Nature Reviews Genetics | Published online 13 Dec 2016; doi:10.1038/nrg.2016.161

EPIGENOMICS

Deciphering non-coding variation with epigenomics

The majority of disease-associated genetic variants identified by genome-wide association studies (GWAS) are in non-coding regions of the genome. This indicates a pathological regulatory role for these variants, but identifying mechanisms of action and their corresponding target genes is often challenging. As part of the International Human Epigenomics Consortium (IHEC), recent studies show how profiling chromosomal interactions and other epigenomic features can provide functional insights into the roles of non-coding variants in disease.

Non-coding regulatory variants often affect enhancer regions, which typically regulate target genes through the formation of chromosomal loops. As enhancer–promoter pairs can be many megabases apart in the primary DNA sequence, predicting the target genes of enhancers based on the closest gene in the primary DNA sequence is suboptimal; hence, strategies are increasingly using experimental characterization of chromosomal contacts, as reported by Javierre *et al.* and Schmitt *et al.*

Javierre *et al.* profiled 17 human haematopoietic cell types using promoter capture Hi-C (PCHi-C), which is based on the Hi-C method for high-throughput characterization of physical chromosomal interactions, but enriches for known promoter sequences so that sequencing capacity is focused on promoter-interacting regions (PIRs). On average, they identified ~175,000 high-confidence interactions per cell type, which comprised a combination of cell-invariant contacts and variable contacts that are likely to regulate cell-type-specific genes during haematopoietic differentiation.

The authors integrated the data with histone modification, chromatin accessibility and gene expression data for a more holistic understanding. As an indication of the functional relevance of the identified PIRs, many were in annotated enhancers for which, in a given cell type, the occurrence of classic histone modifications of enhancer activity and contact with a promoter is correlated with gene activation from that promoter. Interestingly, however, there was some evidence of gene-regulatory activity of PIRs that lack enhancer annotations or typical histone modifications, indicating that these might be a new type of unexplored regulatory element that may need to be considered when genetic variants occur in them.

Javierre *et al.* then used their data sets to interpret non-coding variants from GWAS. They found that haematopoietic PIRs were enriched for single-nucleotide polymorphisms (SNPs) associated with blood-relevant traits and diseases, but not non-blood traits. This validates the physiological relevance and utility of the data sets but indicates that PCHi-C data from other tissues will be needed for broader application to diverse diseases.

The authors integrated PCHi-C data into a bioinformatic pipeline to prioritize likely target genes of disease-associated SNPs and applied it to various autoimmune diseases including rheumatoid arthritis, systemic lupus erythematosus, Crohn's disease and ulcerative colitis. They prioritized >2,500 potential disease-associated genes. Three-quarters of these genes were not previously implicated in disease, including genes regulated by long-range interactions, which would have been missed by methods relying on proximity within the primary DNA sequence. Finally, the authors assembled their target genes into an 'autoimmunity network' for further analysis by the community.

In their study, Schmitt *et al.* carried out Hi-C across 21 diverse primary human tissues and cell types. Analysing an average of 214 million unique chromosome contacts per tissue type, they noticed that some regions displayed particularly high local contact frequencies, which they termed frequently interacting regions (FIREs). FIREs were distinct from previously defined types of chromosome domains such as A/B compartments, topologically associated domains (TADs) and loops, although in general they occurred towards the centre of TADs, partook in numerous intra-TAD interactions and were contained within broader regions of A-compartment active chromatin. Further analyses, including integration with profiles of histone modifications and transcription, revealed that FIREs are highly tissue-type-dependent, frequently occur near (and transcriptionally regulate) cell-identity genes and overlap substantially with chromatin features of active enhancers. Indeed, the overlap was particularly strong for clustered FIREs ('super-FIREs'), of which almost 100% contained clustered enhancers ('super-enhancers') or standard enhancers.

Given the likely gene-regulatory activity of FIREs, Schmitt *et al.* assessed disease relevance, finding that FIREs are enriched for SNPs associated with diseases that affect the particular cell types examined. Analysing pairs of FIREs allowed disease-relevant SNPs to be linked to known and novel target genes.

It will be interesting to further explore the value of FIREs in disease genetic studies to determine, for example, whether the typically short-range (<200 kb) nature of FIRE contacts will allow longer-range regulatory events to be routinely uncovered. However, FIREs might be particularly valuable for studying developmentally dynamic enhancers that regulate multiple target genes.

Overall, the studies provide insights into the mechanisms by which non-coding variants regulate target genes and, ultimately, disease phenotypes. They also provide data sets and methodologies for future mining and adoption. The papers represent just a subset of the ~40 IHEC studies that were coordinately published in Cell Press and other journals, and are available online.

Darren J. Burgess

ORIGINAL ARTICLES Javierre, B. M. *et al.* Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* **167**, 1369–1384 (2016) | Schmitt, A. D. *et al.* A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.* **17**, 2042–2059 (2016)

FURTHER READING Bonev. B. & Cavalli, G. Organization and function of the 3D genome. *Nat. Rev. Genet.* **17**, 661–678 (2016) | Stricker, S. H., Köferle, A. & Beck, S. From profiles to function in epigenomics. *Nat. Rev. Genet.* http://dx.doi.org/10.1038/nrg.2016.138 (2016)