Sakabe *et al. Genome Biology* 2012, **13**:238
http://genomebiology.com/2012/13/1/238

Genome Biology

REVIEW

Transcriptional enhancers in development and disease

Noboru Jo Sakabe*†, Daniel Savic*† and Marcelo A Nobrega

Abstract

Distal transcription enhancers are *cis*-regulatory elements that promote gene expression, enabling spatiotemporal control of genetic programs such as those required in metazoan developmental processes. Because of their importance, their disruption can lead to disease.

Transcription regulation by distal enhancers

Gene expression patterns in metazoans range from widespread expression in multiple cell types, such as expression of genes required for the maintenance of basic cellular functions, to complex spatiotemporal expression of genes with pleiotropic functions. Examples of transcription factors (TFs) involved in such complex regulation are *PAX6*, which is crucial for development of the eye and also of sensory organs and specific neural and epidermal tissues; *Sonic Hedgehog* (*SHH*), which is involved in development of many systems, as diverse as limb and brain; and *TBX5*, which is involved in heart and forelimb development.

The precise and complex spatiotemporal expression of genes often requires the deployment of additional *cis*-regulatory elements, physically displaced from the promoter. These promoter-distal *cis*-regulatory elements bind TFs that are cell-lineage-specific and those that are expressed in the presence of external signals as hormones, for example, at specific time points such as differentiation or proliferation. By integrating different cues, these elements coordinate complex patterns of gene expression in different tissues and time points (Figure 1a).

Enhancers are a class of *cis*-regulatory elements that promote gene expression and often are essential for eliciting the complex expression patterns of developmental genes. These elements typically span a few hundred base pairs (bp) and are composed of clusters of transcription factor binding sites (6- to 20-bp motifs) to which combinations of trans-activating and repressive factors bind in sequence-specific manner. They can be located in intergenic regions, introns and exons, tens to hundreds of kilobases from their target genes ([1], reviewed in [2]).

Although these elements have been studied for decades through careful dissection of individual examples [3], the advent of genome-wide chromatin immunoprecipitation (ChIP), an experimental technique that locates DNA sites where specific proteins are bound (reviewed in [4]) enabled the largely unbiased identification of tens of thousands of putative elements in a single experiment and the discovery of global patterns that are shedding light on how enhancers act (reviewed in [5]). Studies using this technique have confirmed and expanded our appreciation of the importance of *cis*-regulatory elements during development and in adult function, changing the way we view gene regulation in metazoans.

Here we review recent findings, obtained mainly from genome-wide studies, of how enhancers are activated, the role of enhancer features in mammalian development, and the involvement of this class of *cis*-regulatory elements in disease. Although earlier discoveries have attributed enhancer variation to several human diseases (reviewed in [2,6]), these studies have been largely limited to rare Mendelian disorders, which commonly involve single gene disruptions and follow simple patterns of inheritance. We discuss the previously unappreciated role of promoter-distal *cis*-regulatory variation in common disease susceptibility from genome-wide association studies (GWASs) and discuss how a variety of genome annotations can additionally be exploited to expedite discovery of causal variants.

Enhancer activation

Enhancers are recognized by the cellular machinery through a combination of chromatin modifications and sequence-specific binding of TFs. Given that DNA is compacted into chromatin, enhancers must be localized to sites accessible to proteins, that is, in euchromatin regions with exposed DNA. However, enhancers are not

†These authors contributed equally
*Correspondence: dsavic@uchicago.edu; nsakabe@uchicago.edu
Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA

BioMed Central © 2012 BioMed Central Ltd

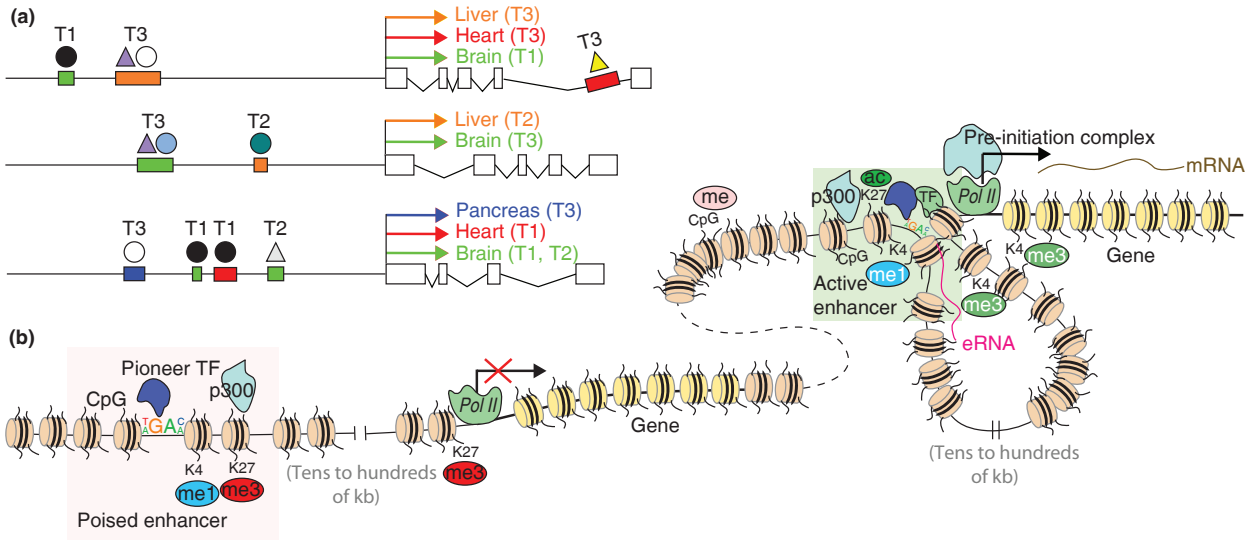

Figure 1. Enhancers in development and their structural features. (a) Enhancers (colored rectangles) modularly drive gene expression at particular time points and in particular tissues (arrows) by integrating inputs (transcription factors (circles), time points when they are expressed (T1, T2, T3) and cell lineages (colors)). For example, the orange enhancers promote expression of the top two genes in liver, but at different time points (T2 and T3). The enhancer in the top gene requires binding of two transcription factors for activation in T3. **(b)** Schematic representation of a segment of DNA wrapped in nucleosomes displaying enhancer genomic features. Histone modifications are represented by blue, green (activating) and red (repressive) ovals attached to nucleosomes. Poised and active enhancers display different histone modifications and are associated with different *trans*-factors. Poised enhancers without H3K27me3 have also been identified [16] but are not shown for clarity. Pioneer TFs pre-specify enhancers that will become active later. Unmethylated CpG islands can be found in pre-specified but inactive enhancers. Enhancers can be located tens to hundreds of kilobases away from genes; active enhancers interact with promoters looping out intervening chromatin. eRNAs are products of transcription of active enhancer sequences by RNA polymerase II (Pol II).

always accessible and may require appropriate stimuli to become 'open'. For example, chromatin containing distal enhancers that have become active has been shown to undergo dynamic nucleosome repositioning following T-cell activation [7], androgen receptor treatment [8] and erythrocyte differentiation [9]. These stimuli and other cellular processes cause nucleosome repositioning, which involves chromatin remodeling complexes such as BAF (reviewed in [10]). The specificity of these complexes to particular enhancers seems to be mediated by 'pioneer' factors, FOXA1 being the best characterized example (reviewed in [11]). These proteins bind to nucleosomal DNA, recruiting chromatin remodelers that facilitate chromatin opening and the subsequent binding of TFs [11].

The binding of chromatin remodelers might also involve chemical groups present in nucleosomes [12,13]. Histones, the proteins that constitute nucleosomes, can be dynamically modified (for example, acetylated, methylated or phosphorylated) at different residues (reviewed in [12]). The role of histone modifications in enhancer function is still unclear. One possibility is that the cell machinery recognizes a code of DNA elements based on combinations of histone modifications [13]. Given that there is a wide assortment of histone modifications, discovering those few that are sufficient to distinguish DNA elements and enhancer states is important. Indeed, recent studies using ChIP have uncovered genome-wide patterns that allow certain DNA elements to be distinguished (reviewed in [5]). For example, whereas trimethylation of lysine 4 of histone 3 (H3K4me3) is predominantly present in active promoters, distal enhancers are associated with monomethylation (H3K4me1) [14], which is largely tissue-specific [15,16].

H3K4me1 was largely accepted as a general enhancer marker, and several studies have used ChIP of H3K4me1 coupled with high-throughput sequencing to locate tens of thousands of distal enhancers in various cells and tissues (for example, [15,17,18]). However, it was found that not all H3K4me1 regions correspond to active enhancers [16-19]. A recent study demonstrated that presence of acetylation of lysine 27 of histone 3 (H3K27ac) was associated with active enhancers identified by H3K4me1 in several cell types, whereas a sub-population comprised seemingly inactive H3K4me1 regions that were devoid of this acetylation and were deemed 'poised' [16]. Histone acetylation is catalyzed by acetyltransferases, such as p300, which are recruited by bound TFs and thought to bind chromatin remodelers [13]. Given its role in enhancer activation, p300 has also been used to locate enhancers [14,15,18,20], but its presence may not distinguish between active and poised

enhancers [16,19], suggesting that factors other than the presence of acetyltransferases are necessary for enhancer activation.

In addition to the active enhancer mark H3K27ac, a recent study found that H3K4me3 is associated with enhancer activation [21], contrary to the widely accepted notion that H3K4me3 is mostly a promoter histone modification. Similarly to H3K27ac, distal enhancers marked by H3K4me1 that became active during T-cell differentiation gained H3K4me3, whereas inactive enhancers remained marked solely by H3K4me1.

The outcome of enhancer activation by acetylation, nucleosome repositioning and TF binding is gene transcription. Active enhancers are believed to initiate gene expression through physical interaction with their target promoters. The prevailing model proposes that they directly contact promoters by looping out intervening chromatin (Figure 1b; reviewed in [22]). This is demonstrated by techniques that allow the determination of physical interactions between segments in the genome, such as chromatin interaction analysis (ChIA) [23] and chromatin conformation capture (3C) and variants [24]. By contacting promoters, enhancers would supply *trans*-factors and activate transcription.

It has been recently shown that at least a fraction of active enhancers are transcribed by RNA polymerase II (Pol II), resulting in 'enhancer RNA' molecules (eRNA) [25,26]. It is unclear whether eRNAs have a regulatory role *per se* or whether they are simply a byproduct associated with Pol II recruitment produced when Pol II passes enhancers as it attempts to recruit methyl- and acetyltransferases [25,26]. Alternatively, assuming that enhancers directly interact with promoters, eRNAs could be the result of transcription of the wrong DNA sequence, with no biological function. This idea is consistent with the dependence of eRNAs on their target promoters, the correlation of eRNAs with mRNA levels and the bi-directionality of eRNA transcription [25,26]. Regardless of their function, eRNAs and presence of Pol II are useful in the identification of active enhancers, in addition to H3K27ac.

In summary, enhancers are epigenetically distinguishable from other DNA elements and undergo activation through chemical modification of specific histone residues, typically acetylation, catalyzed by acetyltransferases such as p300. Recruitment of chromatin remodelers that reposition nucleosomes, through binding either to acetyllysine groups or to pioneer factors that bind nucleosomal DNA, enables sequence-specific binding of TFs to DNA (Figure 1b).

Enhancers in development

Most enhancer features were initially described at developmental loci. One important reason for this identification bias is the dynamic nature of development. Indeed, comparisons between distinct developmental stages might reveal novel features not identifiable in a more static differentiated cell lineage. Later studies might also find some of these features in non-developmental enhancers, but it is possible that they are more frequent among developmental ones, given the complexity and variability of developmental processes.

One possible distinction of developmental enhancers is their enrichment in evolutionarily conserved sequences. Because of the functional importance of *cis*-regulatory elements in general, a significant proportion of enhancer sequences are evolutionarily conserved [27]. However, the conservation of developmental enhancers seems to be even more pronounced. This is alluded to by studies that found a biased association of transcription factors/developmental genes with both higher densities of conserved sequences [28] and the presence of sequences harboring particularly deep conservation [29,30]. More studies are needed to directly compare conservation levels of developmental and other enhancers to fully clarify this issue.

Another feature that might be particular to developmental enhancers is functional redundancy to ensure accurate expression. Shadow enhancers are regulatory elements that drive similar expression patterns to their primary enhancers [31] but together drive more faithful expression, especially under suboptimal conditions [32]. Although shadow enhancers were identified in *Drosophila*, they may not be exclusive to invertebrates, as redundant enhancers have also been observed in mammals [33,34]. However, it remains to be established whether non-developmental genes also rely on shadow enhancers.

The importance of enhancer pre-specification in development

Differentiation of pluripotent cells into terminally differentiated cell lineages involves the expression and repression of diverse gene sets not only through the deployment of tissue-specific TFs but also through the activation of enhancers. The specification of enhancers that will be active in specific tissues occurs during early development, well before the genes they control are expressed, when enhancers are poised or pre-specified by pioneer factors or epigenetic modifications.

Similarly to FOXA1, early binding of the TFs GATA1 [9] and CEBPA [35] has been observed in sites that became functional only upon differentiation. Terminally differentiated cells (macrophages) were also shown to harbor enhancers primed by a TF (SFPI1 or PU.1) and became active following antigen stimulation [18]. Epigenetic pre-specification involves hyperacetylation and windows of hypomethylated CpG dinucleotides, which were seen in tissue-specific enhancers in

embryonic stem cells (ESCs; reviewed in [36]), and possibly H3K4me1.

Pre-specification of enhancers ensures dynamic activation of developmental enhancers in pluripotent cells. During differentiation, specific sets of enhancers control distinct sets of genes in complex spatiotemporal patterns. Therefore, enhancers cannot be constitutively active and must be rapidly turned on or off at specific time points and within particular cell lineages.

Such readiness for activation during development was first observed in promoters and later in enhancers of ESCs. Promoters that concomitantly displayed both the active H3K4me3 and repressive H3K27me3 modifications and controlled genes with low expression levels were deemed poised for transcription [37,38]. These bivalent marks were proposed to be associated with genes expressed during development that would need to be quickly activated or repressed in different contexts, offering an attractive explanation as to how pluripotency is maintained at the genome level [37,38]. In differentiated cells, bivalent marks resolve into either the active H3K4me3 or repressive H3K27me3 [17,39]. Subsequent analyses demonstrated that distal enhancers repressed in ESCs but active in later development were also poised through association with the repressive histone modification H3K27me3, whereas active enhancers lost this modification and gained H3K27ac [19]. A later study also proposed H3K9me3 as a poising modification [40].

Molecular events that occur during early liver/pancreas differentiation are an interesting illustration of the importance of regulatory element poising or pre-specification during development. Endoderm cells are derived precursor cells that give origin to liver, pancreas, colon and other tissues. The choice between pancreas and liver differentiation seems to rely on pre-specification of regulatory elements by pioneer FOXA1 and GATA4 binding (reviewed in [36]) and also on a pre-established epigenetic pattern [41]. In endoderm cells, regulatory elements of the pancreatic determination gene *Pdx1* are poised, whereas elements of the liver-specific *Alb1* gene have low levels of histone modifications. The default fate, pancreas differentiation, is constitutively poised by the repressive H3K27me3 histone modification, whereas *de novo* acetylation of liver-specific elements allows the liver program to unroll. Interfering with the balance between the two types of modifications caused either pancreas or liver buds to spread beyond their original domains, demonstrating the importance of fine regulation of the enhancer's epigenetic state [41].

Although poised promoters and enhancers were initially identified in ESCs, later studies have also found them in more differentiated cells. Poised promoters were found in several tissues [42,43] and in T cells [44], mouse neural progenitor and embryonic fibroblast cells [39] and the human lung fibroblast cell line [45]. Poised enhancers were found in differentiated cells, such as pro-B cells and adult liver [16], 3T3L1 fibroblast-derived adipocytes and bone-marrow-derived macrophages [40]. However, the fact that poised promoters were more numerous in pluripotent cells [42] (no such quantification has yet been performed for enhancers) and no poising was found in tissue-specific enhancers [19] suggests that this mechanism might be more common during development, in line with its dynamic requirements.

Altogether, these observations reveal the sophistication of enhancer specification and activation throughout development. We are only beginning to comprehend how generalized these mechanisms are and understanding how they function in concert will require more analyses.

Unraveling developmental programs genome-wide

The possibility of mapping *cis*-regulatory elements genome-wide with ChIP allows the identification of thousands of genes controlled by a specific TF in a largely unbiased way. Although the roles of several TFs are established in various different developmental processes, knowledge about the networks they regulate is scarce. As part of the effort to fill this gap, one study performed ChIP of the transcription factor GLI1, a zinc finger protein, in mouse neural tube and revealed new GLI1-responsive enhancers and gene targets [46]. Another study targeted GLI3, an important limb development TF, and identified 5,000 new GLI3 binding sites and target genes, greatly enhancing our comprehension of this developmental program and illustrating the power of such genome-wide strategies [47].

Comparison of genes putatively bound by a given TF obtained from ChIP with expression data can improve the identification of active enhancers and the genes they control. One example of this application was a study of the role of EOMES in endoderm differentiation, which identified thousands of genes that are controlled by this TF and that were proposed to coordinate endoderm formation [48]. Applying these genome-wide methods to more TFs and at different developmental stages will allow us to obtain a more dynamic picture of these processes and quickly expand our comprehension of different developmental programs.

Enhancers and disease

Given the importance of regulatory elements during development, the misregulation of these sequences is likely to carry phenotypic consequences. Similar to protein-coding mutations, variation in enhancer elements has been previously attributed to several Mendelian disorders (reviewed in [2,6]). However, the functional impact of mutations in *cis*-regulatory elements

can differ significantly from that of protein-coding mutations, even if both are connected to the same gene. Mutations in enhancers are largely limited to *cis* effects on transcription, whereas those within protein-coding sequences can alter broader aspects of gene regulation, such as mRNA processing and stability, translation initiation and elongation or even protein structure and folding [49]. In addition, as *cis*-regulatory elements are modular and can act independently to regulate their target genes, disruptions to the regulatory elements are restricted to a spatial and temporal subset of the global function of the gene, and they are therefore predicted to result in a less detrimental effect than coding mutations, with which pleiotropic effects could be more prevalent [50,51].

Aside from these constraints, *cis*-regulatory mutations have the potential to generate a plethora of transcriptional alterations through both loss- and gain-of-function effects, leading to a gradient of phenotypic severities. A clear illustration of this is seen in the dysregulation of *SHH* expression and limb malformations. *SHH* expression in a region of limb buds known as the zone of polarizing activity (ZPA) is necessary for limb patterning [2]. This expression pattern is governed by a long-range enhancer element about 1 megabase from *SHH*, known as the ZPA regulatory sequence (ZRS). Point mutations within this element have been linked to a congenital disease leading to extra digits known as preaxial polydactyly [1], whereas deletion of the entire ZRS in mice led to a truncation of limbs [52].

Importantly, these phenotypic hallmarks are not exclusive to enhancer elements but encompass a broader range of regulatory sequences, as is highlighted by a mutation at the α-globin locus in the Melanesian population [53]. The regulatory mutation identified by De Gobbi *et al*. [53] produced a novel GATA1 binding site, leading to the formation of a promoter-like element within the locus that induced a decrease in expression of downstream α-globin genes, leading to α-thalassemia.

Cis-regulatory variation and the common disease common variant model

The aforementioned constraints on regulatory mutations suggest that these types of alterations have lower burdens on fitness than protein-coding mutations, enabling these regulatory variants to reach high frequencies in populations. Interestingly, this prediction is in line with the common disease common variant (CDCV) hypothesis, which postulates that common or complex diseases are caused by DNA sequence variations that are common in populations but that individually carry a modest effect on disease risk [54-57]. The CDCV model was developed to explain the high prevalence of diseases such as type 2 diabetes (T2D) and cardiovascular disease (CVD) that do not follow simple Mendelian patterns of inheritance. Consequently, these common diseases are believed to be polygenic (involving mutations in multiple genes) and the result of complex gene-environment interactions [56].

The CDCV model was one impetus for the use of GWASs to identify genetic predispositions to common diseases [58,59]. GWASs are conducted by genotyping naturally occurring bi-allelic sequence variations known as single nucleotide polymorphisms (SNPs) across the genome in case (with disease) and control (without disease) populations (reviewed in [58,60]). A statistically significant enrichment of one SNP allele in cases compared with controls identifies an interval associated with the disease (Figure 2a). As the genotyped SNP is not necessarily causal but merely tags a haplotype or linkage disequilibrium (LD) block (a sequence of DNA containing a group of SNP alleles that co-segregate), subsequent fine-mapping and additional functional strategies are used to localize the disease-causing variants within the associated interval (Figure 2b). So far, GWASs conducted on over 200 diseases or traits have cataloged over 1,400 associations, the vast majority of which await further characterization [61].

Although regulatory variation has been implicated in several Mendelian disorders (reviewed in [2,6,62]), not until recently has their contribution to common disease risk been extensively explored. The recent characterizations of GWAS intervals have not only confirmed a role, but further hint that *cis*-regulatory variation at enhancer sequences may be a general feature of common disease susceptibility.

Cis-regulatory variation in GWASs

It has been estimated that 40% of loci uncovered by GWASs are restricted to non-coding sequences [62]. This preponderance of non-coding sequence points to a potential role for regulatory variation in common disease predisposition. Indeed, although not all follow-ups to GWASs have implicated *cis*-regulatory alterations [63], several functional studies have uncovered non-coding elements within GWAS intervals that harbor variants associated with several common diseases (Table 1).

Loci at 1p13 and 9p21 have been associated with CVD [64,65]. Through both *in vitro* analyses and animal models, a SNP at the 1p13 locus was identified that altered a CEBPA binding site that regulated *SORT1* expression, thereby uncovering a novel role for this gene in hepatic lipoprotein metabolism [66]. In a follow-up study investigating the association with the 9p21 region, Visel *et al.* [67] demonstrated changes in the cardiac expression of two nearby cyclin-dependent kinase inhibitor genes (*Cdkn2a* and *Cdkn2b*) through the deletion of the association interval in mice. Interestingly, smooth muscle cultures from these mice had phenotypic hallmarks reported in coronary artery disease [67]. The 9p21

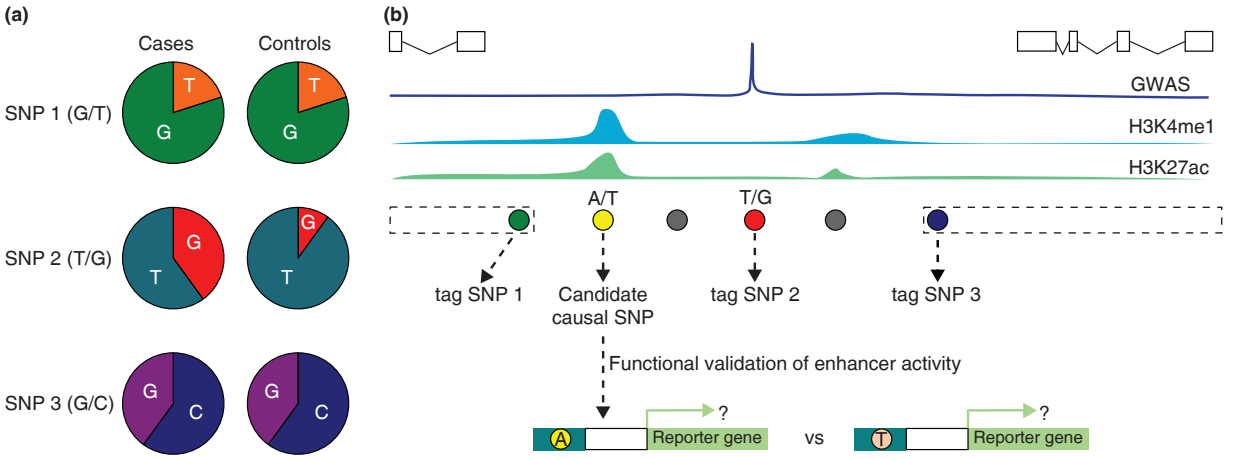

Figure 2. Genome-wide studies identify disease-associated alleles. (a) GWASs compare the allelic frequencies of SNPs across the entire genome in case and control populations. A statistically significant difference in the allelic frequencies between cases and controls constitutes an association with disease. In this example, a guanine (G) at SNP2 is associated with the disease. **(b)** Only a fraction of all known SNPs that are sufficient to tag haplotypes (gray and dashed boxes) are genotyped. The actual causal allele can be another SNP (yellow circle) in the same haplotype block. In the example, the SNP might not be a non-coding variation affecting expression of the gene shown at the top left. Experimental evidence can be used to identify regulatory sequences that harbor SNPs and that are thus likely to have a role in the disease, as illustrated by the H3K27ac and H3K4me1 signals. Functional tests of the candidate sequences spanning the putative disease SNP are performed to identify allelic-specific effects on gene expression using reporter genes.

interval was further shown to harbor 33 enhancers, and disease-associated variation within one enhancer caused the disruption of a STAT1-binding site involved in the interferon-γ (IFN-γ) inflammatory response [68]. Induction of the IFN-γ response in cell lines generated reciprocal changes in expression of *CDKN2B* and *CDKN2B* antisense RNA 1 (*CDKN2BAS*) [68].

Regulatory variation has also been implicated in metabolic disease. The association at the *TCF7L2* locus is the strongest predictor of T2D risk in the human population [69-71]. Using mouse transgenic assays, Savic *et al.* [72] uncovered a variety of *TCF7L2* enhancers within sequences spanning the association interval. Selective deletion of this associated region led to a marked reduction of enhancer activities [72]. Additional functional analyses demonstrated that a repetitive sequence spanning the strongest associated SNP at the *TCF7L2* locus showed allelic-specific enhancer activity in pancreatic beta cell lines [73,74].

Several cancer susceptibility loci have been identified through GWASs. Colorectal cancer susceptibility loci were uncovered at 18q21 [75] and 8q23.3 [76], and the 8q24 region has been implicated in multiple cancers [77-79]. A putative causal variant was fine-mapped and found flanking a conserved non-coding sequence at the 18q21 interval [80]. Sequences encompassing both the conserved region and the SNP displayed allelic-specific enhancer function within the colorectum of *Xenopus laevis* and this element was further proposed to target the neighboring *SMAD7* gene [80]. At the 8p23.3 locus, an associated variant was localized to a transcriptional repressor element that directly acted on the promoter of the nearby eukaryotic translation initiation factor 3, subunit H (*EIF3H*) gene [81]. *In vitro* assays using human colorectal cell lines defined allelic-specific alterations in repressor function [81].

The 8q24 locus upstream harbors intervals independently associated with prostate [77], colorectal [78] and breast cancers [79]. Extensive functional follow-ups have demonstrated that this locus contains regulatory sequences that maintain long-range interactions with the downstream oncogene *MYC* [82,83]. Four studies identified a *MYC* enhancer containing a SNP associated with both colorectal and prostate cancers [83-86]. This SNP disrupted a TCF7L2 binding site [83-85] and demonstrated allelic-specific enhancer properties in Wnt-responsive cell lines [84], colorectal cell lines [85] and mouse prostates [86]. Another investigation uncovered a second enhancer at this locus harboring a regulatory variant implicated in prostate cancer risk [87]. The enhancer was found to be androgen-responsive and the SNP altered the binding of FOXA1 in a prostate cancer cell line [87]. These functional data suggest that predisposition to multiple cancers at the 8q24 locus may use a common mechanism through regulatory variations that lead to alterations in *MYC* expression and potentially the expression of other neighboring genes.

Table 1. Summary of functional studies identifying *cis*-regulatory elements in GWAS intervals

Putative regulatory SNP(s)	Locus	Putative target gene	Protein function	Disease	Reference
rs12740374	1p13	*SORT1*	Intracellular trafficking and endocytosis	Cardiovascular disease	[65]
rs10811656, rs10757278	9p21	*CDKN2A* and *CDKN2B*	Cellular proliferation and senescence	Cardiovascular disease	[66,67]
rs7903146	10q25.2	*TCF7L2*	Transcriptional regulator of cell fate, survival and proliferation	Type 2 diabetes	[71-73]
Novel 1	18q21	*SMAD7*	Cell signaling antagonist	Colorectal cancer	[79]
rs16888589	8q23.3	*EIF3H*	Protein synthesis	Colorectal cancer	[80]
rs6983267	8q24	*MYC*	Transcriptional regulator of cell fate, survival and proliferation	Colorectal cancer	[81-84]
rs6983267, rs11986220	8q24	*MYC*	Transcriptional regulator of cell fate, survival and proliferation	Prostate cancer	[81,82,85,86]

Collectively, the identification of common *cis*-regulatory variation leading to complex disease susceptibility supports the notion these variants can 'compartmentalize' phenotypic effects, ensuring that effects on one tissue or cell type is separable from effects on another. This enables these polymorphisms to reach appreciable frequencies in populations, as was predicted by the CDCV model.

Genome-wide annotations in GWAS functional analyses

Despite several successful post-GWAS investigations, a plethora of GWAS loci await characterization. Although this is a daunting task, these functional follow-ups can be greatly expedited through the use of ever increasing numbers of genome-wide maps for a variety of annotations, such as sequence conservation and variation as well as chromatin and TF binding profiles in diverse cell lines and states.

By exploiting an expanding array of whole genome sequences, annotations of sequence conservation at finer resolutions are being generated. For example, sequence comparisons using 29 eutherian species uncovered selectively constrained sequences comprising 4.2% of the human genome and the further classification of about 60% of these elements [88]. Alongside sequence conservation, the 1000 Genomes Project aims to generate a richer catalog of sequence and structural variation using next-generation sequencing technologies, with the particular aim of targeting rare variants [89]. The applicability of these annotations to post-GWAS analyses stems from their ability to aid in both the fine-mapping and the prioritization of non-coding SNPs to pursue with subsequent functional assays.

Ongoing collective efforts such as the ENCODE project have generated genome-wide maps of histone modifications, TF binding and DNase hypersensitivity in a variety of cell lines [90]. Each of these methods can identify regulatory elements in the non-coding genome. For instance, the epigenetic signatures on histone tails can distinguish diverse *cis*-regulatory sequences such as promoters and enhancers; by mapping and combining a subset of these modifications in 9 human cells, 15 chromatin states were delineated [91]. Moreover, TF binding maps are routinely combined to identify non-coding sequences enriched for multiple binding events that represent putative regulatory elements. Unlike the methods that use ChIP-based technologies, DNase hypersensitivity mapping capitalizes on the marked depletion of nucleosomes at active regulatory sequences, rendering these regions susceptible to digestion [92]. Targeting these genomic catalogs to GWAS disease intervals provides a means for uncovering regulatory sequences that warrant further investigation.

With proper application, these diverse genomic repositories can serve as powerful toolkits for the functional characterization of GWAS loci, accelerating causal variant discovery. Although the annotations can be effective individually, employing them synergistically will provide the most benefit as they can identify regulatory sequences and potential functional variants within such sequences. As these genomic maps will undoubtedly grow, their usage in post-GWAS analyses will become increasingly common and essential.

Future directions

Enhancers are specified and activated through complex mechanisms involving epigenetic modifications and TF binding. By acting as independent gene switches that respond to different cues, enhancers regulate complex expression patterns. Their disruption may cause perturbations in gene expression, leading to disease. Although their role in Mendelian disorders has been previously established, recent GWAS functional analyses have extended their contributions as the underlying cause of several common diseases.

Although a number of studies have greatly expanded our knowledge on the role of enhancers in development, we are only beginning to understand genome-wide aspects of enhancer function. Unraveling developmental

programs on a genome scale will require not only mapping enhancers, but also elucidating the TFs that regulate them. Efforts using ChIP to map binding sites of known TFs in a few cell lineages have been carried out, but more studies will be necessary. Current large-scale projects such as ENCODE will greatly augment the field in this respect [90]. Comparisons of these maps across developmental stages will deliver a full account of how transcriptional regulation during development unrolls.

Given that ChIP is limited to known TFs, it will be necessary to couple proteomics techniques with ChIP to identify partnering TFs *ab initio* in specific contexts [93]. Such studies will expand the number of known pioneer factors and TFs, revealing as-yet unknown regulatory networks. In addition, studies analyzing the phenotypic impact of TF knockouts will be needed to functionally characterize new regulators.

As hinted by the discovery that enhancers can be active or poised, it is likely that enhancers are less mechanistically and functionally homogeneous than we think, and identifying and understanding their different subgroups will be required for a full comprehension of the role of enhancers in gene regulation.

Despite a trend for finding regulatory variation in common disease susceptibilities, additional regulatory mechanisms besides direct sequence variation may be important. As epigenetic states are crucial to *cis*-regulatory function, the misregulation of histone modifications or DNA methylation at an enhancer element could also lead to disease, even if there is no genetic mutation within the element *per se*. This could result in alterations of enhancer accessibility that induce fluctuations in gene expression, similar to the effects of direct sequence variants. In this case, indirect genetic mutations in *trans*-factors or even environmental perturbations could have a role. For example, poor maternal nutrition during gestation in rodents led to epigenetic misregulation at a *Hnf4a* enhancer element in offspring, generating perturbations in promoter-enhancer communication and lowered *Hnf4a* transcriptional output in pancreatic islets [94]. These environmentally mediated transgenerational epigenetic changes have also been demonstrated for hippocampal expression of the glucocorticoid receptor gene (*Nr3c1*), which is involved in stress response [95]. The duration of maternal care affected the degree of DNA methylation, histone acetylation and binding of the nerve growth factor inducible TF NGFI-A at the *Nr3c1* promoter in rodent offspring. Indeed, similar epigenetic modifications have been identified and correlated with childhood abuse at the orthologous promoter in humans [96]. Although these studies have suggested a novel level of regulatory control carrying phenotypic consequences, a more systematic and agnostic strategy is necessary to address the prevalence of such mechanisms on a broader, genome-wide scale. Proposed epigenome-wide association studies may provide much needed information on such regulatory alterations [61].

Although GWASs have been successful in mapping disease-associated regions, the collective genetic effects of these loci contribute only a minority to the heritability of common disease risk, warranting the use of different strategies to identify additional disease variants [97]. Current next-generation sequencing efforts may define rarer, more deleterious *cis*-regulatory mutations, and analyses of structural variation could further uncover genetic disruptions spanning regulatory elements leading to common disease susceptibility.

A combination of these efforts will produce a clearer picture of the role of *cis*-regulatory elements in development as well as their contributions to common disease risk.

Acknowledgments

This work is partially support by grants HG004428, HL088393 and DK078871 (MAN). NJS is a recipient of an American Heart Association post-doctoral fellowship. DS was supported by NIH Genetics and Regulation Training Grant T32GM007197.

Competing interests

The authors declare no competing interests.

Published: 23 January 2012

References

1. Lettice LA, Heaney SJH, Purdie LA, Li L, de Beer P, Oostra BA, Goode D, Elgar G, Hill RE, de Graaff E: **A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly.** *Hum Mol Genet* 2003, **12:**1725-1735.
2. Kleinjan DA, van Heyningen V: **Long-range control of gene expression: emerging mechanisms and disruption in disease.** *Am J Hum Genet* 2005, **76:**8-32.
3. Davidson EH: *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution.* Pasadena: Academic Press; 2006.
4. Kim TH, Ren B: **Genome-wide analysis of protein-DNA interactions.** *Annu Rev Genomics Hum Genet* 2006, **7:**81-102.
5. Sakabe NJ, Nobrega MA: **Genome-wide maps of transcription regulatory elements.** *Wiley Interdiscip Rev Syst Biol Med* 2010, **2:**422-437.
6. Noonan JP, McCallion AS: **Genomics of long-range regulatory elements.** *Annu Rev Genomics Hum Genet* 2010, **11:**1-23.
7. Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, Wei G, Zhao K: **Dynamic regulation of nucleosome positioning in the human genome.** *Cell* 2008, **132:**887-898.
8. He HH, Meyer CA, Shin H, Bailey ST, Wei G, Wang Q, Zhang Y, Xu K, Ni M, Lupien M, Mieczkowski P, Lieb JD, Zhao K, Brown M, Liu XS: **Nucleosome dynamics define transcriptional enhancers.** *Nat Genet* 2010, **42:**343-347.
9. Hu G, Schones DE, Cui K, Ybarra R, Northrup D, Tang Q, Gattinoni L, Restifo NP, Huang S, Zhao K: **Regulation of nucleosome landscape and transcription factor targeting at tissue-specific enhancers by BRG1.** *Genome Res* 2011, **21:**1650-1658.
10. Hargreaves DC, Crabtree GR: **ATP-dependent chromatin remodeling: genetics, genomics and mechanisms.** *Cell Res* 2011, **21:**396-420.
11. Kaestner KH: **The FoxA factors in organogenesis and differentiation.** *Curr Opin Genet Dev* 2010, **20:**527-532.
12. Ruthenburg AJ, Li H, Patel DJ, Allis CD: **Multivalent engagement of chromatin modifications by linked binding modules.** *Nat Rev Mol Cell Biol* 2007, **8:**983-994.
13. Jenuwein T, Allis CD: **Translating the histone code.** *Science* 2001, **293:**1074-1080.
14. 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.
15. 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.
16. Creyghton 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.
17. Cui K, Zang C, Roh T-Y, Schones DE, Childs RW, Peng W, Zhao K: **Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation.** *Cell Stem Cell* 2009, **4:**80-93.
18. Ghisletti S, Barozzi I, Mietton F, Polletti S, De Santa F, Venturini E, Gregory L, Lonie L, Chew A, Wei CL, Ragoussis J, Natoli G: **Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages.** *Immunity* 2010, **32:**317-328.
19. 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.
20. Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Ren B, Rubin EM, Pennacchio LA: **ChIP-seq accurately predicts tissue-specific activity of enhancers.** *Nature* 2009, **457:**854-858.
21. Pekowska A, Benoukraf T, Zacarias-Cabeza J, Belhocine M, Koch F, Holota H, Imbert J, Andrau JC, Ferrier P, Spicuglia S: **H3K4 tri-methylation provides an epigenetic signature of active enhancers.** *EMBO J* 2011, **30:**4198-4210.
22. Zhao H, Dean A: **Organizing the genome: enhancers and insulators.** *Biochem Cell Biol* 2005, **83:**516-524.
23. Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, Orlov YL, Velkov S, Ho A, Mei PH, Chew EG, Huang PY, Welboren WJ, Han Y, Ooi HS, Ariyaratne PN, Vega VB, Luo Y, Tan PY, Choy PY, Wansa KD, Zhao B, Lim KS, Leow SC, Yow JS, Joseph R, Li H, Desai KV, Thomsen JS, Lee YK, *et al.*: **An oestrogen-receptor-alpha-bound human chromatin interactome.** *Nature* 2009, **462:**58-64.
24. Simonis M, Kooren J, de Laat W: **An evaluation of 3C-based methods to capture DNA interactions.** *Nat Methods* 2007, **4:**895-901.
25. 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.
26. De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, Muller H, Ragoussis J, Wei CL, Natoli G: **A large fraction of extragenic RNA pol II transcription sites overlap enhancers.** *PLoS Biol* 2010, **8:**e1000384.
27. Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD, Plajzer-Frick I, Akiyama J, De Val S, Afzal V, Black BL, Couronne O, Eisen MB, Visel A, Rubin EM: **In vivo enhancer analysis of human conserved non-coding sequences.** *Nature* 2006, **444:**499-502.
28. Sironi M, Menozzi G, Comi GP, Cagliani R, Bresolin N, Pozzoli U: **Analysis of intronic conserved elements indicates that functional complexity might represent a major source of negative selection on non-coding sequences.** *Hum Mol Genet* 2005, **14:**2533-2546.
29. Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, Vavouri T, Smith SF, North P, Callaway H, Kelly K, Walter K, Abnizova I, Gilks W, Edwards YJ, Cooke JE, Elgar G: **Highly conserved non-coding sequences are associated with vertebrate development.** *PLoS Biol* 2005, **3:**e7.
30. Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, Haussler D: **Ultraconserved elements in the human genome.** *Science* 2004, **304:**1321-1325.
31. Hong JW, Hendrix DA, Levine MS: **Shadow enhancers as a source of evolutionary novelty.** *Science* 2008, **321:**1314.
32. Frankel N, Davis GK, Vargas D, Wang S, Payre F, Stern DL: **Phenotypic robustness conferred by apparently redundant transcriptional enhancers.** *Nature* 2010, **466:**490-493.
33. Werner T, Hammer A, Wahlbuhl M, Bösl MR, Wegner M: **Multiple conserved regulatory elements with overlapping functions determine Sox10 expression in mouse embryogenesis.** *Nucleic Acids Res* 2007, **35:**6526-6538.
34. Jeong Y, El-Jaick K, Roessler E, Muenke M, Epstein DJ.**A functional screen for sonic hedgehog regulatory elements across a 1Mb interval identifies long-range ventral forebrain enhancers.** *Development* 2006, **133:**761-72.
35. Siersbaek R, Nielsen R, John S, Sung MH, Baek S, Loft A, Hager GL, Mandrup S: **Extensive chromatin remodelling and establishment of transcription factor 'hotspots' during early adipogenesis.** *EMBO J* 2011, **30:**1459-1472.
36. Smale ST: **Pioneer factors in embryonic stem cells and differentiation.** *Curr Opin Genet Dev* 2010, **20:**519-526.
37. Azuara V, Perry P, Sauer S, Spivakov M, Jørgensen HF, John RM, Gouti M, Casanova M, Warnes G, Merkenschlager M, Fisher AG: **Chromatin signatures of pluripotent cell lines.** *Nat Cell Biol* 2006, **8:**532-538.
38. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL, Lander ES: **A bivalent chromatin structure marks key developmental genes in embryonic stem cells.** *Cell* 2006, **125:**315-326.
39. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C, Lander ES, Bernstein BE: **Genome-wide maps of chromatin state in pluripotent and lineage-committed cells.** *Nature* 2007, **448:**553-560.
40. Zentner GE, Tesar PJ, Scacheri PC: **Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions.** *Genome Res* 2011, **21:**1273-1283.
41. Xu CR, Cole PA, Meyers DJ, Kormish J, Dent S, Zaret KS: **Chromatin "prepattern" and histone modifiers in a fate choice for liver and pancreas.** *Science* 2011, **332:**963-966.
42. van Arensbergen J, Garcia-Hurtado J, Moran I, Maestro MA, Xu X, Van de Casteele M, Skoudy AL, Palassini M, Heimberg H, Ferrer J: **Derepression of Polycomb targets during pancreatic organogenesis allows insulin-producing beta-cells to adopt a neural gene activity program.** *Genome Res* 2010, **20:**722-732.
43. Sharov AA, Ko MSH: **Human ES cell profiling broadens the reach of bivalent domains.** *Cell Stem Cell* 2007, **1:**237-238.
44. Roh TY, Cuddapah S, Cui K, Zhao K: **The genomic landscape of histone modifications in human T cells.** *Proc Natl Acad Sci U S A* 2006, **103:**15782-15787.
45. Pan G, Tian S, Nie J, Yang C, Ruotti V, Wei H, Jonsdottir GA, Stewart R, Thomson JA: **Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells.** *Cell Stem Cell* 2007, **1:**299-312.
46. Vokes SA, Ji H, McCuine S, Tenzen T, Giles S, Zhong S, Longabaugh WJ, Davidson EH, Wong WH, McMahon AP: **Genomic characterization of Gli-activator targets in sonic hedgehog-mediated neural patterning.** *Development* 2007, **134:**1977-1989.
47. Vokes SA, Ji H, Wong WH, McMahon AP: **A genome-scale analysis of the cis-regulatory circuitry underlying sonic hedgehog-mediated patterning of the mammalian limb.** *Genes Dev* 2008, **22:**2651-2663.
48. Teo AK, Arnold SJ, Trotter MW, Brown S, Ang LT, Chng Z, Robertson EJ, Dunn NR, Vallier L: **Pluripotency factors regulate definitive endoderm specification through eomesodermin.** *Genes Dev* 2011, **25:**238-250.
49. Sauna ZE, Kimchi-Sarfaty C: **Understanding the contribution of synonymous mutations to human disease.** *Nat Rev Genet* 2011, **12:**683-691.
50. Carroll SB: **Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution.** *Cell* 2008, **134:**25-36.
51. Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, Ingle C, Beazley C, Gutierrez Arcelus M, Sekowska M, Gagnebin M, Nisbett J, Deloukas P, Dermitzakis ET, Antonarakis SE: **Common regulatory variation impacts gene expression in a cell type-dependent manner.** *Science* 2009, **325:**1246-1250.
52. Sagai T, Hosoya M, Mizushina Y, Tamura M, Shiroishi T: **Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb.** *Development* 2005, **132:**797-803.
53. De Gobbi M, Viprakasit V, Hughes JR, Fisher C, Buckle VJ, Ayyub H, Gibbons RJ, Vernimmen D, Yoshinaga Y, de Jong P, Cheng JF, Rubin EM, Wood WG, Bowden D, Higgs DR: **A regulatory SNP causes a human genetic disease by creating a new transcriptional promoter.** *Science* 2006, **312:**1215-1217.
54. Lander ES: **The new genomics: global views of biology.** *Science* 1996, **274:**536-539.
55. Risch N, Merikangas K: **The future of genetic studies of complex human diseases.** *Science* 1996, **273:**1516-1517.
56. Chakravarti A: **Population genetics - making sense out of sequence.** *Nat Genet* 1999, **21(1 Suppl):**56-60.
57. Reich DE, Lander ES: **On the allelic spectrum of human disease.** *Trends Genet* 2001, **17:**502-510.

58. Hirschhorn JN, Daly MJ: **Genome-wide association studies for common diseases and complex traits.** *Nat Rev Genet* 2005, **6:**95-108.
59. Kruglyak L: **The road to genome-wide association studies.** *Nat Rev Genet* 2008, **9:**314-318.
60. Hardy J, Singleton A: **Genomewide association studies and human disease.** *N Engl J Med* 2009, **360:**1759-1768.
61. Rakyan VK, Down TA, Balding DJ, Beck S: **Epigenome-wide association studies for common human diseases.** *Nat Rev Genet* 2011, **12:**529-541.
62. Visel A, Rubin EM, Pennacchio LA: **Genomic views of distant-acting enhancers.** *Nature* 2009, **461:**199-205.
63. Genovese G, Friedman DJ, Ross MD, Lecordier L, Uzureau P, Freedman BI, Bowden DW, Langefeld CD, Oleksyk TK, Uscinski Knob AL, Bernhardy AJ, Hicks PJ, Nelson GW, Vanhollebeke B, Winkler CA, Kopp JB, Pays E, Pollak MR: **Association of trypanolytic ApoL1 variants with kidney disease in African Americans.** *Science* 2010, **329:**841-845.
64. Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissino D, Mannucci PM, Anand S, Engert JC, Samani NJ, Schunkert H, Erdmann J, Reilly MP, Rader DJ, Morgan T, Spertus JA, Stoll M, Girelli D, McKeown PP, Patterson CC, Siscovick DS, O'Donnell CJ, Elosua R, Peltonen L, Salomaa V, Schwartz SM, Melander O, Altshuler D, Ardissino D, Merlini PA, Berzuini C, *et al.*: **Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants.** *Nat Genet* 2009, **41:**334-341.
65. Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, Preuss M, Stewart AF, Barbalic M, Gieger C, Absher D, Aherrahrou Z, Allayee H, Altshuler D, Anand SS, Andersen K, Anderson JL, Ardissino D, Ball SG, Balmforth AJ, Barnes TA, Becker DM, Becker LC, Berger K, Bis JC, Boekholdt SM, Boerwinkle E, Braund PS, Brown MJ, Burnett MS, *et al.*: **Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease.** *Nat Genet* 2011, **43:**333-338.
66. Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, Li X, Li H, Kuperwasser N, Ruda VM, Pirruccello JP, Muchmore B, Prokunina-Olsson L, Hall JL, Schadt EE, Morales CR, Lund-Katz S, Phillips MC, Wong J, Cantley W, Racie T, Ejebe KG, Orho-Melander M, Melander O, Kotelianski V, Fitzgerald K, Krauss RM, Cowan CA, Kathiresan S, Rader DJ: **From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus.** *Nature* 2010, **466:**714-719.
67. Visel A, Zhu Y, May D, Afzal V, Gong E, Attanasio C, Blow MJ, Cohen JC, Rubin EM, Pennacchio LA: **Targeted deletion of the 9p21 non-coding coronary artery disease risk interval in mice.** *Nature* 2010, **464:**409-412.
68. 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-gamma signalling response.** *Nature* 2011, **470:**264-268.
69. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, Zeggini E, Huth C, Aulchenko YS, Thorleifsson G, McCulloch LJ, Ferreira T, Grallert H, Amin N, Wu G, Willer CJ, Raychaudhuri S, McCarroll SA, Langenberg C, Hofmann OM, Dupuis J, Qi L, Segrè AV, van Hoek M, Navarro P, Ardlie K, Balkau B, Benediktsson R, Bennett AJ, Blagieva R, *et al.*: **Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis.** *Nat Genet* 2010, **42:**579-589.
70. Cauchi S, El Achhab Y, Choquet H, Dina C, Krempler F, Weitgasser R, Nejjari C, Patsch W, Chikri M, Meyre D, Froguel P: **TCF7L2 is reproducibly associated with type 2 diabetes in various ethnic groups: a global meta-analysis.** *J Mol Med (Berl)* 2007, **85:**777-782.
71. Lyssenko V: **The transcription factor 7-like 2 gene and increased risk of type 2 diabetes: an update.** *Curr Opin Clin Nutr Metab Care* 2008, **11:**385-392.
72. Savic D, Ye H, Aneas I, Park SY, Bell GI, Nobrega MA: **Alterations in TCF7L2 expression define its role as a key regulator of glucose metabolism.** *Genome Res* 2011, **21:**1417-1425.
73. 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.
74. 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: **Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci.** *Cell Metab* 2010, **12:**443-455.
75. Broderick P, Carvajal-Carmona L, Pittman AM, Webb E, Howarth K, Rowan A, Lubbe S, Spain S, Sullivan K, Fielding S, Jaeger E, Vijayakrishnan J, Kemp Z, Gorman M, Chandler I, Papaemmanuil E, Penegar S, Wood W, Sellick G, Qureshi M, Teixeira A, Domingo E, Barclay E, Martin L, Sieber O, Kerr D, Gray R, Peto J, Cazier JB, Tomlinson I, Houlston RS: **A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk.** *Nat Genet* 2007, **39:**1315-1317.
76. Tomlinson IP, Webb E, Carvajal-Carmona L, Broderick P, Howarth K, Pittman AM, Spain S, Lubbe S, Walther A, Sullivan K, Jaeger E, Fielding S, Rowan A, Vijayakrishnan J, Domingo E, Chandler I, Kemp Z, Qureshi M, Farrington SM, Tenesa A, Prendergast JG, Barnetson RA, Penegar S, Barclay E, Wood W, Martin L, Gorman M, Thomas H, Peto J, Bishop DT, *et al.*: **A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3.** *Nat Genet* 2008, **40:**623-630.
77. Haiman CA, Patterson N, Freedman ML, Myers SR, Pike MC, Waliszewska A, Neubauer J, Tandon A, Schirmer C, McDonald GJ, Greenway SC, Stram DO, Le Marchand L, Kolonel LN, Frasco M, Wong D, Pooler LC, Ardlie K, Oakley-Girvan I, Whittemore AS, Cooney KA, John EM, Ingles SA, Altshuler D, Henderson BE, Reich D: **Multiple regions within 8q24 independently affect risk for prostate cancer.** *Nat Genet* 2007, **39:**638-644.
78. Zanke BW, Greenwood CM, Rangrej J, Kustra R, Tenesa A, Farrington SM, Prendergast J, Olschwang S, Chiang T, Crowdy E, Ferretti V, Laflamme P, Sundararajan S, Roumy S, Olivier JF, Robidoux F, Sladek R, Montpetit A, Campbell P, Bezieau S, O'Shea AM, Zogopoulos G, Cotterchio M, Newcomb P, McLaughlin J, Younghusband B, Green R, Green J, Porteous ME, Campbell H, *et al.*: **Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24.** *Nat Genet* 2007, **39:**989-994.
79. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, Struewing JP, Morrison J, Field H, Luben R, Wareham N, Ahmed S, Healey CS, Bowman R, Meyer KB, Haiman CA, Kolonel LK, Henderson BE, Le Marchand L, Brennan P, Sangrajrang S, Gaborieau V, Odefrey F, Shen CY, Wu PE, Wang HC, Eccles D, Evans DG, Peto J, Fletcher O, *et al.*: **Genome-wide association study identifies novel breast cancer susceptibility loci.** *Nature* 2007, **447:**1087-1093.
80. Pittman AM, Naranjo S, Webb E, Broderick P, Lips EH, van Wezel T, Morreau H, Sullivan K, Fielding S, Twiss P, Vijayakrishnan J, Casares F, Qureshi M, Gómez-Skarmeta JL, Houlston RS: **The colorectal cancer risk at 18q21 is caused by a novel variant altering SMAD7 expression.** *Genome Res* 2009, **19:**987-993.
81. Pittman AM, Naranjo S, Jalava SE, Twiss P, Ma Y, Olver B, Lloyd A, Vijayakrishnan J, Qureshi M, Broderick P, van Wezel T, Morreau H, Tuupanen S, Aaltonen LA, Alonso ME, Manzanares M, Gavilán A, Visakorpi T, Gómez-Skarmeta JL, Houlston RS: **Allelic variation at the 8q23.3 colorectal cancer risk locus functions as a cis-acting regulator of EIF3H.** *PLoS Genet* 2010, **6:**e1001126.
82. Ahmadiyeh N, Pomerantz MM, Grisanzio C, Herman P, Jia L, Almendro V, He HH, Brown M, Liu XS, Davis M, Caswell JL, Beckwith CA, Hills A, Macconaill L, Coetzee GA, Regan MM, Freedman ML: **8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC.** *Proc Natl Acad Sci U S A* 2010, **107:**9742-9746.
83. Sotelo J, Esposito D, Duhagon MA, Banfield K, Mehalko J, Liao H, Stephens RM, Harris TJ, Munroe DJ, Wu X: **Long-range enhancers on 8q24 regulate c-Myc.** *Proc Natl Acad Sci U S A* 2010, **107:**3001-3005.
84. Tuupanen S, Turunen M, Lehtonen R, Hallikas O, Vanharanta S, Kivioja T, Björklund M, Wei G, Yan J, Niittymäki I, Mecklin JP, Järvinen H, Ristimäki A, Di-Bernardo M, East P, Carvajal-Carmona L, Houlston RS, Tomlinson I, Palin K, Ukkonen E, Karhu A, Taipale J, Aaltonen LA: **The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling.** *Nat Genet* 2009, **41:**885-890.
85. 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, Tabernero 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.
86. 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.
87. Jia L, Landan G, Pomerantz M, Jaschek R, Herman P, Reich D, Yan C, Khalid O, Kantoff P, Oh W, Manak JR, Berman BP, Henderson BE, Frenkel B, Haiman CA, Freedman M, Tanay A, Coetzee GA: **Functional enhancers at the gene-poor 8q24 cancer-linked locus.** *PLoS Genet* 2009, **5:**e1000597.
88. Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, Kheradpour P, Ernst J, Jordan G, Mauceli E, Ward LD, Lowe CB, Holloway AK, Clamp M, Gnerre S, Alföldi J, Beal K, Chang J, Clawson H, Cuff J, Di Palma F, Fitzgerald S, Flicek P, Guttman M, Hubisz MJ, Jaffe DB, Jungreis I, Kent WJ, Kostka D, Lara M,

et al.: **A high-resolution map of human evolutionary constraint using 29 mammals.** *Nature* 2011, **478:**476-482.
89. 1000 Genomes Project Consortium: **A map of human genome variation from population-scale sequencing.** *Nature* 2011, **467:**1061-1073.
90. Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P, Boyle PJ, Cao H, Carter NP, Clelland GK, Davis S, Day N, Dhami P, Dillon SC, Dorschner MO, Fiegler H, *et al*.: **Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.** *Nature* 2007, **447:**799-816.
91. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh 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.
92. Sabo PJ, Kuehn MS, Thurman R, Johnson BE, Johnson EM, Cao H, Yu M, Rosenzweig E, Goldy J, Haydock A, Weaver M, Shafer A, Lee K, Neri F, Humbert R, Singer MA, Richmond TA, Dorschner MO, McArthur M, Hawrylycz M, Green RD, Navas PA, Noble WS, Stamatoyannopoulos JA: **Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays.** *Nat Methods* 2006, **3:**511-518.
93. Wu CH, Chen S, Shortreed MR, Kreitinger GM, Yuan Y, Frey BL, Zhang Y, Mirza S, Cirillo LA, Olivier M, Smith LM: **Sequence-specific capture of protein-DNA complexes for mass spectrometric protein identification.** *PLoS One* 2011, **6:**e26217.
94. Sandovici I, Smith NH, Nitert MD, Ackers-Johnson M, Uribe-Lewis S, Ito Y, Jones RH, Marquez VE, Cairns W, Tadayyon M, O'Neill LP, Murrell A, Ling C, Constância M, Ozanne SE: **Maternal diet and aging alter the epigenetic control of a promoter-enhancer interaction at the Hnf4a gene in rat pancreatic islets.** *Proc Natl Acad Sci U S A* 2011, **108:**5449-5454.
95. Weaver IC, Cervoni N, Champagne FA, D'Alessio AC, Sharma S, Seckl JR, Dymov S, Szyf M, Meaney MJ: **Epigenetic programming by maternal behavior.** *Nat Neurosci* 2004, **7:**847-854.
96. McGowan PO, Sasaki A, D'Alessio AC, Dymov S, Labonte B, Szyf M, Turecki G, Meaney MJ: **Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse.** *Nat Neurosci* 2009, **12:**342-348.
97. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM: **Finding the missing heritability of complex diseases.** *Nature* 2009, **461:**747-753.

doi:10.1186/gb-2012-13-1-238
Cite this article as: Sakabe NJ, *et al*.: Transcriptional enhancers in development and disease. *Genome Biology* 2012, **13:**238.