3D genome organization links non-coding disease-associated variants to genes

Gisela Orozco[1,2], Stefan Schoenfelder[3,4], Nicolas Walker[3], Stephan Eyre[1,2] and Peter Fraser[3,5]*

[1]Centre for Genetics and Genomics Versus Arthritis, Division of Musculoskeletal and Dermatological Sciences, School of Biological Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester Academic Health Science Centre, Manchester, United Kingdom, [2]NIHR Manchester Biomedical Research Centre, Manchester University Foundation Trust, Manchester, United Kingdom, [3]Enhanc3D Genomics Ltd., Cambridge, United Kingdom, [4]Epigenetics Programme, The Babraham Institute, Babraham Research Campus, CB22 3AT Cambridge, Cambridge, United Kingdom, [5]Department of Biological Science, Florida State University, Tallahassee, FL, United States

Genome sequencing has revealed over 300 million genetic variations in human populations. Over 90% of variants are single nucleotide polymorphisms (SNPs), the remainder include short deletions or insertions, and small numbers of structural variants. Hundreds of thousands of these variants have been associated with specific phenotypic traits and diseases through genome wide association studies which link significant differences in variant frequencies with specific phenotypes among large groups of individuals. Only 5% of disease-associated SNPs are located in gene coding sequences, with the potential to disrupt gene expression or alter of the function of encoded proteins. The remaining 95% of disease-associated SNPs are located in non-coding DNA sequences which make up 98% of the genome. The role of non-coding, disease-associated SNPs, many of which are located at considerable distances from any gene, was at first a mystery until the discovery that gene promoters regularly interact with distal regulatory elements to control gene expression. Disease-associated SNPs are enriched at the millions of gene regulatory elements that are dispersed throughout the non-coding sequences of the genome, suggesting they function as gene regulation variants. Assigning specific regulatory elements to the genes they control is not straightforward since they can be millions of base pairs apart. In this review we describe how understanding 3D genome organization can identify specific interactions between gene promoters and distal regulatory elements and how 3D genomics can link disease-associated SNPs to their target genes. Understanding which gene or genes contribute to a specific disease is the first step in designing rational therapeutic interventions.

KEYWORDS

3D genomics, GWAS, gene expression, SNP, causality, promoter-capture Hi-C

1 Introduction

1.1 The 3D genome regulates gene expression

1.1.1 Gene expression depends on folding and proximity

Cell type-specific gene expression depends upon promoters, enhancers, silencers, insulators and other gene regulatory elements that determine when genes are switched on, to what level they are expressed, and when they are switched off (Schoenfelder and Fraser, 2019). These regulatory sequences work in concert with trans-acting factors that bind to them to control the flow of genetic information from DNA to RNA to protein, thereby directing developmental, and differentiation cell fate decisions, as well as the maintenance of homeostasis and health. Understanding which regulatory sequences control which gene(s) requires an intricate knowledge of the 3D folding and arrangement of the genome in the cell nucleus, since all these elements carry out their functions through close proximity, direct interaction or contact. For example, enhancer sequences, which can activate or increase expression of a gene, can be located upstream, downstream or in the intron of a gene, and can operate over very large genomic distances to control the expression of a gene millions of base pairs away. Analysis of the sequences in the immediate spatial vicinity of a transcriptionally active gene in the nucleus revealed very close proximity of its distal enhancers, while the intervening sequences were located further away (Carter et al., 2002). This suggests formation of a loop, where regulatory activity of the enhancer is communicated to its target gene through direct interaction with its promoter (Figure 1) (Carter et al., 2002). "Engineered looping" between a gene promoter and distal enhancer sequence, by expression of factors that bind to each and can interact or dimerize, showed that transcription is indeed controlled by the enhancer-promoter loop, mediated by direct contact between factors bound to the regulatory sequences (Figure 1) (Deng et al., 2012; Morgan et al., 2017; Kim et al., 2019).

Enhancers have been mapped throughout the genome in a variety of cell types, using histone marks and cofactor binding. The human genome is thought to contain around one million enhancers (ENCODE project Consortum, 2012; Shen et al., 2012), greatly exceeding the number of protein-coding genes. Depending on the post-translational modifications of nucleosomal histone proteins in the immediate vicinity, enhancers can exist in several states. The neutral/intermediate state is marked by mono-methylation of lysine 4 on histone H3 (H3K4me1). The poised state is marked by H3K4me1 and trimethylation of lysine 27 on histone H3 (H3K27me3) and the active state is marked by H3K4me1 and acetylation of lysine 27 on histone H3 (H3K27ac) (Creyghton et al., 2010; Rada-Iglesias et al., 2011; Zentner et al., 2011). Silencer elements are less well-studied and, in some cases, may act in a manner similar to enhancers, by interacting directly with their target gene promoters (Figure 1). However, other mechanisms for silencers have also been observed or proposed (Segert et al., 2021). Insulators bind the CTCF factor and appear to block enhancer activation of a gene when located in the genomic sequence between the enhancer and the gene.
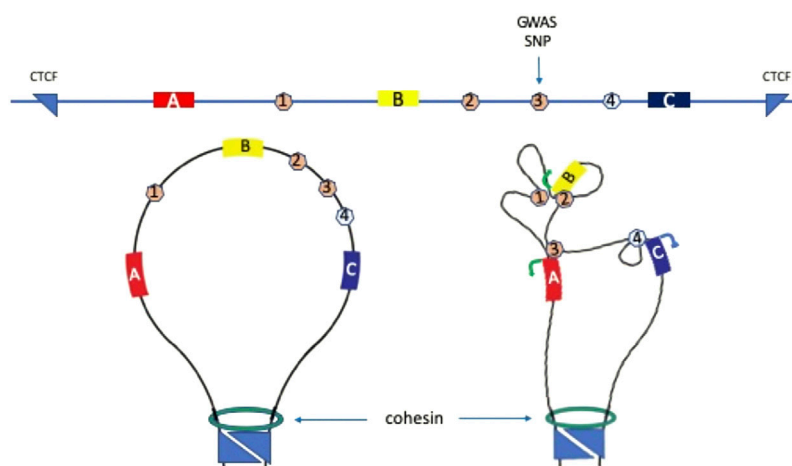
FIGURE 1
Gene regulatory elements and 3D genome conformation. Top line shows schematic of a hypothetical one megabase region of genome with three genes, **(A,B, and C)**; three enhancers 1, 2, and 3; one silencer element 4; and one genome wide association study (GWAS) single nucleotide polymorphism (SNP) (vertical arrow). The locus is flanked by convergent CTCF sites which interact and function as topologically associated domain (TAD) boundaries in cooperation with cohesin (bottom left panel). The GWAS SNP would normally be assigned to gene **(B)** or **(C)** due to proximity, however analysis of 3D genome data shows that SNP-bearing enhancer 3 contacts gene A promoter to activate transcription (bottom right panel). Enhancers 1 and 2 contact gene **(B)** to activate transcription. The silencer element 4 contacts gene **(C)** to silence transcription.

1.2.1 Topologically associated domains; sub-chromosomal organization

CTCF insulator sequences are often found in clusters at the boundaries of large chromatin domains known as topologically associated domains (TADs). TADs average about 1 megabase (Mb) in size and vary little from cell type to cell type in structure, unlike enhancer-promoter loops which are cell type specific (Javierre et al., 2016). Both CTCF and its ring-shaped interaction partner, the cohesin complex, are necessary for TAD formation. The role of cohesin in forming TADs involves loop extrusion, where DNA sequences are thought to be extruded through the ring until they encounter convergently orientated CTCF-bound sites at the boundaries (Sanborn et al., 2015; Fudenberg et al., 2016). The boundaries of TADs often interact due to the CTCF-cohesin complex, and function to locally reduce interactions with neighboring TADs. Thus enhancer-promoter interactions can occur within the TAD in which they reside but appear to be largely insulated or blocked from inappropriate interactions with sequences in neighboring TADs. Perturbation of TADs by large sequence variants overlapping boundaries by can cause dysregulated gene expression due to out-of-place (ectopic) interactions between an enhancer and promoter (Lupiáñez et al., 2015; Narendra et al., 2015; Franke et al., 2016; de Bruijn et al., 2020). Altering distances between enhancers and promoters can have dramatic effects on expression of newly proximal genes and the natural target of enhancers (Dillon et al., 1997). In contrast, deletion of individual CTCF sites or TAD boundaries does not greatly affect gene expression (Despang et al., 2019; Paliou et al., 2019; Williamson et al., 2019). Conditional ablation of CTCF or cohesin leads to loss of TAD structure throughout the genome, however gene expression is largely maintained. Detailed analyses show that interactions between genes in the vicinity of TAD boundaries, or whose interacting sites are located near TAD boundaries are largely lost (Thiecke et al., 2020). So too are interactions between other promoters and their long-range interacting sites which are CTCF/cohesin dependent.

1.3.1 Enhancer-promoter interactions

Interestingly, enhancer-promoter contact does not always drive gene expression (Jin et al., 2013; Ghavi-Helm et al., 2014; Andrey et al., 2017). Analysis of enhancers associated with differentiation-induced genes identified two mechanistic types of enhancer-promoter contacts (Rubin et al., 2017). The first class, referred to as "gained" interactions increased in contact strength during differentiation in concert with enhancer acquisition of the H3K27ac activation mark. The second "stable" class was characterized by pre-established promoter contacts with H3K27ac-marked enhancers in undifferentiated cells. The stable class was associated with cohesin, whereas the gained class was not, implying distinct mechanisms of enhancer-promoter contact formation.

Thus enhancer-promoter interactions and gene expression are maintained by a mix of cohesin-dependent and cohesin-independent mechanisms. Furthermore, although cell differentiation is often characterized by dynamic rewiring of promoter-enhancer loops (Freire-Pritchett et al., 2017; Siersbæk et al., 2017), some specific interactions do not contemporaneously drive gene expression. Rather these may be enhancers in waiting for differentiation, or environmental signals to play a role in enhanced expression (Burren et al., 2017).

Mutations in enhancers and other regulatory sequence and genome rearrangements that disrupt enhancer-promoter interactions often underlie disease and developmental malformations (Lettice et al., 2003; Benko et al., 2009; Bhatia et al., 2013; Uslu et al., 2014; Lupiáñez et al., 2015; Franke et al., 2016). Naturally occurring variation such as single nucleotide polymorphisms (SNPs) in regulatory elements can also lead to variations in gene expression. This can ultimately increase disease susceptibility. Attempts to predict which regulatory elements act on which promoter(s) based solely on proximity in the linear genome sequence often leads to incorrect assignment. Understanding the 3D contacts between regulatory sequences is essential to understand the gene or genes affected by common sequence variation in health and disease.

2 In humans, genetic variation exists in health and in disease

Humans differ from each other by millions of DNA sequence variants. These variants, in combination with environmental triggers, can lead to phenotypic variation between individuals. This can range from variation in morphology (height, body mass index), behavioral traits (alcohol consumption, risk tolerance), predisposition to complex diseases (type 1 diabetes, rheumatoid arthritis) and response to diet or medication.

Genetic variants can be classified as sequence variants or structural variants, according to their composition. SNPs, the most common type of sequence variant, are substitutions of a single nucleotide at a specific position in the genome and occur every 100–300 bases. Less frequent sequence variants include insertions or deletions (indels) of segments of DNA of up to 1 kilobases (kb). On the other hand, structural variants involve larger segments of DNA (1 kb–5 Mb), and include duplications, inversions, translocations and larger indels. Genetic variants can be described as common (minor allele frequency [MAF] > 5%) and rare (MAF < 5%), depending on how often they appear in the population (Rahim et al., 2008).

Technological advances in genotyping and next generation sequencing have enabled all but the rarest human genetic variants to be catalogued. Large scale international efforts such as the HapMap Project (Frazer et al., 2007) and the 1000 Genomes Project (Auton et al., 2015) have densely genotyped thousands of

individuals, providing a comprehensive resource on human genetic variation for a variety of ancestries, freely available in public databases. The NIH-sponsored SNP database (dbSNP) contains over 1,000 million distinct variants, as of build 155 released in June 2021 (Sherry et al., 2001).

Multiple studies assessing the heritability of complex traits have demonstrated that genetic factors are involved in 40%–70% of susceptibility to most common diseases. Therefore, characterizing this relationship between genetic variation and disease risk is an area of enormous interest. Understanding the genetic basis of disease will have a profound impact in the improvement of human health over the next decade, by helping to predict and prevent disease, to implement personalized medicine and to identify new drug targets (Zeggini et al., 2019).

Thanks to a phenomenon called linkage disequilibrium (LD), it is not necessary to directly assay all SNPs across the genome to test their associations with disease. Genetic variants can be inherited together in "blocks" or haplotypes. This means that the presence of a variant at a particular position on the genome can predict or "tag" the presence of other variants that are correlated, or in high LD, with it. International consortia have characterized genetic variants and the architecture of LD blocks across multiple populations (Lander et al., 2001; Frazer et al., 2007; Auton et al., 2015; McCarthy et al., 2016). This, together with technological advances and reduced genotyping costs, has enabled the systematic search for disease risk loci *via* high-throughput genome wide association studies (GWAS).

3 Genome wide association studies identify single nucleotide polymorphisms associated with a disease or trait

GWAS are very large studies that test for differences in the allele frequency of SNPs between individuals with different phenotypes, with the aim of finding associations between genotypes and phenotypes. GWAS require very large sample sizes to pinpoint genome wide associations, since each individual SNP may only contribute a small effect to the overall risk of disease. Microarrays are commonly used for genotyping in GWAS. However, microarrays tend to only capture common variation. Next generation sequencing can be used to genotype rare variants, but this method remains costly for genotyping the large number of individuals required for sufficiently powered case control association studies. After careful quality control, statistical tests are performed to detect significant associations between genotypes and phenotypes. Typically, logistic regression models test for associations if the phenotype is binary (i.e., presence or absence of disease), while linear regression models are used for continuous traits such as blood pressure or height (Uffelmann et al., 2021). Results must be validated in an

independent replication cohort, and meta-analyses are typically conducted to increase sample size and statistical power (Corvin et al., 2010).

For the past 15 years, GWAS have made tremendous advances in identifying risk loci - blocks of correlated SNPs significantly associated with the trait of interest - for most common diseases (Visscher et al., 2017; Claussnitzer et al., 2020). The first GWAS, published in 2005 and studying age-related macular degeneration, demonstrated that hypothesis-free genetic association testing can reveal novel biological mechanisms for complex diseases (Klein et al., 2005). Two years later, the Wellcome Trust Case Control Consortium, 2007 set the standard for the subsequent flurry of successful GWAS studies, with the publication of the largest GWAS at that time, including seven diseases (2007). Since then, more than 5,600 GWAS have been published, identifying over 370,000 associations (Buniello et al., 2019), with samples sizes that now surpass 1 million individuals (Lee et al., 2018; Jansen et al., 2019). Such large GWAS are facilitated by the availability of large publicly available genetic datasets, such as the United Kingdom Biobank (Bycroft et al., 2018).

GWAS have revealed that, for most complex diseases, genetic predisposition arises from the combination of hundreds of mostly common variants with modest individual effect sizes. The next big challenge is to biologically interpret these genetic associations, to realize the potential of GWAS to elucidate mechanistic insights of disease etiology, facilitate personalized medicine and aid the development of novel drugs. There are examples of successful translation of GWAS findings into the clinic. For example, the identification of genetic association of variants in the IL-12/IL-23 pathway with Crohn's disease (Wang et al., 2009) led to the use of drugs targeting this pathway to treat the disease (Moschen et al., 2019). However, several factors have so far limited the widespread use of GWAS findings for clinical benefit (Tam et al., 2019). As described above, LD facilitates the discovery of risk loci, but makes it difficult to pinpoint the causal variant or variants within a given locus. This is complicated by the fact that risk loci can harbor multiple, independent causal variants. In addition, most GWAS variants map to non-coding regions of the genome; approximately 5% of risk SNPs directly affect the coding sequence of a gene (Farh et al., 2015; French and Edwards, 2020). This makes the biological interpretation of GWAS variants challenging. Finally, although GWAS loci have traditionally been annotated to the nearest gene, many risk loci contain multiple genes in the vicinity or map at very large distances from coding genes, complicating the identification of the true causal genes.

Multiple studies have now shown that SNPs identified by GWAS are enriched in enhancers and other regulatory elements, that are active in disease-relevant tissues and cell types (Trynka et al., 2013; Farh et al., 2015; Cano-Gamez and Trynka, 2020). Regulatory elements are generally bound to transcription factors in open chromatin, with nearby marks of active transcription

such as histone modifications. These elements likely use chromatin looping to come into contact with their target genes (Gasperini et al., 2020). Considerable efforts have been made to catalog regulatory elements across the human genome, by mapping these chromatin features using techniques such as ATAC-seq, ChIP-seq and Hi-C, among others (Bernstein et al., 2010; Dunham et al.,2012; Stunnenberg and Hirst, 2016). The recent EpiMap project illustrates how such functional genomics data can be used to predict disease-relevant tissues, causal SNPs and their target genes. Boix et al. used data from the main large scale consortia studies such as ENCODE, Roadmap and IHEC to compile 10,000 epigenomic maps across multiple cell types and tissues, which were used to identify 540 traits with 30,000 associated genetic loci (Boix et al., 2021).

4 Identifying causal variants, cell type and target gene

4.1 Fine mapping of disease-associated regions

A "typical" GWAS locus contains a number of highly correlated SNP variants associated with disease, situated either intronic or intergenic to gene coding regions. To fully utilize and translate GWAS, each disease-associated genetic region must be followed up by identifying the casual variant(s), the target gene(s) and the relevant cell type(s) in which the variants act.

Statistical fine mapping can refine the causal variants in a region. Information from SNPs genotyped *via* microarray is used to "impute" the association from other untyped SNPs. This is achieved using knowledge of the underlying relationship between SNPs from large numbers of sequenced genomes, currently over 85 million on the Michigan Imputation Server (Das et al., 2016). In this way the direct genotyping of a relatively limited number of ~500,000 SNPs can be expanded to assess the association of tens of millions of variants. From here, these SNPs can be statistically fine mapped, using conditional analysis to determine whether the association signal from a locus is best explained with a single, or multiple independent genetic effects.

Bayesian statistical methods are now routinely used to assign a probability as to whether an associated SNP is causal. This provides the scaffold of a locus on which to add functional annotation. For example, a complicated locus may have multiple independent associated signals, each made up of many (>20) SNP variants with roughly equal probability of being causal. Alternatively, the more straightforward loci may well only have a single association, made up of one or two SNPs that are likely to be causal based on Bayesian probabilities. This describes the

"credible SNP set" for a locus, the number of SNPs that make up 99% (or 95%) of the probability (Schaid et al., 2018).
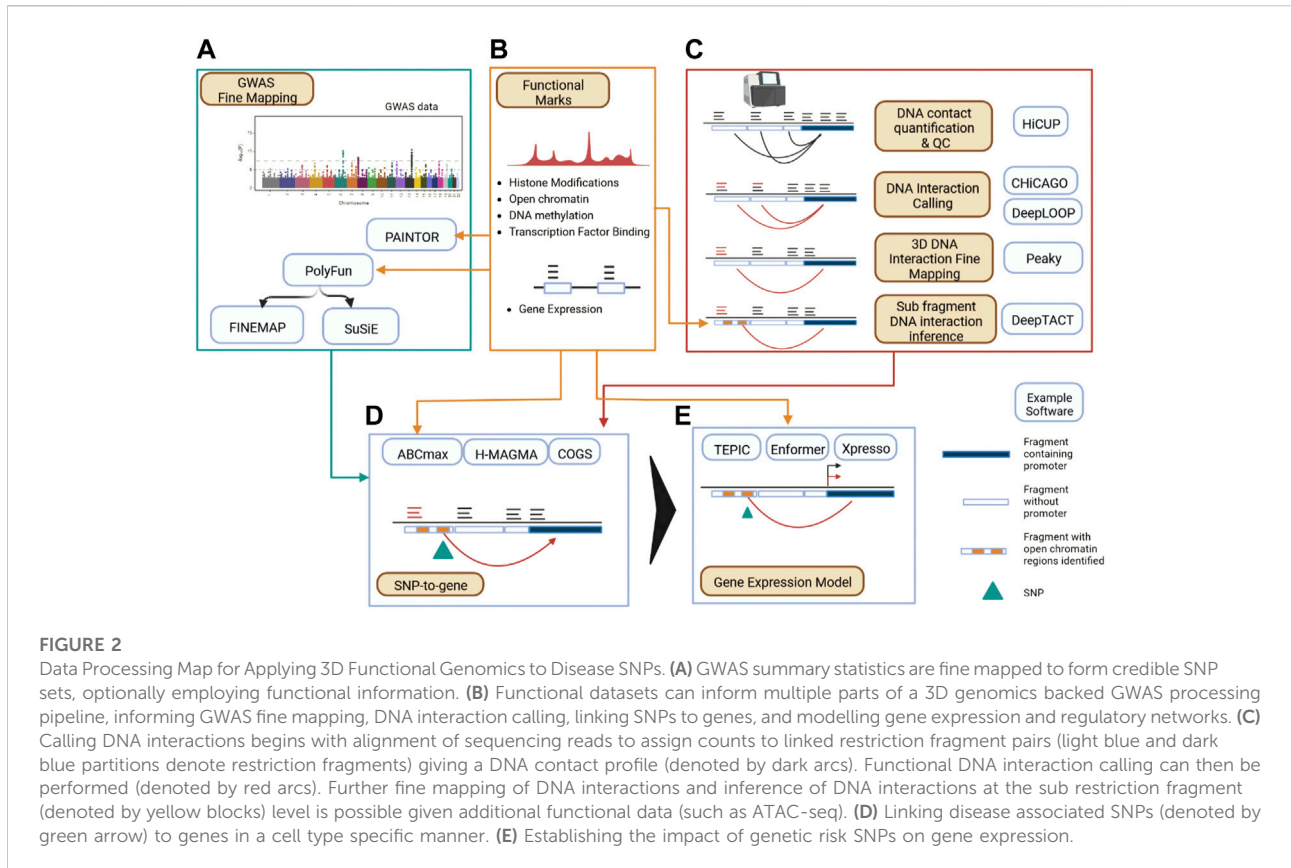
This Bayesian framework incorporates knowledge of the biological nature of the SNPs into the probability models (Udler et al., 2019). For example, higher prior probabilities can be assigned to SNPs that reside in enhancer regions shown to be important in disease such as T cell enhancers in rheumatoid arthritis (Farh et al., 2015), or SNPs that change pivotal transcription factors in a cell type known to be important in disease. This will then inform the SNP(s) that are likely to be important in disease risk, for a given cell type.

4.2 Building the evidence case for causality

As previously described, regulatory regions can act on genes over long distances, often "skipping genes" and not necessarily regulating the closest gene in the linear view of a chromosome. It is therefore important, and non-trivial, to assign the disease-associated SNPs residing within enhancers to the genes they regulate. This can be achieved by combining evidence from a number of genomic technologies to build up evidence of the gene, and cellular context, in which the credible SNPs could act (Delaneau et al., 2019; Ding et al., 2020; Orozco, 2022).

Initially, the credible SNPs can be physically mapped to regions of the genome that are open and active in particular cell types. Open and active non-coding regions of the genome are potentially important in gene regulation. These regions are generally mapped *via* ATAC-seq for openness and ChIP-seq for activity. DNA is tightly wound around histones in the nucleus. In order to be active, the DNA is unwound locally from histones, and the region demarked with histones that are modified, methylated or acetylated, to maintain the open structure. The open DNA is more accessible, such that a transposase enzyme can insert adapters into these regions, so that they are preferentially represented in sequencing analysis (ATAC-seq) (Buenrostro et al., 2015). Histones can be modified to indicate promoters, enhancers, active enhancers or silencing regions. Antibodies that target specific histone modifications (such as H3K27ac for active chromatin, or H3K4me1 for enhancers) can determine these different regulatory genomic regions, by enriching sequencing libraries (ChIP-seq) (Ernst and Kellis, 2017). Using these molecular technologies, it is now possible to map the different chromatin states in a wide range of cell types (Gasperini et al., 2020).

These data can be generated in individual research labs, especially for specialist cell types, conditions or cells from patient samples. General cell type data are available in publicly available databases. Large projects such as the ENCODE (ENCODE Project Consortium, 2012) and Epigenomics Roadmap (Bernstein et al., 2010) document the

FIGURE 2
Data Processing Map for Applying 3D Functional Genomics to Disease SNPs. **(A)** GWAS summary statistics are fine mapped to form credible SNP sets, optionally employing functional information. **(B)** Functional datasets can inform multiple parts of a 3D genomics backed GWAS processing pipeline, informing GWAS fine mapping, DNA interaction calling, linking SNPs to genes, and modelling gene expression and regulatory networks. **(C)** Calling DNA interactions begins with alignment of sequencing reads to assign counts to linked restriction fragment pairs (light blue and dark blue partitions denote restriction fragments) giving a DNA contact profile (denoted by dark arcs). Functional DNA interaction calling can then be performed (denoted by red arcs). Further fine mapping of DNA interactions and inference of DNA interactions at the sub restriction fragment (denoted by yellow blocks) level is possible given additional functional data (such as ATAC-seq). **(D)** Linking disease associated SNPs (denoted by green arrow) to genes in a cell type specific manner. **(E)** Establishing the impact of genetic risk SNPs on gene expression.

state of many cell types. For example, the ENCODE-Roadmap Encyclopedia has generated ten "ground level" data sets, including ChIP-seq and ATAC-seq and Hi-C, on over 500 tissue and cell types (Moore et al., 2020). These data can be viewed in easily accessible websites such as IHEC (https://ihec-epigenomes.org/) (Stunnenberg and Hirst, 2016) or EPIMAP (https://epimap.fr/kremlin-bicetre) (Boix et al., 2021), where SNPs can be readily mapped to areas of activity in different cell types and states.

In this way it is possible to narrow down a range of SNPs from the credible set that are found within regions of activity in certain cell types. Of course, SNPs may exhibit their true causal nature in cell types and states outside these standard ones, for example only under certain stimulatory conditions, in a given chronicity or from a patient with active disease.

The statistical and functional annotation of disease-associated SNPs provides a strong hypothesis as to which of the credible SNPs are likely functional, and the cell type in which they are likely to be active. Overlaying evidence of transcription factor binding sites, from resources such as JASPAR (https://jaspar.genereg.net) (Fornes et al., 2020), could also indicate how the SNP would regulate gene transcription. For example, the destruction of a key transcription factor binding site could impact the regulation of gene expression under certain cellular conditions.

4.3 From single nucleotide polymorphism to target gene and cell type

Next the disease-implicated enhancer/SNP is assigned a likely target gene. Again, many lines of evidence can contribute. As previously described, TADs may well restrict the domain of enhancer/gene contact. Determining the gene expression in the relevant cell type, within the TAD where the implicated enhancer resides, will narrow down the likely target genes. This can be achieved using RNA-seq data from specific cells and exploring cell type-specific expression databases (e.g., the Human Cell Atlas; https://www.humancellatlas.org/and the Single Cell Expression Atlas; https://www.ebi.ac.uk/gxa/sc/home). Correlation between the activity of the enhancer (*via* quantitative ATAC-seq) and the activity of the gene (*via* quantitative RNA-seq or quantitative ATAC-seq of the gene region) can increase confidence as to the gene/enhancer relationship (Gate et al., 2018; Yang et al., 2020).

A stronger relationship between the enhancer SNP and gene expression can be found with expression quantitative trait locus (eQTL) analysis. Here the SNP variant (for example A/G), is correlated with gene expression, where one allele (e.g., A), is consistently associated with increased gene expression (Albert and Kruglyak, 2015). Such compelling evidence suggests that the

TABLE 1 Selected software for integrating 3D functional genomes with disease SNPs.

Step	Software	
DNA contact processing	HiCUP	https://github.com/StevenWingett/HiCUP
	MHi-C	https://github.com/MHi-C
Promoter Interaction Calling	CHiCANE	https://cran.r-project.org/web/packages/chicane
	CHiCAGO	https://bitbucket.org/chicagoTeam/chicago
	CHiCMaxima	https://github.com/yousra291987/ChiCMaxima
	HiCapTools	https://github.com/sahlenlab/HiCapTools
	GOTHiC	http://www.bioconductor.org/packages/release/bioc/html/GOTHiC.html
Statistical and functional fine mapping	PolyFun	https://github.com/omerwe/polyfun
	SuSiE	https://github.com/stephenslab/susieR
	FINEMAP	http://www.christianbenner.com/
	PAINTOR	https://github.com/gkichaev/PAINTOR_V3.0
3D interaction fine mapping	Peaky	https://github.com/broadinstitute/ABC-Enhancer-Gene-Prediction
Gene Prioritization	ABC-max model	https://github.com/broadinstitute/ABC-Enhancer-Gene-Prediction
	COGS	https://github.com/ollyburren/rCOGS
	H-MAGMA	https://github.com/thewonlab/H-MAGMA
Gene expression prediction	TEPIC	https://github.com/SchulzLab/TEPIC
	Xpresso	https://github.com/vagarwal87/Xpresso
	Enformer	https://github.com/deepmind/deepmind-research/tree/master/enformer

variant does indeed influence gene expression, and directly links the SNP and gene. These relationships are best observed in the GTEx consortium data (GTEx, 2013) (https://gtexportal.org). Issues with eQTL exist. For example, it can take hundreds of samples to expose a robust relationship between SNP and gene, and these relationships can change, sometimes dramatically, based on cell type and stimulation (Fairfax et al., 2014). Often RNA-seq experiments are not performed on truly "homogeneous" cell types (e.g., T cells) which could mask the relationship between SNPs and gene in more refined studies (e.g., stimulated T-reg cells). The disease risk variant may also only be highly correlated to the true eQTL functional SNP, so appearing to change the expression, but not being the SNP responsible. In this case, "colocalization" statistical analyses are required to prove that the lead GWAS variant is also the lead eQTL variant (Hormozdiari et al., 2016).

Other QTL analyses can give insight into the function of a disease-associated SNP. For example, the risk allele may be correlated with enhancer activity (ATAC-QTL) (Gate et al., 2018), splice variants (splQTL) (Garrido-Martín et al., 2021), protein levels (pQTL) (Yao et al., 2018) or even histone modifications (hQTL) (McVicker et al., 2013), in specific cell types. As the databases are expanded with more, and different types of samples, these relationships will offer strong clues as to the likely function of variants that increase the risk of disease.

High resolution chromatin capture techniques, such as PCHi-C, can determine the physical links between the GWAS implicated enhancers and the genes they regulate (Martin et al., 2015; Javierre et al., 2016; González-Serna et al., 2022).

Overlaying the ATAC-seq, ChIP-seq, RNA-seq, and QTL data with chromatin interaction data can provide more confidence as to the gene/enhancer relationship, in the identified cellular context. The advances in chromatin technology allow for base pair resolution of DNA interactions, from small numbers of starting cells (Downes et al., 2021). This is important in relating the function of enhancers in particular cell types, such as specific clinical subtypes from patients with active disease, or in remission. This has the power to directly link enhancers to their target genes, in different cellular contexts, and adds a compelling layer of evidence as to how GWAS variants act to increase the risk of disease.

Using this stepwise approach, researchers can refine a complicated structure of many potentially causal variants, to a limited number (statistical fine mapping, functional mapping) in a limited number of cell types (cell type-specific activity of enhancers) impacting a limited number of genes (eQTL, chromatin interactions). This provides a robust hypothesis as to how the variant increases risk of the disease. These hypotheses can be investigated using genetic engineering technologies such as CRISPR in 3D tissue cultures or mice, to provide a model for disease risk.

5 Chromatin conformation capture approaches

Experimental methods to interrogate the 3D folding of the human genome can be divided into microscopy-based and

biochemical approaches. Although throughput is limited, advanced microscopy techniques can capture the temporal and spatial dynamics of 3D genome organization (Boettiger et al., 2016; Williamson et al., 2016; Bintu et al., 2018; Chen et al., 2018; Gu et al., 2018; Gabriele et al., 2022). Here, we focus on high resolution biochemical approaches to map long-range chromosomal contacts, including gene regulatory contacts between enhancers and their target gene promoters. The most commonly used approaches are variants of capturing chromosome conformation (3C) (Dekker et al., 2002) that rely on fixation and proximity ligation, although methods without ligation (Beagrie et al., 2017), without crosslinking (Brant et al., 2016) or without crosslinking and ligation (Redolfi et al., 2019) have also been established.

In 3C, cells are fixed with a short distance crosslinker such as formaldehyde to "freeze" the 3D genome folding in its native state. Subsequently, the chromatin is digested with a restriction enzyme, with the choice of the restriction enzyme determining the resolution of the resulting chromatin interaction maps (Su et al., 2021). Proximity ligation between the overhangs created by restriction digestion then creates hybrid molecules between chromatin regions that were in close proximity at the time of crosslinking. The frequency of these ligation events, revealed by massively parallel sequencing, is a readout for how often the corresponding chromosomal regions are in close spatial proximity in a specific cell type.

Chromosome conformation capture methods can be one-to-one (3C) (Dekker et al., 2002), one-to-all (4C) (Simonis et al., 2006; Zhao et al., 2006), many-to-many (5C) (Dostie et al., 2006) or all-to-all high-throughput (Hi-C) (Lieberman-Aiden et al., 2009). In Hi-C, the ligation step is preceded by integration of biotin at restriction fragment overhangs, enabling the isolation of biotin-marked ligation products. As a result, complex Hi-C libraries contain all the chromosomal interactions within a cell population. This has led to the discovery of key principles and building blocks of 3D chromatin organization, including TADs (Dixon et al., 2012; Nora et al., 2012), and A and B compartments in which active and repressed regions spatially segregate (Lieberman-Aiden et al., 2009). However, Hi-C libraries are enormously complex, with an estimated 10^{11} independent ligation products between ~4 kb fragments of the mouse or human genomes (Belton et al., 2012). This prevents the reliable identification of promoter-enhancer contacts with statistical confidence, unless Hi-C libraries undergo ultra-deep sequencing (Rao et al., 2014; Bonev et al., 2017). This limitation has been addressed by methods have been developed to enrich 3C for specific subsets of interactions, such as Capture-C (Hughes et al., 2014; Davies et al., 2016; Chesi et al., 2019), or Hi-C libraries (Dryden et al., 2014; Schoenfelder et al., 2015a; Mifsud et al., 2015; Sahlén et al., 2015). Promoter-Capture Hi-C (PCHi-C) specifically

enriches promoter-containing ligation products from Hi-C libraries, allowing capture of promoter interacting regions for all (>22 K) promoters across the mouse genome (Schoenfelder et al., 2015a; Sahlén et al., 2015; Wilson et al., 2016; Siersbæk et al., 2017; Comoglio et al., 2018; Novo et al., 2018; Schoenfelder et al., 2018) and the human genome (Mifsud et al., 2015; Javierre et al., 2016; Freire-Pritchett et al., 2017; Rubin et al., 2017; Chovanec et al., 2021). A direct comparison between Capture-C and Capture Hi-C revealed that Capture Hi-C generates two to three times more usable and informative reads (Su et al., 2021), due to the fact that Capture-C enriches for un-ligated fragments (Sahlén et al., 2015).

Variants of 3C approaches have replaced restriction enzymes with micrococcal nuclease (Micro-C) (Hsieh et al., 2015), yielding chromatin interaction maps with nucleosome level resolution. Micro-C libraries can be enriched for interactions with selected bait loci (Micro Capture-C or MCC) (Hua et al., 2021) or continuous genomic regions (Tiled-MCC) (Aljahani et al., 2022), closely mimicking previously developed approaches to enrich 3C or Hi-C libraries for specific interactions.

Further, several methods have been developed to target 3D chromatin interactions of accessible regions in the genome, including OCEAN-C (Li et al., 2018), HiCoP (Zhang et al., 2020), NicE-C (Luo et al., 2022), and HiCAR (Hi-C on accessible regulatory DNA) (Wei et al., 2022).

Alternative approaches have combined proximity ligation with chromatin immunoprecipitation (ChIP) to isolate chromosomal interactions involving specific proteins. In one approach, ChIP is performed, then biotinylation enables isolation of ligation products bound by the protein of interest (Fullwood et al., 2009). This approach is called Chromatin Interaction Analysis by Paired-End Tag sequencing (ChIA-PET). HiChIP, another protein-centric chromatin conformation capture method (Mumbach et al., 2016; Mumbach et al., 2017), combines marking of ligation junctions with biotin, ChIP, and construction of a transposase-mediated library. HiChIP requires a lower input of cells, and generates more informative reads compared to ChIA-PET (Mumbach et al., 2016). Similarly, proximity ligation assisted ChIP-seq (PLAC-seq) seems to outperform ChIA-PET by switching the order of proximity ligation and chromatin shearing steps (Fang et al., 2016).

Importantly, Capture Hi-C does not rely on ChIP (in contrast to ChIA-PET, HiChIP and PLAC-seq), and is therefore capable of interrogating chromosomal interactions irrespective of protein occupancy. This is key for comparing 3D chromatin interactomes in specific genetic backgrounds, for example wild-type to knockout. Indeed, this approach has uncovered a strong inter-chromosomal spatial interaction network between Polycomb-bound and -regulated genes in mouse embryonic stem cells (Schoenfelder et al., 2015b).

FIGURE 3
Overview of workflow for translating GWAS data into clinical benefit for patients.
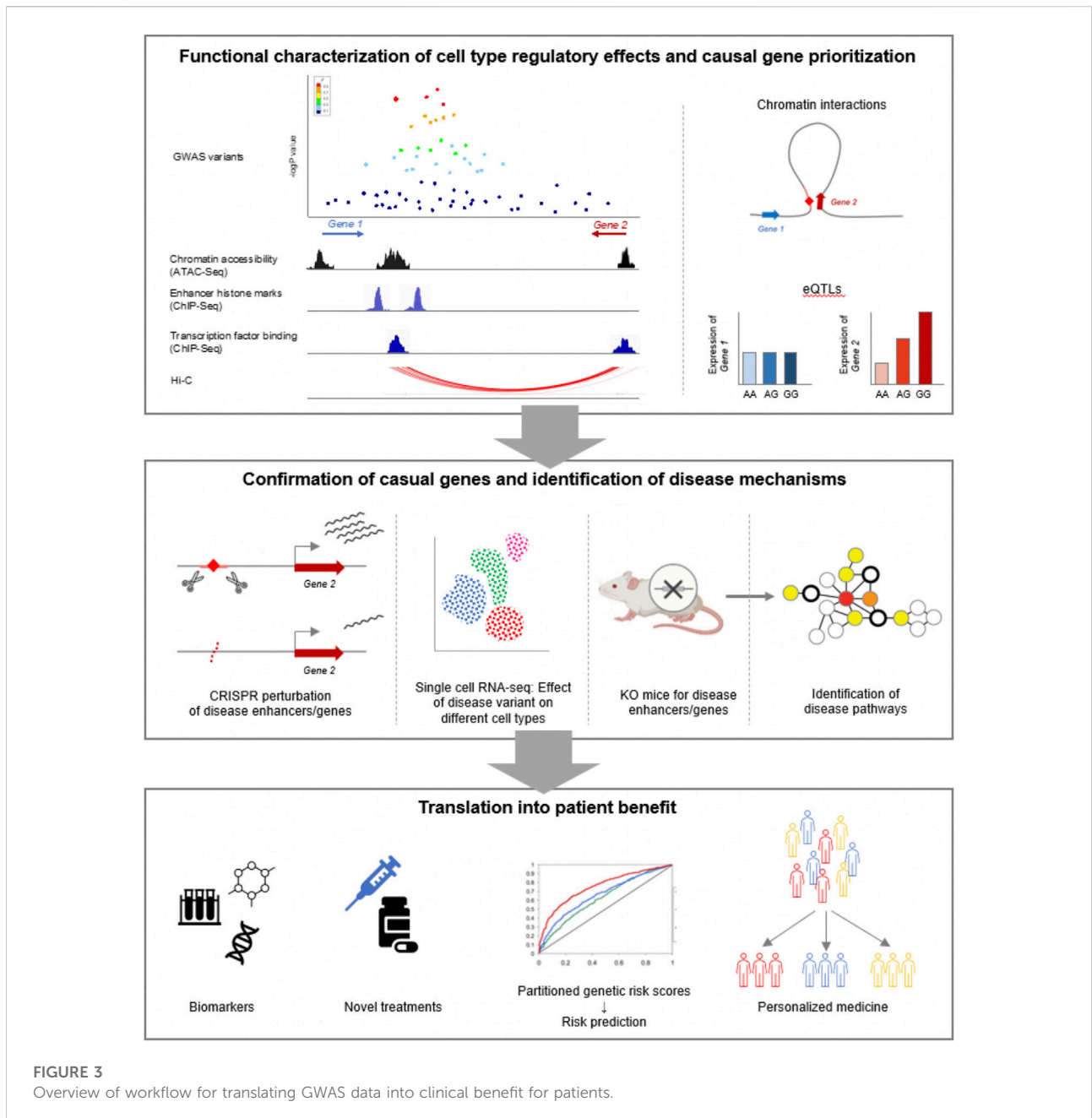
6 Bioinformatic tools to process and integrate genome wide association studies and 3D genomics data

Strategies to integrate GWAS and 3D genomics data as outlined in the previous sections have been supported by a maturing and expanded set of computational and statistical tooling over the last decade (Pal et al., 2019; Gong et al., 2021) (Table 1 ; Figure 2). These tools provide a base to analyze genetic disease risk in the context of the 3D

functional genome; allowing insight into genetic contributions to disease (Gazal et al., 2022) and deep investigation into regulatory dynamics of complex diseases (Jin et al., 2013; Burren et al., 2017; Wang et al., 2018; Su et al., 2020). An analysis pipeline for mapping disease risk variants to target genes *via* 3D genomics seeks to establish confident *cis*-interactions, first by processing proximity ligation sequencing data to DNA contact data (Servant et al., 2015; Wingett et al., 2015) (Zheng et al., 2019), and then inferring functional relevance through the application of statistical models (Cairns

et al., 2016; Mifsud et al., 2017; Ben Zouari et al., 2019; Holgersen et al., 2021). Several studies have evaluated the merits of common pipelines that implement these steps (Forcato et al., 2017; Aljogol et al., 2022), and one commonly used component (CHiCAGO) for calling biologically meaningful *cis*-interactions from DNA contact data generated from PCHi-C has been independently evaluated (Disney-Hogg et al., 2020) and guidance on usage published (Freire-Pritchett et al., 2021). Recent methodological advances allow for improved resolution over the typical restriction fragment level DNA interaction processing pipeline (Figure 2C). In one case, this has been achieved through development of a statistical model of neighboring DNA interactions so as to precisely assign DNA interactions to restriction fragments (Eijsbouts et al., 2019), another approach uses deep learning to infer sub-fragment *cis*-interactions using DNA sequence patterns and chromatin accessibility data (Li et al., 2019). Despite the challenges in calling *cis*-interactions from typically under-sampled Hi-C/CHi-C data (Aljogol et al., 2022), deep learning methods have been developed that provide improved DNA interaction coverage and resolution for sequencing and sample costs (Zhang, 2022).

Multiple reviews cover recent statistical advances that estimate causal genetic risk loci, accounting for LD and functional annotation (Spain and Barrett, 2015; Pasaniuc and Price, 2017; Hutchinson et al., 2020; Wang and Huang, 2022). Briefly, methods have evolved that use GWAS summary statistics rather than individual genotype level data, and relax the single causal variant per locus assumption of earlier landmark fine mapping work (Wakefield, 2009; Maller et al., 2012). Several of these methods allow for the integration of functional marks. For example, the posterior probability for causality computed by the fine mapping methods FINEMAP (Benner et al., 2016) and SuSiE (Wang et al., 2018) can be influenced using prior probabilities computed first using the PolyFun method (Weissbrod et al., 2020), while the PAINTOR method (Kichaev et al., 2014) is specifically designed to integrate functional annotation into the fine mapping process.

Having established credible SNP sets, *cis*-interactions can be used to link disease SNPs to target genes and estimate confidence of association with the disease (often referred to as a SNP-to-gene strategy—"S2G," Figure 2D). One approach is to intersect fine mapped SNPs with Promoter Interacting Regions (PIRs) to establish a set of target genes. For example, Song et al (2019) first compute a window around each SNP such that neighboring SNPs must have LD coefficient of $r^2 > 0.8$, then intersects this window with PIRs to build a set of target genes (Song et al., 2019). Another strategy has been to develop a per-gene disease risk score using a framework that integrates all fine mapped genetic risk linked by PIRs (and other regions) to a particular locus, such as the COGS method (Javierre et al., 2016; Burren et al., 2017) and the more recent H-MAGMA framework (de Leeuw et al., 2015; Sey et al., 2020; Zhang, 2022). With the addition of functional datasets, the Activity By Contact (ABC) model for

enhancer contribution to gene regulation (Fulco et al., 2019) can be employed to translate disease risk to the gene level. Credible SNP sets (for example, obtained from SuSiE) can be overlapped with ABC enhancers, and the gene with the highest ABC enhancer score is chosen (ABC-max) (Nasser et al., 2021). Combining multiple S2G strategies has been demonstrated to be important for biological insight (Giambartolomei et al., 2021) and accuracy (Gazal et al., 2022).

Finally, to estimate how a variant impacts the expressed genome, deep learning approaches that model gene expression from minimal input such as DNA sequence alone have recently undergone substantial improvement in prediction accuracy (Agarwal and Shendure, 2020; Avsec et al., 2021). However, representing long-range enhancer architecture appears to be critical for improving performance to a point where non-coding disease variants can be reliably assessed for their impact on RNA expression (Agarwal and Shendure, 2020). Avsec et al. (2021) also recently demonstrated that allowing for the representation of longer-range interactions in a DNA sequence based deep learning framework significantly improved gene expression prediction. Schmidt et al. (2020) demonstrate that the addition of inter-TAD promoter-enhancer contact data improves gene expression prediction using their existing TEPIC pipeline.

7 Future challenges

Although there has been substantial progress establishing technologies and bioinformatics tooling for 3D genomics over the last decade, multiple challenges and opportunities remain.

Recent evaluation of bioinformatics tooling demonstrates significant differences between tools used for calling *cis*-interactions from Hi-C (Forcato et al., 2017) and PCHi-C (Aljogol et al., 2022). In particular, these studies have highlighted low concordance between technical replicates and reproducibility trading off against resolution (e.g., window size of analysis) and the number of sample replicates. This can be attributed partly to under sampling of complex sequencing libraries from heterogenous cell populations. Technologies that allow for deconvolution of cell heterogeneity and increased efficient use of sequencing space will clearly empower all 3D genome applications. Bioinformatics methods to improve these characteristics are emerging, such as the aforementioned fine mapping of DNA interactions and machine learning derived models that will augment experimental measurement. Methods developed to predict *cis*-interactions from linear genome function marks will likely aid this effort (Piecyk et al., 2022). Thus, a core challenge for the field to enable diagnostic and therapeutic discovery applications is to drive cost-effective construction of 3D genome states at high resolution.

There are substantial improvements to gain in accurately linking SNPs to genes, as demonstrated by recent studies evaluating S2G strategies (Dey et al., 2022; Gazal et al., 2022; Lettre, 2022). These studies point to promoter-enhancer strategies (such as ABC models) along with wider profiling of multiple cell types/states as a promising avenue to pursue. Methodological improvements built on similar frameworks to ABC, combined with functional experimental paradigms may drive progress in this area. Thus, a further challenge in the field is the comprehensive integration of 3D genomes into SNP-to-gene linking strategies.

3D genomes can also aid in inferring the functional impact of disease genetics on linked targets genes, such as the result on gene expression and effect across regulatory networks and pathways. The use of 3D genomics to provide data on long-range interactions (Agarwal and Shendure, 2020) as a base to construct transcriptional networks appears crucial to this goal. Thus, another challenge is the full integration of 3D genomes in the construction of transcriptional regulatory networks and pathways. Support for visualization and interactivity using 3D genomic data has advanced (Yardımcı and Noble, 2017; Oluwadare et al., 2019), but there is significant need for interactive systems focused on exploring disease risk genetics (Wlasnowolski et al., 2020), and to augment human designers of therapeutic and diagnostic strategies. These would allow for interactive exploration of scenarios based on genetic and structural configurations, enabling viewpoints at the base-pair level with the ability to simulate the impact on regulatory dynamics, pathways and cell state seamlessly. Finally, programs that support validation of disease genetics by profiling 3D genomes in multiple patient samples will be crucial to confirm and inform regulatory disease models established using 3D genomics (Figure 3).

Cleary, we are only just beginning to scratch the surface of the complete picture of how human genetic variation contributes to gene expression control, innumerable human traits and disease susceptibility across hundreds of different human tissue-types and their developmental and differentiation precursors. With continued improvement of molecular and computational technologies, and an ever-expanding fountain of new data we should expect to see significant new advances in our understanding of the causes and consequences of disease, which are the critical first steps in rational design of therapeutic interventions.

Author contributions

All authors listed have made an equal, substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

Acknowledgments

Conflict of interest

PF and SS are founders, shareholders and paid consultants of Enhanc3D Genomics Ltd. NW is an employee and shareholder of Enhanc3D Genomics Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Agarwal, V., and Shendure, J. (2020). Predicting mRNA abundance directly from genomic sequence using deep convolutional neural networks. *Cell Rep.* 31 (7), 107663. doi:10.1016/j.celrep.2020.107663

Albert, F. W., and Kruglyak, L. (2015). The role of regulatory variation in complex traits and disease. *Nat. Rev. Genet.* 16 (4), 197–212. doi:10.1038/nrg3891

Aljahani, A., Hua, P., Karpinska, M. A., Quililan, K., Davies, J. O. J., and Oudelaar, A. M. (2022). Analysis of sub-kilobase chromatin topology reveals nano-scale regulatory interactions with variable dependence on cohesin and CTCF. *Nat. Commun.* 13 (1), 2139. doi:10.1038/s41467-022-29696-5

Aljogol, D., Thompson, I. R., Osborne, C. S., and Mifsud, B. (2022). Comparison of capture Hi-C analytical pipelines. *Front. Genet.* 13, 786501. doi:10.3389/fgene.2022.786501

Andrey, G., Schöpflin, R., Jerković, I., Heinrich, V., Ibrahim, D. M., Paliou, C., et al. (2017). Characterization of hundreds of regulatory landscapes in developing limbs reveals two regimes of chromatin folding. *Genome Res.* 27 (2), 223–233. doi:10.1101/gr.213066.116

Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., et al. (2015). A global reference for human genetic variation. *Nature* 526 (7571), 68–74. doi:10.1038/nature15393

Avsec, Ž., Agarwal, V., Visentin, D., Ledsam, J. R., Grabska-Barwinska, A., Taylor, K. R., et al. (2021). Effective gene expression prediction from sequence by integrating long-range interactions. *Nat. Methods* 18 (10), 1196–1203. doi:10.1038/s41592-021-01252-x

Beagrie, R. A., Scialdone, A., Schueler, M., Kraemer, D. C., Chotalia, M., Xie, S. Q., et al. (2017). Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* 543 (7646), 519–524. doi:10.1038/nature21411

Belton, J. M., McCord, R. P., Gibcus, J. H., Naumova, N., Zhan, Y., and Dekker, J. (2012). Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods* 58 (3), 268–276. doi:10.1016/j.ymeth.2012.05.001

Ben Zouari, Y., Molitor, A. M., Sikorska, N., Pancaldi, V., and Sexton, T. (2019). ChiCMaxima: A robust and simple pipeline for detection and visualization of chromatin looping in capture Hi-C. *Genome Biol.* 20 (1), 102. doi:10.1186/s13059-019-1706-3

Benko, S., Fantes, J. A., Amiel, J., Kleinjan, D. J., Thomas, S., Ramsay, J., et al. (2009). Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. *Nat. Genet.* 41 (3), 359–364. doi:10.1038/ng.329

Benner, C., Spencer, C. C., Havulinna, A. S., Salomaa, V., Ripatti, S., and Pirinen, M. (2016). Finemap: Efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* 32 (10), 1493–1501. doi:10.1093/bioinformatics/btw018

Bernstein, B. E., Stamatoyannopoulos, J. A., Costello, J. F., Ren, B., Milosavljevic, A., Meissner, A., et al. (2010). The NIH Roadmap epigenomics mapping consortium. *Nat. Biotechnol.* 28 (10), 1045–1048. doi:10.1038/nbt1010-1045

Bhatia, S., Bengani, H., Fish, M., Brown, A., Divizia, M. T., de Marco, R., et al. (2013). Disruption of autoregulatory feedback by a mutation in a remote, ultraconserved PAX6 enhancer causes aniridia. *Am. J. Hum. Genet.* 93 (6), 1126–1134. doi:10.1016/j.ajhg.2013.10.028

Bintu, B., Mateo, L. J., Su, J. H., Sinnott-Armstrong, N. A., Parker, M., Kinrot, S., et al. (2018). Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* 362, eaau1783. doi:10.1126/science.aau1783

Boettiger, A. N., Bintu, B., Moffitt, J. R., Wang, S., Beliveau, B. J., Fudenberg, G., et al. (2016). Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* 529 (7586), 418–422. doi:10.1038/nature16496

Boix, C. A., James, B. T., Park, Y. P., Meuleman, W., and Kellis, M. (2021). Regulatory genomic circuitry of human disease loci by integrative epigenomics. *Nature* 590 (7845), 300–307. doi:10.1038/s41586-020-03145-z

Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G. L., Lubling, Y., et al. (2017). Multiscale 3D genome rewiring during mouse neural development. *Cell* 171 (3), 557–572. doi:10.1016/j.cell.2017.09.043

Brant, L., Georgomanolis, T., Nikolic, M., Brackley, C. A., Kolovos, P., van Ijcken, W., et al. (2016). Exploiting native forces to capture chromosome conformation in mammalian cell nuclei. *Mol. Syst. Biol.* 12 (12), 891. doi:10.15252/msb.20167311

Buenrostro, J. D., Wu, B., Chang, H. Y., and Greenleaf, W. J. (2015). ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* 109, 21. doi:10.1002/0471142727.mb2129s109

Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 47, D1005–d1012. doi:10.1093/nar/gky1120

Burren, O. S., Rubio García, A., Javierre, B. M., Rainbow, D. B., Cairns, J., Cooper, N. J., et al. (2017). Chromosome contacts in activated T cells identify autoimmune disease candidate genes. *Genome Biol.* 18 (1), 165. doi:10.1186/s13059-017-1285-0

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562 (7726), 203–209. doi:10.1038/s41586-018-0579-z

Cairns, J., Freire-Pritchett, P., Wingett, S. W., Várnai, C., Dimond, A., Plagnol, V., et al. (2016). CHiCAGO: Robust detection of DNA looping interactions in capture Hi-C data. *Genome Biol.* 17 (1), 127. doi:10.1186/s13059-016-0992-2

Cano-Gamez, E., and Trynka, G. (2020). From GWAS to function: Using functional genomics to identify the mechanisms underlying complex diseases. *Front. Genet.* 11, 424. doi:10.3389/fgene.2020.00424

Carter, D., Chakalova, L., Osborne, C. S., Dai, Y. F., and Fraser, P. (2002). Long-range chromatin regulatory interactions *in vivo*. *Nat. Genet.* 32 (4), 623–626. doi:10.1038/ng1051

Chen, H., Levo, M., Barinov, L., Fujioka, M., Jaynes, J. B., and Gregor, T. (2018). Dynamic interplay between enhancer-promoter topology and gene activity. *Nat. Genet.* 50 (9), 1296–1303. doi:10.1038/s41588-018-0175-z

Chesi, A., Wagley, Y., Johnson, M. E., Manduchi, E., Su, C., Lu, S., et al. (2019). Genome-scale Capture C promoter interactions implicate effector genes at GWAS loci for bone mineral density. *Nat. Commun* 10 (1), 1260. doi:10.1038/s41467-019-09302-x

Chovanec, P., Collier, A. J., Krueger, C., Várnai, C., Semprich, C. I., Schoenfelder, S., et al. (2021). Widespread reorganisation of pluripotent factor binding and gene regulatory interactions between human pluripotent states. *Nat. Commun.* 12 (1), 2098. doi:10.1038/s41467-021-22201-4

Claussnitzer, M., Cho, J. H., Collins, R., Cox, N. J., Dermitzakis, E. T., Hurles, M. E., et al. (2020). A brief history of human disease genetics. *Nature* 577 (7789), 179–189. doi:10.1038/s41586-019-1879-7

Comoglio, F., Park, H. J., Schoenfelder, S., Barozzi, I., Bode, D., Fraser, P., et al. (2018). Thrombopoietin signaling to chromatin elicits rapid and pervasive epigenome remodeling within poised chromatin architectures. *Genome Res.* 28 (3), 295–309. doi:10.1101/gr.227272.117

Corvin, A., Craddock, N., and Sullivan, P. F. (2010). Genome-wide association studies: A primer. *Psychol. Med.* 40 (7), 1063–1077. doi:10.1017/S0033291709991723

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* 107 (50), 21931–21936. doi:10.1073/pnas.1016071107

Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A. E., Kwong, A., et al. (2016). Next-generation genotype imputation service and methods. *Nat. Genet.* 48 (10), 1284–1287. doi:10.1038/ng.3656

Davies, J. O., Telenius, J. M., McGowan, S. J., Roberts, N. A., Taylor, S., Higgs, D. R., et al. (2016). Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat. Methods* 13 (1), 74–80. doi:10.1038/nmeth.3664

de Bruijn, S. E., Fiorentino, A., Ottaviani, D., Fanucchi, S., Melo, U. S., Corral-Serrano, J. C., et al. (2020). Structural variants create new topological-associated domains and ectopic retinal enhancer-gene contact in dominant retinitis pigmentosa. *Am. J. Hum. Genet.* 107 (5), 802–814. doi:10.1016/j.ajhg.2020.09.002

de Leeuw, C. A., Mooij, J. M., Heskes, T., and Posthuma, D. (2015). Magma: Generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* 11 (4), e1004219. doi:10.1371/journal.pcbi.1004219

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. *Science* 295 (5558), 1306–1311. doi:10.1126/science.1067799

Delaneau, O., Zazhytska, M., Borel, C., Giannuzzi, G., Rey, G., Howald, C., et al. (2019). Chromatin three-dimensional interactions mediate genetic effects on gene expression. *Science* 364, eaat8266. doi:10.1126/science.aat8266

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P. D., et al. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* 149 (6), 1233–1244. doi:10.1016/j.cell.2012.03.051

Despang, A., Schöpflin, R., Franke, M., Ali, S., Jerković, I., Paliou, C., et al. (2019). Functional dissection of the Sox9-Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. *Nat. Genet.* 51 (8), 1263–1271. doi:10.1038/s41588-019-0466-z

Dey, K. K., Gazal, S., van de Geijn, B., Kim, S. S., Nasser, J., Engreitz, J. M., et al. (2022). SNP-to-gene linking strategies reveal contributions of enhancer-related and candidate master-regulator genes to autoimmune disease. *Cell Genom.* 2, 100145. doi:10.1016/j.xgen.2022.100145

Dillon, N., Trimborn, T., Strouboulis, J., Fraser, P., and Grosveld, F. (1997). The effect of distance on long-range chromatin interactions. *Mol. Cell* 1 (1), 131–139. doi:10.1016/s1097-2765(00)80014-3

Ding, J., Frantzeskos, A., and Orozco, G. (2020). Functional genomics in autoimmune diseases. *Hum. Mol. Genet.* 29 (R1), R59–r65. doi:10.1093/hmg/ddaa097

Disney-Hogg, L., Kinnersley, B., and Houlston, R. (2020). Algorithmic considerations when analysing capture Hi-C data. *Wellcome Open Res.* 5, 289. doi:10.12688/wellcomeopenres.16394.1

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., et al. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485 (7398), 376–380. doi:10.1038/nature11082

Dostie, J., Richmond, T. A., Arnaout, R. A., Selzer, R. R., Lee, W. L., Honan, T. A., et al. (2006). Chromosome conformation capture carbon copy (5C): A massively parallel solution for mapping interactions between genomic elements. *Genome Res.* 16 (10), 1299–1309. doi:10.1101/gr.5571506

Downes, D. J., Cross, A. R., Hua, P., Roberts, N., Schwessinger, R., Cutler, A. J., et al. (2021). Identification of LZTFL1 as a candidate effector gene at a COVID-19 risk locus. *Nat. Genet.* 53 (11), 1606–1615. doi:10.1038/s41588-021-00955-3

Dryden, N. H., Broome, L. R., Dudbridge, F., Johnson, N., Orr, N., Schoenfelder, S., et al. (2014). Unbiased analysis of potential targets of breast cancer susceptibility loci by Capture Hi-C. *Genome Res.* 24 (11), 1854–1868. doi:10.1101/gr.175034.114

Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489 (7414), 57–74. doi:10.1038/nature11247

Eijsbouts, C. Q., Burren, O. S., Newcombe, P. J., and Wallace, C. (2019). Fine mapping chromatin contacts in capture Hi-C data. *BMC Genomics* 20 (1), 77. doi:10.1186/s12864-018-5314-5

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489 (7414), 57–74. doi:10.1038/nature11247

Ernst, J., and Kellis, M. (2017). Chromatin-state discovery and genome annotation with ChromHMM. *Nat. Protoc.* 12 (12), 2478–2492. doi:10.1038/nprot.2017.124

Fairfax, B. P., Humburg, P., Makino, S., Naranbhai, V., Wong, D., Lau, E., et al. (2014). Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* 343 (6175), 1246949. doi:10.1126/science.1246949

Fang, R., Yu, M., Li, G., Chee, S., Liu, T., Schmitt, A. D., et al. (2016). Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. *Cell Res.* 26 (12), 1345–1348. doi:10.1038/cr.2016.137

Farh, K. K., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W. J., Beik, S., et al. (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* 518 (7539), 337–343. doi:10.1038/nature13835

Forcato, M., Nicoletti, C., Pal, K., Livi, C. M., Ferrari, F., and Bicciato, S. (2017). Comparison of computational methods for Hi-C data analysis. *Nat. Methods* 14 (7), 679–685. doi:10.1038/nmeth.4325

Fornes, O., Castro-Mondragon, J. A., Khan, A., van der Lee, R., Zhang, X., Richmond, P. A., et al. (2020). Jaspar 2020: Update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 48 (D1), D87–d92. doi:10.1093/nar/gkz1001

Franke, M., Ibrahim, D. M., Andrey, G., Schwarzer, W., Heinrich, V., Schöpflin, R., et al. (2016). Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* 538 (7624), 265–269. doi:10.1038/nature19800

Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., et al. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449 (7164), 851–861. doi:10.1038/nature06258

Freire-Pritchett, P., Ray-Jones, H., Della Rosa, M., Eijsbouts, C. Q., Orchard, W. R., Wingett, S. W., et al. (2021). Detecting chromosomal interactions in Capture Hi-C data with CHiCAGO and companion tools. *Nat. Protoc.* 16 (9), 4144–4176. doi:10.1038/s41596-021-00567-5

Freire-Pritchett, P., Schoenfelder, S., Várnai, C., Wingett, S. W., Cairns, J., Collier, A. J., et al. (2017). Global reorganisation of cis-regulatory units upon lineage commitment of human embryonic stem cells. *Elife* 6, 21926. doi:10.7554/eLife.21926

French, J. D., and Edwards, S. L. (2020). The role of noncoding variants in heritable disease. *Trends Genet.* 36 (11), 880–891. doi:10.1016/j.tig.2020.07.004

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L. A. (2016). formation of chromosomal domains by loop extrusion. *Cell Rep.* 15 (9), 2038–2049. doi:10.1016/j.celrep.2016.04.085

Fulco, C. P., Nasser, J., Jones, T. R., Munson, G., Bergman, D. T., Subramanian, V., et al. (2019). Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* 51 (12), 1664–1669. doi:10.1038/s41588-019-0538-0

Fullwood, M. J., Liu, M. H., Pan, Y. F., Liu, J., Xu, H., Mohamed, Y. B., et al. (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* 462 (7269), 58–64. doi:10.1038/nature08497

Gabriele, M., Brandão, H. B., Grosse-Holz, S., Jha, A., Dailey, G. M., Cattoglio, C., et al. (2022). Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science* 376 (6592), 496–501. doi:10.1126/science.abn6583

Garrido-Martín, D., Borsari, B., Calvo, M., Reverter, F., and Guigó, R. (2021). Identification and analysis of splicing quantitative trait loci across multiple tissues in the human genome. *Nat. Commun.* 12 (1), 727. doi:10.1038/s41467-020-20578-2

Gasperini, M., Tome, J. M., and Shendure, J. (2020). Towards a comprehensive catalogue of validated and target-linked human enhancers. *Nat. Rev. Genet.* 21 (5), 292–310. doi:10.1038/s41576-019-0209-0

Gate, R. E., Cheng, C. S., Aiden, A. P., Siba, A., Tabaka, M., Lituiev, D., et al. (2018). Genetic determinants of co-accessible chromatin regions in activated T cells across humans. *Nat. Genet.* 50 (8), 1140–1150. doi:10.1038/s41588-018-0156-2

Gazal, S., Weissbrod, O., Hormozdiari, F., Dey, K. K., Nasser, J., Jagadeesh, K. A., et al. (2022). Combining SNP-to-gene linking strategies to identify disease genes and assess disease omnigenicity. *Nat. Genet.* 54 (6), 827–836. doi:10.1038/s41588-022-01087-y

Ghavi-Helm, Y., Klein, F. A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., et al. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* 512 (7512), 96–100. doi:10.1038/nature13417

Giambartolomei, C., Seo, J. H., Schwarz, T., Freund, M. K., Johnson, R. D., Spisak, S., et al. (2021). H3K27ac HiChIP in prostate cell lines identifies risk genes for prostate cancer susceptibility. *Am. J. Hum. Genet.* 108 (12), 2284–2300. doi:10.1016/j.ajhg.2021.11.007

Gong, H., Yang, Y., Zhang, S., Li, M., and Zhang, X. (2021). Application of Hi-C and other omics data analysis in human cancer and cell differentiation research. *Comput. Struct. Biotechnol. J.* 19, 2070–2083. doi:10.1016/j.csbj.2021.04.016

González-Serna, D., Shi, C., Kerick, M., Hankinson, J., Ding, J., McGovern, A., et al. (2022). Functional genomics in primary T cells and monocytes identifies mechanisms by which genetic susceptibility loci influence systemic sclerosis risk. *medRxiv Prepr.* doi:10.1101/2022.05.08.22274711

GTEx (2013). The genotype-tissue expression (GTEx) project. *Nat. Genet.* 45 (6), 580–585. doi:10.1038/ng.2653

Gu, B., Swigut, T., Spencley, A., Bauer, M. R., Chung, M., Meyer, T., et al. (2018). Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. *Science* 359 (6379), 1050–1055. doi:10.1126/science.aao3136

Holgersen, E. M., Gillespie, A., Leavy, O. C., Baxter, J. S., Zvereva, A., Muirhead, G., et al. (2021). Identifying high-confidence capture Hi-C interactions using CHiCANE. *Nat. Protoc.* 16 (4), 2257–2285. doi:10.1038/s41596-021-00498-1

Hormozdiari, F., van de Bunt, M., Segrè, A. V., Li, X., Joo, J. W. J., Bilow, M., et al. (2016). Colocalization of GWAS and eQTL signals detects target genes. *Am. J. Hum. Genet.* 99 (6), 1245–1260. doi:10.1016/j.ajhg.2016.10.003

Hsieh, T. H., Weiner, A., Lajoie, B., Dekker, J., Friedman, N., and Rando, O. J. (2015). Mapping nucleosome resolution chromosome folding in yeast by micro-C. *Cell* 162 (1), 108–119. doi:10.1016/j.cell.2015.05.048

Hua, P., Badat, M., Hanssen, L. L. P., Hentges, L. D., Crump, N., Downes, D. J., et al. (2021). Defining genome architecture at base-pair resolution. *Nature* 595 (7865), 125–129. doi:10.1038/s41586-021-03639-4

Hughes, J. R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., et al. (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat. Genet.* 46 (2), 205–212. doi:10.1038/ng.2871

Hutchinson, A., Asimit, J., and Wallace, C. (2020). Fine-mapping genetic associations. *Hum. Mol. Genet.* 29 (R1), R81–r88. doi:10.1093/hmg/ddaa148

Jansen, P. R., Watanabe, K., Stringer, S., Skene, N., Bryois, J., Hammerschlag, A. R., et al. (2019). Genome-wide analysis of insomnia in 1, 331, 010 individuals identifies new risk loci and functional pathways. *Nat. Genet.* 51 (3), 394–403. doi:10.1038/s41588-018-0333-3

Javierre, B. M., Burren, O. S., Wilder, S. P., Kreuzhuber, R., Hill, S. M., Sewitz, S., et al. (2016). Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* 167 (5), 1369–1384. doi:10.1016/j.cell.2016.09.037

Jin, F., Li, Y., Dixon, J. R., Selvaraj, S., Ye, Z., Lee, A. Y., et al. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* 503 (7475), 290–294. doi:10.1038/nature12644

Kichaev, G., Yang, W. Y., Lindstrom, S., Hormozdiari, F., Eskin, E., Price, A. L., et al. (2014). Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet.* 10 (10), e1004722. doi:10.1371/journal.pgen.1004722

Kim, J. H., Rege, M., Valeri, J., Dunagin, M. C., Metzger, A., Titus, K. R., et al. (2019). Ladl: Light-activated dynamic looping for endogenous gene expression control. *Nat. Methods* 16 (7), 633–639. doi:10.1038/s41592-019-0436-5

Klein, R. J., Zeiss, C., Chew, E. Y., Tsai, J. Y., Sackler, R. S., Haynes, C., et al. (2005). Complement factor H polymorphism in age-related macular degeneration. *Science* 308 (5720), 385–389. doi:10.1126/science.1109557

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409 (6822), 860–921. doi:10.1038/35057062

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., et al. (2018). Gene discovery and polygenic prediction from a genome-wide association study of

educational attainment in 1.1 million individuals. *Nat. Genet.* 50 (8), 1112–1121. doi:10.1038/s41588-018-0147-3

Lettice, L. A., Heaney, S. J., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., et al. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* 12 (14), 1725–1735. doi:10.1093/hmg/ddg180

Lettre, G. (2022). One step closer to linking GWAS SNPs with the right genes. *Nat. Genet.* 54 (6), 748–749. doi:10.1038/s41588-022-01093-0

Li, T., Jia, L., Cao, Y., Chen, Q., and Li, C. (2018). OCEAN-C: Mapping hubs of open chromatin interactions across the genome reveals gene regulatory networks. *Genome Biol.* 19 (1), 54. doi:10.1186/s13059-018-1430-4

Li, W., Wong, W. H., and Jiang, R. (2019). DeepTACT: Predicting 3D chromatin contacts via bootstrapping deep learning. *Nucleic Acids Res.* 47 (10), e60. doi:10.1093/nar/gkz167

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326 (5950), 289–293. doi:10.1126/science.1181369

Luo, Z., Zhang, R., Hu, T., Zhu, Y., Wu, Y., Li, W., et al. (2022). NicE-C efficiently reveals open chromatin-associated chromosome interactions at high resolution. *Genome Res.* 32 (3), 534–544. doi:10.1101/gr.275986.121

Lupiáñez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 161 (5), 1012–1025. doi:10.1016/j.cell.2015.04.004

Maller, J. B., McVean, G., Byrnes, J., Vukcevic, D., Palin, K., Su, Z., et al. (2012). Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat. Genet.* 44 (12), 1294–1301. doi:10.1038/ng.2435

Martin, P., McGovern, A., Orozco, G., Duffus, K., Yarwood, A., Schoenfelder, S., et al. (2015). Capture Hi-C reveals novel candidate genes and complex long-range interactions with related autoimmune risk loci. *Nat. Commun.* 6, 10069. doi:10.1038/ncomms10069

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., et al. (2016). A reference panel of 64, 976 haplotypes for genotype imputation. *Nat. Genet.* 48 (10), 1279–1283. doi:10.1038/ng.3643

McVicker, G., van de Geijn, B., Degner, J. F., Cain, C. E., Banovich, N. E., Raj, A., et al. (2013). Identification of genetic variants that affect histone modifications in human cells. *Science* 342 (6159), 747–749. doi:10.1126/science.1242429

Mifsud, B., Martincorena, I., Darbo, E., Sugar, R., Schoenfelder, S., Fraser, P., et al. (2017). Gothic, a probabilistic model to resolve complex biases and to identify real interactions in Hi-C data. *PLoS One* 12, e0174744. doi:10.1371/journal.pone.0174744

Mifsud, B., Tavares-Cadete, F., Young, A. N., Sugar, R., Schoenfelder, S., Ferreira, L., et al. (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* 47 (6), 598–606. doi:10.1038/ng.3286

Moore, J. E., Purcaro, M. J., Pratt, H. E., Epstein, C. B., Shoresh, N., Adrian, J., et al. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* 583 (7818), 699–710. doi:10.1038/s41586-020-2493-4

Morgan, S. L., Mariano, N. C., Bermudez, A., Arruda, N. L., Wu, F., Luo, Y., et al. (2017). Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. *Nat. Commun.* 8, 15993. doi:10.1038/ncomms15993

Moschen, A. R., Tilg, H., and Raine, T. (2019). IL-12, IL-23 and IL-17 in IBD: Immunobiology and therapeutic targeting. *Nat. Rev. Gastroenterol. Hepatol.* 16 (3), 185–196. doi:10.1038/s41575-018-0084-8

Mumbach, M. R., Rubin, A. J., Flynn, R. A., Dai, C., Khavari, P. A., Greenleaf, W. J., et al. (2016). HiChIP: Efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* 13 (11), 919–922. doi:10.1038/nmeth.3999

Mumbach, M. R., Satpathy, A. T., Boyle, E. A., Dai, C., Gowen, B. G., Cho, S. W., et al. (2017). Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat. Genet.* 49 (11), 1602–1612. doi:10.1038/ng.3963

Narendra, V., Rocha, P. P., An, D., Raviram, R., Skok, J. A., Mazzoni, E. O., et al. (2015). CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* 347 (6225), 1017–1021. doi:10.1126/science.1262088

Nasser, J., Bergman, D. T., Fulco, C. P., Guckelberger, P., Doughty, B. R., Patwardhan, T. A., et al. (2021). Genome-wide enhancer maps link risk variants to disease genes. *Nature* 593 (7858), 238–243. doi:10.1038/s41586-021-03446-x

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485 (7398), 381–385. doi:10.1038/nature11049

Novo, C. L., Javierre, B. M., Cairns, J., Segonds-Pichon, A., Wingett, S. W., Freire-Pritchett, P., et al. (2018). Long-range enhancer interactions are prevalent in mouse embryonic stem cells and are reorganized upon pluripotent state transition. *Cell Rep.* 22 (10), 2615–2627. doi:10.1016/j.celrep.2018.02.040

Oluwadare, O., Highsmith, M., and Cheng, J. (2019). An overview of methods for reconstructing 3-D chromosome and genome structures from Hi-C data. *Biol. Proced. Online* 21, 7. doi:10.1186/s12575-019-0094-0

Orozco, G. (2022). Fine mapping with epigenetic information and 3D structure. *Semin. Immunopathol.* 44 (1), 115–125. doi:10.1007/s00281-021-00906-4

Pal, K., Forcato, M., and Ferrari, F. (2019). Hi-C analysis: From data generation to integration. *Biophys. Rev.* 11 (1), 67–78. doi:10.1007/s12551-018-0489-1

Paliou, C., Guckelberger, P., Schöpflin, R., Heinrich, V., Esposito, A., Chiariello, A. M., et al. (2019). Preformed chromatin topology assists transcriptional robustness of Shh during limb development. *Proc. Natl. Acad. Sci. U. S. A.* 116 (25), 12390–12399. doi:10.1073/pnas.1900672116

Pasaniuc, B., and Price, A. L. (2017). Dissecting the genetics of complex traits using summary association statistics. *Nat. Rev. Genet.* 18 (2), 117–127. doi:10.1038/nrg.2016.142

Piecyk, R. S., Schlegel, L., and Johannes, F. (2022). Predicting 3D chromatin interactions from DNA sequence using Deep Learning. *Comput. Struct. Biotechnol. J.* 20, 3439–3448. doi:10.1016/j.csbj.2022.06.047

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470 (7333), 279–283. doi:10.1038/nature09692

Rahim, N. G., Harismendy, O., Topol, E. J., and Frazer, K. A. (2008). Genetic determinants of phenotypic diversity in humans. *Genome Biol.* 9 (4), 215. doi:10.1186/gb-2008-9-4-215

Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159 (7), 1665–1680. doi:10.1016/j.cell.2014.11.021

Redolfi, J., Zhan, Y., Valdes-Quezada, C., Kryzhanovska, M., Guerreiro, I., Iesmantavicius, V., et al. (2019). DamC reveals principles of chromatin folding *in vivo* without crosslinking and ligation. *Nat. Struct. Mol. Biol.* 26 (6), 471–480. doi:10.1038/s41594-019-0231-0

Rubin, A. J., Barajas, B. C., Furlan-Magaril, M., Lopez-Pajares, V., Mumbach, M. R., Howard, I., et al. (2017). Lineage-specific dynamic and pre-established enhancer-promoter contacts cooperate in terminal differentiation. *Nat. Genet.* 49 (10), 1522–1528. doi:10.1038/ng.3935

Sahlén, P., Abdullayev, I., Ramsköld, D., Matskova, L., Rilakovic, N., Lötstedt, B., et al. (2015). Genome-wide mapping of promoter-anchored interactions with close to single-enhancer resolution. *Genome Biol.* 16 (1), 156. doi:10.1186/s13059-015-0727-9

Sanborn, A. L., Rao, S. S., Huang, S. C., Durand, N. C., Huntley, M. H., Jewett, A. I., et al. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U. S. A.* 112 (47), E6456–E6465. doi:10.1073/pnas.1518552112

Schaid, D. J., Chen, W., and Larson, N. B. (2018). From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.* 19 (8), 491–504. doi:10.1038/s41576-018-0016-z

Schmidt, F., Kern, F., and Schulz, M. H. (2020). Integrative prediction of gene expression with chromatin accessibility and conformation data. *Epigenetics Chromatin* 13 (1), 4. doi:10.1186/s13072-020-0327-0

Schoenfelder, S., and Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. *Nat. Rev. Genet.* 20 (8), 437–455. doi:10.1038/s41576-019-0128-0

Schoenfelder, S., Furlan-Magaril, M., Mifsud, B., Tavares-Cadete, F., Sugar, R., Javierre, B. M., et al. (2015a). The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res.* 25 (4), 582–597. doi:10.1101/gr.185272.114

Schoenfelder, S., Mifsud, B., Senner, C. E., Todd, C. D., Chrysanthou, S., Darbo, E., et al. (2018). Divergent wiring of repressive and active chromatin interactions between mouse embryonic and trophoblast lineages. *Nat. Commun.* 9 (1), 4189. doi:10.1038/s41467-018-06666-4

Schoenfelder, S., Sugar, R., Dimond, A., Javierre, B. M., Armstrong, H., Mifsud, B., et al. (2015b). Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. *Nat. Genet.* 47 (10), 1179–1186. doi:10.1038/ng.3393

Segert, J. A., Gisselbrecht, S. S., and Bulyk, M. L. (2021). Transcriptional silencers: Driving gene expression with the brakes on. *Trends Genet.* 37 (6), 514–527. doi:10.1016/j.tig.2021.02.002

Servant, N., Varoquaux, N., Lajoie, B. R., Viara, E., Chen, C. J., Vert, J. P., et al. (2015). HiC-pro: An optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* 16, 259. doi:10.1186/s13059-015-0831-x

Sey, N. Y. A., Hu, B., Mah, W., Fauni, H., McAfee, J. C., Rajarajan, P., et al. (2020). A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles. *Nat. Neurosci.* 23 (4), 583–593. doi:10.1038/s41593-020-0603-0

Shen, Y., Yue, F., McCleary, D. F., Ye, Z., Edsall, L., Kuan, S., et al. (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature* 488 (7409), 116–120. doi:10.1038/nature11243

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., et al. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 29 (1), 308–311. doi:10.1093/nar/29.1.308

Siersbæk, R., Madsen, J. G. S., Javierre, B. M., Nielsen, R., Bagge, E. K., Cairns, J., et al. (2017). Dynamic rewiring of promoter-anchored chromatin loops during adipocyte differentiation. *Mol. Cell* 66 (3), 420–435. doi:10.1016/j.molcel.2017.04.010

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., et al. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.* 38 (11), 1348–1354. doi:10.1038/ng1896

Song, M., Yang, X., Ren, X., Maliskova, L., Li, B., Jones, I. R., et al. (2019). Mapping cis-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes. *Nat. Genet.* 51 (8), 1252–1262. doi:10.1038/s41588-019-0472-1

Spain, S. L., and Barrett, J. C. (2015). Strategies for fine-mapping complex traits. *Hum. Mol. Genet.* 24 (R1), R111–R119. doi:10.1093/hmg/ddv260

Stunnenberg, H. G., and Hirst, M. (2016). The international human epigenome consortium: A blueprint for scientific collaboration and discovery. *Cell* 167 (5), 1897–1149. doi:10.1016/j.cell.2016.12.002

Su, C., Johnson, M. E., Torres, A., Thomas, R. M., Manduchi, E., Sharma, P., et al. (2020). Mapping effector genes at lupus GWAS loci using promoter Capture-C in follicular helper T cells. *Nat. Commun.* 11 (1), 3294. doi:10.1038/s41467-020-17089-5

Su, C., Pahl, M. C., Grant, S. F. A., and Wells, A. D. (2021). Restriction enzyme selection dictates detection range sensitivity in chromatin conformation capture-based variant-to-gene mapping approaches. *Hum. Genet.* 140 (10), 1441–1448. doi:10.1007/s00439-021-02326-8

Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., and Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* 20 (8), 467–484. doi:10.1038/s41576-019-0127-1

The Wellcome Trust Case Control Consortium (2007). Genome-wide association study of 14, 000 cases of seven common diseases and 3, 000 shared controls. *Nature* 447 (7145), 661–678. doi:10.1038/nature05911

Thiecke, M. J., Wutz, G., Muhar, M., Tang, W., Bevan, S., Malysheva, V., et al. (2020). Cohesin-dependent and -independent mechanisms mediate chromosomal contacts between promoters and enhancers. *Cell Rep.* 32 (3), 107929. doi:10.1016/j.celrep.2020.107929

Trynka, G., Sandor, C., Han, B., Xu, H., Stranger, B. E., Liu, X. S., et al. (2013). Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat. Genet.* 45 (2), 124–130. doi:10.1038/ng.2504

Udler, M. S., McCarthy, M. I., Florez, J. C., and Mahajan, A. (2019). Genetic risk scores for diabetes diagnosis and precision medicine. *Endocr. Rev.* 40 (6), 1500–1520. doi:10.1210/er.2019-00088

Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., et al. (2021). Genome-wide association studies. *Nat. Rev. Methods Prim.* 1 (1), 59. doi:10.1038/s43586-021-00056-9

Uslu, V. V., Petretich, M., Ruf, S., Langenfeld, K., Fonseca, N. A., Marioni, J. C., et al. (2014). Long-range enhancers regulating Myc expression are required for normal facial morphogenesis. *Nat. Genet.* 46 (7), 753–758. doi:10.1038/ng.2971

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., et al. (2017). 10 Years of GWAS discovery: Biology, function, and translation. *Am. J. Hum. Genet.* 101 (1), 5–22. doi:10.1016/j.ajhg.2017.06.005

Wakefield, J. (2009). Bayes factors for genome-wide association studies: Comparison with P-values. *Genet. Epidemiol.* 33 (1), 79–86. doi:10.1002/gepi.20359

Wang, D., Liu, S., Warrell, J., Won, H., Shi, X., Navarro, F. C. P., et al. (2018). Comprehensive functional genomic resource and integrative model for the human brain. *Science* 362, eaat8464. doi:10.1126/science.aat8464

Wang, K., Zhang, H., Kugathasan, S., Annese, V., Bradfield, J. P., Russell, R. K., et al. (2009). Diverse genome-wide association studies associate the IL12/IL23 pathway with Crohn Disease. *Am. J. Hum. Genet.* 84 (3), 399–405. doi:10.1016/j.ajhg.2009.01.026

Wang, Q. S., and Huang, H. (2022). Methods for statistical fine-mapping and their applications to auto-immune diseases. *Semin. Immunopathol.* 44 (1), 101–113. doi:10.1007/s00281-021-00902-8

Wei, X., Xiang, Y., Peters, D. T., Marius, C., Sun, T., Shan, R., et al. (2022). HiCAR is a robust and sensitive method to analyze open-chromatin-associated genome organization. *Mol. Cell* 82 (6), 1225–1238.e6. doi:10.1016/j.molcel.2022.01.023

Weissbrod, O., Hormozdiari, F., Benner, C., Cui, R., Ulirsch, J., Gazal, S., et al. (2020). Functionally informed fine-mapping and polygenic localization of complex trait heritability. *Nat. Genet.* 52 (12), 1355–1363. doi:10.1038/s41588-020-00735-5

Williamson, I., Kane, L., Devenney, P. S., Flyamer, I. M., Anderson, E., Kilanowski, F., et al. (2019). Developmentally regulated Shh expression is robust to TAD perturbations. *Development* 146. doi:10.1242/dev.179523

Williamson, I., Lettice, L. A., Hill, R. E., and Bickmore, W. A. (2016). Shh and ZRS enhancer colocalisation is specific to the zone of polarising activity. *Development* 143 (16), 2994–3001. doi:10.1242/dev.139188

Wilson, N. K., Schoenfelder, S., Hannah, R., Sánchez Castillo, M., Schütte, J., Ladopoulos, V., et al. (2016). Integrated genome-scale analysis of the transcriptional regulatory landscape in a blood stem/progenitor cell model. *Blood* 127 (13), e12–e23. doi:10.1182/blood-2015-10-677393

Wingett, S., Ewels, P., Furlan-Magaril, M., Nagano, T., Schoenfelder, S., Fraser, P., et al. (2015). HiCUP: Pipeline for mapping and processing Hi-C data. *F1000Res.* 4, 1310. doi:10.12688/f1000research.7334.1

Wlasnowolski, M., Sadowski, M., Czarnota, T., Jodkowska, K., Szalaj, P., Tang, Z., et al. (2020). 3D-GNOME 2.0: A three-dimensional genome modeling engine for predicting structural variation-driven alterations of chromatin spatial structure in the human genome. *Nucleic Acids Res.* 48, W170–w176. doi:10.1093/nar/gkaa388

Yang, J., McGovern, A., Martin, P., Duffus, K., Ge, X., Zarrineh, P., et al. (2020). Analysis of chromatin organization and gene expression in T cells identifies functional genes for rheumatoid arthritis. *Nat. Commun.* 11 (1), 4402. doi:10.1038/s41467-020-18180-7

Yao, C., Chen, G., Song, C., Keefe, J., Mendelson, M., Huan, T., et al. (2018). Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease. *Nat. Commun.* 9 (1), 3268. doi:10.1038/s41467-018-05512-x

Yardımcı, G. G., and Noble, W. S. (2017). Software tools for visualizing Hi-C data. *Genome Biol.* 18 (1), 26. doi:10.1186/s13059-017-1161-y

Zeggini, E., Gloyn, A. L., Barton, A. C., and Wain, L. V. (2019). Translational genomics and precision medicine: Moving from the lab to the clinic. *Science* 365 (6460), 1409–1413. doi:10.1126/science.aax4588

Zentner, G. E., Tesar, P. J., and Scacheri, P. C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res.* 21 (8), 1273–1283. doi:10.1101/gr.122382.111

Zhang, S. (2022). A simplified protocol for performing MAGMA/H-MAGMA gene set analysis utilizing high-performance computing environments. *Star. Protoc.* 3 (1), 101083. doi:10.1016/j.xpro.2021.101083

Zhang, Y., Li, Z., Bian, S., Zhao, H., Feng, D., Chen, Y., et al. (2020). HiCoP, a simple and robust method for detecting interactions of regulatory regions. *Epigenetics Chromatin* 13 (1), 27. doi:10.1186/s13072-020-00348-6

Zhao, Z., Tavoosidana, G., Sjölinder, M., Göndör, A., Mariano, P., Wang, S., et al. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.* 38 (11), 1341–1347. doi:10.1038/ng1891

Zheng, Y., Ay, F., and Keles, S. (2019). Generative modeling of multi-mapping reads with mHi-C advances analysis of Hi-C studies. *Elife* 8, e38070. doi:10.7554/eLife.38070

Nomenclature

ABC Activity by Contact

ATAC-seq Assay for Transposase-Accessible Chromatin with high-throughput sequencing

ChIA-PET Chromatin Interaction Analysis by Paired-End Tag sequencing

ChIP Chromatin immunoprecipitation

ChIP-seq Chromatin immunoprecipitation with massively parallel DNA sequencing

COGS Common Omnibus Gene Score

CRISPR Clustered regularly interspaced short palindromic repeats

eQTL Expression quantitative trait locus

GWA Genome wide association

GWAS Genome wide association studies

Hi-C High-throughput chromosome conformation capture

IBD Inflammatory bowel disease

LD Linkage disequilibrium

Micro-C/MCC Micrococcal nuclease capturing chromosome conformation

PCHi-C Promoter-Capture Hi-C

PIP Posterior Inclusion Probability

PIR Promoter interacting regions

PLAC-seq Proximity ligation assisted ChIP-seq

QTL Quantitative trait locus

RNA-seq RNA sequencing

SNP Single nucleotide polymorphism

TAD Topologically associating domain

Tiled-MCC Micro-Capture-C