Current challenges in understanding the role of enhancers in disease

Zaugg, Judith Barbara; Sahlén, Pelin; Andersson, Robin; Alberich-Jorda, Meritxell; de Laat, Wouter; Deplancke, Bart; Ferrer, Jorge; Mandrup, Susanne; Natoli, Gioacchino; Plewczynski, Dariusz; Rada-Iglesias, Alvaro; Spicuglia, Salvatore

Published in:
Nature Structural and Molecular Biology

DOI:
10.1038/s41594-022-00896-3

Publication date:
2022

Document version:
Accepted manuscript

Current challenges in understanding the role of enhancers in disease

Judith Barbara Zaugg[1,*], Pelin Sahlén[2,*], Robin Andersson[3,4,*], Meritxell Alberich-Jorda[5,6], Wouter de Laat[7], Bart Deplancke[8], Jorge Ferrer[9,10,11], Susanne Mandrup[4,12], Gioacchino Natoli[13], Dariusz Plewczynski[14], Alvaro Rada-Iglesias[15], Salvatore Spicuglia[16,*,$]

[1] European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117, Heidelberg

[2] Science for Life Laboratory, KTH - Royal Institute of Technology, Tomtebodavägen 23A, 17165, Solna, Sweden

[3] Section for Computational and RNA Biology, Department of Biology, University of Copenhagen, 2200 Copenhagen N, Denmark

[4] The Novo Nordisk Foundation Center for Genomic Mechanisms of Disease, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

[5] Department of Hemato-oncology, Institute of Molecular Genetics of the CAS, Vídeňská 1083, 14200 Prague 4, Czech Republic

[6] Childhood Leukaemia Investigation Prague, Department of Pediatric Haematology and Oncology, 2nd Faculty of Medicine, Charles University in Prague, University Hospital Motol, V Uvalu 84, Praha 150 06, Czech Republic

[7] Hubrecht Institute-KNAW, Oncode Institute and University Medical Center Utrecht, 3584 CT Utrecht, The Netherlands

[8] Laboratory of Systems Biology and Genetics, Institute of Bio-engineering, School of Life Sciences, EPFL, CH-1015 Lausanne, Switzerland

[9] Centre for Genomic Regulation, the Barcelona Institute of Science and Technology, Barcelona 08003, Spain

[10] Centro de Investigación Biomédica en red Diabetes y enfermedades metabólicas asociadas (CIBERDEM), Madrid 28029, Spain

[11] Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK

[12] Functional Genomics and Metabolism Research Unit, Department of Biochemistry and Molecular Biology, University of Southern Denmark, 5230 Odense M, Denmark

[13] IEO, European Institute of Oncology - IRCCS, Milan, Italy

[14] Laboratory of Functional and Structural Genomics, Centre of New Technologies, University of Warsaw, Banacha 2c, 02-097 Warsaw, Poland

[15] Institute of Biomedicine and Biotechnology of Cantabria (IBBTEC), CSIC/University of Cantabria, Albert Einstein 22, 39011, Santander, Spain

[16] TAGC, UMR 1090, Aix-Marseille University, INSERM, Marseille, France

* These authors contributed equally

[$] To whom correspondence should be addressed, email: salvatore.spicuglia@inserm.fr

Preface

Enhancers play a central role in the spatiotemporal control of gene expression and tend to work in a cell type-specific manner. In addition, they are suggested to be major contributors to phenotypic variation, evolution and disease. There is growing evidence that enhancer dysfunction due to genetic, structural, or epigenetic mechanisms contributes to a broad range of human diseases referred to as enhanceropathies. Such mechanisms often underlie the susceptibility to common diseases, but can also play a direct causal role in cancer or Mendelian diseases. Despite the recent gain of insights into enhancer biology and function, we still have a limited ability to predict how enhancer dysfunction impacts gene expression. Here, we discuss major challenges that need to be overcome when studying the role of enhancers in disease etiology and highlight opportunities and directions for future studies, aiming to disentangle the molecular basis of enhanceropathies.

Main text

Regulation of gene expression is accomplished through the integration of events at regulatory elements that are proximal (promoters) and distal (enhancers) to gene transcription start sites (TSSs). Forty years after their discovery[1], enhancers are recognized as playing a central role in the spatiotemporal control of gene expression underlying human development and homeostasis[2]. Enhancers are short stretches of DNA that act as positive regulators of transcription via their ability to bind key proteins – transcription factors (TFs) – and complexes that control gene expression. Enhancer regulation of genes involves the three-dimensional topology of chromatin, affecting the frequency by which enhancers and gene promoters come into close proximity. Through this topology, several configurations can arise beyond single enhancer-gene pairs, including one-to-many and many-to-many enhancer-gene wirings, which may affect the robustness, strength or specificity of gene expression (**Fig. 1A**).

Enhancer dysfunction has emerged as a central mechanism in the pathogenesis of certain diseases[3,4] (**Table 1**). In particular, the dysfunction of enhancers by either point mutations or structural variants is a significant mechanism underlying aberrant gene regulation in cancer[5] and Mendelian diseases[6]. Moreover, genetic variants associated with common diseases are frequently found in *cis*-regulatory elements including enhancers[7–12]. Depending on the nature of the genetic alteration, enhancer dysfunction can be classified into two main types[4]. The first type involves small single nucleotide variants (SNVs) and indels in the enhancer sequence that lead to changes in enhancer activity (**Fig. 1B**). Such variations can for instance alter the affinity of bound TFs or create new binding sites. The second type involves structural variants that lead to deletion, duplication or relocation of the entire enhancer, which impacts chromatin topology and enhancer function (**Fig. 1C**). Chromosomal rearrangements can lead to rewiring of enhancer-gene connections, which may involve both enhancer adoption/hijacking (gain-of-function; e.g., ref. [13,14]), and enhancer disconnection (loss-of-function; e.g., ref. [15]). Depending on the genomic alteration, enhancer dysfunction may result in either gain or loss of gene expression in a given tissue as well as more complex alterations of expression patterns (**Table 1**).

Elucidating the molecular basis of enhancer function in normal and pathological conditions has far-reaching translational implications. Here, we discuss important challenges that need

to be considered in the study of enhancer dysfunction in disease and highlight critical areas of research to address these challenges in the near future (**Fig. 2**).

Current challenges in characterizing enhancer dysfunction in disease

1) Identifying and validating functional enhancers

A persisting challenge when studying transcriptional regulation in health and disease is to systematically identify functionally meaningful enhancers in a given cell type. Partially, this is because enhancers encompass a diverse group of regulatory sequences, which may utilize various mechanisms to control gene expression[2,16]. Conceptually, a nucleotide sequence can be assigned as a biologically meaningful enhancer once it is experimentally demonstrated that it modulates the transcription of a gene in *cis* in its native context. Unfortunately, there is no high-throughput assay that is able to do this globally outside of cell lines. As a result, enhancers are often defined in an indirect, operational, manner[2]. For example, a sequence may be functionally qualified as an enhancer if it increases the activity of a (minimal) promoter in a plasmid reporter assay, or it may be qualified as an enhancer by association, when linked with chromatin accessibility, transcription of enhancer RNAs (eRNAs), or marked by epigenomic features that have been linked to enhancer activity (e.g. p300, H3K4me1, H3K27ac). Functionally cataloging sequences as candidate enhancers has been boosted by the development of Massively Parallel Reporter Assays (MPRAs) that allow systematic large-scale testing of enhancer activities of any sequence in episomal contexts[2], while genome-wide approaches to find putative enhancers by association were employed by several international consortia (ENCODE[17], ROADMAP[18], FANTOM[7] and BLUEPRINT[19]). While these approaches have greatly expanded our ability to map enhancers, they suffer from some limitations. One is that not all sequences that are predicted to act as enhancers based on MPRAs or association strategies necessarily function to increase gene transcription in their endogenous chromosomal contexts. Furthermore, MPRAs do not account for the effects that linear distance and chromatin environment might have on enhancer activity. In principle, CRISPR-based enhancer screens are able to overcome these limitations and can be used to assess the importance of enhancers in their endogenous context[20–22]. However, these assays may suffer

from the intrinsic redundancy or additive effects of enhancers (see Challenge 3), and high false-negative rates[2]. To further complicate things, some proven biologically meaningful enhancers may lack the expected biochemical marks while others may not show enhancer potential in reporter assays[23,24], perhaps because these assays are based on plasmids and may not reproduce the function of chromatin-dependent enhancers[24,25]. Therefore, it is not surprising that there is no strict overlap[16] between the hundreds of thousands of putative enhancers in the human genome predicted based on indirect assays and the number of biologically confirmed enhancers. Future comparisons between predicted and functionally validated enhancers, including CRISPR screens with high-sensitivity and low false-negative rates, should help improve our definition of the fundamental features of functionally operational enhancers.

2) Defining the spatiotemporal window in which a regulatory variant affects enhancer activity

Enhancer activity is highly specific during development, across cell types, cell states, and stimulatory conditions (e.g., inflammatory response, diet, drug treatment). As a consequence, the effect of enhancer dysfunction (**Fig. 1B**), may only be revealed in specific contexts or tissues. This leads to the challenge of defining the spatiotemporal window in which an enhancer dysfunction has a measurable and meaningful impact. This is particularly important for interpreting disease-associated genetic variants in non-coding regions since the cell type in which an enhancer is active can be informative about the disease mechanism. For instance, the finding that genetic variants associated with Alzheimer's disease, a neurodegenerative condition, overlap enhancers in myeloid cells, rather than neurons[26], has led to a shift in the research focus of the pathology[27]. Similarly, functional assessment of obesity-associated variants has identified putatively causal variants with regulatory properties in both adipose and neuronal cell lines[28].

To capture inducible and context-dependent enhancers, as well as those restricted to rare cell types or developmental stages, efforts in enhancer mapping need to focus on different stimulatory conditions, environmental contexts, developmental stages and rare cell types[29–31]. Single-cell technologies are particularly well-suited for studying rare cell types. Specifically,

single-cell chromatin accessibility assays (scATAC-seq, Assay for Transposase-Accessible CRickelshromatin and sequencing) can serve to operationally predict enhancer activity and have facilitated the functional interpretation of disease-associated non-coding variants in adult and fetal tissues[32–34]. scATAC-seq will be key for expanding putative enhancer maps in diverse (rare) cell types. Furthermore, since chromatin accessibility does not necessarily reflect enhancer activity[7], further development of single-cell technologies that employ orthogonal measures of enhancer activity, e.g., large-scale perturbation assays[35,36], will be crucial to get more confidence in assessing enhancer function for rare cell types.

In line with the importance of cataloging enhancers and their restricted activities, there is an urgent need to assess the functional impact of regulatory variants during development and differentiation. To this end, recent years have seen a promising development of biological models such as transgenic mice and zebrafish[37–39], genetically manipulated human-induced Pluripotent Cells (hiPSC) or immortalized precursor cells[15,27,40,41], human-mouse chimeras[42,43] and organoids[44]. These *in vivo* and *ex vivo* models, in combination with assays to assess developmental and differentiation potential, will facilitate the study of genetic variants and determine their impact in contexts closer to human diseases.

3) Understanding the interplay between cis-regulatory elements

Enhancers are not only highly context-dependent, but they also often work together in regulatory domains to achieve the correct gene expression output. Thus, a major challenge is to understand the interplay between enhancers and other regulatory elements, including promoters, and how the joint activity of a domain is influenced by disruptions of individual enhancers. Multiple enhancers for the same gene may allow distinct enhancers to either be activated under different conditions or to cooperate, both of which can lead to robustness in gene activity[24,37,45–47]. For instance, many developmental genes are associated with “shadow” enhancers with similar transcription factor (TF) binding to ensure robust expression under suboptimal conditions[45,48,49], an observation that has been confirmed by 3D topology-based methods that revealed a complex landscape of multiple enhancer interactions per gene[50–53]. In fact, highly coordinated enhancer activity has been linked to the regulation of cell identity genes[54], signal integration and compartmentalization of the genome[55]. As a consequence of

such regulatory complexity, many enhancers might not, individually, reveal a strong phenotype when disrupted in their endogenous context[23,56], while still possessing endogenous enhancer activity. Thus, the presence of multiple enhancers[57] *per* gene may either additively or synergistically achieve a higher transcriptional output of a gene or provide redundancy and mutational robustness to its expression. Systematic testing of enhancer-promoter compatibilities will help to better understand the still unclear connectivity rules[58,59] that control gene transcription in the human genome.

Elucidating the mechanisms and contexts, including the cell type-specific 3D topology, by which regulatory domains and TFs establish robustness or synergism will therefore be crucial to further our understanding of enhanceropathies. Combinatorial interference or perturbation of multiple enhancers within a regulatory domain will be necessary to understand the principles by which enhancers act together and their effects on gene regulation.

4) Identifying the target genes of enhancers

Enhancers ultimately need to be defined by their role in enhancing endogenous gene expression, which leads to the next challenge: the identification of their target genes. This is particularly challenging for enhancers that are located distally to any gene promoter. It is assumed that distal enhancers have to come into physical proximity to their target gene in order to function, as first demonstrated by chromosome conformation capture (3C) methods in the beta-globin locus[60]. Thus, for the operational mapping of target genes, chromatin-topology assays are key to determine the physical proximity between enhancers and their putative genes. These technologies can map direct contacts (chromatin loops) and at the same time identify larger domains, so-called topologically associating domains (TADs), which have a high density of physical chromatin interactions[61]. The main caveats of using direct contacts for mapping enhancer-gene pairs are that the 3C-technologies typically require large cell numbers (with some exceptions[50,52] and may thus miss enhancer-gene pairs that are looped only in a subset of cells or contacts that are highly transient. TAD-based analyses suffer from low resolution since they typically comprise multiple genes and enhancers and can, on their own at best, restrict the search space for putative target genes[62]. There are a

complementary set of approaches to map enhancer-gene pairs such as targeted Hi-C, where chromatin interactions of regions of interest such as promoters and/or enhancers are captured to increase resolution[55,63], or expression quantitative trait loci (eQTL) mapping, where enhancer genetic variants are associated with mRNA expression changes across individuals[64]. Other approaches use covariation between molecular phenotypes (e.g., histone marks, chromatin accessibility, expression) of enhancers and genes across individuals or cell types[9,65–67] or combine chromatin states and long-range interactions[68], to construct genome-wide maps of enhancer–gene connections in a given cell type. The advantage of these methods is that relying on enhancer-gene co-variation does not assume a specific mechanism of how enhancers regulate gene expression, and can therefore also capture transient enhancer-gene contacts[69]. Here, the caveats are that these methods require molecular data across a large number of individuals or cell types, and they may miss constitutively active enhancers that do not vary much across samples. Given their descriptive (for the 3C technologies) and correlative (for the co-variation methods) nature, all of these approaches provide an operational prediction of putative enhancer-gene pairs. For a functional mapping of target genes, CRISPR-mediated enhancer deletion or inactivation, followed by gene expression analysis[29,68], is the most direct way to search for target genes. However, such CRISPR-based approaches may miss links due to low effect sizes and are often limited to cultured cells. In conclusion, current approaches still have difficulties identifying with high confidence the target genes of enhancers, and likely, the combination of different strategies might improve the efficiency of identifying disease-targeted genes[70]. Recent advances in applying machine learning to predict cell-type specific expression based on DNA sequence[71] show great promise to generate defined and experimentally testable hypotheses. These models were enabled by the vast resources of transcriptomics and genomics data that have been assembled by the community, and additional data, particularly from less accessible cellular states and developmental stages, will further improve the power of these methods.

5) Understanding the grammar of enhancer activity

Gene regulatory elements, including enhancers[7–9], are regulated by TFs, or TF-recruited co-activators, which bind to the enhancer element at any given time and cellular state. It is thus not surprising that genetic variants that disrupt a TF binding motif are enriched among

variants associated with molecular phenotypes, such as histone marks[8] or tissue-specific expression levels[72], and can be disease-causative[73]. For example, a mutation in a *SOX9* enhancer, associated with Pierre Robin Syndrome, disrupts the binding of the TF MSX1[74] (see other examples in **Table 1**). However, the majority of molecular trait-associated Single Nucleotide Polymorphisms (SNPs) do not disrupt known TF binding sites[8], leading to the next open challenge in understanding enhancer dysfunction: to identify the rules by which enhancer sequence determines its activity. Concepts, such as Variable Chromatin Modules (VCMs), where the effects of a lead SNP affecting a local chromatin domain (e.g. through TF binding site disruption) spread into the local vicinity, can explain the missing mechanism to some extent[8,10,75]. Recent studies revealed that flanking regions of TF binding sites are highly informative for some TFs to bind[76] and they impact the enhancer potential of the encompassing regulatory element[77], suggesting we are still missing part of the grammar for TF binding. In line with this, up to 30% of human TFs have no characterized binding motif[78]. Consequently, interpreting regulatory variant-to-phenotype associations requires fundamental insights into the sequence determinants of TF binding and enhancer activity. Here, sequence-based machine learning to model TF binding[76,79,80], enhancer activity[25,77,81] and topologies[71] show promise. However, major challenges remain, including the difficulty to accurately interpret such models, the lack of sufficient training or validation data, and the need to improve accuracy and generalization across cell types/contexts. In parallel, experimental approaches that measure the functional impact of genetic variants on regulatory activity and TF binding in a large-scale, such as MPRA-based approaches[82–86] and SNP-SELEX[87], can provide comprehensive experimental fine-mapping of likely causal variants. Overall, these insights will be crucial for the interpretation of the potential effect and severity of enhancer dysfunction, and thus the potentially implicated genetic variants, within complex regulatory domains.

6) Understanding how TF cooperation defines enhancer activity and specificity

Enhancers integrate non-mutually exclusive layers of molecular information: their function can be impacted by genetic variants/mutations, by epigenetic chromatin remodeling that is typically set up by lineage-specific TFs, or by signaling cascades regulated by stimulus-responsive TFs[88] (**Fig. 1**). Here, we focus on the challenge of understanding the role of TFs and

epigenetics on enhancer dysfunction. Lineage- and developmental-stage specific enhancers, typically regulated by lineage-specific TFs, may define the gene expression potential of a cell, and whether or not it will be able to mount a specific response to a given stimulus[89]. In particular, during development or differentiation, enhancers and whole chromatin domains can be primed in progenitor cells towards certain lineages before gene expression changes are obvious, e.g., during adipogenesis[55]. In contrast, enhancers that are under the control of stimulus-responsive TFs essentially act as signaling response elements and connect cell-extrinsic signals to gene expression programs. Conceptually, lineage-specific TFs and the chromatin accessibility landscape they set up determine the scope of stimulus-regulated TFs. This way, stimulus-responsive TFs can access enhancers that are pre-marked and kept accessible by lineage-specific TFs, thus integrating the two layers of regulation[90]. As a consequence, some response-TFs, such as NF-kB, bind completely different enhancers depending on the cell type in which they are activated[91]. This is consistent with observations that a TF can regulate completely different sets of genes depending on the cell type[92], which is partially explained by the cooperative interaction of TFs[93]. Yet, apart from a couple of well-studied examples, very little is known about the contribution of TF cooperativity, enhancer priming (that can also be TF mediated) and permissive chromatin, which in turn may define the TF regulon (i.e., the set of target genes regulated by a given TF). To fully understand enhancer dysfunction, it is important to study the cell type- or condition-specific TF regulons, and how they are defined by the combinatorial or cooperative binding grammar of enhancer sequences in normal and pathological conditions. Diverse TF-centric studies are even more important given the current literature bias with many studies focusing on a small set of TFs while the majority of TFs are vastly understudied[94].

7) Deciphering the impact and interactions of regulatory mutations in disease

The challenges above culminate in the ultimate challenge of identifying and understanding pathogenic enhancer dysfunction and eventually using this knowledge in clinically relevant studies (**Fig. 2**). The specific challenges that need to be solved for understanding a certain disease depend on the type of enhancer dysfunction and the nature of the genetic alteration (rare vs. common). For rare diseases, few examples of causal enhancer mutations have been established as compared to mutations in the coding genome. It is currently unclear whether

this limited number of reported enhancer mutations in rare disorders is because they do not exist or because we have not been able to find them due to the lack of data and statistical power. Either way, the additional challenge for identifying causal mutations in enhancers vs. coding regions is that each genome carries around 2,000 structural and 8,500 private non-coding variants[95], which are often not even captured since exome-sequencing is still the standard for diagnosing rare diseases. On the other hand, GWA studies have revealed hundreds of non-coding variants of significance for common disease risk, suggesting that the aggregated effect of variants in multiple enhancers modulate common disease risk. Fine-mapping studies aimed at identifying the causal variant(s) among those linked in a haplotype block typically integrate significantly associated variants with experimentally determined enhancer characteristics, as discussed above. While successful for the identification of some causal variants (e.g. [96,97]), this is often difficult because the relevant cell type and trans-acting nuclear environment are not known (challenge 2), the role of the encompassing regulatory domain is not well understood (challenge 3) and the target gene of the affected enhancer is not identified (challenge 4). Fine-mapping of causal signals and effect size predictions can be improved by expanding the battery of GWA studies with cohorts of diverse ancestries[98,99], and computational tools ranking genes based on their dosage-dependent pathogenicity. This allows hypothesis-driven studies where candidate target genes and enhancers are tested simultaneously to measure their combined effects on inferred functions[100]. Furthermore, for common diseases, both genetic and environmental factors contribute to the disease etiology. Therefore, the effects of certain non-coding genetic variants might only or preferentially be manifested under certain environmental conditions. Together with a significant shift for using whole-genome instead of whole-exome sequences as a diagnostic utility, and consequently, an increasing amount of whole-genome data accumulating thanks to biobanks and cohort studies[101], these tools will likely provide much better constraints on assessing disease causality and could pave the way towards systematic prediction of pathogenicity of regulatory variants and mutations for both rare and common variants[102].

Future directions

As disease-associated regulatory mutations at enhancers are increasingly identified, there is an urgent need to fully characterize enhancer mutations to enable their use in functional and

clinical genomic studies (**Box 1**). Besides the complexity of studying enhancer function in normal contexts, the characterization of noncoding variants affecting enhancer activity in disease adds additional challenges ranging from the identification of a credible set of regulatory variants to the identification of tissues and developmental contexts in which variants have an effect. Despite the wealth of data on enhancer activity across multiple cell- and tissue-types, it is challenging to fully utilize the vast potential of such datasets, highlighting the importance of good data-sharing practices. In addition, the majority of available data informing on enhancer activities are derived from populations of cells, disregarding the stochasticity and plasticity of regulatory events across individual cells. However, due to the complexity of the regulatory landscape, we propose that the field should move beyond the generation of enhancer catalogs and invest more in experimental and computational efforts to identify their target genes, in particular for the prioritization of disease-relevant genes susceptible to dysfunction upon misregulation. This can only be uncovered using Systems Biology, computational modeling approaches, and targeted experimental systems. Focused efforts and datasets will enable hypothesis-driven investigations of a set of variants or genes for a given disease phenotype and further inform the modeling of enhancer function from catalog data. Ultimately, the acquired knowledge should allow the implementation of novel strategies to genetically or epigenetically modify enhancer function to treat the associated diseases.

Acknowledgments

All authors are members of the EU-funded Innovative Training Network "Molecular Basis of Human Enhanceropathies" (Enhpathy, www.enhpathy.eu) and received funding from the Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 860002. The Enhpathy network has set up a multidisciplinary scientific consortium in the continuum of basic, translational and clinical research on enhancers and associated diseases.

Competing interests

The authors declare no competing interest.

Table 1. Representative examples of enhancer dysfunction driving disease

Type of disease	Disease	Affected gene(s)	Enhancer*	Type of disruption	Effect on gene exp.	ref
Monogenic (Mendelian)	□-Thalassaemia	□-*globin* genes	LCR	Enhancer deletions	LOE	103,104
	□-Thalassaemia	□-*globin* genes	□-globin enhancers	Deletion or insertion of promoters alter enhancer-gene connectivity	LOE & GOE	105,106
	PDD2	*SHH*	ZRS	Rare variant introducing a TFBS	GOE	107
	HPE	*SHH*	SBE2	Rare variant disrupting a TFBS	LOE	108
	Limb malformations	*PAX3, IHH, WNT6*	EPH4 enhancers	Deletions, duplications and inversions disrupt the boundaries of a TAD containing the *EPH4* enhancer and rewire the connectivity with different genes	GOE	13
	5q14.3 microdeletion syndrome	*MEF2C*	MEF2C enhancers	TAD disruption disconnects MEF2C from associated enhancer	LOE	109
	Pierre Robin syndrome	*SOX9*	SOX9 enhancer	A point mutation in a conserved enhancer disrupts the binding of MSX1	LOE	74
	Cooks syndrome	*SOX9, KCNJ2*	SOX9 enhancers	Duplication of a TAD boundary at the *SOX9* locus causes neo-TAD formation and *KCNJ2* misexpression	GOE	110
	Isolated atrial defect	*TBX5*	90 kb downstream	Rare variant abrogates heart-specific enhancer activity	LOE	111
	Isolated pancreatic agenesis	*PTF1A*	25 kb downstream	Rare variants abolish enhancer activity and disrupt the binding of FOXA2 and PDX1	LOE	41
Common (multifactorial)	Obesity	*IRX3, IRX5*	FTO intronic	Multiple variants on a common haplotype increase the activity of several enhancers	GOE	97
	Type 2 diabetes	*ZFAND3*	Upstream	SNP disrupts the binding of NeuroD1 and decreases enhancer activity	LOE	112
	Vascular diseases	*EDN1*	*PHACTR1* intronic	SNP located in a distal region interacting with *EDN1* enhancer	LOE	96
	HBF level	*BCL11A*	Downstream	SNP disrupts TF binding and diminishes expression in erythroid cells	LOE	113
	Cardiac disorders	*SNC5A*	*SNC10A* intronic	SNP in *SN10A* modulates *SNC5A* expression in the heart	LOE	114
	Hirschsprung disease	*RET*	Several enhancers	Several SNPs located in *RET* enhancers act synergistically to reduce gene expression	LOE	115
	Parkinson	*SNCA*	Intronic	SNP alters bthe inding of EMX2 and NKX6-1	LOE	40
Cancer	Burkitt lymphoma	*MYC*	IgH enhancer	Somatic translocation (enhancer hijacking)	GOE	116,117
	Lung Adenocarcinoma	*MYC*	450 kb downstream	Somatic duplication of the enhancer	GOE	118
	T-ALL	*TAL1*	7 kb upstream	Somatic insertions introduce a MYB binding site and induce the formation of a Neo-enhancer	GOE	119,120
	Ph-like ALL	*GATA3*	Intronic	A rare variant increases enhancer activity	GOE	121
	CLL	*AXIN2*	Upstream	Common variation in the *AXIN2* enhancer modulates CLL susceptibility *via* differential MEF2 binding	GOE	122
	AML	*GATA2, EVI1*	*GATA2* enhancer	Large somatic inversion relocated *GATA2* enhancer in the vicinity of *EVI1*	LOE & GOE	14
	Prostate cancer	*PCAT19, CEACAM21*	*PCAT19* Epromoter	Common variant changes the affinity of TFs and switch promoter and enhancer activities	GOE	123,124

Notes to Table 1: *Enhancer location relates to the regulated gene unless otherwise stated. Abbreviations: LOE: Loss-of-expression; GOE: gain-of-expression; LCR: Locus control Region; PDD2, preaxial polydactyly type II; HPE: holoprosencephaly; HBF: Fetal hemoglobin; T-ALL, T cell Acute Lymphoblastic Leukemia; CLL: Chronic Lymphoblastic Leukemia; MLL: Myeloid Lymphoblastic Leukemia; Ph-like ALL: Philadelphia chromosome-like Acute Lymphoblastic Leukemia; TF, transcription factor; IgH, immunoglobulin heavy chain

Figure legends

Figure 1: Different mechanisms of enhancer function and dysfunction. A) Variations in the interplay between enhancers and target genes. Multiple enhancers can cooperate in a tissue to increase the transcription of a target gene or be active in different tissues to control a complex developmental gene expression pattern. Enhancers can further control multiple genes in a mutually exclusive or shared way. Color code indicates the enhancer activity and gene expression in different tissues or developmental contexts. (**B-C**) Erroneous regulatory wiring between enhancers and genes, by either enhancer disruption **(B)** or altered enhancer-gene connectivity **(C),** can result in dysregulation of gene expression and ultimately cause disease. Enhancer dysfunction can originate from deletions, duplications and mutations, which can result in either loss or gain of gene expression. Altered enhancer-gene connectivity can be caused by chromosomal translocations or large structural variations that can distort or merge Topologically Associating Domains (TAD). As a consequence, enhancer-gene connectivity can be lost or gained resulting in dysregulated gene expression. Changes in gene expression are indicated by the number of arrows.

Figure 2: Challenges to unravel enhancer-associated diseases. Elucidating the molecular basis of enhancer dysfunction in disease requires critical areas of research to be addressed, each corresponding to one of the challenges described in the main text. Resolving challenges I to VI should lead to the ultimate challenge (VII) of identifying the causal variants, the impacted molecular mechanisms as well as the affected genes of a disease. TF: transcription factor. MPRA: Massively Parallel Reporter Assay.

Box 1. Critical areas of research to further our understanding of enhancer dysregulation in disease

We emphasize the following critical areas of research to advance our understanding of enhancer dysregulation in disease and for better translation of enhancer research into clinical practice:

Area 1

The ongoing community-driven comparison and assessment of experimental approaches for the discovery of enhancer activity should be strengthened. This will help improve our definition of the fundamental features of functionally operational enhancers as well as determine the most appropriate assay given a biological or disease context. We particularly

see the benefits of further developments of CRISPR screens that improve sensitivity and allow measurements in non-cultured cells.

Area 2

We foresee major benefits in further efforts towards developing assays that will allow accurate assessment of enhancer activity in single cells. scATAC-seq is key for expanding the enhancer map repertoire, particularly in rare cell types. In addition, further development of single-cell technologies that employ large-scale perturbations of enhancers or TFs will be key to assessing enhancer function for such cell types. The output of such studies will also help building models of enhancer regulation informed by the dynamics and stochasticity of regulatory events as well as discover mechanisms by which their perturbation contributes to pathology.

Area 3

To better understand the rules by which enhancers work together in regulatory domains to achieve robustness, specificity, or synergism, further efforts are needed to derive assays and strategies that allow combinatorial interference or perturbation of multiple enhancers. It is further imperative to develop in vivo (i.e. in situ) assays that allow the study of the activity of an enhancer in isolation or synergy with other enhancers. The outcomes of such studies would enable us to identify the biological mechanisms by which regulatory domains are formed and the rules by which TFs and the interplay between multiple regulatory elements yield robustness or additive effects. These insights will aid the interpretation of the potential effect and severity of regulatory genetic variants and enhancer dysfunction within complex regulatory domains.

Area 4

Experimental disease systems, such as humanized animal models, organoids, and engineered tissues, are becoming increasingly available for genetic engineering and *in situ* or *ex vivo* functional experiments. It will be key to fully employ these advanced disease models for assessing the functional and pathological consequences of non-coding regulatory variants (genetic and structural). Such experimental systems will allow

interrogation of enhancer activity under a relevant internal or external stimulus for their dynamic and contextual assessment.

Area 5

The research community should increase the already promising work towards developing interpretable and generalizable computational models that can accurately predict TF binding, the activity of enhancers and their target genes, using molecular measurements in any given cell type and condition. From these, the main efforts should ideally be focused on deriving the underlying regulatory DNA code, allowing for direct interpretation of the effects of genetic variants across cell types. Relatedly, we foresee great benefits in putting effort into developing approaches to computationally predict dosage-sensitive and responsive genes, as they are more likely to be adversely affected by cis-acting mutations.

Area 6

To fully understand the molecular basis of enhancer dysfunction, we foresee the need to further develop and apply large-scale TF perturbation assays coupled with GRN analysis to study cell type- or condition-specific TF regulons, and how they are defined by the combinatorial or cooperative binding grammar of enhancer sequences in normal and pathological conditions.

Area 7

Last but not least, we foresee great potential for implementing tools (e.g. CRISPR-based) to genetically or epigenetically modify the functions or chromatin contexts of enhancers to treat enhanceropathies. By targeting enhancers, one can avoid the potential pleiotropic effects associated with drugs/tools directed toward proteins or gene promoters.

Highlighted references

Maurano, M. T. et al. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. Science 337, 1190–1195 (2012).
This study provides evidence of the enrichment of GWAS variants in enhancers with tissue- and developmental-specific chromatin accessibility

Lettice, L. A. et al. Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. Proc. Natl. Acad. Sci. 99, 7548–7553 (2002).

This study links mutations of an enhancer to dysregulation of the SHH gene resulting in polydactyly

Lupiáñez, D. G. et al. Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. Cell 161, 1012–1025 (2015).
This study demonstrates links between disruptions of topologically associating domains and limb malformations

Thomas, H. F. *et al.* Temporal dissection of an enhancer cluster reveals distinct temporal and functional contributions of individual elements. *Mol. Cell* **81**, 969-982.e13 (2021).
The study provides evidence that enhancers with low intrinsic activity in episomal assays can collaborate in a highly additive fashion to induce gene expression at the endogenous locus

Soldner, F. *et al.* Parkinson-associated risk variant in distal enhancer of α-synuclein modulates target gene expression. *Nature* **533**, 95–99 (2016).
This is one of the first studies to functionally dissect the impact of a disease-associated genetic variant on enhancer activity

Nasser, J. *et al.* Genome-wide enhancer maps link risk variants to disease genes. *Nature* **593**, 238–243 (2021).
This study suggests a strategy to interpret the functions of GWAS variants based on inferred enhancer–gene maps across many cell types and tissues

Schwartzentruber, J. *et al.* Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk genes. *Nat. Genet.* **53**, 392–402 (2021).
This study identifies genetic variants associated with Alzheimer's disease to overlap enhancers specific to immune cells

Kvon, E. Z. et al. Comprehensive In Vivo Interrogation Reveals Phenotypic Impact of Human Enhancer Variants. Cell 180, 1262-1271.e15 (2020).
This study uses a high-throughput mouse reporter assay to demonstrate that a large majority of genetic variants linked to polydactyly lead to change in reporter gene expression

Hong, J.-W., Hendrix, D. A. & Levine, M. S. Shadow Enhancers as a Source of Evolutionary Novelty. Science 321, 1314–1314 (2008).
This study demonstrates that developmental genes can be regulated by multiple enhancers, which may provide robustness to enhancer deregulation

Hay, D. et al. Genetic dissection of the α-globin super-enhancer in vivo. Nat. Genet. 48, 895–903 (2016).
This study demonstrates that individual enhancers within the α-globin gene locus work independently and in an additive manner

Abell, N. S. *et al.* Multiple causal variants underlie genetic associations in humans. *Science* **375**, 1247–1254 (2022).

This study systematically assesses the effect of genetic variants on regulatory activity by massive parallel reporter assays leading to the identification of causal variants

Yan, J. et al. Systematic analysis of binding of transcription factors to noncoding variants. Nature 591, 147–151 (2021).
This study provides a systematic characterization of the relative affinity of transcription factors to non-coding genetic variants in vitro

Avsec, Ž. *et al.* Base-resolution models of transcription-factor binding reveal soft motif syntax. *Nat. Genet.* **53**, 354–366 (2021).
This study provides a deep learning framework to predict base-resolution profiles of pluripotency factors and interpretation methods to uncover their motifs and syntax

de Almeida, B. P., Reiter, F., Pagani, M. & Stark, A. DeepSTARR predicts enhancer activity from DNA sequence and enables the de novo design of enhancers. *Nat. Genet.* **54**, 613–624 (2022)
This study models the regulatory potential of DNA sequences using deep learning and derives transcription factor motifs and higher-order syntax rules determining enhancer activity

Smemo S, Tena JJ, Kim K-H, Gamazon ER, Sakabe NJ, Gómez-Marín C, Aneas I, Credidio FL, Sobreira DR, Wasserman NF, et al. 2014. Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature 507: 371–375.
This study demonstrates a mechanism of an FTO-associated variant linked to obesity through the derepression of an enhancer leading to increased expression of IRX3 and IRX5

References

1. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299–308 (1981).
2. Gasperini, M., Tome, J. M. & Shendure, J. Towards a comprehensive catalogue of validated and target-linked human enhancers. *Nat. Rev. Genet.* **21**, 292–310 (2020).
3. Claringbould, A. & Zaugg, J. B. Enhancers in disease: molecular basis and emerging treatment strategies. *Trends Mol. Med.* **27**, 1060–1073 (2021).
4. Rickels, R. & Shilatifard, A. Enhancer Logic and Mechanics in Development and Disease. *Trends Cell Biol.* **28**, 608–630 (2018).
5. Bradner, J. E., Hnisz, D. & Young, R. A. Transcriptional Addiction in Cancer. *Cell* **168**, 629–643 (2017).

6. Robson, M. I., Ringel, A. R. & Mundlos, S. Regulatory Landscaping: How Enhancer-Promoter Communication Is Sculpted in 3D. *Mol. Cell* **74**, 1110–1122 (2019).
7. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).
8. Grubert, F. *et al.* Genetic Control of Chromatin States in Humans Involves Local and Distal Chromosomal Interactions. *Cell* **162**, 1051–1065 (2015).
9. Maurano, M. T. *et al.* Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science* **337**, 1190–1195 (2012).
10. Waszak, S. M. *et al.* Population Variation and Genetic Control of Modular Chromatin Architecture in Humans. *Cell* **162**, 1039–1050 (2015).
11. Chen, C., Chang, I.-S., Hsiung, C. A. & Wasserman, W. W. On the identification of potential regulatory variants within genome wide association candidate SNP sets. *BMC Med. Genomics* **7**, 34 (2014).
12. Hnisz, D. *et al.* Super-Enhancers in the Control of Cell Identity and Disease. *Cell* **155**, 934–947 (2013).
13. Lupiáñez, D. G. *et al.* Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* **161**, 1012–1025 (2015).
14. Gröschel, S. *et al.* A Single Oncogenic Enhancer Rearrangement Causes Concomitant EVI1 and GATA2 Deregulation in Leukemia. *Cell* **157**, 369–381 (2014).
15. Laugsch, M. *et al.* Modeling the Pathological Long-Range Regulatory Effects of Human Structural Variation with Patient-Specific hiPSCs. *Cell Stem Cell* **24**, 736-752.e12 (2019).
16. Andersson, R. & Sandelin, A. Determinants of enhancer and promoter activities of regulatory elements. *Nat. Rev. Genet.* **21**, 71–87 (2020).
17. The ENCODE Project Consortium *et al.* Perspectives on ENCODE. *Nature* **583**, 693–698 (2020).
18. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).

19. Stunnenberg, H. G. *et al.* The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. *Cell* **167**, 1145–1149 (2016).
20. Rajagopal, N. *et al.* High-throughput mapping of regulatory DNA. *Nat. Biotechnol.* **34**, 167–174 (2016).
21. Korkmaz, G. *et al.* Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat. Biotechnol.* **34**, 192–198 (2016).
22. Gasperini, M. *et al.* A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. *Cell* **176**, 377-390.e19 (2019).
23. Hnisz, D. *et al.* Convergence of Developmental and Oncogenic Signaling Pathways at Transcriptional Super-Enhancers. *Mol. Cell* **58**, 362–370 (2015).
24. Thomas, H. F. *et al.* Temporal dissection of an enhancer cluster reveals distinct temporal and functional contributions of individual elements. *Mol. Cell* **81**, 969-982.e13 (2021).
25. Sahu, B. *et al.* Sequence determinants of human gene regulatory elements. *Nat. Genet.* **54**, 283–294 (2022).
26. Schwartzentruber, J. *et al.* Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk genes. *Nat. Genet.* **53**, 392–402 (2021).
27. Novikova, G. *et al.* Integration of Alzheimer's disease genetics and myeloid genomics identifies disease risk regulatory elements and genes. *Nat. Commun.* **12**, 1610 (2021).
28. Joslin, A. C. *et al.* A functional genomics pipeline identifies pleiotropy and cross-tissue effects within obesity-associated GWAS loci. *Nat. Commun.* **12**, 5253 (2021).
29. Nasser, J. *et al.* Genome-wide enhancer maps link risk variants to disease genes. *Nature* **593**, 238–243 (2021).
30. Soskic, B. *et al.* Chromatin activity at GWAS loci identifies T cell states driving complex immune diseases. *Nat. Genet.* **51**, 1486–1493 (2019).
31. Ota, M. *et al.* Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases. *Cell* **184**, 3006-3021.e17 (2021).

32. Young, A. M. H. *et al.* A map of transcriptional heterogeneity and regulatory variation in human microglia. *Nat. Genet.* **53**, 861–868 (2021).
33. Corces, M. R. *et al.* Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. *Nat. Genet.* **52**, 1158–1168 (2020).
34. Ulirsch, J. C. *et al.* Interrogation of human hematopoiesis at single-cell and single-variant resolution. *Nat. Genet.* **51**, 683–693 (2019).
35. Replogle, J. M. *et al.* Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell* S0092867422005979 (2022) doi:10.1016/j.cell.2022.05.013.
36. Schraivogel, D. *et al.* Targeted Perturb-seq enables genome-scale genetic screens in single cells. *Nat. Methods* **17**, 629–635 (2020).
37. Osterwalder, M. *et al.* Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* **554**, 239–243 (2018).
38. Aneas, I. *et al.* Asthma-associated genetic variants induce IL33 differential expression through an enhancer-blocking regulatory region. *Nat. Commun.* **12**, 6115 (2021).
39. Bhatia, S. *et al.* Quantitative spatial and temporal assessment of regulatory element activity in zebrafish. *eLife* **10**, e65601 (2021).
40. Soldner, F. *et al.* Parkinson-associated risk variant in distal enhancer of α-synuclein modulates target gene expression. *Nature* **533**, 95–99 (2016).
41. Weedon, M. N. *et al.* Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. *Nat. Genet.* **46**, 61–64 (2014).
42. Hasselmann, J. *et al.* Development of a Chimeric Model to Study and Manipulate Human Microglia In Vivo. *Neuron* **103**, 1016-1033.e10 (2019).
43. Mancuso, R. *et al.* Stem-cell-derived human microglia transplanted in mouse brain to study human disease. *Nat. Neurosci.* **22**, 2111–2116 (2019).
44. de Bruijn, S. E. *et al.* Structural Variants Create New Topological-Associated Domains and Ectopic Retinal Enhancer-Gene Contact in Dominant Retinitis Pigmentosa. *Am. J.*

Hum. Genet. **107**, 802–814 (2020).

45. Kvon, E. Z., Waymack, R., Gad, M. & Wunderlich, Z. Enhancer redundancy in development and disease. *Nat. Rev. Genet.* **22**, 324–336 (2021).

46. Kvon, E. Z. *et al.* Comprehensive In Vivo Interrogation Reveals Phenotypic Impact of Human Enhancer Variants. *Cell* **180**, 1262-1271.e15 (2020).

47. Sabarís, G., Laiker, I., Preger-Ben Noon, E. & Frankel, N. Actors with Multiple Roles: Pleiotropic Enhancers and the Paradigm of Enhancer Modularity. *Trends Genet.* **35**, 423–433 (2019).

48. Cao, K. *et al.* SET1A/COMPASS and shadow enhancers in the regulation of homeotic gene expression. *Genes Dev.* **31**, 787–801 (2017).

49. Hong, J.-W., Hendrix, D. A. & Levine, M. S. Shadow Enhancers as a Source of Evolutionary Novelty. *Science* **321**, 1314–1314 (2008).

50. Beagrie, R. A. *et al.* Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519–524 (2017).

51. Allahyar, A. *et al.* Enhancer hubs and loop collisions identified from single-allele topologies. *Nat. Genet.* **50**, 1151–1160 (2018).

52. Quinodoz, S. A. *et al.* Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* **174**, 744-757.e24 (2018).

53. Zheng, M. *et al.* Multiplex chromatin interactions with single-molecule precision. *Nature* **566**, 558–562 (2019).

54. Blobel, G. A., Higgs, D. R., Mitchell, J. A., Notani, D. & Young, R. A. Testing the super-enhancer concept. *Nat. Rev. Genet.* **22**, 749–755 (2021).

55. Madsen, J. G. S. *et al.* Highly interconnected enhancer communities control lineage-determining genes in human mesenchymal stem cells. *Nat. Genet.* **52**, 1227–1238 (2020).

56. Hay, D. *et al.* Genetic dissection of the α-globin super-enhancer in vivo. *Nat. Genet.* **48**, 895–903 (2016).

57. Sigalova, O. M., Shaeiri, A., Forneris, M., Furlong, E. E. & Zaugg, J. B. Predictive

features of gene expression variation reveal mechanistic link with differential expression. *Mol. Syst. Biol.* **16**, e9539 (2020).

58. Martinez-Ara, M., Comoglio, F., van Arensbergen, J. & van Steensel, B. Systematic analysis of intrinsic enhancer-promoter compatibility in the mouse genome. *Mol. Cell* **82**, 2519-2531.e6 (2022).

59. Bergman, D. T. *et al.* Compatibility rules of human enhancer and promoter sequences. *Nature* **607**, 176–184 (2022).

60. Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. & de Laat, W. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell* **10**, 1453–1465 (2002).

61. Krijger, P. H. L. & de Laat, W. Regulation of disease-associated gene expression in the 3D genome. *Nat. Rev. Mol. Cell Biol.* **17**, 771–782 (2016).

62. Forcato, M. *et al.* Comparison of computational methods for Hi-C data analysis. *Nat. Methods* **14**, 679–685 (2017).

63. Sahlén, P. *et al.* Chromatin interactions in differentiating keratinocytes reveal novel atopic dermatitis- and psoriasis-associated genes. *J. Allergy Clin. Immunol.* **147**, 1742–1752 (2021).

64. Dixon, A. L. *et al.* A genome-wide association study of global gene expression. *Nat. Genet.* **39**, 1202–1207 (2007).

65. Pliner, H. A. *et al.* Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. *Mol. Cell* **71**, 858-871.e8 (2018).

66. Rennie, S. *et al.* Transcription start site analysis reveals widespread divergent transcription in D. melanogaster and core promoter-encoded enhancer activities. *Nucleic Acids Res.* **39**, 311 (2018).

67. Reyes-Palomares, A. *et al.* Remodeling of active endothelial enhancers is associated with aberrant gene-regulatory networks in pulmonary arterial hypertension. *Nat. Commun.* **11**, 1673 (2020).

68. Fulco, C. P. *et al.* Activity-by-contact model of enhancer–promoter regulation from

thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).

69. Huang, Q. Q., Ritchie, S. C., Brozynska, M. & Inouye, M. Power, false discovery rate and Winner's Curse in eQTL studies. *Nucleic Acids Res.* **46**, e133 (2018).

70. Gazal, S. *et al.* Combining SNP-to-gene linking strategies to identify disease genes and assess disease omnigenicity. *Nat. Genet.* **54**, 827–836 (2022).

71. Avsec, Ž. *et al.* Effective gene expression prediction from sequence by integrating long-range interactions. *Nat. Methods* **18**, 1196–1203 (2021).

72. He, Y. *et al.* sn-spMF: matrix factorization informs tissue-specific genetic regulation of gene expression. *Genome Biol.* **21**, 235 (2020).

73. Deplancke, B., Alpern, D. & Gardeux, V. The Genetics of Transcription Factor DNA Binding Variation. *Cell* **166**, 538–554 (2016).

74. Benko, S. *et al.* Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. *Nat. Genet.* **41**, 359–364 (2009).

75. Delaneau, O. *et al.* Chromatin three-dimensional interactions mediate genetic effects on gene expression. *Science* **364**, eaat8266 (2019).

76. Avsec, Ž. *et al.* Base-resolution models of transcription-factor binding reveal soft motif syntax. *Nat. Genet.* **53**, 354–366 (2021).

77. de Almeida, B. P., Reiter, F., Pagani, M. & Stark, A. DeepSTARR predicts enhancer activity from DNA sequence and enables the de novo design of synthetic enhancers. *Nat. Genet.* **54**, 613–624 (2022).

78. Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650–665 (2018).

79. Kelley, D. R. Cross-species regulatory sequence activity prediction. *PLOS Comput. Biol.* **16**, e1008050 (2020).

80. Maslova, A. *et al.* Deep learning of immune cell differentiation. *Proc. Natl. Acad. Sci.* **117**, 25655–25666 (2020).

81. Janssens, J. *et al.* Decoding gene regulation in the fly brain. *Nature* **601**, 630–636 (2022).

82. Abell, N. S. *et al.* Multiple causal variants underlie genetic associations in humans.

650 *Science* **375**, 1247–1254 (2022).
651 83. van Arensbergen, J. *et al.* High-throughput identification of human SNPs affecting
652 regulatory element activity. *Nat. Genet.* **51**, 1160–1169 (2019).
653 84. Tewhey, R. *et al.* Direct Identification of Hundreds of Expression-Modulating Variants
654 using a Multiplexed Reporter Assay. *Cell* **165**, 1519–1529 (2016).
655 85. Ulirsch, J. C. *et al.* Systematic Functional Dissection of Common Genetic Variation
656 Affecting Red Blood Cell Traits. *Cell* **165**, 1530–1545 (2016).
657 86. Bourges, C. *et al.* Resolving mechanisms of immune□mediated disease in primary CD 4
658 T cells. *EMBO Mol. Med.* **12**, (2020).
659 87. Yan, J. *et al.* Systematic analysis of binding of transcription factors to noncoding
660 variants. *Nature* **591**, 147–151 (2021).
661 88. Danek, P. *et al.* β-Catenin–TCF/LEF signaling promotes steady-state and emergency
662 granulopoiesis via G-CSF receptor upregulation. *Blood* **136**, 2574–2587 (2020).
663 89. Ma, S. *et al.* Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and
664 Chromatin. *Cell* **183**, 1103-1116.e20 (2020).
665 90. Glass, C. K. & Natoli, G. Molecular control of activation and priming in macrophages.
666 *Nat. Immunol.* **17**, 26–33 (2016).
667 91. Ghisletti, S. *et al.* Identification and characterization of enhancers controlling the
668 inflammatory gene expression program in macrophages. *Immunity* **32**, 317–328 (2010).
669 92. Bunina, D. *et al.* Genomic Rewiring of SOX2 Chromatin Interaction Network during
670 Differentiation of ESCs to Postmitotic Neurons. *Cell Syst.* **10**, 480-494.e8 (2020).
671 93. Siersbæk, R. *et al.* Transcription factor cooperativity in early adipogenic hotspots and
672 super-enhancers. *Cell Rep.* **7**, 1443–1455 (2014).
673 94. Weidemüller, P., Kholmatov, M., Petsalaki, E. & Zaugg, J. B. Transcription factors:
674 Bridge between cell signaling and gene regulation. *PROTEOMICS* **21**, 2000034 (2021).
675 95. Telenti, A. *et al.* Deep sequencing of 10,000 human genomes. *Proc. Natl. Acad. Sci. U.*
676 *S. A.* **113**, 11901–11906 (2016).
677 96. Gupta, R. M. *et al.* A Genetic Variant Associated with Five Vascular Diseases Is a

Distal Regulator of Endothelin-1 Gene Expression. *Cell* **170**, 522-533.e15 (2017).

97. Smemo, S. *et al.* Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* **507**, 371–375 (2014).

98. Mills, M. C. & Rahal, C. The GWAS Diversity Monitor tracks diversity by disease in real time. *Nat. Genet.* **52**, 242–243 (2020).

99. Polygenic Risk Score Task Force of the International Common Disease Alliance. Responsible use of polygenic risk scores in the clinic: potential benefits, risks and gaps. *Nat. Med.* **27**, 1876–1884 (2021).

100. Mohammadi, P. *et al.* Genetic regulatory variation in populations informs transcriptome analysis in rare disease. *Science* **366**, 351–356 (2019).

101. Tanjo, T., Kawai, Y., Tokunaga, K., Ogasawara, O. & Nagasaki, M. Practical guide for managing large-scale human genome data in research. *J. Hum. Genet.* **66**, 39–52 (2021).

102. The 100,000 Genomes Project Pilot Investigators *et al.* 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care — Preliminary Report. *N. Engl. J. Med.* **385**, 1868–1880 (2021).

103. Kioussis, D., Vanin, E., deLange, T., Flavell, R. A. & Grosveld, F. G. β-Globin gene inactivation by DNA translocation in γβ-thalassaemi. *Nature* **306**, 662–666 (1983).

104. Driscoll, M. C., Dobkin, C. S. & Alter, B. P. Gamma delta beta-thalassemia due to a de novo mutation deleting the 5’ beta-globin gene activation-region hypersensitive sites. *Proc. Natl. Acad. Sci.* **86**, 7470–7474 (1989).

105. De Gobbi, M. *et al.* A Regulatory SNP Causes a Human Genetic Disease by Creating a New Transcriptional Promoter. *Science* **312**, 1215–1217 (2006).

106. Lower, K. M. *et al.* Adventitious changes in long-range gene expression caused by polymorphic structural variation and promoter competition. *Proc. Natl. Acad. Sci.* **106**, 21771–21776 (2009).

107. Lettice, L. A. et al. Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. Proc. Natl. Acad. Sci. 99, 7548–7553 (2002).

706 108. Jeong, Y., El-Jaick, K., Roessler, E., Muenke, M. & Epstein, D. J. A functional screen
707 for sonic hedgehog regulatory elements across a 1 Mb interval identifies long-range
708 ventral forebrain enhancers. *Development* **133**, 761–772 (2006).
709 109. Redin, C. *et al.* The genomic landscape of balanced cytogenetic abnormalities
710 associated with human congenital anomalies. *Nat. Genet.* **49**, 36–45 (2017).
711 110. Franke, M. *et al.* Formation of new chromatin domains determines pathogenicity of
712 genomic duplications. *Nature* **538**, 265–269 (2016).
713 111. Smemo, S. *et al.* Regulatory variation in a TBX5 enhancer leads to isolated congenital
714 heart disease. *Hum. Mol. Genet.* **21**, 3255–3263 (2012).
715 112. Pasquali, L. *et al.* Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-
716 associated variants. *Nat. Genet.* **46**, 136–143 (2014).
717 113. Bauer, D. E. *et al.* An Erythroid Enhancer of *BCL11A* Subject to Genetic Variation
718 Determines Fetal Hemoglobin Level. *Science* **342**, 253–257 (2013).
719 114. van den Boogaard, M. *et al.* A common genetic variant within SCN10A modulates
720 cardiac SCN5A expression. *J. Clin. Invest.* **124**, 1844–1852 (2014).
721 115. Chatterjee, S. *et al.* Enhancer Variants Synergistically Drive Dysfunction of a Gene
722 Regulatory Network In Hirschsprung Disease. *Cell* **167**, 355-368.e10 (2016).
723 116. Dalla-Favera, R. *et al.* Human c-myc onc gene is located on the region of chromosome
724 8 that is translocated in Burkitt lymphoma cells. *Proc. Natl. Acad. Sci.* **79**, 7824–7827
725 (1982).
726 117. Taub, R. *et al.* Translocation of the c-myc gene into the immunoglobulin heavy chain
727 locus in human Burkitt lymphoma and murine plasmacytoma cells. *Proc. Natl. Acad.*
728 *Sci.* **79**, 7837–7841 (1982).
729 118. Zhang, X. *et al.* Identification of focally amplified lineage-specific super-enhancers in
730 human epithelial cancers. *Nat. Genet.* **48**, 176–182 (2016).
731 119. Mansour, M. R. *et al.* An oncogenic super-enhancer formed through somatic mutation
732 of a noncoding intergenic element. *Science* **346**, 1373–1377 (2014).
733 120. Navarro, J.-M. *et al.* Site- and allele-specific polycomb dysregulation in T-cell

leukaemia. *Nat. Commun.* **6**, 6094 (2015).

121. Yang, H. *et al.* Noncoding genetic variation in GATA3 increases acute lymphoblastic leukemia risk through local and global changes in chromatin conformation. *Nat. Genet.* **54**, 170–179 (2022).

122. Llimos, G. *et al.* A leukemia-protective germline variant mediates chromatin module formation via transcription factor nucleation. *Nat. Commun.* **13**, 2042 (2022).

123. Gao, P. *et al.* Biology and Clinical Implications of the 19q13 Aggressive Prostate Cancer Susceptibility Locus. *Cell* **174**, 576-589.e18 (2018).

124. Hua, J. T. *et al.* Risk SNP-Mediated Promoter-Enhancer Switching Drives Prostate Cancer through lncRNA PCAT19. *Cell* **174**, 564-575.e18 (2018).

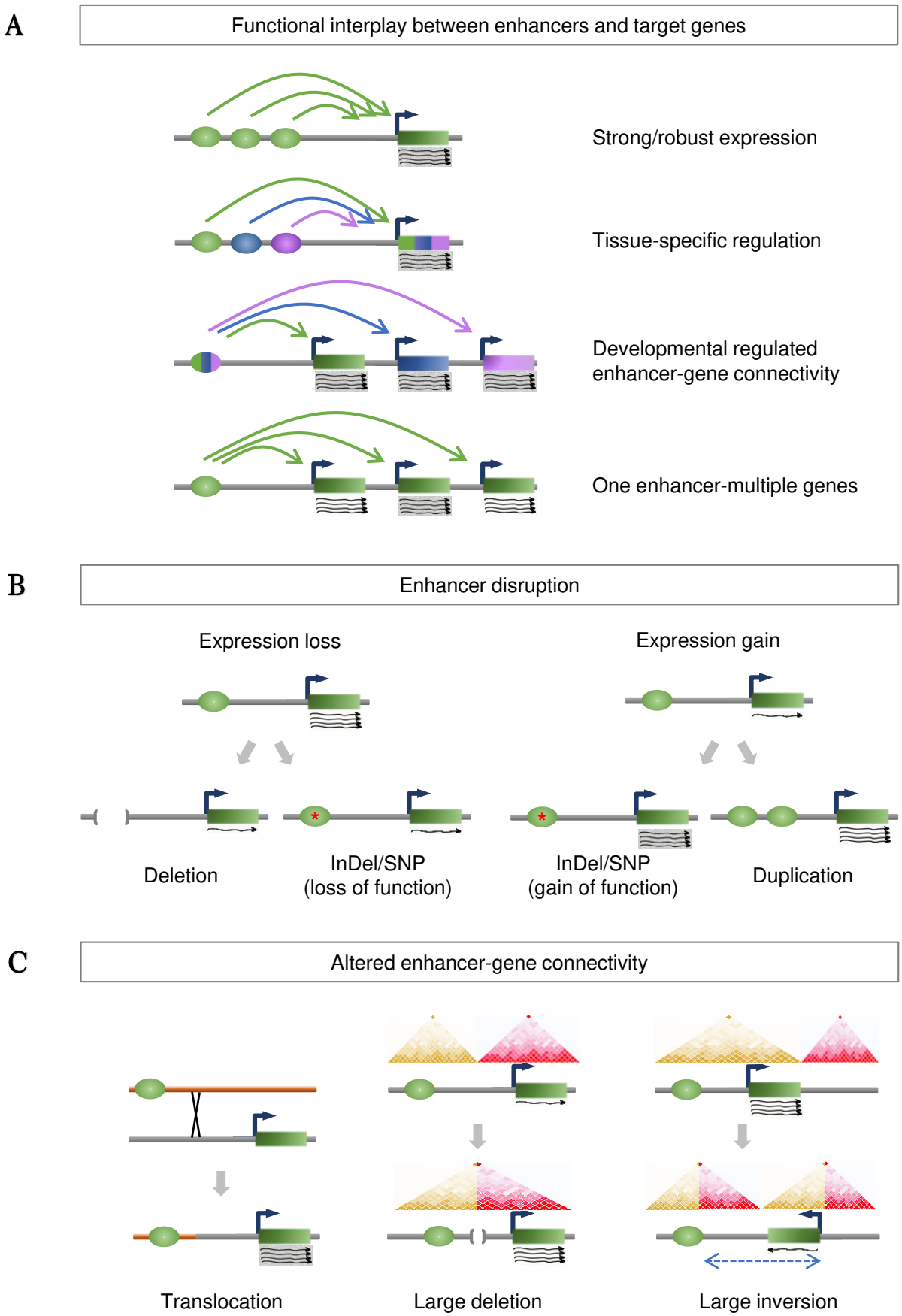

Figure 1 (REVISED)

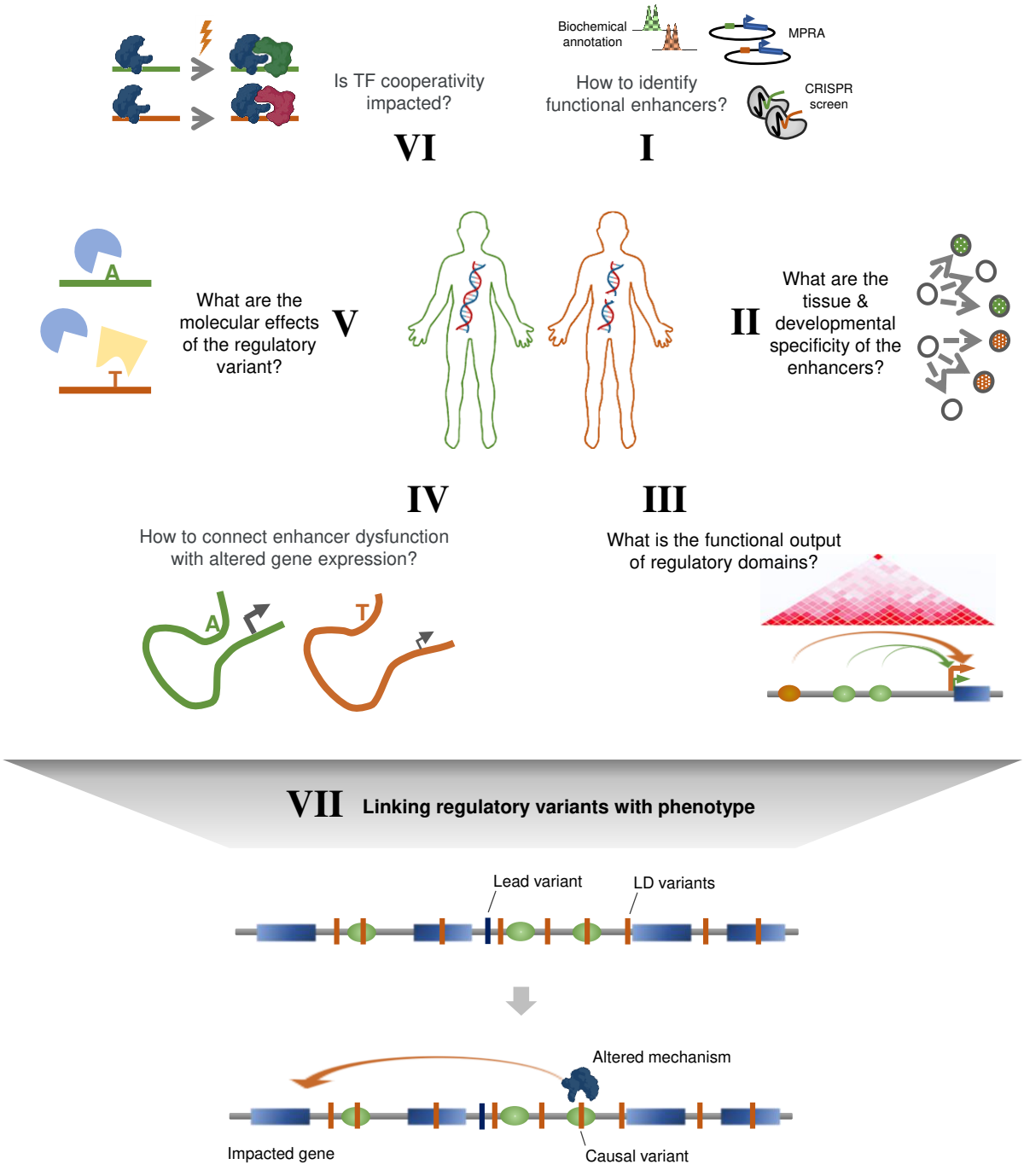
Is TF cooperativity impacted?
VI
Biochemical annotation
MPRA
How to identify functional enhancers?
CRISPR screen
I
A
T
What are the molecular effects of the regulatory variant?
V
II
What are the tissue & developmental specificity of the enhancers?
IV
How to connect enhancer dysfunction with altered gene expression?
A
T
III
What is the functional output of regulatory domains?
VII
Linking regulatory variants with phenotype
Lead variant
LD variants
Altered mechanism
Impacted gene
Causal variant

Figure 2