© 2023. Published by The Company of Biologists Ltd | Disease Models & Mechanisms (2023) 16, dmm049790. doi:10.1242/dmm.049790

REVIEW

GENETIC VARIANCE IN HUMAN DISEASE:
DECODING DIVERSITY TO ADVANCE MODERN MEDICINE

Translating non-coding genetic associations into a better understanding of immune-mediated disease

Christina T. Stankey[1,2] and James C. Lee[1,3,*]

ABSTRACT

Genome-wide association studies have identified hundreds of genetic loci that are associated with immune-mediated diseases. Most disease-associated variants are non-coding, and a large proportion of these variants lie within enhancers. As a result, there is a pressing need to understand how common genetic variation might affect enhancer function and thereby contribute to immune-mediated (and other) diseases. In this Review, we first describe statistical and experimental methods to identify causal genetic variants that modulate gene expression, including statistical fine-mapping and massively parallel reporter assays. We then discuss approaches to characterise the mechanisms by which these variants modulate immune function, such as clustered regularly interspaced short palindromic repeats (CRISPR)-based screens. We highlight examples of studies that, by elucidating the effects of disease variants within enhancers, have provided important insights into immune function and uncovered key pathways of disease.

KEY WORDS: Genetics, Genomics, Immune-mediated disease, Immunology

Introduction

Autoimmune and inflammatory diseases are a heterogeneous set of complex disorders characterised by an aberrant immune response that damages host tissue. These immune-mediated diseases – which range from inflammatory bowel disease to multiple sclerosis to systemic lupus erythematosus – affect up to 10% of individuals in Western countries and are rapidly increasing in incidence in other parts of the world (Freeman et al., 2021; King et al., 2020; Patterson et al., 2019; Safiri et al., 2019; Walton et al., 2020). Almost all of these conditions require better treatments, but the attrition rate of drugs entering clinical development is high – mainly because our understanding of disease aetiology and mechanisms remains incomplete.

Familial patterns of inheritance have long suggested a genetic contribution to immune-mediated diseases. Indeed, despite a growing appreciation of the environmental factors involved in disease development, genetics remains the strongest risk factor for most immune-mediated conditions. Early linkage studies identified a handful of genetic associations, typically with large effect sizes. For example, several studies detected hits at the human leukocyte antigen (HLA) locus, such as the association of *HLA-DQA1* and *HLA-DQB1* haplotypes (see Glossary, Box 1) with coeliac disease and of *HLA-DQB* with type 1 diabetes mellitus (Sollid et al., 1989; Todd et al., 1987). A smaller number of studies also identified novel susceptibility genes, such as nucleotide-binding oligomerisation domain-containing protein 2 (*NOD2*) in Crohn's disease (Hugot et al., 2001). However, the genetic basis of most immune-mediated diseases remained largely unknown until the advent of genome-wide association studies (GWAS; Box 1). This technological advance made it possible to detect disease-associated genetic variants with modest effect sizes that collectively account for a large proportion of disease risk (Visscher et al., 2017). Fuelled by large sample sizes and international collaborations, GWAS rapidly identified hundreds of loci associated with immune-mediated diseases [Anderson et al., 2011; International Genetics of Ankylosing Spondylitis Consortium (IGAS) et al., 2013; Tsoi et al., 2012]. Collectively, these results demonstrated the following: (1) the polygenic nature of most immune-mediated diseases, where many common genetic variants individually contribute a small amount to disease susceptibility; (2) the widespread role of pleiotropy (Box 1), where a genetic haplotype confers susceptibility to multiple immune-mediated diseases; and (3) the importance of core pathways, where immune-mediated disease associations cluster within key biological processes such as T-cell activation or cytokine signalling (Cotsapas et al., 2011; Ellinghaus et al., 2016; Schmiedel et al., 2018). In a few cases, GWAS identified susceptibility variants in pathways not previously thought to be involved in a particular disease and thereby generated new insights into pathobiology. This is best exemplified by the characterisation of defective autophagy in Crohn's disease pathogenesis (Hampe et al., 2007; Rioux et al., 2007). Despite this success, however, the molecular basis of most genetic associations – either with immune-mediated diseases or complex diseases more generally – remains unresolved. Indeed, translating the success of GWAS into an improved understanding of disease mechanisms is arguably the greatest challenge in modern human genetics.

Most immune disease-associated genetic variants map to non-coding regions of the genome and are particularly enriched in immune cell enhancers – regulatory sequences that interact with promoters and transcriptional machinery to increase the expression of target genes (Ernst et al., 2011; Farh et al., 2015; Maurano et al., 2012). Disease-associated variants within enhancers have therefore been hypothesised to disrupt transcriptional circuits and dysregulate key processes in immune cell biology (Javierre et al., 2016). Identifying the molecular and cellular mechanisms involved could provide important insights into immune regulation and disease biology. However, attempts to mechanistically characterise regulatory variants have been hindered by the need to identify the

[1]Genetic Mechanisms of Disease Laboratory, The Francis Crick Institute, London NW1 1AT, UK. [2]Department of Immunology and Inflammation, Imperial College London, London W12 0NN, UK. [3]Institute of Liver and Digestive Health, Royal Free Hospital, University College London, London NW3 2PF, UK.

*Author for correspondence (james.lee@crick.ac.uk)

C.T.S., 0000-0001-5710-1716; J.C.L., 0000-0001-5711-9385

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

Disease Models & Mechanisms (2023) 16, dmm049790. doi:10.1242/dmm.049790

Box 1. Glossary

4C: a high-resolution 'one-versus-all' approach to map chromatin interactions. Following cross-linking, digestion, and ligation, chimeric DNA fragments are circularised and PCR amplified using primers specific to one locus of interest. High-throughput sequencing is then used to identify all fragments interacting with this locus.

Allelic imbalance: a phenomenon in a heterozygous individual where some measure of the two alleles diverges from the expected 1:1 ratio for a particular cellular trait.

Assay for transposase-accessible chromatin sequencing (ATAC-seq): a method to assess chromatin accessibility that uses a hyperactive Tn5 to simultaneously cleave open chromatin and insert adaptors for high-throughput sequencing. Sequenced reads are mapped to identify regions of open chromatin.

Bayesian methods: a set of statistical methods that estimate the probability of a hypothesis based on existing data and the prior probability of the hypothesis. In statistical fine-mapping, Bayesian methods are used to calculate the posterior probability that a variant is causal given the linkage disequilibrium structure of the locus.

Capture-C: a high-resolution 'many-versus-all' approach to map chromatin interactions. Following cross-linking, digestion and ligation, loci of interest are enriched through pulldown with oligonucleotide probes. High-throughput sequencing is then used to identify fragments interacting with these loci.

Chromatin interaction analysis with paired-end tag (ChIA-PET): a high-resolution method to identify interacting chromatin bound by a protein of interest. Chromatin is cross-linked, fragmented and immunoprecipitated to obtain DNA–protein of interest complexes. DNA is ligated with biotinylated adaptors, then is pulled down and sequenced to identify chromatin interactions with specific proteins.

Credible set: the smallest set of variants that contains the true causal variant(s) at a defined probability.

DNase I hypersensitivity sites sequencing (DNase-seq): a method to assess chromatin accessibility whereby nuclei are digested with DNase I, an enzyme that cleaves chromatin at regions free of nucleosomes or transcription factors. Cleaved fragments are sequenced and mapped to identify DNase hypersensitive sites, which denote regions of open chromatin.

Genome-wide association studies (GWAS): an experimental approach to identify associations between genotype and phenotype. Study populations are selected and genotyped using microarrays or whole-genome sequencing, then the genotyped variants are tested for association with the phenotype of interest.

Haplotype: a combination of alleles at a locus that tend to be inherited together.

Hi-C: a lower-resolution 'all-versus-all' approach to map chromatin interactions. Following cross-linking and digestion, biotin-tagged ligation fragments are generated. Biotinylated fragments are enriched by pulldown, and high-throughput sequencing of the fragments is used to identify all interacting pairs of loci.

Insulator: a regulatory element that functions either by blocking interactions of enhancers and promoters or by preventing spread of heterochromatin.

Jurkat-dCas9-VP64 cells: an immortalised human $CD4^+$ T-cell line that constitutively expresses dCas9-VP64.

Mediator: a key transcriptional co-activator complex at enhancers.

Nuclear factor kappa B (NF-κB): a key transcriptional mediator of inflammatory processes across immune lineages.

Pleiotropy: the association of a single genotype with multiple phenotypes.

Protospacer adjacent motif (PAM): a short nucleotide sequence that is necessary for recognition and cleavage of the target sequence by a Cas protein. The PAM sequence depends on the Cas protein used.

Quantitative trait loci (QTL): loci at which a genetic variant is associated with a measurable phenotypic trait.

Regulatory T (Treg) cells: a subpopulation of $CD4^+$ T cells that suppresses inflammatory processes and contributes to maintenance of immunological tolerance.

RNA editing: a process that prevents formation of immunogenic double-stranded RNA molecules.

RNA splicing: a process that generates different transcript isoforms and can affect transcript stability.

Tiling: an approach for designing constructs for a high-throughput screen to cover a genomic region of interest at even intervals. Constructs are usually designed with an offset of one or a few base pairs to ensure high-resolution characterisation of the region.

target gene(s), the cell types in which these gene(s) are dysregulated and the conditions under which this dysregulation occurs.

Recently developed genetic and genomic techniques have now made it possible to elucidate the effects of genetic variants within enhancers and thus provide a means to translate GWAS associations into an understanding of disease biology (Bourges et al., 2020; Nasrallah et al., 2020; Simeonov et al., 2017). Here, we review these techniques in the context of immune-mediated disease. We describe approaches to identify putative causal variants, including statistical methods, such as fine-mapping and colocalisation with chromatin features, and experimental methods, such as high-throughput reporter assays. We also review approaches to resolve the biological effects of regulatory variants in immune-mediated diseases, including chromatin interaction maps and clustered regularly interspaced sport palindromic repeats (CRISPR)-based screens. Furthermore, we provide examples of recent studies that have mechanistically linked genetic variation in enhancers to immune dysregulation, highlighting the potential for genetics to reveal key pathways in disease biology and identify new therapeutic targets.

Challenges to translating associations into disease mechanisms

Two key challenges must be overcome to successfully translate genetic associations into a better understanding of immune disease mechanisms. First, we need to identify the causal variants for which functional consequences directly alter disease risk. This is not a trivial task because an associated locus can contain tens to hundreds of candidate variants. Ironically, this problem results from the haplotype structure of the human genome, which was what enabled the development of GWAS in the first place (Daly et al., 2001). Specifically, the human genome is structured into 'haplotype blocks', genomic regions spanning tens to hundreds of kilobases that are typically inherited as single units during meiosis (Fig. 1) (Gabriel et al., 2002). As such, single-nucleotide polymorphisms (SNPs) within a haplotype block are frequently inherited together, a phenomenon that is known as 'linkage disequilibrium' (LD) (Hästbacka et al., 1992). The allelic variation at any one SNP within a haplotype block can therefore be used to infer the variation at all other SNPs in the block. This observation paved the way for GWAS because it meant that genome-wide genetic variation could

Disease Models & Mechanisms (2023) 16, dmm049790. doi:10.1242/dmm.049790

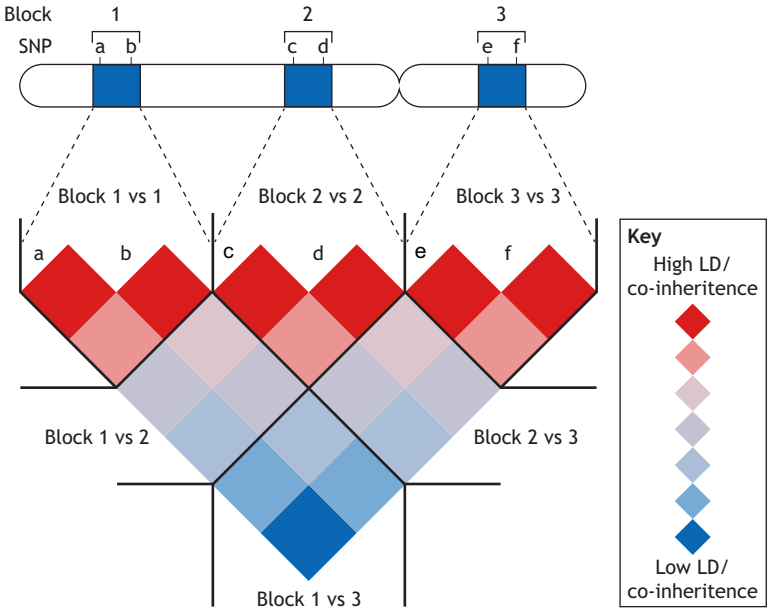


Fig. 1. Haplotype blocks and linkage disequilibrium. Haplotype blocks (blue rectangles, top; 1, 2 and 3) are regions on a chromosome (top) that are inherited as a unit, such that variants or single-nucleotide polymorphisms (SNPs; a, b, c, d, e and f) within the blocks tend to be inherited together. This gives rise to linkage disequilibrium (LD), where inheritance of common genetic variants within a haplotype block is highly correlated (red diamonds; high LD/co-inheritance, e.g. Block 1 vs 1), whereas inheritance of variants in different haplotype blocks is weakly correlated (light-red/light-blue diamonds; weak LD/co-inheritance, e.g. Block 1 vs 2) or not correlated (blue diamonds; low LD/co-inheritance, e.g. Block 1 vs 3).

Disease Models & Mechanisms

be captured by genotyping only a subset of all SNPs (Altshuler et al., 2008). However, it also means that, for any disease-associated locus, there are often many candidate SNPs that could be responsible for the association.

Second, we need to characterise the functional consequences of causal variants on downstream biological pathways. This challenge is not straightforward because only a small proportion of disease-associated variants map to protein-coding regions of the genome (Gusev et al., 2014; Hindorff et al., 2009). Indeed, although we now know that most non-coding disease variants fall within putative enhancers, the specific genes that these enhancers regulate are often unclear, with some variants lying near multiple plausible candidate genes and others residing within intergenic regions that contain no genes (Mifsud et al., 2015; Onengut-Gumuscu et al., 2015). Even when a target gene can be identified, many enhancers exhibit both cell-type- and cell-state-specific activity (Farh et al., 2015). Accordingly, both the appropriate cell type and the disease-relevant external stimuli/conditions must be determined before the functional consequences of a causal variant can be fully resolved.

Challenge 1: identifying the causal variant

A variety of approaches have been developed to identify causal variants at disease-associated loci. These include statistical methods to nominate putative causal variants based on patterns of association, methods to compare the location of candidate variants with known functional genomic regions, and experimental methods to directly assess variants for gene expression-modulating activity. The advantages and limitations of these approaches are summarised in Table 1.

Table 1. Methods to identify causal variants

Technique	Advantages	Limitations	References
Statistical fine-mapping	Narrows GWAS associations to a smaller credible set of variants for functional follow-up.	Rarely refines association to a single putative causal variant.	Huang et al., 2017; Wellcome Trust Case Control Consortium et al., 2012; Wallace et al., 2015
Chromatin annotation	Identifies variants most likely to be functional based on localisation to active regulatory elements. Can implicate cell populations or tissues in disease processes based on selective enrichment of GWAS variants in regulatory elements.	Requires annotations from disease-relevant cellular contexts that are often unavailable. Evidence of causality is correlative.	Degner et al., 2012; Farh et al., 2015; Maurano et al., 2012; Trynka et al., 2012
MPRA	Directly assesses transcriptional activity using large libraries of putative regulatory elements. Identifies variants with expression-modulating effects.	Evaluates transcriptional activity in an episomal context. Depends on transcription factor repertoire of transfected cells so requires appropriate cellular model.	Bourges et al., 2020; Melnikov et al., 2012; Patwardhan et al., 2012; Tewhey et al., 2016
STARR-seq	Directly assesses transcriptional activity of large libraries of putative regulatory elements. Can test longer sequences than possible with MPRA.	Evaluates transcriptional activity in an episomal context. Depends on transcription factor repertoire of transfected cells so requires appropriate cellular model. Can only test regulatory elements and alleles that are present in the source genome.	Arnold et al., 2013; Liu et al., 2017; van Ouwekerk et al., 2020

GWAS, genome-wide association studies; MPRA, massively parallel reporter assay; STARR-seq, self-transcribing regulatory region sequencing.

Disease Models & Mechanisms (2023) 16, dmm049790. doi:10.1242/dmm.049790

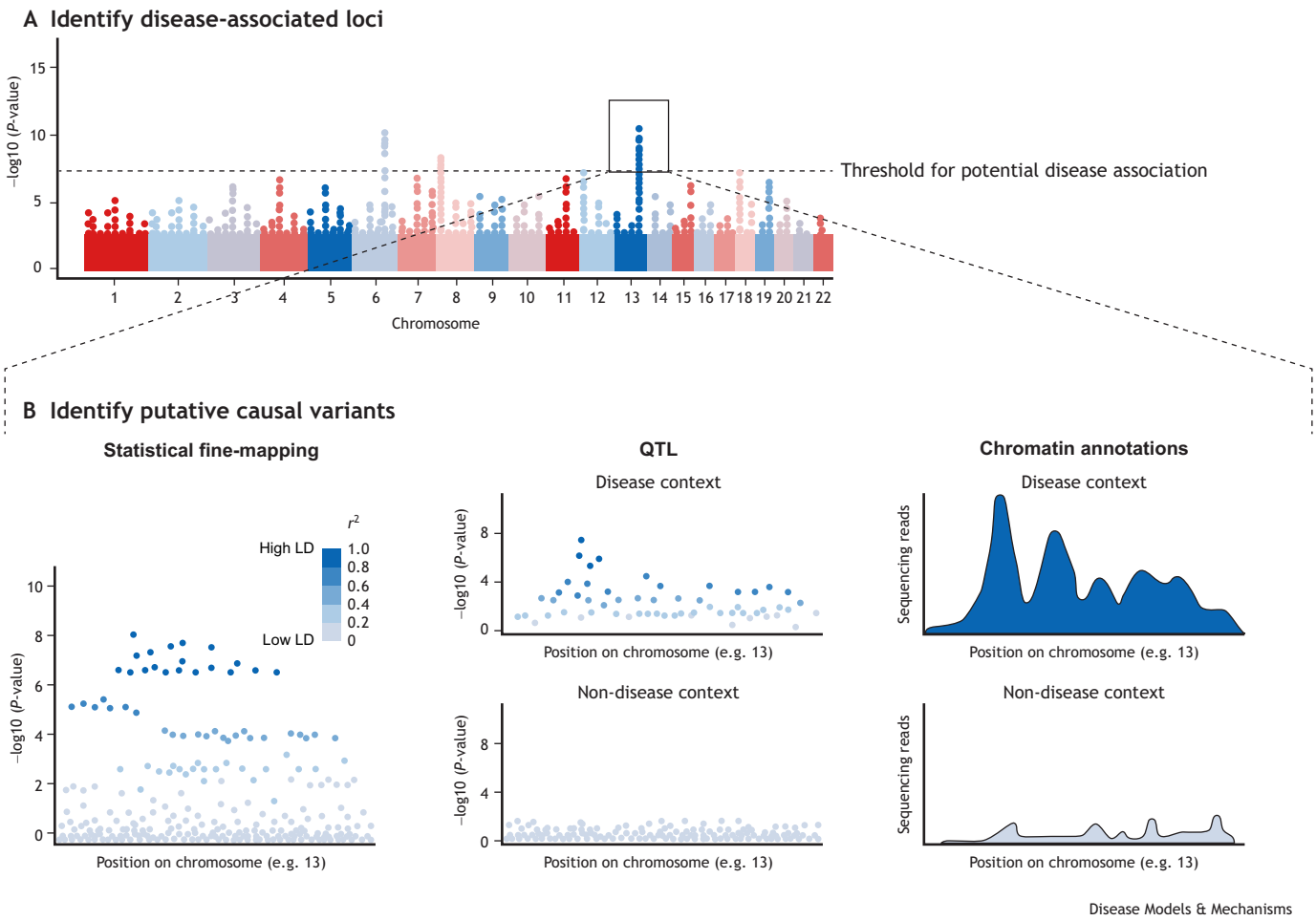


Fig. 2. Statistical methods to identify putative causal variants. (A) First, genome wide-association studies (GWAS) identify loci associated with a disease. Single-nucleotide polymorphisms (SNPs) within these loci that reach a corrected *P*-value threshold (dashed line) are associated with the disease. An association signal (for example on chromosome 13) can be refined through various methods to identify putative causal SNPs, as shown in B. (B) Statistical fine-mapping (left) can be used to refine disease associations. This technique typically uses Bayesian methods and data from SNP microarrays and whole-genome sequencing to evaluate the probability that each variant is causal given the haplotype structure across the locus. SNPs (dots) are coloured according to the r^2 value, which denotes linkage disequilibrium (LD) with the lead SNP at the locus (r^2 values closer to 1 indicate a higher LD). Corrected *P*-values are adjusted for multiple testing. Quantitative trait loci (QTL; middle) are loci in which genetic variation is associated with a cellular trait. In the plots, the corrected *P*-value indicates the association of SNPs with the trait of interest, with higher associations in disease than in non-disease contexts. Chromatin annotations (right) denote, for example, chromatin accessibility or histone marks that are indicative of enhancer activity. In disease context, higher corrected *P*-values indicate higher chromatin accessibility.

Statistical fine-mapping

The most common approach used to identify causal variants at disease-associated loci is statistical fine-mapping (Fig. 2). This technique seeks to refine a genetic association to a smaller subset of credible variants, typically by using Bayesian methods (Box 1) to evaluate the probability that each variant is causal given the haplotype structure across the locus (Wellcome Trust Case Control Consortium et al., 2012; Wallace et al., 2015). Because association statistics for each SNP at the locus must be known, dense genotyping data are required for this analysis. Statistical fine-mapping efforts have therefore benefitted from the development of Immunochip, an SNP microarray that provides dense genotyping of almost 200 human loci associated with at least one autoimmune or inflammatory disease (Cortes and Brown, 2011). Statistical fine-mapping has also benefitted from improvements in imputation reference panels, a set of reference haplotypes generated by whole-genome sequencing (WGS) that are used to predict genotypes of variants not included on the SNP microarray of a GWAS (Kowalski et al., 2019; McCarthy et al., 2016). Use of reference panels – or, as is becoming increasingly common, direct WGS of case-control samples – provides comprehensive genotype data and thereby also enables detection of rare variants (Almöf et al., 2019; Taliun et al., 2021). With these advances, statistical fine-mapping has helped resolve a small number of GWAS loci down to a single variant in diseases such as ankylosing spondylitis, rheumatoid arthritis and ulcerative colitis [International Multiple Sclerosis Genetics Consortium (IMSGC) et al., 2013; International Genetics of Ankylosing Spondylitis Consortium (IGAS) et al., 2013; Eyre et al., 2012; Trynka et al., 2011; Tsoi et al., 2012]. Importantly, however, strong LD (i.e. high rates of co-inheritance) between candidate SNPs has meant that this approach is not possible for most loci [International Multiple Sclerosis Genetics Consortium (IMSGC) et al., 2013; de Lange et al., 2017; Eyre et al., 2012; Huang et al., 2017]. As such, alternative approaches are needed to identify causal variants.

One way to increase the resolution of statistical fine-mapping is to draw on GWAS data from ethnically diverse populations. This approach, referred to as trans-ethnic meta-analysis, was first adopted to compare the genetic architecture of complex diseases across populations and to improve statistical power for locus discovery (Keller et al., 2014; Kim et al., 2015; Liu et al., 2015; Okada et al., 2014; Wang et al., 2010; Yin et al., 2015). Methods for trans-ethnic fine-mapping were later developed to leverage differences in the haplotype structure across different populations to help pinpoint putative causal variants (Kichaev and Pasaniuc, 2015; Mägi et al., 2017; Morris, 2011). These methods exploit the fact that LD structure is determined by ancestral meiotic recombination events and so varies among human populations. At loci at which a disease association is shared but the genetic architecture differs among ancestrally diverse populations, trans-ethnic fine-mapping can decipher the disease association signal by narrowing down the number of candidate variants to only those that are shared in both populations. Although few such studies have been performed, which reflects the fact that most GWAS have been conducted in European populations, these have successfully narrowed credible sets (Box 1) of putative causal variants for several complex traits and diseases, including several immune-mediated diseases (Chen et al., 2020; Kichaev and Pasaniuc, 2015; Liu et al., 2016; Ota et al., 2021; Stuart et al., 2022; van Rooij et al., 2017; Wojcik et al., 2019). For instance, in a GWAS meta-analysis of rheumatoid arthritis, trans-ethnic fine-mapping reduced the average size of credible sets by 20% (Kichaev and Pasaniuc, 2015). This highlights the benefits and importance of recruiting diverse populations for studies of genetic disease risk.

Colocalising candidate SNPs with complementary datasets

Another way to reduce the number of candidate variants at a disease-associated locus is to integrate the association statistics with other relevant genetic data. Aside from facilitating the discovery of disease susceptibility loci, the advent of affordable genome-wide genotyping has made it possible to study the consequences of genetic variation on a range of cellular traits. These include genetic effects on levels of mRNA, proteins, histone modifications, chromatin accessibility and DNA methylation – referred to as quantitative trait loci (QTL; Box 1). Because most GWAS signals are non-coding and likely to act by perturbing gene regulation, data from QTL studies, which frequently relate to, or reflect, altered enhancer activity, can help refine credible sets of putative causal variants and investigate their functional consequences.

One common approach is to use colocalisation analysis to investigate the overlap between GWAS associations and QTL data (Fig. 2). Colocalisation methods formally test the hypothesis that individual signals, such as a disease association and an effect on gene expression, co-occur at a given locus because of a shared causal variant(s) – a finding that would implicate both the variant and the biological effect in disease pathology (Giambartolomei et al., 2014; Hormozdiari et al., 2016; Wen et al., 2017). This approach has been successfully used to nominate candidate genes for dozens of GWAS loci in immune-mediated diseases, but has been less successful in identifying causal variants because both the GWAS and QTL datasets remain constrained by LD between candidate SNPs (de Lange et al., 2017; Guo et al., 2015; Okada et al., 2014; Peters et al., 2016; Wallace et al., 2012). A further challenge is that in immune cells most of the cellular features measured by QTL studies are highly dependent on both cell lineage and activation state (Chun et al., 2017; Fairfax et al., 2014; Kim-Hellmuth et al., 2017; Yazar et al., 2022). Accordingly, GWAS–QTL colocalisation studies need to be performed in the relevant cell type under suitable stimulation conditions to avoid missing evidence of a shared causal variant. This is an important limitation, as most QTL datasets have been generated from whole-blood samples or from immortalised cell lines that poorly reflect the physiology of primary immune cells (Andreu et al., 2017; Bartelt et al., 2009). Moreover, even when studies in specific immune cell types have been conducted, most have used unstimulated cells, which often differ in physiology from that of cells exposed to disease-relevant stimuli (Kerimov et al., 2021).

Combining GWAS data with chromatin features

A complementary set of methods have used known genomic features, often based on chromatin annotations, to try to interpret GWAS signals (Fig. 2) (Maurano et al., 2012; Pickrell, 2014). Because transcriptional regulation often relies on the binding of transcription factors, co-activators and other transcriptional machinery, chromatin that is inaccessible to these molecules (i.e. heterochromatinised) is unlikely to harbour SNPs that causally affect gene regulation. An important caveat, however, is that this assumes that the correct cell type is being studied and that the causal variant affects transcription factor binding, rather than other aspects of transcriptional control, such as insulator (Box 1) function. Nevertheless, high-throughput methods to annotate chromatin states have emerged as a useful tool to refine lists of candidate SNPs. Two general approaches have been adopted. One approach is to identify chromatin that is accessible to transcription factor binding using sequencing-based methods, such as DNase I hypersensitivity sites sequencing (DNase-seq; Box 1) or assay for transposase-accessible chromatin sequencing (ATAC-seq; Box 1) (Boyle et al., 2008; Buenrostro et al., 2015; Song et al., 2011; Thurman et al., 2012; Tong et al., 2016). A second approach is to use histone modifications that are associated with specific regulatory features, such as histone H3 lysine 27 acetylation (H3K27ac) or histone H3 lysine 4 monomethylation (H3K4me1), to identify enhancers (Creyghton et al., 2010; Heintzman et al., 2009). Both of these approaches have shown that the chromatin landscape is also cell-type specific, prompting the development of models to help prioritise pathogenic cell types based on selective enrichment of disease-associated variants in active regulatory regions (Ernst et al., 2011; Song et al., 2011; Thurman et al., 2012). This in turn led to the realisation that T-cell subsets frequently show the strongest enrichment for immune disease-associated variants (Farh et al., 2015; Maurano et al., 2012; Onengut-Gumuscu et al., 2015; Pickrell, 2014; Schmidt et al., 2015; Trynka et al., 2011). As with QTLs, however, the enrichment of immune disease-associated variants is most pronounced – and in some cases, only detectable – in certain primary cell types under specific stimulation conditions (Calderon et al., 2019; Glinos et al., 2020; Soskic et al., 2019). This finding again reflects the context-specific function of many enhancers and highlights the need for GWAS associations to be interpreted in disease-relevant cells under appropriate conditions, many of which remain incompletely catalogued at present.

Testing candidate SNPs for expression-modulating effects

An important caveat to the methods described above, which seek to refine lists of candidate SNPs using external datasets, is that the results produced are correlative and do not show whether a putative causal variant actually modulates gene expression. To address this issue, individual genetic reporter assays have been used to directly assess the transcriptional effects of putative causal SNPs (Elsby et al., 2010). However, these assays were traditionally labour

Disease Models & Mechanisms (2023) 16, dmm049790. doi:10.1242/dmm.049790

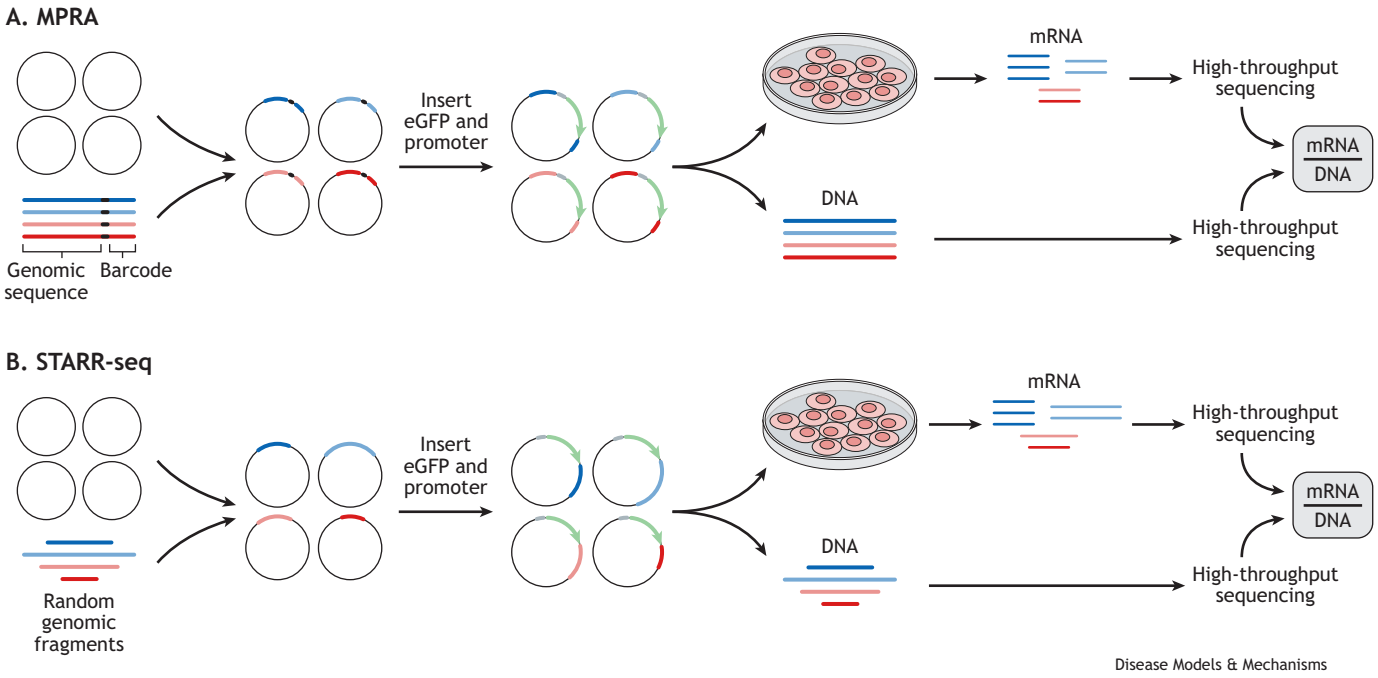


Fig. 3. Experimental methods to identify variants that modulate gene expression. (A,B) Massively parallel reporter assay (MPRA; A) and self-transcribing active regulatory region sequencing (STARR-seq; B) are used to directly assess the transcriptional effects of putative regulatory elements and/or putative causal variants. A library of barcoded putative enhancer sequences (MPRA) or genomic fragments (STARR-seq) are cloned into empty vectors (circles). A promoter (grey) and enhanced green fluorescent protein (eGFP; green arrow) are also inserted. The resulting plasmid is transfected into cells, RNA is extracted, and mRNA barcode counts (MPRA) or mRNA genomic fragment counts (STARR-seq) are obtained by high-throughput sequencing. mRNA counts are then normalised to their corresponding DNA counts within the transfected plasmid pool. This provides a measure of expression-modulating activity (mRNA/DNA).

intensive and low throughput, often assessing one allele at a time, and were therefore unsuitable for evaluating the possible functional consequences of thousands of candidate variants across many disease-associated loci. More recently, high-throughput assays of enhancer activity have been developed that can simultaneously test the regulatory effects of multiple non-coding sequences. Two main approaches have been used (Fig. 3). The first method, massively parallel reporter assay (MPRA), involves synthesising a library of short DNA oligonucleotides that can be systematically tested for enhancer/repressor activity (Melnikov et al., 2012; Patwardhan et al., 2012). This is achieved by cloning the DNA oligonucleotides into a reporter plasmid such that the putative enhancer sequence lies immediately upstream of a promoter and a reporter gene, similar to a luciferase assay. The key difference is that each sequence is paired with a unique oligonucleotide barcode positioned within the 3′ untranslated region (UTR) of the reporter gene, to allow the RNA molecules transcribed from individual plasmids to be matched to the putative enhancer sequence that modulated their expression using barcode sequencing (Fig. 3A) (Kheradpour et al., 2013; Tewhey et al., 2016). In this way, large numbers of putative enhancer sequences can be simultaneously assessed for transcriptional activity by normalising the RNA barcode counts that are obtained from cells following transfection to the corresponding DNA barcode counts within the transfected plasmid pool (Melnikov et al., 2014). The second method, self-transcribing regulatory region sequencing (STARR-seq), adopts a similar approach, but obtains putative enhancer sequences by randomly shearing genomic DNA to create smaller fragments, rather than by synthesising oligonucleotides (Arnold et al., 2013). These genomic fragments are ligated with adaptors and amplified with PCR before being cloned into reporter vectors and transfected into cells. The fragments are then sequenced directly rather than using oligonucleotide barcodes (Fig. 3B). STARR-seq allows longer DNA sequences to be tested than is possible with MPRA, owing to technical limitations with oligonucleotide synthesis, but is constrained by the genotype of the source genome and can only test for expression-modulating effects if both alleles are present.

Both MPRA and STARR-seq have the advantage of directly measuring the process by which non-coding variants principally act, and have been successfully used to distinguish expression-modulating variants from co-inherited non-functional variants (Bourges et al., 2020; Choi et al., 2020; Liu et al., 2017; Lu et al., 2021; Tewhey et al., 2016; Ulirsch et al., 2016; van Ouwekerk et al., 2020). However, although these assays represent powerful tools for characterising the functional effects of disease-associated variants, they too have limitations (Table 1). For example, both MPRA and STARR-seq assess the transcriptional activity of candidate enhancer sequences in an episomal vector using a standard promoter, and so do not fully recapitulate the interactions that take place between specific enhancers and their target promoters within the human genome. This might not matter, however, because the observed transcriptional effects do not appear to depend on the promoter used (Ferreira et al., 2016) – an observation consistent with the recent finding that most enhancers activate all promoters by similar amounts (Bergman et al., 2022). Another important consideration is that any expression-modulating activity is determined by the interaction between the tested DNA sequences and the transcription factors present within the transfected cells. For this reason, the cellular context in which disease-associated variants are tested is a key experimental parameter. To date, most MPRA and STARR-seq studies have been performed using immortalised cell lines, which are highly amenable to transfection but differ

substantially in physiology and transcriptional regulation from the primary cells that they seek to model (Andreu et al., 2017; Bartelt et al., 2009). To address this, Bourges et al. (2020) developed an adapted MPRA for use in primary human $CD4^+$ T cells, having demonstrated that the conventional MPRA vector was unsuitable for use in primary cells. This study confirmed that putative causal variants could be identified via their expression-modulating activity in primary immune cells, even at loci that were unresolvable by statistical fine-mapping, and that the results differed from those obtained in commonly used cell-line models, such as Jurkat T cells. The adapted MPRA system therefore provides a means to identify putative causal variants in disease-relevant cells by directly assaying the process by which they alter disease risk. This approach, however, does not identify the genes affected by any expression-modulating effects, for which complementary methods are required.

Statistical methods to prioritise putative causal variants and experimental methods to determine the transcriptional effects of individual SNPs have thus meant that it is now possible to identify putative causal variants at many disease-associated loci. The challenge then becomes characterising the mechanisms by which these variants drive disease processes.

Challenge 2: identifying disease mechanisms

Several complementary methods have been developed to characterise the functional consequences of non-coding variants. For immune-mediated diseases, frequently used approaches include descriptive methods to identify putative target genes affected by disease-associated loci, such as maps of three-dimensional chromatin structure and/or colocalisation analysis, as well as experimental methods to establish functional links between enhancers and target genes with CRISPR-based approaches. The advantages and limitations of these approaches are summarised in Table 2.

Combining GWAS data with three-dimensional chromatin interactions

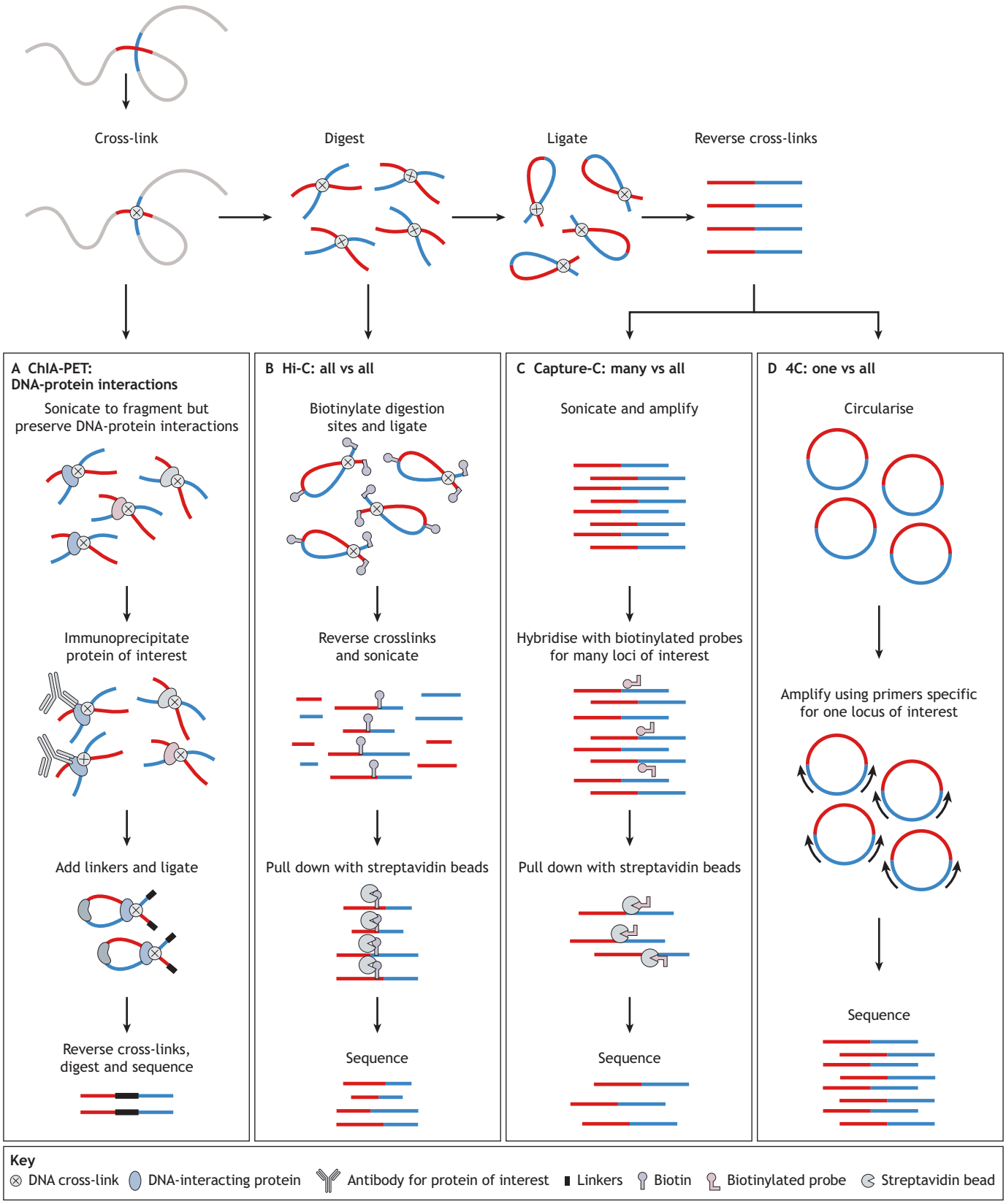
Although early genetic studies typically nominated the closest gene to a disease-associated locus as most likely to be causal, it is now clear that linear distance is not the best predictor of regulatory interactions. Indeed, it is well recognised that chromatin forms three-dimensional loops that can bring enhancers and promoters, which are separated by hundreds of kilobases, into close spatial proximity (Larke et al., 2021; Lieberman-Aiden et al., 2009; Sanyal et al., 2012; Schoenfelder et al., 2015). These interactions can be identified using a series of methods derived from chromatin conformation capture (3C), a technique using which chromatin is cross-linked, enzymatically digested and DNA ligated to produce chimeric fragments containing sequences from loci that were in close spatial proximity (Fig. 4) (Dekker et al., 2002). Common methods include (1) 4C (Box 1), a ‘one-versus-all’ approach to identify all regions interacting with a locus of interest; (2) Capture-C (Box 1), a ‘many-versus-all’ approach to determine the interaction partners for hundreds of loci of interest; (3) Hi-C (Box 1), an ‘all-versus-all’ approach to map all chromatin interactions throughout the genome; and (4) chromatin interaction analysis with paired-end tag (ChIA-PET; Box 1), an approach to assess chromatin interactions anchored by a particular protein (Fig. 4) (Fullwood et al., 2009; Hughes et al., 2014; Li et al., 2012; Lieberman-Aiden et al., 2009; Simonis et al., 2006; Zhao et al., 2006). By enabling unbiased and high-throughput detection of chromatin interactions, these methods have mapped hundreds of non-coding loci to their putative target genes (Hughes et al., 2014; Martin et al., 2015; Meddens et al., 2016; Schoenfelder et al., 2015). Moreover, integrating GWAS data with chromatin interaction maps has shown

Table 2. Methods to characterise disease mechanisms

Technique	Advantages	Limitations	References
Three-dimensional chromatin mapping	Links a non-coding locus to its target gene by showing that the locus physically interacts with a gene promoter. Can demonstrate allele-specific differences in chromatin contacts.	Technical challenges in generating maps often limit data to low resolution genome-wide profiles or higher-resolution profiles at selected loci. Evidence of causality is correlative.	Calderon et al., 2019; Javierre et al., 2016; Mumbach et al., 2017; Nasser et al., 2021
Colocalisation (of GWAS associations and eQTLs)	Assigns a non-coding locus to its target gene by showing that a GWAS association and eQTL likely share a causal variant at the locus. Shows directionality of effect and relevant cell context.	Requires eQTL data from disease-relevant cellular contexts that are often unavailable. eQTL can be linked to multiple genes or opposite effects on expression in different contexts. Evidence of causality is correlative.	de Lange et al., 2017; Fairfax et al., 2014; Giambartolomei et al., 2014; Hormozdiari et al., 2016; Okada et al., 2014
CRISPR knockout screens	Generates detailed maps of enhancer domains. Validates the regulatory effects of specific variants. Establishes causal role for enhancer in regulating its target gene.	Limited by small effect sizes of indels or low editing efficiencies of larger deletions. Cannot target sequences that lack PAM sites.	Canver et al., 2015; Gasperini et al., 2017; Dixit et al., 2016; Sanjana et al., 2016
CRISPRa/i screens	Facilitates discovery and functional characterization of enhancers. Offers a non-destructive approach and avoids potential off-target effects of double-strand breaks. Demonstrates that an enhancer is sufficient to drive expression of a target gene (CRISPRa) or is necessary for its expression (CRISPRi).	Epigenetic modifications can be unphysiological and can affect several kilobases surrounding the target. Difficulties in using the most relevant model (primary cells) because of requirement for sustained expression of modified Cas9. Cannot target sequences that lack PAM sites.	Fulco et al., 2016; Simeonov et al., 2017; Thakore et al., 2015; Schmidt et al., 2022

CRISPR, clustered regularly interspaced short palindromic repeats; CRISPRa, CRISPR activation; CRISPRi, CRISPR interference; eQTL, expression quantitative trait loci; GWAS, genome wide association studies; indel, insertion or deletion; PAM, protospacer adjacent motif.

Disease Models & Mechanisms (2023) 16, dmm049790. doi:10.1242/dmm.049790



Disease Models & Mechanisms

Fig. 4. See next page for legend.

Disease Models & Mechanisms

Fig. 4. Methods to map three-dimensional chromatin interactions. Methods to map three-dimensional chromatin interactions are based on chromatin conformation capture, where chromatin (grey line with red and blue interacting segments) is cross-linked. (A) In chromatin interaction analysis with paired-end tag (ChIA-PET), cross-linked chromatin is sonicated to obtain complexes of DNA fragments and proteins. The protein of interest (blue oval) is immunoprecipitated using the antibody specific for the protein of interest along with the cross-linked DNA. Linkers are added and the fragments are ligated (grey shape). The cross-links are then removed, the proteins are digested, and the DNA is prepared for sequencing to identify fragments of DNA interacting with the protein of interest. (B) In Hi-C, following chromatin digestion to generate cross-linked DNA fragments, biotin is added at the digested ends of DNA fragments, and the fragments are ligated. The cross-links are reversed, then the fragments are sonicated. Streptavidin beads are used to pull down biotinylated fragments. Biotin is removed, and the DNA is prepared for sequencing to identify all interacting pairs of loci. (C,D) In Capture-C and 4C, chromatin is digested to generate cross-linked DNA fragments, which are self-ligated and the cross-links are removed. (C) In Capture-C, ligated fragments are sonicated and amplified by PCR. Biotinylated oligonucleotide probes are added to hybridise to DNA sequences of interest, which are pulled down using streptavidin beads and sequenced to identify fragments interacting with these loci. (D) In 4C, ligated fragments are circularised then are amplified using PCR primers (black arrows) for a locus of interest. High-throughput sequencing is then done to identify all fragments interacting with this locus.

that disease-associated variants are enriched within interacting enhancers and promoters in disease-relevant cell types (Javierre et al., 2016; Mifsud et al., 2015; Mumbach et al., 2017; Nasser et al., 2021). Allele-specific mapping of disease variants at these loci has identified instances of allelic imbalance (Box 1), where chromatin interactions or accessibility profiles differ for risk and non-risk alleles (Calderon et al., 2019; Mumbach et al., 2017). This allelic imbalance is consistent with the hypothesis that disease variants disrupt transcription factor binding, ultimately leading to altered chromatin state and enhancer activity. Indeed, mechanistic insights from recent chromatin-mapping studies support a model of transcriptional regulation in which master transcription factors facilitate the formation of cell-type-specific regulatory hubs that can encompass many enhancers and promoters and that together drive complex gene expression programs (Di Giammartino et al., 2019; Hsieh et al., 2020; Hua et al., 2021; Oudelaar et al., 2018). Thus, these studies not only implicate genes affected by disease-associated loci but also suggest putative mechanisms of disease, indicating that risk variants within enhancers can perturb the formation of regulatory hubs and reduce or augment transcription of target genes, resulting in dysregulation of cellular processes.

Identifying causal genes via eQTL data

Another common method used to identify causal genes is to assess the colocalisation of GWAS signals with expression quantitative trait loci (eQTL). eQTL are SNP haplotypes where allelic variants are associated with levels of gene expression (Fairfax et al., 2014). Although such colocalisation analyses cannot always pinpoint causal variants (for reasons described in the earlier section), this approach can provide statistical evidence that a GWAS association and an eQTL are likely to share a mechanistic basis and thus a target gene. An important, limiting, feature of eQTL effects is that they are largely cell-type and cell-state dependent, and so could be missed if the wrong context or cell type is used (Fairfax et al., 2014; Guo et al., 2015; Kim-Hellmuth et al., 2017). Indeed, although GWAS associations are enriched for eQTLs, only a small proportion have been matched to their target gene through colocalisation analysis, likely because eQTL data do not exist for all relevant cell types or conditions. Recently, however, a growing number of datasets have profiled primary human immune cell subpopulations under relevant stimulation conditions, leading to improved detection of colocalisation (Chen et al., 2016; Ota et al., 2021; Schmiedel et al., 2018). This has benefitted from use of single-cell RNA sequencing (scRNA-seq) to identify transcriptional changes that might only occur in very small subsets of cells (Soskic et al., 2022; Yazar et al., 2022). Interestingly, eQTL studies have occasionally uncovered contrasting effects in different cell types, either where different genes are eQTLs for the same GWAS loci or where opposite directional effects are observed (Ota et al., 2021; Peters et al., 2016; Schmiedel et al., 2018). Resolving causal genes in this situation will inevitability require complementary methods.

Mapping enhancer activity across the genome using CRISPR

In recent years, CRISPR has emerged as a powerful tool for studying genome function, including functionally linking enhancers to target genes. This was initially applied *in vitro* via a conventional knockout approach, in which single-guide RNAs (sgRNAs) were designed to disrupt either non-coding regions around a gene of interest or putative binding sites of a given transcription factor (Canver et al., 2015; Korkmaz et al., 2016; Rajagopal et al., 2016; Sanjana et al., 2016). These studies generated detailed maps of functional enhancers in primary cells and cell lines, facilitating investigation of how specific variants might affect enhancers and establishing mechanistic links between enhancers and their target genes. However, this approach was often limited by the minor effects that small insertions or deletions (indels) have on enhancer function and by the inability to target sequences that lacked protospacer adjacent motif (PAM) (Box 1) sites. To overcome this limitation, subsequent studies used pairs of sgRNAs to create larger deletions, often several hundred bases apart, although this method is, in turn, limited by the lower efficiency of deleting large DNA segments (Diao et al., 2017; Gasperini et al., 2017). Other studies have used a pooled approach to make knockout screens more scalable and better able to detect complex phenotypes. This approach delivers multiple sgRNAs per cell then uses scRNA-seq to determine the effect of each perturbation at a transcript level (Datlinger et al., 2017; Dixit et al., 2016; Jaitin et al., 2016).

An alternative approach was subsequently developed using a nuclease-deactivated Cas9 (dCas9) that was fused to an effector domain that could alter chromatin state at target sites, either activating transcription [CRISPR activation (CRISPRa)] or repressing it [CRISPR interference (CRISPRi)] (Gilbert et al., 2014). This approach has the advantage of being non-destructive and, therefore, allows for assessment of regulatory function while minimising off-target effects due to double-strand breaks. In CRISPRa, dCas9 is fused to a strong transcriptional activator, such as VP64 or VP64-p65-Rta (VPR), making it possible to discover putative enhancers that can drive expression of target genes, including those not typically active under the screening conditions (Dai et al., 2021; Li et al., 2020; Simeonov et al., 2017). In contrast, CRISPRi uses dCas9 fused to a lysine-specific demethylase 1 (LSD1) or Krüppel associated box (KRAB) domain to recruit repressive chromatin modifiers and promote heterochromatin formation (Kearns et al., 2015; Thakore et al., 2015). This approach reduces the activity of enhancers that are functional under the screening conditions and can identify regions necessary for gene expression (Fulco et al., 2016; Klann et al., 2017).

CRISPR-based screens offer a powerful and unbiased approach to identify enhancers in an endogenous context, but also have several limitations (Table 2). First, the epigenetic modifications

induced by CRISPRi and CRISPRa are often unphysiological and can alter several kilobases of surrounding chromatin. This makes high-resolution identification of functional enhancer sequences challenging (Thakore et al., 2015). Second, although CRISPR-based screens best recapitulate physiological conditions when conducted in primary cells, most studies rely on immortalised cell lines that can be engineered to stably express the Cas9 or dCas9 fusion proteins. Indeed, the difficulty of delivering Cas9 has previously hindered efforts to adapt CRISPR screens for primary cells, as the large transgene results in low transduction efficiency when delivered using lentivirus and high toxicity when delivered as a plasmid (Lesueur et al., 2016; Shifrut et al., 2018). CRISPR knockout screens, which only require transient expression of Cas9, have circumvented this issue by delivering Cas9 protein following lentiviral transduction of the sgRNA library (Shifrut et al., 2018; Ting et al., 2018). However, CRISPRi and CRISPRa screens, which induce reversible epigenetic modifications, require more sustained expression of the effector molecules. A recent study has reported a possible solution to this: the authors used an optimised transduction method to efficiently deliver dCas9 fusion proteins via lentiviral vectors into primary human T cells, thereby enabling CRISPRi and CRISPRa screens in these cells (Schmidt et al., 2022). This approach facilitated the identification of genes that could modulate cytokine production in T cells by either activating or interfering with promoter activity – highlighting the potential for similar approaches in other primary cell types to provide insights into key regulatory circuits.

The development of methods to link enhancers to their target genes represents a key advance in efforts to understand how non-coding loci mediate disease risk. Indeed, the ability to identify causal variants within regulatory elements and delineate their downstream functional consequences will build a better understanding of disease mechanisms.

Translating non-coding SNPs to immune disease mechanisms

Advances in genetic and genomic techniques now provide an unprecedented opportunity to functionally dissect and systematically characterise non-coding genetic variation in immune-mediated disease. Using these approaches, several recent studies have uncovered mechanisms by which causal variants within enhancers perturb regulatory activity and drive immune disease processes. These studies offer important insights into key pathways in immune-mediated disease – some of which may be potential therapeutic targets – and illustrate general principles by which common genetic variants in enhancers can contribute to pathology.

IL2RA: temporal expression alters immune cell differentiation

The interleukin 2 receptor alpha (*IL2RA*) gene has been identified in multiple GWAS studies as a key susceptibility locus for over a dozen immune-mediated diseases (Carr et al., 2009; Ellinghaus et al., 2016; Lowe et al., 2007; Stahl et al., 2010). This locus encodes IL-2Ra, a subunit of the heterotrimeric high-affinity IL-2 receptor that is constitutively expressed by regulatory T (Treg) cells (Box 1) and upon activation by conventional T cells (Sakaguchi et al., 1995; Waysbort et al., 2013). IL-2 signalling plays a critical role in the maintenance of Treg cells, as well as in the expansion and differentiation of naïve T cells (Fontenot et al., 2005; Liao et al., 2011; Liu et al., 2015; Pipkin et al., 2010). Although coding mutations that ablate *IL2RA* expression result in severe immunodeficiency and autoimmunity, the functional consequences of non-coding variation at the *IL2RA* locus have proven more difficult to characterise (Caudy et al., 2007; Goudy et al., 2013; Sharfe et al., 1997). Statistical fine-mapping indicates that the human *IL2RA* locus contains multiple independent signals, including one association refined to a single putative causal variant (rs61839660 C>T) that paradoxically confers risk for Crohn's disease but protection against type 1 diabetes mellitus (Burren et al., 2017; Huang et al., 2017; Maier et al., 2009; Onengut-Gumuscu et al., 2015). This complex pattern of association reflects the complicated regulatory landscape of the locus, which contains a cluster of enhancers that cooperatively tune *IL2RA* expression in response to T-cell receptor (TCR) stimulation and signalling from multiple cytokines, including IL-2 (Busse et al., 2010; John et al., 1996; Li et al., 2017; Liao et al., 2011).

Simeonov et al. (2017) devised a CRISPRa screen to better understand the regulatory landscape of this locus by identifying enhancers that could induce *IL2RA* expression in resting T cells, which was the first time that this methodology had been used to characterise enhancers rather than promoters. By tiling (Box 1) a 178-kilobase region around *IL2RA* in Jurkat-dCas9-VP64 cells (Box 1), the screen uncovered six putative enhancers, including an intronic region containing rs61839660 – the SNP associated with Crohn's disease and type 1 diabetes mellitus. Further investigation in murine $CD4^+$ T cells demonstrated that this conserved region is an enhancer that specifically regulates TCR stimulation-induced *IL2RA* expression, and thereby influences the balance of differentiation to either a pro-inflammatory T-helper 17 (Th17) phenotype or a tolerogenic Treg phenotype under IL-2-restricted polarisation conditions. This study additionally showed that, in mice, introduction of the risk allele for rs61839660 leads to a temporal delay in IL-2Ra expression by $CD4^+$ T cells. Together, these findings suggest that the causal variant alters Th17/Treg balance by disrupting the timing of *IL2RA* induction – a plausible mechanism by which the variant could contribute to disease. However, the molecular mechanism by which the causal variant delays expression and the reason this has opposing effects on susceptibility to two immune-mediated diseases were not explored in this study. Subsequent work described in a preprint (Simeonov et al., 2020), however, indicated that this intronic enhancer specifically governs IL-2Ra expression in stimulated conventional T cells, whereas a different upstream enhancer controls its expression in Tregs. These observations help parse the enhancer code that governs *IL2RA* expression and provide insights into the nuanced role of IL-2 signalling in maintaining immune homeostasis. The findings can help inform clinical studies by indicating that low-dose IL-2 might induce Treg expansion and suppress aberrant inflammation (Dong et al., 2021).

TNFAIP3: regulation of T-cell activation depends on super-enhancer formation

Among the most challenging disease-associated variants to characterise are those located within loci that contain no protein-coding genes ('gene deserts'), because there are no immediate indications as to which genes/pathways underpin the associations. Bourges et al. (2020) designed an MPRA to assess regulatory activity of candidate variants from 14 gene deserts associated with ten immune-mediated diseases and leveraged these MPRA findings to investigate disease biology. This study – the first to develop an MPRA for primary immune cells – resolved putative causal variants within these gene deserts via their transcriptional effects in primary human $CD4^+$ T cells. The authors investigated the 6q23 gene desert further because of its association with at least six immune-mediated

diseases, highlighting the biological importance of this locus, observing that the variant with the greatest expression-modulating effect (rs6927172 C>G) disrupted binding of nuclear factor kappa B (NF-κB; Box 1) at the 6q23 locus in stimulated T cells. The disruption of NF-κB binding reduced enhancer strength across the region. This led to the discovery that the 6q23 region harbours a super-enhancer, which is formed upon T-cell activation and drives expression of tumour necrosis factor alpha-induced protein 3 (*TNFAIP3*). *TNFAIP3* encodes A20, a key inhibitor of NF-κB that suppresses inflammatory responses. This suggested that, in response to cellular stimulation, a regulatory mechanism is induced that limits T-cell activation and prevents uncontrolled inflammation. This study validated the molecular mechanism by which the causal variant at the 6q23 locus disrupts super-enhancer activity using CRISPR in primary $CD4^+$ T cells. The authors also demonstrated that deletion of this critical NF-κB binding site leads to unrestrained T-cell responses and thereby provided an explanation for the pleiotropic immune disease risk. These results mirror those of a simultaneous study, which showed that the same rs6927172 risk allele is associated with altered chromatin remodelling and with reduced downstream expression of *TNFAIP3* in stimulated primary human $CD4^+$ T cells (Calderon et al., 2019).

The finding that the causal variant at the 6q23 locus broadly compromises the activity of surrounding enhancers (aside from its effect on its cognate enhancer) provides further evidence of the sensitivity of super-enhancers to perturbation of their constituent parts. Although the precise nature and unique functions of super-enhancers remain debated, these elements are generally defined as long stretches of highly active enhancers that are marked by extensive histone modifications and by high densities of Mediator (Box 1), master transcription factors, and other co-factors that regulate the expression of genes central to cell identity (Farooq et al., 2021; Moorthy et al., 2017; Parker et al., 2013; Whyte et al., 2013). Studies that have deleted or mutated component enhancers within super-enhancer loci have found that the effects of such disruption range from mild decreases to near-complete ablation of target gene expression (Hay et al., 2016; Moorthy et al., 2017; Shin et al., 2016). Yet, how disease-associated variants that are enriched in super-enhancers of relevant cell types might affect super-enhancer activity remained largely uncharacterised (Hnisz et al., 2013; Oldridge et al., 2015; Vahedi et al., 2015). Bourges et al. (2020) therefore elucidated a mechanism by which a single variant can disrupt super-enhancer function and contribute directly to multiple diseases.

LRRC32: an enhancer mediates the immunosuppressive activity of Treg cells

Although the 11q13.5 locus is associated with allergic disease, inflammatory bowel disease and type 1 diabetes mellitus, its target gene and the mechanism by which it contributes to disease had not been definitively established (Anderson et al., 2011; Barrett et al., 2008; Esparza-Gordillo et al., 2009; Marenholz et al., 2015; Onengut-Gumuscu et al., 2015). Recently, Nasrallah et al. (2020) identified chromatin features at this locus that were suggestive of enhancer activity in Treg cells. Investigation of this conserved region in mice found that deleting the locus increased susceptibility to colitis and decreased expression of leucine rich repeat containing 32 (*Lrrc32*), the gene encoding glycoprotein A repetitions predominant (GARP). Further study established that the disease-associated locus is a Treg-specific enhancer, which, in response to IL-2 and TCR signalling, is bound by signal transducer and activator of transcription 5 (STAT5) and NF-κB and drives expression of GARP. A small CRISPRa screen in primary human CD4 T cells further indicated that rs11236797, or a nearby variant, is likely to be causal. Colocalisation analysis at this locus also showed that the disease risk alleles colocalise with variants associated with reduced levels of the histone modification H3K27ac in Treg cells. However, the precise mechanism responsible for impaired enhancer activity remains undefined. With these results, the authors nominated GARP as a potential therapeutic target to enhance the immunosuppressive effects of Treg cells. Indeed, a growing number of studies support a key role for GARP in immune regulation, with knockdown or inhibition of GARP reducing Treg-mediated suppression of T-effector cell activity and rare coding mutations in *LRRC32* being associated with primary immunodeficiency and atopic dermatitis (Cuende et al., 2015; Lehmkuhl et al., 2021; Manz et al., 2016; Wang et al., 2010). By characterising the contribution of the 11q13.5 locus to *LRRC32* regulation in Tregs, the study by Nasrallah et al. (2020) helped delineate the pathways that establish Treg-specific expression of *LRRC32*, which had previously been proposed to be caused by either TCR stimulation or signalling from multiple cytokines (Kuhn et al., 2017; Manz et al., 2016; Zhou et al., 2013). This study therefore provides important insights into the regulation of GARP and its central role in Treg cell function.

The examples above illustrate how resolving the biological mechanisms by which non-coding genetic variants contribute to disease can provide important insights into transcriptional regulation, gene function and fundamental immunology, as well as disease mechanisms. Ongoing work to resolve disease associations at non-coding loci promises to yield further insights into immune biology and should uncover novel therapeutic targets.

Future perspectives

Translating GWAS associations into a better understanding of disease biology remains arguably the most important challenge in modern genetics. Tremendous progress has been made by using recently developed tools and approaches to characterise the functions of the non-coding genome. However, much more work is required to fulfil the potential of GWAS and elucidate the molecular and cellular mechanisms that cause immune-mediated diseases. Several key challenges remain.

First, there is a clear need to develop more accurate cellular models to study disease processes. One area of focus is to adapt high-throughput screening assays and related methods for use in primary human immune cells because these cells best recapitulate effector pathways in disease. Although challenges in obtaining sufficient numbers of primary cells and difficulties in transfecting or transducing these cells have long prohibited their use in high-throughput screens and assays, recent studies have adapted these methods for use in primary human T cells (Bourges et al., 2020; Hiatt et al., 2022; Schmidt et al., 2022). The next step will be to extend these methods into other primary immune cell types that play key roles in disease.

A related area of focus is to identify stimulation conditions that are most relevant to immune-mediated diseases. Appropriate stimulation conditions are essential for studying disease processes because the stimuli used, their dose and duration all determine gene expression programs and cell activation state (Xue et al., 2014). Currently, *in vitro* studies with primary cells often rely on simplistic differentiation and activation conditions – in large part because disease-relevant stimuli have not been fully characterised. Resolving this challenge will therefore ensure that future studies

can investigate disease mechanisms using conditions that recapitulate those encountered by cells in diseased tissue.

Second, a better understanding of the molecular mechanisms by which non-coding variants contribute to disease is urgently required. Non-coding variants within enhancers have been shown to predispose to disease by altering transcription factor binding and reducing or augmenting expression of target genes (Bourges et al., 2020; Musunuru et al., 2010). However, we currently have a limited ability to predict the effects of genetic variation on transcription factor binding because many binding motifs are unknown and even well-characterised transcription factors often bind at sequences other than their canonical motifs (Jolma et al., 2013; Weirauch et al., 2014). Moreover, the mechanisms by which altered transcription factor binding leads to differential gene expression are incompletely understood. Indeed, transcription factor binding facilitates processes ranging from formation of enhancer-promoter contacts to chromatin remodelling and DNA methylation, and perturbation of any of these processes can modulate levels of transcription (Di Giammartino et al., 2019; Martin-Trujillo et al., 2020). A further consideration is that some disease-associated non-coding variants lie outside of enhancer regions and must therefore contribute to immunopathology through other mechanisms. The observation that these variants are enriched at sites of post-translational modifications has given rise to several hypotheses of their function. These include the following: (1) that non-coding variants influence RNA splicing (Box 1); (2) that non-coding variants affect RNA editing (Box 1); and (3) that non-coding variants alter the 3′ UTR, which can influence transcript stability and translation (Griesemer et al., 2021; Li et al., 2022, 2016). However, the exact contribution of each of these processes to immune-mediate disease susceptibility remains to be characterised.

Ultimately, addressing these challenges will help translate genetic associations into a deeper knowledge of disease biology. This should reveal novel druggable targets and enable subsequent drug discovery efforts to be directly focused on validated disease mechanisms, thereby improving the treatment of patients with immune-mediated diseases.

Competing interests

The authors declare no competing or financial interests.

Funding

This work was supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK (CC2219), the UK Medical Research Council (CC2219) and the Wellcome Trust (CC2219). J.C.L. is a Lister Institute of Preventive Medicine Prize Fellow. For the purpose of Open Access, the authors have applied a CC BY public copyright licence to any author accepted manuscript version arising from this submission.

References

Altshuler, D., Daly, M. J. and Lander, E. S. (2008). Genetic mapping in human disease. *Science* **322**, 881-888. doi:10.1126/science.1156409

Almlöf, J. C., Nystedt, S., Leonard, D., Eloranta, M. L., Grosso, G., Sjöwall, C., Bengtsson, A. A., Jönsen, A., Gunnarsson, I., Svenungsson, E. et al. (2019). Whole-genome sequencing identifies complex contributions to genetic risk by variants in genes causing monogenic systemic lupus erythematosus. *Hum. Genet.* **138**, 141-150. doi:10.1007/s00439-018-01966-7

Andreu, N., Phelan, J., de Sessions, P. F., Cliff, J. M., Clark, T. G. and Hibberd, M. L. (2017). Primary macrophages and J774 cells respond differently to infection with Mycobacterium tuberculosis. *Sci. Rep.* **7**, 42225. doi:10.1038/srep42225

Anderson, C. A., Boucher, G., Lees, C. W., Franke, A., D'Amato, M., Taylor, K. D., Lee, J. C., Goyette, P., Imielinski, M., Latiano, A. et al. (2011). Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat. Genet.* **43**, 246-252. doi:10.1038/ng.764

Arnold, C. D., Gerlach, D., Stelzer, C., Boryń, Ł. M., Rath, M. and Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**, 1074-1077. doi:10.1126/science.1232542

Barrett, J. C., Hansoul, S., Nicolae, D. L., Cho, J. H., Duerr, R. H., Rioux, J. D., Brant, S. R., Silverberg, M. S., Taylor, K. D., Barmada, M. M. et al. (2008). Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat. Genet.* **40**, 955-962. doi:10.1038/ng.175

Bartelt, R. R., Cruz-Orcutt, N., Collins, M. and Houtman, J. C. (2009). Comparison of T cell receptor-induced proximal signaling and downstream functions in immortalized and primary T cells. *PLoS ONE* **4**, e5430. doi:10.1371/journal.pone.0005430

Bergman, D. T., Jones, T. R., Liu, V., Ray, J., Jagoda, E., Siraj, L., Kang, H. Y., Nasser, J., Kane, M., Rios, A. et al. (2022). Compatibility rules of human enhancer and promoter sequences. *Nature* **607**, 176-184. doi:10.1038/s41586-022-04877-w

Bourges, C., Groff, A. F., Burren, O. S., Gerhardinger, C., Mattioli, K., Hutchinson, A., Hu, T., Anand, T., Epping, M. W., Wallace, C. et al. (2020). Resolving mechanisms of immune-mediated disease in primary CD4 T cells. *EMBO Mol. Med.* **12**, e12112. doi:10.15252/emmm.202012112

Boyle, A. P., Davis, S., Shulha, H. P., Meltzer, P., Margulies, E. H., Weng, Z., Furey, T. S. and Crawford, G. E. (2008). High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**, 311-322. doi:10.1016/j.cell.2007.12.014

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**, 21.29.1-21.29.9. doi:10.1002/0471142727.mb2129s109

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

Busse, D., de la Rosa, M., Hobiger, K., Thurley, K., Flossdorf, M., Scheffold, A. and Höfer, T. (2010). Competing feedback loops shape IL-2 signaling between helper and regulatory T lymphocytes in cellular microenvironments. *Proc. Natl. Acad. Sci. USA* **107**, 3058-3063. doi:10.1073/pnas.0812851107

Calderon, D., Nguyen, M. L. T., Mezger, A., Kathiria, A., Müller, F., Nguyen, V., Lescano, N., Wu, B., Trombetta, J., Ribado, J. V. et al. (2019). Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nat. Genet.* **51**, 1494-1505. doi:10.1038/s41588-019-0505-9

Canver, M. C., Smith, E. C., Sher, F., Pinello, L., Sanjana, N. E., Shalem, O., Chen, D. D., Schupp, P. G., Vinjamur, D. S., Garcia, S. P. et al. (2015). BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* **527**, 192-197. doi:10.1038/nature15521

Carr, E. J., Clatworthy, M. R., Lowe, C. E., Todd, J. A., Wong, A., Vyse, T. J., Kamesh, L., Watts, R. A., Lyons, P. A. and Smith, K. G. (2009). Contrasting genetic association of IL2RA with SLE and ANCA-associated vasculitis. *BMC Med. Genet.* **10**, 22. doi:10.1186/1471-2350-10-22

Caudy, A. A., Reddy, S. T., Chatila, T., Atkinson, J. P. and Verbsky, J. W. (2007). CD25 deficiency causes an immune dysregulation, polyendocrinopathy, enteropathy, X-linked-like syndrome, and defective IL-10 expression from CD4 lymphocytes. *J. Allergy Clin. Immunol.* **119**, 482-487. doi:10.1016/j.jaci.2006.10.007

Chen, L., Ge, B., Casale, F. P., Vasquez, L., Kwan, T., Garrido-Martín, D., Watt, S., Yan, Y., Kundu, K., Ecker, S. et al. (2016). Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell* **167**, 1398-1414.e24. doi:10.1016/j.cell.2016.10.026

Chen, M. H., Raffield, L. M., Mousas, A., Sakaue, S., Huffman, J. E., Moscati, A., Trivedi, B., Jiang, T., Akbari, P., Vuckovic, D. et al. (2020). Trans-ethnic and ancestry-specific blood-cell genetics in 746,667 individuals from 5 global populations. *Cell* **182**, 1198-1213.e14. doi:10.1016/j.cell.2020.06.045

Choi, J., Zhang, T., Vu, A., Ablain, J., Makowski, M. M., Colli, L. M., Xu, M., Hennessey, R. C., Yin, J., Rothschild, H. et al. (2020). Massively parallel reporter assays of melanoma risk variants identify MX2 as a gene promoting melanoma. *Nat. Commun.* **11**, 2718. doi:10.1038/s41467-020-16590-1

Chun, S., Casparino, A., Patsopoulos, N. A., Croteau-Chonka, D. C., Raby, B. A., De Jager, P. L., Sunyaev, S. R. and Cotsapas, C. (2017). Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. *Nat. Genet.* **49**, 600-605. doi:10.1038/ng.3795

Cortes, A. and Brown, M. A. (2011). Promise and pitfalls of the Immunochip. *Arthritis Res. Ther.* **13**, 101. doi:10.1186/ar3204

Cotsapas, C., Voight, B. F., Rossin, E., Lage, K., Neale, B. M., Wallace, C., Abecasis, G. R., Barrett, J. C., Behrens, T., Cho, J. et al. (2011). Daly MJ; FOCiS Network of Consortia. Pervasive sharing of genetic effects in autoimmune disease. *PLoS Genet.* **7**, e1002254. doi:10.1371/journal.pgen.1002254

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

Cuende, J., Liénart, S., Dedobbeleer, O., van der Woning, B., De Boeck, G., Stockis, J., Huygens, C., Colau, D., Somja, J., Delvenne, P. et al. (2015). Monoclonal antibodies against GARP/TGF-β1 complexes inhibit the

immunosuppressive activity of human regulatory T cells in vivo. *Sci Transl Med* **7**, 284ra56. doi:10.1126/scitranslmed.aaa1983

Dai, Z., Li, R., Hou, Y., Li, Q., Zhao, K., Li, T., Li, M. J. and Wu, X. (2021). Inducible CRISPRa screen identifies putative enhancers. *J. Genet. Genomics* **48**, 917-927. doi:10.1016/j.jgg.2021.06.012

Daly, M. J., Rioux, J. D., Schaffner, S. F., Hudson, T. J. and Lander, E. S. (2001). High-resolution haplotype structure in the human genome. *Nat. Genet.* **29**, 229-232. doi:10.1038/ng1001-229

Datlinger, P., Rendeiro, A. F., Schmidl, C., Krausgruber, T., Traxler, P., Klughammer, J., Schuster, L. C., Kuchler, A., Alpar, D. and Bock, C. (2017). Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **14**, 297-301. doi:10.1038/nmeth.4177

de Lange, K. M., Moutsianas, L., Lee, J. C., Lamb, C. A., Luo, Y., Kennedy, N. A., Jostins, L., Rice, D. L., Gutierrez-Achury, J., Ji, S. G. et al. (2017). Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat. Genet.* **49**, 256-261. doi:10.1038/ng.3760

Degner, J. F., Pai, A. A., Pique-Regi, R., Veyrieras, J. B., Gaffney, D. J., Pickrell, J. K., De Leon, S., Michelini, K., Lewellen, N., Crawford, G. E. et al. (2012). DNaseI sensitivity QTLs are a major determinant of human expression variation. *Nature* **482**, 390-394. doi:10.1038/nature10808

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

Di Giammartino, D. C., Kloetgen, A., Polyzos, A., Liu, Y., Kim, D., Murphy, D., Abuhashem, A., Cavaliere, P., Aronson, B., Shah, V. et al. (2019). KLF4 is involved in the organization and regulation of pluripotency-associated three-dimensional enhancer networks. *Nat. Cell Biol.* **21**, 1179-1190. doi:10.1038/s41556-019-0390-6

Diao, Y., Fang, R., Li, B., Meng, Z., Yu, J., Qiu, Y., Lin, K. C., Huang, H., Liu, T., Marina, R. J. et al. (2017). A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. *Nat. Methods* **14**, 629-635. doi:10.1038/nmeth.4264

Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C. P., Jerby-Arnon, L., Marjanovic, N. D., Dionne, D., Burks, T., Raychowdhury, R. et al. (2016). Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167**, 1853-1866.e17. doi:10.1016/j.cell.2016.11.038

Dong, S., Hiam-Galvez, K. J., Mowery, C. T., Herold, K. C., Gitelman, S. E., Esensten, J. H., Liu, W., Lares, A. P., Leinbach, A. S., Lee, M. et al. (2021). The effect of low-dose IL-2 and Treg adoptive cell therapy in patients with type 1 diabetes. *JCI Insight* **6**, e147474. doi:10.1172/jci.insight.147474

Ellinghaus, D., Jostins, L., Spain, S. L., Cortes, A., Bethune, J., Han, B., Park, Y. R., Raychaudhuri, S., Pouget, J. G., Hübenthal, M. et al. (2016). Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat. Genet.* **48**, 510-518. doi:10.1038/ng.3528

Elsby, L. M., Orozco, G., Denton, J., Worthington, J., Ray, D. W. and Donn, R. P. (2010). Functional evaluation of TNFAIP3 (A20) in rheumatoid arthritis. *Clin. Exp. Rheumatol.* **28**, 708-714.

Ernst, J., Kheradpour, P., Mikkelsen, T. S., Shoresh, N., Ward, L. D., Epstein, C. B., Zhang, X., Wang, L., Issner, R., Coyne, M. et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43-49. doi:10.1038/nature09906

Esparza-Gordillo, J., Weidinger, S., Fölster-Holst, R., Bauerfeind, A., Ruschendorf, F., Patone, G., Rohde, K., Marenholz, I., Schulz, F., Kerscher, T. et al. (2009). A common variant on chromosome 11q13 is associated with atopic dermatitis. *Nat. Genet.* **41**, 596-601. doi:10.1038/ng.347

Eyre, S., Bowes, J., Diogo, D., Lee, A., Barton, A., Martin, P., Zhernakova, A., Stahl, E., Viatte, S., McAllister, K. et al. (2012). High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat. Genet.* **44**, 1336-1340. doi:10.1038/ng.2462

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

Farh, K. K., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W. J., Beik, S., Shoresh, N., Whitton, H., Ryan, R. J., Shishkin, A. A. et al. (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337-343. doi:10.1038/nature13835

Farooq, U., Saravanan, B., Islam, Z., Walavalkar, K., Singh, A. K., Jayani, R. S., Meel, S., Swaminathan, S. and Notani, D. (2021). An interdependent network of functional enhancers regulates transcription and EZH2 loading at the INK4a/ARF locus. *Cell Rep* **34**, 108898. doi:10.1016/j.celrep.2021.108898

Ferreira, L. M., Meissner, T. B., Mikkelsen, T. S., Mallard, W., O'Donnell, C. W., Tilburgs, T., Gomes, H. A., Camahort, R., Sherwood, R. I., Gifford, D. K. et al. (2016). A distant trophoblast-specific enhancer controls HLA-G expression at the maternal-fetal interface. *Proc. Natl. Acad. Sci. USA* **113**, 5364-5369. doi:10.1073/pnas.1602886113

Freeman, K., Ryan, R., Parsons, N., Taylor-Phillips, S., Willis, B. H. and Clarke, A. (2021). The incidence and prevalence of inflammatory bowel disease in UK primary care: a retrospective cohort study of the IQVIA Medical Research Database. *BMC Gastroenterol.* **21**, 139. doi:10.1186/s12876-021-01716-6

Fontenot, J. D., Rasmussen, J. P., Gavin, M. A. and Rudensky, A. Y. (2005). A function for interleukin 2 in Foxp3-expressing regulatory T cells. *Nat. Immunol.* **6**, 1142-1151. doi:10.1038/ni1263

Fulco, C. P., Munschauer, M., Anyoha, R., Munson, G., Grossman, S. R., Perez, E. M., Kane, M., Cleary, B., Lander, E. S. and Engreitz, J. M. (2016). Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* **354**, 769-773. doi:10.1126/science.aag2445

Fullwood, M. J., Liu, M. H., Pan, Y. F., Liu, J., Xu, H., Mohamed, Y. B., Orlov, Y. L., Velkov, S., Ho, A., Mei, P. H. et al. (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* **462**, 58-64. doi:10.1038/nature08497

Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M. et al. (2002). The structure of haplotype blocks in the human genome. *Science* **296**, 2225-2229. doi:10.1126/science.1069424

Gasperini, M., Findlay, G. M., McKenna, A., Milbank, J. H., Lee, C., Zhang, M. D., Cusanovich, D. A. and Shendure, J. (2017). CRISPR/Cas9-mediated scanning for regulatory elements required for HPRT1 expression via thousands of large, programmed genomic deletions. *Am. J. Hum. Genet.* **101**, 192-205. doi:10.1016/j.ajhg.2017.06.010

Gilbert, L. A., Horlbeck, M. A., Adamson, B., Villalta, J. E., Chen, Y., Whitehead, E. H., Guimaraes, C., Panning, B., Ploegh, H. L., Bassik, M. C. et al. (2014). Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159**, 647-661. doi:10.1016/j.cell.2014.09.029

Glinos, D. A., Soskic, B., Williams, C., Kennedy, A., Jostins, L., Sansom, D. M. and Trynka, G. (2020). Genomic profiling of T-cell activation suggests increased sensitivity of memory T cells to CD28 costimulation. *Genes Immun.* **21**, 390-408. doi:10.1038/s41435-020-00118-0

Goudy, K., Aydin, D., Barzaghi, F., Gambineri, E., Vignoli, M., Ciullini Mannurita, S., Doglioni, C., Ponzoni, M., Cicalese, M. P., Assanelli, A. et al. (2013). Human IL2RA null mutation mediates immunodeficiency with lymphoproliferation and autoimmunity. *Clin. Immunol.* **146**, 248-261. doi:10.1016/j.clim.2013.01.004

Griesemer, D., Xue, J. R., Reilly, S. K., Ulirsch, J. C., Kukreja, K., Davis, J. R., Kanai, M., Yang, D. K., Butts, J. C., Guney, M. H. et al. (2021). Genome-wide functional screen of 3'UTR variants uncovers causal variants for human disease and evolution. *Cell* **184**, 5247-5260.e19. doi:10.1016/j.cell.2021.08.025

Gusev, A., Lee, S. H., Trynka, G., Finucane, H., Vilhjálmsson, B. J., Xu, H., Zang, C., Ripke, S., Bulik-Sullivan, B., Stahl, E. et al. (2014). Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am. J. Hum. Genet.* **95**, 535-552. doi:10.1016/j.ajhg.2014.10.004

Giambartolomei, C., Vukcevic, D., Schadt, E. E., Franke, L., Hingorani, A. D., Wallace, C. and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383. doi:10.1371/journal.pgen.1004383

Guo, H., Fortune, M. D., Burren, O. S., Schofield, E., Todd, J. A. and Wallace, C. (2015). Integration of disease association and eQTL data using a Bayesian colocalisation approach highlights six candidate causal genes in immune-mediated diseases. *Hum. Mol. Genet.* **24**, 3305-3313. doi:10.1093/hmg/ddv077

Hampe, J., Franke, A., Rosenstiel, P., Till, A., Teuber, M., Huse, K., Albrecht, M., Mayr, G., De La Vega, F. M., Briggs, J. et al. (2007). A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat. Genet.* **39**, 207-211. doi:10.1038/ng1954

Hästbacka, J., de la Chapelle, A., Kaitila, I., Sistonen, P., Weaver, A. and Lander, E. (1992). Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nat. Genet.* **2**, 204-211. doi:10.1038/ng1192-204

Hay, D., Hughes, J. R., Babbs, C., Davies, J. O. J., Graham, B. J., Hanssen, L., Kassouf, M. T., Marieke Oudelaar, A. M., Sharpe, J. A., Suciu, M. C. et al. (2016). Genetic dissection of the α-globin super-enhancer *in vivo*. *Nat. Genet.* **48**, 895-903. doi:10.1038/ng.3605

Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., Ye, Z., Lee, L. K., Stuart, R. K., Ching, C. W. et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108-112. doi:10.1038/nature07829

Hiatt, J., Hultquist, J. F., McGregor, M. J., Bouhaddou, M., Leenay, R. T., Simons, L. M., Young, J. M., Haas, P., Roth, T. L., Tobin, V. et al. (2022). A functional map of HIV-host interactions in primary human T cells. *Nat. Commun.* **13**, 1752. doi:10.1038/s41467-022-29346-w

Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S. and Manolio, T. A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* **106**, 9362-9367. doi:10.1073/pnas.0903103106

Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-André, V., Sigova, A. A., Hoke, H. A. and Young, R. A. (2013). Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-947. doi:10.1016/j.cell.2013.09.053

Hormozdiari, F., van de Bunt, M., Segrè, A. V., Li, X., Joo, J. W. J., Bilow, M., Sul, J. H., Sankararaman, S., Pasaniuc, B. and Eskin, E. (2016). Colocalization

of GWAS and eQTL signals detects target genes. *Am. J. Hum. Genet.* **99**, 1245-1260. doi:10.1016/j.ajhg.2016.10.003

Hsieh, T. S., Cattoglio, C., Slobodyanyuk, E., Hansen, A. S., Rando, O. J., Tjian, R. and Darzacq, X. (2020). Resolving the 3D landscape of transcription-linked mammalian chromatin folding. *Mol. Cell* **78**, 539-553.e8. doi:10.1016/j.molcel.2020.03.002

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

Huang, H., Fang, M., Jostins, L., Umićević Mirkov, M., Boucher, G., Anderson, C. A., Andersen, V., Cleynen, I., Cortes, A., Crins, F. et al. (2017). Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature* **547**, 173-178. doi:10.1038/nature22969

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

Hugot, J. P., Chamaillard, M., Zouali, H., Lesage, S., Cézard, J. P., Belaiche, J., Almer, S., Tysk, C., O'Morain, C. A., Gassull, M. et al. (2001). Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411**, 599-603. doi:10.1038/35079107

International Genetics of Ankylosing Spondylitis Consortium (IGAS), Cortes, A., Hadler, J., Pointon, J. P., Robinson, P. C., Karaderi, T., Leo, P., Cremin, K., Pryce, K., Harris, J. et al. (2013). Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nat. Genet.* **45**, 730-738. doi:10.1038/ng.2667

International Multiple Sclerosis Genetics Consortium (IMSGC), Beecham, A. H., Patsopoulos, N. A., Xifara, D. K., Davis, M. F., Kemppinen, A., Cotsapas, C., Shah, T. S., Spencer, C. et al. (2013). Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat. Genet.* **45**, 1353-1360. doi:10.1038/ng.2770

Jaitin, D. A., Weiner, A., Yofe, I., Lara-Astiaso, D., Keren-Shaul, H., David, E., Salame, T. M., Tanay, A., van Oudenaarden, A. and Amit, I. (2016). Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-seq. *Cell* **167**, 1883-1896.e15. doi:10.1016/j.cell.2016.11.039

Javierre, B. M., Burren, O. S., Wilder, S. P., Kreuzhuber, R., Hill, S. M., Sewitz, S., Cairns, J., Wingett, S. W., Várnai, C., Thiecke, M. J. et al. (2016). Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* **167**, 1369-1384.e19. doi:10.1016/j.cell.2016.09.037

John, S., Robbins, C. M. and Leonard, W. J. (1996). An IL-2 response element in the human IL-2 receptor alpha chain promoter is a composite element that binds Stat5, Elf-1, HMG-I(Y) and a GATA family protein. *EMBO J.* **15**, 5627-5635. doi:10.1002/j.1460-2075.1996.tb00947.x

Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K. R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G. et al. (2013). DNA-binding specificities of human transcription factors. *Cell* **152**, 327-339. doi:10.1016/j.cell.2012.12.009

Kearns, N. A., Pham, H., Tabak, B., Genga, R. M., Silverstein, N. J., Garber, M. and Maehr, R. (2015). Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat. Methods* **12**, 401-403. doi:10.1038/nmeth.3325

Keller, M. F., Reiner, A. P., Okada, Y., van Rooij, F. J., Johnson, A. D., Chen, M. H., Smith, A. V., Morris, A. P., Tanaka, T., Ferrucci, L. et al. (2014). Trans-ethnic meta-analysis of white blood cell phenotypes. *Hum. Mol. Genet.* **23**, 6944-6960. doi:10.1093/hmg/ddu401

Kerimov, N., Hayhurst, J. D., Peikova, K., Manning, J. R., Walter, P., Kolberg, L., Samoviča, M., Sakthivel, M. P., Kuzmin, I., Trevanion, S. J. et al. (2021). A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nat. Genet.* **53**, 1290-1299. doi:10.1038/s41588-021-00924-w

Kheradpour, P., Ernst, J., Melnikov, A., Rogov, P., Wang, L., Zhang, X., Alston, J., Mikkelsen, T. S. and Kellis, M. (2013). Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res.* **23**, 800-811. doi:10.1101/gr.144899.112

Kichaev, G. and Pasaniuc, B. (2015). Leveraging functional-annotation data in trans-ethnic fine-mapping studies. *Am. J. Hum. Genet.* **97**, 260-271. doi:10.1016/j.ajhg.2015.06.007

Kim, K., Bang, S. Y., Lee, H. S., Cho, S. K., Choi, C. B., Sung, Y. K., Kim, T. H., Jun, J. B., Yoo, D. H., Kang, Y. M. et al. (2015). High-density genotyping of immune loci in Koreans and Europeans identifies eight new rheumatoid arthritis risk loci. *Ann. Rheum. Dis.* **74**, e13. doi:10.1136/annrheumdis-2013-204749

Kim-Hellmuth, S., Bechheim, M., Pütz, B., Mohammadi, P., Nédélec, Y., Giangreco, N., Becker, J., Kaiser, V., Fricker, N., Beier, E. et al. (2017). Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. *Nat. Commun.* **8**, 266. doi:10.1038/s41467-017-00366-1

King, J. A., Jeong, J., Underwood, F. E., Quan, J., Panaccione, N., Windsor, J. W., Coward, S., deBruyn, J., Ronksley, P. E., Shaheen, A. A. et al. (2020). Incidence of celiac disease is increasing over time: a systematic review and meta-analysis. *Am. J. Gastroenterol.* **115**, 507-525. doi:10.14309/ajg.0000000000000523

Klann, T. S., Black, J. B., Chellappan, M., Safi, A., Song, L., Hilton, I. B., Crawford, G. E., Reddy, T. E. and Gersbach, C. A. (2017). CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. *Nat. Biotechnol.* **35**, 561-568. doi:10.1038/nbt.3853

Korkmaz, G., Lopes, R., Ugalde, A. P., Nevedomskaya, E., Han, R., Myacheva, K., Zwart, W., Elkon, R. and Agami, R. (2016). Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat. Biotechnol.* **34**, 192-198. doi:10.1038/nbt.3450

Kowalski, M. H., Qian, H., Hou, Z., Rosen, J. D., Tapia, A. L., Shan, Y., Jain, D., Argos, M., Arnett, D. K., Avery, C. et al. (2019). Use of >100,000 NHLBI trans-omics for precision medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations. *PLoS Genet.* **15**, e1008500. doi:10.1371/journal.pgen.1008500

Kuhn, C., Rezende, R. M., M'Hamdi, H., da Cunha, A. P. and Weiner, H. L. (2017). IL-6 inhibits upregulation of membrane-bound TGF-β 1 on CD4+ T cells and blocking IL-6 enhances oral tolerance. *J. Immunol.* **198**, 1202-1209. doi:10.4049/jimmunol.1600921

Larke, M. S. C., Schwessinger, R., Nojima, T., Telenius, J., Beagrie, R. A., Downes, D. J., Oudelaar, A. M., Truch, J., Graham, B., Bender, M. A. et al. (2021). Enhancers predominantly regulate gene expression during differentiation via transcription initiation. *Mol. Cell* **81**, 983-997.e7. doi:10.1016/j.molcel.2021.01.002

Lehmkuhl, P., Gentz, M., de Otezya AC, G., Grimbacher, B., Schulze-Koops, H. and Skapenko, A. (2021). Dysregulated immunity in PID patients with low GARP expression on Tregs due to mutations in LRRC32. *Cell. Mol. Immunol.* **18**, 1677-1691. doi:10.1038/s41423-021-00701-z

Lesueur, L. L., Mir, L. M. and André, F. M. (2016). Overcoming the specific toxicity of large plasmids electrotransfer in primary cells in vitro. *Mol. Ther. Nucleic Acids* **5**, e291. doi:10.1038/mtna.2016.4

Li, G., Ruan, X., Auerbach, R. K., Sandhu, K. S., Zheng, M., Wang, P., Poh, H. M., Goh, Y., Lim, J., Zhang, J. et al. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**, 84-98. doi:10.1016/j.cell.2011.12.014

Li, K., Liu, Y., Cao, H., Zhang, Y., Gu, Z., Liu, X., Yu, A., Kaphle, P., Dickerson, K. E., Ni, M. et al. (2020). Interrogation of enhancer function by enhancer-targeting CRISPR epigenetic editing. *Nat. Commun.* **11**, 485. doi:10.1038/s41467-020-14362-5

Li, P., Mitra, S., Spolski, R., Oh, J., Liao, W., Tang, Z., Mo, F., Li, X., West, E. E., Gromer, D. et al. (2017). STAT5-mediated chromatin interactions in superenhancers activate IL-2 highly inducible genes: Functional dissection of the *Il2ra* gene locus. *Proc. Natl. Acad. Sci. USA* **114**, 12111-12119. doi:10.1073/pnas.1714019114

Li, Q., Gloudemans, M. J., Geisinger, J. M., Fan, B., Aguet, F., Sun, T., Ramaswami, G., Li, Y. I., Ma, J. B., Pritchard, J. K. et al. (2022). RNA editing underlies genetic risk of common inflammatory diseases. *Nature* **608**, 569-577. doi:10.1038/s41586-022-05052-x

Li, Y. I., van de Geijn, B., Raj, A., Knowles, D. A., Petti, A. A., Golan, D., Gilad, Y. and Pritchard, J. K. (2016). RNA splicing is a primary link between genetic variation and disease. *Science* **352**, 600-604. doi:10.1126/science.aad9417

Liao, W., Lin, J. X., Wang, L., Li, P. and Leonard, W. J. (2011). Modulation of cytokine receptors by IL-2 broadly regulates differentiation into helper T cell lineages. *Nat. Immunol.* **12**, 551-559. doi:10.1038/ni.2030

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O. et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293. doi:10.1126/science.1181369

Liu, C. T., Raghavan, S., Maruthur, N., Kabagambe, E. K., Hong, J., Ng, M. C., Hivert, M. F., Lu, Y., An, P., Bentley, A. R. et al. (2016). Trans-ethnic meta-analysis and functional annotation illuminates the genetic architecture of fasting glucose and insulin. *Am. J. Hum. Genet.* **99**, 56-75. doi:10.1016/j.ajhg.2016.05.006

Liu, S., Liu, Y., Zhang, Q., Wu, J., Liang, J., Yu, S., Wei, G. H., White, K. P. and Wang, X. (2017). Systematic identification of regulatory variants associated with cancer risk. *Genome Biol.* **18**, 194. doi:10.1186/s13059-017-1322-z

Liu, Z., Gerner, M. Y., Van Panhuys, N., Levine, A. G., Rudensky, A. Y. and Germain, R. N. (2015). Immune homeostasis enforced by co-localized effector and regulatory T cells. *Nature* **528**, 225-230. doi:10.1038/nature16169

Lowe, C. E., Cooper, J. D., Brusko, T., Walker, N. M., Smyth, D. J., Bailey, R., Bourget, K., Plagnol, V., Field, S., Atkinson, M. et al. (2007). Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the IL2RA region in type 1 diabetes. *Nat. Genet.* **39**, 1074-1082. doi:10.1038/ng2102

Lu, X., Chen, X., Forney, C., Donmez, O., Miller, D., Parameswaran, S., Hong, T., Huang, Y., Pujato, M., Cazares, T. et al. (2021). Global discovery of lupus genetic risk variant allelic enhancer activity. *Nat. Commun.* **12**, 1611. doi:10.1038/s41467-021-21854-5

Mägi, R., Horikoshi, M., Sofer, T., Mahajan, A., Kitajima, H., Franceschini, N., McCarthy, M. I., COGENT-Kidney Consortium, T2D–GENES Consortium and Morris, A. P. (2017). Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum. Mol. Genet.* **26**, 3639-3650. doi:10.1093/hmg/ddx280

Maier, L. M., Lowe, C. E., Cooper, J., Downes, K., Anderson, D. E., Severson, C., Clark, P. M., Healy, B., Walker, N., Aubin, C. et al. (2009). IL2RA genetic heterogeneity in multiple sclerosis and type 1 diabetes susceptibility and soluble interleukin-2 receptor production. *PLoS Genet.* **5**, e1000322. doi:10.1371/journal.pgen.1000322

Manz, J., Rodríguez, E., ElSharawy, A., Oesau, E. M., Petersen, B. S., Baurecht, H., Mayr, G., Weber, S., Harder, J., Reischl, E. et al. (2016). Targeted resequencing and functional testing identifies low-frequency missense variants in the gene encoding GARP as significant contributors to atopic dermatitis risk. *J. Invest. Dermatol.* **136**, 2380-2386. doi:10.1016/j.jid.2016.07.009

Marenholz, I., Esparza-Gordillo, J., Rüschendorf, F., Bauerfeind, A., Strachan, D. P., Spycher, B. D., Baurecht, H., Margaritte-Jeannin, P., Sääf, A., Kerkhof, M. et al. (2015). Meta-analysis identifies seven susceptibility loci involved in the atopic march. *Nat. Commun.* **6**, 8804. doi:10.1038/ncomms9804

Martin, P., McGovern, A., Orozco, G., Duffus, K., Yarwood, A., Schoenfelder, S., Cooper, N. J., Barton, A., Wallace, C., Fraser, P. 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

Martin-Trujillo, A., Patel, N., Richter, F., Jadhav, B., Garg, P., Morton, S. U., McKean, D. M., DePalma, S. R., Goldmuntz, E., Gruber, D. et al. (2020). Rare genetic variation at transcription factor binding sites modulates local DNA methylation profiles. *PLoS Genet.* **16**, e1009189. doi:10.1371/journal.pgen.1009189

Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., Reynolds, A. P., Sandstrom, R., Qu, H., Brody, J. et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190-1195. doi:10.1126/science.1222794

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., Kang, H. M., Fuchsberger, C., Danecek, P., Sharp, K. et al. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279-1283. doi:10.1038/ng.3643

Meddens, C. A., Harakalova, M., van den Dungen, N. A., Foroughi Asl, H., Hijma, H. J., Cuppen, E. P., Björkegren, J. L., Asselbergs, F. W., Nieuwenhuis, E. E. and Mokry, M. (2016). Systematic analysis of chromatin interactions at disease associated loci links novel candidate genes to inflammatory bowel disease. *Genome Biol.* **17**, 247. doi:10.1186/s13059-016-1100-3

Melnikov, A., Murugan, A., Zhang, X., Tesileanu, T., Wang, L., Rogov, P., Feizi, S., Gnirke, A., Callan, C. G., Jr, Kinney, J. B. et al. (2012). Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.* **30**, 271-277. doi:10.1038/nbt.2137

Melnikov, A., Zhang, X., Rogov, P., Wang, L. and Mikkelsen, T. S. (2014). Massively parallel reporter assays in cultured mammalian cells. *J. Vis. Exp.* **90**, 51719. doi:10.3791/51719

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

Moorthy, S. D., Davidson, S., Shchuka, V. M., Singh, G., Malek-Gilani, N., Langroudi, L., Martchenko, A., So, V., Macpherson, N. N. and Mitchell, J. A. (2017). Enhancers and super-enhancers have an equivalent regulatory role in embryonic stem cells through regulation of single or multiple genes. *Genome Res.* **27**, 246-258. doi:10.1101/gr.210930.116

Morris, A. P. (2011). Transethnic meta-analysis of genomewide association studies. *Genet. Epidemiol.* **35**, 809-822. doi:10.1002/gepi.20630

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

Musunuru, K., Strong, A., Frank-Kamenetsky, M., Lee, N. E., Ahfeldt, T., Sachs, K. V., Li, X., Li, H., Kuperwasser, N., Ruda, V. M. et al. (2010). From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* **466**, 714-719. doi:10.1038/nature09266

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

Nasrallah, R., Imianowski, C. J., Bossini-Castillo, L., Grant, F. M., Dogan, M., Placek, L., Kozhaya, L., Kuo, P., Sadiyah, F., Whiteside, S. K. et al. (2020). A distal enhancer at risk locus 11q13.5 promotes suppression of colitis by T_{reg} cells. *Nature* **583**, 447-452. doi:10.1038/s41586-020-2296-7

Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S. et al. (2014). Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376-381. doi:10.1038/nature12873

Oldridge, D. A., Wood, A. C., Weichert-Leahey, N., Crimmins, I., Sussman, R., Winter, C., McDaniel, L. D., Diamond, M., Hart, L. S., Zhu, S. et al. (2015). Genetic predisposition to neuroblastoma mediated by a LMO1 super-enhancer polymorphism. *Nature* **528**, 418-421. doi:10.1038/nature15540

Onengut-Gumuscu, S., Chen, W. M., Burren, O., Cooper, N. J., Quinlan, A. R., Mychaleckyj, J. C., Farber, E., Bonnie, J. K., Szpak, M., Schofield, E. et al. (2015). Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat. Genet.* **47**, 381-386. doi:10.1038/ng.3245

Ota, M., Nagafuchi, Y., Hatano, H., Ishigaki, K., Terao, C., Takeshima, Y., Yanaoka, H., Kobayashi, S., Okubo, M., Shirai, H. et al. (2021). Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases. *Cell* **184**, 3006-3021.e17. doi:10.1016/j.cell.2021.03.056

Oudelaar, A. M., Davies, J. O. J., Hanssen, L. L. P., Telenius, J. M., Schwessinger, R., Liu, Y., Brown, J. M., Downes, D. J., Chiariello, A. M., Bianco, S. et al. (2018). Single-allele chromatin interactions identify regulatory hubs in dynamic compartmentalized domains. *Nat. Genet.* **50**, 1744-1751. doi:10.1038/s41588-018-0253-2

Parker, S. C., Stitzel, M. L., Taylor, D. L., Orozco, J. M., Erdos, M. R., Akiyama, J. A., van Bueren, K. L., Chines, P. S., Narisu, N., NISC Comparative Sequencing Program et al. (2013). Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl. Acad. Sci. USA* **110**, 17921-17926. doi:10.1073/pnas.1317023110

Patterson, C. C., Karuranga, S., Salpea, P., Saeedi, P., Dahlquist, G., Soltesz, G. and Ogle, G. D. (2019). Worldwide estimates of incidence, prevalence and mortality of type 1 diabetes in children and adolescents: Results from the international diabetes federation diabetes Atlas, 9th edition. *Diabetes Res. Clin. Pract.* **157**, 107842. doi:10.1016/j.diabres.2019.107842

Patwardhan, R. P., Hiatt, J. B., Witten, D. M., Kim, M. J., Smith, R. P., May, D., Lee, C., Andrie, J. M., Lee, S. I., Cooper, G. M. et al. (2012). Massively parallel functional dissection of mammalian enhancers in vivo. *Nat. Biotechnol.* **30**, 265-270. doi:10.1038/nbt.2136

Peters, J. E., Lyons, P. A., Lee, J. C., Richard, A. C., Fortune, M. D., Newcombe, P. J., Richardson, S. and Smith, K. G. (2016). Insight into genotype-phenotype associations through eQTL mapping in multiple cell types in health and immune-mediated disease. *PLoS Genet.* **12**, e1005908. doi:10.1371/journal.pgen.1005908

Pickrell, J. K. (2014). Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am. J. Hum. Genet.* **94**, 559-573. doi:10.1016/j.ajhg.2014.03.004

Pipkin, M. E., Sacks, J. A., Cruz-Guilloty, F., Lichtenheld, M. G., Bevan, M. J. and Rao, A. (2010). Interleukin-2 and inflammation induce distinct transcriptional programs that promote the differentiation of effector cytolytic T cells. *Immunity* **32**, 79-90. doi:10.1016/j.immuni.2009.11.012

Rajagopal, N., Srinivasan, S., Kooshesh, K., Guo, Y., Edwards, M. D., Banerjee, B., Syed, T., Emons, B. J., Gifford, D. K. and Sherwood, R. I. (2016). High-throughput mapping of regulatory DNA. *Nat. Biotechnol.* **34**, 167-174. doi:10.1038/nbt.3468

Rioux, J. D., Xavier, R. J., Taylor, K. D., Silverberg, M. S., Goyette, P., Huett, A., Green, T., Kuballa, P., Barmada, M. M., Datta, L. W. et al. (2007). Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat. Genet.* **39**, 596-604. doi:10.1038/ng2032

Safiri, S., Kolahi, A. A., Hoy, D., Smith, E., Bettampadi, D., Mansournia, M. A., Almasi-Hashiani, A., Ashrafi-Asgarabad, A., Moradi-Lakeh, M., Qorbani, M. et al. (2019). Global, regional and national burden of rheumatoid arthritis 1990-2017: a systematic analysis of the Global Burden of Disease study 2017. *Ann. Rheum. Dis.* **78**, 1463-1471. doi:10.1136/annrheumdis-2019-215920

Sakaguchi, S., Sakaguchi, N., Asano, M., Itoh, M. and Toda, M. (1995). Immunologic self-tolerance maintained by activated T cells expressing IL-2 receptor alpha-chains (CD25). Breakdown of a single mechanism of self-tolerance causes various autoimmune diseases. *J. Immunol.* **155**, 1151-1164. doi:10.4049/jimmunol.155.3.1151

Sanjana, N. E., Wright, J., Zheng, K., Shalem, O., Fontanillas, P., Joung, J., Cheng, C., Regev, A. and Zhang, F. (2016). High-resolution interrogation of functional elements in the noncoding genome. *Science* **353**, 1545-1549. doi:10.1126/science.aaf7613

Sanyal, A., Lajoie, B. R., Jain, G. and Dekker, J. (2012). The long-range interaction landscape of gene promoters. *Nature* **489**, 109-113. doi:10.1038/nature11279

Sharfe, N., Dadi, H. K., Shahar, M. and Roifman, C. M. (1997). Human immune disorder arising from mutation of the alpha chain of the interleukin-2 receptor. *Proc. Natl. Acad. Sci. USA* **94**, 3168-3171. doi:10.1073/pnas.94.7.3168

Schmiedel, B. J., Singh, D., Madrigal, A., Valdovino-Gonzalez, A. G., White, B. M., Zapardiel-Gonzalo, J., Ha, B., Altay, G., Greenbaum, J. A., McVicker, G. et al. (2018). Impact of genetic polymorphisms on human immune cell gene expression. *Cell* **175**, 1701-1715.e16. doi:10.1016/j.cell.2018.10.022

Schmidt, E. M., Zhang, J., Zhou, W., Chen, J., Mohlke, K. L., Chen, Y. E. and Willer, C. J. (2015). GREGOR: evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. *Bioinformatics* **31**, 2601-2606. doi:10.1093/bioinformatics/btv201

Schmidt, R., Steinhart, Z., Layeghi, M., Freimer, J. W., Bueno, R., Nguyen, V. Q., Blaeschke, F., Ye, C. J. and Marson, A. (2022). CRISPR activation and interference screens decode stimulation responses in primary human T cells. *Science* **375**, eabj4008. doi:10.1126/science.abj4008

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

Shifrut, E., Carnevale, J., Tobin, V., Roth, T. L., Woo, J. M., Bui, C. T., Li, P. J., Diolaiti, M. E., Ashworth, A. and Marson, A. (2018). Genome-wide CRISPR screens in primary human T cells reveal key regulators of immune function. *Cell* **175**, 1958-1971. e15. doi:10.1016/j.cell.2018.10.024

Shin, H. Y., Willi, M., HyunYoo, K., Zeng, X., Wang, C., Metser, G. and Hennighausen, L. (2016). Hierarchy within the mammary STAT5-driven Wap super-enhancer. *Nat. Genet.* **48**, 904-911. doi:10.1038/ng.3606

Simeonov, D. R., Gowen, B. G., Boontanrart, M., Roth, T. L., Gagnon, J. D., Mumbach, M. R., Satpathy, A. T., Lee, Y., Bray, N. L., Chan, A. Y. et al. (2017). Discovery of stimulation-responsive immune enhancers with CRISPR activation. *Nature* **549**, 111-115. doi:10.1038/nature23875

Simeonov, D. R., Wong, H. S., Cortez, J. T., Young, A., Li, Z., Nguyen, V., Park, K., Umhoefer, J., Indart, A. C., Woo, J. M. et al. (2020). T cell subset-selective IL2RA enhancers shape autoimmune diabetes risk. *bioRxiv*. doi:10.1101/2020.07.22.216564

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

Sollid, L. M., Markussen, G., Ek, J., Gjerde, H., Vartdal, F. and Thorsby, E. (1989). Evidence for a primary association of celiac disease to a particular HLA-DQ alpha/beta heterodimer. *J. Exp. Med.* **169**, 345-350. doi:10.1084/jem.169.1.345

Song, L., Zhang, Z., Grasfeder, L. L., Boyle, A. P., Giresi, P. G., Lee, B. K., Sheffield, N. C., Gräf, S., Huss, M., Keefe, D. et al. (2011). Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res.* **21**, 1757-1767. doi:10.1101/gr.121541.111

Soskic, B., Cano-Gamez, E., Smyth, D. J., Rowan, W. C., Nakic, N., Esparza-Gordillo, J., Bossini-Castillo, L., Tough, D. F., Larminie, C. G. C., Bronson, P. G. et al. (2019). Chromatin activity at GWAS loci identifies T cell states driving complex immune diseases. *Nat. Genet.* **51**, 1486-1493. doi:10.1038/s41588-019-0493-9

Soskic, B., Cano-Gamez, E., Smyth, D. J., Ambridge, K., Ke, Z., Matte, J. C., Bossini-Castillo, L., Kaplanis, J., Ramirez-Navarro, L., Lorenc, A. et al. (2022). Immune disease risk variants regulate gene expression dynamics during $CD4^+$ T cell activation. *Nat. Genet.* **54**, 817-826. doi:10.1038/s41588-022-01066-3

Stahl, E. A., Raychaudhuri, S., Remmers, E. F., Xie, G., Eyre, S., Thomson, B. P., Li, Y., Kurreeman, F. A., Zhernakova, A., Hinks, A. et al. (2010). Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat. Genet.* **42**, 508-514. doi:10.1038/ng.582

Stuart, P. E., Tsoi, L. C., Nair, R. P., Ghosh, M., Kabra, M., Shaiq, P. A., Raja, G. K., Qamar, R., Thelma, B. K., Patrick, M. T. et al. (2022). Transethnic analysis of psoriasis susceptibility in South Asians and Europeans enhances fine-mapping in the MHC and genomewide. *HGG Adv.* **3**, 100069. doi:10.1016/j.xhgg.2021.100069

Taliun, D., Harris, D. N., Kessler, M. D., Carlson, J., Szpiech, Z. A., Torres, R., Taliun, S. A. G., Corvelo, A., Gogarten, S. M., Kang, H. M. et al. (2021). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290-299. doi:10.1038/s41586-021-03205-y

Thakore, P. I., D'Ippolito, A. M., Song, L., Safi, A., Shivakumar, N. K., Kabadi, A. M., Reddy, T. E., Crawford, G. E. and Gersbach, C. A. (2015). Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat. Methods* **12**, 1143-1149. doi:10.1038/nmeth.3630

Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E., Sheffield, N. C., Stergachis, A. B., Wang, H., Vernot, B. et al. (2012). The accessible chromatin landscape of the human genome. *Nature* **489**, 75-82. doi:10.1038/nature11232

Tewhey, R., Kotliar, D., Park, D. S., Liu, B., Winnicki, S., Reilly, S. K., Andersen, K. G., Mikkelsen, T. S., Lander, E. S., Schaffner, S. F. et al. (2016). Direct identification of hundreds of expression-modulating variants using a multiplexed reporter assay. *Cell* **165**, 1519-1529. doi:10.1016/j.cell.2016.04.027

Ting, P. Y., Parker, A. E., Lee, J. S., Trussell, C., Sharif, O., Luna, F., Federe, G., Barnes, S. W., Walker, J. R., Vance, J. et al. (2018). Guide Swap enables genome-scale pooled CRISPR-Cas9 screening in human primary cells. *Nat. Methods* **15**, 941-946. doi:10.1038/s41592-018-0149-1

Todd, J. A., Bell, J. I. and McDevitt, H. O. (1987). HLA-DQ beta gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. *Nature* **329**, 599-604. doi:10.1038/329599a0

Tong, A. J., Liu, X., Thomas, B. J., Lissner, M. M., Baker, M. R., Senagolage, M. D., Allred, A. L., Barish, G. D. and Smale, S. T. (2016). A stringent systems approach uncovers gene-specific mechanisms regulating inflammation. *Cell* **165**, 165-179. doi:10.1016/j.cell.2016.01.020

Trynka, G., Hunt, K. A., Bockett, N. A., Romanos, J., Mistry, V., Szperl, A., Bakker, S. F., Bardella, M. T., Bhaw-Rosun, L., Castillejo, G. et al. (2011). Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat. Genet.* **43**, 1193-1201. doi:10.1038/ng.998

Tsoi, L. C., Spain, S. L., Knight, J., Ellinghaus, E., Stuart, P. E., Capon, F., Ding, J., Li, Y., Tejasvi, T., Gudjonsson, J. E. et al. (2012). Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. *Nat. Genet.* **44**, 1341-1348. doi:10.1038/ng.2467

Ulirsch, J. C., Nandakumar, S. K., Wang, L., Giani, F. C., Zhang, X., Rogov, P., Melnikov, A., McDonel, P., Do, R., Mikkelsen, T. S. et al. (2016). Systematic functional dissection of common genetic variation affecting red blood cell traits. *Cell* **165**, 1530-1545. doi:10.1016/j.cell.2016.04.048

Vahedi, G., Kanno, Y., Furumoto, Y., Jiang, K., Parker, S. C., Erdos, M. R., Davis, S. R., Roychoudhuri, R., Restifo, N. P., Gadina, M. et al. (2015). Super-enhancers delineate disease-associated regulatory nodes in T cells. *Nature* **520**, 558-562. doi:10.1038/nature14154

van Ouwerkerk, A. F., Bosada, F. M., Liu, J., Zhang, J., van Duijvenboden, K., Chaffin, M., Tucker, N. R., Pijnappels, D., Ellinor, P. T., Barnett, P. et al. (2020). Identification of functional variant enhancers associated with atrial fibrillation. *Circ. Res.* **127**, 229-243. doi:10.1161/CIRCRESAHA.119.316006

van Rooij, F. J. A., Qayyum, R., Smith, A. V., Zhou, Y., Trompet, S., Tanaka, T., Keller, M. F., Chang, L. C., Schmidt, H., Yang, M. L. et al. (2017). Genome-wide trans-ethnic meta-analysis identifies seven genetic loci influencing erythrocyte traits and a role for RBPMS in erythropoiesis. *Am. J. Hum. Genet.* **100**, 51-63. doi:10.1016/j.ajhg.2016.11.016

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

Wallace, C., Rotival, M., Cooper, J. D., Rice, C. M., Yang, J. H., McNeill, M., Smyth, D. J., Niblett, D., Cambien, F., Cardiogenics Consortium et al. (2012). Statistical colocalization of monocyte gene expression and genetic risk variants for type 1 diabetes. *Hum. Mol. Genet.* **21**, 2815-2824. doi:10.1093/hmg/dds098

Wallace, C., Cutler, A. J., Pontikos, N., Pekalski, M. L., Burren, O. S., Cooper, J. D., García, A. R., Ferreira, R. C., Guo, H., Walker, N. M. et al. (2015). Dissection of a complex disease susceptibility region using a bayesian stochastic search approach to fine mapping. *PLoS Genet.* **11**, e1005272. doi:10.1371/journal.pgen.1005272

Walton, C., King, R., Rechtman, L., Kaye, W., Leray, E., Marrie, R. A., Robertson, N., La Rocca, N., Uitdehaag, B., van der Mei, I. et al. (2020). Rising prevalence of multiple sclerosis worldwide: Insights from the Atlas of MS, third edition. *Mult. Scler.* **26**, 1816-1821. doi:10.1177/1352458520970841

Wang, K., Baldassano, R., Zhang, H., Qu, H. Q., Imielinski, M., Kugathasan, S., Annese, V., Dubinsky, M., Rotter, J. I., Russell, R. K. et al. (2010). Comparative genetic analysis of inflammatory bowel disease and type 1 diabetes implicates multiple loci with opposite effects. *Hum. Mol. Genet.* **19**, 2059-2067. doi:10.1093/hmg/ddq078

Waysbort, N., Russ, D., Chain, B. M. and Friedman, N. (2013). Coupled IL-2-dependent extracellular feedbacks govern two distinct consecutive phases of CD4 T cell activation. *J. Immunol.* **191**, 5822-5830. doi:10.4049/jimmunol.1301575

Weirauch, M. T., Yang, A., Albu, M., Cote, A. G., Montenegro-Montero, A., Drewe, P., Najafabadi, H. S., Lambert, S. A., Mann, I., Cook, K. et al. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431-1443. doi:10.1016/j.cell.2014.08.009

Wellcome Trust Case Control Consortium, Maller, J. B., McVean, G., Byrnes, J., Vukcevic, D., Palin, K., Su, Z., Howson, J. M., Auton, A., Myers, S. et al. (2012). Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat. Genet.* **44**, 1294-1301. doi:10.1038/ng.2435

Wen, X., Pique-Regi, R. and Luca, F. (2017). Integrating molecular QTL data into genome-wide genetic association analysis: Probabilistic assessment of enrichment and colocalization. *PLoS Genet.* **13**, e1006646. doi:10.1371/journal.pgen.1006646

Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., Rahl, P. B., Lee, T. I. and Young, R. A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319. doi:10.1016/j.cell.2013.03.035

Wojcik, G. L., Graff, M., Nishimura, K. K., Tao, R., Haessler, J., Gignoux, C. R., Highland, H. M., Patel, Y. M., Sorokin, E. P., Avery, C. L. et al. (2019). Genetic analyses of diverse populations improves discovery for complex traits. *Nature* **570**, 514-518. doi:10.1038/s41586-019-1310-4

Xue, J., Schmidt, S. V., Sander, J., Draffehn, A., Krebs, W., Quester, I., De Nardo, D., Gohel, T. D., Emde, M., Schmidleithner, L. et al. (2014). Transcriptome-based network analysis reveals a spectrum model of human macrophage activation. *Immunity* **40**, 274-288. doi:10.1016/j.immuni.2014.01.006

Yazar, S., Alquicira-Hernandez, J., Wing, K., Senabouth, A., Gordon, M. G., Andersen, S., Lu, Q., Rowson, A., Taylor, T. R. P., Clarke, L. et al. (2022). Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. *Science* **376**, eabf3041. doi:10.1126/science.abf3041

Yin, X., Low, H. Q., Wang, L., Li, Y., Ellinghaus, E., Han, J., Estivill, X., Sun, L., Zuo, X., Shen, C. et al. (2015). Genome-wide meta-analysis identifies multiple novel associations and ethnic heterogeneity of psoriasis susceptibility. *Nat. Commun.* **6**, 6916. doi:10.1038/ncomms7916

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

Zhou, A. X., Kozhaya, L., Fujii, H. and Unutmaz, D. (2013). GARP-TGF-β complexes negatively regulate regulatory T cell development and maintenance of peripheral CD4+ T cells in vivo. *J. Immunol.* **190**, 5057-5064. doi:10.4049/jimmunol.1300065