Published in final edited form as:

Semin Cell Dev Biol. 2016 September ; 57: 40–50. doi:10.1016/j.semcdb.2016.05.014.

Decoding transcriptional enhancers: Evolving from annotation to functional interpretation

Krysta L. Engel¹, Mark Mackiewicz¹, Andrew A. Hardigan^{1,2}, Richard M. Myers¹, and Daniel Savic^{1,3}

¹HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806

²Department of Genetics, University of Alabama at Birmingham, Birmingham, AL 35294

Abstract

Deciphering the intricate molecular processes that orchestrate the spatial and temporal regulation of genes has become an increasingly major focus of biological research. The differential expression of genes by diverse cell types with a common genome is a hallmark of complex cellular functions, as well as the basis for multicellular life. Importantly, a more coherent understanding of gene regulation is critical for defining developmental processes, evolutionary principles and disease etiologies. Here we present our current understanding of gene regulation by focusing on the role of enhancer elements in these complex processes. Although functional genomic methods have provided considerable advances to our understanding of gene regulation, these assays, which are usually performed on a genome-wide scale, typically provide correlative observations that lack functional interpretation. Recent innovations in genome editing technologies have placed gene regulatory studies at an exciting crossroads, as systematic, functional evaluation of enhancers and other transcriptional regulatory elements can now be performed in a coordinated, high-throughput manner across the entire genome. This review provides insights on transcriptional enhancer function, their role in development and disease, and catalogues experimental tools commonly used to study these elements. Additionally, we discuss the crucial role of novel techniques in deciphering the complex gene regulatory landscape and how these studies will shape future research.

Introduction

Easily accessible, inexpensive DNA sequencing technologies have led to fundamental changes in our understanding of gene regulation and genome function. Although closer attention has traditionally been paid to protein-coding sequences, recognition of the importance of non-coding DNA segments has shifted our focus to the remaining ~98% of the genome. Numerous large, multi-center efforts such as the Encyclopedia of DNA Elements consortium (ENCODE) [1] and the Roadmap Epigenomics Project [2] have helped to make significant progress towards assigning function to the non-coding genome. The further discovery of diverse, functional, non-coding elements in genomes has gained increased attention as we have begun to more fully appreciate the roles of these distinct

³Corresponding Author: Daniel Savic (dsavic@hudsonalpha.org, daniel.d.savic@gmail.com).

regulatory features (i.e. insulators, silencers, enhancers, etc.) and associated non-coding RNAs (i.e., siRNA, lncRNA, eRNAs, etc.; reviewed in [3–5]) in transcriptional control and higher-order genome architecture. These non-coding elements confer an added level of genetic control by regulating the spatial and temporal expression of genes, as well as the degree of transcriptional activation.

Enhancers are an important class of regulatory elements that up-regulate or “enhance” the expression levels of target genes. A hallmark of enhancers is their ability to communicate across long distances, as many as hundreds of kilobases, to direct gene expression [6, 7]. Despite the identification of the first eukaryotic enhancers decades ago [8–10], the ubiquitous use of these regulatory elements is becoming more appreciated [11]. This focused look at regulatory elements has been greatly facilitated by the advent of next generation sequencing (NGS) platforms and associated functional genomic approaches for large-scale, genome-wide annotation. However, we are currently reaching the horizon of another revolution as new, highly efficient genome editing technologies have become available. These innovative methodologies will enable the first large-scale, functional interpretation of genome function and structure, and will allow for the elucidation of the underlying molecular mechanisms governing coordinated, genome-wide enhancer activity across diverse biological contexts. In this review, we describe the current understanding of enhancer function, including a brief discussion of the role of these elements during development and in disease susceptibility. We further delineate diverse, genome-wide experimental approaches that are available to researchers and conclude with how these new technologies will drive future advances in the field.

A brief view of enhancer function

Enhancer sequence structure directs its function

The specific DNA sequence composition of an enhancer contains the information necessary for imparting its composite functional effect [8]. Notably, although enhancers maintain gene regulatory functions, the evolutionary constraints placed on their sequences are less stringent than those observed in protein coding loci and gene promoters [12–14]. In fact, alteration of enhancer sequences is thought to occur rapidly and drive much of the phenotypic divergence observed between species (reviewed in [15]). The nucleotide changes that occur within enhancers during evolution are thought to alter the binding affinities of specific transcription factor (TF) proteins [16], and may even generate or ablate entire binding sites or enhancer elements. Multiple studies have provided examples in which alteration of enhancers at the sequence level lead to phenotype differences during evolution [12, 13, 17–20]. For instance, a set of alterations in enhancer sequences is thought to explain the differences in human and chimpanzee forebrain structure [20]. Single nucleotide changes in enhancers are also sufficient to alter phenotypes in *Drosophila* [21–23]. Despite the pervasive use of enhancers to regulate gene expression, a large knowledge gap concerning the complexities of enhancer mechanisms exist. This knowledge gap renders prediction of functional enhancers and their mechanisms difficult. Additionally, predicting the functional effects of different sequence variants within enhancers is further complicated by the recently described phenomenon that pairs of factors cooperatively binding to a DNA segment influence motif preference and

binding kinetics [24]. Furthermore,, although enhancer function is traditionally believed to be independent of DNA strand orientation, recent studies have challenged this notion [23, 25].

The occupancy of specific TF proteins and associated cofactor proteins generate macromolecular complexes involved in transcriptional regulation (Figure 1; reviewed in [26]). These proteins also serve as a molecular bridge by physically tethering enhancers to target gene promoters through long-range contacts (referred to as the “looping model”). However, additional models have been proposed, including tracking and variations of tracking [27, 28]. Long-range looping contacts are believed to activate target gene transcription by increasing the local concentration of the RNA polymerase II (Pol II) machinery, transcription factors, and chromatin modifying enzymes, while further precluding transcriptional inhibitors [29]. In prokaryotes, Mark Ptashne advanced the idea that *cis*-regulatory regions increase gene expression by providing the Pol II transcription machinery additional binding sites, effectively increasing the local concentration of Pol II and transcription factors [30]. Because transcription initiation is thought to be rate-limiting for gene expression [31, 32], increasing Pol II occupancy at promoters would presumably augment gene expression. A similar model has been adopted for the role of enhancers in eukaryotes. In support of this, Pol II transcription machinery and the Mediator complex occupy promoter-distal enhancer elements [33–37].

Enhancer activity is regulated at multiple layers

In order to control gene regulation in a coordinated manner, enhancers themselves are subject to dynamic regulation via epigenetic modifications (Figure 1). Chromatin marks are commonly used as indicators of enhancer “state”, which is defined as the ability for an enhancer to increase expression of target genes, and are categorized into inactive, poised and active states. Inactive enhancers reside in closed chromatin conformations that are characterized by the presence of histone-3 lysine-27 tri-methylation (H3K27me3) [38, 39]. By contrast, poised enhancers are co-occupied by histone-3 lysine-4 mono-methylation (H3K4me1) and H3K27me3 histone modifications [39–41]. During activation, histone-3 lysine-27 acetylation (H3K27ac) replaces the H3K27me3 mark [38, 41–43]. Apart from histone modifications, DNA methylation [5-methylcytosine (5mC)] is also implicated in the regulation of enhancer activity. DNA hypomethylation and 5-hydroxymethylcytosine (5hmC) appear to correlate with enhancer activity [44, 45]. Because DNA methylation typically inhibits TF binding [46], 5hmC has been suggested to prime enhancers for later use by preventing 5mC [47]. This epigenetic mechanism is used throughout development and even during early developmental stages [48–50] (see section entitled “Gene expression during development is influenced by enhancer utilization”). The DNA methylation landscape surrounding enhancers is likely a complex occurrence, and recent studies suggest that DNA methylation of sub regions within enhancers positively correlate with H3K27ac and negatively correlate with TF binding sites (Charlet, diumich, lay et al. Mol Cell vol 62, issue 3, 2016). Future elucidation of enhancer regulation will most assuredly uncover additional intricate methods by which the chromatin landscape enables enhancers to fine-tune gene expression of their target loci.

Enhancers are also transcribed by Pol II to produce enhancer RNAs (eRNAs; see Figure 1) [11]. It is unclear whether eRNA synthesis is merely a result of the close proximity to Pol II machinery or if eRNA production serves a role during gene activation. Multiple studies indicate a direct role for eRNAs in enhancer function. A recent report from the Young laboratory showed that eRNAs bind TFs and increase the local concentration of YY1, a sequence-specific TF [51]. Further supporting a functional role for eRNAs in enhancer activity, the presence of eRNA is also involved in the occupancy of CTCF and CP190 DNA-binding proteins [52, 53], and ablation of various eRNAs have illustrated their importance for proper gene expression [54–56]. However, other studies have shown that eRNAs may not be required for enhancer activity [57]. Interestingly, enhancer-promoter contact through looping was not disrupted from eRNA depletion [55]. Despite these conflicting results, it is clear that eRNA production is positively correlated with active regulatory elements [7, 58].

Further complicating the elucidation of the role(s) that RNA derived from regulatory elements play is the delineation between different types of non-coding RNA, specifically enhancer-derived RNA (eRNA) and long non-coding RNA (lncRNA) [59]. The predominant view in the field is that eRNAs are not polyadenylated, are degraded quickly, and are transcribed from an enhancer; whereas lncRNAs are polyadenylated, stable, and are transcribed from their own promoter. However, recent evidence shows that *Lockd*, a well-studied polyadenylated lncRNA, is actually transcribed from an enhancer and that the RNA itself is not required for enhancer function [59]. These findings challenge the classifications of eRNA and lncRNA as completely separate categories and further confounds our ability to predict enhancer elements based on the type of RNA produced from them.

Gene expression is regulated across 3-dimensional space

An intriguing property of enhancers is their capacity to act over long distances. This long-range activity imparts challenges, as the frequency with which two DNA segments randomly interact decays linearly with increased distance [60, 61]. To overcome the negative impact of distance between enhancers and target promoter loci, chromatin forms 3-dimensional (3D) topologically associated domains (TADs) that enhance the frequency of interactions between distant loci (Figure 1) [62–67]. TADs typically span hundreds of kilobases [62, 63, 68] and their formation compartmentalizes sets of regulatory elements by bringing them into close spatial proximity [63, 69]. CTCF and cohesin are key factors involved in the regulation of TAD borders, as well as contacts formed within TADs [29]. Interestingly, direct contacts between two DNA segments have also been proposed to influence 3D genome architecture [70]. Multiple studies have reported that intra-TAD chromatin interactions are dynamic [33, 71, 72] and occur more frequently than interactions between TADs, while the disruption of intra-TAD interaction networks can affect gene expression [73–75]. Interestingly, CTCF and cohesin also bind diverse regulatory elements, including enhancers and insulators [29], which indicates a high level of coordination between different types of loci for the maintenance of proper transcription control.

There is also a connection between higher-order nuclear localization and gene expression. For instance, lamina associated domains (LADs) are regions of chromatin that reside near the nuclear periphery and are typically silenced (Figure 1) [62, 71]. Upon activation, gene

loci physically move from the nuclear periphery to the center of the nucleus into “transcription factories” prior to gene activation (Figure 1) [76–79]. The choice of transcription factory is non-random and co-regulated genes are thought to exist within a single factory [80]. In support of the latter, the disruption of a gene within a transcription factory also affects other genes within the same factory [81].

A related question concerns how enhancer-promoter specificity is achieved. Enhancer-promoter specificity is a complex phenomenon as enhancers contact an average of 2 promoters, whereas promoters contact an average of 4–5 enhancers each [6, 82, 83]. These interactions may also be context dependent, with distinct cell-types or developmental stages harboring unique combinations of interactions [6, 7, 84, 85]. The precision of enhancer-promoter interactions is driven, at least in part, by the occupancy of sequence- and tissue-specific TFs at promoters and regulatory loci [86]. Enhancers also display a preference towards interacting with different types of promoters [87–91]. In light of the role of 3D chromatin structure in gene regulation described above, nuclear architecture likely provides an additional level of control by prohibiting specific loci from interacting [38, 39].

Enhancers in development and disease

Gene expression during development is dynamically regulated by enhancer utilization

Fundamental developmental processes are orchestrated by the spatial and temporal regulation of enhancer element activity [1, 92–95]. This spatiotemporal genetic control is prevalent even at the onset of metazoan development. Indeed, pluripotency of embryonic stem cells (ESCs) is defined by two distinct transitional states: a “naïve” state, which corresponds to pre-implanted cells of the inner cell mass or epiblast, and the “primed” state, which represents the post-implanted epiblast [96]. Although cells in both states express the same master regulatory TFs (Oct4, Sox2, and Nanog) critical for maintaining “stemness”, which describes the ability of cells to both self-renew and differentiate, the gene expression profiles and cellular requirements for each cellular state are divergent, pointing to the utilization of distinct gene regulatory programs [97–99]. Chromatin state and global TF occupancy further confirm differential enhancer usage as a key driver of the disparate gene expression profiles between naïve and primed ESC states [100, 101]. In general, active enhancers in ESCs, as is typical for enhancers in general, are marked by low nucleosome density, H3K27ac and H3K4me1, and are bound by the EP300 histone acetyltransferase. Inactive but poised enhancers in ESCs are primed with H3K4me1 and marked by H3K27me3, and are dynamically activated during early development by differential TF binding and deposition of active chromatin marks to drive cell and lineage-specific expression programs [38,39].

Differential DNA methylation between cell types could also explain enhancer-promoter specificity and transcriptional activity in different contexts. 5mC regulation occurs most frequently in a CpG dinucleotide context, but non-CpG methylation has been shown to account for as much as 25% of 5mC in embryonic stem cells (ESCs) [102]. These non-CpG context 5mC sites are preferentially depleted in active ESC enhancers relative to CpG sites and are lost during differentiation, though they can be restored upon somatic-cell conversion to iPSCs [102].

During development, the DNA methylome is dynamically remodeled, with up to 21.8% of autosomal CpGs in over 700,000 unique differentially methylated regions (DMRs) exhibiting altered methylation between cell types with various levels of differentiation [103]. Approximately 42.3% of DMRs were shown to overlap with DNaseI hypersensitivity sites and 26.1% were located in putative enhancer elements. Additionally, adult tissue specific DMRs (tsDMRs) were shown to comprise 6.7% of the mouse genome, the majority of which were located near distal regulatory elements such as enhancers [104]. These tsDMRs are hypomethylated and enriched for lineage-specific master regulator transcription factors. Developmental enhancers can remain active or become inactive by means of chromatin modification and DNA methylation. One method by which inactivation is thought to occur is through removal of H3K4me1 by LSD1 [105]. Subsequent DNA methylation after deactivation of enhancers could also help prevent binding of pioneer transcription factors and re-activation of the enhancer. However, recent work has shown that some enhancers retain “epigenetic memory” as vestigial enhancers, which gain closed chromatin modifications to transition from developmentally active to inactive states but remain hypomethylated in adult tissue [104].

Highly coordinated gene regulatory control also persists during later tissue development, as recently demonstrated by Nord and colleagues [106]. Through temporal epigenomic profiling and transgenic reporter assay validation, tens of thousands of putative developmental enhancers were identified across seven developmental stages in three distinct mouse tissues [106]. A majority of these developmental enhancers exhibited tissue-specific activity, as well as rapid and tightly controlled temporal changes in usage. Supporting the biological relevance of these observations, these regulatory changes mirrored differential gene expression profiles during the developmental time course [106]. By further evaluating sequence conservation at these regulatory elements, these analyses supported the “hourglass” model of developmental evolution [106], where the largest constraint in gene regulatory control is observed during early embryogenesis [107–109].

The selection of these cell- and developmental state-specific enhancer programs is largely influenced by the activity of lineage-determining TFs called pioneer factors [110]. These factors displace nucleosome complexes, allowing additional TF proteins subsequent access for binding to sequence motifs within the enhancer element [111]. The selection of the enhancer repertoire by pioneer factors can also depend on the cell’s surrounding environment. Indeed, macrophages reside in many organs throughout the body and exhibit unique gene expression patterns at the distinct resident tissues they populate [112, 113]. Notably, this activity is mediated through the differential activation of enhancers [114]. Collectively, as highly dynamic and context-specific elements, enhancers seem to act as molecular rheostats, governing developmental processes while further maintaining physiological equilibrium in response to extracellular cues.

Sequence variation in enhancer elements contributes to disease phenotypes

Although the importance of enhancers in orchestrating gene regulatory programs during cellular differentiation and development is well established, it has also become increasingly clear that disease pathogenesis can regularly arise from mutations in enhancer sequences.

This notion is best exemplified by Genome-Wide Association Studies (GWAS) that frequently identify risk loci for common diseases or traits in non-coding genomic DNA segments [115–118]. Several informative reviews have described these association results as well as downstream functional validations [115, 116, 119]. In fact, it is estimated that more than 90% of trait-associated variants reside in non-coding regions of the human genome [120, 121]. Moreover, those variants that lie in annotated enhancer elements are thought to explain a larger proportion of the heritability for some disorders compared to protein-coding variants [121–124].

To determine the percentage of disease- or trait-associated variants situated within non-coding regulatory elements, Hnisz and colleagues mapped more than five thousand single-nucleotide polymorphisms (SNPs) identified across 1,675 GWAS analyses and used chromatin state information from various human cells and tissues in an integrated analysis [125]. They observed that the majority of SNPs mapped to non-coding regions of the genome, with 64% of SNPs localizing to putative enhancer elements [125]. Notably, these SNPs were also enriched at super-enhancer elements, larger enhancer sequences (~10–20 kilobases in length) that are believed to be drivers of cell identity [126, 127]. Mutations in these large developmental elements are also implicated in several complex diseases, including Alzheimer's disease, type 1 diabetes, systemic lupus erythematosus, rheumatoid arthritis and multiple sclerosis [125, 128]. In addition to *cis* mutations within enhancers, altered regulation of enhancer activity can also occur *in trans*, through mutations in TFs or transcriptional cofactors that bind to regulatory elements [46, 115, 116, 129, 130]. These *trans* mutational effects have been observed for various Mendelian developmental disorders, such as Cornelia de Lange, Lujan and Rubenstein-Taybi syndromes [128].

Experimental methods for studying gene regulation – a historical perspective

Conservation predicts function

The catalog of diverse whole genome sequences that has been generated during the last 15 years has provided the first picture of genome composition. These insights have further highlighted the daunting task of identifying functional relevance for the ~98% of higher eukaryotic genomes represented by non-coding sequences [131–133]. Unlike the case for protein-coding sequences, there is no discernable genetic code for experimental exploitation of non-coding sequences. Despite this hurdle, the genome datasets available from numerous species provide an initial solution as their evolutionary histories can be harnessed to identify functional sites in non-coding sequence [134, 135]. Indeed, studies of sequence conservation provide a rational strategy as sequence alterations at key regulatory sequences would presumably, and under many circumstances, lead to negative effects on organismal fitness. As a result, DNA sequence conservation reflects an underlying biological function. A variety of studies support the utility of evolutionary conservation for identifying non-coding regulatory sequences [136–138]. This simple approach even proved valuable for elucidating the genetic causes of rare human diseases. For instance, a sonic hedgehog (*Shh*) enhancer element situated one megabase away from the *Shh* gene locus was identified based on sequence conservation and mutations within this element were subsequently linked to

preaxial polydactyly [139]. Despite the initial success of sequence conservation, the power of this method is directly related to the degree of nucleotide constraint and is therefore not feasible for identifying less conserved, yet still critical regulatory elements. This is a significant concern, considering the higher sequence turnover rate in enhancers [12–14].

The emergence of functional genomics

As a complement to evolutionary conservation, diverse, orthogonal strategies are available that use NGS technologies to detect regulatory elements [140, 141]. These functional genomic tools provide a largely unbiased, genome-wide picture of gene regulation and genome function [142], and can assess genome-wide DNA methylation, gene expression, protein-DNA interactions and identify regions of open chromatin. These techniques have revolutionized genomics research and many of the insights concerning enhancer function described above are derived from functional genomic experimentation. DNA methylation maps can be ascertained through Whole Genome Bisulfite Sequencing (WGBS) or Reduced Representation Bisulfite Sequencing (RRBS) [143]. Meanwhile, transcriptome profiles can be generated by using RNA sequencing (RNA-seq) [144], while several, related techniques are available for identifying regulatory elements. The genome-wide capture of DNA-binding proteins directly associated with genomic sequences is possible through Chromatin Immunoprecipitation followed by NGS (ChIP-seq) [145], while DNaseI based assays [146, 147] provide a detailed map of open chromatin that is agnostic to DNA-binding protein information. Apart from linear annotation, 3D genome structure can also be profiled using variations of the Chromatin Conformation Capture (3C) methodology [148], such as Chromosome Conformation Capture Carbon Copy (5C) [149] and Hi-C [150]. Chromatin Interaction Analysis by Paired-End Tag sequencing (ChIA-PET) [151], a combination of ChIP- and 3C-based techniques, can even be used to identify 3D interactions of individual DNA-binding proteins.

Despite the success of these genomic techniques, challenges remain. For instance, ChIP-seq relies on the availability of suitable ChIP-seq grade antibodies [152]; current estimates from the ENCODE Project indicate that less than 10% of commercial antibodies are sufficiently specific or efficacious enough to be useful in ChIP-seq assays (our unpublished observations). To circumvent these problems, epitope tagging techniques have been applied where diverse DNA-binding proteins can be annotated using a single, high-quality antibody [153]. To further combat the difficulties of working with primary tissues, alternative preparation strategies have been used for downstream ChIP-seq characterizations [154]. As DNaseI-derived methods can be technically challenging, a transposase-based strategy called Assay for Transposase-Assessible Chromatin with high-throughput sequencing (ATAC-seq) has also recently gained wide appeal for its simplicity [155]. For more dynamic analyses of gene regulation, Genomic Run-On sequencing (GRO-seq) has been utilized for identifying actively transcribed regions [156]. Recent alterations have also been applied to 3D genome mapping. As a way to limit the genome space assessed by Hi-C and therefore the necessary NGS read depth to obtain relevant information, Capture Hi-C has been applied [157].

A primary limitation of several functional genomic assays, including ChIP-seq, open chromatin and 3D mapping, stems from the realization that these assays only provide

annotation and cannot reliably predict regulatory element function or even relevance, such as a role in gene regulation. These methods can also lead to false positive results and conclusions. For example, these genomic assays have consistently highlighted the presence of thousands of regulatory elements in the genome [1, 158]. However, many of these TF binding events do not appear to be directly involved in transcriptional regulation [159–161]. As a result, these genomic approaches are useful for annotation, but cannot provide a comprehensive genomic analysis where truly functional elements are differentiated from passive sites.

On the road to high-throughput functional interpretation

The genomics of gene regulation is at an exciting crossroads with the recent advent of systematic, high-throughput assays that enable true functional validation of genome-wide observations. Although NGS had been used for more traditional functional genomic assays, various high-throughput reporter-based assays have recently capitalized on this technology [162–165]. These massively parallel reporter assays provide a direct assessment of regulatory activity for a large number of DNA segments. Two predominant approaches have been used to date. Several methods, such as Cis-Regulatory Element analysis by sequencing (CRE-seq, [166]), utilize a barcode-based strategy [162–164] where regulatory activity is measured through NGS of barcodes within the 3'-untranslated region (3'-UTR). As this strategy relies on oligonucleotides harboring the test element, barcode and restriction enzyme sites, oligonucleotide synthesis limitations place restrictions on the size of elements that can be tested. Consequently, the accurate, functional assessment of longer enhancers may prove problematic. Alternative strategies such as Self-Transcribing Active Regulatory Region sequencing (STARR-seq) involve cloning putative test DNA regulatory segments directly within the 3'-UTR of reporter genes and measuring regulatory activity by NGS of test element derived RNA [165]. Compared to other techniques, STARR-seq can test longer DNA fragments and also provides a more straightforward experimental cloning design. However, the placement of longer test sequences within the 3'-UTR may impact RNA stability, limiting biological interpretation. Despite these drawbacks, diverse next-generation reporter assays have generated notable functional validations of observations from genomic studies [58, 164, 167].

Although massively parallel reporter assays provide a controlled system for functional interpretation, the reliance on an artificial DNA construct may generate confounding effects. Alternative techniques that allow for the functional characterization of DNA sequences within their endogenous genomic setting are now possible with diverse genome editing technologies. In light of their simple design, high efficiency and multiplexing capabilities, Clustered Regularly Interspaced Short Palindromic Repeat/CRISPR-associated protein-9 nuclease (CRISPR/Cas9) genome editing [168–170] is rapidly replacing alternative strategies such as zinc finger nucleases (ZFNs) [171] and transcription activator-like effector nucleases (TALENs) [172]. Collectively, massively parallel reporter assays and CRISPR/Cas9 provide two complementary platforms for large-scale, functional interpretation of genome-wide annotations.

The future of gene regulation studies

Facilitated by diverse technological advances, genomics-based research has generated a wealth of information on gene regulatory control. Despite this initial success, the connections between simple genome sequence structure and complex cellular processes remain elusive. A lack of gene regulatory understanding is further compounded as the transition is made from simple, cellular analyses into more intricate, inter-tissue, whole organismal studies. Moreover, although non-coding sequences have been carefully annotated across distinct cell types using various genomic assays, biological and functional roles have yet to be assigned to most of these loci. Despite these challenges, there are compelling reasons to be excited about the future of studies aimed at understanding gene regulation.

As NGS techniques become more accessible and cost-effective, functional genomic analyses will become more integrative and allow for the elucidation of information from multiple levels of transcriptional control. By combining DNA methylation, chromatin state, DNA-binding protein occupancy, eRNA production and gene expression with long-range, 3D interactions, key elements will be identified for subsequent functional analyses, and the critical features that serve as hallmarks of active regulatory elements will become increasingly clear (Figure 2A). To control for cellular and clonal heterogeneity that may confound biological interpretation [173, 174], single cell functional genomic approaches will also become increasingly popular. Indeed, various single-cell genomic strategies have recently been developed for RNA-seq [175–177], bisulfite sequencing [178] and ATAC-seq [179, 180]. Data from these studies has highlighted extensive variability across cellular populations, confirming the potential problems with using heterogeneous samples [181–183]. Although robust single-cell strategies for ChIP-seq assays are still lacking, several techniques that are amenable for low cell number experimentations have been recently developed [184, 185].

CRISPR/Cas9 genome editing is at the forefront of future gene regulatory studies as this technology can provide functional interpretation for thousands of enhancer elements in a high-throughput manner. In fact, large-scale CRISPR/Cas9 screens have recently been performed through efficient viral-based delivery methods in cell lines [186–188] as well as in animal models [189]. These screens have provided key insight regarding novel genes and cellular pathways involved in pharmacological drug resistance [187], immunological response [188] and tumor metastasis [189]. CRISPR/Cas9 technology has also been applied for epigenetic manipulations through the use of nuclease-deactivated Cas9 proteins fused to various transcriptional activator and repressor domains [190–193]. The engineering of diverse Cas9 editing systems [194, 195] will also provide added flexibility to study designs. Moreover, this platform offers a straightforward approach for the functional interrogation of hundreds of disease-associated GWAS-identified loci [196], as well as a viable path for cataloging the first phenotypic map for all enhancers in a single cell or across diverse tissues. As these and related high-throughput screens become more widely applied, the underlying molecular mechanisms governing complex developmental and physiological processes will become more apparent.

Although GWAS studies have played a pinnacle role for identifying common disease-causing variants, rare variants are garnering more attention in hopes of explaining a higher fraction of the genetic susceptibility to complex diseases [197, 198]. The declining cost of NGS will therefore have a profound effect on the design of disease studies [199–201]. The future will see a plethora of rarer, non-coding variant data available for functional interpretation in disease-related research and in clinical settings; prioritizing and annotating the thousands of non-coding sequence variations that will be identified, including discriminating active from passive variants, will provide additional challenges [202]. For these analyses, variant pathogenicity estimate methods that utilize diverse information, including functional genomic datasets, will prove highly beneficial [203]. As these efforts aim to identify rarer causal variants and make treatments more personalized, patient-derived induced pluripotent stem cells (iPSC) will become important in modeling both variants as well as disease states [204, 205] (see Figure 2A). Following integrative functional genomic assay annotation in these cells, genome editing strategies and high-throughput reporter assays will provide researchers with the tools necessary for functional interpretation (see Figures 2A and 2B). In fact, massively parallel reporter assays are beginning to combat design limitations (see above) [206], and these next-generation reporter assays have already been adapted for large-scale functional interpretation of variants at disease-associated loci [207].

Following the complete sequence map of diverse species' genomes, a high degree of gene regulatory complexity within genomes has been uncovered. As we transition into a functional phase of genomic discovery, many outstanding questions in development and disease remain to be addressed. Ongoing efforts capitalizing on the current repertoire of novel experimental approaches will undoubtedly generate unexpected results, elucidate new mechanisms, and lead to a greater appreciation for the non-coding, regulatory genome.

Acknowledgments

This work was supported by NIH grant U54 HG006998-0 (to RMM). We thank Nick Cochran, Jessica Woolnough and Sarah Meadows for helpful suggestions and edits.

References

1. ENCODE. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012; 489:57–74. [PubMed: 22955616]
2. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015; 518:317–30. [PubMed: 25693563]
3. Pek JW, Okamura K. Regulatory RNAs discovered in unexpected places. *Wiley Interdiscip Rev RNA*. 2015; 6:671–86. [PubMed: 26424536]
4. Orom UA, Shiekhattar R. Noncoding RNAs and enhancers: complications of a long-distance relationship. *Trends Genet*. 2011; 27:433–9. [PubMed: 21831473]
5. Holoch D, Moazed D. RNA-mediated epigenetic regulation of gene expression. *Nat Rev Genet*. 2015; 16:71–84. [PubMed: 25554358]
6. Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature*. 2013; 503:290–4. [PubMed: 24141950]
7. Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. *Nature*. 2012; 489:109–13. [PubMed: 22955621]

8. Collis P, Antoniou M, Grosveld F. Definition of the minimal requirements within the human beta-globin gene and the dominant control region for high level expression. *EMBO J.* 1990; 9:233–40. [PubMed: 2295312]
9. Tuan D, Kong S, Hu K. Transcription of the hypersensitive site HS2 enhancer in erythroid cells. *Proc Natl Acad Sci U S A.* 1992; 89:11219–23. [PubMed: 1454801]
10. Ashe HL, Monks J, Wijgerde M, Fraser P, Proudfoot NJ. Intergenic transcription and transinduction of the human beta-globin locus. *Genes Dev.* 1997; 11:2494–509. [PubMed: 9334315]
11. Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, et al. Widespread transcription at neuronal activity-regulated enhancers. *Nature.* 2010; 465:182–7. [PubMed: 20393465]
12. Degner JF, Pai AA, Pique-Regi R, Veyrieras JB, Gaffney DJ, Pickrell JK, et al. DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature.* 2012; 482:390–4. [PubMed: 22307276]
13. Shibata Y, Sheffield NC, Fedrigo O, Babbitt CC, Wortham M, Tewari AK, et al. Extensive evolutionary changes in regulatory element activity during human origins are associated with altered gene expression and positive selection. *PLoS Genet.* 2012; 8:e1002789. [PubMed: 22761590]
14. Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, et al. Enhancer evolution across 20 mammalian species. *Cell.* 2015; 160:554–66. [PubMed: 25635462]
15. Wray GA. The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet.* 2007; 8:206–16. [PubMed: 17304246]
16. Maurano MT, Haugen E, Sandstrom R, Vierstra J, Shafer A, Kaul R, et al. Large-scale identification of sequence variants influencing human transcription factor occupancy in vivo. *Nat Genet.* 2015
17. Arnold CD, Gerlach D, Spies D, Matts JA, Sytnikova YA, Pagani M, et al. Quantitative genome-wide enhancer activity maps for five *Drosophila* species show functional enhancer conservation and turnover during cis-regulatory evolution. *Nat Genet.* 2014; 46:685–92. [PubMed: 24908250]
18. Cotney J, Leng J, Yin J, Reilly SK, DeMare LE, Emera D, et al. The evolution of lineage-specific regulatory activities in the human embryonic limb. *Cell.* 2013; 154:185–96. [PubMed: 23827682]
19. Hare EE, Peterson BK, Iyer VN, Meier R, Eisen MB. Sepsid even-skipped enhancers are functionally conserved in *Drosophila* despite lack of sequence conservation. *PLoS Genet.* 2008; 4:e1000106. [PubMed: 18584029]
20. McLean CY, Reno PL, Pollen AA, Bassan AI, Capellini TD, Guenther C, et al. Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature.* 2011; 471:216–9. [PubMed: 21390129]
21. Frankel N, Erezylmaz DF, McGregor AP, Wang S, Payre F, Stern DL. Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. *Nature.* 2011; 474:598–603. [PubMed: 21720363]
22. Jeong S, Rebeiz M, Andolfatto P, Werner T, True J, Carroll SB. The evolution of gene regulation underlies a morphological difference between two *Drosophila* sister species. *Cell.* 2008; 132:783–93. [PubMed: 18329365]
23. Williams TM, Selegue JE, Werner T, Gompel N, Kopp A, Carroll SB. The regulation and evolution of a genetic switch controlling sexually dimorphic traits in *Drosophila*. *Cell.* 2008; 134:610–23. [PubMed: 18724934]
24. Jolma A, Yin Y, Nitta KR, Dave K, Popov A, Taipale M, et al. DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature.* 2015
25. Jeong S, Rokas A, Carroll SB. Regulation of body pigmentation by the Abdominal-B Hox protein and its gain and loss in *Drosophila* evolution. *Cell.* 2006; 125:1387–99. [PubMed: 16814723]
26. Arnosti DN, Kulkarni MM. Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *J Cell Biochem.* 2005; 94:890–8. [PubMed: 15696541]
27. Bulger M, Groudine M. Functional and mechanistic diversity of distal transcription enhancers. *Cell.* 2011; 144:327–39. [PubMed: 21295696]
28. Dean A. On a chromosome far, far away: LCRs and gene expression. *Trends Genet.* 2006; 22:38–45. [PubMed: 16309780]

29. Cubenas-Potts C, Corces VG. Architectural proteins, transcription, and the three-dimensional organization of the genome. *FEBS Lett.* 2015; 589:2923–30. [PubMed: 26008126]
30. Ptashne M. Gene regulation by proteins acting nearby and at a distance. *Nature.* 1986; 322:697–701. [PubMed: 3018583]
31. Kugel JF, Goodrich JA. Promoter escape limits the rate of RNA polymerase II transcription and is enhanced by TFIIE, TFIIH, and ATP on negatively supercoiled DNA. *Proc Natl Acad Sci U S A.* 1998; 95:9232–7. [PubMed: 9689063]
32. Hoopes BC, LeBlanc JF, Hawley DK. Contributions of the TATA box sequence to rate-limiting steps in transcription initiation by RNA polymerase II. *J Mol Biol.* 1998; 277:1015–31. [PubMed: 9571019]
33. Phillips-Cremins JE, Sauria ME, Sanyal A, Gerasimova TI, Lajoie BR, Bell JS, et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell.* 2013; 153:1281–95. [PubMed: 23706625]
34. Hou C, Dale R, Dean A. Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proc Natl Acad Sci U S A.* 2010; 107:3651–6. [PubMed: 20133600]
35. Lin C, Garruss AS, Luo Z, Guo F, Shilatifard A. The RNA Pol II elongation factor Ell3 marks enhancers in ES cells and primes future gene activation. *Cell.* 2013; 152:144–56. [PubMed: 23273992]
36. Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, et al. Mediator and cohesin connect gene expression and chromatin architecture. *Nature.* 2010; 467:430–5. [PubMed: 20720539]
37. Maston GA, Landt SG, Snyder M, Green MR. Characterization of enhancer function from genome-wide analyses. *Annu Rev Genomics Hum Genet.* 2012; 13:29–57. [PubMed: 22703170]
38. Creighton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A.* 2010; 107:21931–6. [PubMed: 21106759]
39. Rada-Iglesias A, Bajpai R, Swigut T, Bruggmann SA, Flynn RA, Wysocka J. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature.* 2011; 470:279–83. [PubMed: 21160473]
40. Herz HM, Mohan M, Garruss AS, Liang K, Takahashi YH, Mickey K, et al. Enhancer-associated H3K4 monomethylation by Trithorax-related, the *Drosophila* homolog of mammalian Mll3/Mll4. *Genes Dev.* 2012; 26:2604–20. [PubMed: 23166019]
41. Zentner GE, Tesar PJ, Scacheri PC. Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res.* 2011; 21:1273–83. [PubMed: 21632746]
42. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet.* 2007; 39:311–8. [PubMed: 17277777]
43. Ferrari KJ, Scelfo A, Jammula S, Cuomo A, Barozzi I, Stutzer A, et al. Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. *Mol Cell.* 2014; 53:49–62. [PubMed: 24289921]
44. Pulakanti K, Pinello L, Stelloh C, Blinka S, Allred J, Milanovich S, et al. Enhancer transcribed RNAs arise from hypomethylated, Tet-occupied genomic regions. *Epigenetics.* 2013; 8:1303–20. [PubMed: 24135681]
45. Stroud H, Feng S, Morey Kinney S, Pradhan S, Jacobsen SE. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biol.* 2011; 12:R54. [PubMed: 21689397]
46. Smith E, Shilatifard A. Enhancer biology and enhanceropathies. *Nat Struct Mol Biol.* 2014; 21:210–9. [PubMed: 24599251]
47. Pastor WA, Aravind L, Rao A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat Rev Mol Cell Biol.* 2013; 14:341–56. [PubMed: 23698584]
48. Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Scholer A, et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature.* 2011; 480:490–5. [PubMed: 22170606]
49. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. The accessible chromatin landscape of the human genome. *Nature.* 2012; 489:75–82. [PubMed: 22955617]

50. Varley KE, Gertz J, Bowling KM, Parker SL, Reddy TE, Pauli-Behn F, et al. Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res.* 2013; 23:555–67. [PubMed: 23325432]
51. Sigova AA, Abraham BJ, Ji X, Molinie B, Hannett NM, Guo YE, et al. Transcription factor trapping by RNA in gene regulatory elements. *Science.* 2015
52. Lei EP, Corces VG. RNA interference machinery influences the nuclear organization of a chromatin insulator. *Nat Genet.* 2006; 38:936–41. [PubMed: 16862159]
53. Yao H, Brick K, Evrard Y, Xiao T, Camerini-Otero RD, Felsenfeld G. Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. *Genes Dev.* 2010; 24:2543–55. [PubMed: 20966046]
54. Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, et al. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature.* 2013; 498:516–20. [PubMed: 23728302]
55. Melo CA, Drost J, Wijchers PJ, van de Werken H, de Wit E, Oude Vrielink JA, et al. eRNAs are required for p53-dependent enhancer activity and gene transcription. *Mol Cell.* 2013; 49:524–35. [PubMed: 23273978]
56. Hah N, Murakami S, Nagari A, Danko CG, Kraus WL. Enhancer transcripts mark active estrogen receptor binding sites. *Genome Res.* 2013; 23:1210–23. [PubMed: 23636943]
57. Kaikkonen MU, Spann NJ, Heinz S, Romanoski CE, Allison KA, Stender JD, et al. Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol Cell.* 2013; 51:310–25. [PubMed: 23932714]
58. Savic D, Roberts BS, Carleton JB, Partridge EC, White MA, Cohen BA, et al. Promoter-distal RNA polymerase II binding discriminates active from inactive CCAAT/enhancer-binding protein beta binding sites. *Genome Res.* 2015
59. Paralkar VR, Taborda CC, Huang P, Yao Y, Kossenkova AV, Prasad R, et al. Unlinking an lncRNA from Its Associated cis Element. *Mol Cell.* 2016; 62:104–10. [PubMed: 27041223]
60. Fudenberg G, Mirny LA. Higher-order chromatin structure: bridging physics and biology. *Curr Opin Genet Dev.* 2012; 22:115–24. [PubMed: 22360992]
61. Tark-Dame M, van Driel R, Heermann DW. Chromatin folding--from biology to polymer models and back. *J Cell Sci.* 2011; 124:839–45. [PubMed: 21378305]
62. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature.* 2012; 485:376–80. [PubMed: 22495300]
63. Hou C, Li L, Qin ZS, Corces VG. Gene density, transcription, and insulators contribute to the partition of the *Drosophila* genome into physical domains. *Mol Cell.* 2012; 48:471–84. [PubMed: 23041285]
64. Li L, Lyu X, Hou C, Takenaka N, Nguyen HQ, Ong CT, et al. Widespread rearrangement of 3D chromatin organization underlies polycomb-mediated stress-induced silencing. *Mol Cell.* 2015; 58:216–31. [PubMed: 25818644]
65. Phillips-Cremins JE, Corces VG. Chromatin insulators: linking genome organization to cellular function. *Mol Cell.* 2013; 50:461–74. [PubMed: 23706817]
66. Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, et al. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell.* 2012; 148:458–72. [PubMed: 22265598]
67. Van Bortle K, Nichols MH, Li L, Ong CT, Takenaka N, Qin ZS, et al. Insulator function and topological domain border strength scale with architectural protein occupancy. *Genome Biol.* 2014; 15:R82. [PubMed: 24981874]
68. Sofueva S, Yaffe E, Chan WC, Georgopoulou D, Vietri Rudan M, Mira-Bontenbal H, et al. Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J.* 2013; 32:3119–29. [PubMed: 24185899]
69. Lupianez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell.* 2015; 161:1012–25. [PubMed: 25959774]
70. Cournac A, Koszul R, Mozziconacci J. The 3D folding of metazoan genomes correlates with the association of similar repetitive elements. *Nucleic Acids Res.* 2015

71. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*. 2012; 485:381–5. [PubMed: 22495304]
72. Williamson I, Berlivet S, Eskeland R, Boyle S, Illingworth RS, Paquette D, et al. Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization. *Genes Dev*. 2014; 28:2778–91. [PubMed: 25512564]
73. Ing-Simmons E, Seitan VC, Faure AJ, Flicek P, Carroll T, Dekker J, et al. Spatial enhancer clustering and regulation of enhancer-proximal genes by cohesin. *Genome Res*. 2015; 25:504–13. [PubMed: 25677180]
74. Seitan VC, Faure AJ, Zhan Y, McCord RP, Lajoie BR, Ing-Simmons E, et al. Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. *Genome Res*. 2013; 23:2066–77. [PubMed: 24002784]
75. Zuin J, Dixon JR, van der Reijden MI, Ye Z, Kolovos P, Brouwer RW, et al. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc Natl Acad Sci U S A*. 2014; 111:996–1001. [PubMed: 24335803]
76. Kosak ST, Skok JA, Medina KL, Riblet R, Le Beau MM, Fisher AG, et al. Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science*. 2002; 296:158–62. [PubMed: 11935030]
77. Chambeyron S, Bickmore WA. Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev*. 2004; 18:1119–30. [PubMed: 15155579]
78. Volpi EV, Chevret E, Jones T, Vatcheva R, Williamson J, Beck S, et al. Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. *J Cell Sci*. 2000; 113(Pt 9):1565–76. [PubMed: 10751148]
79. Lee HY, Johnson KD, Boyer ME, Bresnick EH. Relocalizing genetic loci into specific subnuclear neighborhoods. *The Journal of biological chemistry*. 2011; 286:18834–44. [PubMed: 21398517]
80. Schoenfelder S, Sexton T, Chakalova L, Cope NF, Horton A, Andrews S, et al. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat Genet*. 2010; 42:53–61. [PubMed: 20010836]
81. Fanucchi S, Shibayama Y, Burd S, Weinberg MS, Mhlanga MM. Chromosomal contact permits transcription between coregulated genes. *Cell*. 2013; 155:606–20. [PubMed: 24243018]
82. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. *Nature*. 2014; 507:455–61. [PubMed: 24670763]
83. Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell*. 2012; 148:84–98. [PubMed: 22265404]
84. Agelopoulos M, McKay DJ, Mann RS. Developmental regulation of chromatin conformation by Hox proteins in *Drosophila*. *Cell reports*. 2012; 1:350–9. [PubMed: 22523743]
85. Andrey G, Montavon T, Mascrez B, Gonzalez F, Noordermeer D, Leleu M, et al. A switch between topological domains underlies HoxD genes collinearity in mouse limbs. *Science*. 2013; 340:1234167. [PubMed: 23744951]
86. van Arensbergen J, van Steensel B, Bussemaker HJ. In search of the determinants of enhancer-promoter interaction specificity. *Trends Cell Biol*. 2014; 24:695–702. [PubMed: 25160912]
87. Butler JE, Kadonaga JT. Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. *Genes Dev*. 2001; 15:2515–9. [PubMed: 11581157]
88. Juven-Gershon T, Hsu JY, Kadonaga JT. Caudal, a key developmental regulator, is a DPE-specific transcriptional factor. *Genes Dev*. 2008; 22:2823–30. [PubMed: 18923080]
89. Ohtsuki S, Levine M, Cai HN. Different core promoters possess distinct regulatory activities in the *Drosophila* embryo. *Genes Dev*. 1998; 12:547–56. [PubMed: 9472023]
90. Akbari OS, Bae E, Johnsen H, Villaluz A, Wong D, Drewell RA. A novel promoter-tethering element regulates enhancer-driven gene expression at the bithorax complex in the *Drosophila* embryo. *Development*. 2008; 135:123–31. [PubMed: 18045839]

91. Kwon D, Mucci D, Langlais KK, Americo JL, DeVido SK, Cheng Y, et al. Enhancer-promoter communication at the *Drosophila* engrailed locus. *Development*. 2009; 136:3067–75. [PubMed: 19675130]
92. Levine M. Transcriptional enhancers in animal development and evolution. *Curr Biol*. 2010; 20:R754–63. [PubMed: 20833320]
93. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*. 2009; 459:108–12. [PubMed: 19295514]
94. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. The accessible chromatin landscape of the human genome. *Nature*. 2012; 489:75–82. [PubMed: 22955617]
95. Kieffer-Kwon KR, Tang Z, Mathe E, Qian J, Sung MH, Li G, et al. Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. *Cell*. 2013; 155:1507–20. [PubMed: 24360274]
96. Nichols J, Smith A. Naive and primed pluripotent states. *Cell Stem Cell*. 2009; 4:487–92. [PubMed: 19497275]
97. Tesar PJ, Chenoweth JG, Brook FA, Davies TJ, Evans EP, Mack DL, et al. New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature*. 2007; 448:196–9. [PubMed: 17597760]
98. Boiani M, Scholer HR. Regulatory networks in embryo-derived pluripotent stem cells. *Nat Rev Mol Cell Biol*. 2005; 6:872–84. [PubMed: 16227977]
99. Van Bortle K, Corces VG. Lost in transition: dynamic enhancer organization across naive and primed stem cell states. *Cell Stem Cell*. 2014; 14:693–4. [PubMed: 24905156]
100. Factor DC, Corradin O, Zentner GE, Saiakhova A, Song L, Chenoweth JG, et al. Epigenomic comparison reveals activation of “seed” enhancers during transition from naive to primed pluripotency. *Cell Stem Cell*. 2014; 14:854–63. [PubMed: 24905169]
101. Buecker C, Srinivasan R, Wu Z, Calo E, Acampora D, Faial T, et al. Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell*. 2014; 14:838–53. [PubMed: 24905168]
102. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*. 2009; 462:315–22. [PubMed: 19829295]
103. Ziller MJ, Gu H, Muller F, Donaghey J, Tsai LT, Kohlbacher O, et al. Charting a dynamic DNA methylation landscape of the human genome. *Nature*. 2013; 500:477–81. [PubMed: 23925113]
104. Hon GC, Rajagopal N, Shen Y, McCleary DF, Yue F, Dang MD, et al. Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat Genet*. 2013; 45:1198–206. [PubMed: 23995138]
105. Whyte WA, Bilodeau S, Orlando DA, Hoke HA, Frampton GM, Foster CT, et al. Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature*. 2012; 482:221–5. [PubMed: 22297846]
106. Nord AS, Blow MJ, Attanasio C, Akiyama JA, Holt A, Hosseini R, et al. Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell*. 2013; 155:1521–31. [PubMed: 24360275]
107. Domazet-Loso T, Tautz D. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature*. 2010; 468:815–8. [PubMed: 21150997]
108. Kalinka AT, Varga KM, Gerrard DT, Preibisch S, Corcoran DL, Jarrells J, et al. Gene expression divergence recapitulates the developmental hourglass model. *Nature*. 2010; 468:811–4. [PubMed: 21150996]
109. Stergachis AB, Neph S, Reynolds A, Humbert R, Miller B, Paige SL, et al. Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. *Cell*. 2013; 154:888–903. [PubMed: 23953118]
110. Heinz S, Romanoski CE, Benner C, Glass CK. The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol*. 2015; 16:144–54. [PubMed: 25650801]

111. Zaret KS, Watts J, Xu J, Wandzioch E, Smale ST, Sekiya T. Pioneer factors, genetic competence, and inductive signaling: programming liver and pancreas progenitors from the endoderm. *Cold Spring Harb Symp Quant Biol.* 2008; 73:119–26. [PubMed: 19028990]
112. Gautier EL, Shay T, Miller J, Greter M, Jakubzick C, Ivanov S, et al. Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. *Nat Immunol.* 2012; 13:1118–28. [PubMed: 23023392]
113. Okabe Y, Medzhitov R. Tissue-specific signals control reversible program of localization and functional polarization of macrophages. *Cell.* 2014; 157:832–44. [PubMed: 24792964]
114. Gosselin D, Link VM, Romanoski CE, Fonseca GJ, Eichenfield DZ, Spann NJ, et al. Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. *Cell.* 2014; 159:1327–40. [PubMed: 25480297]
115. Sakabe NJ, Savic D, Nobrega MA. Transcriptional enhancers in development and disease. *Genome Biol.* 2012; 13:238. [PubMed: 22269347]
116. Corradin O, Scacheri PC. Enhancer variants: evaluating functions in common disease. *Genome Med.* 2014; 6:85. [PubMed: 25473424]
117. Barr CL, Misener VL. Decoding the Non-Coding Genome: Elucidating Genetic Risk Outside the Coding Genome. *Genes Brain Behav.* 2015
118. Visel A, Rubin EM, Pennacchio LA. Genomic views of distant-acting enhancers. *Nature.* 2009; 461:199–205. [PubMed: 19741700]
119. Edwards SL, Beesley J, French JD, Dunning AM. Beyond GWASs: illuminating the dark road from association to function. *Am J Hum Genet.* 2013; 93:779–97. [PubMed: 24210251]
120. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, et al. A map of human genome variation from population-scale sequencing. *Nature.* 2010; 467:1061–73. [PubMed: 20981092]
121. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science.* 2012; 337:1190–5. [PubMed: 22955828]
122. Ernst J, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature.* 2011; 473:43–9. [PubMed: 21441907]
123. Trynka G, Sandor C, Han B, Xu H, Stranger BE, Liu XS, et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat Genet.* 2013; 45:124–30. [PubMed: 23263488]
124. Gusev A, Lee SH, Trynka G, Finucane H, Vilhjalmsdottir BJ, Xu H, et al. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am J Hum Genet.* 2014; 95:535–52. [PubMed: 25439723]
125. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, et al. Super-enhancers in the control of cell identity and disease. *Cell.* 2013; 155:934–47. [PubMed: 24119843]
126. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell.* 2013; 153:307–19. [PubMed: 23582322]
127. Pott S, Lieb JD. What are super-enhancers? *Nat Genet.* 2015; 47:8–12. [PubMed: 25547603]
128. Niederriter AR, Varshney A, Parker SC, Martin DM. Super Enhancers in Cancers, Complex Disease, and Developmental Disorders. *Genes (Basel).* 2015; 6:1183–200. [PubMed: 26569311]
129. Lowe WL Jr, Reddy TE. Genomic approaches for understanding the genetics of complex disease. *Genome Res.* 2015; 25:1432–41. [PubMed: 26430153]
130. Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. *Nat Rev Genet.* 2015; 16:197–212. [PubMed: 25707927]
131. McPherson JD, Marra M, Hillier L, Waterston RH, Chinwalla A, Wallis J, et al. A physical map of the human genome. *Nature.* 2001; 409:934–41. [PubMed: 11237014]
132. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature.* 2001; 409:860–921. [PubMed: 11237011]

133. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, et al. Initial sequencing and comparative analysis of the mouse genome. *Nature*. 2002; 420:520–62. [PubMed: 12466850]
134. Boffelli D, Nobrega MA, Rubin EM. Comparative genomics at the vertebrate extremes. *Nat Rev Genet*. 2004; 5:456–65. [PubMed: 15153998]
135. Ahituv N, Rubin EM, Nobrega MA. Exploiting human--fish genome comparisons for deciphering gene regulation. *Hum Mol Genet*. 2004; 13(Spec No 2):R261–6. [PubMed: 15358733]
136. Nobrega MA, Ovcharenko I, Afzal V, Rubin EM. Scanning human gene deserts for long-range enhancers. *Science*. 2003; 302:413. [PubMed: 14563999]
137. Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, et al. In vivo enhancer analysis of human conserved non-coding sequences. *Nature*. 2006; 444:499–502. [PubMed: 17086198]
138. Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, et al. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*. 2011; 478:476–82. [PubMed: 21993624]
139. Lettice LA, Heaney SJ, Purdie LA, Li L, de Beer P, Oostra BA, et al. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet*. 2003; 12:1725–35. [PubMed: 12837695]
140. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008; 456:53–9. [PubMed: 18987734]
141. Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol*. 2008; 26:1135–45. [PubMed: 18846087]
142. Wold B, Myers RM. Sequence census methods for functional genomics. *Nat Methods*. 2008; 5:19–21. [PubMed: 18165803]
143. Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res*. 2005; 33:5868–77. [PubMed: 16224102]
144. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*. 2008; 5:621–8. [PubMed: 18516045]
145. Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of in vivo protein-DNA interactions. *Science*. 2007; 316:1497–502. [PubMed: 17540862]
146. Crawford GE, Holt IE, Whittle J, Webb BD, Tai D, Davis S, et al. Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). *Genome Res*. 2006; 16:123–31. [PubMed: 16344561]
147. John S, Sabo PJ, Canfield TK, Lee K, Vong S, Weaver M, et al. Genome-scale mapping of DNase I hypersensitivity. *Curr Protoc Mol Biol*. 2013; Chapter 27(Unit 21):7.
148. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science*. 2002; 295:1306–11. [PubMed: 11847345]
149. Dostie J, Dekker J. Mapping networks of physical interactions between genomic elements using 5C technology. *Nat Protoc*. 2007; 2:988–1002. [PubMed: 17446898]
150. Belton JM, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods*. 2012; 58:268–76. [PubMed: 22652625]
151. Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, et al. An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature*. 2009; 462:58–64. [PubMed: 19890323]
152. Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res*. 2012; 22:1813–31. [PubMed: 22955991]
153. Savic D, Partridge EC, Newberry KM, Smith SB, Meadows SK, Roberts BS, et al. CETCh-seq: CRISPR epitope tagging ChIP-seq of DNA-binding proteins. *Genome Res*. 2015; 25:1581–9. [PubMed: 26355004]
154. Savic D, Gertz J, Jain P, Cooper GM, Myers RM. Mapping genome-wide transcription factor binding sites in frozen tissues. *Epigenetics Chromatin*. 2013; 6:30. [PubMed: 24279905]

155. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*. 2013; 10:1213–8. [PubMed: 24097267]
156. Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*. 2008; 322:1845–8. [PubMed: 19056941]
157. Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L, et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat Genet*. 2015; 47:598–606. [PubMed: 25938943]
158. Stamatoiyannopoulos JA, Snyder M, Hardison R, Ren B, Gingeras T, et al. ENCODE. An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biol*. 2012; 13:418. [PubMed: 22889292]
159. Reddy TE, Pauli F, Sprouse RO, Neff NF, Newberry KM, Garabedian MJ, et al. Genomic determination of the glucocorticoid response reveals unexpected mechanisms of gene regulation. *Genome Res*. 2009; 19:2163–71. [PubMed: 19801529]
160. Gertz J, Reddy TE, Varley KE, Garabedian MJ, Myers RM. Genistein and bisphenol A exposure cause estrogen receptor 1 to bind thousands of sites in a cell type-specific manner. *Genome Res*. 2012; 22:2153–62. [PubMed: 23019147]
161. Cheng Y, Wu W, Kumar SA, Yu D, Deng W, Tripic T, et al. Erythroid GATA1 function revealed by genome-wide analysis of transcription factor occupancy, histone modifications, and mRNA expression. *Genome Res*. 2009; 19:2172–84. [PubMed: 19887574]
162. Patwardhan RP, Hiatt JB, Witten DM, Kim MJ, Smith RP, May D, et al. Massively parallel functional dissection of mammalian enhancers in vivo. *Nat Biotechnol*. 2012; 30:265–70. [PubMed: 22371081]
163. White MA, Myers CA, Corbo JC, Cohen BA. Massively parallel in vivo enhancer assay reveals that highly local features determine the cis-regulatory function of ChIP-seq peaks. *Proc Natl Acad Sci U S A*. 2013; 110:11952–7. [PubMed: 23818646]
164. Kheradpour P, Ernst J, Melnikov A, Rogov P, Wang L, Zhang X, et al. Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res*. 2013; 23:800–11. [PubMed: 23512712]
165. Arnold CD, Gerlach D, Stelzer C, Boryn LM, Rath M, Stark A. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science*. 2013; 339:1074–7. [PubMed: 23328393]
166. Kwasniewski JC, Mogno I, Myers CA, Corbo JC, Cohen BA. Complex effects of nucleotide variants in a mammalian cis-regulatory element. *Proc Natl Acad Sci U S A*. 2012; 109:19498–503. [PubMed: 23129659]
167. Kwasniewski JC, Fiore C, Chaudhari HG, Cohen BA. High-throughput functional testing of ENCODE segmentation predictions. *Genome Res*. 2014; 24:1595–602. [PubMed: 25035418]
168. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science*. 2013; 339:819–23. [PubMed: 23287718]
169. Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J. RNA-programmed genome editing in human cells. *Elife*. 2013; 2:e00471. [PubMed: 23386978]
170. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-guided human genome engineering via Cas9. *Science*. 2013; 339:823–6. [PubMed: 23287722]
171. Miller JC, Holmes MC, Wang J, Guschin DY, Lee YL, Rupniewski I, et al. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol*. 2007; 25:778–85. [PubMed: 17603475]
172. Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, et al. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol*. 2011; 29:143–8. [PubMed: 21179091]
173. Altschuler SJ, Wu LF. Cellular heterogeneity: do differences make a difference? *Cell*. 2010; 141:559–63. [PubMed: 20478246]
174. Marusyk A, Polyak K. Tumor heterogeneity: causes and consequences. *Biochim Biophys Acta*. 2010; 1805:105–17. [PubMed: 19931353]
175. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods*. 2009; 6:377–82. [PubMed: 19349980]

176. Hashimshony T, Wagner F, Sher N, Yanai I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.* 2012; 2:666–73. [PubMed: 22939981]
177. Kolodziejczyk AA, Kim JK, Svensson V, Marioni JC, Teichmann SA. The technology and biology of single-cell RNA sequencing. *Mol Cell.* 2015; 58:610–20. [PubMed: 26000846]
178. Smallwood SA, Lee HJ, Angermueller C, Krueger F, Saadeh H, Peat J, et al. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat Methods.* 2014; 11:817–20. [PubMed: 25042786]
179. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature.* 2015; 523:486–90. [PubMed: 26083756]
180. Cusanovich DA, Daza R, Adey A, Pliner HA, Christiansen L, Gunderson KL, et al. Epigenetics. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science.* 2015; 348:910–4. [PubMed: 25953818]
181. Kim KT, Lee HW, Lee HO, Kim SC, Seo YJ, Chung W, et al. Single-cell mRNA sequencing identifies subclonal heterogeneity in anti-cancer drug responses of lung adenocarcinoma cells. *Genome Biol.* 2015; 16:127. [PubMed: 26084335]
182. Buettner F, Natarajan KN, Casale FP, Proserpio V, Scialdone A, Theis FJ, et al. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat Biotechnol.* 2015; 33:155–60. [PubMed: 25599176]
183. Pott S, Lieb JD. Single-cell ATAC-seq: strength in numbers. *Genome Biol.* 2015; 16:172. [PubMed: 26294014]
184. Lara-Astiaso D, Weiner A, Lorenzo-Vivas E, Zaretzky I, Jaitin DA, David E, et al. Immunogenetics. Chromatin state dynamics during blood formation. *Science.* 2014; 345:943–9. [PubMed: 25103404]
185. Brind'Amour J, Liu S, Hudson M, Chen C, Karimi MM, Lorincz MC. An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. *Nat Commun.* 2015; 6:6033. [PubMed: 25607992]
186. Wang T, Wei JJ, Sabatini DM, Lander ES. Genetic screens in human cells using the CRISPR-Cas9 system. *Science.* 2014; 343:80–4. [PubMed: 24336569]
187. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science.* 2014; 343:84–7. [PubMed: 24336571]
188. Parnas O, Jovanovic M, Eisenhaure TM, Herbst RH, Dixit A, Ye CJ, et al. A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. *Cell.* 2015; 162:675–86. [PubMed: 26189680]
189. Chen S, Sanjana NE, Zheng K, Shalem O, Lee K, Shi X, et al. Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell.* 2015; 160:1246–60. [PubMed: 25748654]
190. Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, Joung JK. CRISPR RNA-guided activation of endogenous human genes. *Nat Methods.* 2013; 10:977–9. [PubMed: 23892898]
191. Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, et al. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods.* 2013; 10:973–6. [PubMed: 23892895]
192. Hilton IB, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, Reddy TE, et al. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol.* 2015; 33:510–7. [PubMed: 25849900]
193. Thakore PI, D'Ippolito AM, Song L, Safi A, Shivakumar NK, Kabadi AM, et al. Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat Methods.* 2015
194. Esvelt KM, Mali P, Braff JL, Moosburner M, Yaung SJ, Church GM. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat Methods.* 2013; 10:1116–21. [PubMed: 24076762]
195. Kleinstiver BP, Prew MS, Tsai SQ, Topkar VV, Nguyen NT, Zheng Z, et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature.* 2015; 523:481–5. [PubMed: 26098369]

196. Spisak S, Lawrenson K, Fu Y, Csabai I, Cottman RT, Seo JH, et al. CAUSEL: an epigenome- and genome-editing pipeline for establishing function of noncoding GWAS variants. *Nat Med.* 2015; 21:1357–63. [PubMed: 26398868]
197. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature.* 2009; 461:747–53. [PubMed: 19812666]
198. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, et al. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet.* 2010; 11:446–50. [PubMed: 20479774]
199. Anderson MW, Schrijver I. Next generation DNA sequencing and the future of genomic medicine. *Genes (Basel).* 2010; 1:38–69. [PubMed: 24710010]
200. Gullapalli RR, Desai KV, Santana-Santos L, Kant JA, Becich MJ. Next generation sequencing in clinical medicine: Challenges and lessons for pathology and biomedical informatics. *J Pathol Inform.* 2012; 3:40. [PubMed: 23248761]
201. Lin E, Chien J, Ong FS, Fan JB. Challenges and opportunities for next-generation sequencing in companion diagnostics. *Expert Rev Mol Diagn.* 2015; 15:193–209. [PubMed: 25249308]
202. Cooper GM, Shendure J. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nat Rev Genet.* 2011; 12:628–40. [PubMed: 21850043]
203. Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet.* 2014; 46:310–5. [PubMed: 24487276]
204. Brennand KJ, Simone A, Jou J, Gelboin-Burkhart C, Tran N, Sangar S, et al. Modelling schizophrenia using human induced pluripotent stem cells. *Nature.* 2011; 473:221–5. [PubMed: 21490598]
205. Israel MA, Yuan SH, Bardy C, Reyna SM, Mu Y, Herrera C, et al. Probing sporadic and familial Alzheimer’s disease using induced pluripotent stem cells. *Nature.* 2012; 482:216–20. [PubMed: 22278060]
206. Shen SQ, Myers CA, Hughes AE, Byrne LC, Flannery JG, Corbo JC. Massively parallel cis-regulatory analysis in the mammalian central nervous system. *Genome Res.* 2015
207. Vockley CM, Guo C, Majoros WH, Nodzenski M, Scholtens DM, Hayes MG, et al. Massively parallel quantification of the regulatory effects of noncoding genetic variation in a human cohort. *Genome Res.* 2015; 25:1206–14. [PubMed: 26084464]

Figure 1. Multiple levels of genome architecture and gene regulation

A schematic of distinct levels of gene regulatory control and nuclear architecture is given. Structural features of the nucleus including lamina associated domains (LADs) and transcription factories are shown. A topologically associated domain (TAD) within a transcription factory is given below. Intra-TAD enhancer-promoter contacts are displayed. A closer look at enhancer looping between intra-TAD enhancer-promoter interactions is given in the last panel. Chromatin state near active regulatory elements and promoter sequences are altered through histone modifications and correlate with enhancer state as well as gene expression levels. Enhancer elements expressing eRNAs and bound by transcription factors (TF) and associated transcriptional cofactors and chromatin remodeling enzymes (COF) associate with distal promoter sequences through long-range looping interactions, leading to increased expression of target genes.

Figure 2. Next-generation sequencing based disease studies

Experimental approaches that can be applied for assessing the role of enhancers or other non-coding, regulatory variants in diseases are shown. (A) A schematic of patient-derived iPSC models is given. Reprogrammed patient iPSCs into relevant cell types can be used for functional genomic studies and cellular phenotypic analyses. These functional genomic studies can identify key elements using an integrative approach that incorporates complementary information, including DNA methylation (DNA Methyl), DNA-binding protein occupancy (TF), regions of open chromatin (Chrom), histone modifications (Histone), eRNA production (eRNA), long range interactions (3D) and RNA expression (mRNA) to identify regulatory elements (1). Subsequent CRISPR genome editing of key regulatory elements is used to validate target genes (2), whereas regulatory element swapping can be used to predict the functional effect of regulatory sequence variants (3). (B) A high-throughput reporter assay screen is shown. After construction and transfection of a complex pool of reporter plasmids harboring hundreds of enhancer sequences from case and control populations, NGS can identify both rare and common variants within enhancer sequences that lead to functional effects on gene expression. (P = Promoter; GFP = Green Fluorescent Protein).