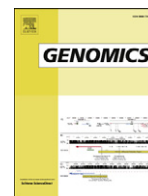# University of Massachusetts Amherst

September 6, 2015

Recent advances in functional assays of transcriptional enhancers

Courtney Babbitt, *University of Massachusetts - Amherst*
Michele Markstein, *University of Massachusetts - Amherst*
Jesse Gray, *Harvard Medical School*

Recent advances in functional assays of transcriptional enhancers

Courtney C. Babbitt [a,*], Michele Markstein [a], Jesse M. Gray [b]

[a] Department of Biology, University of Massachusetts, Amherst, United States
[b] Department of Genetics, Harvard Medical School, United States

ARTICLE INFO

ABSTRACT

In this special edition of Genomics, we present reviews of the current state of the field in identifying and functionally understanding transcriptional enhancers in cells and developing tissues. Typically several enhancers coordinate the expression of an individual target gene, each controlling that gene's expression in specific cell types at specific times. Until recently, identifying each gene's enhancers had been challenging because enhancers do not occupy prescribed locations relative to their target genes. Recently there have been powerful advances in DNA sequencing and other technologies that make it possible to identify the majority of enhancers in virtually any cell type of interest. The reviews in this edition of Genomics highlight some of these new and powerful approaches.

© 2015 Elsevier Inc. All rights reserved.

In this special edition of *Genomics*, we present reviews of the current state of the field in identifying and functionally understanding transcriptional enhancers in cells and developing tissues. Enhancers are short (50–1000) bp DNA sequences that precisely regulate the expression of target genes. Along with promoters they likely constitute the vast majority of the regulatory sequences in metazoan genomes. They are essential for development and function in animals and plants, although the first enhancer to be identified was a viral enhancer from the SV40 genome [4].

Typically several enhancers coordinate the expression of an individual target gene, each controlling that gene's expression in specific cell types at specific times. Until recently, identifying each gene's enhancers had been challenging because enhancers do not occupy prescribed locations relative to their target genes. Enhancers may be upstream, downstream, or within their target genes. They may be close to the transcriptional start site or as far as 1 million base pairs away. In the past, enhancers were painstakingly identified one-by-one. However, a number of powerful techniques have been recently developed that make it relatively straightforward to identify putative and actual enhancers on a genome-wide scale.

A central challenge in identifying enhancers in a given tissue or cell type is effectively isolating the relevant cells. Recently there have been powerful advances combining cell isolation and DNA sequencing that make it possible to identify the majority of enhancers in virtually any cell type. Bowman's review, "*Discovering enhancers by mapping chromatin features in primary tissue*," discusses the wide range of techniques to isolate cells or nuclei from primary tissues. These techniques range from bench-top affinity purification to traditional fluorescence activated cell sorting (FACS). Of particular excitement are developments in bench-top methods, which make it possible for individual laboratories to conduct analyses of tissue specific enhancers directly, without needing to invest money or time at a FACS facility. These approaches are gaining traction, as exemplified by the widespread use of the "Isolation of Nuclei Tagged in specific Cell Types," INTACT, method [8], which has now been used in *Xenopus*, *Arabidopsis*, *Drosophila*, and *Caenorhabditis elegans*.

Once cells are isolated, enhancers can be identified across the entire genome based on unique biochemical signatures. For example, enhancers tend to be largely nucleosome-free and therefore can be identified by methods that detect open chromatin, such as DNase-Seq, which detects DNase-accessible sites [6], and ATAC-Seq, which detects transpose-accessible sites [5]. A more refined method is to identify loci bound by modified histones that are better correlated with enhancer activity: histone H3 monomethylated at lysine 4 (H3K4me1) [9], and histone H3 acetylated at lysine 27 (H3K27ac) [7].

An important question is which of these putative enhancers are really functional in a cell or tissue? Recently developed methods such as STARR-Seq and FIREWACh can identify functional enhancers at a genome-wide scale. Murdter et al. describe Self-Transcribing Active Regulatory Region sequencing (STARR-seq) in their article "*STARR-seq — Principles and Applications.*" STARR-Seq is a quantitative method that screens in parallel millions of DNA fragments for their enhancer activity. In STARR-Seq, DNA libraries of putative enhancers are cloned into the 3′ UTR of a reporter construct. In this way, each putative enhancer serves as a unique barcoded-reporter for itself. After transfection into cells, the activity of millions of putative enhancers can be measured in parallel using sequencing. DNA libraries of putative enhancers can be obtained from arbitrary sources of DNA, including genomic DNA, TF binding sites, or

* Corresponding author.
  E-mail address: cbabbitt@bio.umass.edu (C.C. Babbitt).

predicted enhancers. STARR-Seq has been performed in *Drosophila* cells, leading to the identification of thousands of functionally active enhancers [3]. Attempts to apply genome-wide STARR-Seq in mammals have been complicated by the larger size of the mammalian genome, which requires construction of much larger libraries and sequencing to much greater depths. As a result, newer methods that focus STARR-SEQ on selected portions of the mammalian genome, such as the CAPSTARR-seq [12], are likely to become more widely used in mammals.

In Dailey's article ("*High Throughput Technologies for the Functional Discovery of Mammalian Enhancers*: *New Approaches for Understanding Transcriptional Regulatory Network Dynamics*") she describes Functional Identification of Regulatory Elements Within Accessible Chromatin (FIREWACh), a method in which DNase-accessible regions are cloned and subsequently screened for their activity. FIREWACh is more feasible in mammalian cells because it narrows the screen to to DNase-accessible regions. This eliminates the cloning, screening, and sequencing of extraneous sequences, dramatically simplifying the screening process. Moving forward, STARR-Seq, FIREWACh, and related methods are likely to uncover the dynamics of enhancers over the course of development and disease, such as in malignant transformation of mammalian cells.

Once enhancers have been found, it remains a challenge to discover the nucleotides within them that are required for their function. Systematic mutagenesis is impractical: even a single 100 bp sequence of DNA comprises far more possible sequence variants than can fit within the genomes of all the human beings on the planet. An emerging tool called Massively Parallel Reporter Assays (MPRAs) or CRE-seq, described both by Inoue and Ahituv ("*Decoding enhancers using massively parallel reporter assays*") and by White ("*Understanding how cis-regulatory function is encoded in DNA sequence using massively parallel reporter assays and designed sequences*"), is enabling this sequence space to be explored. In MPRA experiments, designer libraries of oligonucleotides can be commercially obtained and tested for their enhancer activity using RNA-Seq. While the throughput of MPRA does not allow for brute force mutagenesis of multiple enhancers, it nonetheless enables searches for guiding principles (or "regulatory grammar"). These assays have demonstrated large-scale functional validation of putative *cis*-regulatory elements (CREs), exhaustive mutational analyses of individual regulatory sequences, and tests of large libraries of synthetic CREs. Synthetic enhancers also allow for testing combinations of transcription factor binding sites that do not occur in the genome, thereby exploring non-natural "sequence space" to better understand the regulatory logic involved in *cis*-regulatory element function. Variants in enhancer regions affect phenotype and are implicated in human diseases [1]. High-throughput reporter assays suffer from a few current limitations. While MPRAs enable the investigation of the function of programmable sequences, they are limited to sequences of up to several hundred nucleotides, which is likely shorter than many regulatory sequences.

Another critical question in understanding functional enhancers is what dictates enhancer-promoter interactions. Enhancer function was originally defined based on plasmid experiments as being distance- and orientation-independent [4], but how does distance affect enhancer function across local genomic environments or at genome-wide scales? Here, Pindyurin et al. ("*TRIP through the chromatin*: *a high throughput exploration of enhancer regulatory landscapes*") discuss a new direction in the field to study the cumulative effects of many enhancers within their genomic contexts. The approach they discuss, called "Thousands of Reporters Integrated in Parallel" ("TRIP") involves integrating a barcoded reporter at random throughout the genome in tissue culture cells. In an early use of the technique, a constitutively active enhancer–promoter–reporter construct was integrated into thousands of locations, each of which was monitored by a unique barcode in the 3′UTR of the reporter [2]. Interestingly, the expression of the reporter varied by as much as 3 orders of magnitude depending in large part on its distance to the nearest enhancer. This result upends the original idea that enhancer function is distance-independent. Instead, whether an enhancer

has the potential to interact with a promoter depends almost entirely on its distance from that promoter. Enhancers in mammalian genomes that are within tens to hundreds of kilobases of promoters are able to interact, whereas those beyond these distances — or outside of topologically associated domains (TADs) — are not. This idea has been corroborated by genome-wide scans of chromatin interactions using Hi-C, which show that DNA–DNA interactions decline precipitously with linear genomic distance [11].

We can now identify functional enhancers in specific cells, but what about assaying enhancer function during organismal development? One powerful organismal system for understanding enhancer function as it relates to development and especially mammalian disease is the mouse. Reviews by Nord, ("*Learning about mammalian gene regulation from functional enhancer assays in the mouse*") and Kvon ("*Using transgenic reporter assays to functionally characterize enhancers in animals*") describe the original "enhancer trap" strategy to screen for regulatory sequences that drive expression in a cell- or tissue-specific manner. The VISTA database is a remarkable resource containing expression data from more than 2000 mouse and human sequences at embryonic day 11.5. The ability in mice to perform precise manipulations of endogenous genomic loci and examine resulting organismal phenotypes — now being done with increasing frequency — has confirmed that individual enhancers can influence limb development, facial morphology, brain structure, and body size.

In addition to the mammalian VISTA database, a database which is bound to have a major impact on our understanding of developmental enhancers is FlyEnhancers. As discussed in Kvon's review, FlyEnhancers currently reports the activity of over 7700 2-kb DNA fragments cloned from the *Drosophila* genome and tested in transgenic reporter assays for activity in *Drosophila* embryos [10]. The fragments constitute 13.5% of the *Drosophila* noncoding non-repetitive genome, representing the most comprehensive examination to date of how enhancer activity is encoded across a genome. Over 45% of the tested fragments were found to act as enhancers in fly embryos, suggesting that the fly genome may have 50,000–100,000 developmentally active enhancers. Interestingly, 10–20% of the enhancers reported appear to skip over their nearest neighboring genes to regulate the expression of a more distal gene. Based on similar extrapolations, it has been estimated that the mouse genome may have on the order of 1 million developmentally active enhancers. As discussed by Kvon, the positive and negative data from FlyEnhancers will serve as both a guide and inspiration for continued approaches to functionally identify enhancers across animal genomes.

High-throughput functional assays of enhancers, whether highly laborious or highly parallel, have proven to be powerful in generating knowledge of enhancer function. These assays all treat the enhancer as a fundamental unit of regulation. Treating the enhancer as a unit is fully appropriate to study how transcription factors cooperate at an enhancer or to screen for genetic variants that alter enhancer function. Unfortunately, the throughput of functional assays of enhancers drops precipitously once the unit of regulation becomes the gene instead of the enhancer — yet it is gene regulation that ultimately matters. There is great potential in extending high-throughput assays of enhancers to high-throughput assays of their function in regulating genes in the genome. Perhaps most promisingly, success in parallelizing Cas9-based genome editing to assess the effects of mutating many enhancers simultaneously could bring the power of massively parallel reporter assays to the genome.

References

[1] N. Ahituv (Ed.), Gene Regulatory Sequences and Human Disease, Springer, New York, 2012 (283 pp.).

[2] W. Akhtar, A.V. Pindyurin, J. de Jong, L. Pagie, J. Ten Hoeve, A. Berns, L.F. Wessels, B. van Steensel, M. van Lohuizen, Using TRIP for genome-wide position effect analysis in cultured cells, Nat. Protoc. 9 (6) (2014) 1255–1281, http://dx.doi.org/10.1038/nprot.2014.072 (Epub 2014 May 8).

[3] C.D. Arnold, D. Gerlach, C. Stelzer, Ł.M. Boryń, M. Rath, A. Stark, Genome-wide quantitative enhancer activity maps identified by STARR-seq, Science 339 (6123) (2013) 1074–1077, http://dx.doi.org/10.1126/science.1232542.

[4] J. Banerji, S. Rusconi, W. Schaffner, Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences, Cell (1981) 299–308.

[5] J.D. Buenrostro, P.G. Giresi, L.C. Zaba, H.Y. Chang, W.J. Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position, Nat. Methods 10 (2013) 1213–1218, http://dx.doi.org/10.1038/nmeth.2688.

[6] G.E. Crawford, et al., DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays, Nat. Methods 3 (2006) 503–509, http://dx.doi.org/10.1038/nmeth888.

[7] M.P. Creyghton, et al., Histone H3K27ac separates active from poised enhancers and predicts developmental state, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 21931–21936, http://dx.doi.org/10.1073/pnas.1016071107.

[8] R.B. Deal, S. Henikoff, The INTACT method for cell type-specific gene expression and chromatin profiling in *Arabidopsis thaliana*, Nat. Protoc. (2011) 56–68, http://dx.doi.org/10.1038/nprot.2010.175.

[9] N.D. Heintzman, et al., Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome, Nat. Genet. 39 (2007) 311–318.

[10] E.Z. Kvon, T. Kazmar, G. Stampfel, J.O. Yáñez-Cuna, M. Pagani, K. Schernhuber, B.J. Dickson, A. Stark, Genome-scale functional characterization of *Drosophila* developmental enhancers in vivo, Nature 512 (7512) (2014) 91–95, http://dx.doi.org/10.1038/nature13395.

[11] N.L. van Berkum, et al., Hi-C: a method to study the three-dimensional architecture of genomes, J. Vis. Exp. (2010) http://dx.doi.org/10.3791/1869.

[12] L. Vanhille, A. Griffon, M.A. Maqbool, J. Zacarias-Cabeza, L.T. Dao, N. Fernandez, B. Ballester, J.C. Andrau, S. Spicuglia, High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq, Nat. Commun. 6 (2015) 6905, http://dx.doi.org/10.1038/ncomms7905.