About

NSFC/RGC Workshop on Single-Cell Data Science

Emerging Technology and Statistical Challenges

Rapid recent advances in single-cell technology provide an unprecedented opportunity to understand cellular heterogeneity, genetic substructure, cell-cell interactions and temporal dynamics of complex tissues and tumours. The large volume and complexity of these high dimensional data pose unique statistical and computational challenges. This two-day NSFC/RGC workshop aims to bring together single-cell researchers in Hong Kong, mainland China and beyond to discuss emerging challenges and solutions in single-cell data science in an open and collaborative environment.

General Information

Venue: Via Zoom
Date: 1-2 June 2022
Website: https://singlecell2022.hku.hk/
Contact: singlecell2022@gmail.com

Organizing committee

Dr. Yuanhua Huang, HKU
Dr. Joshua Ho, HKU
Dr. Xi Chen, SUSTech
Dr. Zhixiang Lin, CUHK

Keynote Speakers

Large scale single-cell multi-sample multi-condition data integration

Prof. Jean Yang KS

University of Sydney, Sydney, Australia

Advances such as large-scale single-cell profiling such as single-cell RNA-seq (scRNA-seq) have exploded in recent years and enabled unprecedented insight into the identity and function of individual cells. The recent emergence of multiple individual cohort studies further allows researchers to investigate cells from different individuals. Effective integration of multiple collections cohort studies promises further biological insights into cells under different conditions that can not be uncovered with individual studies. Here, we present scMerge2, a scalable algorithm that allows data integration of large-scale multi-sample multi-condition single-cell studies. We have generalised scMerge2 to enable the merging of millions of cells from single-cell studies. Using a large COVID-19 data collection with over two million cells from over seven hundred individuals from multiple studies globally, we demonstrate that scMerge2 enables multi-sample multi-condition scRNA-seq data integration from multiple cohorts and reveals distinct cell states of COVID-19 patients with varying degrees of severity. We further illustrate the extraction of interpretable cellular representation (scFeatures) for each individual and the potential of these signatures for discriminating between moderate and severe patients.

From genotype to phenotype with single-cell resolution

Prof. Oliver Stegle KS

The German Cancer Research Center, Heidelberg, Germany
European Molecular Biology Laboratory

The study of genetic effects on gene expression and other molecular traits using bulk sequencing has allowed for the functional annotation of disease variants in diverse human tissues. Advances in single-cell RNA sequencing and multi-omics protocols provide for unprecedented opportunities to greatly increase the resolution of such genetic analyses, allowing to assess gene regulatory effects at the resolution of cell types, cell states and even in individual cells in human tissues. In this talk, I will present computational strategies for analyzing and integrating population-scale multi-omics dataset. I will then describe applications of these strategies to population-scale single-cell sequencing dataset from genetically diverse human iPSCs across differentiation towards a neuronal fate. Our data provide unprecedented opportunities to map regulatory variants and human disease variants both in major cell types but also in subtle subtypes and across cellular differentiation, revealing dynamic changes of regulatory variants. We describe novel disease-relevant linkages of several of the regulatory variants identified, thereby illustrating how this technology can open novel opportunities to study risk factors for human diseases.

A comparison of computational methods for selecting marker genes in single-cell RNA sequencing data

Dr. Davis McCarthy

St Vincent's Institute of Medical Research, University of Melbourne, Melbourne, Australia

A common step in the analysis of scRNA-seq data is the selection of so-called marker genes, most commonly to enable the annotation of the biological cell types present in the sample. We benchmarked 41 computational methods for selecting marker genes in scRNA-seq data. The performance of the methods was compared using 10 real scRNA-seq datasets and over 170 additional simulated datasets. The methods were compared based on their ability to recover simulated and expert-annotated marker genes, the predictive performance of the gene sets they select, their memory usage and speed, the quality of their implementations, and the characteristics of the marker genes they select. In addition, various case studies were used to highlight issues and inconsistencies in the most commonly used methods. Overall, we present a comprehensive evaluation of methods for selecting marker genes in scRNA-seq data. Our results highlight the efficacy of simple methods, especially the Wilcoxon rank-sum test, Student's t-test and logistic regression.

hECA: Human Ensemble Cell Atlas as a Virtual Body for *"In Data"* Cellular Experiments

Prof. Xuegong Zhang

Tsinghua University, Beijing, China

Profiling the molecular features of all cells with their anatomical and functional attributes is essential for understanding the human body in health and diseases. In recent years, scientists have been enthusiastic in building such atlases of human cells using single-cell omics technologies, led by consortiums and programs like HCA, HuBMAP and HDCA. In the meanwhile, the whole community has been more and more single-cell studies with the rapid development and popularization single-cell RNA-sequencing and other technologies. Tremendous amount of single-cell data are accumulating in the public domain. This suggests the possibility of an alternative approach for building cell atlases by assembling such "shot-gun" data in scattered publications. We have been studying this possibility and the informatics solutions in recent years, and realized that the key challenges in atlas assembly are not unique to scattered data, but also applies data generated by consortium efforts. The task of cell atlas assembly will be of "shot-gun" manner in nature as the spatial and temporal destiny of each cell is not deterministic at the microscopic level. The information complexity and volume are many magnitudes larger than that of the human genome project. We proposed a unified information framework for assembling atlases from data of various sources and built a human Ensemble Cell Atlas (hECA). It includes an infrastructure for storing and retrieving data that are large in both depth and width, and an information graph for unifying and representing multifaceted annotations of cells. We developed an *"in data"* cell sorting scheme that allows extracting cells using logic formula from the "virtual human body" to investigate scientific questions involving multiple organs and cell types.

Differential Inference for Single-cell RNA-sequencing Data

Dr. Yingying Wei

The Chinese University of Hong Kong, HKSAR, China

With the wide-adoption of single-cell RNA-seq (scRNA-seq) technologies, scRNA-seq experiments are becoming more and more complicated with multiple treatments or biological conditions. However, despite the active research on batch effects correction, cell type clustering, and missing data imputation for scRNA-seq data, rigorous statistical methods to compare scRNA-seq experiments under different conditions are still lacking. Here, we propose a Bayesian hierarchical model, DIFferential Inference for Single-cell RNA-sequencing Data (DIFseq), to rigorously quantify the treatment effects on both cellular compositions and cell-type-specific gene expression levels for scRNA-seq data. We derive conditions for the model identifiability, which provides guidelines on the experimental design for comparative scRNA-seq studies. We implement a highly scalable Monte Carlo Expectation-Maximization algorithm to handle the large number of cells. Application of DIFseq to a pancreatic study demonstrates that considering the biological conditions of samples in the analysis substantially boosts the clustering accuracy as compared to traditional analysis pipeline for scRNA-seq data and identifies cell-type-specific and condition-specific differentially expressed genes.

Cell clustering for spatial transcriptomics data with graph neural network

Dr. Ye Yuan

Department of Automation, Shanghai Jiaotong University, Shanghai, China

Spatial transcriptomics data can provide high-throughput gene expression profiling and spatial structure of tissues simultaneously. An essential question of its initial analysis is cell clustering. However, most existing studies rely on only gene expression information and cannot utilize spatial information efficiently. Taking advantages of two recent technical development, spatial transcriptomics and graph neural network, we thus introduce CCST, Cell Clustering for Spatial Transcriptomics data with graph neural network, an unsupervised cell clustering method based on graph convolutional network to improve ab initio cell clustering and discovering of novel sub cell types based on curated cell category annotation. CCST is a general framework for dealing with various kinds of spatially resolved transcriptomics. With application to five in vitro and in vivo spatial datasets, we show that CCST outperforms other spatial cluster approaches on spatial transcriptomics datasets, and can clearly identify all four cell cycle phases from MERFISH data of cultured cells, and find novel functional sub cell types with different micro-environments from seqFISH+ data of brain, which are all validated experimentally, inspiring novel biological hypotheses about the underlying interactions among cell state, cell type and micro-environment.

The Graphical R2D2 Estimator for the Biological Networks

Dr. Dora Zhang

The University of Hong Kong, HKSAR, China

Biological networks are important for the analysis of human diseases, which summarize the regulatory interactions and other relationships between different molecules. Understanding and constructing networks for molecules, such as DNA, RNA and proteins, can help elucidate the mechanisms of complex biological systems. The Gaussian Graphical Models (GGMs) are popular tools for the estimation of gene regulatory networks because of their biological interpretability. Nonetheless, reconstructing GGMs from high-dimensional datasets is still challenging and current methods cannot handle the sparsity and high-dimensionality issues arising from datasets very well. Here we will talk about a new GGM, called the graphical R2D2 (R2-induced Dirichlet Decomposition), based on the R2D2 priors for linear models. When the true precision matrix is sparse and of high dimension, the graphical R2D2 provides the estimates with smallest information divergence from the sampling model. We will also provide breast cancer gene network analysis example using the graphical R2D2 estimator.

Measuring Protein-DNA Interaction for Decoding Epigenome in Single Cells

Prof. Aibin He

Peking University, Beijing, China

Recent advances in single-cell epigenomic technologies are transforming our understanding of gene regulation. Here I will talk about our current progresses on developing single-cell methods for measuring protein-DNA interaction for decoding epigenome during development, including sc-itChIP-seq and CoBATECH, as well as multimodal omics in CoTECH. Further, a promising method, SoMUCH, to simultaneously profile multiple histone modifications in the same cell will be mentioned.

scONE-seq: A one-tube single-cell multi-omics method enables simultaneous dissection of molecular phenotype and genotype heterogeneity from frozen tumors

Prof. Angela Wu

The Hong Kong University of Science and Technology, HKSAR, China

Genomic and transcriptomic heterogeneity both play important roles in normal cellular function as well as in disease development. To be able to characterize these different forms of cellular heterogeneity in diverse sample types, we developed scONE-seq, which enables simultaneous transcriptome and genome profiling in a one-tube reaction. Previous single-cell-whole-genome-RNA-sequencing (scWGS-RNA-seq) methods require physical separation of DNA and RNA, often by physical separation of the nucleus from the cytoplasm. These methods are labor-intensive and technically demanding, time-consuming, or require special devices, and they are not applicable to frozen samples that cannot generate intact single-cell suspensions. scONE-seq is a one-tube reaction, thus is highly scalable and is the first scWGS-RNA-seq method compatible with frozen biobanked tissue. We benchmarked scONE-seq against existing methods using cell lines and lymphocytes from a healthy donor, and we applied it to a 2-year-frozen astrocytoma sample profiling over 1,200 nuclei, subsequently identifying a unique transcriptionally normal-like tumor clone. scONE-seq makes it possible to perform large-scale single-cell multi-omics interrogation with ease on the vast quantities of biobanked tissue, which could transform the scale of future multi-omics single-cell cancer profiling studies.

ISSAAC-seq enables sensitive and flexible multimodal profiling of chromatin accessibility and gene expression in single cells

Dr. Chen Xi

Southern University of Science And Technology, Shenzhen, China

Joint profiling of chromatin accessibility and gene expression from the same single cell/nucleus provides critical information about cell types in a tissue and cell states during a dynamic process. These emerging multi-omics techniques help the investigation of cell-type resolved gene regulatory mechanisms. However, many methods are currently limited by low sensitivity, low throughput or complex workflow. Here, we developed in situ SHERRY after ATAC-seq (ISSAAC-seq), a highly sensitive and flexible single cell multi-omics method to interrogate chromatin accessibility and gene expression from the same single nucleus. We demonstrated that ISSAAC-seq is sensitive and provides high quality data.

Inferring genetic models from cell landscapes

Dr. Guoji Guo

Zhejiang University, Hangzhou, China

Despite extensive efforts to sequence different genomes, genetic models to interpret gene regulation and cell fate decisions are lacking for most species. Here, we constructed cross-species cell landscapes covering representative metazoan species to study gene regulation through evolution. We developed a deep learning–based model, NvWA, to predict expression landscapes and decipher regulatory elements at the single-cell level. Although entirely derived from the genome sequences, NvWA recognizes global regulatory signals. We provide evidence that regulatory elements are more conserved among vertebrates and invertebrates than the gene expression.

Cellular and microbial niches of the human intestinal tract

Dr. Kylie James

Garvan Institute of Medical Research, Sydney, Australia

The human intestinal cell landscape is dynamic, changing in response to functional requirements and environmental exposures. Our recent work has comprehensively mapped cell lineages from distinct anatomical regions of the healthy human intestinal tract using single cell RNA-seq. Using a systems integrated analysis approach we have identified transcriptionally-distinct epithelial cells across the intestinal tract and an increasing gradient of plasma cell activation in response to a changing neighbouring microbiota. Together, this work provides an unprecedented catalogue of intestinal cells, and new insights into the complexity of cellular programs in gut homeostasis.

A human embryonic limb cell atlas resolved in space and time

Prof. Hongbo Zhang

Sun Yat-sen University, Guangzhou, China

Human limbs emerge during the fourth post-conception week as mesenchymal buds which develop into fully-formed limbs over the subsequent months. Limb development is orchestrated by numerous temporally and spatially restricted gene expression programmes, making congenital alterations in phenotype common. Decades of work with model organisms has outlined the fundamental processes underlying vertebrate limb development, but an in-depth characterisation of this process in humans has yet to be performed. Here we detail the development of the human embryonic limb across space and time, using both single-cell and spatial transcriptomics. We demonstrate extensive diversification of cells, progressing from a restricted number of multipotent progenitors to myriad mature cell states, and identify several novel cell populations, including perineural fibroblasts and multiple distinct mesenchymal states. We uncover two waves of human muscle development, each characterised by different cell states regulated by separate gene expression programmes. We identify musculin (MSC) as a key transcriptional repressor maintaining muscle stem cell identity and validate this by performing MSC knock down in human embryonic myoblasts, which results in significant upregulation of late myogenic genes. Spatially mapping the cell types of the limb across a range of gestational ages demonstrates a clear anatomical segregation between genes linked to brachydactyly and polysyndactyly, and uncovers two transcriptionally and spatially distinct populations of the progress zone, which we term “outer” and “transitional” layers. The latter exhibits a transcriptomic profile similar to that of the chondrocyte lineage, but lacking the key chondrogenic transcription factors SOX5,6 & 9. Finally, we perform scRNA-seq on murine embryonic limbs to facilitate cross-species developmental comparison at single-cell resolution, finding substantial homology between the two species.

Integrative Spatial Transcriptome Analysis for Embryo Development

Dr. Guangdun Peng

Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou, China

Single-cell sequencing has revolutionized biomedicine research in unprecedented scales. However current single-cell sequencing has been limited by the loss of spatial information. Spatial omics aim to survey the natural state of cells in native tissues, to identify the location-defined cell types and to understand how the cells are communicating within their community. The integrative analysis for spatial transcriptome and single-cells will allow for studying cellular heterogeneity at different scales and for discovering new layers of molecular connectivity between the genome and its functional output, and leading to many innovative discoveries. Here, we present a combination of experimental and computational pipelines in uncovering spatial variances for embryonic tissue organizations, particularly during gastrulation and mouse brain development.

From one to many: the making of multi-ciliated cells

Dr. Mu He

School of Biomedical Sciences, University of Hong Kong, HKSAR, China

The airway is a complex organ vital for all air breathing animals. Diseases of the respiratory system are amongst the leading causes of morbidity and mortality worldwide. The early manifestation of many airway diseases raises the question of the cellular origins for diseases during embryonic and fetal development. In our recent study, we systematically characterized the developmental landscape of the mouse airway using single-cell RNA sequencing and identified remarkably conserved cellular programs operating during human fetal development. In addition, trajectory analyses combined with lineage tracing experiments allowed us to uncover a unique cilia-secretory hybrid cell state highly relevant in development and disease.

Cells of the developing human lung

Dr. Peng He

EMBL-EBI, Sanger Institute, University of Cambridge, Cambridge, UK

We present a multiomic cell atlas of human lung development that combines single cell RNA and ATAC sequencing, high throughput spatial transcriptomics and single cell imaging. Coupling single cell methods with spatial analysis has allowed a comprehensive cellular survey of the epithelial, mesenchymal, endothelial and erythrocyte/leukocyte compartments from 5-22 post conception weeks. We identify new cell states in all compartments. These include developmental-specific secretory progenitors that resemble cells in adult fibrotic lungs and a new subtype of neuroendocrine cell related to human small cell lung cancer; observations which strengthen the connections between development and disease/regeneration. Finally, to illustrate its general utility, we use our cell atlas to generate predictions about cell-cell signalling and transcription factor hierarchies which we test using organoid models.

Sample demultiplexing, multiplet detection, cell-type classification and verification in large-scale single cell sequencing

Dr. Lian Qiuyu

Shanghai Jiaotong University, Shanghai, China

Identifying and removing multiplets and the multiplet-induced artificial cell types are essential to improving the scalability and the reliability of single cell RNA sequencing (scRNA-seq). Multiplets create artificial cell types in scRNA-seq datasets and complicate the automation of cell surface phenotyping. We propose a Gaussian mixture model-based multiplet identification method, GMM-Demux, and an artificial-cell-type aware surface marker clustering method, CITE-sort, to enable accurate identification and removal of multiplets in large-scale scRNA-seq datasets, as well as accurate cell type clustering by surface markers. Together, they accurately identify and remove multiplets and artificial cell types through joint RNA-and-protein sequencing (CITE-seq) and sample barcoding technologies (cell hashing or MULTI-seq). We benchmarked GMM-Demux and CITE-sort on two in-house cell-hashing-enabled CITE-seq datasets. We showed that GMM-Demux is more stable and accurate than existing state-of-the-art sample demultiplexing methods and can recognize 9 multiplet-induced fake cell types in a PBMC dataset. We demonstrated that CITE-sort produces the best clustering performance in large scale CITE-seq datasets with substantial multiplet footprints, when compared against a suite of canonical, multiplet-unaware clustering methods. Together, GMM-Demux and CITE-sort enable accurate identification and removal of multiplets and artificial cell types; hence increase the scale of scRNA-seq experiments from lower than 10K to beyond 30K cells.

FlowGrid: A python package for fast clustering for millions of single cell transcriptomic profiles

Ms. Fang Xiunan

University of Hong Kong, HKSAR, China

Recent technological advances have revolutionized the scale of single cell analyses where limited solutions exist. The number of reported cells in each single cell study increased from several hundreds to several millions. The increasing scale of scRNA-seq analysis urges the need for a scalable and reliable clustering algorithm that can handle single cell data set on the scale of millions. FlowGrid combines the benefit of DBSCAN (a density-based clustering algorithm for large spatial databases) and a grid-based approach to achieve scalability. The key idea of FlowGrid algorithm is to replace the calculation of density from individual points to discrete bins as defined by a uniform grid. This way, the clustering step of the algorithm will scale with the number of non-empty bins, which is significantly smaller than the number of points in lower dimensional data sets. FlowGrid package is implemented in python and made compatible with Scanpy platform with the functions including clustering, cluster result report and outlier detection.

UniTVelo: temporally unified RNA velocity reinforces single-cell trajectory inference

Mr. Gao Mingze

School of Biomedical Sciences, University of Hong Kong, HKSAR, China

The recent breakthrough of single-cell RNA velocity methods brings attractive promises to automatically identify directed trajectory on cell differentiation, states transition and response to perturbations, which is uniquely demanded in in-vivo applications and abnormal conditions. However, the existing RNA velocity methods, including scVelo, are often found to return erroneous results, partly due to model violation of complex expression profiles or lack of temporal regularization. Here, we present UniTVelo, a statistical framework of RNA velocity that models the flexible transcription dynamics of spliced and unspliced RNAs via a spliced RNA oriented framework. Uniquely, it also supports the effective inference of unified latent time across genes and orders cells on individual genes in the phase portrait, especially for multiple-rate kinetics genes and those with stable and monotonic changes across the transcriptome. With ten datasets, we demonstrate that UniTVelo returns the expected trajectory in different biological systems, including hematopoietic differentiation and those even with weak kinetics or complex branches. Specifically, UniTVelo correctly identifies the differentiation trajectories of the human bone marrow development, from hematopoietic stem cells to three distinct branches. This system is complex and cannot be fully resolved by other currently available RNA velocity methods.

Differential composition analysis of single-cell data

Ms. Lin Xinyi

School of Biomedical Sciences, University of Hong Kong, HKSAR, China

Single cell profiling technology such as single-cell RNA-seq (scRNA-seq) enables high throughput discovery and characterisation of diverse cell types or cell states in a population of cells. This ability has given rise to new statistical problems in robust quantification and comparison of cell-type proportion.. It remains challenging to effectively detect differential compositions of cell types when comparing samples coming from different conditions or along with continuous covariates, partly due to the small number of replicates and high uncertainty of cell clustering. Here, we introduce a new statistical model, DCATS, for differential composition analysis in single cells in a framework of beta-binomial regression. It leverages a confusion matrix to correct the bias of clustering and allows pre-estimated parameters across all cell types to account for its uncertainty. It also allows us to regress out the influence of confounding covariates except for the condition factor. Through multiple simulated and experimental data sets, we demonstrate the high effectiveness of DCATS in identifying variable cell types in various experiment designs. Combining the differential genes analysis, cell-cell interaction analysis, and other scRNA-seq analysis, DCATS deepens our understanding of cell types differential composition and gain biological insight.

Spatial-temporal transcriptomics analysis of Mesodermal Lineage Organoids (MLOs) reveals human developmental hematopoiesis

Mr. Xiang Yang

School of Biomedical Sciences, University of Hong Kong, HKSAR, China

Organoids are valuable tools in studying developmental biology and disease modeling. However, the lack of an organoid platform with vasculature and immune cells hinders the understanding of developmental hematopoiesis, the origin of the blood system in the embryo. Here we established mesodermal lineage organoids (MLOs) from human pluripotent stem cells by step-by-step exposure to morphogens and cytokines. We hypothesized that MLOs could recapitulate developmental hematopoiesis. To validate that the MLOs followed the early hematopoiesis consistent with human embryos, time-series single-cell RNA-seq combined with spatial transcriptomics revealed the signatures of early hematopoiesis. To be specific, erythro-megakaryocyte commitment occurred first, followed by monocyte commitment and expansion. These observations reinforced previous knowledge of early hematopoiesis in human embryos, demonstrating that MLOs recapitulate developmental hematopoiesis.

Single cell analysis reveals CHOP as an early marker for conventional chondrosarcoma

Dr. Su Nelson, Zezhuo

University of Hong Kong, HKSAR, China

Chondrosarcoma consisting of cartilage producing cells is a relatively rare cancer. Conventional chondrosarcoma (CCS) is the most common form of chondrosarcoma that can arise from benign tumour. Benign tumour is managed conservatively, while surgery is the main form of treatment for CCS. However, there is currently no clear diagnostic marker for malignant transformation and clinical decision making. We used single-cell RNA sequencing (scRNA-seq) to characterize primary tumour tissues from five patients with CCS, one patient with benign tumour, as well as a chondroblastic osteosarcoma and a fetal femur as controls. This atlas of 25,739 single cells were analysed to identify tumour heterogeneity, malignant transformation, and early markers of CCS. Based on inter-tumoral heterogeneity at single-cell resolution, we found that differentiated tumours contained for two distinct neoplastic cell clusters. One cluster resembled normal resting chondrocytes from foetal femur while the other cluster, which was subjected to cellular reprograming and ER stress, indicated malignant transformation. Further single cell analysis on additional central, peripheral, dedifferentiated chondrosarcomas revalidated ER stress as a specific early marker for central chondrosarcoma. In addition, we experimental revalidated the cancer development trajectory that NFkB pathway regulated the survival of chondrosarcoma cells through alleviating ER stress induced apoptosis. Importantly, coupled with bulk expression profiles of large cohort, we proposed a prognostic model and identified DNA Damage Inducible Transcript 3 (DDIT3), commonly known as CHOP which mediates ER stress induced apoptosis, as an early marker for CCS. Overall, these findings deciphered tumour heterogeneity and malignant transformation mechanisms and provided molecular signature for diagnosis of malignant transformation and grading of CCS.

Single-cell RNA sequencing of lung adenocarcinoma patients reveals age-dependent tumour microenvironment alterations that facilitate response to immunotherapy

Mr. Zhu Xiaoqiang

University of Hong Kong, HKSAR, China

Background: Immunotherapy has revolutionized the treatment of lung adenocarcinoma (LUAD). While the increase in cancer incidence with age and age-associated changes in immune function have been well characterized, age-related differences in the intratumoral immune populations and how this affects immunotherapy response remain unknown.
Methods: We collected clinical data from 3,544 samples treated with immune checkpoint blockade (ICB) at the pan-cancer level. Single-cell RNA sequencing data from 19 LUAD samples were integrated and analyzed. Cell types were identified based on canonical marker gene expression.
Results: Aged individuals responded more efficiently to ICB treatment at the pan-cancer level. By focusing on LUAD, we showed that patients aged above 46 responded better and had longer progression-free survival than younger counterparts. Further, we presented a 58,031-cell catalog of age-associated single-cell transcriptomes from LUAD patients. Our analysis revealed the accumulation of immune response and chronic inflammation with increasing age. Aged tissues (> 60 years) displayed dominant exhausted natural killer (NK)/T cells and T regulatory cells and were enriched with two subtypes of tumour-associated macrophages marked by SPP1 and CX3CR1 expression, respectively. Furthermore, the ageing dynamics of stroma cells created a pro-tumor microenvironment reflected by increased angiogenesis and collagen formation.
Conclusions: These results revealed an increasing immune-suppressive tumour microenvironment with age, supporting an observation of improved ICB response in older LUAD patients. Our extensive single-cell analysis enhances the understanding of age-dependent molecular and cellular dynamics in LUAD. Our results highlight the importance of considering age as a factor for ICB response and provide a rich resource for developing therapeutic targets in LUAD-microenvironment interactions.

Investigation of Thymic Nursing Cell Complexes by Single-Cell RNA Sequencing

Mr. Velayutham Sampath Kannan

Department of Pharmacology and Pharmacy, University of Hong Kong, HKSAR, China

Thymic nurse cell complexes (TNC) provide important microenvironment for T-cell development in thymus. However, there is little information on the composition of cells within TNC. Here, TNC were isolated from thymus of 7-weeks old female wild type (WT) and adiponectin knock out (AKO) mice. Single-cell RNA sequencing was performed using the NovaSeq 6000 system in Centre for PanorOmic Sciences . Cell Ranger, Seurat, SingleR and ClusterMap algorithm were applied for identifying and comparing the different types of cells based on Cell Marker and cellkb databases. The results demonstrated that the cellular composition of TNC was significantly different between WT and AKO. The latter contained significantly reduced percentages of T lymphocytes, epithelial, macrophage and dendritic cells, but increased number of B cell progenitors and gamma-delta T cells, suggesting a role of adiponectin in TNC-mediated T cell development and selection.

Thursday 2nd June

Population genetics meets cellular genomics

Prof. Joseph Powell

Garvan Institute of Medical Research, Sydney, Australia
University of New South Wales, Sydney, Australia

Common diseases are the most prevalent cause of disease burden. These diseases, which are sometimes referred to as complex diseases, are characterised by the underlying genetic architecture that is highly polygenetic. There may be hundreds to thousands of independent risk loci for a given disease. An essential feature of these diseases is that genetic loci predominately act by changing how the genome is 'regulated' rather than changing protein sequences. A challenge in understanding the genetic mechanism underlying disease is that while DNA is the same in every cell, how genetic variation contributes to genome regulation is not the same in every cell. There are significant 'cell-type' specific effects. Our research is focused on studying this phenomenon at scale, utilising cellular genomics, cellular phenotyping, and population genetics.

CoSpar identifies early cell fate biases from single cell transcriptomic and lineage information

Prof. Shuowen Wang

Westlake University, Hangzhou, China
Harvard University, Massachusetts, USA

A goal of single cell genome-wide profiling is to reconstruct dynamic transitions during cell differentiation, disease onset, and drug response. Single cell assays have recently been integrated with lineage tracing, a set of methods that identify cells of common ancestry to establish bona fide dynamic relationships between cell states. These integrated methods have revealed unappreciated cell dynamics, but their analysis faces recurrent challenges arising from noisy, dispersed lineage data. Here, we develop coherent, sparse optimization (CoSpar) as a robust computational approach to infer cell dynamics from single-cell transcriptomics integrated with lineage tracing. Built on assumptions of coherence and sparsity of transition maps, CoSpar is robust to severe down-sampling and dispersion of lineage data, which enables simpler experimental designs and requires less calibration. In datasets representing hematopoiesis, reprogramming, and directed differentiation, CoSpar identifies early fate biases not previously detected, predicting transcription factors and receptors implicated in fate choice. Documentation and detailed examples for common experimental designs are available at https://cospar.readthedocs.io/.

Adversarial domain translation networks for integrating large-scale atlas-level single-cell datasets

Prof. Can Yang

Hong Kong University of Science & Technology, HKSAR, China

The rapid emergence of large-scale atlas-level single-cell RNA-seq datasets presents remarkable opportunities for broad and deep biological investigations through integrative analyses. However, harmonizing such datasets requires integration approaches to be not only computationally scalable, but also capable of preserving a wide range of fine-grained cell populations. We created Portal, a unified framework of adversarial domain translation to learn harmonized representations of datasets. When compared to other state-of-the-art methods, Portal achieves better performance for preserving biological variation during integration, while achieving the integration of millions of cells in minutes with low memory consumption. We show that Portal is widely applicable to integrating datasets across different samples, platforms and data types. We also apply Portal to the integration of cross-species datasets with limited shared information among them, elucidating biological insights into the similarities and divergences in the spermatogenesis process among mouse, macaque and human.

Profiling of transcribed cis-regulatory elements in single cells

Dr. Chung-chau Hon

RIKEN IMS Center for Integrative Medical Sciences, Japan

Profiling of cis-regulatory elements (CREs, mostly promoters and enhancers) in single cells allows the interrogation of the cell-type and cell-state-specific contexts of gene regulation and genetic predisposition to diseases. Here we demonstrate single-cell RNA-5'end-sequencing (sc-end5-seq) methods can detect transcribed CREs (tCREs), enabling simultaneous quantification of gene expression and enhancer activities in a single assay at no extra cost. We showed enhancer RNAs can be detected using sc-end5-seq methods with either random or oligo(dT) priming. To analyze tCREs in single cells, we developed SCAFE (Single Cell Analysis of Five-prime Ends) to identify genuine tCREs and analyze their activities (`https://github.com/chung-lab/scafe`). As compared to accessible CRE (aCRE, based on chromatin accessibility), tCREs are more accurate in predicting CRE interactions by co-activity, more sensitive in detecting shifts in alternative promoter usage and more enriched in diseases heritability. Our results highlight additional dimensions within sc-end5-seq data which can be used for interrogating gene regulation and disease heritability.

Tracing T cell development and T cell activation using single cell RNA-seq

Dr. Wenfei Jin

Southern University of science and Technology, Shenzhen, China

T cells are lymphocyte immune cells that protect us from pathogens and cancer cells. Tracing its development and activation could significantly facilitate our understanding of its function. We analyzed single-cell RNA-seq data of human CD34+ cells that represent hematopoietic stem and progenitor cells (HSPCs), from multiple tissues. We found most HSPCs subsets in bone marrow have counterparts in peripheral blood except B cell lineages, indicating that the HSPCs migration between bone marrow and blood might be very dynamics. We further interrogated 6 bead-enriched T cell subsets from peripheral blood and re-clustered them into 9 single cell clustered population (scCPops) using scRNA-seq. We found that Naïve T showed the highest similarity to its scCPop counterpart among all T cell subsets. Interestingly, we discovered a T cell subpopulation that highly expressed Interferon Signaling Associated Genes (ISAGs), which is named ISAGhi T and may contribute to quick T cell activation. After T cell activation, There is a fraction of stimulated T cells expressing some activated T cell markers, but their expression profile was similar to that of resting T cells, and we called them inert T cells. Compared with resting T cells, inert T cells showed increased activity, cell proliferation, and cytokine expression.

Identification and characterization of pathogen-specific T cells in paired single-cell RNA and TCR sequencing data

Dr. David Shih

University of Hong Kong, HKSAR, China

Investigating the phenotypic profiles of pathogen-specific T cells is critical to understanding T cell responses against pathogens as well as improving the efficacy of therapeutics and vaccines. However, current methodologies for identifying pathogen-reactive T cells are limited in scope, throughput, or specificity. We have developed an integrative approach to identify pathogen-specific T cells in blood samples and characterize their single-cell transcriptomes. Our approach involves first identifying pathogen-specific T cells by modeling the temporal expansion trajectories in longitudinal bulk TCR-seq data, and then using TCR sequences as barcodes to label the identified pathogen-specific T cells in the matching single-cell data. Applying our approach to a clinical study of an experimental vaccine against human cytomegalovirus, we are able to characterize the single-cell transcriptomes of vaccine-specific T cells and discover transcriptional signatures of transient and durable T cell response to cytomegalovirus. Accordingly, our methodology can greatly facilitate the study of T cell responses to vaccines and pathogens.

Tumor-associated monocytes promote glioma progression via EGFR signaling

Prof. Jiguang Wang

Hong Kong University of Science & Technology, HKSAR, China

TBD

Microfluidics for single-cell printing and live-cell elasticity measurement

Dr. Huaying Chen

Harbin Institute of Technology, Shenzhen, China

Microfluidic chips are of unique advantages in single-cell manipulation and have been extensively employed in single-cell analysis. This article mainly introduces single-cell printing and elasticity measurement using microfluidics. A microfluidic chip integrating two pneumatic microvalves was developed to print single cells with the ability of dynamic size screening. The microvalves were designed to regulate the clearance of the flow channel by air pressure, and then achieve the size selection of single cells. The front and rear valves respectively defined the upper and lower limits of the cell size. Individual cells that meet the size requirements were quickly printed into a 384-well plate to achieve one cell per well. The viability of cells after printing was 97.2%. The printing process did not affect cell viability in comparison to the control group without size selection or printing.Besides, our team has also developed a microfluidic chip that integrates particle separation and pressure sensors for the precise measurement of live-cell elasticity. The chip can effectively separate the large and long particles in the cell suspension to avoid the blockage of the cell deformation channel. When a cell was driven into the microchannel, it deformed and finally pass through. Meanwhile, the protrusion of cells and pressure inducing cell deformation were recorded simultaneously. Finally, the elastic modulus and viscosity of each cell can be calculated using the cell protrusion length and the corresponding pressure in combination with a power-law rheological model. The elasticity of K562 and human umbilical vein cells were measured as 64.2 ± 33.3 Pa and 383.4 ± 226.7 Pa, respectively. The microfluidic chips introduced in this talk will be of significant application in stem cell studies.

Massive single-cell image-based profiling and analytics: Expect the unexpected

Prof. Kevin KM Tsia

Department of Electrical & Electronic Engineering, University of Hong Kong, HKSAR, China

Recent advances in optical microscopy have revolutionized our ability to visualize multifaceted morphological signatures (biophysical & molecular) of cells without the onerous and destructive sample handling commonly used in molecular profiling. This talk will cover our latest developments in combining ultrahigh-throughput microfluidic single-cell imaging (at a throughput up to 100,000 cells/sec) with different supervised/unsupervised image analytic strategies for crafting large-scale morphological “fingerprints” (profiles) of single cells. The talk will further discuss the new opportunities and challenges of how to integrate this rapidly growing image-based repertoire with the single-cell multi-omics data - altogether formulating holistic and integrative single-cell analysis strategies for dissecting the complex biological system.

Live single-cell imaging reveals a polyploid tumour cell subset upon bidirectional tumour-macrophage interaction

Prof. Alice Wong

University of Hong Kong, HKSAR, China

Macrophages closely associate with cancer progression and promote metastatic spread However, how macrophages interact with and are functionally altered by different cancer cell subsets in a heterogeneous cancer cell population is unknown. Here, using both an isogenic ovarian cancer cell line model with opposite metastatic abilities and humanized mice, we show that highly metastatic (HM) cells, but not non-metastatic (NM) cells, have a selective advantage in skewing the interacting macrophages to tumour-associated macrophages (TAM). Live single-cell imaging reveals that these TAMs subsequently induce failed cytokinesis in a significant subset of HM, leading to the formation of polyploid tumour cells that are pivotal for tumour aggressiveness. This bidirectional interaction between HM and macrophages are mediated by metadherin expression induced by β-catenin on HM via trans-acting CEACAM1 expressed on macrophages. Using the orthotopic humanized mice model, we confirm that knockdown of β-catenin or metadherin in HM cells reduced tumour burden and the number of tumor-infiltrating CD163+ TAMs. Expression of β-catenin is also found to be positively correlate with expression of metadherin in clinical samples. The induced polyploid metastatic subset represents a previously underappreciated therapeutic target.

Histology image-based spatial characterizations of tumor microenvironment for cancer diagnosis, prognosis and molecular subtyping

Dr. Xin Wang

Chinese University of Hong Kong, HKSAR, China

Histopathological images derived from tissue slides contain rich information about cell morphologies and tissue structures, providing a cost-efficient and easily accessible tool to dissect the tumor microenvironment for clinical decision-making. Recently, deep learning based on convolutional neural networks (CNNs) has been widely adopted as a powerful tool to delineate the cell morphologies, tissue architecture, and develop novel image-based biomarkers. More strikingly, H&E-stained histology images can be used to directly predict the status of molecular characteristics such as gene mutations, microsatellite instability, and even molecular subtypes. In this talk, I will introduce deep learning frameworks we recently developed for automated tissue classification of routine H&E stained whole-slide images. We employed a multi-scale quantification approach to calculate spatial organization features (SOFs) at different magnification levels, resulting in "tissue-omics" profiles including hundreds of morphometric and subvisual morphometric features. Using several case studies, I will demonstrate the clinical relevance of SOFs and their potential applications in cancer diagnosis, prognosis, and molecular subtyping.

Decitabine response in acute myeloid leukemia

Dr. Asif Javed

University of Hong Kong, HKSAR, China

Decitabine is a hypomethylation agent which can induce strong albeit transient remission in myeloid malignancies. Patient response to decitabine varies considerably with mutations in P53 gene being a key positive prognostic indicator. Over time, even the responsive patients acquire resistance to the treatment. The mechanism of de novo or acquired decitabine resistance remains elusive. We leverage the power of single cell omics to analyze the ebbs and flows of decitabine treatment. To this end, treatment sensitive and resistant patients in a decitabine clinical trial were recruited. Serial samples from these patients at prognostically relevant timepoints were evaluated. Single cell RNA sequencing analysis guided by clinical and genomic features highlighted molecular characteristics of drug response in leukemic and immune compartments. Through this analysis we aim to identify early prognostic markers of decitabine treatment and unearth possibilities for a more long term remission.

Multiomic single-cell analysis of the differentiation trajectories of acute myeloid leukemia

Prof. Feng Liu

Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China

Cancer differentiation therapy aims to induce the maturation of neoplastic cells, yet it is unclear how cell fate is determined in an oncogenic cellular context. Recently, we have applied multiomic single-cell analysis the differentiation trajectories of acute myeloid leukemia in vitro. We show that, under the induction of all-trans retinoid acid (ATRA), the NB-4 (PML-RARA+ APL/AML-M3-subtype cell line) cells navigate through a cell fate bifurcation point, with only one branch leading to mature granulocytes. Co-cis-accessibility network analyses, coupled with CRISPR- and CRISPRi-genome editing experiments, indicate that SPI1 and CEBPE, two hematopoietic lineage-determining transcription factors, are directly activated by ATRA signaling via discrete PML-RARA-targeting enhancers to promote terminal granulopoiesis. By contrast, in HL-60 cells, a cell line with AML-M2-subtype features, ATRA fails to activate strong SPI1 and CEBPE expression, and ATRA-induced differentiation is not only incomplete but also promiscuous, which is characterized by weak coin-duction of both granulopoiesis and lymphopoiesis gene expression programs. Together, our study uncovers significant heterogeneity in the differentiation trajectories of AML cells and suggests that the therapeutic efficacy of ATRA in certain AML-subtypes may be compromised by therapy-induced lineage promiscuity.

Mutational landscapes to identify the origin of cancers at single cell resolution

Dr. Jason Wong

School of Biomedical Sciences, University of Hong Kong, HKSAR, China

One of the major determinants of cancer cell phenotype is its cell type of origin. Recently, it was shown that the somatic mutational landscape of cancers could be used to infer cell type of origin as chromatin structure has a significant influence on the distribution of somatic mutations across cancer genomes. In this presentation, I will describe our efforts to integrate single-cell RNA-seq data and cancer somatic mutations to infer the cell type of origin of hepatocellular carcinoma and other cancers.

Clustering single-cell RNA sequencing data using copy number alterations without an healthy reference

Mr. Salvatore Milite

Human Technopole, Milano, Italy

Cancers are constituted by heterogeneous populations of cells that show complex genotypes and phenotypes which we can read out by sequencing. Many attempts at deciphering the clonal process that drives these populations are focusing on single-cell technologies to resolve genetic and phenotypic intra-tumor heterogeneity. While the ideal technologies for these investigations are multi-omics assays, unfortunately, these types of data are still too expensive and have limited scalability. We can resort to single-molecule assays, which are cheaper and scalable, and statistically emulate a joint assay, only if we can integrate measurements collected from independent cells of the same sample. In this work we follow this intuition and construct a new Bayesian method to genotype copy number alterations on single-cell RNA sequencing data, therefore integrating DNA and RNA measurements. Our method is unsupervised and leverages a segmentation of the input DNA to determine the sample subclonal composition at the copy number level, together with clone-specific phenotypes defined from RNA counts. By design our probabilistic method works without a reference RNA expression profile, and therefore can be applied in cases where this information may not be accessible. We implement and test our model on both simulated and real data, showing its ability to determine copy number associated clones and their RNA phenotypes in tumour data from 10x and Smart-Seq assays.

Single cell mitochondrial DNA mutations enable clonal tracking in osteosarcoma

Dr. Xue Yan Sharon

University of Hong Kong, HKSAR, China

Cancer progression from early oncogenic transformation to distant metastasis is driven by a process of clonal evolution. Reconstructing clonal dynamics is critical to understand cancer biology and implement effective therapies. MAESTER (Mitochondrial Alteration Enrichment from Single-cell Transcriptomes to Establish Relatedness) has been shown as a powerful technology to resolve clonal population through mitochondrial variant enrichment. Here, we explored the osteosarcoma clonality by adapting MAESTER in high-throughput single-cell RNA sequencing. The results showed that osteosarcoma mitochondrial transcripts were successfully enriched from full length cDNA barcoded scRNA-seq libraries generated by the 10X Genomics platform. The higher and more even coverage of mitochondrial transcript reads significantly enhanced the ability to detect clone-specific somatic mtDNA mutations at the single cell level, thus increased the feasibility to infer clonal dynamics in osteosarcoma progression. In conclusion, we have established the platform to enable discoveries in cancer biology by expanding the use of naturally occurring barcodes created by mtDNA alterations.

Single-cell computational analysis of CRISPR/Cas9-based cellular lineage tracing system

Ms. Chao Yiming

School of Biomedical Sciences, University of Hong Kong, HKSAR, China

The fundamental question in developmental biology is when and how the cells acquire a functional state. Single-cell lineage tracing reveals the stage of cells accruing functionality. The recently established molecular barcoding systems generated robust datasets of lineage tracing. However, the computational pipeline to simultaneously annotate cell-type and lineage trajectory has been challenging. Here, we analyzed one of the molecular barcoding systems, the CARLIN (CRISPR Array Repair LINeage tracing) system established in 2020, and developed a computational pipeline to deconvolute lineage tracing of murine hematopoiesis in vivo. Based on CellRanger for sequencing reads pre-processing and Seurat package for normal downstream analysis to annotate cell-type, we developed a new computational pipeline for sequence alignment, indel calling, and barcode genotyping. The barcode information defines cell clonality and lineage trajectory. Our pipeline unbiasedly identified and clustered the CARLIN barcodes, revealing the multi-lineage contribution of hematopoietic stem cells. In summary, we have developed a computational analysis pipeline at a single-cell level with user-friendly tools. We foresee that the new computational pipeline will address fundamental mechanisms in functional cell type specification.

XClone: Statistical modelling of copy number variations in single cells

Ms. Huang Rongting

School of Biomedical Sciences, University of Hong Kong, HKSAR, China

Somatic copy number variation (CNVs) are major mutations in various cancers for their development and clonal evolution. Analysing CNV in single-cell RNA-seq data is of critical importance for both detecting the CNV states in tumour cells and revealing its impact on transcriptional phenotypes. However, the intrinsic low coverage and high noise properties in scRNA-seq make it difficult to call the CNVs accurately. A few computational methods (inferCNV, CopyKAT, HoneyBADGER, CaSpER) have been recently proposed to analyse CNV from scRNA-seq data, but their accuracy and computational efficiency have not been well benchmarked. Here we present a statistical method, XClone, that integrates expression levels and allelic balance to enhance the detection of haplotype-aware CNVs from scRNA-seq data and the reconstruction of tumour clonal phylogeny. Compared to commonly used methods, XClone is found to be a promising tool for accurate CNV analysis across multiple data sets, including a well-characterized and verified gastric cancer sample that covers copy loss, gain, and loss of heterozygosity.

MQuad enables clonal substructure discovery using single cell mitochondrial variants

Mr. Kwok Aaron Wing Cheung

School of Biomedical Sciences, University of Hong Kong, HKSAR, China

Mitochondrial mutations are increasingly recognised as informative endogenous genetic markers that can be used to reconstruct cellular clonal structure using single-cell RNA or DNA sequencing data. However, identifying informative mtDNA variants in noisy and sparse single-cell sequencing data is still challenging with few computation methods available. Here we present an open source computational tool MQuad that accurately calls clonally informative mtDNA variants in a population of single cells, and an analysis suite for complete clonality inference, based on single cell RNA, DNA or ATAC sequencing data. Through a variety of simulated and experimental single cell sequencing data, we showed that MQuad can identify mitochondrial variants with both high sensitivity and specificity, outperforming existing methods by a large extent. Furthermore, we demonstrate its wide applicability in different single cell sequencing protocols, particularly in complementing single-nucleotide and copy-number variations to extract finer clonal resolution.