

Spatiotemporal Reconstruction of Gene Expression

Guilherme de Sena Brandine

April 11, 2017

1 Introduction

Many biological experiments are capable of giving detailed insights on the gene expression mechanisms governing complex systems at a molecular level. Two particular technologies in this context are:

- Imaging RNA *in situ* through fluorescent hybridizations (FISH), which yields a quantitative measure of specific marker genes that allows for a clear picture of its spatial distribution across a tissue.
- Droplet-Based Single Cell RNA Sequencing (Drop-Seq), which allows for the profiling of the entire gene expression of all the tissue's single cells.

The two methods complement each other in terms of resolution: While imaging gives us the detailed expression distribution across the entire space, it can only be done for a small subset of marker genes. Conversely, Drop-Seq is comprehensive in profiling the whole transcriptome, but the only known spatial information from these cells is the region of the tissue from which it was originally obtained.

A question of great interest in developmental biology is: Given a cell in a certain position coordinate in a fixed developmental timepoint of a tissue, what is the probability that the cell will be of a particular phenotype? (eg: a stem cell, a differentiating cell or a specialized cell) Furthermore, what is the probability distribution of a particular gene or set of genes given that the cell is of a given phenotype? Having such information at hand would give us a better understanding of the correlation between genes and spacing and allow for the discovery of pathways associated with cell specialization in developing tissues, organs and systems.

Rigorously speaking, all this information can be obtained separately by two aforementioned methods, the remaining problem being to accurately integrate the two to obtain the detailed full picture. The computational challenge is thus to be able to merge the high resolution information from both methods by matching equivalent data points using their low resolution counterparts. Specifically: For a particular region imaged with FISH from which a small set of relative fluorescence values of marker genes is known, we'd like use the correlation of the sequenced marked genes with the fluorescence values to infer both phenotype and gene expression probabilities and confidence regions in each space and time point for the whole transcriptome.