# Gene Expression Trees in Blood Cell Development

**Ivan G. Costa***, **Stefan Roepcke***, **Alexander Schliep**

Dept. of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Berlin, Germany * Authors contributed equally to this work.

We present a novel statistical framework for analyzing transcription during cell development. In particular, we focus on data from mouse lymphoid lineages [1], which is part of blood cell development. Gene expression data of cells of several distinguishable developmental stages fosters the elucidation of the underlying molecular processes, which change gradually over time and lock cells in certain lineages. Large-scale analysis of this data requires a computational framework for tasks ranging from visualization, querying, and finding clusters of similar genes, to answering detailed questions about the functional roles of individual genes and their similarities and differences.

In development, we have temporal sequences, in which one cell type undergoes a given differentiation process and turns into a new cell type. Furthermore, there can be branching points at which cells at particular developmental stages differentiate into two or more cell types, and subsequently follow different differentiation paths. To address the tree-like structure of these processes, we choose to represent mRNA expression data during cell development and differentiation with tree models [2], which model dependencies during differentiation, and to combine several of these models in a mixture. Together, we obtain a robust and flexible statistical model for analyzing and clustering genome-wide mRNA expression data sets, in which the inherent dependencies between stages can be seen and overlapping clusters are allowed. Additionally, we combine sequence information, as given by finding microRNA binding sites [3], and genes with similar expression profiles, as given by the clustering results, to perform a more detailed microRNA target prediction (see Figure 1).

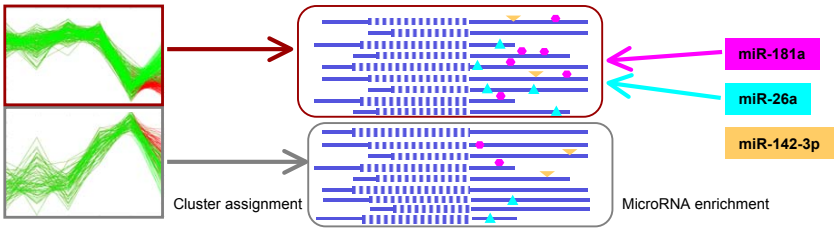


Figure 1: Strategy to identify microRNAs and their target genes overrepresented in groups of co-expressed genes (indicated left) as part of a post-transcriptional regulatory mechanism. In the middle mRNAs clustered according to our clustering results are depicted and potential microRNA binding sites in their 3'UTRs are symbolized.

Our work concentrates on two detailed studies covering several stages of the B and T cell development [1]. We recover well-known biological facts and also identify putative but convincing regulatory elements, genes and functional assignments. In particular, our results suggests that some microRNAs, which have been previously related to hematopoiesis [4], have a regulatory role in reducing the transcript levels of genes that are important for cell proliferation.

[1] R. Hoffmann and F. Melchers. A genomic view of lymphocyte development. *Curr Opin Immunol*, 15(3):239-245, 2003.

[2] C. Chow and C. Liu. Approximating discrete probability distributions with dependence trees. *IEEE Trans. Info. Theory*,14(3):462-467, 1968.

[3] A. J. Enright, *et al.*. MicroRNA targets in Drosophila. *Genome Biol*,5(1):R1, 2003.

[4] S. Monticelli, *et al.*. MicroRNA profiling of the murine hematopoietic system. *Genome Biol*, 6(8):R71,2005.