



Session 1

Session chair: John Marioni

10:00–10:10 am ET Introduction & Welcome by Vijay Kuchroo, Shing-Tung Yau, and Martin Hemberg

10:10-10:40 am ET Peter Kharchenko, Harvard University

Title: “Bayesian segmentation of spatially resolved transcriptomics data”

Abstract: Spatial transcriptomics is an emerging stack of technologies, which adds spatial dimension to conventional single-cell RNA-sequencing. New protocols, based on in situ sequencing or multiplexed RNA fluorescent in situ hybridization register positions of single molecules in fixed tissue slices. Analysis of such data at the level of individual cells, however, requires accurate identification of cell boundaries. While many existing methods are able to approximate cell center positions using nuclei stains, current protocols do not report robust signal on the cell membranes, making accurate cell segmentation a key barrier for downstream analysis and interpretation of the data. To address this challenge, we developed a tool for Bayesian Segmentation of Spatial Transcriptomics Data (Baysor), which optimizes segmentation considering the likelihood of transcriptional composition, size and shape of the cell. The Bayesian approach can take into account nuclear or cytoplasm staining, however can also perform segmentation based on the detected transcripts alone. We show that Baysor segmentation can in some cases nearly double the number of the identified cells, while reducing contamination. Importantly, we demonstrate that Baysor performs well on data acquired using five different spatially-resolved protocols, making it a useful general tool for analysis of high-resolution spatial data.

10:40–10:45am ET Break

10:45-11:15 am ET Smita Krishnaswamy, Yale University

Title: “Geometric and Topological Approaches to Representation Learning in Biomedical Data”

Abstract: High-throughput, high-dimensional data has become ubiquitous in the biomedical, health and social sciences as a result of breakthroughs in measurement technologies and data collection. While these large datasets containing millions of observations of cells, peoples, or brain voxels hold great potential for understanding generative state space of the data, as well as drivers of differentiation, disease and progression, they also pose new challenges in terms of noise, missing data, measurement artifacts, and the so-called “curse of dimensionality.” In this talk, I will cover data geometric and topological approaches to understanding the shape and structure of the data. First, we show how diffusion geometry and deep learning can be used to obtain useful representations of the data that enable denoising (MAGIC), dimensionality reduction (PHATE), and factor analysis (Archetypal Analysis Network) of the data. Next we will show how to learn dynamics from static snapshot data by using a manifold-regularized neural ODE-based optimal transport (TrajectoryNet). Finally, we cover a novel approach to combine diffusion geometry with topology to extract multi-granular features from the data (Diffusion Condensation and Multiscale PHATE) to assist in differential and predictive analysis. On the flip side, we also create a manifold geometry from topological descriptors, and show its applications to neuroscience. Together, we will show a complete framework for exploratory and unsupervised analysis of big biomedical data.

11:15–11:20 am ET Break

11:20-11:50 am ET Meromit Singer, Harvard Medical School

Title: “Utilizing coupled single-cell RNA-seq and TCR-seq to reveal Th17 systemic dynamics during homeostasis and disease”

Abstract: In this talk we will describe a systemic study of the transcriptional and clonal characteristics and function of Th17 cells throughout multiple mouse organs, as revealed by coupled single-cell RNA-seq and TCR-seq, and validated in follow-up experiments in the lab. We will describe how we utilized 84,000 tissue Th17 cells profiled during homeostasis and disease to characterize their heterogeneity, plasticity, and migration at homeostasis and during CNS autoimmunity. We discovered a homeostatic Th17 cell population, that is induced by the intestinal microbiota, is present in both lymphoid organs and the intestine, and expresses IL-17. We discovered that during EAE this homeostatic population gives rise to a pathogenic Th17 cell population, that migrates specifically through the draining lymph nodes and the spleen to the CNS, and highly expresses a specific subset of cytokines.

In this talk we will emphasize how coupled single-cell RNA-seq and TCR data was used to generate hypotheses regarding cell subtype characterization and T cell clonality and migration, and how such hypotheses were followed-up on experimentally.

11:50–11:55 am ET Break

11:55-12:25 pm ET John Marioni, EMBL-EBI

Title: “Analysis of multi-modal single-cell data”

12:25–1:25 pm ET Lunch Break

Session 2

Session chair: Martin Hemberg

1:25-1:55pm ET Uri Alon, Weizmann Institute

Title: “Design principles of hormone circuits”

1:55–2:00 pm ET Break

2:00-2:30pm ET Eran Segal, Weizmann Institute

Title: “Harnessing big data for personalized medicine”

Abstract: The recent availability of diverse health data resources on large cohorts of human individuals presents many challenges and opportunities. I will present our work aimed at developing machine learning algorithms for predicting future onset of disease and identifying causal drivers of disease based on nationwide electronic health record data as well as data from high-throughput omics profiling technologies such as genetics, microbiome, and metabolomics. Our models provide novel insights into potential drivers of obesity, diabetes, and heart disease, and identify hundreds of novel markers at the microbiome, metabolite, and immune system level. Overall, our predictive models can be translated into personalized disease prevention and treatment plans, and to the development of new therapeutic modalities based on metabolites and the microbiome.

2:30–2:35 pm ET Break

2:35-3:05 pm ET Martin Hemberg, Brigham and Women’s Hospital

Title: “Searching for alien DNA – characterization of sequences that are not present in the DNA”

Abstract: Nullomers and nullpeptides are short DNA or amino acid sequences that are absent from a genome or proteome, respectively. One potential cause for their absence could be that they have a detrimental impact on an organism. Here, we identified all possible nullomers and nullpeptides in the genomes and proteomes of over thirty species and show that a significant proportion of these sequences are under negative selection. We assign nullomers to different functional categories (coding sequences, exons, introns, 5’UTR, 3’UTR, regulatory regions and promoters) and show that nullomers from coding sequences and promoters are most likely to be selected against. Similarly, we find that regulatory regions and transcription factor binding sites harbor more mutations resulting in nullomers than expected. Further analysis of coding regions also reveals specific pathways where mutations are more likely to result in nullomers or nullpeptides. Utilizing variants in the human population, we annotate variant-associated nullomers, highlighting their potential use as DNA ‘fingerprints’.

3:05-3:10 pm ET Break



3:10-3:40 pm ET Elana Fertig, Johns Hopkins

Title: “Uncovering hidden sources of transcriptional dysregulation arising from inter- and intra-tumor heterogeneity”

Abstract: Heterogeneity poses a major challenge in translational research. For example, inter-tumor heterogeneity limits the biomarker discovery and intra-tumor heterogeneity enables therapeutic resistance. Moreover, in some cancers driver mutations are insufficient to account for the widespread transcriptional variation responsible for these outcomes. Thus, new computational tools to model transcriptional variation are essential. To address this we develop an innovative computational framework, Expression Variation Analysis (EVA), to model transcriptional dysregulation in cancer. Briefly, EVA quantifies transcriptional heterogeneity for one set of samples or cells from one phenotype using the expected dissimilarity between pairs of expression profiles. U-statistics theory can then quantify the statistical significance of the difference in transcriptional heterogeneity between phenotypes. We apply EVA to perform a comprehensive characterization of transcriptional variation in head and neck squamous cell carcinoma (HNSCC). At a pathway level, transcriptional variation in HNSCC tumors is higher than normal controls. Applying EVA to integrate ChIP-seq data with RNA-seq reveals that these pervasive transcriptional differences occur in enhancers. Similarly, applying EVA at a gene level to model splicing reveals more heterogeneity in transcript usage in tumor samples than normals. HPV- HNSCC tumors are unique in having mutations in genes that regulate the splicing machinery, and the HPV- tumors with these alterations have a greater number of dysregulated splice variants than those without. Nonetheless, the EVA analysis identifies a similar number of alternative splice variants in HPV+ as HPV- tumors suggesting an alternative mechanism of transcriptional heterogeneity in HPV+ disease. Adapting EVA to single cell data demonstrates that increased fibroblast composition is associated with greater variation in immune pathway activity in HNSCC. Moreover, we observe greater transcriptional heterogeneity in HNSCC primary tumors than lymph node metastasis consistent with a clonal outgrowth. We demonstrate that the statistical framework from EVA enables differential heterogeneity analysis in HNSCC ranging from pathway dysregulation, splice variation, epigenetic regulation, and single cell analysis. This algorithm provides a critical framework to model the hidden multi-molecular mechanisms underlying the complex patient outcomes that are pervasive in cancer.

3:40–3:50 pm ET Closing Remarks