

REVIEW

# Enhancer variants: evaluating functions in common disease

Olivia Corradin<sup>1</sup> and Peter C Scacheri<sup>1,2\*</sup>

## Abstract

Gene enhancer elements are noncoding segments of DNA that play a central role in regulating transcriptional programs that control development, cell identity, and evolutionary processes. Recent studies have shown that noncoding single nucleotide polymorphisms (SNPs) that have been associated with risk for numerous common diseases through genome-wide association studies frequently lie in cell-type-specific enhancer elements. These enhancer variants probably influence transcriptional output, thereby offering a mechanistic basis to explain their association with risk for many common diseases. This review focuses on the identification and interpretation of disease-susceptibility variants that influence enhancer function. We discuss strategies for prioritizing the study of functional enhancer SNPs over those likely to be benign, review experimental and computational approaches to identifying the gene targets of enhancer variants, and highlight efforts to quantify the impact of enhancer variants on target transcript levels and cellular phenotypes. These studies are beginning to provide insights into the mechanistic basis of many common diseases, as well as into how we might translate this knowledge for improved disease diagnosis, prevention and treatments. Finally, we highlight five major challenges often associated with interpreting enhancer variants, and discuss recent technical advances that may help to surmount these challenges.

## Introduction

Transcriptional enhancer elements are noncoding stretches of DNA that have a central role in controlling gene expression programs in cells. Rather than on-off switches, enhancers are hypothesized to function as transcriptional rheostats to fine-tune target transcript levels. Higher-order

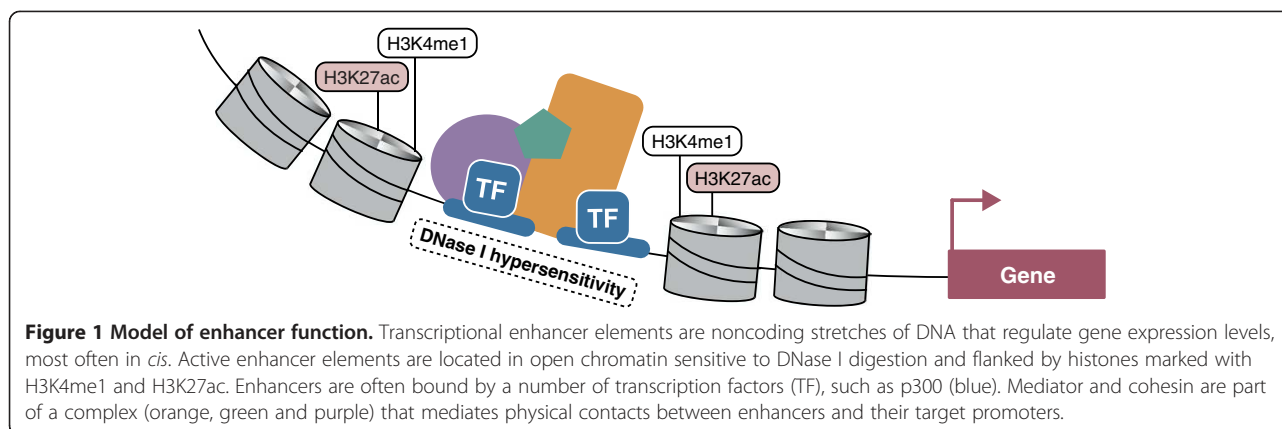
three-dimensional organization of chromatin facilitates physical interactions between enhancers and their target promoters. Interactions between enhancers and their targets may occur on the same chromosome (in *cis*) or on different chromosomes (in *trans*) (Figure 1) [1-3]. In any given mammalian cell type, the number of putative enhancer elements ranges from 50,000 to 100,000, and therefore far exceeds the number of protein-coding genes.

In the last decade, more than 1,900 genome-wide association studies (GWASs) have been published, identifying loci associated with susceptibility to over 1,000 unique traits and common diseases [4]. With the eventual goal of finding new therapies and preventative measures for common diseases, efforts are now focused on determining the functional underpinnings of these associations. Several groups have associated GWAS risk variants, mostly SNPs, with newly annotated cell-type-specific gene enhancer elements identified through epigenomic profiling studies. These enhancer variants probably play an important part in common disease susceptibility by influencing transcriptional output. Of all the genetic risk variants discovered to date, the number that impact enhancer function is estimated to far exceed the number that affect protein-coding genes or disrupt promoter function (Figure 2). Additionally, disease-associated variants in noncoding regions, particularly those that lie in cell-type-specific enhancer elements, have been estimated to explain a greater proportion of the heritability for some disorders than variants in coding regions [5]. This review focuses on the identification and interpretation of disease-associated variants that affect enhancer function. We consider the latest approaches for evaluating enhancer variants and identifying their gene targets, and highlight successful cases in which risk variants have been shown to alter gene expression by disrupting enhancer function. In addition, we discuss the remaining challenges to delineating the impact of noncoding variants, such as the identification of enhancer activity, validation of causal variants and identification of responsible genes. Future efforts to surmount these challenges should help to remove the barrier between the discovery of disease

\* Correspondence: [pxs183@case.edu](mailto:pxs183@case.edu)

<sup>1</sup>Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH 44122, USA

<sup>2</sup>Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH 44106, USA



associations and the translation of this knowledge for improved diagnosis and treatment of many common diseases.

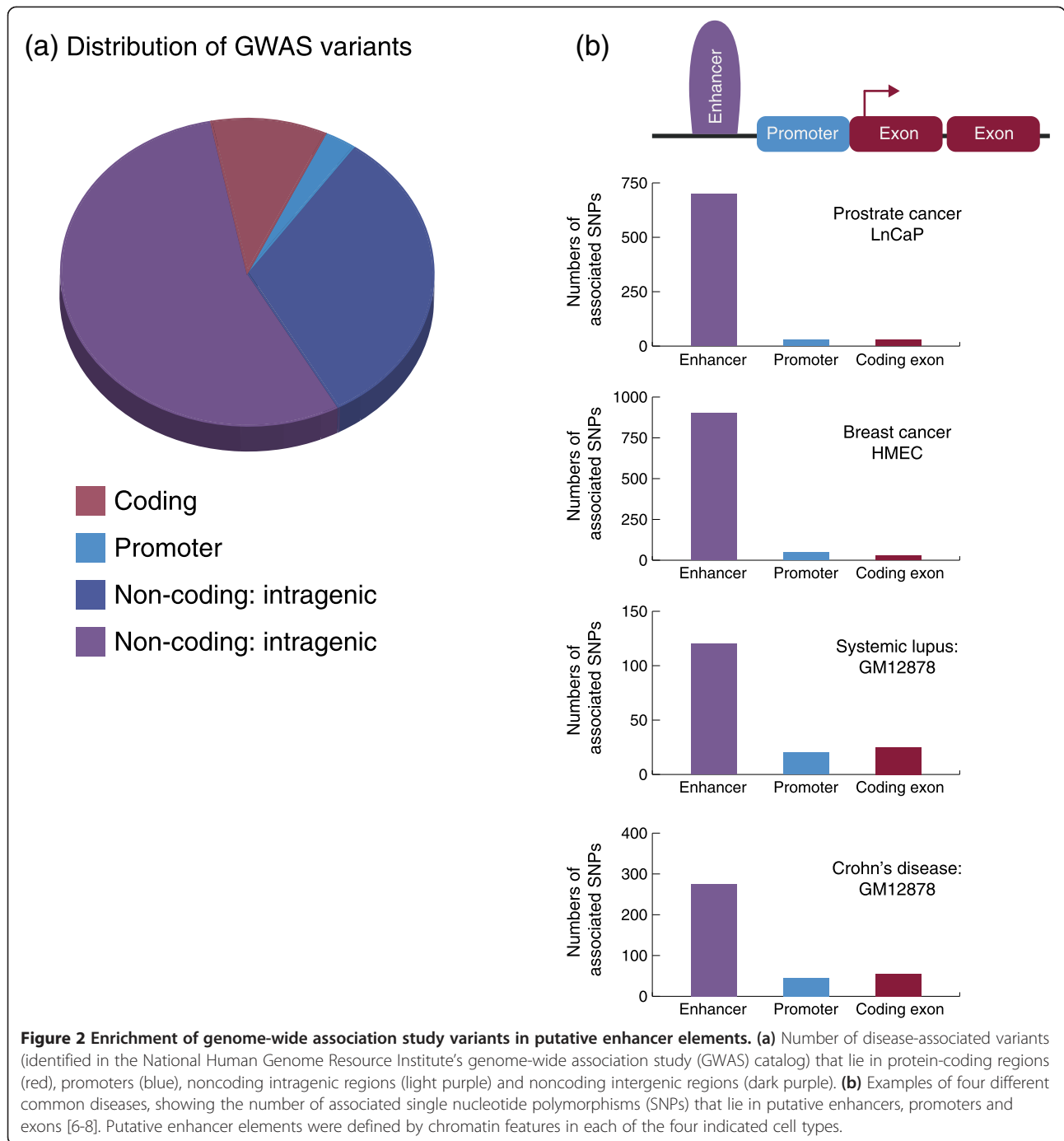
### Genetic risk variants are enriched in cell-type-specific enhancer elements defined by signature chromatin features

The locations of enhancer elements coincide with DNase I hypersensitive regions of open chromatin flanked by nucleosomes marked with the mono- and/or di-methylated forms of lysine 4 at histone H3 (H3K4me1/2) [9,10]. Enhancers can be active or repressed, and each state generally correlates with the presence of additional histone marks, such as H3K27ac and H4K16ac which are associated with active chromatin, or H3K27me3 and H3K9me3 which are associated with repressed chromatin [11-14]. Active enhancers are bi-directionally transcribed and capped at their 5' end [15,16]. Most enhancer elements are located in introns and intergenic regions, although some are exonic [17-19]. Relative to promoters, the distribution of enhancers across the epigenome is highly cell-type specific. Some of the first studies to associate GWAS variants with enhancer elements integrated genetic risk variants with regulatory element maps generated through epigenomic profiling (using chromatin immunoprecipitation combined with massively parallel DNA sequencing (ChIP-seq) and the profiling of DNase I hypersensitive sites (DHSs)) [20-22]. Two major themes emerged from these studies. First, loci with signature enhancer features (DHSs, H3K4me1, H3K27ac) are highly enriched for genetic risk variants relative to other chromatin-defined elements such as promoters and insulators [21]. Second, risk variants preferentially map to enhancers specific to disease-relevant cell types in both cancer and other common diseases [21]. For example, type 2 diabetes-associated variants preferentially map to pancreatic islet enhancers [22-25], and SNPs predisposing to colon cancer are enriched in enhancer elements in colon cancer cells and colon crypts, from which colon cancer is derived [26]. Further assessment of the

effects of enhancer risk variants has shown that they can alter transcription-factor-binding sites (TFBSs) and impact the affinity of transcription factors for chromatin, and/or induce allele-specific effects on target gene expression [6,27-40]. These studies illustrate the utility of epigenomic profiling for identifying risk variants that lie in putative enhancer elements and for identifying disease-relevant cell types in which the enhancer variants could exert their regulatory effects.

### Super-enhancers, stretch enhancers, and enhancer clusters: hotspots for genetic risk variants

Four studies recently demonstrated correlations between genetic risk variants and large clusters of active enhancers, similar to locus control regions. These features have been called 'super-enhancers' [41,42], 'stretch enhancers' [24], 'multiple enhancers' [7] and 'enhancer clusters' [23], and are similar but not identical between studies, although many of these features overlap. The methods used to identify these clusters are distinct. Super-enhancers, for example, are defined by identifying the top-ranking enhancers on the basis of the levels of associated transcription factors or chromatin marks identified through ChIP studies. Stretch enhancers are defined by stretches of open chromatin more densely and more broadly marked with enhancer-histone modifications than typical enhancers. Despite these differences, many of the defined features overlap. These enhancer clusters are highly cell-type specific and have been proposed to play a predominant role in regulating the cell-type-specific processes that define the biology of a given cell type. Moreover, they are disproportionately enriched for genetic risk variants compared to typical enhancers, and the enrichment is biased toward disease-relevant cell types. These results further support the notion that variants that influence cell-type-specific gene regulation are major contributors to common disease risk, and extend this concept to demonstrate that altering the expression of genes under exquisite regulation can frequently lead to increased risk. Enhancer cluster



identification provides a means of detecting highly regulated genes and may help to prioritize noncoding variants that are likely to be functional.

A typical locus identified through a GWAS contains dozens to thousands of SNPs in linkage disequilibrium (LD) with the 'lead' SNP that is reported to be associated with the disease in question. Any SNP in LD with the lead SNP may be causal, and the prevailing assumption is that only one is causal. Indeed, this scenario has been reported

to be the case for some risk loci involving enhancers [34,43], and there are several examples of Mendelian disorders in which a single enhancer variant causes congenital disease [44-50]. However, it is equally plausible that more than one SNP is causal, particularly at GWAS loci harboring enhancer clusters. In these instances, several variants distributed among multiple enhancers throughout the locus, rather than a single SNP, may combine to affect expression of their gene targets and confer susceptibility

to common traits. This has been called the ‘multiple enhancer variant’ (MEV) hypothesis. Corradin and colleagues provided support for the MEV hypothesis for six common autoimmune disorders, including rheumatoid arthritis, Crohn’s disease, celiac disease, multiple sclerosis, systemic lupus erythematosus and ulcerative colitis. The extent of MEVs across additional common diseases is not yet known [7,28,37].

### Interpreting enhancer variants

Given that risk variants lie in cell-type-specific enhancer elements, it is critical to utilize a disease-relevant cell type to identify potential enhancer variants. SNPs associated with a particular disease can be compared to enhancer elements to identify cell types whose active enhancers are disproportionately enriched for disease variants. Variant set enrichment is a permutation-based method that compares the enrichment of genetic risk-variant sets within any functional element (such as H3K4me1-marked putative enhancers) to randomly generated matched genetic risk-variant sets [26,38]. This type of analysis provides an unbiased way of evaluating the utility of a cell type for studying the impact of variants on enhancer elements.

Several computational programs are currently available to integrate chromatin landscapes with GWAS risk variants to identify candidate regulatory SNPs and evaluate their disease-causing potential. These include IGR [38], RegulomeDB [51], HaploReg [52], FunciSNP [53] and FunSeq [54]. These programs are particularly useful for prioritizing SNPs for functional analyses, which may include transcription factor ChIP or electrophoretic mobility shift assays to test whether a given SNP influences a transcription factor’s ability to bind to the enhancer, and *in vitro* and *in vivo* gene reporter assays to test the SNP’s effect on enhancer activity. Additionally, allele-specific expression can be utilized to quantify the impact of enhancer variants within a specific cell type. Finally, DNA editing strategies involving CRISPR/Cas9-based methods can be employed to evaluate the effect of a variant. Following the identification of a functional enhancer variant, the next major challenge is to identify its target and to test the effect of the SNP(s) on target transcript levels. Many enhancer elements are located within 100 kilobases (kb) of the genes that they regulate, but can also be located more than a megabase away, or even on separate chromosomes. Enhancers can regulate genes or long noncoding RNAs. Most genes are regulated by more than one enhancer, and many enhancers regulate more than one target gene [55,56]. The problem is further complicated by our limited knowledge of barrier elements, which block enhancer-gene interactions. The most common method of assigning an enhancer to its nearest gene is inaccurate, with false discovery rate (FDR) estimates ranging from 40% to 73% [55,57]. Refining methods for identifying the

nearest gene to looking for the ‘nearest expressed gene’ still results in a high FDR, with 53% to 77% [55,58] of distal elements bypassing the nearest active transcription start site to interact with a distant gene. Clearly, proximity alone cannot be used to accurately identify the target of an enhancer SNP.

### Methods of identifying gene targets of enhancer variants

To identify enhancer targets, DNA fluorescence *in situ* hybridization (FISH) [59,60], as well as chromatin association methods (chromosome conformation capture (3C)) [61], can be employed. These are powerful approaches for evaluating whether a region of interest interacts with a specific genomic target, but they suffer from the limitation that the regions of interest must be pre-specified, that is, they are ‘one-by-one’ approaches. 4C (circular chromosome conformation capture), an extension of 3C, can capture all regions that physically contact a site of interest, without prior knowledge of the regions that contact that site being necessary [62] (that is, a ‘one-to-all’ approach). Higher-throughput methods include carbon-copy chromosome conformation capture (5C, many-to-many), a high-throughput expansion of 3C, Hi-C (all-to-all) and chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) (for detailed comparison of these methods, see reviews [63,64]). These global approaches can enable the identification of loci that directly and indirectly contact enhancers of interest, and can reveal complex interactions in which dozens to hundreds of loci aggregate, so-called transcriptional hubs or enhanceosomes [65]. These types of high-order interactions have been recently described by several studies [55,56,58]. The extent by which they overlap risk loci remains unexplored. Unfortunately, these approaches tend to be expensive and difficult for most labs to execute, and their resolution often prohibits their use for interrogating GWAS loci. Until recently, for example, the resolution of Hi-C was limited to capturing interactions separated by more than one megabase; 5 to 10 times greater than the distance by which most enhancer-gene interactions occur. Despite the limitations, ‘C’-based methods have been implemented to successfully identify targets of enhancer-risk variants and to quantify their functional effects. For example, Cowper-Salari and colleagues utilized 3C and allele-specific expression to demonstrate the impact of the breast cancer risk SNP rs4784227 on expression of *TOX3*, thought to have a role in chromatin regulation [38]. Bauer and co-workers utilized 3C to identify *BCL11A* as the gene target of an erythroid enhancer, and then further demonstrated the impact of enhancer variants on transcription factor binding and expression. Gene editing strategies have also been employed to demonstrate that this enhancer is essential for erythroid gene expression [28]. Finally, we highlight



a study by Smemo and colleagues in which 4C-seq was used to identify *IRX3* as the target of an enhancer SNP located in intron 1 of the *FTO* gene, which was originally thought to be the target and therefore the causal gene for increased risk of obesity. Functional studies in mice were used to verify that *IRX3* is the most likely causal gene, not *FTO* [30].

### Computational approaches to identify gene targets of enhancer elements

As alternatives to experimental approaches, several groups have developed computational techniques for determining the targets of enhancers [7,16,21,66-70]. These methods are similar in that they compare patterns of regulatory activity across multiple cell types to predict interactions between enhancers and genes. However, they vary significantly in the type of data required to generate enhancer-gene predictions, the specific approaches used to generate and validate the predictions, and their availability (Table 1). The method described by Ernst and colleagues identifies H3K4me1/2 and H3K27ac sites that co-vary with expressed genes within 125 kb of the gene locus, and uses this to predict enhancer-gene interactions [21]. Thurman and co-workers utilized DHS exclusively to predict interactions. Enhancers were assigned to genes by correlating the cross-cell-type DNase I signal at each DHS site with all promoters located within 500 kb [66]. The method developed by Sheffield and colleagues also uses DHS profiles, but additionally incorporates genome-wide expression data [70]. Rather than employing a fixed distance-based model, Shen and colleagues apply chromatin conformation data from Hi-C experiments to guide the association of enhancers to genes marked by H3K4me1, H3K27ac and RNA Pol II [67]. As an alternative to methods based on chromatin structure, Andersson and colleagues leverage cap analysis of gene expression (CAGE) data to correlate transcription at enhancers with gene expression [16]. There are two computational approaches that are publicly available and executable through website or command-line programs: predicting specific tissue interactions of genes and enhancers (PreSTIGE) [7] and integrated methods for predicting enhancer targets (IM-PET) [69]. PreSTIGE identifies enhancers and genes that demonstrate quantitative cell-type specificity based on H3K4me1 and RNA sequencing (RNA-seq), and can process data from human and mouse cell types [68]. IM-PET, like previously discussed methods, considers the proximity of an enhancer to potential gene targets and the correlation of enhancer and promoter activity, along with measures of transcription factor activity and evolutionary conservation.

When the appropriate datasets are available, computational approaches can offer a relatively fast and cost-effective way of identifying putative enhancer-gene interactions in a given cell type. However, they are generally limited to detecting a subset of enhancer-promoter

interactions within a given cell type, and none are capable of identifying *trans* interactions. Methods that rely on cell-type specificity or concordant changes in enhancers and genes across cell types may lack the sensitivity to predict interactions for ubiquitously expressed genes or to delineate interactions in domains with a high density of cell-type-specific genes. There is no standard or 'reference' dataset to validate the accuracy of gene-enhancer predictions. Thus, each study utilizes a different approach to evaluate accuracy, which makes it difficult to determine which method is most accurate. This necessitates experimental validation of enhancer-gene interactions determined using prediction-based methods. Despite these limitations, computational approaches can help to identify the targets of enhancer-risk variants. The method developed by Thurman and colleagues was applied to all GWAS loci and predicted gene targets of 419 disease-associated risk variants [20], most of which were located more than 100 kb from the risk SNP. PreSTIGE was utilized to predict gene targets of 122 noncoding loci associated with six immune disorders: rheumatoid arthritis, Crohn's disease, celiac disease, multiple sclerosis, lupus and ulcerative colitis. Furthermore, at several of the autoimmune-disease-associated loci, the effect of the risk allele on target gene expression was quantified.

### Utilizing expression quantitative trait loci studies to evaluate the impact of enhancer variants

Expression quantitative trait loci (eQTL) studies enable the identification of genetic variants that influence gene expression. eQTL studies involve stratifying a panel of individuals based on their particular SNP genotypes and then determining whether transcript levels differ between individuals based on the specific SNP genotypes. Genome-wide eQTL studies have identified transcripts that differ in expression on the basis of the genotype of the risk allele at GWAS loci. Alternatively, eQTL-based analyses can be applied to candidate interactions between SNPs and gene targets identified through the experimental or computational approaches described above. In both instances, genetic variation inherent in the human population is utilized to reveal the quantitative and directional effect of SNPs on gene expression (that is, the degree to which expression is upregulated or downregulated).

eQTL studies can locate SNPs within a given GWAS locus that influence target transcript levels, but caution must be taken when interpreting results. First, eQTLs, like enhancers, are cell-type specific. Thus, the effect of a SNP on transcription may only occur in disease-relevant cell types [71,72]. Second, the SNP associated with transcript levels may not be the causal SNP: SNPs in LD with the eQTL SNP may be driving the association. Third, the results are correlative and may reflect indirect associations

**Table 1 Computational approaches to predicting gene targets of enhancer elements**

Reference or method	Input data required	Gene expression	Linear model	Number of genes with predictions (per cell line)	FDR	Species	Publicly available
Nearest gene	None	Not considered	Nearest gene	NA	~40% to 73%	Any	NA
Nearest expressed gene	Gene expression	Considered	Nearest expressed gene	NA	~53% to 77%	Any	NA
Ernst <i>et al.</i> [21]	H3K4me1, H3K4me2, H3K27ac, RNA-seq	Considered	Distance based (125 kb)	NA	Not determined	Human	No
Thurman <i>et al.</i> [66]	DNase I hypersensitivity	Not considered	Distance based (500 kb)	NA	Not determined	Human	No
Sheffield <i>et al.</i> [70]	DNase I hypersensitivity and RNA-seq	Considered	100 kb	NA	Not determined	Human	Predicted interactions: [ <a href="http://dnase.genome.duke.edu/">http://dnase.genome.duke.edu/</a> ]
Shen <i>et al.</i> [67]	H3K4me1, H3K27ac, RNA Pol II	Not considered	Topological domain based	5,000 to 8,000	Not determined	Mouse	No
Andersson <i>et al.</i> [16]	CAGE	Considered	500 kb	NA	Not determined	Human	No
PreSTIGE [7]	H3K4me1	Considered	Distance (100 kb) and CTCF based	3,000 to 5,000	~13% to 23%	Human	Predicted interactions: <a href="http://genetics.case.edu/prestige">http://genetics.case.edu/prestige</a> Method application: <a href="http://prestige.case.edu">http://prestige.case.edu</a>
PreSTIGEouse [68]	H3K4me1	Considered	Distance based (100 kb)	3,000 to 5,000	Not determined	Mouse	Predicted interactions: <a href="http://genetics.case.edu/prestige">http://genetics.case.edu/prestige</a> Method application: <a href="http://prestige.case.edu">http://prestige.case.edu</a>
IM-PET [69]	H3K4me1, H3K27ac, H3K4me3 and RNA-seq*	Considered	Distance (2 Mb)	7,000 to 10,000	~1%	Human	Method application: <a href="http://www.healthcare.uiowa.edu/labs/tan/IM-PET.html">http://www.healthcare.uiowa.edu/labs/tan/IM-PET.html</a>

\*Input data utilized in publication, other input options exist. CAGE, cap analysis of gene expression; CTCF, CCCTC-binding factor (zinc finger protein demonstrated to function as an insulator protein); FDR, false discovery rate; Mb, megabases; NA, not applicable; RNA-seq, RNA sequencing.

between SNPs and genes. Fourth, the effects on gene expression must be robust in order to be identified over the confounding effects of the genetic background. This poses a challenge for detecting functional variants that have modest effects, as has been proposed for most enhancer variants [7,33,73,74]. Fifth, eQTL analyses rarely consider the combinatorial effects of multiple SNPs at a given locus. Last, because eQTL studies are typically performed on healthy individuals, the impact of the SNP on the quantitative trait may differ in response to disease-specific stimuli. This was observed in a survey of enhancer SNPs associated with prostate cancer. Here, the effect of a SNP on enhancer function was only observed in the presence of the androgen dihydrotestosterone [6]. Additionally, a study by Harismendy and co-workers demonstrated that the chromatin interaction between an enhancer locus associated with coronary artery disease and the gene target *IFNA21* was significantly remodeled by treatment with interferon- $\gamma$  [31].

### Transcriptional effects of enhancer variants

Studies that delineate the impact of disease-associated enhancer variants (Table 2) reveal the relatively modest effect of enhancer variants on gene expression. The effect of enhancer variants has also been evaluated with massively parallel reporter assays in which the impact of mutations in enhancer sequences is determined through heterologous barcoding and high-throughput sequencing (reviewed in [75]). These high-throughput assays show that most variants that impact transcription induce 1.3- to 2-fold differences in target gene expression [73,74]. These findings align with the notion that enhancers modulate or fine-tune gene expression, analogous to a rheostat. Despite their modest transcriptional effects, enhancer variants can have large effects on downstream phenotypes. As an example, we highlight a SNP (rs12821256) associated with blond hair color in Europeans. This SNP lies in an enhancer that drives *KITLG* expression in developing hair follicles [33]. The blond-hair-associated SNP was shown to reduce enhancer activity by only 22% *in vitro*. Nonetheless, when the blond hair and ancestral alleles were evaluated in transgenic mice, the reduction in enhancer activity associated with the blond hair allele was sufficient to yield mice of visibly lighter coat color than mice generated with the ancestral allele [33]. Whether or not the blond-hair-associated SNP represents a special instance of a more general mechanism in which enhancer variants with modest functional effects have robust phenotypic effects remains to be seen.

### Implications for disease and medicine

The strategies discussed above (summarized in Figure 3) have been utilized to interpret the transcriptional effects of enhancer variants associated with several traits and common diseases. The continued application of these and

other emerging strategies will have important implications for disease and medicine. These studies should not only help to reveal the gene targets of noncoding risk variants, but also provide information on whether these risk variants increase or decrease expression of the target gene. This information will be essential for identifying appropriate therapeutic targets and determining whether inhibitors or activators of these targets would be most effective. Knowledge of gene targets may also reveal pathways that are commonly altered among affected individuals, which could also guide treatment strategies and rational drug design.

### Conclusions and future challenges

We have reviewed approaches for the identification and interpretation of common-disease-associated variants that impact enhancer function, citing examples in which these methods have been successfully implemented (Figure 3, Table 2). We highlight three main conclusions. First, cell-type-specific enhancer variants are highly prevalent among loci associated with the majority of common diseases identified through GWASs. Second, GWAS-identified enhancer variants are disproportionately enriched in enhancer clusters, which control genes with highly specialized cell-type-specific functions. Third, these enhancer variants can have modest but significant effects on target gene expression, which can have robust effects on phenotype. Thus, interpreting the functional effects of enhancer variants requires rational experiment design that takes these characteristics into account. Furthermore, although current methods have enabled the thorough characterization of enhancer variants at some GWAS loci, high-throughput methods are needed, given the huge number of disease-associated enhancer variants. Here, we discuss additional lessons learned from these studies, and note five remaining challenges (Figure 4).

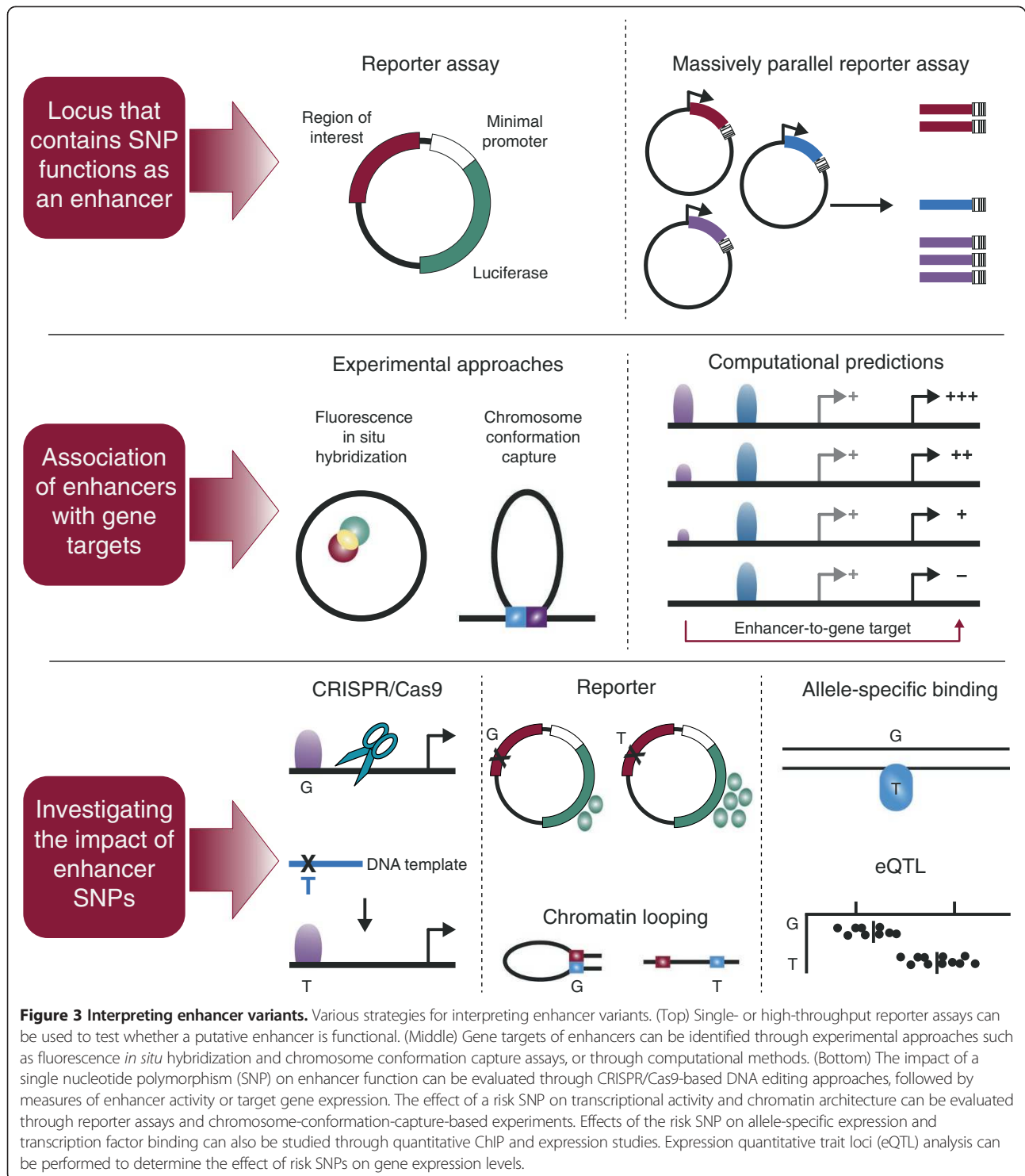
First, chromatin landscapes vary considerably between cell types and are highly dynamic, capable of altering in response to internal and external environmental stimuli. Given the spacial, temporal, environmental and epigenetic complexity of gene regulation, it is essential that the appropriate human cell type or model is utilized when trying to draw inferences between risk alleles and enhancer elements. Integrating risk variants with the chromatin landscapes of cell types or conditions that are insufficient models for a disorder will likely give misleading results. This is highlighted by eQTL studies. Even in comparisons of relatively similar cell types (monocytes and T cells [72] or B cells and monocytes [71]), noncoding variants that impact expression in one cell type often had no effect in the other cell type. Additionally, in a study of *cis*-regulation in colon cancer, the impact of some SNPs on expression was seen amongst colon cancer samples, but not observed in normal colon from the same patients, implying

**Table 2 Functional enhancer studies of GWAS risk loci**

Disease/trait	Reference	Lead SNP	Proposed functional SNP	Gene target	How gene target was selected	Cell type	Data supporting SNP enhancer function
Blond hair color	Guenther <i>et al.</i> [33]	rs12821256	rs12821256	<i>KITLG</i>	Phenotype in mouse model	Developing hair follicles, HaCaT keratinocyte cell line	Allele-specific luciferase activity, allele-specific ChIP, effect of SNP on mouse phenotype
Breast cancer	Cowper-Salari <i>et al.</i> [38]	rs4784227	rs4784227	<i>TOX3</i>	3C	MCF7	Binding motif disruption, allele-specific ChIP, allele-specific 3C, allele-specific expression, eQTL
Colorectal cancer	Pomerantz <i>et al.</i> [34]	rs6983267	rs6983267	<i>c-MYC</i>	3C	Colo205 and LS174T	Allele-specific luciferase activity, allele-specific ChIP
Colorectal cancer	Wright <i>et al.</i> [39]	rs6983267	rs6983267	<i>c-MYC</i>	3C	DLD1 and HCT116	Allele-specific ChIP, allele-specific expression
Colorectal cancer	Tuupanen <i>et al.</i> [36]	rs67491583	rs67491583	<i>c-MYC</i>	Nearest gene	HeLa	Binding motif disruption, allele-specific ChIP, allele-specific luciferase activity
Prostate cancer	Wasserman <i>et al.</i> [35]	rs6983267	rs6983267	<i>c-MYC</i>	Nearest gene	Prostate tissue (mouse)	Allele-specific activity, LacZ enhancer assay (mouse)
Coronary artery disease	Harismendy <i>et al.</i> [31]	rs10757278	rs10811656/ rs10757278	<i>CDKN2B</i> , <i>CDKN2BAS</i> , <i>IFNA21</i> , <i>MTAP</i>	3C and FISH ( <i>IFNA21</i> )	HUVEC	Binding motif disruption, allele-specific ChIP
Coronary heart disease	Miller <i>et al.</i> [29]	rs12190287	rs12190287	<i>TCF21</i>	Nearest gene, eQTL gene	HCASMC	Binding motif disruption, allele-specific luciferase activity, EMSA, allele-specific ChIP and allele-specific expression
Fetal hemoglobin level	Bauer <i>et al.</i> [28]	rs1427407, rs7606173	rs1427407, rs7606173	<i>BCL11A</i>	3C	Primary human erythroblasts	Binding motif disruption, allele-specific ChIP, allele-specific expression, LacZ enhancer assay (mouse), deletion by TALEN
Multiple sclerosis	Alcina <i>et al.</i> [27]	rs658115	rs10877013	<i>FAM119B</i> , <i>AVIL</i> , <i>TSFM</i> , <i>TSPAN31</i>	eQTL	LCLs and monocytes	Allele-specific luciferase activity, eQTL
Obesity	Smemo <i>et al.</i> [30]	rs9930506	NA	<i>IRX3</i>	4C-seq, 3C, ChIA-PET, Hi-C	Whole mouse embryo and adult mouse brain	eQTL mapping
Prostate cancer	Hazelett <i>et al.</i> [6]	rs5945619	rs4907792	<i>NUDT1</i>	Nearest gene, eQTL gene	LNCaP	Allele-specific luciferase activity, eQTL
Prostate cancer	Hazelett <i>et al.</i> [6]	rs10486567	rs10486567	<i>JAZF1</i>	Nearest gene	LNCaP	Allele-specific luciferase activity, binding motif disruption
QT interval	Kapoor <i>et al.</i> [32]	rs12143842	rs7539120	<i>NOS1AP</i>	eQTL gene, genetic association	Cardiac tissues	Allele-specific luciferase activity, eQTL, enhancer assay (zebrafish embryos)
Restless leg syndrome	Spieler <i>et al.</i> [37]	rs12469063	rs13469063	<i>MEIS1</i>	PreSTIGE prediction method, ChIA-PET, Hi-C	Telencephalon	Allele-specific expression of reporter gene in zebrafish, Allele-specific LacZ (mouse), EMSA, binding motif disruption, effect of decreased gene expression on phenotype
Systemic lupus erythematosus	Wang <i>et al.</i> [40]	rs2230926	rs148314165, rs200820567	<i>TNFAIP3</i>	3C	LCLs	EMSA, allele-specific luciferase activity, allele-specific 3C
Type 2 diabetes	Gaulton <i>et al.</i> [76]	rs7903146	rs7903146	<i>TCF7L2</i>	Nearest gene	Pancreatic islet cells	Allele-specific luciferase activity, allele-specific FAIRE

3C, chromosome conformation capture; 4C-seq, circular chromosome conformation capture followed by sequencing; ChIA-PET, chromatin interaction analysis by paired-end tag sequencing; ChIP, chromatin immunoprecipitation; EMSA, electrophoretic mobility shift assay; eQTL, expression quantitative trait loci; FAIRE, formaldehyde-assisted isolation of regulatory elements; FISH, fluorescence *in situ* hybridization; LCLs, lymphoblastoid cell lines; NA, not applicable; SNP, single nucleotide polymorphism.

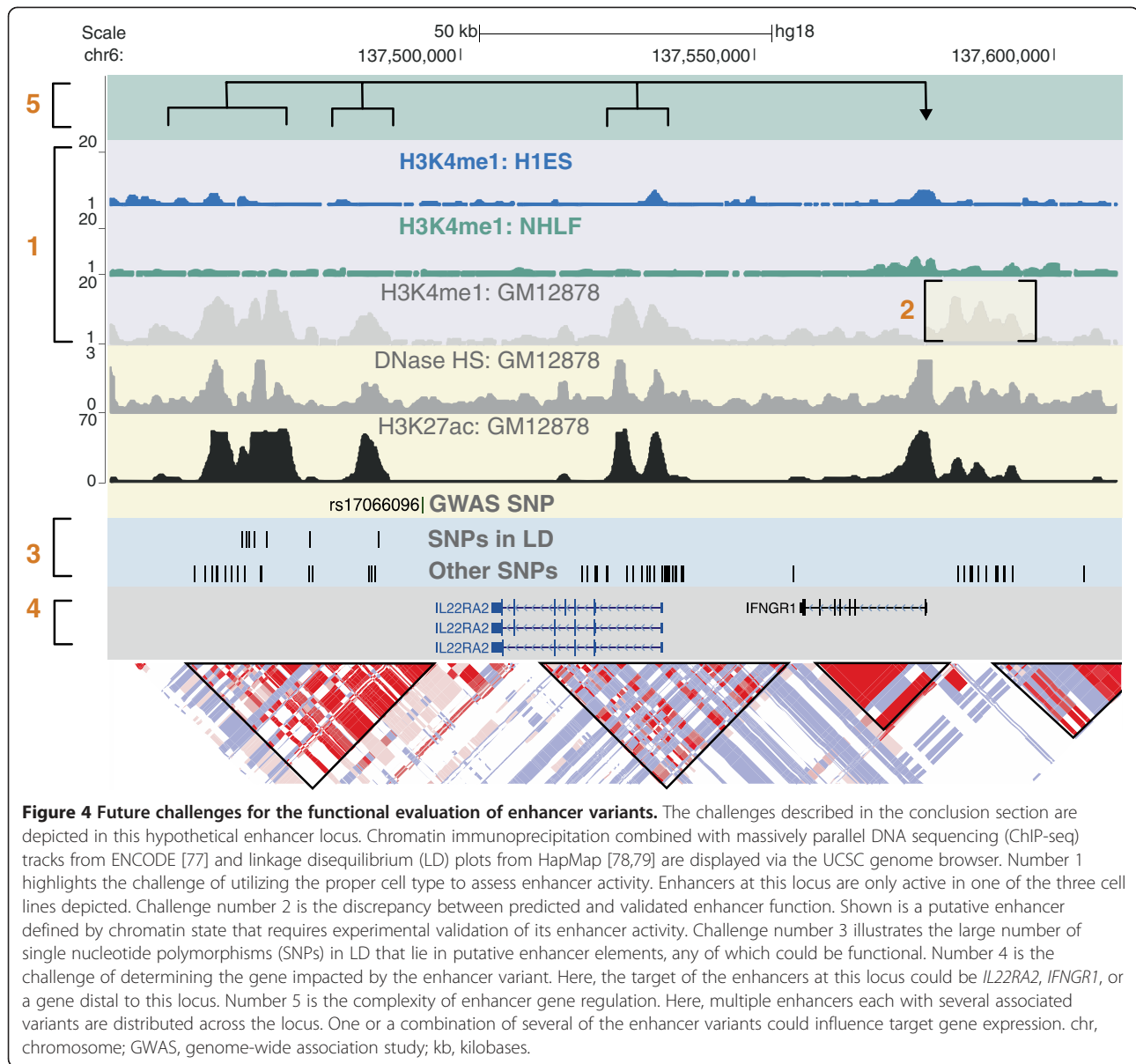




that the impact of the variant is dependent on disease-specific environmental factors [80]. The effect of non-coding variants on expression was also observed to be strongly context dependent in a study of monocytes under diverse types and durations of stimuli. Fairfax and colleagues demonstrated that 43% of identified eQTLs were

associated with an effect on expression only after treatment with immune response stimuli lipopolysaccharide or interferon- $\gamma$  [81].

Second, there remains a gap between the prediction and functional validation of putative enhancer elements. Thus, if a risk SNP is localized to a putative enhancer



element defined through chromatin profiling, it is essential that the putative enhancer is functionally validated. *In vitro* and *in vivo* reporter assays can help in this regard. However, these assays are relatively low throughput and usually involve the use of a general promoter such as SV40 rather than the enhancer's endogenous promoter, which complicates the interpretation of negative results. Additionally, most genes are regulated by more than one enhancer, yet typically only one enhancer is tested in a reporter assay.

Third, at a given GWAS locus, the SNP with the most significant association (that is, the lowest *P* value) with the disease is usually reported as the 'lead' SNP. Except in rare instances, such as the SNP rs6983267, which influences the MYC enhancer and confers risk for multiple

cancers [34,35], the SNP with the lowest *P* value is not necessarily causal. Any SNP in LD with the lead SNP may be causal, and there may be dozens to thousands of candidates. Fine mapping studies can help narrow the locus and reduce the number of candidates. Additionally, as discussed above, identifying SNPs that co-localize with enhancer-chromatin features or TFBSs in an appropriate human cell type can help prioritize candidate functional variants [30,38]. Indeed, Claussnitzer and colleagues developed a method, phylogenetic module complexity analysis (PMCA), which utilizes conserved co-occurring TFBS patterns to identify functional regulatory variants [82]. However, hundreds of candidate SNPs may remain even after prioritization, especially when the locus harbors an enhancer cluster. This was

illustrated in a recent survey of breast cancer risk loci, which showed that 921 SNPs co-localize with putative enhancers in human mammary epithelial cells across 71 risk loci [8]. Similarly, 663 enhancer SNPs were identified for 77 prostate risk loci [6]. Furthermore, while some enhancer variants influence transcription factor binding [6,28,29,34], SNPs do not necessarily have to reside within a TFBS to influence transcription factor binding or enhancer activity [33,73,74,83]. It is clear that massively parallel reporter assays (discussed above) will be necessary to help distinguish functional variants from those that are passengers.

Fourth, in order to determine the phenotypic effect of an enhancer variant, it is essential that an enhancer variant is demonstrated to influence the levels of its target transcript. The target may be a gene, or could alternatively be a noncoding RNA. However, enhancers frequently regulate multiple genes. Even if the levels of a given transcript correlate with the genotype of an enhancer risk variant, it does not necessarily mean that the correlated gene is causal. Functional assays, and ultimately *in vivo* models, are needed to verify that the gene is directly involved in the development of the disease. CRISPR/Cas9 technology would enable such studies by altering single SNPs in the genome of a model organism while maintaining the native genomic context of the variant. Alternatively, single-site integration of the risk or non-risk alleles into a model organism, as utilized for the enhancer variant associated with blond hair color [33], could be employed. Although CRISPR/Cas9 can be utilized to demonstrate the functional impact of a given variant, the complex phenotypes of many diseases are not easily modeled *in vitro* and thus the determination of causality will often not be trivial.

Lastly, genes are frequently regulated by multiple enhancer elements or clusters of enhancer elements. Thus, the independent effect of a single enhancer or variant may be below the sensitivity threshold of standard assays. This, in addition to the demonstration that multiple enhancer SNPs can act in combination to impact gene expression, suggests that epistatic effects between noncoding variants may play a particularly important role for enhancer loci, especially when enhancer variants of the same gene are inherited independently. The impact of the interaction between SNPs on transcription and ultimately on clinical risk for disease remains to be seen.

We have discussed the strategies for, and challenges associated with, the interpretation of noncoding putative enhancer SNPs as applied to the study of common variants identified by GWAS studies of common diseases and traits. As whole-genome sequencing becomes more prevalent, these same strategies will be necessary to elucidate the impact of rare noncoding mutations and to distinguish damaging from innocuous enhancer alterations.

## Abbreviations

3C: Chromosome conformation capture; 4C: Circular chromosome conformation capture; 5C: Carbon-copy chromosome conformation capture; CAGE: Cap analysis of gene expression; ChIA-PET: Chromatin interaction analysis by paired-end tag sequencing; ChIP-seq: Chromatin immunoprecipitation with massively parallel DNA sequencing; DHS: DNase I hypersensitivity site; eQTL: Expression quantitative trait loci; FDR: False discovery rate; FISH: Fluorescence *in situ* hybridization; GWAS: Genome-wide association study; H3K27ac: Acetylation of lysine 27 on histone 3 [as an example]; H3K4me: Methylation of lysine 4 on histone 3 [as an example]; IM-PET: Integrated methods for predicting enhancer targets; kb: Kilobases; LD: Linkage disequilibrium; MEV: Multiple enhancer variant; PMCA: Phylogenetic module complexity analysis; PreSTIGE: Predicting specific tissue interactions of genes and enhancers; RNA-seq: RNA sequencing; SNP: Single nucleotide polymorphism; TFBS: Transcription-factor-binding site; VSE: Variant set enrichment.

## Competing interests

The authors declare that they have no competing interests.

## Acknowledgements

This work was supported by NIH grants R01CA160356, R01DC009410, R01DE018470 (PCS) and CWRU Cellular and Molecular Biology training grant T32GM008056 (OC).

Published online: 28 October 2014

## References

1. Sasaki-Iwaoka H, Maruyama K, Endoh H, Komori T, Kato S, Kawashima H: **A trans-acting enhancer modulates estrogen-mediated transcription of reporter genes in osteoblasts.** *J Bone Miner Res* 1999, **14**:248–255.
2. Muller HP, Schaffner W: **Transcriptional enhancer can act *in trans*.** *Trends Genet* 1990, **6**:300–304.
3. Muller HP, Sogo JM, Schaffner W: **An enhancer stimulates transcription *in trans* when attached to the promoter via a protein bridge.** *Cell* 1989, **58**:767–777.
4. **A Catalog of Published Genome-Wide Association Studies.** [http://www.genome.gov/gwasstudies]
5. Gusev A, Hong Lee S, Neale BM, Trynka G, Vilhjalmsson BJ, Finucane H, Xu H, Zang C, Ripka S, Stahl E, Schizophrenia Working Group of the PGC, SWE-SCZ Consortium, Kahler AK, Hultman CM, Purcell SM, McCarrroll SA, Daly M, Pasaniuc B, Sullivan PF, Wray NR, Raychaudhuri S, Price AL: **Regulatory variants explain much more heritability than coding variants across 11 common diseases.** *bioRxiv* 2014, doi:10.1101/004309.
6. Hazelett DJ, Rhie SK, Gaddis M, Yan C, Lakeland DL, Coetzee SG, Henderson BE, Noushmehr H, Cozen W, Kote-Jarai Z, Eeles RA, Easton DF, Haiman CA, Lu W, Farnham PJ, Coetzee GA, Ellipse/GAME-ON consortium, Practical consortium: **Comprehensive functional annotation of 77 prostate cancer risk loci.** *PLoS Genet* 2014, **10**:e1004102.
7. Corradin O, Saiakhova A, Akhtar-Zaidi B, Myeroff L, Willis J, Cowper-Sal Lari R, Lupien M, Markowitz S, Scacheri PC: **Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits.** *Genome Res* 2014, **24**:1–13.
8. Rhie SK, Coetzee SG, Noushmehr H, Yan C, Kim JM, Haiman CA, Coetzee GA: **Comprehensive functional annotation of seventy-one breast cancer risk loci.** *PLoS One* 2013, **8**:e63925.
9. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, Wang W, Weng Z, Green RD, Crawford GE, Ren B: **Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome.** *Nat Genet* 2007, **39**:311–318.
10. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, Ching KA, Antosiewicz-Bourget JE, Liu H, Zhang X, Green RD, Lobanenkov VV, Stewart R, Thomson JA, Crawford GE, Kellis M, Ren B: **Histone modifications at human enhancers reflect global cell-type-specific gene expression.** *Nature* 2009, **459**:108–112.
11. Creghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R: **Histone H3K27ac separates active from poised enhancers and predicts developmental state.** *Proc Natl Acad Sci U S A* 2010, **107**:21931–21936.

12. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J: **A unique chromatin signature uncovers early developmental enhancers in humans.** *Nature* 2011, **470**:279–283.
13. Zentner GE, Tesar PJ, Scacheri PC: **Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions.** *Genome Res* 2011, **21**:1273–1283.
14. Taylor GC, Eskeland R, Hekimoglu-Balkan B, Pradeepa MM, Bickmore WA: **H4K16 acetylation marks active genes and enhancers of embryonic stem cells, but does not alter chromatin compaction.** *Genome Res* 2013, **23**:2053–2065.
15. Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, Markenscoff-Papadimitriou E, Kuhl D, Bito H, Worley PF, Kreiman G, Greenberg ME: **Widespread transcription at neuronal activity-regulated enhancers.** *Nature* 2010, **465**:182–187.
16. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T, Ntini E, Arner E, Valen E, Li K, Schwarzfischer L, Glatz D, Raithe J, Lilje B, Rapin N, Bagger FO, Jørgensen M, Andersen PR, Bertin N, Rackham O, Burroughs AM, Baillie JK, Ishizu Y, Shimizu Y, Furuhashi E, Maeda S, et al: **An atlas of active enhancers across human cell types and tissues.** *Nature* 2014, **507**:455–461.
17. Birnbaum RY, Clowney EJ, Agamy O, Kim MJ, Zhao J, Yamanaka T, Pappalardo Z, Clarke SL, Wenger AM, Nguyen L, Gurrieri F, Everman DB, Schwartz CE, Birk OS, Bejerano G, Lomvardas S, Ahituv N: **Coding exons function as tissue-specific enhancers of nearby genes.** *Genome Res* 2012, **22**:1059–1068.
18. Stergachis AB, Haugen E, Shafer A, Wenqing F, Vernot B, Reynolds A, Raubitschek A, Ziegler S, LeProust EM, Akey JM, Stamatoyannopoulos JA: **Exonic transcription factor binding directs codon choice and affects protein evolution.** *Science* 2013, **342**:1367–1372.
19. Tumpel S, Cambroero F, Sims C, Krumlauf R, Wiedemann LM: **A regulatory module embedded in the coding region of Hoxa2 controls expression in rhombomere 2.** *Proc Natl Acad Sci U S A* 2008, **105**:20077–20082.
20. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, Shafer A, Neri F, Lee K, Kutayin T, Stehling-Sun S, Johnson AK, Canfield TK, Giste E, Diegel M, Bates D, Hansen RS, Neph S, Sabo PJ, Heimfeld S, Raubitschek A, Ziegler S, Cotsapas C, Sotoodehnia N, Glass I, Sunyaev SR et al: **Systematic localization of common disease-associated variation in regulatory DNA.** *Science* 2012, **337**:1190–1195.
21. Ernst J, Kheradpour P, Mikkelson TS, Shores N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M, Bernstein BE: **Mapping and analysis of chromatin state dynamics in nine human cell types.** *Nature* 2011, **473**:43–49.
22. Trynka G, Sandor C, Han B, Xu H, Stranger BE, Liu XS, Raychaudhuri S: **Chromatin marks identify critical cell types for fine mapping complex trait variants.** *Nat Genet* 2013, **45**:124–130.
23. Pasquali L, Gaulton KJ, Rodriguez-Segui SA, Mularoni L, Miguel-Escalada I, Akerman I, Tena JJ, Morán I, Gómez-Marín C, van de Bunt M, Ponsa-Cobas J, Castro N, Nammo T, Cebola I, García-Hurtado J, Maestro MA, Pattou F, Piemonti L, Berney T, Gloyn AL, Ravassard P, Gómez-Skarmeta JL, Müller F, McCarthy MI, Ferrer J: **Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants.** *Nat Genet* 2014, **46**:136–143.
24. Parker SC, Stitzel ML, Taylor DL, Orozco JM, Erdos MR, Akiyama JA, van Bueren KL, Chines PS, Narisu N, Black BL, Visel A, Pennacchio LA, Collins FS, NISC Comparative Sequencing Program, National Institutes of Health Intramural Sequencing Center Comparative Sequencing Program Authors, NISC Comparative Sequencing Program Authors: **Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants.** *Proc Natl Acad Sci U S A* 2013, **110**:17921–17926.
25. Stitzel ML, Sethupathy P, Pearson DS, Chines PS, Song L, Erdos MR, Welch R, Parker SC, Boyle AP, Scott LJ, Margulies EH, Boehnke M, Furey TS, Crawford GE, Collins FS, NISC Comparative Sequencing Program: **Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci.** *Cell Metab* 2010, **12**:443–455.
26. Akhtar-Zaidi B, Cowper-Sal-lari R, Corradin O, Saiakhova A, Bartels CF, Balasubramanian D, Myeroff L, Lutterbaugh J, Jarrar A, Kalady MF, Willis J, Moore JH, Tesar PJ, Laframboise T, Markowitz S, Lupien M, Scacheri PC: **Epigenomic enhancer profiling defines a signature of colon cancer.** *Science* 2012, **336**:736–739.
27. Alcina A, Fedetz M, Fernández Ó, Saiz A, Izquierdo G, Lucas M, Leyva L, García-León JA, Abad-Grau Mdel M, Alloza I, Antigüedad A, García-Barcina MJ, Vandenbroeck K, Varadé J, de la Hera B, Arroyo R, Comabella M, Montalban X, Petit-Marty N, Navarro A, Otaegui D, Olasoaga J, Blanco Y, Urcelay E, Matesanz F: **Identification of a functional variant in the KIF5A-CYP27B1-METTL1-FAM119B locus associated with multiple sclerosis.** *J Med Genet* 2012, **50**:25–33.
28. Bauer DE, Kamran SC, Lessard S, Xu J, Fujiwara Y, Lin C, Shao Z, Canver MC, Smith EC, Pinello L, Sabo PJ, Vierstra J, Voit RA, Yuan GC, Porteus MH, Stamatoyannopoulos JA, Lettrec G, Orkin SH: **An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level.** *Science* 2013, **342**:253–257.
29. Miller CL, Anderson DR, Kundu RK, Raiesdana A, Nurnberg ST, Diaz R, Cheng K, Leeper NJ, Chen CH, Chang IS, Schadt EE, Hsiung CA, Assimes TL, Quertermous T: **Disease-related growth factor and embryonic signaling pathways modulate an enhancer of TCF21 expression at the 6q23.2 coronary heart disease locus.** *PLoS Genet* 2013, **9**:e1003652.
30. Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, Gomez-Marín C, Aneas I, Credidio FL, Sobreira DR, Wasserman NF, Lee JH, Puvindran V, Tam D, Shen M, Son JE, Vakili NA, Sung HK, Naranjo S, Acemel RD, Manzanares M, Nagy A, Cox NJ, Hui CC, Gomez-Skarmeta JL, Nóbrega MA: **Obesity-associated variants within FTO form long-range functional connections with IRX3.** *Nature* 2014, **507**:371–375.
31. Harismendy O, Notani D, Song X, Rahim NG, Tanasa B, Heintzman N, Ren B, Fu XD, Topol EJ, Rosenfeld MG, Frazer KA: **9p21 DNA variants associated with coronary artery disease impair interferon-γ signalling response.** *Nature* 2011, **470**:264–268.
32. Kapoor A, Sekar RB, Hansen NF, Fox-Talbot K, Morley M, Pihur V, Chatterjee S, Brandimarto J, Moravec CS, Pulit SL, Pfeufer A, Mullikin J, Ross M, Green ED, Bentley D, Newton-Cheh C, Boerwinkle E, Tomaselli GF, Cappola TP, Arking DE, Halushka MK, Chakravarti A, QT Interval-International GWAS Consortium: **An enhancer polymorphism at the cardiomyocyte intercalated disc protein NOS1AP locus is a major regulator of the QT interval.** *Am J Hum Genet* 2014, **94**:854–869.
33. Guenther CA, Tasic B, Luo L, Bedell MA, Kingsley DM: **A molecular basis for classic blond hair color in Europeans.** *Nat Genet* 2014, **46**:748–752.
34. Pomerantz MM, Ahmadiyeh N, Jia L, Herman P, Verzi MP, Doddapaneni H, Beckwith CA, Chan JA, Hills A, Davis M, Yao K, Kehoe SM, Lenz HJ, Haiman CA, Yan C, Henderson BE, Frenkel B, Barretina J, Bass A, Taberero J, Baselga J, Regan MM, Manak JR, Shivdasani R, Coetzee GA, Freedman ML: **The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer.** *Nat Genet* 2009, **41**:882–884.
35. Wasserman NF, Aneas I, Nobrega MA: **An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a MYC enhancer.** *Genome Res* 2010, **20**:1191–1197.
36. Tuupanen S, Yan J, Turunen M, Gylfe AE, Kaasinen E, Li L, Eng C, Culver DA, Kalady MF, Pennison MJ, Pasche B, Manne U, de la Chapelle A, Hampel H, Henderson BE, Le Marchand L, Hautaniemi S, Askthorab H, Smoot D, Sandler RS, Keku T, Kupfer SS, Ellis NA, Haiman CA, Taipale J, Aaltonen LA: **Characterization of the colorectal cancer-associated enhancer MYC-335 at 8q24: the role of rs67491583.** *Cancer Genet* 2012, **205**:25–33.
37. Spieler D, Kaffe M, Knauf F, Bessa J, Tena JJ, Giesert F, Schormair B, Tilch E, Lee H, Horsch M, Czamara D, Karbalai N, von Toerne C, Waldenberger M, Gieger C, Lichtner P, Claussnitzer M, Naumann R, Müller-Myhsok B, Torres M, Garrett L, Rozman J, Klingenspor M, Gailus-Durner V, Fuchs H, Hrabec de Angelis M, Beckers J, Höltner SM, Meitinger T, Hauck SM, et al: **Restless legs syndrome-associated intronic common variant in Meis1 alters enhancer function in the developing telencephalon.** *Genome Res* 2014, **24**:592–603.
38. Cowper-Sal-lari R, Zhang X, Wright JB, Bailey SD, Cole MD, Eeckhoutte J, Moore JH, Lupien M: **Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression.** *Nat Genet* 2012, **44**:1191–1198.
39. Wright JB, Brown SJ, Cole MD: **Upregulation of c-MYC in cis through a large chromatin loop linked to a cancer risk-associated single-nucleotide polymorphism in colorectal cancer cells.** *Mol Cell Biol* 2010, **30**:1411–1420.
40. Wang S, Wen F, Wiley GB, Kinter MT, Gaffney PM: **An enhancer element harboring variants associated with systemic lupus erythematosus engages the TNFAIP3 promoter to influence A20 expression.** *PLoS Genet* 2013, **9**:e1003750.
41. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA: **Master transcription factors and mediator establish super-enhancers at key cell identity genes.** *Cell* 2013, **153**:307–319.
42. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, Hoke HA, Young RA: **Super-enhancers in the control of cell identity and disease.** *Cell* 2013, **155**:934–947.



43. Zhang X, Cowper-Salari R, Bailey SD, Moore JH, Lupien M: **Integrative functional genomics identifies an enhancer looping to the SOX9 gene disrupted by the 17q24.3 prostate cancer risk locus.** *Genome Res* 2012, **22**:1437–1446.
44. Grice EA, Rochelle ES, Green ED, Chakravarti A, McCallion AS: **Evaluation of the RET regulatory landscape reveals the biological relevance of a HSCR-implicated enhancer.** *Hum Mol Genet* 2005, **14**:3837–3845.
45. Emison ES, McCallion AS, Kashuk CS, Bush RT, Grice E, Lin S, Portnoy ME, Cutler DJ, Green EG, Chakravarti A: **A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk.** *Nature* 2005, **434**:857–863.
46. Weedon MN, Cebola I, Patch AM, Flanagan SE, De Franco E, Caswell R, Rodriguez-Segui SA, Shaw-Smith C, Cho CH, Lango Allen H, Houghton JA, Roth CL, Chen R, Hussain K, Marsh P, Vallier L, Murray A, Ellard S, Ferrer J, Hattersley AT, International Pancreatic Agenesis Consortium: **Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis.** *Nat Genet* 2014, **46**:61–64.
47. Furniss D, Lettice LA, Taylor IB, Critchley PS, Giele H, Hill RE, Wilkie AO: **A variant in the sonic hedgehog regulatory sequence (ZRS) is associated with triphalangeal thumb and deregulates expression in the developing limb.** *Hum Mol Genet* 2008, **17**:2417–2423.
48. Lettice LA: **Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits.** *Hum Mol Genet* 2003, **12**:1725–1735.
49. Kiousis D, Vanin E, deLange T, Flavell RA, Grosveld FG:  **$\beta$ -Globin gene inactivation by DNA translocation in  $\gamma\beta$ -thalassaemia.** *Nature* 1983, **306**:662–666.
50. Semenza GL, Delgrosso K, Poncz M, Malladi P, Schwartz E, Surrey S: **The silent carrier allele:  $\beta$  thalassaemia without a mutation in the  $\beta$ -globin gene or its immediate flanking regions.** *Cell* 1984, **39**:123–128.
51. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S, Cherry JM, Snyder M: **Annotation of functional variation in personal genomes using RegulomeDB.** *Genome Res* 2012, **22**:1790–1797.
52. Ward LD, Kellis M: **HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants.** *Nucleic Acids Res* 2012, **40**:D930–D934.
53. Coetzee SG, Rhie SK, Berman BP, Coetzee GA, Noushmehr H: **FunciSNP: an R/bioconductor tool integrating functional non-coding data sets with genetic association studies to identify candidate regulatory SNPs.** *Nucleic Acids Res* 2012, **40**:e139.
54. Khurana E, Fu Y, Colonna V, Mu XJ, Kang HM, Lappalainen T, Sboner A, Lochovsky L, Chen J, Harmanci A, Das J, Abyzov A, Balasubramanian S, Beal K, Chakravarty D, Challis D, Chen Y, Clarke D, Clarke L, Cunningham F, Evani US, Flicek P, Fragoza R, Garrison E, Gibbs R, Gümüs ZH, Herrero J, Kitabayashi N, Kong Y, Lage K, et al: **Integrative annotation of variants from 1092 humans: application to cancer genomics.** *Science* 2013, **342**:1235587.
55. Sanyal A, Lajoie BR, Jain G, Dekker J: **The long-range interaction landscape of gene promoters.** *Nature* 2012, **489**:109–113.
56. Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, Yen CA, Schmitt AD, Espinoza CA, Ren B: **A high-resolution map of the three-dimensional chromatin interactome in human cells.** *Nature* 2013, **503**:290–294.
57. Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J, Sim HS, Peh SQ, Mulawadi FH, Ong CT, Orlov YL, Hong S, Zhang Z, Landt S, Raha D, Euskirchen G, Wei CL, Ge W, Wang H, Davis C, Fisher-Aylor KI, Mortazavi A, Gerstein M, Gingeras T, Wold B, Sun Y, et al: **Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation.** *Cell* 2012, **148**:84–98.
58. Zhang Y, Wong CH, Birnbaum RY, Li G, Favaro R, Ngan CY, Lim J, Tai E, Poh HM, Wong E, Mulawadi FH, Sung WK, Nicolis S, Ahituv N, Ruan Y, Wei CL: **Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations.** *Nature* 2013, **504**:306–310.
59. Bickmore WA: **The spatial organization of the human genome.** *Ann Rev Genomics Hum Genet* 2013, **14**:67–84.
60. Williamson I, Eskeland R, Lettice LA, Hill AE, Boyle S, Grimes GR, Hill RE, Bickmore WA: **Anterior-posterior differences in HoxD chromatin topology in limb development.** *Development* 2012, **139**:3157–3167.
61. Dekker J, Rippe K, Dekker M, Kleckner N: **Capturing chromosome conformation.** *Science* 2002, **295**:1306–1311.
62. Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, van Steensel B, de Laat W: **Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C).** *Nat Genet* 2006, **38**:1348–1354.
63. de Wit E, de Laat W: **A decade of 3C technologies: insights into nuclear organization.** *Genes Dev* 2012, **26**:11–24.
64. Sajan SA, Hawkins RD: **Methods for identifying higher-order chromatin structure.** *Ann Rev Genomics Hum Genet* 2012, **13**:59–82.
65. Panne D: **The enhanceosome.** *Curr Opin Struct Biol* 2008, **18**:236–242.
66. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, Garg K, John S, Sandstrom R, Bates D, Boatman L, Canfield TK, Diegel M, Dunn D, Ebersol AK, Frum T, Giste E, Johnson AK, Johnson EM, Kutayin T, Lajoie B, Lee BK, Lee K, London D, Lotakis D, Neph S, et al: **The accessible chromatin landscape of the human genome.** *Nature* 2012, **489**:75–82.
67. Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, Wagner U, Dixon J, Lee L, Lobanenkov VV, Ren B: **A map of the cis-regulatory sequences in the mouse genome.** *Nature* 2012, **488**:116–120.
68. Factor DC, Corradin O, Zentner GE, Saiakhova A, Song L, Chenoweth JG, McKay RD, Crawford GE, Scacheri PC, Tesar PJ: **Epigenomic comparison reveals activation of “seed” enhancers during transition from naive to primed pluripotency.** *Cell Stem Cell* 2014, **14**:854–863.
69. He B, Chen C, Teng L, Tan K: **Global view of enhancer-promoter interactome in human cells.** *Proc Natl Acad Sci U S A* 2014, **111**:E2191–E2199.
70. Sheffield NC, Thurman RE, Song L, Safi A, Stamatoyannopoulos JA, Lenhard B, Crawford GE, Furey TS: **Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions.** *Genome Res* 2013, **23**:777–788. <http://dnase.genome.duke.edu/>.
71. Fairfax BP, Makino S, Radhakrishnan J, Plant K, Leslie S, Dilthey A, Ellis P, Langford C, Vannberg FO, Knight JC: **Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles.** *Nat Genet* 2012, **44**:502–510.
72. Raj T, Rothamel K, Mostafavi S, Ye C, Lee MN, Replogle JM, Feng T, Lee M, Asinowski N, Frohlich I, Imboywa S, Von Korff A, Okada Y, Patsopoulos NA, Davis S, McCabe C, Paik HI, Srivastava GP, Raychaudhuri S, Hafler DA, Koller D, Regev A, Hacohen N, Mathis D, Benoist C, Stranger BE, De Jager PL: **Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes.** *Science* 2014, **344**:519–523.
73. Patwardhan RP, Hiatt JB, Witten DM, Kim MJ, Smith RP, May D, Lee C, Andrie JM, Lee SI, Cooper GM, Ahituv N, Pennacchio LA, Shendure J: **Massively parallel functional dissection of mammalian enhancers in vivo.** *Nat Biotechnol* 2012, **30**:265–270.
74. Kheradpour P, Ernst J, Melnikov A, Rogov P, Wang L, Zhang X, Alston J, Mikkelsen TS, Kellis M: **Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay.** *Genome Res* 2013, **23**:800–811.
75. Shlyueva D, Stampfel G, Stark A: **Transcriptional enhancers: from properties to genome-wide predictions.** *Nat Rev Genet* 2014, **15**:272–286.
76. Gaulton KJ, Nammo T, Pasquali L, Simon JM, Giresi PG, Fogarty MP, Panhuis TM, Mieczkowski P, Secchi A, Bosco D, Berney T, Montanya E, Mohlke KL, Lieb JD, Ferrer J: **A map of open chromatin in human pancreatic islets.** *Nat Genet* 2010, **42**:255–259.
77. The ENCODE Project Consortium: **Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.** *Nature* 2007, **447**:799–816.
78. The International HapMap Consortium: **A second generation human haplotype map of over 3.1 million SNPs.** *Nature* 2007, **449**:851–861.
79. The International HapMap 3 Consortium: **Integrating common and rare genetic variation in diverse human populations.** *Nature* 2010, **467**:52–58.
80. Ongen H, Andersen CL, Bramsen JB, Oster B, Rasmussen MH, Ferreira PG, Sandoval J, Vidal E, Whiffin N, Planchon A, Padialeu I, Bielser D, Romano L, Tomlinson I, Houlston RS, Esteller M, Orntoft TF, Dermitzakis ET: **Putative cis-regulatory drivers in colorectal cancer.** *Nature* 2014, **512**:87–90.
81. Fairfax BP, Humberg P, Makino S, Naranbhai V, Wong D, Lau E, Jostins L, Plant K, Andrews R, McGee C, Knight JC: **Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression.** *Science* 2014, **343**:1246949.
82. Claussnitzer M, Dankel SN, Klocke B, Grallert H, Glunk V, Berulava T, Lee H, Oskolkov N, Fadista J, Ehlers K, Wahl S, Hoffmann C, Qian K, Rönn T, Riess H, Müller-Nurasyid M, Bretschneider N, Schroeder T, Skurk T, Horsthemke B, Spieler D, Klingenspor M, Seifert M, Kern MJ, Mejhert N, Dahlman I, Hansson O, Hauck SM, DIAGRAM + Consortium, et al: **Leveraging cross-species**



transcription factor binding site patterns: from diabetes risk loci to disease mechanisms. *Cell* 2014, **156**:343–358.

83. Kilpinen H, Waszak SM, Gschwind AR, Raghav SK, Witwicki RM, Orioli A, Migliavacca E, Wiederkehr M, Gutierrez-Arcelus M, Panousis NI, Yurovsky A, Lappalainen T, Romano-Palumbo L, Planchon A, Bielser D, Bryois J, Padioleau I, Udin G, Thurnheer S, Hacker D, Core LJ, Lis JT, Hernandez N, Reymond A, Deplancke B, Dermitzakis ET: **Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription.** *Science* 2013, **342**:744–747.

doi:10.1186/s13073-014-0085-3

**Cite this article as:** Corradin and Scacheri: **Enhancer variants: evaluating functions in common disease.** *Genome Medicine* 2014 **6**:85.