A Long-Distance Chromatin Affair

Annette Denker[1] and Wouter de Laat[1,*]
[1]Hubrecht Institute-KNAW and University Medical Center Utrecht, 3584 CT Utrecht, the Netherlands
*Correspondence: w.delaat@hubrecht.eu
http://dx.doi.org/10.1016/j.cell.2015.08.022

Changes in transcription factor binding sequences result in correlated changes in chromatin composition locally and at sites hundreds of kilobases away. New studies demonstrate that this concordance is mediated via spatial chromatin interactions that constitute regulatory modules of the human genome.

The majority of disease-causing genetic variations occur in non-coding regulatory sequences that presumably control the transcriptional output of target genes. Such enhancer regions and their encompassed transcription factor (TF) binding sites display high variability among individuals. Disease-associated sequence variations (quantitative trait loci; QTL) alter local chromatin states and affect binding of transcription factors, DNase I hypersensitivity (DHS), nucleosome positioning, histone modifications, and, ultimately, enhancer activity (McVicker et al., 2013; Kasowski et al., 2013; Kilpinen et al., 2013). Surprisingly, such correlated changes in chromatin state are not limited to the local environment but can, in some cases, affect loci up to 200 kb distal to the QTL (McVicker et al., 2013; Kilpinen et al., 2013). Previous work has suggested that chromatin architecture may be involved in this phenomenon (e.g., McVicker et al., 2013); however, definitive proof and functional understanding of this process had been missing. In this issue, Waszak et al. (2015) and Grubert et al. (2015) explain such coordinated chromatin variability in light of the three-dimensional (3D) organization of our genome.

The authors use human lymphoblastoid cell lines to generate ChIP-seq profiles for the histone modifications H3K4me3 (demarcating promoters), H3K4me1 (enhancers), and H3K27ac (promoters and enhancers), as well as for the regulatory TF PU.1 and RNA polymerase II, which they integrate with DHS and gene expression data. Both studies demonstrate that distinct, and often disease-relevant, genetic variations, especially at the level of TF motifs, can serve as QTLs, causing local and distal allelic variation in histone marks, chromatin accessibility, and/or gene expression. Subsequently, they compare the observed molecular associations with previously published (Rao et al., 2014) and newly generated chromatin conformation data (Grubert et al., 2015) to show that long-range genetic regulation of chromatin variation often involves specific 3D contacts between pairs of regulatory modules (Figure 1). Regions of coordinated chromatin variation thereby form an intricate network of enhancer-enhancer, enhancer-promoter, and promoter-promoter interactions, which are spatially organized into "variable chromatin modules" (VCMs; Waszak et al., 2015) up to several hundred kilobases in size. Chromatin QTLs have weaker effects on distal interaction partners than on local sequences, consistent with the idea that distal interactions, as observed by chromatin conformation capture techniques, represent transient events.

How do these regulatory micro-environments fit into our current understanding of chromatin organization? Recent studies on genome-wide chromatin topology have revealed the existence of megabase-sized "topologically associated domains" (TADs)—chromosomal regions within which sequences preferentially contact each other. Most TADs appear conserved across cell types and species and can be further divided into topological subdomains that show a median size of nearly 200 kb. These chromosomal domains have been described to display distinct patterns of histone marks (Rao et al., 2014) and exert unique regulatory activity (Symmons et al., 2014), indicating that they demarcate not only spatial but also functional entities. Waszak et al. and Grubert et al. now show that genetic control of chromatin states occurs within TADs and their smaller subdomains.

Both studies also report that histone QTLs (hQTLs; affecting the local and distal chromatin state) are enriched in common (auto-)immune disease variants, consistent with the cell type under study, a finding that emphasizes the medical relevance of this phenomenon. Consequentially, Grubert et al. and Waszak et al. propose using chromatin QTL mapping in genome-wide association studies (GWAS) to help identify putative target genes of disease-associated variants. It is still unclear whether many disease-associated single-nucleotide polymorphisms in non-coding DNA are disease causative ("drivers") or merely act as "passengers" and how, exactly, they alter genome functioning. Paired chromatin variation between an enhancer near or at a GWAS hit and a distal promoter may help identify the disease-relevant gene and elucidate the regulatory network driving its expression.

While these findings will certainly launch new possibilities to assign function to the collection of GWAS hits, which remains largely descriptive, both papers show that concerted variation in chromatin states between distal sites is linked to, and probably caused by, their spatial interactions. The advantage of chromatin QTL mapping over analysis of enhancer-promoter interactions by chromosome conformation capture techniques therefore remains to be determined. Several recent papers have demonstrated the usefulness of high-resolution chromatin contact maps to link human GWAS variants to target genes (e.g., Dixon et al., 2015). The ever increasing resolution of Hi-C maps can serve to first identify the (sub-)TAD encompassing the risk variant. The genes co-occupying this domain are prime candidate target genes. Analysis of chromatin loops formed by their gene promoters

CrossMark

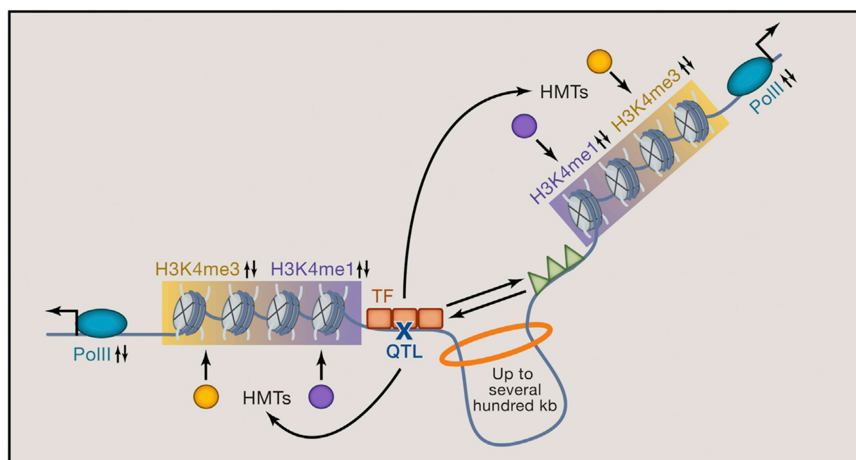

Figure 1. Genomic Variants Alter Local and Distal Chromatin
A change in a transcription factor (TF) binding site (quantitative trait locus [QTL] indicated by a blue cross; TF shown in red) can result in altered affinities and consequentially dynamic changes in the local chromatin composition—for example, by affecting the rate of recruitment of modifying enzymes such as histone methyltransferases (HMTs). (Depicted here are changes in H3K4 mono- and trimethylation; however H3K27ac, DNase I hypersensitivity, and recruitment of further TFs can also be affected.) Eventually, altered chromatin states can enhance or reduce RNA polymerase II (Pol II) recruitment to nearby promoters and therefore affect transcriptional output. Grubert et al. (2015) and Waszak et al. (2015) demonstrate that QTLs can also affect chromatin composition concordantly at distal but spatially interacting genomic sequences, thereby encompassing subdomains of correlated histone marks at the level of tens to hundreds of kilobases. A hypothetical cohesin-mediated loop (orange) is shown with potential interacting proteins (green).

with the regulatory modules at or near the GWAS hit may subsequently enable identification of disease-relevant genes.

As the authors propose, hQTL mapping can also reveal the cell type implicated in disease, a first step often needed to uncover relevant target genes. GWAS benefit from the tissue invariance of inherited genetic variation, but disease-associated SNPs identified in genetic material of white blood cells, for instance, will often be functional only in a given other tissue and possibly at another stage of development. Their chromatin makeup is not tissue invariant and can likely reveal the cell type in which they exert their action, in which case distal chromatin QTL mapping may uncover the linked, disease-relevant gene.

Practically, this seems an ambitious enterprise, as it would require the isolation of relatively pure populations of a plethora of cell types from at least dozens (the current two studies) and probably even more individuals, which could become even more challenging in the investigation of developmental diseases. In the future, systematic analysis of organoids derived from many individuals may enable creation of an hQTL reference database for various tissues and cell types. However, to cause phenotypic variability, sequence alterations are expected to affect target gene expression, and eQTL (variants affecting gene expression) data sets of these cell types may therefore prove more meaningful. A combination of eQTL and high-resolution chromatin contact maps across tissues might eventually be the most powerful strategy to identify GWAS target genes. Recently, 55%–75% of chromatin loops have been found to be conserved between cell types (Rao et al., 2014), and further analyses in primary cells will likely refine this estimate. However, if the majority of enhancer-promoter loops are indeed pre-formed, or "permissive," (existing across multiple tissues), rather than formed de novo upon "instructive" tissue-specific cues (de Laat and Duboule, 2013), it might not be necessary to create contact maps of every single cell type in order to identify putative target genes contacting risk variants.

Regardless of whether or not chromatin QTLs prove instrumental in detecting new disease genes and networks, the structural micro-environments described in these two studies provide an explanation for regulatory variability in the absence of proximal sequence change. The genetic status of one regulatory module can affect the chromatin state of a proximal module, as it has recently been demonstrated for Polycomb domains in *Drosophila*, where recruitment of Polycomb group proteins to binding sites, even if weak, is substantially enhanced if they are linearly close to other strong Polycomb binding sites (Schuettengruber et al., 2014). The studies by Waszak et al. and Grubert et al. now demonstrate that such cooperative effects are also at play at distant but contacting genomic loci and represent a new regulatory dimension of complementary chromatin states.

ACKNOWLEDGMENTS

A.D. is supported by a Long-Term Postdoctoral Fellowship of the Human Frontier Science Program (LT000709/2013). W.d.L. is supported by grants from the Dutch Scientific Organization (NWO) (91204082 and 935170621) and a EU grant 2010-259743 (MODHEP). W.d.L. is a founder and SAB member of Cergentis (the Netherlands).

REFERENCES

de Laat, W., and Duboule, D. (2013). Nature *502*, 499–506.

Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., et al. (2015). Nature *518*, 331–336.

Grubert, F., Zaugg, J.B., Kasowski, M., Ursu, O., Spacek, D.V., Martin, A.R., Greenside, P., Srivas, R., Phanstiel, D.H., Pekowska, A., et al. (2015). Cell *162*, this issue, 1051–1065.

Kasowski, M., Kyriazopoulou-Panagiotopoulou, S., Grubert, F., Zaugg, J.B., Kundaje, A., Liu, Y., Boyle, A.P., Zhang, Q.C., Zakharia, F., Spacek, D.V., et al. (2013). Science *342*, 750–752.

Kilpinen, H., Waszak, S.M., Gschwind, A.R., Raghav, S.K., Witwicki, R.M., Orioli, A., Migliavacca, E., Wiederkehr, M., Gutierrez-Arcelus, M., Panousis, N.I., et al. (2013). Science *342*, 744–747.

McVicker, G., van de Geijn, B., Degner, J.F., Cain, C.E., Banovich, N.E., Raj, A., Lewellen, N., Myrthil, M., Gilad, Y., and Pritchard, J.K. (2013). Science *342*, 747–749.

Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). Cell *159*, 1665–1680.

Schuettengruber, B., Oded Elkayam, N., Sexton, T., Entrevan, M., Stern, S., Thomas, A., Yaffe, E., Parrinello, H., Tanay, A., and Cavalli, G. (2014). Cell Rep. *9*, 219–233.

Symmons, O., Uslu, V.V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., Ettwiller, L., and Spitz, F. (2014). Genome Res. *24*, 390–400.

Waszak, S.M., Delaneau, O., Gschwind, A.R., Kilpinen, H., Raghav, S.K., Witwicki, R.M., Orioli, A., Wiederkehr, M., Panousis, N.I., Yurovsky, A., et al. (2015). Cell *162*, this issue, 1039–1050.