PROJECT TITLE:     **RNA-sequencing as a stochastic process**

SUPERVISOR:        Daniel Hebenstreit (Life Sciences)
Co-SUPERVISOR:     Barbel Finkenstadt (Statistics)

BACKGROUND

RNA-sequencing is an experimental technique that has revolutionized molecular biology over the last few years. It is based on 'next-generation-sequencing' and produces several millions of short sequencing tags that are mapped to mRNA sequences; it allows quantification of mRNA expression levels for virtually all expressed genes of a cell. It has recently been developed into a technique with single-cell resolution, which, in principal, also allows analysis of phenotypic variation on the level of single cells ('biological noise'). However, RNA-sequencing suffers from issues with regards to sensitivity and precision, with unclear implications for research findings that are obtained with it. The lack of precision is largely due to multiple experimental steps that are carried out during sample preparation and which are incompletely understood.

RESEARCH OBJECTIVES OF THE PROJECT

The principal objective of the project is it to understand details with regards to the experimental steps during RNA-sequencing sample preparation.  By modeling it as a stochastic process, we will be able to analyze how the datasets resulting from RNA-sequencing are affected by various steps in the protocol. This will help gaining insights into the actual reactions underlying sample preparation, which in turn will yield insights into bottlenecks, biases, etc. that affect precision of the method. The project will encompass the following steps:

1. We will take existing (assumed) knowledge about the biochemical reactions that take place during the sample preparation and will translate this into a probabilistic framework using conditional probabilities that link the individual steps in the protocol.
2. We will study this framework with regards to its sensitivity and dependence on parameters related to the experimental reactions.
3. We will construct a markov chain monte carlo (mcmc) sampler that allows 'reverse-engineering' of the experimental reactions based on existing datasets of our lab.
4. We will optimize computational aspects of the mcmc sampler and convert the scripts into an R library.

Preliminary work for the project has been carried out, which should allow for a smooth and straightforward introduction to the topic. Steps 1 and 2 are easily achievable within the time frame, steps 3 and 4 depend are subject to progress during the project. Our lab routinely carries out RNA-sequencing and several fully processes datasets are available, with no need for further bioinformatics.

PROJECT DELIVERABLES / IMPACT

The project will yield a stochastic model that describes the experimental steps of the RNA-sequencing method. Depending on the rate of progress during the project work, the work can be developed into a functional parameter estimation pipeline, and might even include the completion of a software package for general use (e.g. an R library). In that case, authorship on a publication is likely.

Due to the enormous impact and importance of RNA-sequencing in molecular biology, any improvements of and novel insights into experimental procedures are highly sought after; the global market for next-generation-sequencing, which includes RNA-sequencing along with genome sequencing as the two main applications, is expected to reach a value of $27.8 billion by 2022, according to recent market research.

TECHNIQUES

Stochastic modelling, programming, statistics, markov chain monte carlo sampling.

Literature:

1. Stochastic Methods: A Handbook for the Natural and Social Sciences (Springer Series in Synergetics), Crispin Gardiner.

2. Stochastic Modelling for Systems Biology (CRC Press), Darren J Wilkinson.

3. Single-cell RNA-seq: advances and future challenges. Saliba AE, Westermann AJ, Gorski SA, Vogel J. Nucleic Acids Res. 2014 Aug;42(14):8845-60. doi: 10.1093/nar/gku555.

4. Validation of noise models for single-cell transcriptomics. Grün D, Kester L, van Oudenaarden A. Nat Methods. 2014 Jun;11(6):637-40. doi: 10.1038/nmeth.2930.

5. Nature, nurture, or chance: stochastic gene expression and its consequences. Raj A, van Oudenaarden A. Cell. 2008 Oct 17;135(2):216-26.