Genes 2023

Single-Cell Genomics Moving Forward

29–31 March 2023 | Barcelona, Spain

Program and Abstract Book

Organizers

MDPI Academic Open Access Publishing since 1996

genes

Sponsors

Parse BIOSCIENCES

Atrandi BIOSCIENCES

MGI

Media Partners

DNA

SEBBM

Genes 2023
Single-Cell Genomics Moving Forward

Barcelona, Spain
29 – 31 March 2023

Conference Chairs
Dr. Anelia Horvath
Prof. Peter Young

Organised by

Conference Secretariat
Paula Navarro
Íñigo Aristizabal
Ana Sanchís
Pablo Velázquez

CONTENTS

Conference Programme

Wednesday 29 March 2023

13:00 – 13:30	Registration Desk Open (Check-in)
13:30 – 13:45	Welcome from the Chairs

Session 1. Part I
Single Cell Multi-Omics Technologies and Analyses

13:45 – 14:25	**Keynote Talk: Manel Esteller** - Using Single Cell Multiomics in Cancer: From Knowledge to Applications
14:25 – 14:40	**Silvia Ogbeide** - Single-cell multi-omics profiling in the study of colorectal cancer evolution
14:40 – 14:55	**Francesca Nadalin** - Single-cell multi-omic lineage tracing uncouples tumour initiation and drug tolerance in breast cancer
14:55 – 15:10	**Enrico Sebastiani** - An Investigation of the origin of neuroblastoma with single-cell transcriptomic analyses
15:10 – 15:25	**Igor Filippov** - Single-cell re-analysis of human transcriptomes reveals shared characteristics of immune ageing
15:25 – 15:50	**Invited Talk: Anelia Horvath** - Identification and analysis of cell-specific expressed genetic variants from scRNA-seq data
15:50 – 16:20	**Coffee Break**

Session 2. Part I
Artificial Intelligence in Single Cell Genomics

16:20 – 16:50	**Invited Talk: Nacho Molina** - Biophysics-informed interpretable deep learning to model gene regulation from single-cell and single-molecule genomic data
16:50 – 17:05	**Kyoung Jae Won** - Legible Gene Regulatory Network Reconstruction using Multivariate Transfer Entropy Over Single Cell Transcriptomics Data
17:05 – 17:20	**Bogac Aybey** - Immune cell type signature discovery and random forest classification for analysis of single cell gene expression datasets
17:20 – 18:00	**Keynote Talk: Sandrine Dudoit** - Learning from Data in Single-Cell Transcriptomics
20:00	**Conference Dinner**

Thursday 30 March 2023

Session 1. Part II
Single Cell Multi-Omics Technologies and Analyses

08:30 – 09:00	**Invited Talk: Merja Heinäniemi** - Multimodal methods in single cell immunogenomics
09:00 – 09:15	**Kalle Rytkönen** - Gene regulatory network analysis of decidual stromal cells and natural killer cells
09:15 – 09:30	**Alexandre Gaspar-Maia** - Single cell multiomic analysis of Clonal Hematopoiesis reveals enhancer deregulation and increased COVID-19 inflammation severity
09:30 – 09:45	**Sara Terzoli** - Expansion of memory Vδ2 T cells following repeated exposure to mRNA SARS-CoV-2 vaccination
09:45 – 10:00	**Alejandro Jiménez Sánchez** - Optimized tumor metastatic phenotypes among distal metastatic lesions in pancreatic ductal adenocarcinoma
10:00 – 10:30	**Invited Talk: Merav Cohen** - The Functional Role of Immune-Related Intercellular Signaling Networks during Tissue Development and Cancer
10:30 – 11:00	**Coffee Break**

Session 1. Part III
Single Cell Multi-Omics Technologies and Analyses

11:00 – 11:15	**Enrico Gaffo** - Benchmarking of methods for assessing circular RNA differential expression in single-cell RNA-seq data
11:15 – 11:30	**Amy Hamilton** - Unlocking the potential for single-cell sequencing at scale with combinatorial indexing
11:30 – 11:45	**Celia Alda Catalinas** - Mapping the functional impact of immune disease associated regulatory elements through single-cell CRISPR-based screens
11:45 – 12:00	**Rapolas Zilionis** - Robust and versatile ultra-high-throughput single-microbe genome sequencing

12:00 – 12:15	**Veredas Coleto Alcudia** - TOTEM: A Web Tool For Tissue Enrichment On Gene Lists With Single Cell Resolution
12:15 – 12:30	**Emma Busarello** - Harmonizing the annotation of single cells in normal and aberrant hematopoiesis with the Cell Marker Accordion
12:30 – 13:45	**Lunch**

Session 3
Spatial Genomics and Transcriptomics

13:45 – 14:00	**Kyoung Jae Won** - Image processing approach to spatial genomics data identifies tissue architecture and cell-contact specific gene regulation
14:00 – 14:15	**Biao Lu** - Alteration of transcriptions loaded in engineered extracellular vesicles using transmembrane scaffolds
14:15 – 14:30	**Susmita Datta** - Inferring cell-cell communications from spatially resolved transcriptomics data
14:30 – 14:45	**Alba Garrido Trigo** - Macrophage and neutrophil heterogeneity at single-cell spatial resolution in inflammatory bowel disease
14:45 – 15:00	**Kang Jin** - Gene colocalization-aware segmentation of image-based spatial transcriptomics using Graph Neural Networks
15:00 – 15:30	**Coffee Break**

Session 1. Part IV
Single Cell Multi-Omics Technologies and Analyses

15:30 – 16:00	**Invited Talk: Björn Reinius** - Dosage compensation dynamics unmasked by allele-specific single-cell transcriptomics
16:00 – 16:15	**Evgenia Ulianova** - Cell-Level Analysis of Telomeric Transcripts and Telomere-Associated Gene Variants
16:15 – 16:30	**Meltem Omur** - Characterization of gene regulatory networks and combinatorial transcription factor interactions during pancreatic β-cell differentiation
16:30 – 16:45	**Florina Moldovan** - Estrogen Regulation of POC5 centriolar protein by ERafla is Altered in Human Scoliotic Cells
16:45 – 17:00	**Efthalia Preka** - Dynamic transcriptomic alterations of microglia following cranial irradiation in the juvenile mouse hippocampus

17:00 – 17:30 **Invited Talk: Antonio Herrera Camacho** - Spatiotemporal transcriptomic mapping of the mouse embryonic brain under folate-deficient conditions

Friday 31 March 2023

Session 2. Part II
Artificial Intelligence in Single Cell Genomics

08:30 – 09:10 **Keynote Talk: Oliver Stegle** - From genotype to phenotype with single-cell resolution

09:10 – 09:25 **Piotr Słowiński** - Model-driven AI for multi-omics data analysis

09:25 – 09:40 **Yunhee Jeong** - scMaui: variational autoencoders combined with adversarial learning reveal cellular heterogeneity from single-cell multiomics data and handle multiple batch effects independently

09:40 – 10:10 **Invited Talk: Krasimira Tsaneva Atanasova** - Topological data analysis for single cell sequencing data

10:00 – 10:30 **Coffee Break**

Session 1. Part V
Single Cell Multi-Omics Technologies and Analyses

10:40 – 11:10 **Invited Talk: Irene Papatheodorou** - Computational analysis of cell atlases across species

11:10 – 11:20 **Karin Engström** - Comparison between scRNA-seq and snRNA-seq as sequencing strategies in five different tissues

11:20 – 11:35 **Martina Höckner** - Single-cell analysis reveal Cd-related effects on immune cells (coelomocytes) from the earthworm Lumbricus terrestris

11:35 – 11:50 **Ana Hernández de Sande** - Cell-type-specific characterization of miRNA gene dynamics in immune cell subpopulations during aging and atherosclerosis disease progression at single-cell resolution

11:50 – 12:05 **Nadine Bestard Cuche** - Marked regional heterogeneity of white matter glia in the human frontal lobe, cerebellum and spinal cord

12:05 – 12:20	**Oksana Ivanova** - LMNA R482L mutation-specific impairments of skeletal muscle metabolism are associated with pathological accumulation of reactive oxygen species
12:20 – 12:50	**Invited Talk: Abel González** - Somatic mutations in cancer and normal tissues
12:40	**Award Ceremony and Closing Remarks**

Genes is a peer-reviewed, open access journal of genetics and genomics published monthly online by MDPI. The Spanish Society for Biochemistry and Molecular Biology (SEBBM) is affiliated with Genes and their members receive discounts on the article processing charges.

Journal Webpage: https://www.mdpi.com/journal/genes
Impact factor: **4.141** (2021); 5-Year Impact Factor: 4.474 (2021)

Welcome from the Chairs

Dear Colleagues,

It is our pleasure to invite scientists, academicians, young researchers and students from all over the world to attend in person and present their research at the International Conference "Single Cell Genomics Moving Forward" which will be held as a three-day event in Barcelona, Spain, from 29 to 31 March 2023.

Our goal is to facilitate innovative collaborative research on the interface of single cell genomics, spatial transcriptomics and artificial intelligence. The demand for integrating these fields is quickly growing, and by bringing together multi-disciplinary groups of scientists to present and exchange breakthrough ideas, we hope to stimulate advances in the Single Cell research field.

Our program includes keynote lectures, invited talks, and collaborative discussion sessions on recent outstanding achievements in Single Cell research. We particularly focus on identifying future trends and needs of the field and on proposing and discussing innovative interdisciplinary approaches to them.

The Conference is initiated and sponsored by the journal *Genes* (Impact Factor 4.141) with the goal of establishing a regular event to provide a venue for initiating collaborations, and discuss the latest achievements and challenges in single cell biology. Importantly, we hope to inspire young scientists to join this emerging and exciting field.

It is only through an exchange of the widest variety of research that we can offer the best forward solutions in this challenging field.

We look forward to meeting you and discussing your exciting single cell research in the beautiful city of Barcelona!

Dr. Anelia Horvath
Conference Chair

George Washington University, Washington, USA

Prof. Peter Young
Conference Chair

Emeritus Professor, University of York, UK

General Information

Genes 2023 - Single-Cell Genomics Moving Forward will be held at the AXA Convention Centre, Barcelona, on 29 – 31 March 2023.

Conference Venue

AXA Convention Centre
Carrer Deu i Mata, 111, 335, 08029 Barcelona

Registration Desk

The desk for registration, information and distribution of documents will be open from 13:00h on 29 March 2023.

Certificate of Attendance

Participants of the event will be able to downlod an electronic Certificate of Attendance by accessing their dashboards on Sciforum.net once the event is concluded. The certificates will be found under "My Certificates" category.

Disclaimer

Delegates will receive a name-badge at the Information Desk, upon registration. The badge must be worn prominently in order to gain access to the congress area during all scientific and social events. Admission will be refused to anyone not in possession of an appropriate badge.

Insurance

The organizers do not accept liability for personal accidents, loss, or damage to private property incurred as a result of participation in **Genes 2023 - Single-Cell Genomics Moving Forward**. Delegates are advised to arrange appropriate insurance to cover travel, cancellation costs, medical, and theft or damage of belongings.

The AXA Convention Centre

Address: Carrer Deu i Mata, 111, 335, 08029 Barcelona

The AXA Convention Centre is located in a vibrant modern zone of the city, next to the main road and with easy access from the airport, as well as the urban and suburban areas.

It is part of the "L'Illa Diagonal", a modern complex which include a shopping centre, two 4-stars hotels, several offices, a sports centre, a public park and a parking with 2500 places.

Contact persons during the event

Elena González

Email: elena.gonzalez@mdpi.com

Pablo Velázquez

Email: pablo.velazquez@mdpi.com

Emergency Information

All emergencies in Spain: 112 (no area code needed)
Ambulance (Ambulancia) and health emergencies: 061 or 112
Fire brigade (Cuerpo de bomberos): 080 or 112
Spanish National Police (Policía nacional): 091

Abstracts

Talks

Using Single-Cell Multiomics in Cancer: From Knowledge to Applications

Manel Esteller

Josep Carreras Leukaemia Research Institute, Badalona, Barcelona, Spain

Bulk multiomics have represented an important step in our understanding of cancer initiation and progression. However, those methodologies do not exhibit the necessary granularity to interrogate in detail the many axes and pathways involved in tumor biology such as celular microenvironment, clonal heterogeneity, intratumoral signatures and spatial organization that can also be critical to define metastasis and drug-resistance mechanisms. In recent times years, single-cell approached have irrupted as disruptive technologies that allow the identification of particular cell populations within the tumor that could be useful for translational purposes such as improving diagnosis, prognosis or providing new actionable genes, networks and cell types for therapies. The latest achievements in this field enable now the combination of different single-cell “omics” in one. In the presentation, beyond the above described points, I will address studies of our research in the area dealing with the use of spatial single-cell transcriptomics to map the hallmarks of cancer in human malignancies; the application of this technique in the diagnosis of cancer of unknown primary; and the combination of single-cell genomics and proteomics to predict the clinical behaviour of myelodysplastic syndrome, a form of pre-leukemia.

Single-cell multi-omics profiling in the study of colorectal cancer evolution

Silvia Ogbeide

Earlham Institute, Macaulay group, UK

The heterogenic nature of cancers limits the efficacy of targeted therapies and poses a significant challenge for precision medicine. This is mainly due to the introduction of artificial selection by chemotherapy and radiation, which inadvertently exert a selective pressure that eliminates sensitive cancer phenotypes while promoting the proliferation of variant cells with higher Darwinian fitness and malignant potential. Looking at evolutionary cues in cancer will help us understand the disease and provide insights that can help design novel therapeutic strategies that account for its adaptive nature, promote treatment responsiveness, and delay therapeutic resistance.

Current knowledge of genomic abnormalities in cancer is mostly derived from high-throughput sequencing analyses performed on bulk tumour specimens. However, bulk-sequencing methods are not sensible enough to detect genomic changes present in minority subclones. For this reason, single-cell sequencing methods are preferable to understand cell-to-cell variability.

We have applied single-cell genome and transcriptome sequencing (G&T-seq) on patient-derived organoids (PDOs) of metastatic colorectal cancer samples to characterise the cellular diversity of organoids, follow the evolution of clones under two AKT-inhibitors (MK-2206 and AZD5363), and explore the mechanisms by which chemoresistance is achieved. We first analysed the transcriptomes of 316 cells of organoid-derived single cells from control and treated organoids. We observed a shift in gene expression in cells isolated from drug-resistant PDOs, with one cluster expanding and another decreasing in the resistant PDOs compared to the untreated organoid. These resistant cell clusters express genes that may have a role in the development of resistance to chemotherapy in colorectal cancer.

Intratumor heterogeneity is closely related to cancer progression. The combined examination of a single cell's genome and transcriptome in PDOs offers scientists the opportunity to model tumour progression in vitro, focusing on understanding the clinical relevance of resistant subclones to drug therapies.

Single-cell multi-omic lineage tracing uncouples tumour initiation and drug tolerance in breast cancer

Francesca Nadalin[1, 2], Matteo Marzi[1], Patricio Fuentes[1], Maria Pirra Piscazzi[1], Simone Procaccia[1], Montserrat Climent[1], Paola Bonetti[1], Irene Papatheodorou[2], John Marioni[2, 3], Francesco Nicassio[1]

1 Center for Genomic Science of IIT@SEMM, Fondazione Istituto Italiano di Tecnologia, Milano, Italy
2 European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, UK
3 Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK

Cancer is a heterogeneous disease in which multiple, phenotypically distinct subpopulations co-exist. Tumour evolution is a result of a complex interplay between genetic and epigenetic factors. Predicting the different cancer phenotypes requires linking the molecular state of a clone to its fate with high precision and without affecting the system composition. To this aim, we combine multi-omic single-cell sequencing with phenotypic assays and propose GALILEO, a framework providing lineage tracing, transcriptomic, and chromatin accessibility information simultaneously at single-cell resolution. We applied the GALILEO workflow to a triple-negative breast cancer model, with the aim of characterising the pre-existing cellular subpopulations along with their propensity to sustain aggressive cancer phenotypes, namely tumour initiation and tolerance to chemotherapy. Upon tumour engraftment in mice, we trace cancer cells and observe that the highly tumorigenic pool segregates with two transcriptionally distinct subpopulations, one (S1) displaying specific and stable marker expression and the other one (S3) characterised by genomic amplification. GALILEO identifies a major epigenetic module common to the two subpopulations, which may explain the phenotypic relationship between them. However, when exposed to cytotoxic treatment, the S1 subset shows lower drug sensitivity both in vitro and in vivo. Finally, we trace the clonal response to chemotherapy across time and identify two distinct paths to drug tolerance, each showing memory of their pretreatment state. We anticipate that our approach will make an important step forward in the understanding of the complex interplay of genetic and epigenetic factors involved in the evolution of cancer diseases as well as in the prediction of aggressive cancer phenotypes.

An investigation of the origin of neuroblastoma with single-cell transcriptomic analyses

Enrico Sebastiani, Angelo Guarniero, Emanuele Rosatti, Chiara Maria Argento, Alessandro Quattrone

Translational Genomics Laboratory, department of Cellular, Computational and Integrative Biology (CIBIO), University of Trento, Italy

Neuroblastoma is the most common extracranial solid tumor diagnosed in the first year of life, accounting for 15% of all pediatric deaths caused by cancers. Neuroblastoma arises during the development of the sympathetic nervous system from neural-crest-derived cells undergoing a defective differentiation. Although the biological aspects of neuroblastoma development and progression have been studied for a long time, the mechanisms of its genesis are still largely unknown, and the cells from which this tumor originates have not yet been reliably identified. In this work, we studied the transcriptomes of 54 single-cell neuroblastoma datasets and 7 normal fetal adrenal glands to characterize the putative neuroblastoma cell of origin. At first, we define the differentiation trajectories that occur during normal adrenal gland development from Schwann cell precursors to sympathoblasts and chromaffin cells. Secondly, we demonstrated that neuroblastoma cells resemble the transcriptome of undifferentiated sympathoblasts. Finally, we studied the transcription factors and coregulated genes driving the differentiation process to identify a possible differentiation therapy that can help malignant cells to overcome the differentiation blocks.

Single-cell re-analysis of human transcriptomes reveals shared characteristics of immune ageing

Igor Filippov[1,2], Vivi R. Gregersen[1], Leif Schauser[1], Pärt Peterson[2]

1 QIAGEN Aarhus A/S
2 Tartu University

Ageing is accompanied by the immune system remodeling characterized by changes in both the abundance and phenotype of various immune cell subsets, as well as by low-grade systemic inflammation (inflammageing). This results in increased infection susceptibility and altered vaccine responses in old individuals. Furthering the understanding of cell-intrinsic immune changes regulators and pathways will aid the development of interventions that improve health outcomes in the elderly. The advances in next-generation sequencing allow generation of massive high-dimensional single-cell datasets that capture the phenotypic differences between immune cell populations. These data can be combined to mine for novel insights across donor cohorts. We performed an integrative analysis of several recently published single-cell transcriptomic datasets from old and young subjects to define the ageing immune transcriptome. We assed the utility of previously published senescence- and ageing-associated gene sets (SenMayo, GenAge, and others) in distinguishing granular immune cell subsets from old and young individuals. In addition, we leveraged the Qiagen Knowledge Base through Ingenuity Pathway Analysis to get a mechanistic understanding of ageing-associated gene signatures based on curated findings from the literature. Our analysis provides a comprehensive characterisation of immune pathways and regulators in ageing at the single-cell level.

Identification and analysis of cell-specific expressed genetic variants from scRNA-seq data

Allen Kim[1], Kai Saito[2], Zhe Yu[1], Hovhannes Arestakesyan[1], Evgenia Ulianova[1], Nathan Edwards[3], Anelia Horvath[1,3]

1 McCormick Genomics and Proteomics Center, School of Medicine and Health Sciences, The George Washington University, USA
2 Northeastern University, Boston, USA
3 Department of Biochemistry and Molecular Medicine, The George Washington University, USA
4 Department of Biochemistry and Molecular & Cellular Biology, Georgetown University, USA

To date, most genetic variation is studied in bulk sequencing datasets, where low (cellular) frequency variants are difficult to distinguish from sequencing errors and other artifacts. Low cellular frequency variants may indicate pre- or early-somatic clonality in cancer and normal tissues, or cell-specific RNA post-transcriptional modifications.
To address challenges posed by low frequency variation events, we have developed a set of methods and a framework for identification and analysis of Single Cell-specific Expressed Single Nucleotide Variants (sceSNVs) from single cell RNA-sequencing (scRNA-seq) data. Central for the framework is our new tool SCExecute, which enables the execution of software designed for bulk sequencing data on barcode-stratified, extracted on-the-fly, single-cell alignments.
Applying SCExecute in conjunction with variant callers developed for bulk sequencing data – GATK and Strelka2 - we explored, for the first-time, expressed genetic variation at cell-level across 28 publicly available cancer and normal datasets, including prostate cancer, non-small cell lung carcinoma, cholangiocarcinoma, neuroblastoma, normal fetal adrenal and normal embryo. This analysis identified over 100,000 previously unreported expressed SNVs, including somatic mutations and RNA-originating posttranscriptional variants. Applying statistical models, we outlined SNVs preferentially occurring or expressed in particular cell clusters along progressive pseudotime trajectories, including cells at trajectory branching positions, as well as cells at cell-fate end-points. Single-cell RNA e-QTL (scReQTL) analysis showed that the expression of these SNVs – defined as variant-carrying transcripts over all transcripts at the position of interest – correlated with the expression of their harboring gene. Furthermore, differential gene expression analysis between cells expressing these SNVs and the co-clustered cells expressing the reference allele, showed specific deregulated sets of genes, many of which in known functional networks with the SNV-harboring gene.
Our findings suggest that there is an unappreciated repertoire of cell-level expressed genetic variation, possibly recurrent and common across samples, that participates in transcriptome function and dynamics in both cancer and normal cells. Their appearance and, for some, relationship to certain gene-sets and cell types, suggests novel mechanisms and function for the expressed genetic variation, including in tissue evolution and cell fate.

Biophysics-informed interpretable deep learning to model gene regulation from single-cell and single-molecule genomic data

Nacho Molina

IGBMC - CNRS - University of Strasbourg, France

Biology is currently undergoing an incredible revolution. With the advent of single-cell genomics, it is now possible to characterize all cell types in the human body, thus gaining a systematic understanding of collective cell function in both health and disease. This presents exciting challenges in computational biology, as new approaches are required to extract valuable information and produce reliable predictions. Deep learning methods have been proven to be powerful tools and have been used for clustering, denoising data, and learning reduced dimensional latent spaces. However, the black-box nature of these methods hinders the interpretability of the latent variables and limits their predictability power. To overcome this limitation, we developed an interpretable variational autoencoder combined with biophysical modeling to characterize gene regulation from single-cell sequencing data. Our model was trained on large-scale datasets, and we identified key regulators responsible for the specific gene program of each cell type. Moreover, we were able to infer gene-specific non-linear response functions that can capture complex combinatorial regulations. Finally, we applied our model to gain a deeper quantitative understanding of gene expression dynamics throughout the cell cycle, leading to the estimation of cell-cycle dependent transcription and degradation rates for each gene and identifying key transcription factors underlying the transcriptional dynamics. In summary, we believe that our approach provides a powerful tool for analysing and interpreting single-cell sequencing data, enabling us to gain deeper insights into the mechanisms of gene regulation.

Legible Gene Regulatory Network Reconstruction using Multivariate Transfer Entropy Over Single-Cell Transcriptomics Data

Kyoung Jae Won

Department of Computational Biomedicine, Cedars-Sinai Medical Center

Computational approaches to infer gene regulation from transcriptomic data have been popular in the field of systems biology. Traditionally, transcriptomic information obtained from bulk cells (using microarray or RNA sequencing(RNAseq)) from various conditions has been used to understand the relationships between genes. Single-cell RNA sequencing (scRNAseq), by providing expression levels for millions of cells simultaneously, has replaced classical bulk RNAseq in reconstructing GRNs. Numerous algorithms have been developed to use the power of scRNAseq.

However, even the recent algorithmic development using scRNAseq can easily end up with a hair-ball structure in which multiple genes are interrelated to each other. The results of network inference are often influenced by the quality of data and indirect gene regulation and contain excessive false prediction.

To include multiple genes in reconstructing GRNs from scRNAseq, we developed scmTE. scmTE incorporated multivariate transfer entropy (MTE) to quantify potential causal relationships among a group of genes. Compared with our previous approaches using transfer entropy (TE) to measure the potential causal relationships between a pair of genes from the scRNAseq aligned along the pseudo-time, G-mTE can consider the relationships among multiple genes and produce interpretable results.

By utilizing multivariate transfer entropy, scmTE accounts for gene-to-gene interdependence when quantifying regulatory relationships. Benchmarking against other methods using synthetic data manifested that scmTE is a unique algorithm that did not produce a hair-ball structure (due to too many predictions) and recapitulated known ground-truth relationships with high accuracy. In silico knockdown experiments show that scmTE assigns higher scores for specific interactions important for differentiation processes. We applied scmTE to T-cell differentiation, myelopoiesis, and pancreatic development and identified known and novel regulatory interactions. scmTE provides a robust approach to infer interpretable networks by effectively removing unwanted indirect relationships.

Immune-cell-type signature discovery and random forest classification for the analysis of single-cell gene-expression datasets

Bogac Aybey[1, 2], Benedikt Brors[3, 4], Sheng Zhao[2], Eike Staub[2]

1 Faculty of Biosciences, Heidelberg University, Heidelberg, Germany.
2 Merck Healthcare KGaA, Clinical Measurement Sciences, Oncology Bioinformatics, Darmstadt, Germany.
3 Division of Applied Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany.
4 German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ) Heidelberg, Germany.

Background
Robust immune cell gene expression signatures are central to the analysis of single-cell studies in immuno-oncology. Nearly all known sets of immune cell signatures have been derived by making use of only single gene expression datasets. Utilizing the power of multiple integrated datasets could lead to high-quality immune cell signatures, which could be used as superior inputs to machine-learning-based cell-type classification approaches.
Results
We established a novel gene expression similarity-based workflow for the discovery of immune-cell-type signatures that leverages multiple datasets, here four single-cell expression datasets from three different cancer types. We used our immune cell signatures to train random forest classifiers for immune cell type assignment. We obtained similar or better prediction results compared to commonly used methods for cell-type assignment in two independent benchmarking datasets. Our gene signature set yields higher prediction scores than other published immune-cell-type gene sets in our random forest approach.
Discussion and conclusion
We demonstrated the quality of our immune cell signatures and their strong performance in a random-forest-based cell-typing approach. We argue that classifying cells based on our comparably slim sets of genes accompanied by a random-forest-based approach not only matches or outperforms widely used published approaches but also enables unbiased downstream statistical analyses of differential gene expression for 90% of all genes whose expression profiles have influenced the cell-type classification.
Comments
Please check that the intended meaning has been retained throughout.

Learning from Data in Single-Cell Transcriptomics

Sandrine Dudoit

Department of Statistics, Division of Biostatistics, School of Public Health University of California, Berkeley, USA

The ability to measure gene expression levels for individual cells (vs. pools of cells) is crucial to address many important biological questions, such as the study of stem cell differentiation, the detection of rare mutations in cancer, or the discovery of cellular subtypes in the brain. Single-cell transcriptome sequencing (RNA-Seq) allows the high-throughput measurement of gene expression levels for entire genomes at the resolution of single cells. RNA-Seq studies provide a great example of the range of questions one encounters in a Data Science workflow, where the data are complex in a variety of ways, there are multiple analysis steps, and drawing on rigorous statistical principles and methods is essential to derive reliable and interpretable biological results. In this talk, I will provide a survey of statistical questions related to the analysis of single-cell RNA-Seq data to investigate the differentiation of stem cells in the brain, including, exploratory data analysis, dimensionality reduction, normalization, expression quantitation, cluster analysis, and the inference of cellular lineages.

Multimodal methods in single cell immunogenomics

Merja Heinäniemi

University of Eastern Finland, Finland

Bone marrow represents a highly complex tissue that gives rise to multiple blood and immune cell populations. Current methods to analyse blood cells in clinical setting are based on surface protein profiling with flow cytometry and microscopy. Single cell methods can now capture all RNAs, large panels of cell surface proteins and examine chromatin and DNA sequence from the same cell. In this talk, I will present neural network model-based analysis to capture the full molecular fingerprint of bone marrow cell phenotypes and present computational methods that allow characterizing the regulatory networks that underlie cell state evolution in immune cells during aging, disease and cancer.

Gene regulatory network analysis of decidual stromal cells and natural killer cells

Kalle T. Rytkönen[1, 2], Nigatu Adossa[1], Sebastián Zúñiga Norman[1], Tapio Lönnberg[1], Matti Poutanen[2], Laura L. Elo[1, 3]

1 Turku Bioscience Centre, University of Turku and Åbo Akademi University
2 Institute of Biomedicine, Research Centre for Integrative Physiology and Pharmacology, University of Turku
3 Institute of Biomedicine, University of Turku

Human reproductive success depends on a properly decidualized (differentiated) uterine endometrium that allows implantation, the formation of the placenta, and maintenance of the pregnancy. Two crucial decidual uterine cell types in this process include stromal cells (ESF and DSC) and natural killer (dNK) cells. Recent single-cell RNA sequencing (scRNA-seq) of cultured stromal cells and assembloids has provided detailed insights into the differentiation process of these cells. Here, we reassessed the stromal and NK cell subpopulations using single-cell data from first-trimester selected terminations of pregnancy and conducted gene regulatory network (GRN) analysis of decidual stromal and NK cells. For stromal cells, we discovered specific first-trimester cell populations corresponding to previously in vitro-described DSCs and senescence DSC (snDSC). With GRN analysis, we discovered previously in vitro-described stromal decidualization regulators for early pregnancy cellular states, and novel ones including DDIT3 and BRF2, which are major regulators of oxidative stress protection. For the dNK cells, we describe transcription factors involved in the previously detected immunotolerant (dNK1) subpopulation, including FOXP2, IRX3, and RELB, a repressor of the NFKB pathway, and contrasted these regulators to the less immunotolerant population (dNK3) with predicted TBX21 (T-bet)-, IRF2-, and IRF7-mediated regulation. We then studied the translational relevance of our results by intersecting the target genes from the detected regulons with cell-type-specific differentially regulated genes from preeclampsia and recurrent pregnancy loss. Notably, specific dNK1 regulon targets were enriched with pregnancy disorder downregulated gene sets, whereas specific dNK3 regulon targets were enriched with pregnancy disorder upregulated gene sets. In summary, our findings extend the previously defined subcluster-specific transcriptional patterns by robust prediction of master transcription factors directing these states.

Single cell multiomic analysis of Clonal Hematopoiesis reveals enhancer deregulation and increased COVID-19 inflammation severity

Alexandre Gaspar-Maia

Department of Lab Medicine and Pathology, Mayo Clinic

Coronavirus disease 2019 (COVID-19) is associated with significant morbidity and mortality, albeit with considerable heterogeneity among affected individuals. It remains unclear which factors determine severity of illness and long-term outcomes. Emerging evidence points towards an important role of preexisting host factors, such as a deregulated inflammatory response at the time of infection. Here, we demonstrate the negative impact of clonal hematopoiesis, a prevalent clonal disorder of ageing individuals, on COVID-19-related cytokine release severity and mortality. We show that mutations in the gene coding for the methylcytosine dioxygenase TET2 promotes amplification of classical and intermediate monocyte subsets. Using single cell multiomic sequencing approaches, we identify cell-specific gene expression changes associated with the loss of TET2 and significant epigenomic deregulation affecting enhancer accessibility of a subset of transcription factors involved in monocyte differentiation. We further identify EGR1 down-regulation secondary to TET2-mediated hypermethylation, resulting in deregulation of hematopoietic stem cell differentiation and monocyte lineage commitment. Together, these data provide a mechanistic insight to the poor prognostic value of clonal hematopoiesis in patients infected with Sars-COV2.

Expansion of memory Vδ2 T cells following repeated exposure to mRNA SARS-CoV-2 vaccination

Sara terzoli

Laboratory of Clinical and Experimental Immunology, IRCCS Humanitas Research Hospital, Rozzano, Milan, Italy
Department of Biomedical Sciences, Humanitas University, Pieve Emanuele, Milan, Italy

Current studies investigating the cellular targets of mRNA SARS-CoV-2 vaccines focus on αβ T cell memory. Alternatively, γδ T cells also provide rapid cellular immunity against pathogens. Here, we used single-cell RNA-sequencing combined with matched γδTCR sequencing to delineate the molecular changes in human γδ T cells in a longitudinal study following two doses of vaccination with Pfizer-BioNTech BNT162b2. At single-cell resolution, a heterogenous Vδ2 T cell population shows multiple grades of responsiveness to the SARS-CoV-2 vaccine. While the first dose of mRNA SARS-CoV-2 vaccine induces a first Vδ2 T cell activation, it is the second vaccine administration that significantly boosts the immune response exerted by these cells. In particular, double vaccination promotes the effectorness of the Vδ2 T cell subset together with their clonal expansion and the development of memory features. This kinetic of vaccine-induced activation of Vδ2 T cells is also confirmed in vitro with T cells harvested from PBMC and stimulated with SARS-CoV-2 spike-peptides. Indeed, the second vaccine challenge triggers a significantly higher production of IFNγ secretion by Vδ2 T cells. Taken together, our findings indicate that the activation and expansion of Vδ2 T cells parallel the αβ T cell memory responses after the administration of two doses of mRNA SARS-CoV-2 vaccine.

Optimized tumor metastatic phenotypes among distal metastatic lesions in pancreatic ductal adenocarcinoma

Alejandro Jiménez Sánchez[1], Yubin Xie[1], Roshan Sharma[1], Shigeaki Umeda[2], Wungki Park[2], Christine Iacobuzio-Donahue[2], Dana Pe'er[1]

1 Computational and Systems Biology Department, Sloan Kettering Institute
2 Human Oncology and Pathogenesis Program, Sloan Kettering Institute

Computational methods that provide insights beyond cell-type classification and where assumptions are not violated in cancer are still lacking. Here, we employed archetype analysis to find optimized cancer phenotypes in a patient with metastatic pancreatic ductal adenocarcinoma. We collected multiple metastatic lesions and multiple samples from each lesion. We collected 86,000 high-quality nuclei from 6 distal metastatic sites and the primary tumor. Using these data, we employed archetype analysis, a statistical method that aims to find extreme data points in multidimensional data; we found well-defined extreme tumor phenotypes that relate to different cell type fundamental functions, namely: pancreatic epithelium, cell cycle, collagen and extracellular matrix reorganization, fatty acid metabolism and tricarboxylic acid (TCA) cycle metabolic shift, metal ion response, senescence-like phenotype, immune-activated, HLA class II, angiogenic, and epithelial–mesenchymal-transition (EMT). We find that the pancreatic epithelium archetype has clear ductal and acinar transcriptomic phenotypes that support the acinar-to-ductal metaplasia process, considered the main origin of pancreatic pre-neoplastic lesions that eventually develop into PDAC. To our surprise, we find independent archetypes characterized by cell-cycle gene expression. Differential expression analysis revealed that two of these archetypes differ by their potential mode of spread according to the metastatic site, related to direct spread or hematogenous spread. We also find archetypes of cells showing a metabolic shift towards using fatty acids as a potential energy source. Finally, through a multivariate Cox hazard model survival analysis, adjusting for clinical covariates using PDAC TCGA data, we find that two EMT- and MET-related archetypal signatures confer a higher or lower risk of PDAC-related death, respectively. We have adapted archetype analysis to the use of single-cell data in a cancer-specific context and showed that this computational approach could aid in the detection of highly specialized cancer cell phenotypes with clinical value.

The Functional Role of Immune-Related Intercellular Signaling Networks during Tissue Development and Cancer

Merav Cohen

The Department of Clinical Microbiology and Immunology, Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

Intercellular signaling networks drive tissue fundamental processes. Immune cells present a wide range of versatile functions and distinctive plasticity, which position them as tissue signaling-hubs during tissue development, homeostasis and cancer. Recent genomic advances have greatly improved our understanding of cell composition and states; however, investigation of the molecular signatures of intercellular crosstalk at the single-cell level remains limited. By applying novel single-cell RNA-sequencing technologies, as physically interacting-cell sequencing (PIC-seq), we were able to investigate cell states and gene signatures of the interactions of different cell types, and in various conditions. We hypothesize that investigating the immune-non-immune bidirectional crosstalk during tissue development will shed light on novel molecular pathways that control cancer induction and progression. By performing dense sampling and transcriptional profiling of immune and non-immune cells, we were able to identify differences in immune, stromal, and epithelial cell types and states along tissue development and cancer progression. Projection of whole-tissue signaling networks along development and cancer revealed dynamics in immune cell composition and intercellular communication. Together, exploring tissue development, homeostasis and cancer from the point of view of immune-controlled signaling networks, has the potential to reveal novel diagnostic and therapeutic candidates with high biological and medical impact.

Benchmarking of methods for assessing circular RNA differential expression in single-cell RNA-seq data

Enrico Gaffo[1], Alessia Buratin[1], Silvia Orsi[2], Stefania Bortoluzzi[1]

1 Department of Molecular Medicine, University of Padova, Padova - Italy
2 Department of Biology, University of Padova, Padova - Italy

Circular RNAs (circRNAs) are transcripts with the 5' and 3' ends covalently closed by a backsplicing event and are today regarded as critical regulators of cellular processes and drivers of disease mechanisms. CircRNA expression is generally lower than the expression of linear RNAs, and their quantification from RNA-seq experiments is based on detecting sequenced reads spanning the backsplicing junction.

In bulk RNA-seq, we previously observed that such characteristics of circRNA expression and its estimation inflate zero and small counts in the expression matrices and hinder the performance of standard methods for differential expression assessment.

We analysed circRNA expression in single-cell RNA-seq (scRNA-seq) data and observed that the fraction of zero counts (> 95%) is even more prominent than in bulk RNA-seq (up to 72%).

We evaluated 38 different parameter configurations of 14 differential expression methods (DEMs) on scRNA-seq circRNA expression data. The DEMs were applied to circRNA expression datasets simulated from real datasets of the CircSC database with a semiparametric approach and evaluated according to computational time, type I error control, the false discovery rate (FDR), and sensitivity.

The DEMs ran in a reasonable time and scaled well as the number of cells in the dataset increased. Many DEMs crashed with several dataset instances, especially when the number of cells was smaller than 100. A few DEMs adequately controlled the type I error, including glmGamPoi, edgeR, NBID, ROTS, SAMseq, and the Wilcoxon test. However, all DEMs performed poorly and worse than in bulk RNA-seq, showing high FDRs and low sensitivity.

Our results confirm the challenging nature of circRNA expression analysis in scRNA-seq and prompt the development of new tools to study circRNAs at the single-cell level.

Unlocking the potential for single-cell sequencing at scale with combinatorial indexing

Amy Hamilton, Maggie Nakamoto, Lin Lin, Jerushah Thomas, Dmitry Pokholok, Felix Schlesinger, Toumy Guettouche, Frank Steemers

ScaleBio

The ability to examine populations at the single-cell level has led to novel advances across many fields. Cell throughput of single-cell experiments has not scaled rapidly due to high costs, making studies that require large numbers of cells prohibitively expensive. To address this, we use combinatorial indexing technology, which uses multiple rounds of in situ barcoding to achieve high throughput without instrumentation. Here, we show two workflows using kits from ScaleBio to profile chromatin accessibility and transcriptome data.

In the first workflow, used for scATAC profiling, barcodes are added upstream of an on-market system. Nuclei from up to 24 samples are distributed across 24 wells for barcoded tagmentation. Nuclei are then pooled and can be superloaded onto existing capture systems. Despite loading 100,000 nuclei per channel, the additional barcode creates a low effective doublet rate.

In the second workflow, used for scRNA profiling, cells are taken through three rounds of sequential barcoding in an entirely plate-based workflow. A reduction in the hands-on and total workflow time is achieved through streamlined reagents and the use of smart plastics for pooling and spinning between barcoding steps, creating a fast workflow yielding up to 125,000 cells per experiment.

Prior to beginning both workflows, human and mouse nuclei (scATAC) or cells (scRNA) were mixed. Despite loading 100k nuclei (scATAC) or 125k cells (scRNA), the effective doublet rates remain low (5%), with effective recovery of nuclei/cells across multiple sites and users (~55-60% from the final barcoding step). Analysis shows a clean distinction between human and mouse populations with good sensitivity and low background noise. PBMCs with similar loading rates show high barcode and mapping rates, good recovery of nuclei/cells, and successful identification of major PBMC subsets.

Taken together, these data show that ScaleBio single-cell sequencing kits can massively increase the throughput while decreasing the cost per cell, enabling high-throughput single-cell analysis.

Mapping the functional impact of immune disease associated regulatory elements through single-cell CRISPR-based screens

Celia Alda Catalinas[1], Ximena Ibarra-Soria[1], Jorge Esparza Gordillo[1], Christina Flouri[1], Bin Sun[1], Sebastian Ullrich[1], Sunir Malla[1, 2], Muhammad Maqbool[1], Andrew Adams[1], Alison Acevedo[1], Carl Fishwick[1], Gerard Drewes[1], Radu Rapiteanu[1]

1 Genomic Sciences, GlaxoSmithKline Research and Development, Stevenage, UK
2 MedDesign BioDisc Antibody Selections, GlaxoSmithKline Research and Development, Stevenage, UK

Drug targets with human genetic evidence are expected to increase clinical success by at least two-fold. Yet, translating disease-associated genetic variants into functional knowledge remains a fundamental challenge of early drug discovery. Immune disease associated variants are enriched within regulatory elements, such as distal enhancers, found in cell type-specific open chromatin regions. CD4+ T cells play a central role in regulating the immune system in health and disease. To identify the genes and, thus, the molecular programs modulated by immune disease associated regulatory elements, we developed crisprQTL, a CRISPR-based, single-cell functional screening approach (CROPseq) in primary human CD4+ T cells that enables interrogation of transcriptomic changes induced by modulation of regulatory elements at scale. Firstly, we enabled highly efficient CRISPRi in primary CD4+ T cells via CROPseq optimised vectors and successfully profiled single-cell transcriptomes. Subsequently, we performed a pilot screen targeting 31 previously validated intergenic regulatory elements and 26 transcription start sites (TSSs), profiling approximately 280,000 CD4+ T cell single-cell transcriptomes. After systematically interrogating different analytical pipelines, we identified the expected cis effects for >75% of the intergenic elements perturbed, validating our method. Finally, we used this framework to screen 244 immune disease-associated regulatory elements in primary CD4+ T cells derived from 5 donors, facilitating the large-scale mapping of regulatory elements to genes. In conclusion, we present a functional genomics approach that enables the generation of enhancer-gene maps at scale in a primary cell in vitro model, opening up exciting new avenues for disease target identification with genetic evidence.

Robust and versatile ultra-high-throughput single-microbe genome sequencing

Vaidotas Kiseliovas, Duygucan Gul, Giedrius Zolubas, Rapolas Zilionis

Droplet Genomics

Whole-genome and targeted sequencing open a window to understanding the function of unculturable microorganisms. On one hand, metagenomic sequencing is attractive for its straightforward sequencing library preparation from bulk environmental samples but only offers limited resolution into individual species. On the other hand, single-microbe sequencing offers true single-clone resolution but can only meaningfully address the high biological diversity expected in environmental samples if performed on thousands of individual cells in parallel. To satisfy the need to study such large numbers of single-microbes per sample, well- and droplet-based approaches keep evolving in parallel to provide single-cell compartmentalization required during sequencing library preparation. However, these approaches suffer from a fundamental trade-off between throughput and versatility. Being individually addressable, microwells enable multi-step processing but are not scalable. Droplets offer a throughput of up to a million cells per experiment but only allow a limited number of processing steps to be performed. The later limitation is severely felt went it comes to microbial research, as harsh conditions typically required for lysis are incompatible with downstream nucleic acid barcoding steps. Our Semi-Permeable Capsule (SPC) technology combines the throughput of droplets with the versatility of wells by enabling a virtually unlimited number of processing steps on genetic material from millions of individual microbes in parallel. We demonstrate the use of SPCs for barcoding >100,000 individual microbial genomes to obtain single-microbe whole genome sequencing data of unprecedented quality.

TOTEM: A Web Tool for Tissue Enrichment on Gene Lists with Single Cell Resolution

Veredas Coleto-Alcudia, Fidel Lozano-Elena, Gonzalo Vera and Ana I. Caño-Delgado

Center for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Campus UAB, Bellaterra, Spain

Single cell sequencing techniques have brought an unprecedented resolution to transcriptomics. In plants, the well-organized cell lineages of the Arabidopsis primary root offer a suitable model to study the developmental trajectories and specialized transcriptional programs of different cell populations. Recent studies have prompted the application of single cell technologies to other Arabidopsis organs, like shoots and leaves (Shaw et al., 2021; Minne et al., 2022). However, despite the increased spatio-temporal resolution provided by those studies, it is difficult to capture all the different developmental stages existing at the same time in a certain organ, as well as certain populations with few number of cells (McFaline-Figueroa et al., 2020). Therefore, the combination of these datasets into a single atlas would allow the identification of rare and specific cell populations and their spatio-temporal expression patterns (Shahan et al., 2021) that drive specific processes of plant growth and adaptation to the environment. Because of the manipulation of this type of data can be difficult for non-experienced users and available tools only offer single or pair gene queries into not very high spatial resolution datasets, we created TOTEM, a web TOol for Tissue-EnrichMent. TOTEM is a user-friendly web tool designed to detect tissue-specific gene overrepresentations given a list of genes of interest. TOTEM cope the limitations of individual single cell datasets by incorporating root and leaves expression atlases integrating different available scRNAseq datasets. This enables the highest spatio-temporal resolution so far possible. Additional features include the capability to identify tissue-specific genes for a given expression atlas from different species or conditions. Overall, TOTEM offers a user-friendly tool to investigate tissue-specific processes and highlight the amenability of using plants for investigating biological processes with single cell resolution.

Acknowledgements & Funding

A.I.C.-D. has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 683163). A.I.C.-D. is a recipient of project BIO2016-78150-P funded by MCIN/AEI/ 10.13039/501100011033 and by "ERDF A way of making Europe"; V.C-A is funded by Severo Ochoa PhD Fellowship. We acknowledge financial support from the Spanish Ministry of Science and Innovation-State Research Agency (AEI), through the "Severo Ochoa Programme for Centres of Excellence in R&D" SEV-2015-0533 and CEX2019-000902-S funded by MCIN/AEI/10.13039/501100011033 and support from the CERCA Programme/Generalitat de Catalunya; F.L-E. received funding from project BIO2016-78150-P funded by MCIN/AEI/ 10.13039/501100011033 and by "ERDF A way of making Europe" and PIRSES-GA-2013-612583 and ERC-2015-CoG–683163 granted to A.I.C.-D. laboratory.

Harmonizing the annotation of single cells in normal and aberrant hematopoiesis with the Cell Marker Accordion

Emma Busarello[1], Zuhairia Ibnat[1], Giulia Biancon[2], Jennifer VanOudenhove[2], Fabio Lauria[3], Kristin Roseth Aass[4], Therese Standal[4], Gabriella Viero[3], Stephanie Halene[2], Toma Tebaldi[1, 2]

1 Department of Cellular, Computational and Integrative Biology (CIBIO), University of Trento, Italy
2 Yale Comprehensive Cancer Center, Yale University School of Medicine, New Haven, CT, USA
3 Institute of Biophysics, CNR Unit at Trento, Italy
4 Norwegian University of Science and Technology (NTNU), Trondheim, Norway

In recent years, single-cell and spatial technologies have significantly improved our knowledge of cellular heterogeneity and architecture in health and disease, providing unprecedented insight into hematopoiesis.
A crucial and challenging step in single-cell data analysis is the annotation of cell types and states. Marker genes provide an effective way to achieve accurate cell-type identification. The recent increase in databases collecting markers potentially simplifies the automatic annotation of cells. Nevertheless, each database contains dissimilar marker sets for the same cell type, leading to inconsistent annotations depending on the source choice. Furthermore, the lack of standard classification of cell types and validation of marker sets in the databases increases the annotation discrepancies.
Here, we developed the Cell Marker Accordion, a novel R Shiny web app addressing the need for robust and reproducible identification of hematopoietic cell types in single-cell datasets. As data sources, we integrated multiple databases of gene expression and cell surface markers for normal and aberrant hematopoiesis, obtaining a comprehensive set of 7650 markers associated with 145 cell types. Markers can be weighted by their specificity and evidence consistency scores, which measure the agreement of different annotation sources.
Importantly, we validated our approach on flow cytometry sorted blood cell populations and human bone marrow samples profiled with multi-omics single-cell methods (CITE-seq and Abseq). In all cases, the Cell Marker Accordion significantly improved the annotation accuracy of cell types with respect to any of the single-source databases.
Furthermore, we demonstrated the efficacy of the Cell Marker Accordion in identifying disease-critical cells in acute myeloid leukemia and multiple myeloma patient-derived single-cell datasets.
The Cell Marker Accordion is indeed a user-friendly and flexible tool that can be exploited to improve the annotation of hematopoietic populations in single-cell datasets focused on the study of human pathologies.

Image processing approach to spatial genomics data identifies tissue architecture and cell-contact-specific gene regulation

Kyoung Jae Won

Department of Computational Biomedicine, Cedars-Sinai Medical Center

Characterization of the tissue architecture promises to deliver insights into tissue development, cell communication, and disease. In silico spatial-domain-retrieval methods have been developed for spatial genomics data, assuming the transcriptional similarity of neighboring barcodes. However, domain-retrieval approaches with this assumption cannot work in complex tissues composed of multiple cell types. This task becomes especially challenging in cellular-resolution spatial transcriptomics methods. We developed Vesalius to decipher tissue anatomy from spatial genomics data by applying image processing technology. To utilise image processing, spatial genomics data are converted into RGB code.

Vesalius uniquely detected territories composed of multiple cell types and successfully recovered tissue structures in high-resolution spatial transcriptomics data including in the mouse brain, embryo, liver, and colon. Vesalius also detected the tissue territory from spatial epigenomics data (spatial CUT&Tag data and spatial ATACseq). Utilizing this tissue architecture, Vesalius identified tissue-morphology-specific gene expression and region-specific gene expression changes for astrocytes, interneurons, oligodendrocytes, and entorhinal cells in the mouse brain.

We further show a new strategy to study cell-contact-specific interactions from spatial genomics data. Our strategy identified a unique set of genes differentially expressed due to cell contact.

Alteration of transcriptions loaded in engineered extracellular vesicles using transmembrane scaffolds

Christopher Olson[1], Darin Boyes[2], Joy Ku[1], Renceh Angelo Flojo[1], Pengyang Zhang[1], Biao Lu[1]

1 Department of Bioengineering, Santa Clara University
2 Department of Biology, Santa Clara Universitiy

Transmembrane scaffolds play a critical role in engineering cell-derived extracellular vesicles (EVs) for targeted drug delivery and therapy1. Although a number of transmembrane scaffolds have been successfully used for such purposes, their potential impact on the loading of transcript cargoes remains largely unknown. Using a combination of gene fusion, molecular imaging, vesicle isolation, and deep sequencing, we investigated the effects of various transmembrane scaffolds, including human tetraspanins (CD9, CD61, CD81) and vesicular stomatitis virus glycoprotein (VSVG), on transcript loading into engineered EVs in cultured human cells. Transfection, fluorescence imaging, and characterization of isolated EVs confirmed the successful incorporation of each scaffold into the membrane of engineered EVs. Following the isolation of the total transcripts from EVs, we performed reverse transcription of total RNAs and subsequently conducted a deep sequencing analysis. Our results showed that all 15 types of RNA species were detected in non-engineered EVs. This included types of RNAs with high abundance (rRNA 25.6%, other noncoding RNA 18.7%, RefSeq 14.9%, RefSeq_antisense 12.8%, rfam 10.3%), moderate abundance (LincRNA 8.5%, miRNA 4.3%, Other_ncRNA_antisense 1.1%, tRNA 1.04%), and low abundance (lincRNA_antisense 0.95, tRNA_like 0.45%, piRNA 0.08%, CDBox 0.06%, scaRNA 0.007%, HAcaBox 0.0008%). Comparative analysis revealed the alteration of the abundance of RNA transcripts in EVs modified by transmembrane scaffolds as compared with non-engineered controls. Sorting and ranking by raw counts identified the top 100 RNA species for both RefSeq and miRNA loaded in EVs. Again, comparative analysis revealed discordance among the top 100 RNAs between the scaffold-engineered vs. non-engineered EVs. Importantly, a dramatic increase in the levels of scaffold transcripts was evident, with 1.7-, 133-, and 81.1-fold increases in CD9-, CD63-, and CD81-modified as compared to non-engineered EVs. Altogether, our study results demonstrate that EV engineering via transmembrane scaffolds alters the global RNA profiles and levels of both endogenous and exogenous RNA species loaded into EVs.

Inferring cell–cell communications from spatially resolved transcriptomics data

Dongyuan Wu[1], Jeremy Gaskins[2], Michael Sekula[2], Susmita Datta[1]

1 Department of Biostatistics, University of Florida
2 Department of Bioinformatics and Biostatistics, University of Louisville

Cellular communication through biochemical signaling is fundamental to every biological activity. Investigating cell signaling diffusions across cell types can further help us to understand biological mechanisms. In recent years, this has become a rather important research topic with the maturity of single-cell sequencing technologies. However, cell signaling activities are spatially constrained, and the single-cell data cannot provide spatial information for each cell. This issue may cause a high false discovery rate, and using spatial transcriptomics data is necessary. On the other hand, as far as we know, most existing methods focus on providing an arbitrary measurement to estimate intercellular communication instead of relying on a statistical model. It is undeniable that descriptive statistics are straightforward and more accessible, while a suitable statistical model can provide a more accurate and reliable estimate.

We propose a generalized linear regression model to infer cellular communications based on spatially resolved transcriptomics data, especially spot-based data. Our Bayesian approach estimates the communication scores between cell types with the consideration of their corresponding distances. Due to the property of the regression model, the proposed method naturally provides the direction of the effect of cell types on a specific signaling pathway that other approaches cannot offer. In summary, our novel model can fill in gaps in the inference of cell–cell communication and provide robust and straightforward results.

Macrophage and neutrophil heterogeneity at single-cell spatial resolution in inflammatory bowel disease

Alba Garrido-Trigo[1], Ana Corraliza[1], Marisol Veny[1], Isabella Dotti[1], Elisa Melón-Ardanaz[1], Aina Rill[2], Helena Crowell[3], Ángel Corbí[4], Victoria Gudiño[1], Miriam Esteller[1], Iris Teubel[1], Daniel Aguilar[1], M Carme Masamunt[1], Emily Killingbeck[5], Michael Leon[5], Youngmi Kim[5], Sudha Visvanathan[6], Domenica Marchese[7], Ginevra Caratù[7], Albert Martin-Cardona[8], Maria Esteve[8], Julian Panés[1], Elena Ricart[1], Elisabetta Mereu[2], Holger Heyn[7], Azucena Salas[1]

[1] Institut d'Investigacions Biomèdiques August Pi I Sunyer (IDIBAPS), Hospital Clínic, Barcelona, Spain
[2] Josep Carreras Leukaemia Research Institute (IJC), Badalona, Spain
[3] Department of Molecular Life Sciences, University of Zurich, Switzerland. SIB Swiss Institute of Bioinformatics, Zurich, Switzerland
[4] Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas (CSIC), Madrid
[5] NanoString Technologies, Seattle, WA, USA
[6] Translational Medicine and Clinical Pharmacology, Boehringer-Ingelheim Pharmaceuticals Inc, Ridgefield, CT, United States of America.
[7] Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona
[8] Department of Gastroenterology, Hospital Universitari Mútua Terrassa, Universitat de Barcelona

Crohn's disease and ulcerative colitis are chronic inflammatory bowel diseases that show perplexing heterogeneity in prognosis and response to treatment. However, the cellular basis for this heterogeneity remains uncharacterized. Here, we applied single-cell RNA sequencing and CosMxTM spatial molecular imaging to the human colon and found the highest diversity in cellular composition in the myeloid compartment of patients suffering from inflammatory bowel disease. We first detected, in both healthy and inflamed colons, a novel resident macrophage population (M0) that differs from the previously described M2. In addition, we identified a variety of classical inflammatory M1 macrophages and an inflammation-dependent alternative (IDA) macrophage population, all of which are unique to inflamed tissues. By integrating single-cell transcriptomic and spatial data, we identified two populations of IDA macrophages: subepithelial IDA macrophages highly expressing NRG1, and NRG1low IDA macrophages, which were expanded within the submucosa and in Crohn's disease granulomas. Spatial analysis unravels their proximity to abundant inflammatory fibroblasts, suggesting that activated fibroblasts may promote macrophage polarization in patients. Moreover, epithelial organoid culture showed the upregulation of OLFM4 and downregulation of LGR5 upon neuregulin 1 stimulation, demonstrating epithelial crosstalk. In addition, for the first time, we characterized the transcriptional signatures of intestinal neutrophils, finding three different neutrophil states, including a CXCR4+ population that is not found in the blood. Our study combines single-cell sequencing and spatial analysis and unravels the complexity of the macrophage and neutrophil populations in inflammatory bowel disease, which, we argue, could contribute to patient-to-patient heterogeneity.

Gene colocalization-aware segmentation of image-based spatial transcriptomics using Graph Neural Networks

Kang Jin[1], Zuobai Zhang[2], Bruce J. Aronow[1], Jian Shu[3]
1 Cincinnati Children's Hospital Medical Center
2 Mila - Québec AI Institute
3 Massachusetts General Hospital

Single-cell spatial transcriptomics technologies based on in situ hybridization or sequencing provide detailed transcriptomics organization at subcellular levels. However, inaccurate segmentation severely restricts cell detection and limits potential downstream insights. Here, we propose Bering, a Graph Neural Network (GNN) model utilizing transcript colocalization relationships in alignment with morphological characteristics for instance segmentation in spatial transcriptomics data. Using GNN, Bering improves inference of cell labels and boundaries by learning effective transcript-colocalization representation and modeling local cell label dependency through label propagation. We first demonstrate that transcript-colocalization is indicative of cell labels and their specific morphological features. To evaluate performance, we compared Bering with other methods and observed significantly improved detection of cells and recovery of cell type-specific transcripts in coarsely-segmented datasets. Furthermore, we demonstrated satisfying performances in four different protocols and proved the generalizability of Bering across various spatial transcriptomic technologies and tissue environments. In summary, Bering is a gene colocalization-aware deep learning approach that enhances single-cell spatial transcriptomics segmentation performance in a variety of scenarios.

Dosage compensation dynamics unmasked by allele-specific single-cell transcriptomics

Antonio Lentini, Huaitao Cheng, JC Noble, Natali Papanicolaou, Christos Coucoravas, Nathanael Andrews, Qiaolin Deng, Martin Enge, Björn Reinius

Karolinska Institutet

X-chromosome inactivation and X-upregulation are the fundamental modes of chromosome-wide gene regulation that collectively achieve dosage compensation in mammals, but the regulatory link between the two remains elusive and the X-upregulation dynamics were unknown. We used allele-resolved single-cell RNA-seq combined with chromatin accessibility profiling to dissect their separate effects on RNA levels during mouse development. Surprisingly, we uncoved that X-upregulation elastically tunes expression dosage in a sex- and lineage-specific manner, and moreover along varying degrees of X-inactivation progression. Male blastomeres achieve X-upregulation upon zygotic genome activation while females experience two distinct waves of upregulation, upon imprinted and random X-inactivation; and ablation of Xist impedes female X-upregulation. Female cells carrying two active X chromosomes lack upregulation, yet their collective RNA output exceeds that of a single hyperactive allele. Importantly, this conflicts the conventional dosage compensation model in which naïve female cells are initially subject to biallelic X-upregulation followed by X-inactivation of one allele to correct the X dosage. Crucially, these discoveries could only be made at allelic resolution in single cells. Together, our study provides key insights to the chain of events of dosage compensation, explaining how transcript copy numbers can remain remarkably stable across developmental windows wherein severe dose imbalance would otherwise be experienced by the cell.

Cell-Level Analysis of Telomeric Transcripts and Telomere-Associated Gene Variants

Evgenia Ulianova[1], Christian Dillard[1], Prashant NM[1,2], Hongyu Liu[1,3], Anelia Horvath[1]

1 McCormick Genomics and Proteomics Center, School of Medicine and Health Sciences, The George Washington University, Washington, DC, 20037, USA
2 Departments of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
3 Division of Animal Sciences, University of Missouri, Columbia, MO, 65211, USA

Telomeres are repetitive DNA regions located at the termini of eukaryotic chromosomes. These structures are critical for the maintenance of an organism's genomic integrity. Factors critical to the function of telomeres include the telomeric repeat–containing RNA (TERRA) transcribed from telomere regions. TERRA has been shown to exhibit cross-regulation with a number of telomere maintenance factors, including shelterin and telomerase. Additionally, TERRA is known to modulate telomeric chromatin as well as recruit a number of complexes to the telomeric DNA. Still, much of TERRA biology remains to be elucidated. In this study, we sought to (1) identify and quantify TERRA expressions at the level of a single cell, (2) analyze gene-expression patterns associated with TERRA expression, and (3) investigate distinctions in cell phenotypes based on TERRA expression. To perform these analyses, we applied existing and newly developed tools on 25 tumor and normal publicly available single-cell RNA sequencing (scRNA-seq) datasets from prostate cancer, cholangiocarcinoma, neuroblastoma, normal fetal adrenal, and normal embryo samples. Applying our novel method, for the first time, we quantitatively detect TERRA expression at the single-cell level and show that TERRA-expressing cells exhibit whole-transcriptome profiles distinct from the rest of the cells in a dataset. This observation was consistent across different cancer and normal datasets. We show that evaluating TERRA expression at the level of a cell holds great potential for a deeper understanding of telomere biology.

Characterization of gene regulatory networks and combinatorial transcription factor interactions during pancreatic β-cell differentiation

Meltem Eda Omur[1,2,3], Samantha Mar[1,2], Francis Lynn[1,2], Wyeth Wasserman[1,2,3]

1 University of British Columbia, Canada
2 BC Children's Hospital Research Institute, Canada
3 Centre for Molecular Medicine and Therapeutics, Canada

Generating an infinite source of insulin-producing pancreatic β-cells using hPSCs is an active pursuit in type 1 diabetes (T1D) research towards treatment. T1D, the focus of our research, is characterized by pancreatic β-cell destruction by the body's immune system. While insulin replacement by injection or pumps is effective, the restoration of healthy β-cells could be curative. As the transplantation of donor cells is limited by supply, human pluripotent stem cells (hPSCs) could be the source as they can be differentiated into β-like cells under appropriate conditions.

Directed differentiation protocols include a series of transition steps to pancreatic lineage-specific progenitor cells that can be further differentiated into β-cells. Cells exhibit diverse characteristics, which are determined, in part, by underlying gene regulatory networks (GRNs). These networks are composed of specific combinations of transcription factors (TFs) and their target genes. Deciphering GRNs underlying hPSC-derived pancreatic β-cells differentiation offers opportunities to unravel the factors that define cellular phenotypes and improve differentiation protocols.

The presented research will describe the application of computational approaches to scRNA-seq datasets from different stages of pancreatic β-cells differentiation to gain insights into the mechanism driving transcriptomic changes during differentiation processes at a single-cell level. The characteristics and quality of the data were assessed, demonstrating the presence of cells of diverse characteristics from the differentiation protocol. Focused analyses of the cell populations explored the GRNs present, including both sets of genes with coordinated expression and TFs with enriched evidence of interaction with their target genes. Interactions between the TFs and across the GRNs will be presented, with a focus on combinatorial interactions of TFs in β-cell differentiation. These findings allow us to further develop models for gene regulatory network analysis as well as to improve existing differentiation protocols.

Estrogen Regulation of POC5 centriolar protein by ERafla is Altered in Human Scoliotic Cells

Amani Hassan[1], Edward T Bagu[2], Shunmoogum A Patten[3], Stefan Parent[4], Soraya Barchi[5], André Tremblay[6], Florina Moldovan[7]

1 McGill University, Montreal
2 Department of Basic Biomedical Sciences, Sanford Medical School, University of South Dakota, USA
3 INRS-Institut Armand-Frappier
4 Sainte Justine Research Center
5 CHU Sainte Justine Research Center
6 CHU Sainte Justine Research Center and Université de Montréal
7 Université de Montréal

Adolescent idiopathic scoliosis (AIS) is a complex three-dimensional spinal deformity. The incidence of AIS in females is 8.4 times higher than in males. Several hypotheses on the role of estrogen have been postulated for the progression of AIS. Recently, POC5 gene was identified as a causative gene in AIS. POC5 is a centriolar protein that is important for cell cycle progression and centriole elongation. However, the hormonal regulation of POC5 remains to be determined. Here, we identify POC5 gene as an estrogen-responsive gene under the regulation of estrogen receptor ERalfa in normal osteoblasts (NOB) and other ERalfa positive cells. Using promoter activity, gene and protein expression assays, we found that the POC5 gene was upregulated by treatment of osteoblasts with estradiol (E2) through direct genomic signaling. We observed different effects of E2 in NOB and mutated POC5A429V AIS osteoblasts. Using promoter assays, we identified an estrogen response element (ERE) in the proximal promoter of POC5 which conferred estrogen responsiveness through ERalfa. Recruitment of ERalfa to the ERE of the POC5 promoter was also potentiated by estrogen. Collectively, these findings suggest that estrogen is an etiological factor in scoliosis through deregulation of POC5.

Dynamic transcriptomic alterations of microglia following cranial irradiation in the juvenile mouse hippocampus

Efthalia Preka[1], Alejandro Lastra Romero[1], Ying Sun[2], Christer Bestholtz[3], Adamantia Fragkopoulou[1], Ahmed Osman[1], Klas Blomgren[1, 4]

1 Department of Women´s and Children´s Health, Karolinska Institutet, Stockholm, Sweden.
2 Department of Immunology, University of Uppsala, Stockholm, Sweden
3 Department of Developmental Genetics, University of Uppsala, Stockholm, Sweden
4 Pediatric oncology, Karolinska University Hospital, Stockholm, Sweden

Cranial irradiation (IR) is an important modality in the treatment of high-grade brain tumors and metastases, but it is also considered a major risk factor for cognitive impairment in survivors, particularly in children. The underlying mechanisms are not fully understood, but a suggested cause is ablation of hippocampal neurogenesis, partly due to direct precursor cell death, partly due to persisting microglia-induced neuroinflammation. Microglia are key players in maintaining brain homeostasis and after IR they undergo a series of distinct transcriptomic and morphological alterations. We previously reported an acute IR-induced inflammatory response in the hippocampus 6 hours post-IR associated with neural progenitor cell death. Here, we extended our work and analyzed long-term effects of IR on microglia. Mice were subjected to whole brain IR with a single dose of 8 Gy and hippocampal tissue was collected 2 and 6 weeks thereafter. Single-cell RNA sequencing revealed an inflammatory response in microglia 2 weeks after IRR, dominated by type-I interferon (IFN) related genes. At this time point, RNA velocity analysis identified trajectories demonstrating that the IR-induced IFN cluster drives the transcriptome of microglia into either an apoptotic or a senescent-like phenotype. Two of the putative driver genes were Cathepsin B (Ctsb), which is a senescence-associated marker, and Signal transducer and activator of transcription 1 (Stat1), which is an IFN-related gene. 6 weeks after IR, we found that the vast majority of microglia had transitioned to a senescent-like state. In conclusion, we show that IR induces dynamic changes in microglia, shuttling them into a senescent and pro-inflammatory phenotype long after IR that negatively impacts neurogenesis and cognitive performance, suggesting elimination of senescent microglia as a potential therapeutic target to prevent the cognitive impairments in children after IR.

Spatiotemporal transcriptomic mapping of the mouse embryonic brain under folate-deficient conditions

Antonio Herrera Camacho

EPFL, Switzerland

Folate is a vitamin (B9) required for de novo purine synthesis pathway and DNA methylation. This dual requirement involves folate in essential functions during development, such as proliferation and epigenetic regulation of gene expression. Although the components and reactions involved in folate metabolism have been extensively studied, the differential sensitivity and impact of folate deficiency on embryonic cellular heterogeneity have yet to be characterized at the single-cell level. In recent years, spatial transcriptomics methods have rapidly emerged as a promising technology to study the spatial complexity of the heterogeneity of cell types and states described by single-cell omics. However, most available spatial transcriptomics analysis methods compromise accuracy and spatial throughput, limiting their use for generating large-scale comparative cellular maps across tissues under perturbed conditions.

Motivated by the unprecedented resolution provided by sc-omics, we use joint single-cell RNA + ATAC sequencing to define the molecular events conducting to brain malformation under folate-deficient conditions. Then, based on sc-RNA-seq data, we selected 200 genes to accurately map the expression of patterning markers across the aberrant brain structures induced by folate deficiency using hybridization-based in situ sequencing (HybISS) and a new and enhanced spot detection and decoding tool (CBAIDT). The integration of the different technologies allows us the spatiotemporal characterization of the neural cell populations involved in the neurodevelopmental malformations driven by the lack of folate during pregnancy.

From genotype to phenotype with single-cell resolution

Oliver Stegle

Division Computational Genomics and Systems Genetics, German Cancer Research Center (DKFZ), Germany

The study of genetic effects on gene expression using bulk-RNA sequencing has enabled the identification of quantitative trait loci (eQTL), which link inherited genetic variants to gene expression changes across different human tissues. Advances in single-cell RNA sequencing (scRNA-seq) now provide for unprecedented opportunities to greatly increase the resolution of eQTL mapping approaches, thereby assaying gene regulatory effects at the resolution of cell types, cell states and even individual cells contained in human tissues. In this talk, I will present novel experimental strategies and analytical methods to enable the required population-scale single-cells sequencing. I will then describe applications of these approaches to perform population-scale single-cell profiling of human induced pluripotent stem cells derived from more than one hundred donors across a differentiation trajectory towards endoderm and neuronal linages. Our data allow for the identification of temporal changes of gene regulatory effects across cell state changes and differentiation, and we describe novel disease-relevant linkages of eQTL variants that are relevant for human disease traits.

Model-driven AI for multi-omics data analysis

Piotr Słowiński

University of Exeter

I present two statistical models developed to study probability distributions of variant allele frequencies and genetic risk scores. The models are used to generate synthetic/surrogate data samples with known biologically interpretable properties. The real data are then compared with synthetic data by means of the earth mover's distance (EMD). Intuitively, the EMD is the minimal cost of work required to transform one 'pile of earth' into another, with each 'pile of earth' representing a probability distribution. Mathematically, the EMD is a Wasserstein distance and has been widely used in computer and data sciences. The EMD allows the similarity between observed and synthetic data to be rigorously quantified. The degree of similarity is used to infer properties of the real data sample and answer research questions. E.g., identifying genomic alterations or determining the disease prevalence within a cohort. The advantage of the proposed approach is its interpretability and ability to provide conceptual and mechanistic insights into underlying biological processes. In addition, if properly implemented, it can be used to verify that the analysed dataset satisfies mathematical assumptions used to develop the model. Finally, I discuss how the data analysis approaches based on stochastic models and synthetic data can be adapted for single-cell multi-omic data analysis and potential methods for combining model-driven and data-driven AI methods.

For the purpose of the presentation, I interpret AI in the broadest sense as a set of tools, techniques, and methodologies used to automate data analysis.

scMaui: variational autoencoders combined with adversarial learning reveal cellular heterogeneity from single-cell multiomics data and handle multiple batch effects independently

Yunhee Jeong[1,2], Jonathan Ronen[3], Wolfgang Kopp[3], Pavlo Lutsik[1], Altuna Akalin[3]

1 Division of Cancer Epigenomics, German Cancer Research (DKFZ), Germany
2 Faculty of Mathematics and Informatics, Heidelberg University, Germany
3 Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association, Berlin Institute for Medical Systems Biology, Germany

Integrating single-cell multiomics data calls for sophisticated computational models due to the high complexity. Successful single-cell multiomics integration models should not be biased by differences in feature dimension, variance, and sparsity across assays. Undesired sources of variability that obscure the cellular heterogeneity caused by experimental and data-processing workflows or individual-specific effects (e.g., batch effects) must also be taken into account.

We present a new variational autoencoder-based single-cell multiomics integration model, scMaui. It summarises cellular heterogeneity into a reduced dimensional latent space and identifies relevant features over different assays. In particular, scMaui supports two novel functions compared to previous methods: correction for multiple batch effects and assay-specific reconstruction loss. While most multiomics integration methods require multiple types of batch effect information merged in one discrete categorical vector, scMaui can deal with those independently via adversarial learning and accepts both categorical and continuous values. Since the distribution of omics features can differ depending on the type of assay and preprocessing, scMaui provides a wide range of reconstruction loss functions covering negative binomial, Poisson distributions, and so on. These characteristics ensure the high flexibility and usability of scMaui.

According to our benchmarking, scMaui not only outperformed other methods in population classification, clustering, and imputation tasks but also demonstrated the ability to reveal novel cellular heterogeneity and detect cell sub-populations that remain hidden in individual omics layers. Moreover, scMaui provides a user-friendly command line tool that does not require any additional programming skills and produces results in an interpretable format automatically saved in the designated directory. Consequently, we believe that the flexibility and accessibility of scMaui will extend its usage to a broader range of researchers in biology and medicine who want to analyse their own data.

Topological data analysis for single cell sequencing data

Krasimira Tsaneva-Atanasova[1], Vijai Anhad Dharmalingam[1], Anelia Horvath[2]

1 University of Exeter, UK
2 The George Washington Universiy, USA

Recently topological data analysis (TDA) has been rapidly gaining popularity in applications to complex biomedical data sets. Single cell sequencing data is no exception as topology offers a suite of techniques and tools that could be applied to a diverse set of measurements. This enables a holistic approach to robustly identifying multidimensional properties and relationships within a given high-dimensional dataset. Topological methods also naturally lend themselves to visualisation rendering them useful for applications that require interpretability and explainability. TDA offers more unbiased and rigorous approach to analysing complex data sets as it does not depend on prior hypotheses nor focus on pairwise relationships within the data in contrast to other established approaches such as supervised clustering and association analysis. A popular implementation of TDA, known as the Mapper algorithm, represents a high-dimensional data set as a network projected in a low-dimensional space usually obtained via a dimensionality reduction technique, e.g. tSNE or UMAP. In this setting each “node” could represent a group of cells “similar to” each other in the multiple dimensions on which the TDA is based, e.g. expression of genes and/or mutations. Links or “edges” in the network represent individual cells that are shared between nodes (groups of cells). Once the TDA network has been created further clustering, graph-theoretical and statistical measures and techniques could be applied to investigate network substructures and/or reveal emergent properties of the data set. Moreover, TDA provides a geometric representation of the data that allows visualisation for readily identifying meaningful subgroups and discovering complex relationships and signals within the data in a less supervised, data-driven manner.

Comparison between scRNA-seq and snRNA-seq as sequencing strategies in five different tissues.

Karin Engström

EPI@LUND, Division of Occupational and Environmental Medicine, Department of Laboratory Medicine, Lund University, Lund, Sweden

Single-cell RNA sequencing (scRNA-seq) allows the comparison of transcriptomes of individual cells. However, there are some limitations with scRNA-seq, e.g., the fact that single-cell dissociation introduces stress-induced transcriptional artifacts and that scRNA-seq is incompatible with frozen archived material. Single-nucleus RNA-seq (snRNA-seq) is compatible with frozen samples and shows fewer technical issues due to dissociation. The current project aims to compare scRNA-seq and snRNA-seq in five different tissues and evaluate whether snRNA is a reliable tool to evaluate transcriptomics and thus can be a robust alternate sequencing strategy for RNA from these tissues. This was evaluated by comparing snRNA-seq and scRNA-seq data from tissues from two donors. Sequencing was performed using the 10X Chromium platform. The Seurat package in R was used, and different strategies/parameters were evaluated, such as employing standard Seurat normalization vs. SCTransform, using unintegrated datasets or integrating the datasets via Seurat integration or Harmony, and using supervised vs. unsupervised clustering. We evaluated protocols by cell/nucleus quality (e.g., genes detected per cell/nucleus and the percent of UMIs that were from mitochondrial genes) as well as cellular composition (e.g., the number of cell types captured and the relative abundance of each cell type). Preliminary results from the first tissue evaluated showed that scRNA-Seq and snRNA-Seq recovered the same cell types, but in different proportions. The choice of integration method had a large impact on the results.

Single-cell analysis reveals Cd-related effects on immune cells (coelomocytes) from the earthworm Lumbricus terrestris

Martina Höckner[1], Francesca Finotello[2], Lorenzo Merotto[2]

1 University of Innsbruck, Centre for Molecular Biosciences, Institute of Zoology
2 University of Innsbruck, Department of Molecular Biology, Digital Science Center

Immune cells, so-called coelomocytes, are one important part of innate immunity in earthworms since they lack an adaptive immune system. Coelomocytes are released through dorsal pores and present an excellent defense system against stress and soil pathogens.

The latter mechanism can be exploited to easily and non-invasively retrieve a single-cell suspension of coelomocytes—a perfect model system for single-cell analysis with no further need for treatment, as is the case for tissue samples, for example.

Coelomocytes do not only take part in immunity but are important for several other cellular processes such as the detoxification of toxic metals; we can, therefore, conclude that these cells play an indispensable role in the stress response to changing environmental conditions.

Morphological and molecular studies revealed already different coelomocyte cell types. Using single-cell analyses, we were able to detect 12 cell clusters. Exposure of earthworms to the toxic metal cadmium before harvesting of coelomocytes revealed a shift in cell clusters, suggesting a huge impact on the cell-type composition.

These results might explain in more detail the adverse effect that cadmium has on the earthworm immune system. A contribution not only to the effect of environmental stress but also to basic knowledge of the cell composition and function of coelomocytes in earthworms can be observed, which will further our understanding of how earthworms, as one of the most important soil-dwelling organisms, cope with natural and man-made challenges such as soil pollution through toxic metals such as cadmium.

Cell-type-specific characterization of miRNA gene dynamics in immune cell subpopulations during aging and atherosclerosis disease progression at single-cell resolution

Ana Hernández de Sande[1], Tanja Turunen[1], Celia Tunidor Centeno[1], Maria-Bouvy Liivrand[1], Tiit Örd[2], Tapio Lönnberg[3], Senthil Palani[3], Anne Roivainen[3], Minna Kaikkonen-Määttä[2], Merja Heinäniemi[1]

1 School of Medicine, University of Eastern Finland, Kuopio, North-Savo 70200, Finland
2 A. I. Virtanen Institute, University of Eastern Finland, Kuopio, North-Savo 70200, Finland
3 Turku Bioscience Centre, University of Turku and Åbo Akademi University, 20520, Turku, Finland

Aging is a step-wise process characterized by changes in tissue distribution and cellular heterogeneity. Specifically, changes in the distribution of adipose depots throughout the body dramatically affect tissue growth, plasticity, and function, leading to metabolic dysfunction and low-grade inflammation. These features participate in the development of aging diseases including type 2 diabetes, atherosclerosis, dyslipidemia, thermal dysregulation, and skin ulcers, among others.

Mature miRNAs that originated from their miRNA genes control gene expression at the post-transcriptional level to maintain cellular homeostasis. An altered mature miRNA expression is associated with aging and disease development. Similar to protein-coding genes, miRNA genes are transcriptionally regulated. Nevertheless, genome-wide miRNA gene annotation and quantification are rarely available, hindering the possibility of linking miRNA genes to mature miRNA-mediated contributions to gene regulatory networks.

To enable the measurement of miRNA gene expression and its contribution to gene regulatory networks that regulate cell state changes upon aging and disease transition, here, we provide a novel approach to quantify miRNA genes from single-cell RNA-sequencing data. We applied our approach to identify the differential expression of miRNA genes in immune cell subpopulations in the spleen and fat tissue during aging. Moreover, we linked the miRNA gene dynamics with gene regulatory network changes in fat tissue during atherosclerosis disease progression, providing new insight into lipid-associated macrophage biology. Collectively, our work enables the measurement of miRNA gene expression and reveals TF-miRNA regulatory networks in aging and disease.

Marked regional heterogeneity of white matter glia in the human frontal lobe, cerebellum and spinal cord

Luise A. Seeker[1], Nadine Bestard Cuche[2], Sarah Jakel[3], Nina Lydia Kazakou[2], Sunniva M. K. Bostrand[1], Alastair M Kilpatrick[1], Mukund Kabbe[4], Fabio Baldivia Pohl[5], Zahra Moslehi[5], Neil C Henderson[2], Catalina Vallejos[2], Gioele La Manno[6], Goncalo Castelo-Branco[5], Anna Williams[2]

1 University of Edinburgh, Centre for Regenerative Medicine, Edinburgh, United Kingdom
2 University of Edinburgh
3 2Ludwig-Maximilians-Universität, Institute for Stroke and Dementia Research, Munich, Germany,
4 3Karolinska Institutet, Laboratory of Molecular Neurobiology, Stockholm, Sweden,
5 Karolinska Institutet, Laboratory of Molecular Neurobiology, Stockholm, Sweden,
6 École Polytechnique Fédérale de Lausanne (EPFL), Laboratory of Neurodevelopmental Systems Biology, Lausanne, Switzerland

The Human Cell Atlas (HCA) is an international collaborative effort that aims to define all human cell types in terms of distinctive molecular profiles. In this HCA study, we conducted single-nuclei RNAseq analysis of healthy human post-mortem white matter of the central nervous system, with the aim of further understanding glial populations. The white matter tracts are rich in glia including oligodendrocytes and their oligodendrocyte precursor cells (OPCs), astrocytes and microglia.

We hypothesised that heterogeneity in white matter glia with region, age and/or sex leads to the variations observed in demyelinating and neurodegenerative diseases, where glia is commonly affected.

The most marked glial heterogeneity we found was among tissue regions (white matter from the primary motor cortex, cerebellum, and cervical spinal cord). We performed differential abundance analysis that revealed tissue-specific cell populations of OPCs, and differential expression indicated that they segregate as they maintain markers of their developmental origin (anterior: FOXG1; posterior: PAX3). We also detect region-specific astrocyte populations that are characterised by a clear separation in marker gene expression (anterior: ADGRV1, posterior: FAT3). Another statistically significantly enriched glial population includes a SPARC-expressing oligodendrocyte type that is more abundant in the human spinal cord. Integration with previously published mouse spinal cord data revealed that this population is human-specific. Finally, spinal cord microglia, but not astrocytes, show a more activated gene expression compared to the brain.
These regional effects combined with age and sex differences that we find will be a basis for understanding the human brain in health and also in diseases with an anatomical pattern, which is essential knowledge for therapeutic strategies. All the data, code used, and results will be accessible through an interactive application, following HCA commitment to open data, code, and community.

LMNA R482L mutation-specific impairments of skeletal muscle metabolism are associated with pathological accumulation of reactive oxygen species

Oksana Ivanova, Elena Ignatieva, Margarita Sorokina, Anna Kostareva, Renata Dmitrieva

Research Centre of Unknown, Rare and Genetically Determined Diseases, World-Class Research Centre for Personalized Medicine

Background: Familial partial lipodystrophy Dunnigan variety (FPLD) is a rare inherited severe metabolic disorder caused by mutations in the LMNA gene that leads to cardiovascular and skeletal muscle complications. FPLD is characterized by a decrease in the expression of cytochrome c oxidase IV and excessive production of reactive oxygen species (ROS). Muscle cells are sensitive to changes in redox homeostasis due to the large amount of ROS produced during muscle contraction. However, how LMNA mutations associated with FPLD affect the redox status and bioenergetics of skeletal muscles is still poorly understood.

Aim: To investigate the redox status and molecular mechanisms in skeletal muscles harbouring the LMNA-R482L mutation causing FPLD.

Methods: The C2C12 mouse myoblast cell line was transduced with the lentivirus-bearing human LMNA gene with mutation R482L and WT; myoblasts were differentiated into myotubes. The ROS levels were measured by cell-permeant 2',7'-dichlorodihydrofluorescein diacetate staining. Transcriptome sequencing was performed for C2C12 cells. Cell-type-specific gene expression was assessed using single-cell RNA-seq (10X Genomics).

Results: LMNA-R482L myoblasts show significantly higher levels of ROS, exhibiting oxidative stress. Comprehensive analysis of the transcriptome and metabolic networks of mutant myoblasts revealed a specific molecular module that could be responsible for the accumulation of ROS–glutathione metabolism: both branches of this pathway that regulate the balance of oxidized and reduced glutathione (GSH/GSSG) were upregulated. Analysis of single-cell RNA-seq data of myoblasts revealed 10 cell populations; 3 of them were mainly presented in LMNA-R482L cells. These clusters express genes associated with chondrogenesis, osteogenesis, and mineralization, which indicates pathological impairment of the myoblast differentiation program. These mutation-specific clusters are enriched with the ROS metabolic process and response to oxidative stress pathways.

Conclusion: We showed the pathological accumulation of ROS associated with specific cell subpopulations and metabolic modules resulting in redox imbalance in muscle cells bearing the LMNA-R482L mutation causing FPLD.

Somatic Mutations in Cancer and Normal Tissues

Abel González Pérez

IRB Barcelona, Spain

Somatic mutations occur in our tissues from fertilization onwards. The exposure to endogenous and external DNA damaging agents are responsible for different mutational processes across normal tissues. The study of these mutational processes thus increases our knowledge about the underlying mechanisms of DNA damage and the cellular processes charged with their repair. It also allows us to explore basic molecular biology processes through their interference with DNA repair, which results in specific mutational patterns. In our group, through the study of this interplay between DNA damage and repair in mutagenesis, we have revealed unexpected effects of proteins bound to DNA, such as transcription factors and histones on the aforementioned mutational patterns. The study of somatic mutations is also key to understanding the mechanisms of tumorigenesis across cancer types. In our group, we have also dedicated ourselves to develop methods to identify cancer driver genes and mutations. We also explore the translation of this knowledge into clinical practice.

NOTES

NOTES

NOTES

NOTES

NOTES

NOTES


MDPI