Unsupervised Detection of Hematopoietic Stem Cell Subpopulations from Single Cell Time Lapse Movies

Interdisciplinary Project Description

Quantitative Single Cell Dynamics
Institute of Computational Biology
Helmholtz Zentrum München

February 20, 2014

Processed by: Linda Leidig (TUM: 03608416), André Seitz (TUM: 03622870)
From February 20, 2014 to July 31, 2014

Medicine - field of application

The vertebrate blood system consists of different cell types with common progenitors, referred to as hematopoietic stem cells (HSC). Due to their ability to differentiate into certain hematopoietic cell types, as well as their self renewability, HSCs are in the focus of research [1]. Medicine is interested in these cells as investigation of stem cells may improve the understanding of cell biology as well as the understanding and treatment of certain diseases like cancer [7]. To investigate the cells in a certain state, cell identification plays a major role in preparation of biological experiments and analytical methods. Stem cell identification shall be improved in scope of this project.

The data, on which this project is based, consists of three single cell time lapse movies (SCTLM) of sorted hematopoietic stem cells (HSC), further referred to as starting cells. These cells have been purified according to certain biological markers by fluorescence-activated cell sorting (FACS) before they were cultivated [5]. However, these cells and their descendants revealed that purification with respect to the biological markers only lead to purity of 60%. The movies were generated in the lab of Timm Schroeder. They have three different channels, each showing either out of focus brightfield signals or fluorescence signals emitted from certain marked proteins.

Project

The aim of this project is to develop an unsupervised detection method for HSCs based on features that can be extracted from the existing movies like cell size, cell shape, and cell movement [3].

The project consists of two parts: first, the cells in the movies have to be segmented and tracked over time. This might be achieved automatically. For most of the cells it will be sufficient to track only the first or the first two generations, what will reduce the amount of work to a feasible degree. As there already exist various methods to perform segmentation [9] and tracking [6, 4],

the major work will be to apply and validate promising methods, integrate the best ones into a pipeline, and use this pipeline. The result of this first part will be a sufficiently segmented and tracked dataset of starting cells in the movies. Moreover, the students will assess how good auto tracking performs for starting cells. Additionally, this task will allow the students to get a deeper understanding and sense for the data.
The second task is to implement, use, and validate a number of different unsupervised machine learning approaches in order to differentiate between HSCs and non-HSCs in the starting cell population. Some approaches already exist [2] which maybe have to be supplemented, validated, and integrated into a workflow. Available deep tracked trees can be taken into account for either learning or validation too. As, in order to certainly determine the type of one cell, considerable effort in biological experiments is required, the validation will mainly be based on the general biological meaningfulness of the results e.g. certain cell symmetries, cell cycle lengths, or tracking tree symmetries within clusters [8]. In addition, the results can be roughly compared against the assumed FACS purity according to previous investigations.
The benefit of this method, if successful, is to optimize the subpopulation profiling and increase the HSC sorting purity which may lead to an improvement of purifying methods and thereby to an improvement of subsequent methods and analyses.

Informatics
As stated above, the informatics part of this project is the implementation of unsupervised machine learning methods using MATLAB to detect HSCs. Therefore various methods (like one-class SVM, PCA and further clustering methods) will be implemented, depending on the first statistical evaluation of features of the tracked cells.

Milestones

20.02.2014	start of project
24.02.2014	familiarize with data
26.02.2014	track 10 trees manually
03.03.2014	autotracking and comparison

References

[1] M. Bethesda. "Hematopoietic Stem Cells". In: *Stem Cell Information*. 2011. Chap. 5.

[2] C. M. Bishop. *Pattern Recognition and Machine Learning*. Ed. by M Jordan, J Kleinberg, and B Schölkopf. Vol. 4. Information science and statistics 4. Springer, 2006. Chap. Graphical, p. 738. arXiv: `0-387-31073-8`.

[3] R. Chakravorty et al. "Labour-efficient in vitro lymphocyte population tracking and fate prediction using automation and manual review." In: *PloS one* 9.1 (Jan. 2014), e83251.

[4] H. M. Eilken, S.-I. Nishikawa, and T. Schroeder. "Continuous single-cell imaging of blood generation from haemogenic endothelium." In: *Nature* 457.7231 (Feb. 2009), pp. 896–900.

[5] J. Picot et al. "Flow cytometry: retrospective, fundamentals and recent instrumentation." In: *Cytotechnology* 64.2 (Mar. 2012), pp. 109–30.

[6] M. A. Rieger et al. "Hematopoietic cytokines can instruct lineage choice." In: *Science (New York, N.Y.)* 325.5937 (July 2009), pp. 217–8.

[7] A. K. Saxena, D. Singh, and J. Gupta. "Role of stem cell research in therapeutic purpose–a hope for new horizon in medical biotechnology." In: *Journal of experimental therapeutics & oncology* 8.3 (Jan. 2010), pp. 223–33.

[8] T. Schroeder. “Hematopoietic stem cell heterogeneity: subtypes, not unpredictable behavior.” In: *Cell stem cell* 6.3 (Mar. 2010), pp. 203–7.

[9] M. Schwarzfischer et al. “Efficient and reliable long-term single-cell tracking and quantification of cellular and molecular behaviour in time-lapse microscopy”.