

CLUSTERING IN RNA-Seq DATA USING GRADE of MEMBERSHIP MODELS

Kushal K Dey¹, Chiaowen Joyce Hsiao, Matthew Stephens^{1,2}

¹ Department of Statistics, University of Chicago, Chicago, Illinois 60637, USA;

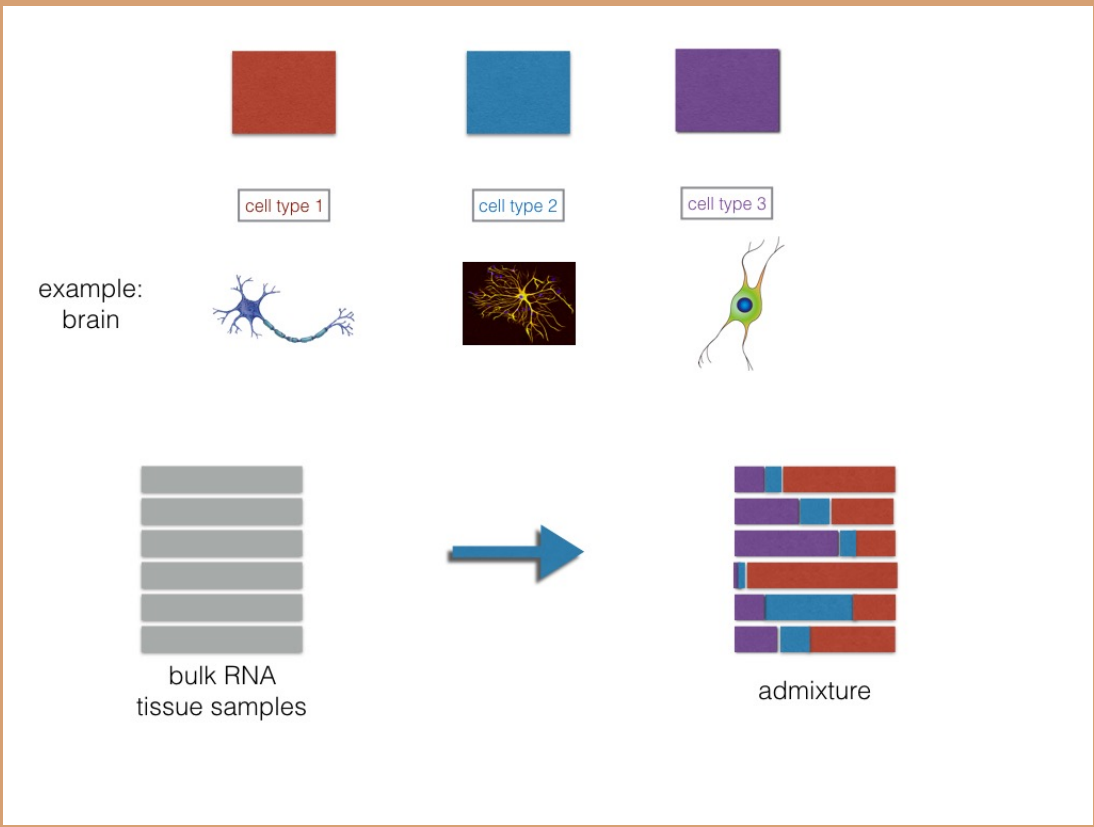
² Department of Human Genetics, University of Chicago, Chicago, Illinois 60637, USA



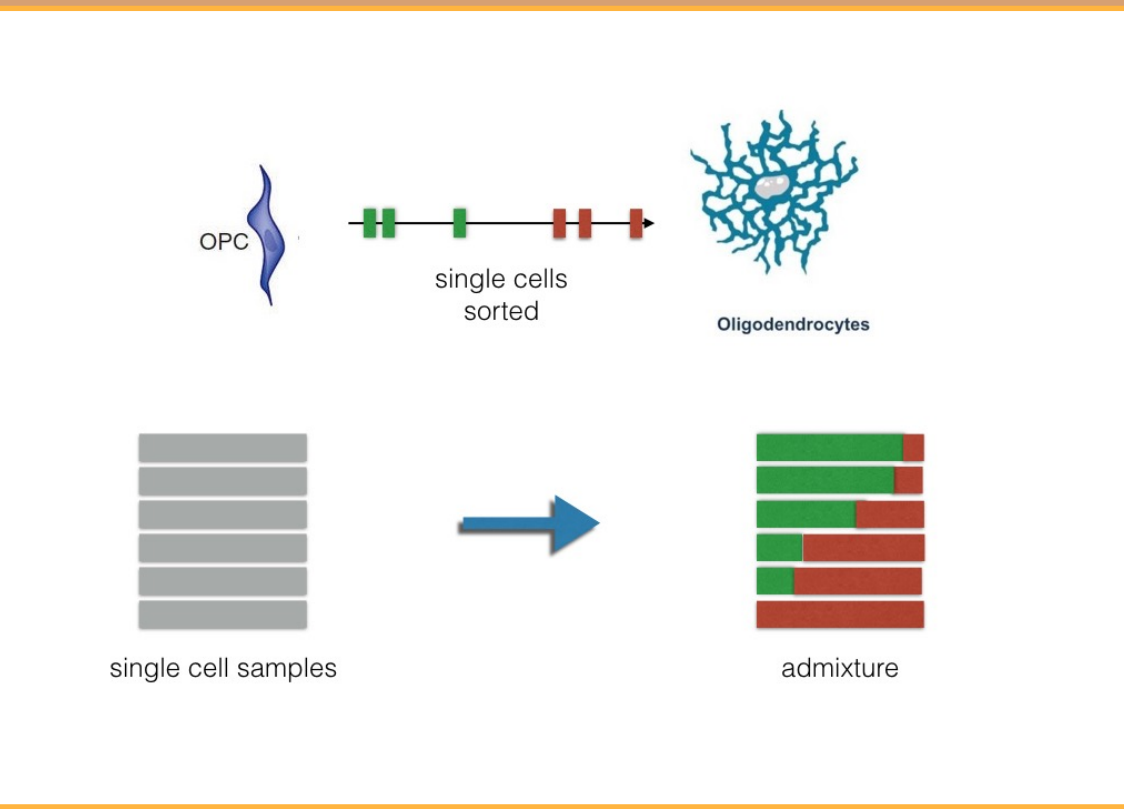
Abstract

We use a model based clustering method, similar to “admixture model” in population genetics for interpreting structure in bulk and single cell RNA-seq read counts data.

Each tissue sample in bulk RNA-seq may have memberships in multiple clusters, which are taken as representatives of different cell types.

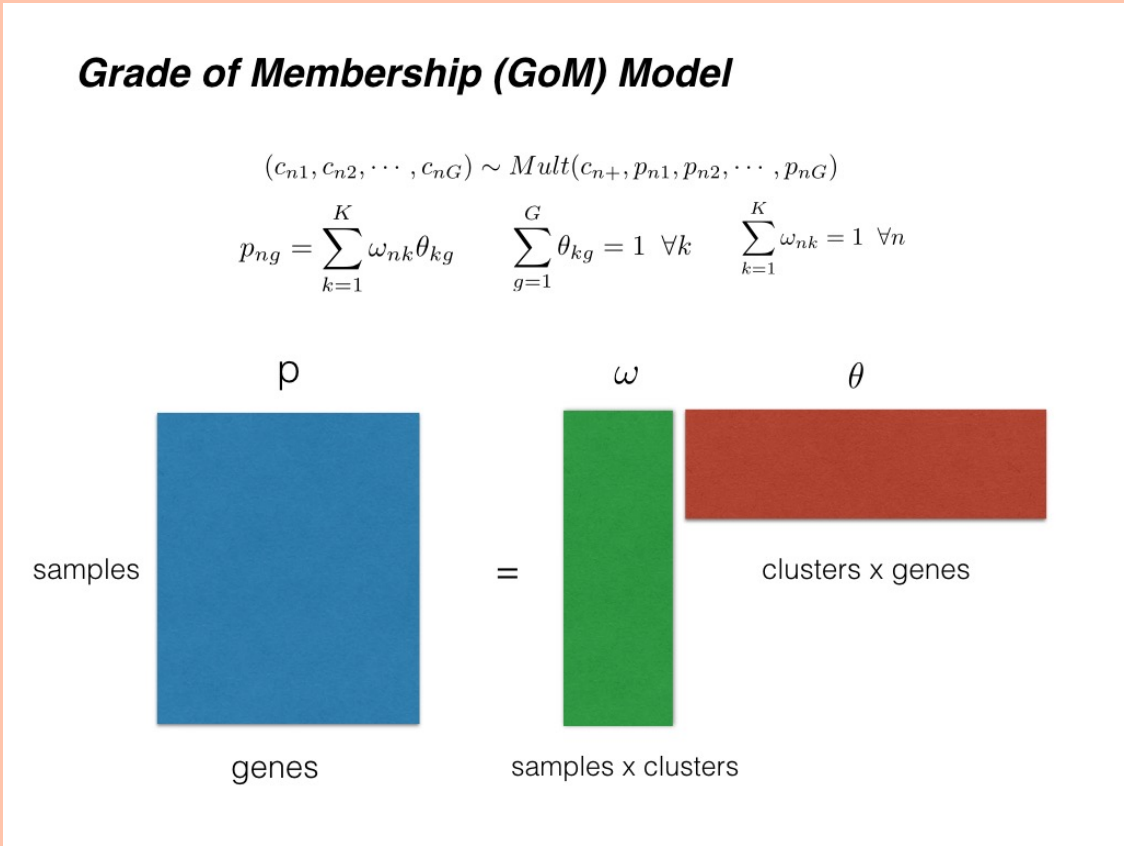


In single cell RNA-seq, each cell may have grades of memberships in the extreme stages of development.



Model

c_{ng} : reads mapped to gene g for sample n
 c_{n+} : library size sample n

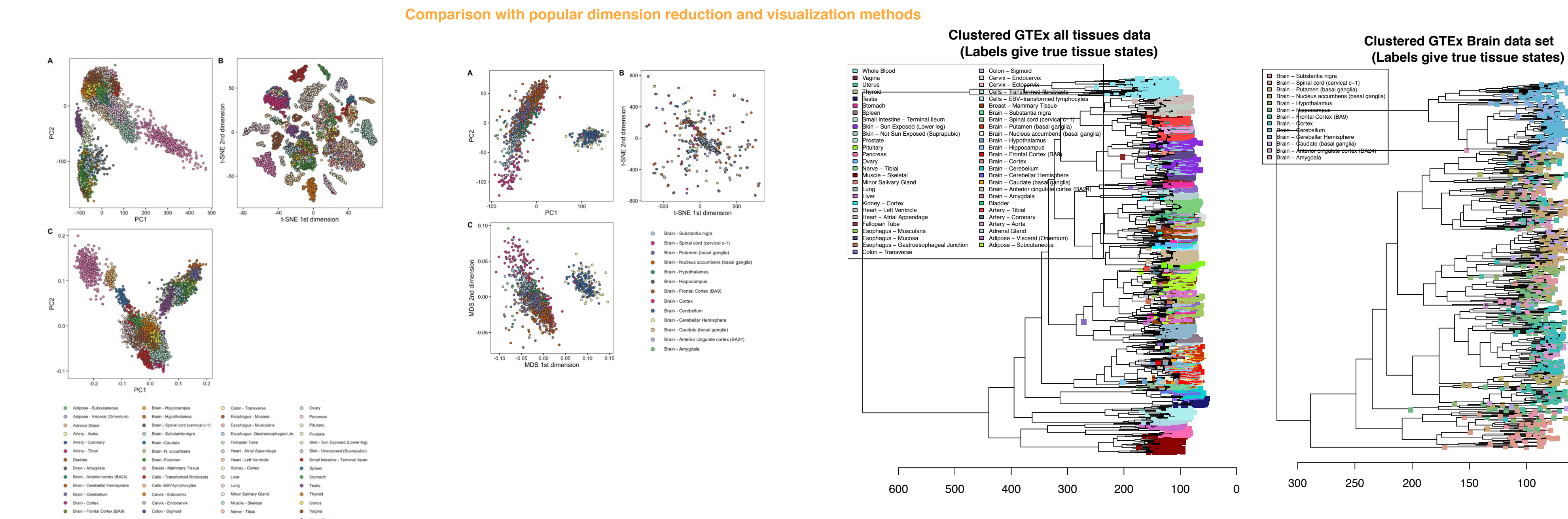
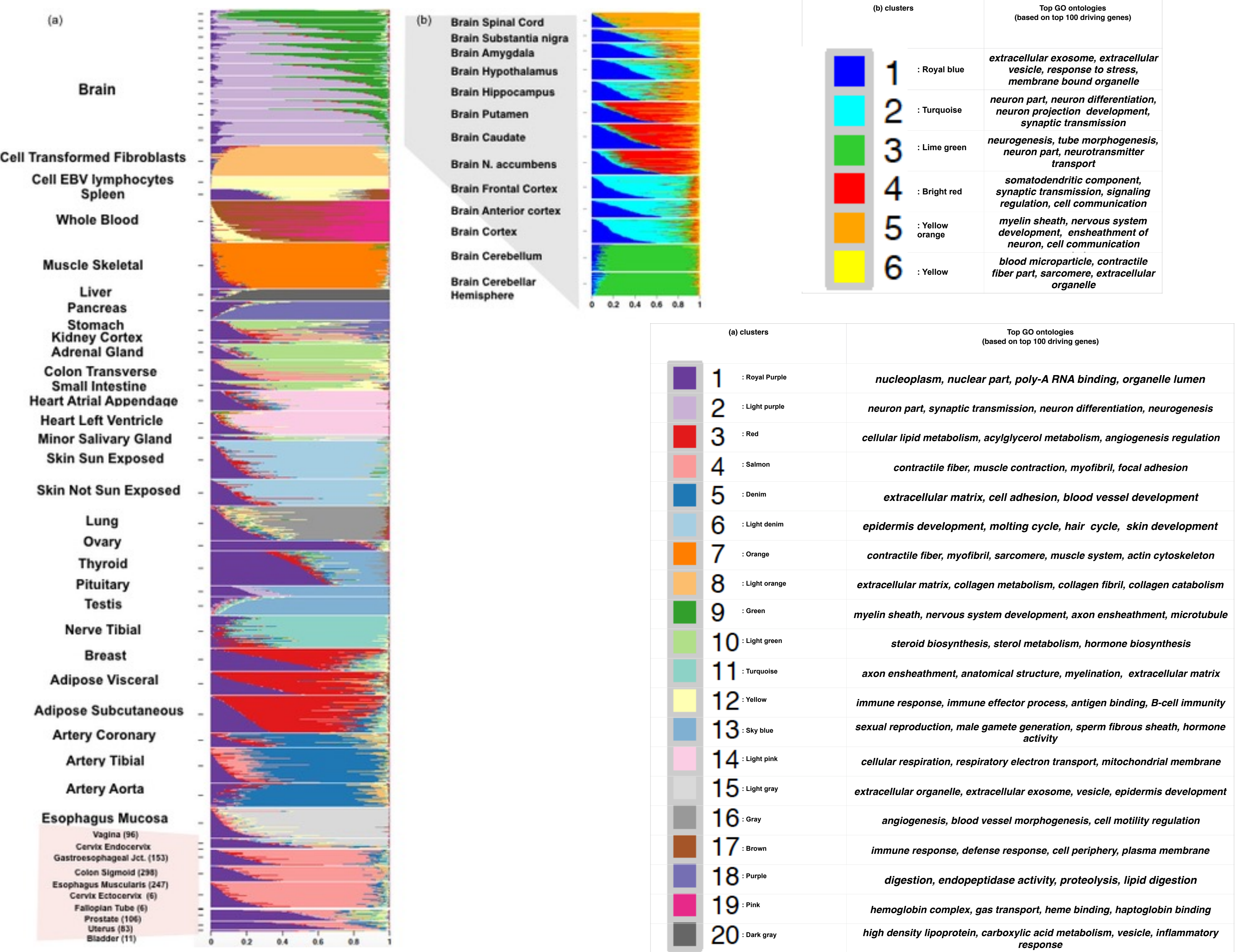


Applications

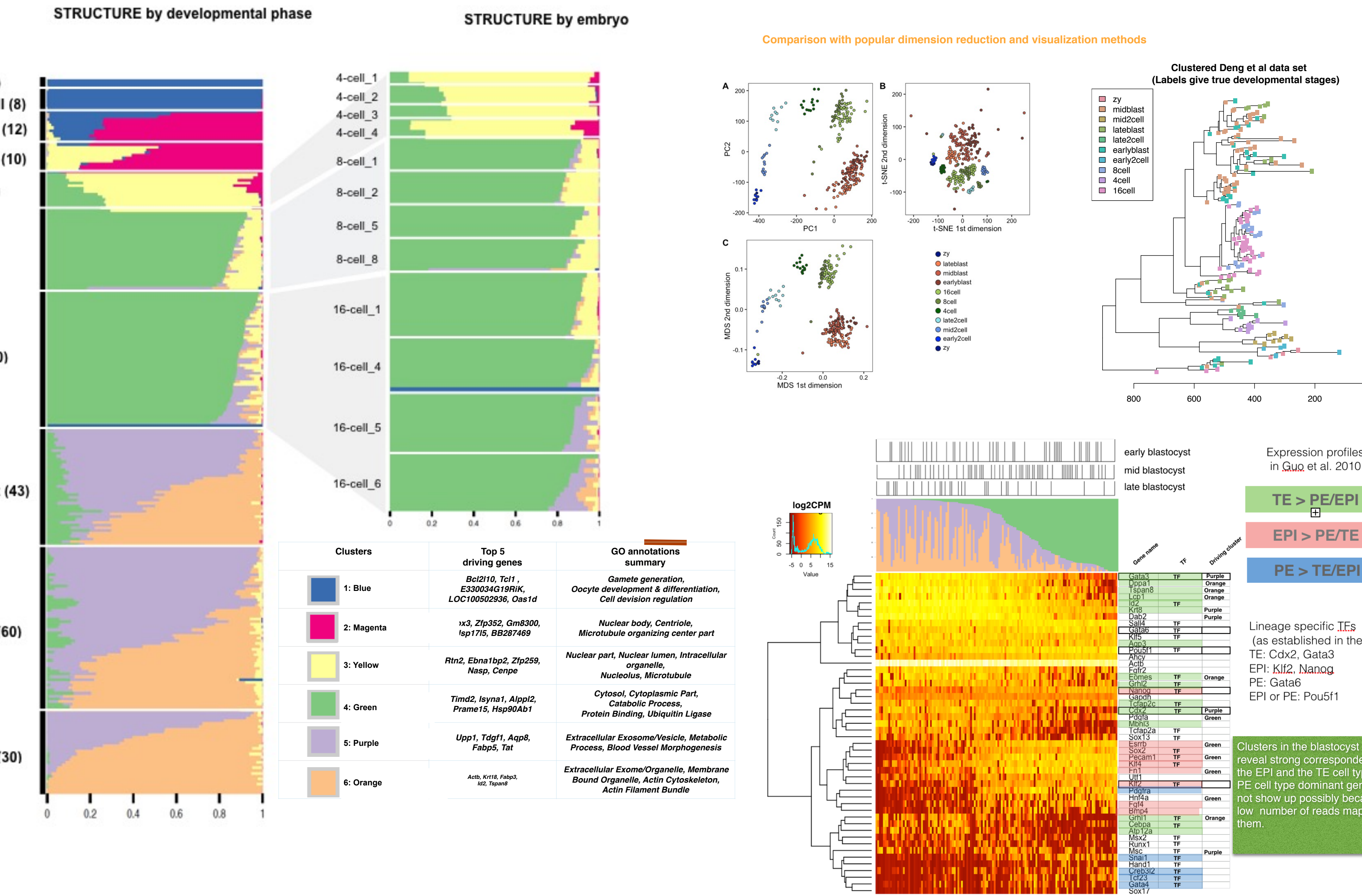
1. Bulk RNA-seq (GTEx - Genotype tissue expression project bulk RNA samples across tissues)
2. Single cell RNA-seq (Deng et al 2016 - single cell developmental phase data collected from mouse embryo across multiple developmental stages)

The model fit and the visualization using the Structure plot stacked bar chart representation is performed by the **Bioconductor** package **CountClust**

Bulk RNA-seq (GTEx data) analysis



Single cell RNA-seq analysis (Deng et al 2016)



References:
1. K Dey, C Hsiao, and M Stephens. Clustering RNA-seq expression data using grade of membership model. <http://bioRxiv.org/content/early/2016/05/03/051631>.
2. Deng Q, Ramsdell D, Reinus B, Sandberg R. 2014. Single-Cell RNA-Seq Reveals Dynamic, Random Monoallelic Gene Expression in Mammalian Cells. Science. 343 (6167): 193-196.
3. Guo G, Hsiao M, Tong DG, Wang C, Sun LL, Clarke ND, Holston P. Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. Developmental Cell. 18(4): 675-686.
4. The GTEx Consortium. 2013. The Genotype-Tissue Expression (GTEx) project. Nature genetics. 45(6): 580-585. doi:10.1038/ng.2653.
* The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the U.S. Department of Health and Human Services, the National Institutes of Health. The paper is supported by the grant U01CA198835 from the NIH R01CA program.