessive personal use of resources.

Breaches or Violations of the Code of Conduct

Cold Spring Harbor Laboratory aims to maintain a conference environment in accordance with the principles and expectations outlined in the Code of Conduct. Meeting organizers are tasked with providing leadership during each meeting, and may be approached informally about any breach or violation. Breaches or violations should also be reported to program leadership in person or by email:

- Dr. David Stewart, Grace Auditorium Room 204, 516-367-8801 or x8801 from a campus phone, stewart@cshl.edu

- Dr. Charla Lambert, Hershey Laboratory Room 214, 516-367-5058 or x5058 from a campus phone, clambert@cshl.edu

The Laboratory will take action as needed to resolve the matter, up to and including immediate expulsion of the offending participant(s) from the meeting, dismissal from the Laboratory, and exclusion from future academic events offered by CSHL.