



**HAL**  
open science

## Current challenges in understanding the role of enhancers in disease

Judith Barbara Zaugg, Pelin Sahlén, Robin Andersson, Meritxell Alberich-Jorda, Wouter de Laat, Bart Deplancke, Jorge Ferrer, Susanne Mandrup, Gioacchino Natoli, Dariusz Plewczynski, et al.

► **To cite this version:**

Judith Barbara Zaugg, Pelin Sahlén, Robin Andersson, Meritxell Alberich-Jorda, Wouter de Laat, et al.. Current challenges in understanding the role of enhancers in disease. *Nature Structural and Molecular Biology*, 2022, 29 (12), pp.1148-1158. 10.1038/s41594-022-00896-3 . hal-03934940

**HAL Id: hal-03934940**

**<https://amu.hal.science/hal-03934940>**

Submitted on 13 Jan 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# 1 **Current challenges in understanding the role of enhancers in disease**

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

5 <sup>1</sup> European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117, Heidelberg

6 <sup>2</sup> Science for Life Laboratory, KTH - Royal Institute of Technology, Tomtebodavägen 23A,  
7 17165, Solna, Sweden

8 <sup>3</sup> Section for Computational and RNA Biology, Department of Biology, University of  
9 Copenhagen, 2200 Copenhagen N, Denmark

10 <sup>4</sup> The Novo Nordisk Foundation Center for Genomic Mechanisms of Disease, Broad Institute  
11 of MIT and Harvard, Cambridge, MA 02142, USA

12 <sup>5</sup> Department of Hemato-oncology, Institute of Molecular Genetics of the CAS, Vídeňská 1083,  
13 14200 Prague 4, Czech Republic

14 <sup>6</sup> Childhood Leukaemia Investigation Prague, Department of Pediatric Haematology and Oncology,  
15 2nd Faculty of Medicine, Charles University in Prague, University Hospital Motol, V Uvalu 84,  
16 Praha 150 06, Czech Republic

17 <sup>7</sup> Hubrecht Institute-KNAW, Oncode Institute and University Medical Center Utrecht, 3584 CT  
18 Utrecht, The Netherlands

19 <sup>8</sup> Laboratory of Systems Biology and Genetics, Institute of Bio-engineering, School of Life  
20 Sciences, EPFL, CH-1015 Lausanne, Switzerland

21 <sup>9</sup> Centre for Genomic Regulation, the Barcelona Institute of Science and Technology,  
22 Barcelona 08003, Spain

23 <sup>10</sup> Centro de Investigación Biomédica en red Diabetes y enfermedades metabólicas asociadas  
24 (CIBERDEM), Madrid 28029, Spain

25 <sup>11</sup> Department of Metabolism, Digestion and Reproduction, Imperial College London, London,  
26 UK

27 <sup>12</sup> Functional Genomics and Metabolism Research Unit, Department of Biochemistry and  
28 Molecular Biology, University of Southern Denmark, 5230 Odense M, Denmark

29 <sup>13</sup> IEO, European Institute of Oncology - IRCCS, Milan, Italy

30 <sup>14</sup> Laboratory of Functional and Structural Genomics, Centre of New Technologies, University  
31 of Warsaw, Banacha 2c, 02-097 Warsaw, Poland

32 <sup>15</sup> Institute of Biomedicine and Biotechnology of Cantabria (IBBTEC), CSIC/University of  
33 Cantabria, Albert Einstein 22, 39011, Santander, Spain

34 <sup>16</sup> TAGC, UMR 1090, Aix-Marseille University, INSERM, Marseille, France

35 \* These authors contributed equally

36 ‡ To whom correspondence should be addressed, email: [salvatore.spicuglia@inserm.fr](mailto:salvatore.spicuglia@inserm.fr)

37

## 38 **Preface**

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

50

51 **Main text**

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

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

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

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

83

#### 84 **Current challenges in characterizing enhancer dysfunction in disease**

85

##### 86 ***1) Identifying and validating functional enhancers***

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

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

122

## 123 ***2) Defining the spatiotemporal window in which a regulatory variant affects enhancer*** 124 ***activity***

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

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

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

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

157

### 158 **3) Understanding the interplay between cis-regulatory elements**

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

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

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

184

#### 185 ***4) Identifying the target genes of enhancers***

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



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

227

## 228 ***5) Understanding the grammar of enhancer activity***

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

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

257

## 258 **6) Understanding how TF cooperation defines enhancer activity and specificity**

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

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

286

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

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

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

321

## 322 **Future directions**

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

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

344

#### 345 **Acknowledgments**

346 All authors are members of the EU-funded Innovative Training Network “Molecular Basis of  
347 Human Enhanceropathies” (Enhpathy, [www.enhpathy.eu](http://www.enhpathy.eu)) and received funding from the  
348 Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant  
349 agreement No 860002. The Enhpathy network has set up a multidisciplinary scientific  
350 consortium in the continuum of basic, translational and clinical research on enhancers and  
351 associated diseases.

352

#### 353 **Competing interests**

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

356 to Table 1: \*Enhancer location relates to the regulated gene unless otherwise stated. Abbreviations: LOE: Loss-of-expression; GOE: gain-  
357 expression; LCR: Locus control Region; PDD2, preaxial polydactyly type II; HPE: holoprosencephaly; HBF: Fetal hemoglobin; T-ALL, T  
358 Acute Lymphoblastic Leukemia; CLL: Chronic Lymphoblastic Leukemia; MLL: Myeloid Lymphoblastic Leukemia; Ph-like ALL:  
359 Philadelphia chromosome-like Acute Lymphoblastic Leukemia; TF, transcription factor; IgH, immunoglobulin heavy chain  
360

361

362

363

363 **Figure legends**

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

378

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

385

386

**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.

387

#### **388 Highlighted references**

389 Maurano, M. T. et al. Systematic Localization of Common Disease-Associated Variation in  
390 Regulatory DNA. *Science* 337, 1190–1195 (2012).

391 **This study provides evidence of the enrichment of GWAS variants in enhancers with**  
392 **tissue- and developmental-specific chromatin accessibility**

393

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

396 **This study links mutations of an enhancer to dysregulation of the SHH gene resulting**  
397 **in polydactyly**  
398  
399 Lupiáñez, D. G. et al. Disruptions of Topological Chromatin Domains Cause Pathogenic  
400 Rewiring of Gene-Enhancer Interactions. *Cell* 161, 1012–1025 (2015).  
401 **This study demonstrates links between disruptions of topologically associating**  
402 **domains and limb malformations**  
403  
404 Thomas, H. F. et al. Temporal dissection of an enhancer cluster reveals distinct temporal  
405 and functional contributions of individual elements. *Mol. Cell* 81, 969-982.e13 (2021).  
406 **The study provides evidence that enhancers with low intrinsic activity in episomal**  
407 **assays can collaborate in a highly additive fashion to induce gene expression at the**  
408 **endogenous locus**  
409  
410 Soldner, F. et al. Parkinson-associated risk variant in distal enhancer of  $\alpha$ -synuclein  
411 modulates target gene expression. *Nature* 533, 95–99 (2016).  
412 **This is one of the first studies to functionally dissect the impact of a**  
413 **disease-associated genetic variant on enhancer activity**  
414  
415 Nasser, J. et al. Genome-wide enhancer maps link risk variants to disease genes. *Nature*  
416 593, 238–243 (2021).  
417 **This study suggests a strategy to interpret the functions of GWAS variants based on**  
418 **inferred enhancer–gene maps across many cell types and tissues**  
419  
420 Schwartzenuber, J. et al. Genome-wide meta-analysis, fine-mapping and integrative  
421 prioritization implicate new Alzheimer's disease risk genes. *Nat. Genet.* 53, 392–402 (2021).  
422 **This study identifies genetic variants associated with Alzheimer's disease to overlap**  
423 **enhancers specific to immune cells**  
424  
425 Kvon, E. Z. et al. Comprehensive In Vivo Interrogation Reveals Phenotypic Impact of Human  
426 Enhancer Variants. *Cell* 180, 1262-1271.e15 (2020).  
427 **This study uses a high-throughput mouse reporter assay to demonstrate that a large**  
428 **majority of genetic variants linked to polydactyly lead to change in reporter gene**  
429 **expression**  
430  
431 Hong, J.-W., Hendrix, D. A. & Levine, M. S. Shadow Enhancers as a Source of Evolutionary  
432 Novelty. *Science* 321, 1314–1314 (2008).  
433 **This study demonstrates that developmental genes can be regulated by multiple**  
434 **enhancers, which may provide robustness to enhancer deregulation**  
435  
436 Hay, D. et al. Genetic dissection of the  $\alpha$ -globin super-enhancer in vivo. *Nat. Genet.* 48,  
437 895–903 (2016).  
438 **This study demonstrates that individual enhancers within the  $\alpha$ -globin gene locus**  
439 **work independently and in an additive manner**  
440  
441 Abell, N. S. et al. Multiple causal variants underlie genetic associations in humans. *Science*  
442 375, 1247–1254 (2022).

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

446

447 *Yan, J. et al. Systematic analysis of binding of transcription factors to noncoding variants.*  
448 *Nature* 591, 147–151 (2021).

449 **This study provides a systematic characterization of the relative affinity of**  
450 **transcription factors to non-coding genetic variants in vitro**

451

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

454 **This study provides a deep learning framework to predict base-resolution profiles of**  
455 **pluripotency factors and interpretation methods to uncover their motifs and syntax**

456

457 *de Almeida, B. P., Reiter, F., Pagani, M. & Stark, A. DeepSTARR predicts enhancer activity*  
458 *from DNA sequence and enables the de novo design of enhancers.* *Nat. Genet.* **54**, 613–624  
459 (2022)

460 **This study models the regulatory potential of DNA sequences using deep learning**  
461 **and derives transcription factor motifs and higher-order syntax rules determining**  
462 **enhancer activity**

463

464 *Smemo S, Tena JJ, Kim K-H, Gamazon ER, Sakabe NJ, Gómez-Marín C, Aneas I, Credidio*  
465 *FL, Sobreira DR, Wasserman NF, et al. 2014. Obesity-associated variants within FTO form*  
466 *long-range functional connections with IRX3.* *Nature* 507: 371–375.

467 **This study demonstrates a mechanism of an FTO-associated variant linked to obesity**  
468 **through the derepression of an enhancer leading to increased expression of IRX3 and**  
469 **IRX5**

470

## 471 **References**

472 1. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a  $\beta$ -globin gene is enhanced by  
473 remote SV40 DNA sequences. *Cell* **27**, 299–308 (1981).

474 2. Gasperini, M., Tome, J. M. & Shendure, J. Towards a comprehensive catalogue of  
475 validated and target-linked human enhancers. *Nat. Rev. Genet.* **21**, 292–310 (2020).

476 3. Claringbould, A. & Zaugg, J. B. Enhancers in disease: molecular basis and emerging  
477 treatment strategies. *Trends Mol. Med.* **27**, 1060–1073 (2021).

478 4. Rickels, R. & Shilatifard, A. Enhancer Logic and Mechanics in Development and  
479 Disease. *Trends Cell Biol.* **28**, 608–630 (2018).

480 5. Bradner, J. E., Hnisz, D. & Young, R. A. Transcriptional Addiction in Cancer. *Cell* **168**,  
481 629–643 (2017).

- 482 6. Robson, M. I., Ringel, A. R. & Mundlos, S. Regulatory Landscaping: How Enhancer-  
483 Promoter Communication Is Sculpted in 3D. *Mol. Cell* **74**, 1110–1122 (2019).
- 484 7. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues.  
485 *Nature* **507**, 455–461 (2014).
- 486 8. Grubert, F. *et al.* Genetic Control of Chromatin States in Humans Involves Local and  
487 Distal Chromosomal Interactions. *Cell* **162**, 1051–1065 (2015).
- 488 9. Maurano, M. T. *et al.* Systematic Localization of Common Disease-Associated Variation  
489 in Regulatory DNA. *Science* **337**, 1190–1195 (2012).
- 490 10. Waszak, S. M. *et al.* Population Variation and Genetic Control of Modular Chromatin  
491 Architecture in Humans. *Cell* **162**, 1039–1050 (2015).
- 492 11. Chen, C., Chang, I.-S., Hsiung, C. A. & Wasserman, W. W. On the identification of  
493 potential regulatory variants within genome wide association candidate SNP sets. *BMC*  
494 *Med. Genomics* **7**, 34 (2014).
- 495 12. Hnisz, D. *et al.* Super-Enhancers in the Control of Cell Identity and Disease. *Cell* **155**,  
496 934–947 (2013).
- 497 13. Lupiáñez, D. G. *et al.* Disruptions of Topological Chromatin Domains Cause Pathogenic  
498 Rewiring of Gene-Enhancer Interactions. *Cell* **161**, 1012–1025 (2015).
- 499 14. Gröschel, S. *et al.* A Single Oncogenic Enhancer Rearrangement Causes Concomitant  
500 EVI1 and GATA2 Deregulation in Leukemia. *Cell* **157**, 369–381 (2014).
- 501 15. Laugsch, M. *et al.* Modeling the Pathological Long-Range Regulatory Effects of Human  
502 Structural Variation with Patient-Specific hiPSCs. *Cell Stem Cell* **24**, 736-752.e12  
503 (2019).
- 504 16. Andersson, R. & Sandelin, A. Determinants of enhancer and promoter activities of  
505 regulatory elements. *Nat. Rev. Genet.* **21**, 71–87 (2020).
- 506 17. The ENCODE Project Consortium *et al.* Perspectives on ENCODE. *Nature* **583**, 693–  
507 698 (2020).
- 508 18. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human  
509 epigenomes. *Nature* **518**, 317–330 (2015).

- 510 19. Stunnenberg, H. G. *et al.* The International Human Epigenome Consortium: A Blueprint  
511 for Scientific Collaboration and Discovery. *Cell* **167**, 1145–1149 (2016).
- 512 20. Rajagopal, N. *et al.* High-throughput mapping of regulatory DNA. *Nat. Biotechnol.* **34**,  
513 167–174 (2016).
- 514 21. Korkmaz, G. *et al.* Functional genetic screens for enhancer elements in the human  
515 genome using CRISPR-Cas9. *Nat. Biotechnol.* **34**, 192–198 (2016).
- 516 22. Gasperini, M. *et al.* A Genome-wide Framework for Mapping Gene Regulation via  
517 Cellular Genetic Screens. *Cell* **176**, 377-390.e19 (2019).
- 518 23. Hnisz, D. *et al.* Convergence of Developmental and Oncogenic Signaling Pathways at  
519 Transcriptional Super-Enhancers. *Mol. Cell* **58**, 362–370 (2015).
- 520 24. Thomas, H. F. *et al.* Temporal dissection of an enhancer cluster reveals distinct  
521 temporal and functional contributions of individual elements. *Mol. Cell* **81**, 969-982.e13  
522 (2021).
- 523 25. Sahu, B. *et al.* Sequence determinants of human gene regulatory elements. *Nat.*  
524 *Genet.* **54**, 283–294 (2022).
- 525 26. Schwartzenuber, J. *et al.* Genome-wide meta-analysis, fine-mapping and integrative  
526 prioritization implicate new Alzheimer's disease risk genes. *Nat. Genet.* **53**, 392–402  
527 (2021).
- 528 27. Novikova, G. *et al.* Integration of Alzheimer's disease genetics and myeloid genomics  
529 identifies disease risk regulatory elements and genes. *Nat. Commun.* **12**, 1610 (2021).
- 530 28. Joslin, A. C. *et al.* A functional genomics pipeline identifies pleiotropy and cross-tissue  
531 effects within obesity-associated GWAS loci. *Nat. Commun.* **12**, 5253 (2021).
- 532 29. Nasser, J. *et al.* Genome-wide enhancer maps link risk variants to disease genes.  
533 *Nature* **593**, 238–243 (2021).
- 534 30. Soskic, B. *et al.* Chromatin activity at GWAS loci identifies T cell states driving complex  
535 immune diseases. *Nat. Genet.* **51**, 1486–1493 (2019).
- 536 31. Ota, M. *et al.* Dynamic landscape of immune cell-specific gene regulation in immune-  
537 mediated diseases. *Cell* **184**, 3006-3021.e17 (2021).

- 538 32. Young, A. M. H. *et al.* A map of transcriptional heterogeneity and regulatory variation in  
539 human microglia. *Nat. Genet.* **53**, 861–868 (2021).
- 540 33. Corces, M. R. *et al.* Single-cell epigenomic analyses implicate candidate causal  
541 variants at inherited risk loci for Alzheimer’s and Parkinson’s diseases. *Nat. Genet.* **52**,  
542 1158–1168 (2020).
- 543 34. Ulirsch, J. C. *et al.* Interrogation of human hematopoiesis at single-cell and single-  
544 variant resolution. *Nat. Genet.* **51**, 683–693 (2019).
- 545 35. Replogle, J. M. *et al.* Mapping information-rich genotype-phenotype landscapes with  
546 genome-scale Perturb-seq. *Cell* S0092867422005979 (2022)  
547 doi:10.1016/j.cell.2022.05.013.
- 548 36. Schraivogel, D. *et al.* Targeted Perturb-seq enables genome-scale genetic screens in  
549 single cells. *Nat. Methods* **17**, 629–635 (2020).
- 550 37. Osterwalder, M. *et al.* Enhancer redundancy provides phenotypic robustness in  
551 mammalian development. *Nature* **554**, 239–243 (2018).
- 552 38. Aneas, I. *et al.* Asthma-associated genetic variants induce IL33 differential expression  
553 through an enhancer-blocking regulatory region. *Nat. Commun.* **12**, 6115 (2021).
- 554 39. Bhatia, S. *et al.* Quantitative spatial and temporal assessment of regulatory element  
555 activity in zebrafish. *eLife* **10**, e65601 (2021).
- 556 40. Soldner, F. *et al.* Parkinson-associated risk variant in distal enhancer of  $\alpha$ -synuclein  
557 modulates target gene expression. *Nature* **533**, 95–99 (2016).
- 558 41. Weedon, M. N. *et al.* Recessive mutations in a distal PTF1A enhancer cause isolated  
559 pancreatic agenesis. *Nat. Genet.* **46**, 61–64 (2014).
- 560 42. Hasselmann, J. *et al.* Development of a Chimeric Model to Study and Manipulate  
561 Human Microglia In Vivo. *Neuron* **103**, 1016-1033.e10 (2019).
- 562 43. Mancuso, R. *et al.* Stem-cell-derived human microglia transplanted in mouse brain to  
563 study human disease. *Nat. Neurosci.* **22**, 2111–2116 (2019).
- 564 44. de Bruijn, S. E. *et al.* Structural Variants Create New Topological-Associated Domains  
565 and Ectopic Retinal Enhancer-Gene Contact in Dominant Retinitis Pigmentosa. *Am. J.*

- 566 *Hum. Genet.* **107**, 802–814 (2020).
- 567 45. Kvon, E. Z., Waymack, R., Gad, M. & Wunderlich, Z. Enhancer redundancy in  
568 development and disease. *Nat. Rev. Genet.* **22**, 324–336 (2021).
- 569 46. Kvon, E. Z. *et al.* Comprehensive In Vivo Interrogation Reveals Phenotypic Impact of  
570 Human Enhancer Variants. *Cell* **180**, 1262-1271.e15 (2020).
- 571 47. Sabarís, G., Laiker, I., Preger-Ben Noon, E. & Frankel, N. Actors with Multiple Roles:  
572 Pleiotropic Enhancers and the Paradigm of Enhancer Modularity. *Trends Genet.* **35**,  
573 423–433 (2019).
- 574 48. Cao, K. *et al.* SET1A/COMPASS and shadow enhancers in the regulation of homeotic  
575 gene expression. *Genes Dev.* **31**, 787–801 (2017).
- 576 49. Hong, J.-W., Hendrix, D. A. & Levine, M. S. Shadow Enhancers as a Source of  
577 Evolutionary Novelty. *Science* **321**, 1314–1314 (2008).
- 578 50. Beagrie, R. A. *et al.* Complex multi-enhancer contacts captured by genome architecture  
579 mapping. *Nature* **543**, 519–524 (2017).
- 580 51. Allahyar, A. *et al.* Enhancer hubs and loop collisions identified from single-allele  
581 topologies. *Nat. Genet.* **50**, 1151–1160 (2018).
- 582 52. Quinodoz, S. A. *et al.* Higher-Order Inter-chromosomal Hubs Shape 3D Genome  
583 Organization in the Nucleus. *Cell* **174**, 744-757.e24 (2018).
- 584 53. Zheng, M. *et al.* Multiplex chromatin interactions with single-molecule precision. *Nature*  
585 **566**, 558–562 (2019).
- 586 54. Blobel, G. A., Higgs, D. R., Mitchell, J. A., Notani, D. & Young, R. A. Testing the super-  
587 enhancer concept. *Nat. Rev. Genet.* **22**, 749–755 (2021).
- 588 55. Madsen, J. G. S. *et al.* Highly interconnected enhancer communities control lineage-  
589 determining genes in human mesenchymal stem cells. *Nat. Genet.* **52**, 1227–1238  
590 (2020).
- 591 56. Hay, D. *et al.* Genetic dissection of the  $\alpha$ -globin super-enhancer in vivo. *Nat. Genet.* **48**,  
592 895–903 (2016).
- 593 57. Sigalova, O. M., Shaeiri, A., Forneris, M., Furlong, E. E. & Zaugg, J. B. Predictive

- 594 features of gene expression variation reveal mechanistic link with differential  
595 expression. *Mol. Syst. Biol.* **16**, e9539 (2020).
- 596 58. Martinez-Ara, M., Comoglio, F., van Arensbergen, J. & van Steensel, B. Systematic  
597 analysis of intrinsic enhancer-promoter compatibility in the mouse genome. *Mol. Cell*  
598 **82**, 2519-2531.e6 (2022).
- 599 59. Bergman, D. T. *et al.* Compatibility rules of human enhancer and promoter sequences.  
600 *Nature* **607**, 176–184 (2022).
- 601 60. Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. & de Laat, W. Looping and  
602 interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell* **10**,  
603 1453–1465 (2002).
- 604 61. Krijger, P. H. L. & de Laat, W. Regulation of disease-associated gene expression in the  
605 3D genome. *Nat. Rev. Mol. Cell Biol.* **17**, 771–782 (2016).
- 606 62. Forcato, M. *et al.* Comparison of computational methods for Hi-C data analysis. *Nat.*  
607 *Methods* **14**, 679–685 (2017).
- 608 63. Sahlén, P. *et al.* Chromatin interactions in differentiating keratinocytes reveal novel  
609 atopic dermatitis- and psoriasis-associated genes. *J. Allergy Clin. Immunol.* **147**, 1742–  
610 1752 (2021).
- 611 64. Dixon, A. L. *et al.* A genome-wide association study of global gene expression. *Nat.*  
612 *Genet.* **39**, 1202–1207 (2007).
- 613 65. Pliner, H. A. *et al.* Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell  
614 Chromatin Accessibility Data. *Mol. Cell* **71**, 858-871.e8 (2018).
- 615 66. Rennie, S. *et al.* Transcription start site analysis reveals widespread divergent  
616 transcription in *D. melanogaster* and core promoter-encoded enhancer activities.  
617 *Nucleic Acids Res.* **39**, 311 (2018).
- 618 67. Reyes-Palomares, A. *et al.* Remodeling of active endothelial enhancers is associated  
619 with aberrant gene-regulatory networks in pulmonary arterial hypertension. *Nat.*  
620 *Commun.* **11**, 1673 (2020).
- 621 68. Fulco, C. P. *et al.* Activity-by-contact model of enhancer–promoter regulation from



- 622 thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).
- 623 69. Huang, Q. Q., Ritchie, S. C., Brozynska, M. & Inouye, M. Power, false discovery rate  
624 and Winner’s Curse in eQTL studies. *Nucleic Acids Res.* **46**, e133 (2018).
- 625 70. Gazal, S. *et al.* Combining SNP-to-gene linking strategies to identify disease genes and  
626 assess disease omnigenicity. *Nat. Genet.* **54**, 827–836 (2022).
- 627 71. Avsec, Ž. *et al.* Effective gene expression prediction from sequence by integrating long-  
628 range interactions. *Nat. Methods* **18**, 1196–1203 (2021).
- 629 72. He, Y. *et al.* sn-spMF: matrix factorization informs tissue-specific genetic regulation of  
630 gene expression. *Genome Biol.* **21**, 235 (2020).
- 631 73. Deplancke, B., Alpern, D. & Gardeux, V. The Genetics of Transcription Factor DNA  
632 Binding Variation. *Cell* **166**, 538–554 (2016).
- 633 74. Benko, S. *et al.* Highly conserved non-coding elements on either side of SOX9  
634 associated with Pierre Robin sequence. *Nat. Genet.* **41**, 359–364 (2009).
- 635 75. Delaneau, O. *et al.* Chromatin three-dimensional interactions mediate genetic effects  
636 on gene expression. *Science* **364**, eaat8266 (2019).
- 637 76. Avsec, Ž. *et al.* Base-resolution models of transcription-factor binding reveal soft motif  
638 syntax. *Nat. Genet.* **53**, 354–366 (2021).
- 639 77. de Almeida, B. P., Reiter, F., Pagani, M. & Stark, A. DeepSTARR predicts enhancer  
640 activity from DNA sequence and enables the de novo design of synthetic enhancers.  
641 *Nat. Genet.* **54**, 613–624 (2022).
- 642 78. Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650–665 (2018).
- 643 79. Kelley, D. R. Cross-species regulatory sequence activity prediction. *PLOS Comput.*  
644 *Biol.* **16**, e1008050 (2020).
- 645 80. Maslova, A. *et al.* Deep learning of immune cell differentiation. *Proc. Natl. Acad. Sci.*  
646 **117**, 25655–25666 (2020).
- 647 81. Janssens, J. *et al.* Decoding gene regulation in the fly brain. *Nature* **601**, 630–636  
648 (2022).
- 649 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.*  $\beta$ -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

- 678 Distal Regulator of Endothelin-1 Gene Expression. *Cell* **170**, 522-533.e15 (2017).
- 679 97. Smemo, S. *et al.* Obesity-associated variants within FTO form long-range functional  
680 connections with IRX3. *Nature* **507**, 371–375 (2014).
- 681 98. Mills, M. C. & Rahal, C. The GWAS Diversity Monitor tracks diversity by disease in real  
682 time. *Nat. Genet.* **52**, 242–243 (2020).
- 683 99. Polygenic Risk Score Task Force of the International Common Disease Alliance.  
684 Responsible use of polygenic risk scores in the clinic: potential benefits, risks and gaps.  
685 *Nat. Med.* **27**, 1876–1884 (2021).
- 686 100. Mohammadi, P. *et al.* Genetic regulatory variation in populations informs transcriptome  
687 analysis in rare disease. *Science* **366**, 351–356 (2019).
- 688 101. Tanjo, T., Kawai, Y., Tokunaga, K., Ogasawara, O. & Nagasaki, M. Practical guide for  
689 managing large-scale human genome data in research. *J. Hum. Genet.* **66**, 39–52  
690 (2021).
- 691 102. The 100,000 Genomes Project Pilot Investigators *et al.* 100,000 Genomes Pilot on  
692 Rare-Disease Diagnosis in Health Care — Preliminary Report. *N. Engl. J. Med.* **385**,  
693 1868–1880 (2021).
- 694 103. Kioussis, D., Vanin, E., deLange, T., Flavell, R. A. & Grosveld, F. G.  $\beta$ -Globin gene  
695 inactivation by DNA translocation in  $\gamma\beta$ -thalassaemi. *Nature* **306**, 662–666 (1983).
- 696 104. Driscoll, M. C., Dobkin, C. S. & Alter, B. P. Gamma delta beta-thalassemia due to a de  
697 novo mutation deleting the 5' beta-globin gene activation-region hypersensitive sites.  
698 *Proc. Natl. Acad. Sci.* **86**, 7470–7474 (1989).
- 699 105. De Gobbi, M. *et al.* A Regulatory SNP Causes a Human Genetic Disease by Creating a  
700 New Transcriptional Promoter. *Science* **312**, 1215–1217 (2006).
- 701 106. Lower, K. M. *et al.* Adventitious changes in long-range gene expression caused by  
702 polymorphic structural variation and promoter competition. *Proc. Natl. Acad. Sci.* **106**,  
703 21771–21776 (2009).
- 704 107. Lettice, L. A. *et al.* Disruption of a long-range cis-acting regulator for Shh causes  
705 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

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

735 121. Yang, H. *et al.* Noncoding genetic variation in GATA3 increases acute lymphoblastic  
736 leukemia risk through local and global changes in chromatin conformation. *Nat. Genet.*  
737 **54**, 170–179 (2022).

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

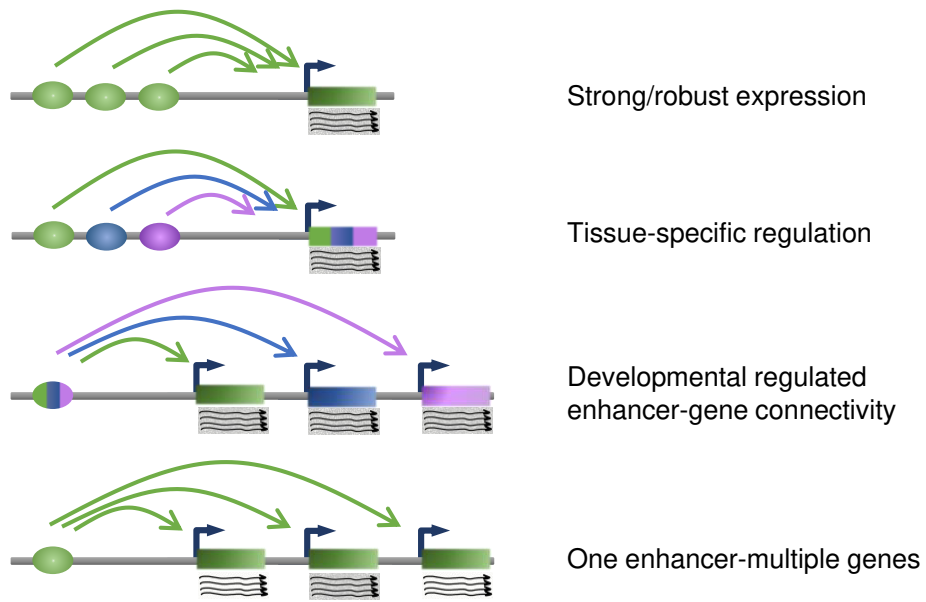
740 123. Gao, P. *et al.* Biology and Clinical Implications of the 19q13 Aggressive Prostate  
741 Cancer Susceptibility Locus. *Cell* **174**, 576-589.e18 (2018).

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

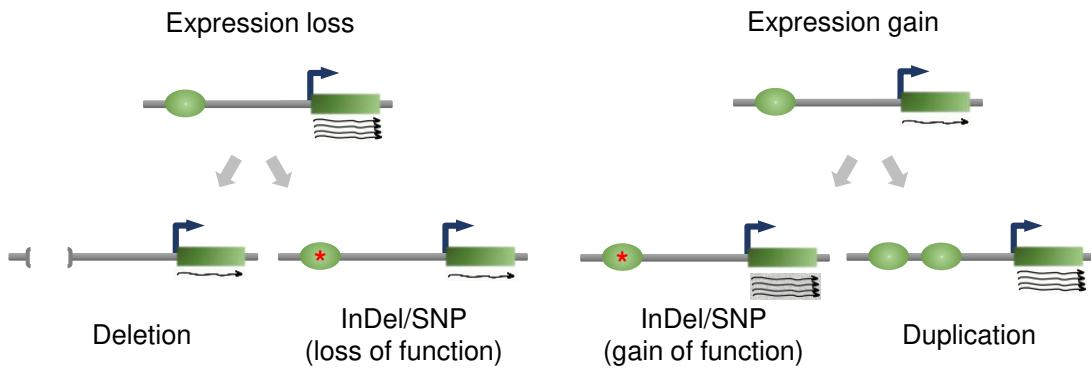
744

**A**

Functional interplay between enhancers and target genes

**B**

Enhancer disruption

**C**

Altered enhancer-gene connectivity

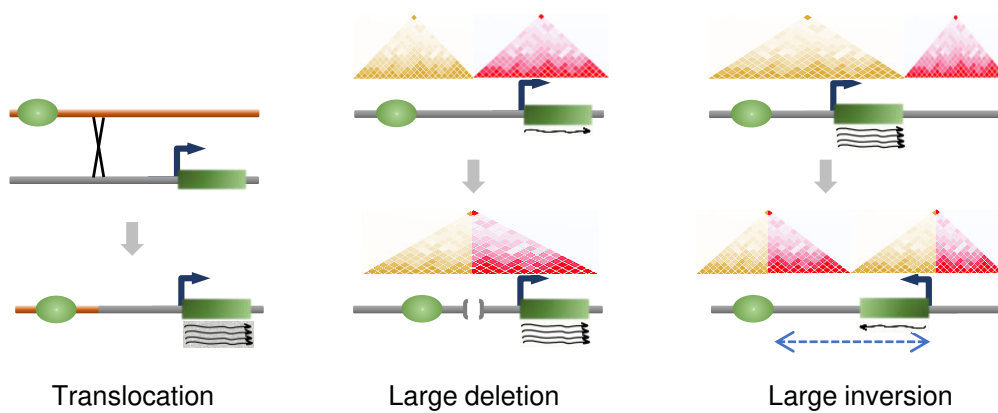


Figure 1 (REVISED)

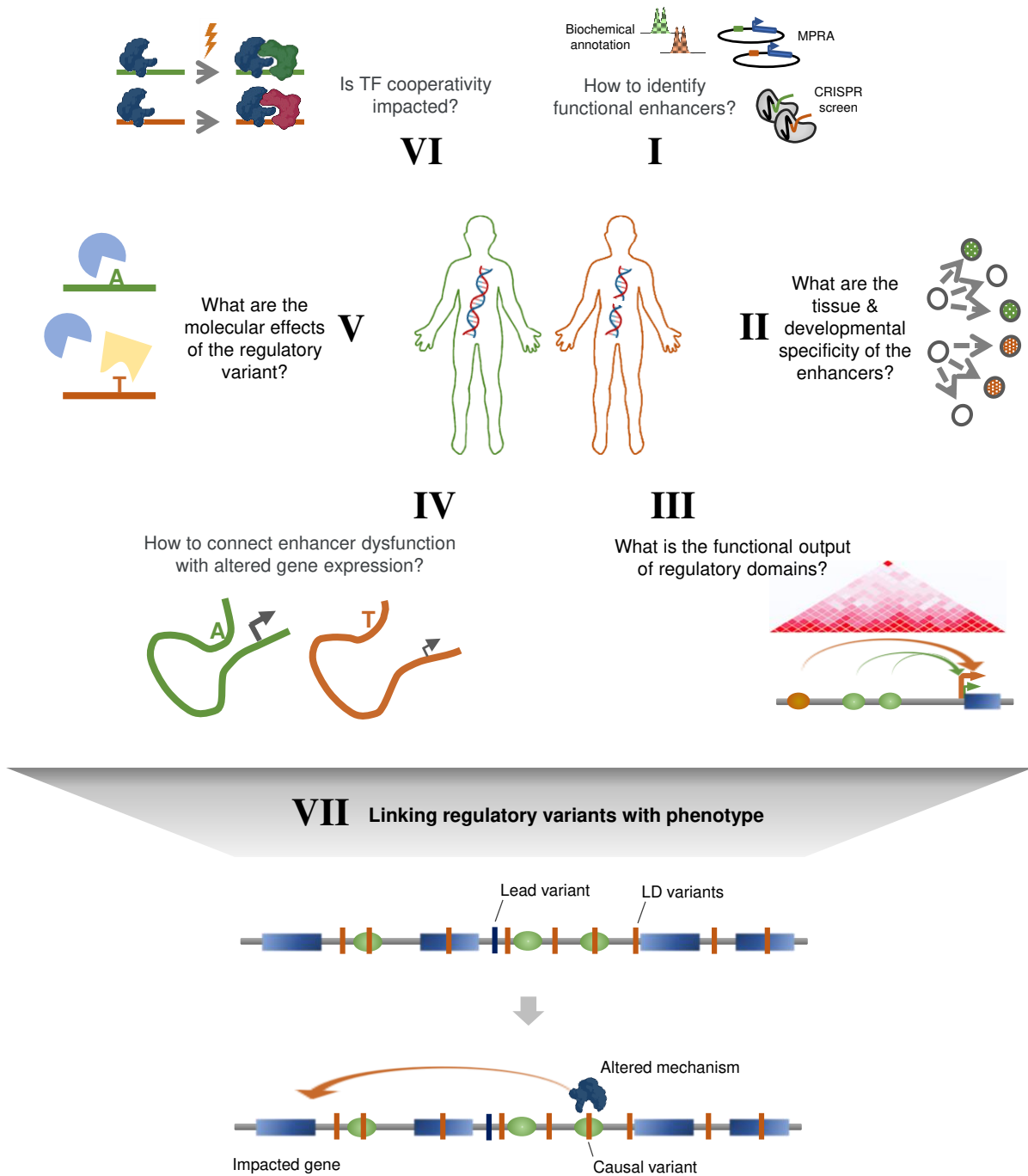


Figure 2