4th Challenges in Computational Biology meeting: Single Cell Data Analysis

Mainz, Germany, 18 September 2020

A virtual satellite meeting of the German Conference on Bioinformatics

BOOK OF ABSTRACTS

Organizing committee:
Miguel Andrade, JGU Mainz, andrade@uni-mainz.de
Ina Koch, Goethe University Frankfurt, ina.koch@bioinformatik.uni-frankfurt.de
Jean-Fred Fontaine, JGU Mainz, fontaine@uni-mainz.de
Pablo Mier, JGU Mainz, munoz@uni-mainz.de

Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM.

Luca Pinello.
Harvard University.

Abstract:
Single-cell transcriptomic assays have enabled the de novo reconstruction of lineage differentiation trajectories, along with the characterization of cellular heterogeneity and state transitions. Several methods have been developed for reconstructing developmental trajectories from single-cell transcriptomic data, but efforts on analyzing and visualizing single-cell epigenomic and proteomic data remain limited. Here we present STREAM, an interactive pipeline capable of disentangling and visualizing complex branching trajectories from both single-cell transcriptomic, epigenomic and proteomic data.

Molecular views into cellular functions by in-cell cryo-electron tomography.

Julia Mahamid.
EMBL-Heidelberg.

Abstract:
Most structural biology focuses on the structure and function of individual macromolecular complexes, but falls short of revealing how they come together to give rise to cellular functions. Here, cryo-electron tomography (cryo-ET) provides a unique opportunity for obtaining in situ structural information across a wide range of scales - from whole cells to individual macromolecules. Recent methodological developments in cryo-ET now allow assignment of molecular structures directly from three-dimensional stills of intact cells and reveal their molecular sociology. Using the genome-reduced human pathogen *Mycoplasma pneumoniae* as a model system, we demonstrate the potential synergistic application of whole-cell crosslinking mass spectrometry, cellular cryo-electron tomography, and integrative modelling, and determine an in-cell architecture of a transient transcribing and translating expressome at sub-nanometer resolution. Computational breakthroughs further allow visualization of small molecule antibiotics bound to their active cite within the intact pathogen. These findings highlight the enormous discovery potential of structural cell biology at the single cell level.

Reprogramming cell identity in the nervous system: what single-cell transcriptomics can teach us.

Benedikt Berninger.
King's College.

Abstract:
Reprogramming cell identity across cell lineages emerges as a novel strategy to repair organs with little intrinsic self-repair. This is of particular interest in the case of the brain, where highly vulnerable neurons are embedded in a tissue containing various types of glial cells. Recent work has shown that these glial cells can be converted into induced neurons by forced expression of neurogenic transcription factors such as the proneural factors Neurog2 and Ascl1. Single-cell technologies such as single-cell RNA sequencing allow shedding unprecedented light on the molecular and cellular processes that underlie this fascinating process of cellular metamorphosis. They have revealed that the process of glia-to-neuron conversion is anything but straightforward and harbours exciting surprises which in turn may offer molecular handles of further improving conversion efficiency and accuracy.

Asynchronous cellular aging revealed by machine learning based classification of single-cell data.

Sumeet Pal Singh.
Université Libre de Bruxelles.

Abstract:
Aging is a fact of life. Though aging is universal, the rate of aging is personal. Moreover, the rate of aging is asynchronous between the different organs of an individual. For instance, a person who smokes might have ‘older’ lungs, but ‘normal’ liver. Can we measure the biological age of individual cells and hence predict the health status of organs?

The techniques for determining cellular age are limited and rely on a limited set of histological markers and lack predictive power. Models that can quantify and predict cellular age would provide the ability to detect premature aging, providing a window for preventive therapies against age-related diseases.

Here, we implement GERAS (GEnetic Reference for Age of Single-cell), a machine learning based framework capable of assigning individual cells to chronological stages based on their transcriptomes. GERAS displays greater than 90% accuracy in classifying the chronological stage of zebrafish and human pancreatic cells. The framework demonstrates robustness against biological and technical noise, as evaluated by its performance on independent samplings of single-cells. Additionally, GERAS determines the impact of differences in calorie intake and BMI on the aging of zebrafish and human pancreatic cells, respectively.

We further harness the classification ability of GERAS to identify molecular factors that are potentially associated with the aging of beta-cells. We show that one of these factors, junba, is necessary to maintain the proliferative state of juvenile beta-cells. Our results showcase the applicability of a machine learning framework to classify the chronological stage of heterogeneous cell populations, while enabling detection of candidate genes associated with aging.

Cell, cell type and cell line handling in a database of pluripotent stem cells.

Nancy Mah, Stefanie Seltmann, Andreas Kurtz.
Charité - Universitätsmedizin Berlin.

Abstract:
Pluripotent stem cells (PSC) can differentiate into all cell types of the body. However, their identity is often difficult to ascertain as hPSC - derived cells frequently maintain immature features. To determine whether PSC-derived cell types represent states of natural developmental processes and plasticity, or are cell culture derived novel cell type with novel phenotypes remains a challenge. In addition to that, different PSC-lines are each obtained from a different individual, and when cultivated in vitro diversify genetically. The summary authentication as pluripotent stem cells is based on a set of minimal phenotypic characteristics. One of the challenges to define a cell type to PSC and their derivatives is the lack of a comprehensive and generally accepted cell type model, which can be used to generate a matrix of naturally and perhaps also artificially occurring cell types. A prerequisite of such a model would be the assignment of unique identifiers for cell types. The human pluripotent stem cell registry (hPSCreg) has developed unique identifiers for human PSC - lines and aims to extend this towards other cells. To that end, the International Cell Type Authentication Committee (ICTAC) was formed to develop a list and identifier of all known human cell types.

It takes two to tango: exploration and interpretation of RNA-seq data, made interactive and reproducible

Federico Marini.
Center for Thrombosis and Hemostasis Mainz (CTH), Mainz, Germany and
Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI), Mainz, Germany.

Abstract:
The exploration and the interpretation of RNA-seq transcriptome data can be complex tasks, yet these are essential to better generate hypotheses and communicate results.

Making these workflows accessible to a wide spectrum of researchers is key to empower many users in extracting knowledge out of large datasets, while still keeping up to the best practices for guaranteeing reproducibility of the analyses.

I will illustrate these concepts by describing the fundamental aspects of two R/Bioconductor packages I developed, iSEE (and its companion iSEEu to extend this exploration framework) (https://bioconductor.org/packages/iSEE/ &
https://bioconductor.org/packages/iSEEu/) and GeneTonic
(https://bioconductor.org/packages/GeneTonic/), showcasing their features and how these are integrated in the user interfaces for effective usage.
Interactivity can play an essential role in simplifying the way how one accesses and digests RNA-seq data analysis in a comprehensive way, and the exposure to the code that generates the desired output can have the pleasant side effect of learn-by-doing, also leading to a better appreciation of the final data product, to be further shared and extended.

Low amount ChIP-Seq using tagmentation assisted fragmentation of chromatin.

Junaid Akhtar.
Johannes Gutenberg University, Mainz.

Abstract:
The conventional approach of ChIP-Seq is not amenable for limited number of cells, often being the major impediment in the study of small cell populations. Here, we present a Tagmentation-Assisted Fragmentation ChIP (TAF-ChIP) and sequencing method to generate high-quality histone profiles from low cell numbers. Owing to the high signal to noise ratio, the data obtained from the TAF-ChIP can be subjected to standard bioinformatic data analysis workflow. The epigenetic profiles obtained from TAF-ChIP approach showed high degree of agreement with conventional ChIP-Seq datasets, highlighting the utility of this approach.

Modelling cellular signalling variability based on single-cell data: the TGFb/SMAD signaling pathway.

Stefan Legewie.
Institute of Molecular Biology, Mainz.

Abstract:
Non-genetic heterogeneity is key to cellular decisions, as even genetically identical cells respond in very different ways to the same external stimulus, e.g., during cell differentiation or therapeutic treatment of disease. Strong heterogeneity is typically already observed at the level of signaling pathways that are the first sensors of external inputs and transmit information to the nucleus where decisions are made. Since heterogeneity arises from random fluctuations of cellular components, mathematical models are required to fully understand the dynamics of heterogeneous cell populations. In my talk, I will discuss imaging-based modeling approaches of cellular signaling heterogeneity, with special focus on the TGFb/SMAD signaling pathway which plays a key role in tissue homeostasis.

Single cell-based multiscale modelling in stem cell research.

Antonio del Sol.
Université du Luxembourg.

Abstract:
A number of challenges in stem cell research can be addressed with the development of multiscale computational models. Indeed, with the increasing amount of available single-cell data, especially scRNA-seq data, we are now in the position of developing computational models at different levels of complexity, including intracellular and cell–cell communication network-based models. Here, I will present examples of recently developed single cell-based computational models at different scales of biological organization in order to address a variety of scientific questions in the stem cell field and to guide experimentalists in the design of new strategies for stem cell therapies and treatment.

Single-cell analysis revealed the factors controlling specification and differentiation of the gastric epithelial progenitors.

Natalia Soshnikova.
Institute for Molecular Medicine,
University Medical Center of the Johannes Gutenberg University, Mainz

Abstract:
Adult gastric stem cells are responsible for continuous renewal of the stomach epithelium in homeostasis. Gastric stem cells are heterogeneous in their molecular markers and proliferative capacity. Whether the stem cells come from common or molecularly and functionally distinct embryonic progenitors was unknown. We used single-cell RNA sequencing, genetic lineage tracing and organoid assays to characterise gastric progenitors in the developing mouse stomach. scRNA-sequencing predicted a role for several signalling pathways in cell fate determination of the gastric epithelial progenitors. By inhibiting the BMP, Notch and WNT pathways in vivo and ex vivo, we found that WNT and Notch signals are necessary for the maintenance of the corpus identity, whereas BMP signals are sufficient to promote the antrum identity of the embryonic gastric epithelium. Moreover, WNT signals are required for the differentiation of the embryonic epithelium along enzyme-producing zymogenic cell lineage, whereas BMP signals promote the differentiation of the gastric epithelium towards acid-producing parietal cells. In contrast, both signals block differentiation of the embryonic progenitors along mucous-producing pit lineage. Finally, inhibition of Notch signalling leads to differentiation of the embryonic gastric progenitors along all secretory lineages.

Analysis and integration of single cell epigenomics data.

Maria Colomé-Tatché.
Technical University of Munich.

Abstract:
In this talk I will present epiScanpy, a computational framework for the analysis of single-cell DNA methylation and single-cell ATAC-seq data. EpiScanpy makes the many existing RNA-seq workflows from scanpy available to large-scale single-cell data from other -omics modalities. I will introduce and compare multiple feature space constructions for epigenetic data and show the feasibility of common clustering, dimension reduction and trajectory learning techniques for both single-cell DNA methylation data and scATAC-seq data.

Currently, a large number of scATAC-seq datasets are being generated by different laboratories, using different experimental protocols. To leverage the full potential of that data, it is necessary to integrate it all in a single analysis. We have benchmarked different scRNA-seq atlas integration tools in their ability to integrate scATAC-seq datasets. Our results show that scATAC-seq integration is more challenging than that of scRNA-seq data, and that there is trade-off between batch effect removal and conservation of biological information.