



Published in final edited form as:

Cell. 2016 November 17; 167(5): 1163–1166. doi:10.1016/j.cell.2016.10.054.

## The Ties That Bind: Mapping the Dynamic Enhancer-Promoter Interactome

Cailyn H Spurrell<sup>1</sup>, Diane E Dickel<sup>1</sup>, and Axel Visel<sup>1,2,3,#</sup>

<sup>1</sup> MS 84-171, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

<sup>2</sup> U.S. Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA.

<sup>3</sup> School of Natural Sciences, University of California, Merced, CA 95343, USA.

### Abstract

Coupling chromosome conformation capture to molecular enrichment for promoter-containing DNA fragments enables the systematic mapping of interactions between individual distal regulatory sequences and their target genes. In this Minireview, we describe recent progress in the application of this technique and related complementary approaches to gain insight into the lineage- and cell type-specific dynamics of interactions between regulators and gene promoters.

Distal regulatory elements, such as enhancers, play a central role in controlling expression in mammalian genomes. Enhancer sequences act as substrates for binding of tissue-specific transcription factors and drive transcription through physical interaction with gene promoters (Spitz and Furlong, 2012). Recent chromatin profiling studies reveal the exceptional cell type and temporal specificity of enhancer activity, which exceeds that of other classes of gene regulatory sequences (Ernst and Kellis, 2010; Nord et al., 2013). This stunning specificity, alongside advances in sequencing technologies and the increasingly recognized importance of non-coding sequences in human development and disease, have driven large-scale efforts to annotate regulatory elements and gene transcription in the human genome under a wide variety of conditions. The International Human Epigenome Consortium (IHEC) (Bae, 2013) connects many of these projects, with the goal of characterizing 1000 epigenomes from different human cell types at diverse developmental stages and disease states.

New studies, published in this issue of *Cell* and in *Cell Reports* and described in more detail throughout the following sections of this Minireview, build upon IHEC efforts to explore the role of cell type-specific regulation and begin to address several important challenges in the field (Schmitt et al., 2016; Javierre et al., 2016; Breeze et al., 2016; Pellacani et al., 2016). Briefly, Pellacani *et al.* (2016) tackle the question of cell type specificity of enhancers across the individual cell types that make up heterogeneous tissues. The authors use chromatin

# Correspondence should be addressed to: [AVisel@lbl.gov](mailto:AVisel@lbl.gov).

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

profiling methods to identify regulatory elements active in the distinct cell populations that comprise mammary tissue. While chromatin profiling is powerful for identifying predicted enhancer sequences, it is limited in its ability to elucidate the gene target(s) of the predicted enhancers. To address this challenge, Javierre *et al.* (2016) and Schmitt *et al.* (2016) use cutting-edge chromosome conformation capture techniques to map enhancer-promoter interactions in a variety of human tissues and primary cell types. Finally, disease-associated variants identified in genome-wide association studies (GWAS) are overwhelmingly non-coding (Altshuler *et al.*, 2010; Visel *et al.*, 2009) and enriched in non-coding loci harboring regulatory functions (Maurano *et al.*, 2012), but specific examples of non-coding sequence variants conclusively and mechanistically linked to disease remain limited. The functional genome annotations from the series of new papers (Schmitt *et al.*, 2016; Javierre *et al.*, 2016; Pellacani *et al.*, 2016) along with a computational algorithm capable of integrating epigenomic findings described in Breeze *et al.* (2016) provide handy tools for addressing the gap between disease-associated non-coding variants and their regulatory gene targets. Using these complementary techniques to explore the regulatory landscape in human tissues and isolated primary cell populations, these studies report insights and resources that will be instrumental in linking variants with causal mechanisms of disease.

### Insights into Cell Type-Specific Regulation

Histone ChIP-seq has now become a standard method to identify regulatory regions genome-wide (Park, 2009). ChIP-seq combines chromatin immunoprecipitation of histone modifications with high-throughput sequencing to identify active enhancers and other regulatory features. While the underlying DNA sequence does not vary between cell types, histone modifications mark regions that are active or repressed *in vivo* in a tissue-specific manner. When paired with technologies for capturing specific cell types, ChIP-seq can be used to identify differential regulation in cell populations derived from heterogeneous tissue. An elegant example of this approach is provided by Pellacani *et al.* (2016), who generate histone ChIP-seq, DNA methylation and gene expression data to identify cell type-specific regulatory elements in primary human mammary tissue. Consistent with previous findings (Gascard *et al.*, 2015), their results show widespread differences among the different cell types isolated from this heterogeneous tissue and relative to previous results from immortalized mammary cell lines. The biological relevance of these observations is reinforced by the findings that differential enhancer utilization in mammary cell types is consistent with cell-specific gene expression and that cell type-specific enhancers are enriched for unique transcription factor binding sites. This high-resolution developmental specificity of enhancer activity mirrors results from previous chromatin profiling studies, and these data allow the authors to derive insights into the cells that make up a complex tissue.

### 3D Chromatin Structure Links Enhancers to Genes

While ChIP-seq can identify differential activity of regulatory elements across tissues and cell types, it does not provide evidence that formally links individual distal regulatory elements to their respective target genes. Tools based on Chromosome Conformation Capture (3C) enable the identification of genomic regions that can be far apart in the linear

genome sequence but are proximate in three-dimensional space within the nucleus. Hi-C, one variant of 3C, identifies these distal yet interacting partners on a global genomic scale by digesting cross-linked chromatin and ligating physically interacting fragments together (Lieberman-Aiden et al., 2009). The resulting libraries are sequenced without further molecular enrichment for marks associated with any particular functional class of genomic elements, thereby creating a largely unbiased genome-wide map of chromatin architecture. Because the spatial resolution along the linear genome correlates with sequencing depth, the high complexity of these libraries requires deep sequencing to identify statistically significant interactions. Thus, the approach was initially used to identify megabase-scale Topologically Associated Domains (TADs) of chromosome organization (Dixon et al., 2012). This high level architecture tends to be conserved across cell types and mammalian species, but the library complexity masks intra-TAD variation and less robust interactions. Efforts to create higher resolution maps of chromatin interactions require an order of magnitude more sequencing but are able to detect smaller conserved domains (Rao et al., 2014).

A new paper by Schmitt *et al.* reports traditional Hi-C on 14 primary human tissues and describes computational methods to identify new features of genomic architecture. The authors designed an algorithm to normalize sequencing depth variation across tissues, which allows them to identify both TADs and cell-specific interactions. Consistent with the results from previous cell-based studies, the authors observed that TAD structure is stable across different human tissues. Beyond the resolution of TADs, however, high resolution chromatin loops have been described to partition the genome into smaller domains within the TAD structure (Rao et al., 2014). Reinforcing these previous observations, a subset of the interactions reported by Schmitt *et al.* represent a distinct set of such sub-TAD regulatory networks. The chromatin interactions within TADs show a remarkable degree of tissue specificity; approximately 40% of interactions are unique to one tissue type. These tissue-specific interaction regions tend to be located near genes with tissue-specific expression, and they are enriched for marks of active enhancers. These findings can begin to be used to directly link genes with some of their non-coding regulatory elements and further demonstrate the diverse regulatory landscape across human tissues.

A second paper, by Javierre *et al.* (2016), defines even more specific chromatin interaction architecture using a variant of Hi-C that employs RNA oligonucleotides to enrich for interactions involving promoter sequences (Schoenfelder et al., 2015). This Promoter Capture Hi-C (PCHi-C) technology results in libraries with far lower complexity than standard Hi-C, greatly reducing the amount of sequencing required, and resulting in high-resolution maps showing interactions between promoters and other loci. Javierre *et al.* applied this method to 17 primary human cell types from the hematopoietic lineage to further characterize the types of loci that interact with promoters and to understand how long-range interactions between promoters and other loci evolve during cell differentiation.

The observed interactions anchored on promoters span a median distance of ~300 kb, and the distal interacting partners do not always link to the closest gene by linear distance. Consistent with the Schmitt *et al.* (2016) study, these distal regions identified as interacting with promoters are enriched for chromatin marks associated with active enhancers. Javierre

*et al.* (2016) further investigate the biological role of promoter-interacting regions by comparing them to previously reported expression quantitative trait loci (eQTLs). Expression QTLs are identified by measuring gene expression in a population of cells and linking expression differences to alleles of a sequence variant (Cookson et al., 2009). Using published eQTL data from several cell types, the authors observe an enrichment for eQTLs in the promoter-interacting regions from the same cell types. In particular, distal regions are enriched for eQTLs that associate with the same interacting gene. This result supports that promoter-interacting regions have a functional regulatory role and that variation within promoter-interacting regions can be connected to potential gene targets.

One important finding from Javierre *et al.* (2016) is that in the hematopoietic lineage, chromatin architecture is highly dynamic, and lineage-specific interactions delineate the myeloid and lymphoid regulatory landscape. The regulatory complexities of the promoter-interacting regions are schematically outlined in **Figure 1**. The first column is an example of an invariant interaction between a single promoter and multiple enhancers across all cell types. While invariant interactions are abundant, many interactions vary by cell type. Clustering the promoter-enhancer interactions shows a general divergence between interactions found in the myeloid and lymphoid lineages. Schematic examples of myeloid- and lymphoid-specific interactions are represented in columns 2 and 3 of **Figure 1**. These interactions are invariant within each lineage but divergent between the two cell lineages. Column 4 shows a CD4+ T cell-specific interaction, representative of cell type-specific interactions, which were also observed in other individual cell types examined. Surprisingly, approximately 80% of promoters had lineage- or cell type- specific interactions. Further showing the complexity of the regulatory network, in cells of the myeloid and lymphoid lineages the same promoter may be regulated through different enhancer interactions (column 5), and one enhancer can interact with different promoters in a lineage-specific manner (column 6). Javierre *et al.* (2016) cluster these highly specific interactions to create a detailed lineage tree of all 17 hematopoietic cell types that recapitulates the known relationships between different cell populations. Consistent with this, promoter-associated enhancers are predicted to be active in a manner that mirrors the cell type specificity of expression of the interacting gene. The authors combined their chromatin interaction data with enhancer annotations and clustered genes according to enhancer specificity for each cell type. This analysis identifies sets of genes that are dynamically regulated in different cell types across the hematopoietic tree. The correlation between cell type-specific enhancer activity and gene expression supports a functional role for these interactions in regulating cell fate and differentiation.

## Interpretation of Genetic and Epigenetic Variation in Disease

Elucidating the mechanistic role of non-coding sequence variation in human disease remains an unmet challenge. Tissue- and cell type-specific annotations of regulatory elements generated by CHIP-seq are now widely available through the work of the IHEC members and individual investigators. These efforts represent an important first step in bridging this gap, and work is now being done to integrate these diverse maps together into high-confidence enhancer annotations to identify which disease-associated variants are most likely to impact gene regulatory sequences (Dickel et al., 2016). Chromosome conformation capture

techniques complement these datasets by linking tissue-specific enhancers with candidate gene targets, and such approaches are increasingly being used to interpret non-coding disease-associated variation (Martin et al., 2015; Won et al., 2016). Most studies thus far have focused on one specific cell type or tissue to prioritize GWAS variants. In contrast, Javierre *et al.* (2016) and Schmitt *et al.* (2016) analyze genome interactions across many tissue types or cell populations, increasing the specificity of the regulatory candidates. These elegant papers show that lineage- and cell type-specific regulatory regions are enriched for genetic variation from association studies of phenotypes with similar cell specificity. Javierre *et al.* (2016) also use lineage-specific interactions elucidated by PCHi-C to create a prioritized list of genes that may be implicated in disease through interactions with disease-associated non-coding regions identified by GWAS. Their analysis combines genome interaction data with GWAS results to elucidate and prioritize candidate genes and pathways that may underlie human phenotypes. One type of interaction diagrammed in **Figure 1** is “lineage-specific promoter interactions”. Hypothetically, the presence of a phenotype-associated variant in an enhancer that interacts with two promoters in a relevant cell lineage would prioritize these genes over other nearby genes, thereby helping to narrow down the list of genes whose misregulation might underlie the phenotype. Javierre *et al.* (2016) outline how this strategy based on PCHi-C data can be used to complement eQTL-based approaches, which require variants to have detectable effects on gene expression in order to link a regulatory sequence to a target gene (Guo et al., 2015). Their results highlight the strength of using physical interaction data to link disease-relevant genes and enhancers.

Complementary to GWAS, epigenome-wide association studies (EWAS) identify changes in the epigenome that are associated with disease susceptibility. For example, previous EWAS studies have found associations between specific changes in DNA methylation and phenotypic status (Liu et al., 2013). Building upon the success of the FORGE software (Dunham et al., 2014), which intersects GWAS results with maps of DNase hypersensitive sites to determine which disease-associated variants fall into regulatory sequences, a new paper (Breeze *et al.*, 2016) describes eFORGE, software designed to perform similar analyses for EWAS results. The new tool maps regions of differential methylation that have been implicated in disease through EWAS to regulatory regions genome-wide. Thus, eFORGE identifies potential mechanistic links between cell type-specific distal regulation and epigenome-wide association studies, information which could aid in the development of disease treatments.

The compelling new studies presented here use epigenomic data to assess the regulatory architecture across an impressive range of primary human cells and tissues. Their findings emphasize the cell type specificity of regulatory interactions and the dynamic nature of the regulatory networks, and this information will be valuable for the interpretation of human disease findings. While this Minireview focused on assessing non-coding variants from GWAS, cell-type specific interactions can also be used to interpret rare non-coding variation from whole genome sequencing studies (Weedon et al., 2014), a technology that is being adopted with increasing frequency for human disease studies. The computational and experimental resources from these epigenomic studies will be valuable for understanding

chromatin structure, as well as for facing the considerable challenge of linking non-coding variation with cell-specific mechanisms of disease.

## Acknowledgments

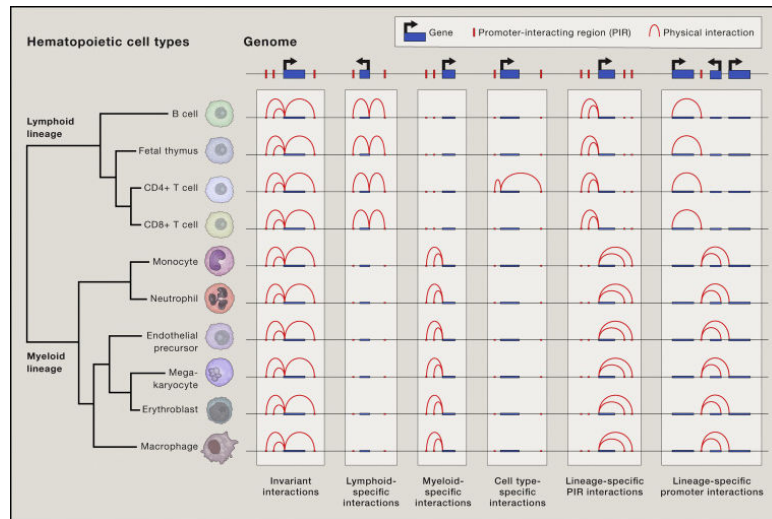
This work was supported by National Institutes of Health grants R01HG003988, U54HG006997, U01DE024427, R24HL123879, and UM1HL098166. Research conducted at the E.O. Lawrence Berkeley National Laboratory was performed under Department of Energy Contract DE-AC02-05CH11231, University of California.

## References

[Copy editor: Please confirm that citations of papers in press at Cell and Cell Reports correspond to final accepted versions, author lists were taken from pre-final versions]

- Altshuler D, Lander E, Ambrogio L. *Nature*. 2010; 476:1061–1073.
- Anthony D, Schmitt, Hu Ming, Jung Inkyung, Xu Zheng, Qiu Yunjiang, Tan Catherine L. Li Yun, Barr Cathy L. *B.R. Cell*. 2016
- Bae J-B. *Genomics Inform*. 2013; 11:7–14. [PubMed: 23613677]
- Javierre, Biola M.; Burren, Oliver S.; Wilder, Steven P.; Kreuzhuber, R.; Hill, Steven M.; Sewitz, Sven; Cairns, Jonathan; Wingett, S.W.; Várnai, Csilla; Thiecke, Michiel J.; Burden, Frances; Farrow, S.; Cutler, Antony J.; Rehnstrom, Karola; Downes, Kate; Grassi, L.; Kostadima, Myrto; Freire-Pritchett I, Paula; Fan Wang, T.; BLUEPRINT Consortium. Stunnenberg, Hendrik G.; Todd, John A.; Zerbino, DR.; Stegle, Oliver; Ouwehand, Willem H.; Frontini, Mattia; C.; Wallace, MS.; P.F. *Cell*. 2016
- Breeze, Charles E.; Paul, Dirk S.; van Dongen, Jenny; Butcher, Lee M.; Ambrose, John C.; Barrett, James E.; Lowe, Robert; Rakyán, Vardhman K.; Iotchkova, Valentina; Frontini, Mattia; Downes, Kate; Ouwehand, Willem H.; Laperle, Jonathan; Jacques, Pierre-Étienne; Guilla, S.B.. *Cell Reports*. 2016
- Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M. *Nat. Rev. Genet*. 2009; 10:184–194. [PubMed: 19223927]
- Pellacani, Davide; Bilenky, Misha; Kannan, Nagarajan; Heravi-Moussavi, Alireza; Knapp, David J.H.F.; Gakkhar, Sitanshu; Moksa, Michelle; Carles, Annaick; Moore, Richard; Mungall, Andrew J.; Marra, Marco A.; Jones, Steven J.M.; Aparicio, Samuel; Martin Hirst, CJE. *Cell Reports*. 2016
- Dickel DE, Barozzi I, Zhu Y, Fukuda-Yuzawa Y, Osterwalder M, Mannion BJ, May D, Spurrell CH, Plajzer-Frick I, Pickle CS, et al. *Nat. Commun*. 2016; 7:12923. [PubMed: 27703156]
- Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. *Nature*. 2012; 485:376–380. [PubMed: 22495300]
- Dunham I, Kulesha E, Iotchkova V, Morganella S, Birney E. *bioRxiv*. 2014
- Ernst J, Kellis M. *Nat. Biotechnol*. 2010; 28:817–825. [PubMed: 20657582]
- Gascard P, Bilenky M, Sigaroudinia M, Zhao J, Li L, Carles A, Delaney A, Tam A, Kamoh B, Cho S, et al. *Nat. Commun*. 2015; 6:6351. [PubMed: 25690954]
- Guo H, Fortune MD, Burren OS, Schofield E, Todd JA, Wallace C. *Hum. Mol. Genet*. 2015; 24:3305–3313. [PubMed: 25743184]
- Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. *Science* (80-. ). 2009:326.
- Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, Reinius L, Acevedo N, Taub M, Ronninger M, et al. *Nat. Biotechnol*. 2013; 31:142–147. [PubMed: 23334450]
- Martin P, McGovern A, Orozco G, Duffus K, Yarwood A, Schoenfelder S, Cooper NJ, Barton A, Wallace C, Fraser P, et al. *Nat. Commun*. 2015; 6:10069. [PubMed: 26616563]
- Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, et al. *Science*. 2012; 337:1190–1195. [PubMed: 22955828]
- Nord AS, Blow MJ, Attanasio C, Akiyama JA, Holt A, Hosseini R, Phouanavong S, Plajzer-Frick I, Shoukry M, Afzal V, et al. *Cell*. 2013; 155:1521–1531. [PubMed: 24360275]

- Park PJ. *Nat. Rev. Genet.* 2009; 10:669–680. [PubMed: 19736561]
- Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, et al. *Cell.* 2014; 159:1665–1680. [PubMed: 25497547]
- Schoenfelder S, Furlan-Magaril M, Mifsud B, Tavares-Cadete F, Sugar R, Javierre B-M, Nagano T, Katsman Y, Sakthidevi M, Wingett SW, et al. *Genome Res.* 2015; 25:582–597. [PubMed: 25752748]
- Spitz F, Furlong EEM. *Nat. Rev. Genet.* 2012; 13:613–626. [PubMed: 22868264]
- Visel A, Rubin EM, Pennacchio LA. *Nature.* 2009; 461:199–205. [PubMed: 19741700]
- Weedon MN, Cebola I, Patch A-M, Flanagan SE, De Franco E, Caswell R, Rodríguez-Seguí SA, Shaw-Smith C, Cho CH-H, Lango Allen H, et al. *Nat. Genet.* 2014; 46:61–64. [PubMed: 24212882]
- Won H, de la Torre-Ubieta L, Stein JL, Parikshak NN, Huang J, Opland CK, Gandal MJ, Sutton GJ, Hormozdiari F, Lu D, et al. *Nature.* 2016



**Figure 1. Lineage-specific interactions of promoters with non-coding regulatory sequences**  
 Left: Javierre *et al.* (2016) used promoter capture Hi-C (PCHi-C) to systematically map interactions between promoters and promoter-interacting regions (PIRs) across 17 primary human hematopoietic cell types (left, 10 representative examples shown). Right: Comparison of PIR-promoter interactions across cell types reveals that some interactions are invariant across all cell types examined, while others are specific to major lineages or individual cell types. Importantly, some promoters interact with different sets of PIRs depending on cell type and, vice versa, some PIRs interact with different sets of promoters in a cell-type-specific manner.