Abstract: Motivation: Single-cell experiments of cells from the early mouse embryo yield gene expression data for different developmental stages from zygote to blastocyst. To better understand cell fate decisions during differentiation, it is desirable to analyse the high-dimensional gene expression data and assess differences in gene expression patterns between different developmental stages as well as within developmental stages. Conventional methods include univariate analyses of distributions of genes at different stages or multivariate linear methods such as principal component analysis (PCA). However, these approaches often fail to resolve important differences as each lineage has a unique gene expression pattern which changes gradually over time yielding different gene expressions both between different developmental stages as well as heterogeneous distributions at a specific stage. Furthermore, to date, no approach taking the temporal structure of the data into account has been presented. Results: We present a novel framework based on Gaussian process latent variable models (GPLVMs) to analyse single-cell qPCR expression data of 48 genes from mouse zygote to blastocyst as presented by (Guo et al., 2010). We extend GPLVMs by introducing gene relevance maps and gradient plots to provide interpretability as in the linear case. Furthermore, we take the temporal group structure of the data into account and introduce a new factor in the GPLVM likelihood which ensures
that small distances are preserved for cells from the same developmental stage. Using our novel framework, it is possible to resolve differences in gene expressions for all developmental stages. Furthermore, a new subpopulation of cells within the 16-cell stage is identified which is significantly more trophectoderm-like than the rest of the population. The trophectoderm-like subpopulation was characterized by considerable differences in the expression of Id2, Gata4 and, to a smaller extent, Klf4 and Hand1. The relevance of Id2 as early markers for TE cells is consistent with previously published results.

Zeitschriftentitel: Bioinformatics

Jahr: 2012

Band: 28

Heft / Issue: 18

Seiten: i626-i632


entries: