

STREAM : A Practical Workbench for the Quantitative Modeling of Transcription

Denis C. Bauer¹ **Timothy L. Bailey**¹
d.bauer@imb.uq.edu.au t.bailey@imb.uq.edu.au

¹ Institute for Molecular Bioscience, The University of Queensland, Brisbane, Qld. 4072 Australia

Abstract

Understanding the transcriptional regulation of a gene in detail is a crucial step towards uncovering and ultimately utilizing the regulatory grammar of the genome. Here, we present STREAM, the first publicly available framework for modeling, visualizing, and predicting the regulation of the transcription rate of a target gene. Given the concentrations of a set of transcription factors, the transcription factor binding sites in a regulatory DNA region, and the transcription rate of the target gene, STREAM will optimize its parameters to generate a model that best fits the input data. This trained model can then be used to (a) validate that the given set of TFs is able to regulate the target gene and (b) to predict the transcription rate under different conditions (e.g. different tissues, knockout/additional TFs or mutated/missing TFBSs).

The platform independent executable of STREAM is available at <http://bioinformatics.org.au/stream/>.

Keywords:

Thermodynamic model, bioinformatics application, gene regulation, predicting transcriptional output

1 Introduction

Transcription of a gene can be induced by the binding of specific transcription factors (TFs) to so-called cis-regulatory-modules (CRMs). The frequency and duration of the binding events are influenced by the concentrations of the TFs, the binding affinities and location of the transcription factor binding sites (TFBSs) in the CRM and the properties of the TFs themselves (e.g. effectiveness, competitive interaction with other TFs).

An increasingly successful approach to mathematically simulating these binding events and ultimately the resulting transcriptional regulation is using thermodynamic models [3, 2, 4, 5]. These approaches model interaction of TF and DNA using kinetic equations and predict the transcriptional response of the target gene as mediated by the CRM and the TFs. A training algorithm is used to minimize the difference between the observed and predicted transcriptional response.

Building mathematical models to associate a specific occupation of a specific CRM with an observed transcriptional response promotes a better understanding of the transcriptional regulation and enables *in-silico* hypothesis-testing about postulated TFs or mechanisms.

2 Usage and Application

Here we present STREAM, a Java-implemented framework to calculate and visualize transcriptional regulation using thermodynamic modeling approaches. STREAM offers several optimization methods including gradient descent and simulated annealing for adjusting the internal parameters of the model

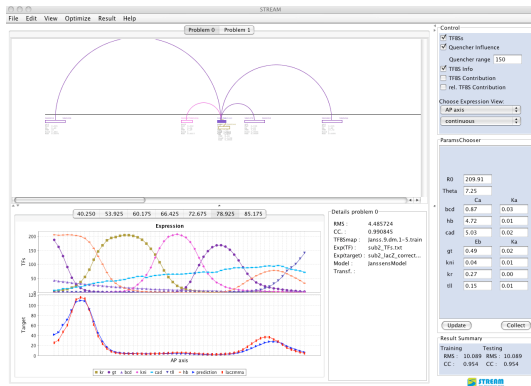


Figure 1: Screen-shot of the graphical user interface of STREAM.

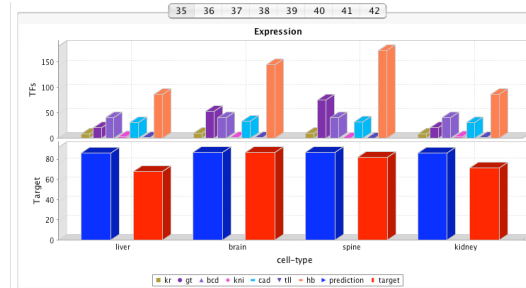


Figure 2: Screen-shot of the data representation in the categorical display mode.

to best fit the user's input data. STREAM can optimize one model to the data of more than one CRM. Being able to accurately fit data of several CRMs, which are suspected to have the same regulatory TFs, increases the confidence in the produced model.

STREAM can be executed using a graphical user interface (GUI) as well as via the command-line. The GUI is shown in Fig. 1 and Fig. 2.

STREAM has been tested extensively on the even-skipped gene (*eve*) in *Drosophila melanogaster* and was used to predict the location of the specific regulatory elements from the homologous genes of *eve* in six other *Drosophila* species [1].

3 Discussions

Besides the implemented thermodynamic model, we plan to provide additional models with enhanced functionality to simulate interacting TFs in a more detailed way, e.g., by incorporating TF-TF cooperation. Our framework can provide the environment to directly compare different modeling approaches.

References

- [1] Bauer, D. C. , Bailey, T. L. Studying the functional conservation of cis-regulatory modules and their transcriptional output. *BMC Bioinformatics*, 2008, 9, 220
- [2] Janssens, H., Hou, S., Jaeger, J., Kim, A., Myasnikova, E., Sharp, D. , Reinitz, J. Quantitative and predictive model of transcriptional control of the *Drosophila melanogaster* even skipped gene. *Nat Genet*, 2006, 38, 1159-1165
- [3] Reinitz, J., Hou, S. , Sharp, D. H. Transcriptional Control in *Drosophila* *Complexus*, 2003, 1, 54-64
- [4] Segal, E., Raveh-Sadka, T., Schroeder, M., Unnerstall, U. , Gaul, U. Predicting expression patterns from regulatory sequence in *Drosophila* segmentation. *Nature*, 2008, Jan
- [5] Zinzen, R. P., Senger, K., Levine, M. , Papatsenko, D. Computational models for neurogenic gene expression in the *Drosophila* embryo. *Curr Biol*, 2006, 16, 1358-1365