View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by MDC Repository

OPEN ACCESS

Repository of the Max Delbrück Center for Molecular Medicine (MDC) in the Helmholtz Association

https://edoc.mdc-berlin.de/17071

Redundant regulation

Lacadie S.A., Ohler U.

This is the final version of the accepted manuscript. The original article has been published in final edited form in:

Nature Ecology & Evolution
2018 MAR ; 2(3): 418-419
2018 JAN 29 (first published online: final version)
doi: 10.1038/s41559-018-0479-5

URL: https://www.nature.com/articles/s41559-018-0479-5

Publisher: Macmillan Publishers (Springer Nature)

Copyright © 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Publisher's Notice
This is a post-peer-review, pre-copyedit version of an article published in *Nature Ecology & Evolution*. The final authenticated version is available online at: https://dx.doi.org/10.1038/s41559-018-0479-5

Gene expression: Redundant regulation

Scott A. Lacadie[1,3§], and Uwe Ohler[1,2,3§]

1 Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine, 13125 Berlin, Germany
2 Department of Biology, Humboldt Universitaet zu Berlin, 10115 Berlin, Germany
3 Berlin Institute of Health (BIH), Berlin, Germany
[§] Co-corresponding

Control of gene activity through transcriptional regulatory elements is a major driving force in human evolution. A new study measures nascent transcription directly and shows that sequence, activity and 3D organization of transcriptional regulatory elements, all contribute to the evolution of gene activity within primate CD4+ T-cells.

Spatiotemporal determination of gene expression is controlled by transcriptional regulatory elements (TREs) that are both proximal (promoters) and distal (enhancers) to genes. While it is generally thought that transcription factor binding sites (TFBSs) within TREs evolve to dictate when and where genes are expressed, previous reports have shown surprisingly low levels of mammalian TRE conservation[1–3]. Furthermore, few evolutionary studies to date have utilized measurements of three-dimensional chromatin structure known to correspond with TRE function. Writing in *Nature Ecology & Evolution*, Danko et al. report that TFBS sequences within human-specific TREs are under positive selection, TFBS sequence changes drive changes in transcription levels, TRE activity levels can explain gene activity levels, and evolution of both promoters and enhancers is influenced by the number of three dimensional interactions they participate in. These observations not only shed light on transcription regulatory evolution, but also have implications for basic mechanisms of how enhancers and promoters control gene output.

Accurate identification of active TREs is a critical first step to any comparative study of gene regulation. Currently used methodologies for genome-wide assessment of transcriptional regulation have limitations such as poor resolution, high false discovery rates, and indirect quantification of activities. Such methodologies mature in parallel with our basic understanding of all aspects of gene regulation.

To bypass many of the aforementioned pitfalls, Danko and colleagues use precision run-on followed by sequencing (PRO-seq)[4], a method consisting of nuclear isolation to trap RNA polymerases engaged on the DNA, followed by "run-on" transcription with labeled nucleotides and physical isolation of newly synthesized RNA. PRO-seq allowed the authors to 1) separate nascent, direct transcription from effects of post-transcriptional gene regulation, and 2) measure activity of both promoter and enhancer TREs via their production of divergently oriented transcripts (Figure 1). Indeed, research in the last decade has shown that a majority of active mammalian promoters and enhancers can initiate transcription on both DNA strands, in an upstream antisense orientation, from opposite edges of accessible, nucleosome-depleted regions of chromatin[5]. Defining TREs by the space in between these transcription events narrows in on the DNA sequence most relevant to TF binding, which holds promise for more accurate descriptions of regulatory sequence grammars. The authors applied PRO-seq to resting and activated primary T-cells from humans, chimpanzees, rhesus macaques, mice, and rats due to their ease of isolation and their known high rate of evolution.

The findings of the study support the hypothesis that sequence changes in transcription factor biding sites are a major driving force for the evolution of transcription levels, which they confirm in detail for one candidate locus resulting from their analysis (*SGPP2*) using both reporter assays and CRISPRi[6]. Making use of their previously described probabilistic tool INSIGHT[7] to assess aspects of natural selection from the distributions of polymorphism and divergence across the identified TREs and TFBSs, the authors also find that human-specific promoters and enhancers generally evolve under positive selection. Importantly, using a machine learning linear modeling approach, changes in coding gene transcription levels were substantially explained via changes in non-coding transcription levels of putative TREs assigned to them, an observation that has been somewhat elusive in the field using other measures such as TF binding or histone modification levels[1]. Lastly, Danko et al incorporate published datasets measuring three-dimensional chromatin interactions to make two key observations: 1) promoter functional and sequence conservation correlate with the number of distal interactions they participate in, and 2) distal TRE functional and sequence conservation inversely correlate with the number of interactions their targets participate in. Together, these results imply that for cases in which a promoter interacts with few enhancers, it requires these enhancers to be enriched for sequence and positional cues, and in cases where many enhancers participate, these may individually be under less stringent selection due to a distribution of the regulatory information. This model potentially explains the overall low levels of enhancer conservation in mammals reported in this and previous studies[1–3].

The PRO-seq assay combined with the author's own approach for predicting TREs from such data (dREG)[8] is a powerful strategy that likely captures the majority of relevant TREs. What remains open is the extent of its dependency on the relative strength and balance of TRE divergent transcription events, and whether it may overlook some candidate TREs that display transcription predominantly in one direction. There is a wide distribution of transcription initiation "directionality" in mammalian TREs, and a recent study found that highly conserved human promoters displayed more directional skewing toward the gene compared to more recently evolved promoters[9]. Furthermore, there may be types of regulatory elements not displaying transcription at all and, therefore, simply not candidates for detection by the PRO-seq/dREG approach, and in some systems the amount of material required to perform PRO-seq could preclude its use. Regardless, PRO-seq has many strong advantages for identifying active TREs and should be seriously considered as a choice assay for those undertaking such studies.

A natural next step to follow up on the work by Danko et al is to produce interaction data, such as ChIA-PET or Hi-C-seq, in each of the cell populations used for comparisons. It would be especially satisfying to directly observe the evolutionary dynamics of enhancer-promoter interactions as predicted by the current analysis. Furthermore, as isolation of homologous cell-types is such a critical aspect of comparative functional genomics, the field will benefit from quickly-evolving single-cell genomics methods which give higher confidence in the precise identity or similarity of distinct cell populations[10]. Lastly, further advances in computation approaches, such as dREG, for use with cutting-edge developments in assays for regulatory architecture will propel our understanding of basic transcriptional regulatory mechanisms and their dynamics throughout evolution.

Figure Legend.

PRO-seq signal identifies TREs and quantifies transcription activity directly.

a. Gene distal and proximal transcriptional regulatory elements (TREs) consist of nucleosome depleted regions (NDRs; boxed region) from which divergently orientated transcription initiates (red and blue arrows) from core promoter DNA sequences (yellow bars) close to flanking nucleosomes (grey cylinders). The PRO-seq assay can detect these highly unstable divergent RNAs (red and blue lines) whose levels are thought to reflect TRE activity. The author's previously described tool dREG[8] identifies the space between divergently oriented PRO-seq signal (grey rectangle) known to be enriched for transcript factor (TF) binding sites. **b.** PRO-seq quantifies gene transcription directly, avoiding effects of post-transcriptional gene regulation reflected in steady-state RNA levels. This is analogous to a bathtub of water where the level of water (RNA steady-state levels) is determined by a balance of the amount coming from the tap (RNA transcription; measured directly by PRO-seq) and the amount leaving the drain (RNA degradation).

1. Villar, D., Flicek, P. & Odom, D. T. Evolution of transcription factor binding in metazoans-mechanisms and functional implications. *Nat. Rev. Genet.* **15,** 221–233 (2014).
2. Siepel, A. & Arbiza, L. Cis-regulatory elements and human evolution. *Curr. Opin. Genet. Dev.* **29,** 81–89 (2014).
3. Reilly, S. K. & Noonan, J. P. Evolution of Gene Regulation in Humans. *Annu. Rev. Genomics Hum. Genet.* **17,** 45–67 (2016).
4. Kwak, H., Fuda, N. J., Core, L. J. & Lis, J. T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339,** 950–3 (2013).
5. Lacadie, S. A., Ibrahim, M. M., Gokhale, S. A. & Ohler, U. Divergent transcription and epigenetic directionality of human promoters. *FEBS J.* **283,** 4214–4222 (2016).
6. Thakore, P. I. *et al.* Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat. Methods* 1–9 (2015). doi:10.1038/nmeth.3630
7. Arbiza, L. *et al.* Genome-wide inference of natural selection on human transcription factor binding sites. *Nat. Genet.* **45,** 723–729 (2013).
8. Danko, C. G. *et al.* Identification of active transcriptional regulatory elements from GRO-seq data. *Nat. Methods* **12,** 433–438 (2015).
9. Jin, Y., Eser, U., Struhl, K. & Churchman, L. S. The Ground State and Evolution of Promoter Region Directionality. *Cell* **170,** 889–898 (2017).
10. Stubbington, M. J. T., Rozenblatt-Rosen, O., Regev, A. & Teichmann, S. A. Single-cell transcriptomics to explore the immune system in health and disease. *Science (80-.).* **358,** 58–63 (2017).

a

TRE

nucleosomes

transcription initiation sites

Variable width

NDR

precision run-on transcripts (PRO-seq)

dREG prediction

TF binding sites

b