Modelling Transcriptomics Data of the Developing Enteric Nervous System

Jens Kleinjung, Reena Lasrado and Vassilis Pachnis
The Francis Crick Institute

London
jens.kleinjung@crick.ac.uk
https://jkleinj.github.io

Introduction

The Enteric Nervous System (ENS) comprises in the order of 10^8 cells that perform sensory functions, control intestinal muscle movement and regulate enzyme secretion. The ENS develops from neural crest cell progenitors that proliferate and differentiate to glia and neurons along with the expansion of the gastrointestinal tract during development. This spatio-temporal self-organisation of the ENS to form a functional neural network despite a seemingly disordered distribution of neurons is only marginally understood. Using transcriptomics, in particular single-cell RNA-Seq data, we have studied ENS cell lineages at embryonic time E13 [1]. More recently we have performed a developmental time series from stage E11 to the adult to reveal the transcriptomic variability and to characterise cellular lineages through their gene expression levels.

Cell Lineage Identification at embryonic time E13

Key factors of the analysis were the progenitor/gliogenic (Erbb3, Sox10, Fabp7 and Plp1) and progenitor/neurogenic (Tubb3, Elavl4, Ret and Phox2b) marker genes.

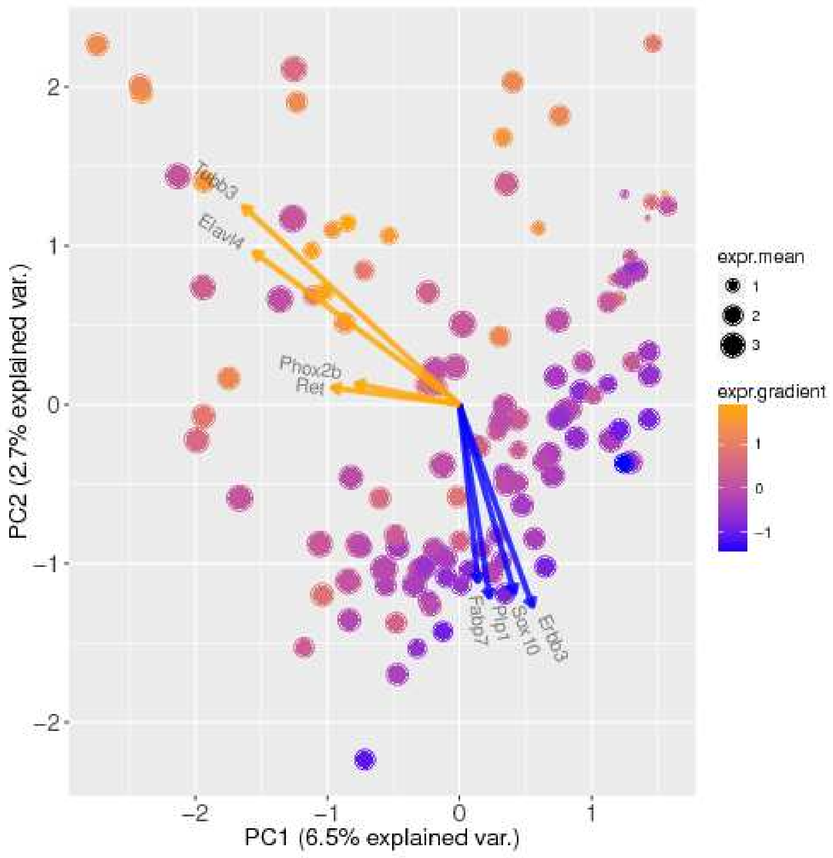
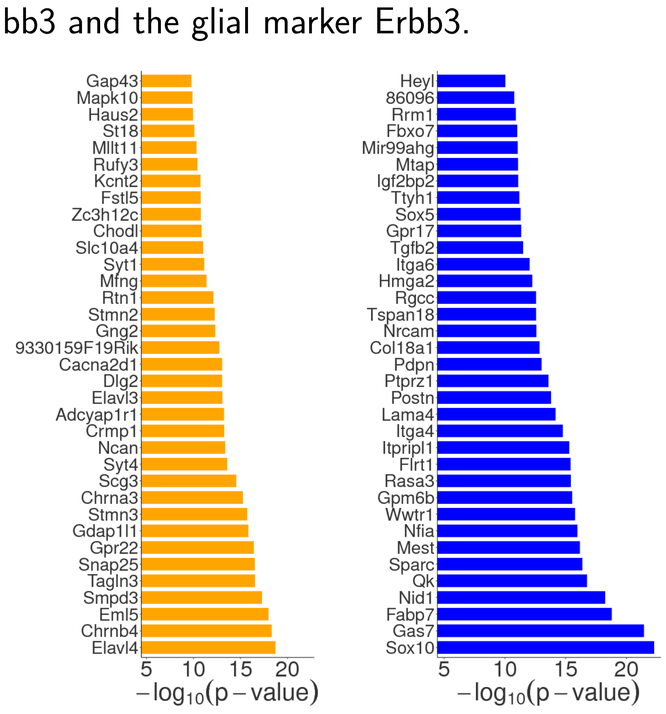


Fig. 1: Biplot of single-cell gene expression. The underlying data set comprises the most divergently expressed genes. The gradient of marker expression is illustrated as colour shade between in blue (glial) and orange (neuronal). Total marker expression is indicated by the size of the sphere symbol. Arrows show the direction and magnitude of marker gene vectors in the given PC1/PC2 plane.

Detection of Lineage-Specific Genes

Using the expression profile of marker genes across all single-cells, a linear regression model was applied and the resulting fit was assessed by ANOVA. The set of best fitting genes is listed below for the neuronal marker Tubb3 and the glial marker Erbb3.



(a) Tubb3 (b) Erbb3

Fig. 2: Lists of genes with expression profiles similar to those of (a) the neuronal marker Tubb3 and (b) the glial marker Erbb3.

Cross-Sectional Time Series

Bulk RNA-Seq data were collected for embryonic stages E11, E13 and E17 and for adult lineages RET, CHAT and NOS. A PCA plot resolves all stages/lineages in the first three PC dimensions.

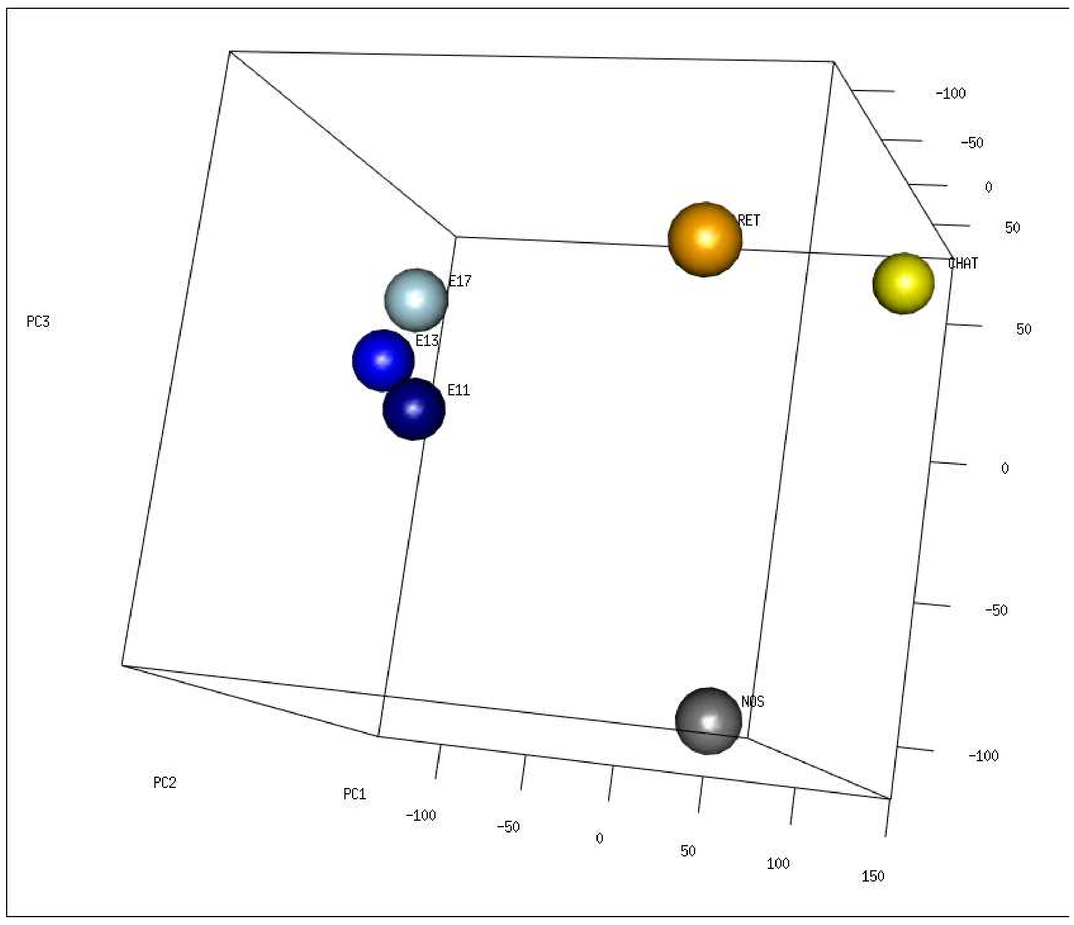
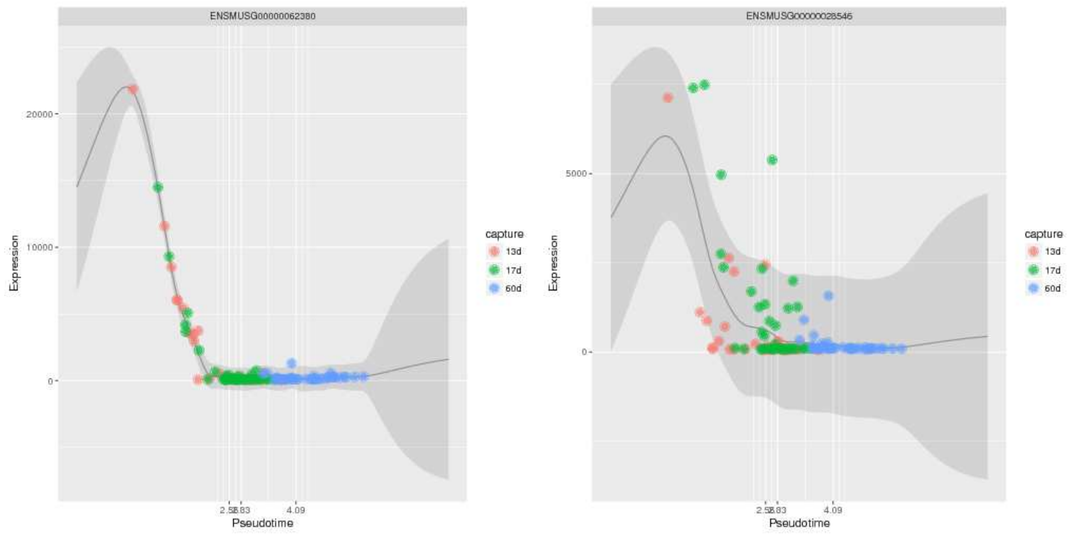


Fig. 3: 3-Dimensional PCA plot of a cross-sectional time series of ENS gene expression. The progression from embryonic cells (blue colours) at times E11, E13 and E17 to adult cells (gray, orange and yellow) is well resolved. The adult cell lineages RET, CHAT and NOS show very distinct expression profiles.

Pseudo-time Estimation

The temporal behaviour of gene expression can be modeled as a Gaussian Process, where a multivariate Gaussian is fitted to a given set of data within a Bayesian framework [2]. The examples in Fig.4 illustrate the down-regulation of embryonic neuronal genes. The shade around the regression line indicates the prediction uncertainty.



(a) Tubb3 (b) Elavl4

Fig. 4: Pseudo-time estimates of the single-cell expression profiles of the marker genes (a) Tubb3 and (b) Elavl4.

Conclusions

Single-cell RNA-Seq data provide a detailed picture of the heterogeneity of gene expression within differentiating cell populations of the ENS. Gene expression profiles allow for the discriminative analysis of cell lineages and the detection of gene clusters. Time-dependent gene expression can be modelled as Gaussian Process, yielding a pseudo-temporal description of the underlying biological processes.

References

[1] R. Lasrado, W. Boesmans, J. Kleinjung, C. Pin, D. Bell, L. Bhaw, S. McCallum, H. Zong, L. Luo, H. Clevers, P. V. Berghe, and V. Pachnis.
Lineage-dependent spatial and functional organization of the mammalian enteric nervous system.
Science, 356(6339):722–726, 2017.

[2] J. E. Reid and L. Wernisch.
Pseudotime estimation: deconfounding single cell time series.
Bioinformatics, 32(19):2973–2980, 2016.