BITS :: Call for Abstracts 2021 - Oral communication

Type	Oral communication
Session	Multiomics and Single Cell Analysis
Title	Genotyping Copy Number Alterations from single-cell RNA sequencing
All Authors	Salvatore Milite(1), Riccardo Bergamin(1), Giulio Caravagna(1)
Affiliation	
(1) Dept. Mathematics and Geosciences, University of Trieste, Italy	

Motivation

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 focus 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 data types 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.

Methods

In this work we follow this intuition and construct CONGAS, a Bayesian method to genotype CNA calls from single-cell RNAseq data, and cluster cells into subpopulations with the same CNA profile.
The main idea is that subclonal populations (Figure 1c) with a copy-number event in a given region will have, in the very same region, a change in the total number of counts (Figure 1d) proportional to the CNA.
CONGAS is based on a Poisson Mixture model, able to jointly model the change in copy-number as a latent variable and cluster subclones in RNA space (Figure 1e).
The method uses, as input absolute counts of transcripts from single-cell RNAseq and a segmentation of the tumour genome (Figure 1a).
Our tool is unsupervised, and leverages on a segmentation of the input DNA to determine the sample subclonal composition at the copy number level (Figure 1b), 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.
CONGAS is implemented in 2 open-source packages.
The CONGAS package implements all the fitting procedures in the Python probabilistic programming language Pyro, exploiting its backend to fit the model by stochastic variational inference, running on both CPU and GPU. An extra R package RCONGAS wraps functions for data pre- and post-processing, visualisation and model investigation around CONGAS.
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, as well as in data from the Human Cell Atlas project. We also validate our prediction against a state-of-the-art tool for subclonal genotyping that uses scDNA-seq (Figure 1i).

Results

Our tool can be used to detect CNA-associated subclones (Figure 1f,h), and measure their precise differential expression patterns (Figure 1g), a key step to study how selective pressures shape genotypes and phenotypes evolution in distinct populations of cells. In this first work, we also show how to determine clone-specific differentially expressed genes which can only be partially explained by copy number segments, pointing to complex non-trivial regulatory mechanisms that link genotype states with expression patterns.
Our method provides a solid statistical framework to approach this type of inference, which is crucial to investigate clonal dynamics in disease progression, as well as cell plasticity and patterns of drug response from the large wealth of single-cell data available nowadays.

Info

-

Figure

Availability	https://www.biorxiv.org/content/10.1101/2021.02.02.429335v2

Corresponding Author

Name, Surname	Giulio, Caravagna
Email	gcaravagna@units.it
Submitted on	28.04.2021