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
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, distant 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-acting transcription 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 (SFP11 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 intervals 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...
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]. This 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.
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

| 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] |
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 unfolds.

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 supported 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, Yabara R, Northrup D, Tang Q, Gattinnoni 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, Barbera 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, Jurat-Fossé E, Zhang X, Green RD, Lobanenkov VV, Stewart R, Thomson JA, Crawford GE, Mellis K, Ren B: Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 2009, 459:108-112.

11. Creighton MR, Cheng AW, Welstead GG, Kooistra T, Carey EW, 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. Cui K, Zhang C, Roh TY, 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.

13. Ghisletti S, Barozzi I, Mietton F, Polletti S, Amati B, Palossi G, Heinberg L, Nissly-Perlman J, Lane D, Ferrando AA: A unique chromatin signature uncovers early developmental enhancers in humans. Nature 2011, 470:279-283.

14. Visel A, Blow M, Li Z, Zhang T, Akiyama JA, Holt A, Fraga MF, 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, 460:1173-1177.

15. Pekowska A, Benoukraf T, Zacarias-Cabeza J, Belhocine M, Koch F, Holota H, MNguồn S, Smit S, Visel D, Iovine V, Schmitt F, van der Ploeg L, van der Greef J, Eilers P: In vivo enhancer analysis using intronic conserved elements indicates that functional complexity might be underestimated. Genome Biol 2007, 8:R5.

16. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J: Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. Immunity 2010, 32:517-528.

17. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J: Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. Immunity 2010, 32:517-528.

18. Ghisletti S, Barozzi I, Mietton F, Polletti S, Amati B, Palossi G, Heinberg L, Nissly-Perlman J, Lane D, Ferrando AA: A unique chromatin signature uncovers early developmental enhancers in humans. Nature 2011, 470:279-283.

19. Visel A, Blow M, Li Z, Zhang T, Akiyama JA, Holt A, Fraga MF, 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, 460:1173-1177.

20. Pekowska A, Benoukraf T, Zacarias-Cabeza J, Belhocine M, Koch F, Holota H, M sağlayan S, Smit S, Visel D, Iovine V, Schmitt F, van der Ploeg L, van der Greef J, Eilers P: In vivo enhancer analysis using intronic conserved elements indicates that functional complexity might be underestimated. Genome Biol 2007, 8:R5.

21. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J: Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. Immunity 2010, 32:517-528.

22. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J: Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. Immunity 2010, 32:517-528.

23. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J: Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. Immunity 2010, 32:517-528.

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, Laptevici R, Baranovskaya A, Belhocine M, Koch F, Holota H, Iovine V, Schmitt F, van der Ploeg L, van der Greef J, Eilers P: In vivo enhancer analysis using intronic conserved non-coding sequences. Nature 2009, 460:678-684.

26. Simonis M, Kooren J, de Laat W: An evaluation of 3C-based methods to capture DNA interactions. Nat Methods 2007, 4:895-901.

27. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J: Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. Immunity 2010, 32:517-528.

28. Simonis M, Kooren J, de Laat W: An evaluation of 3C-based methods to capture DNA interactions. Nat Methods 2007, 4:895-901.

29. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J: Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. Immunity 2010, 32:517-528.

30. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J: Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. Immunity 2010, 32:517-528.
58. Hirschhorn JN, Daly MJ: Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 2005, 6:505-10.

59. Kruglyak L: The road to genome-wide association studies. Nat Rev Genet 2008, 9:134-138.

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, 12:259-264.

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, Lecocq L, Uzureau P, Friedman BI, Bowden DW, Langeland CD, Olesky TK, Ucinski Knobl A, Bernhardy AJ, Hicks PJ, Nelson JW, Vanholbeck B, Winkler CA, Kopp JB, Pays E, Pollack MR: Association of tripanolytic ApoA1 variants with kidney disease in African Americans. Science 2010, 329:841-845.

64. Kathiresan S, Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, Zeggini E, Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, Preuss M, Visel A, Rubin EM, Pennacchio LA: Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 2005, 6:505-10.

65. Kruglyak L: The road to genome-wide association studies. Nat Rev Genet 2008, 9:134-138.

66. Hardy J, Singleton A: Genomewide association studies and human disease. N Engl J Med 2009, 360:1759-1768.

67. Rakyan VK, Down TA, Balding DJ, Beck S: Epigenome-wide association studies for common human diseases. Nat Rev Genet 2011, 12:259-264.

68. Visel A, Rubin EM, Pennacchio LA: Genomic views of distant-acting enhancers. Nature 2009, 461:199-205.

69. Genovese G, Friedman DJ, Ross MD, Lecocq L, Uzureau P, Friedman BI, Bowden DW, Langeland CD, Olesky TK, Ucinski Knobl A, Bernhardy AJ, Hicks PJ, Nelson JW, Vanholbeck B, Winkler CA, Kopp JB, Pays E, Pollack MR: Association of tripanolytic ApoA1 variants with kidney disease in African Americans. Science 2010, 329:841-845.

70. Kathiresan S, Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, Zeggini E, Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, Preuss M, Visel A, Rubin EM, Pennacchio LA: Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 2005, 6:505-10.

71. Kruglyak L: The road to genome-wide association studies. Nat Rev Genet 2008, 9:134-138.

72. Hardy J, Singleton A: Genomewide association studies and human disease. N Engl J Med 2009, 360:1759-1768.

73. Rakyan VK, Down TA, Balding DJ, Beck S: Epigenome-wide association studies for common human diseases. Nat Rev Genet 2011, 12:259-264.

74. Visel A, Rubin EM, Pennacchio LA: Genomic views of distant-acting enhancers. Nature 2009, 461:199-205.
et al.: A high-resolution map of human evolutionary constraint using 29 mammals. Nature 2011, 478:476-482.

1000 Genomes Project Consortium: A map of human genome variation from population-scale sequencing. Nature 2011, 467:1061-1073.

89. 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, Malikhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P, Boyle PI, Cao H, Carter NP, Clelland GK, Davis S, Day N, Dhana 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.

90. Ernst J, Kheradpour P, Mikkelsen 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.

91. Sabo PJ, Kuehn MS, Thurman R, Johnson BE, Johnson EM, Cao H, Yu M, Rosenzweig E, Gold 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.

92. Wu CH, Chen S, Shortreed MR, Kreitzer GM, Yuan Y, Frey BL, Zhang Y, Mirza S, Cirillo LA, Oliver M, Smith LM: Sequence-specific capture of protein-DNA complexes for mass spectrometric protein identification. PLoS One 2011, 6:e26217.

93. Sandovici I, Smith NH, Nitert MD, Ackers-Johnson M, Uribe-Lewis S, Ito Y, Jones RH, Marquez VE, Cairns W, Tadayon 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 Hnf4α gene in rat pancreatic islets. Proc Natl Acad Sci U S A 2011, 108:5449-5454.

94. Weaver IC, Cervoni N, Champagne FA, D’Alessio AC, Sharma S, Seckl JR, Dymov S, Styf M, Meaney MJ: Epigenetic programming by maternal behavior. Nat Neurosci 2004, 7:847-854.

95. McGowan PO, Sasaki A, D’Alessio AC, Dymov S, Labonte B, Sztyf M, Turecki G, Meaney MJ: Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. Nat Neurosci 2009, 12:342-348.

96. 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, Slaktin 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.

Cite this article as: Sakabe NJ, et al.: Transcriptional enhancers in development and disease. Genome Biology 2012, 13:238.