

REGULATORY GENOMICS – DECODING DROSOPHILA REGULATORY SEQUENCES

Stark A¹

¹IMP Vienna, Austria
stark@starklab.org

Abstract: *The precise regulation of gene expression is crucial for all animals. The regulatory information about when and where genes are to be expressed is located in defined genomic elements called enhancers that are able to activate gene expression in precise spatial and temporal patterns. We have developed a novel method to find such enhancers and measure their activity quantitatively across entire genomes. This allows us to trace enhancer activity across evolution and study its sequence basis.*

Keywords: *transcriptional regulation, enhancers, genomics, computational biology, transcription factors & motifs*

Introduction

In higher eukaryotes, genes are expressed dynamically in complex spatial and temporal patterns, which are progressively refined to set up body plans and define specific cell-types.

Transcriptional regulatory information is encoded in enhancers, discrete regions within the non-coding part of the genome. As enhancers lie at variable positions within and around genes their discovery has been challenging and only relatively few enhancers has been described and functionally characterized.

Methods

As methods to identify transcriptional enhancers genome-wide based on their activity have been lacking, we developed STARR-seq (self-transcribing active regulatory region sequencing, [1]). In STARR-seq, candidate sequences are positioned downstream of a minimal or core promoter such that sequences with enhancer activity transcribe themselves and enhancer activity is reflected by the sequences' presence among cellular RNA. This allows the quantification of enhancer activity for millions of candidate sequences from arbitrary sources of input DNA and enables screens across entire genomes.

We applied STARR-seq to the *Drosophila melanogaster* genome in two different *Drosophila* cell lines and to a 1 megabase (Mb) region of the human genome in human HeLa cells.

Results

Applied to two *Drosophila* cell-types, STARR-seq reveals thousands of enhancers with cell-type specific activity across a wide range of strengths. Surprisingly, about one third of the enhancers lie in inaccessible chromatin and

are marked by H3K27me3, suggesting that their endogenous loci are silenced or correspond to a poised enhancer state – even though such enhancers function in luciferase assays using ectopic or genomically integrated reporters.

STARR-seq reveals a surprising complexity of gene regulation with several independently functioning enhancers for single genes, including both developmental regulators (e.g. transcription factors) and broadly expressed genes such as actin. Similarly, the genome-wide strongest enhancers are located near both functional classes of genes.

Using STARR-seq, we are screening the genomes of five closely related *Drosophila* species in *Drosophila melanogaster* S2 cells to trace cis-regulatory function across evolution in a constant trans-regulatory environment. This will allow us to assess conservation of enhancer function in orthologous sequences as well as the number of enhancers that emerged de novo since the *D. melanogaster* and *yakuba* split about 10 million years ago.

Finally, we compare the sequences of enhancers with similar and different activities across different cell-types and tissues to determine regulatory motifs that are shared in functionally related sequences and are required for regulatory function.

Discussion

STARR-seq will be widely applicable to screen DNA from arbitrary sources in any cell-type of interest. This includes human HeLa cells, for which we demonstrated that STARR-seq detects enhancers that function in classical luciferase assays independent of their chromatin states and thus more reliably than previous methods.

Acknowledgement

We thank the IMP/IMBA BioOptics and Graphics Services for help. Deep sequencing was performed at the CSF Next-Generation Sequencing Unit (<http://csf.ac.at>). This work is supported by a European Research Council (ERC) Starting Grant (no. 242922). Basic research at the IMP is supported by Boehringer Ingelheim GmbH.

Bibliography

- [1] C. Arnold et. al.: Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science*, vol. 339 (6123), pp. 1074-1077, March 2013.