Transcriptional Enhancers: Bridging the Genome and Phenome

BING REN1 AND FENG YUE2

1Ludwig Institute for Cancer Research, University of California, San Diego School of Medicine, Department of Cellular and Molecular Medicine, Institute of Genomic Medicine, and Moores Cancer Center, La Jolla, California 92037-0653
2Department of Biochemistry and Molecular Biology and Institute for Personalized Medicine, Pennsylvania State University College of Medicine, Hershey, Pennsylvania 17033

Correspondence: biren@ucsd.edu

Enhancers play a major role in animal development by modulating spatiotemporal expression of genes. They interact with sequence-specific transcriptional regulators in response to internal and external cues to bring about transcriptional changes, thus serving as the critical link between an organism’s genome and its phenotypic traits. Deciphering the biology of enhancers is a key to understanding the genetic basis of common human diseases. Although a large number of candidate enhancers have been annotated through genome-wide analyses of chromatin accessibility, transcription factor binding, and histone modification in diverse cell types, efforts to characterize their biological roles in human diseases have only begun. Recent experiments have suggested a role for the three-dimensional chromatin architecture in regulation of gene expression by enhancers.

The genome of each person differs from one another at millions of nucleotides, and these sequence variants together are responsible for the spectrum of phenotypic traits and disease risks of that individual (Carlson et al. 2004). Although it has become commonplace these days to sequence one’s genome, predicting the specific phenotypic traits of each individual from DNA still seems an insurmountable challenge. This is because >98% of the human genome is non-protein-coding and generally without a clearly defined biological function (Lander 2011). On the other hand, these “junk DNAs,” as previously called, have now been shown to harbor numerous functional elements with important roles in gene regulation and disease pathogenesis (ENCODE Project Consortium 2012). In particular, scattered in these sequences are millions of putative cis-regulatory elements responsible for spatiotemporal gene expression during development. Furthermore, a large number of sequence variants in the cis-regulatory elements are believed to confer risks to various common human diseases (Ernst et al. 2011; Maurano et al. 2012). Therefore, identifying and characterizing the cis-regulatory elements in the human genome have the potential to significantly enhance our ability to link DNA variations to phenotypic traits.

Enhancers are a class of cis-regulatory DNA sequences that play a major role in cell type–specific gene expression during animal development (Levine 2010; Bulger and Groudine 2011). They are capable of modulating the expression of target genes from a distance that could extend more than 1 million base pairs (bp) (Levine et al. 2014). Their mechanisms of function involve the combinatorial binding of sequence-specific transcription factors, which in turn recruit chromatin remodeling complexes, alteration of local chromatin structure, formation of chromatin interactions with target gene promoters, and transcriptional initiation or elongation of RNA (Fig. 1; Calo and Wysocka 2013; Smith and Shilatifard 2014). Therefore, recognition of specific enhancer sequences by distinct sets of sequence-specific transcription factors is the primary mechanism by which DNA sequences dictate gene expression programs.

Thanks to the development of high-throughput technologies (Hawkins et al. 2010), hundreds of thousands of candidate enhancers have been annotated in the human genome, compromising at least 12% of the total DNA sequences (ENCODE Project Consortium 2012; Roadmap Epigenomics Consortium et al. 2015). A majority of these sequences display chromatin accessibility or characteristic patterns of histone modifications in a cell type–specific fashion, supporting their roles in mediating cell type–specific gene expression programs. Further, the sequences are enriched in transcription factor–binding motifs, providing clues to the lineage-specific transcription factors that work at these sequences. Of particular interest to the biomedical researchers, the identified enhancers have been shown to harbor a significant fraction of the noncoding sequence variants associated with increased risks of common human diseases, highlighting the importance of studying these elements in biomedical research (Ernst et al. 2011; Maurano et al. 2012).

Despite these major advances in identifying candidate enhancers in the human genome, a number of important questions remain. For example, what are the biological functions of the annotated enhancers? How do they contribute to cell type–specific gene expression programs? How are their activities regulated in response to extracel-
ular and intracellular signaling to achieve spatiotemporal regulation of target genes during development? How do sequence variants affect their function and contribute to physiological traits and disease risks? The following sections attempt to provide a general context of the problems and summarize the latest efforts to address these questions.

Identifying Enhancers in the Genome

Enhancers were initially identified as a class of sequences that can activate a heterologous promoter from a distance in an orientation-independent manner (Banerji et al. 1981; Moreau et al. 1981). This functional definition is not easy to use in the genome-wide identification of enhancers—it is simply impractical to test all human DNA sequences across all cell types in the body under all possible physiological conditions to determine whether any of them functions as an enhancer.

Early on, it was recognized that cis-regulatory sequences including enhancers generally display DNase I hypersensitivity (Tuan and London 1984). This strategy led to the discovery of a number of mammalian enhancers, including the β-globin locus control region (LCR) (Grosveld et al. 1993; Martin et al. 1996) critical for the developmental control of β-globin genes in mammals. This approach has undergone several iterations over the years, from the laborious Southern blot analysis (Weintraub and Groudine 1976; Herbomel et al. 1981; Tuan and London 1984) to the high-throughput microarray experiments (Sabó et al. 2006) and the ultrahigh-throughput DNA-sequencing analysis that has been now broadly applied to hundreds of cell types (Boyle et al. 2008; Hesselberth et al. 2009; Maurano et al. 2012; Vierstra et al. 2014). The latest iteration of this assay is the use of a genetically engineered Tn5 transposase, instead of DNase I, in a highly streamlined experimental protocol called ATAC-seq (assay for transposase-accessible chromatin with high-throughput sequencing) (Buenrostro et al. 2013) that is especially amendable to very small cell numbers and even single cells (Buenrostro et al. 2015; Cusanovich et al. 2015).

However, DNase I hypersensitivity is not unique to enhancers. Other cis elements such as promoters and insulators also show this property (ENCODE Project Consortium 2012). Therefore, just the presence of DNase I hypersensitivity alone is insufficient to distinguish these different classes of cis elements. Are there other biochemical features that could be used to separate different classes of cis elements?

With the development of ChIP-chip and recently ChIP-seq to map transcription factor–binding and chromatin modification patterns in the genome (Kim et al. 2005; Barski et al. 2007; Johnson et al. 2007; Mikkelsen et al. 2007), it was possible to show that promoters, enhancers, and insulators are indeed differentially occupied by transcription factors and are associated with distinct chromatin modification patterns (Heintzman et al. 2007; Kim et al. 2007; Creyghton et al. 2010; Rada-Iglesias et al. 2011). This then allowed genome-wide determination of candidate enhancers. For example, binding sites of p300, a coactivator protein and histone acetyltransferase that acts at enhancers, led to annotation of several thousand enhancers in embryonic brain and limbs in the mouse (Visel et al. 2009). Mapping of DNA occupancy of sequence-specific transcription factors such as Sox2, Oct4, and Nanog also provided the first map of enhancers in the human embryonic stem cells (Boyer et al. 2005; Chen et al. 2008). With the increased understanding of the role of histone modification processes in nucleosome dynamics and chromatin organization, it was natural to examine the chromatin modification patterns in the genome, since different classes of cis elements may recruit different chromatin remodeling proteins that generate distinct patterns of histone modifications. Indeed, this hypothesis was supported by the observation that enhancers are as-
associated with high levels of H3K4me1 and low levels of H3K4me3, while the opposite is true for promoters (Heintzman et al. 2007). The finding led to the development of chromatin signature–based strategy for mapping enhancers, and identification of such DNA in both mouse and human genomes (Heintzman et al. 2007; Ernst and Kellis 2012; Hoffman et al. 2012; Rajagopal et al. 2013). Computational algorithms have been developed to identify chromatin modification patterns in an unsupervised manner, and then match the different patterns to different classes of candidate functional elements, resulting in a chromatin state–based vocabulary for “reading” the human genome (Hon et al. 2008; Ernst and Kellis 2012). This later strategy has been applied to the chromatin modification maps from more than 120 human tissue and cell types, providing the most comprehensive annotation of the human genome to date (Table 1; Roadmap Epigenomics Consortium et al. 2015). Similarly, the same approach has also resulted in mapping of cis elements in the mouse genome (Fig. 2; Table 1; Shen et al. 2012; Yue et al. 2014).

Among the large number of candidate enhancers predicted in the human and mouse genomes, a small fraction of them drew special attention. These sequences, referred to as super-enhancers, are dense clusters of enhancers and are found in virtually every mammalian cell and tissue type examined (Hnisz et al. 2013). They are bound by an extraordinary amount of transcription factors and cofactors and are typically associated with high levels of transcription activities (Whyte et al. 2013) in specific cell types. Found particularly enriched at cell identity genes, super-enhancers are regarded as the key control elements of lineage specification.

Enhancers also show cell type–specific hypomethylation of cytosine (Stadler et al. 2011; Hon et al. 2013; Table 1). Public resources for enhancer annotations in the human and mouse genomes

| Species | Paper Method | Link |
|---------|--------------|------|
| Human | Roadmap Epigenomics Consortium et al. 2015 | ChromHMM (Ernst and Kellis 2012) | http://egg2.wustl.edu/roadmap/web_portal/ |
| Human | ENCODE Project Consortium 2012 | Joint call by chromHMM (Ernst and Kellis 2012) + Segway (Hoffman et al. 2012) | http://genome.ucsc.edu/cgi-bin/hgHubConnect click Connect: ENCODE Analysis Hub |
| Human | Unpublished | DNase-I hypersensitive sites (DHS) + H3K27ac | https://www.encodeproject.org/data/annotations/ http://promoter.bx.psu.edu/ENCODE/ |
| Mouse | Yue et al. 2014 | RFECS (Rajagopal et al. 2013) | http://www.mouseencode.org |
| Human | Andersson et al. 2014 | Cap analysis of gene expression (CAGE) | http://fantom.gsc.riken.jp/5/data/ |

Table 1. Public resources for enhancer annotations in the human and mouse genomes

Figure 2. ChIP-seq analysis identifies candidate cis-regulatory elements in the mouse genome. Shown in the UCSC (University of California, Santa Cruz) genome browser are ChIP-seq data for several chromatin modification marks, transcription factor CTCF (CCCTC-binding factor), coactivator protein p300, and RNA polymerase II binding in mouse embryonic stem cell, heart, and liver. The input normalized ChIP-seq data is shown. (Adapted from Shen et al. 2012.)
This DNA hypomethylation at enhancers likely involves oxidation of methylcytosine that is subsequently converted to unmethylated cytosine in a Tet-dependent process (Yu et al. 2012; Shen et al. 2013; Song et al. 2013; Hon et al. 2014). Thus, cell type–specific, differentially methylated DNA sequences can be used to define enhancers in different cell types (Stadler et al. 2011; Hon et al. 2013; Ziller et al. 2013; Schultz et al. 2015). Indeed, applying this strategy to several dozen human tissues and cell types, hundreds of thousands of enhancers have been predicted in the human genome and mouse genome (Hon et al. 2013; Ziller et al. 2013; Schultz et al. 2015).

**FUNCTIONAL CHARACTERIZATION OF ANNOTATED ENHANCERS**

Now that a large number of candidate enhancers have been annotated in the human genome, the next pressing question is whether these predicted enhancers are truly functional, and, if so, how do they contribute to gene regulation and development. Specifically, in what biological contexts do they function? What genes do they control and how much do they contribute to the target genes’ expression in vivo?

Given the large number of candidate enhancers annotated in the human genome, high-throughput assays are needed to characterize the function of these sequences. To this end, massively parallel reporter assays (MPRAs) have been developed to validate the identity of enhancers (Melnikov et al. 2012; Patwardhan et al. 2012; Arnold et al. 2013; Smith et al. 2013; White et al. 2013). MPRA and related methods involve the construction of a library of reporter plasmids containing the test sequence placed either upstream of or downstream from a barcode-containing reporter gene driven by a heterologous minimal promoter. The barcode sequences, designed to be unique for each tested element, are transcribed along with the reporter gene. Their abundances could be accurately determined by sequencing and used to indicate activities of the corresponding enhancer. Such assays have confirmed that a major fraction of the predicted enhancers are indeed capable of driving reporter genes in cultured cells, providing strong evidence that biochemical signature–defined enhancers are actually what they are thought to be (Kheradpour et al. 2013; Yue et al. 2014).

One drawback of using cultured cells lies in the fact that they are unlikely to fully capture the native context where an enhancer functions. Overcoming this difficulty is the mouse transgenic assays (Pennacchio et al. 2006). In this assay, a reporter construct is injected into fertilized mouse eggs, which are then implanted back into the womb to allow embryonic development to proceed. At a certain developmental stage, say Day 11.5, the embryos are dissected and examined for reporter expression. This assay, although of modest throughput, permits the determination of tissue-specific activities of enhancers in all the embryonic tissues. So far, several thousand elements have been tested in a systematic manner using this strategy, and a database, VISTA enhancer browser, has been set up to distribute the results (Visel et al. 2007). A significant percentage of enhancers predicted using p300 or H3K27ac was shown to possess bona fide enhancer function in vivo (Visel et al. 2009; Nord et al. 2013).

Reporter-based assays, whether in cultured cells or transgenic animals, provide only limited functional information about an enhancer. Neither the target genes of the enhancer nor the degree of transcriptional modulation on target gene expression can be determined. To address this problem, old-fashioned knockout technologies or more recent genome editing approaches are necessary (Attanasio et al. 2013). In particular, the CRISPR/cas9 genome editing technologies have greatly facilitated the deployment of this strategy (Doudna and Charpentier 2014; Shalem et al. 2015). The function of a growing number of candidate enhancers in the mouse genome has now been tested (Li et al. 2014; Zhou et al. 2014; Hnisz et al. 2015). One important consideration in the experimental design is to correctly separate direct cis-regulatory effects from indirect effects of down-regulation of the target genes. A strategy to rule out indirect effects involves the construction of reciprocal, monoallelic enhancer deletion and analysis of allelic gene expression patterns in the resulting mutants (Fig. 3). Such analysis was used to determine the mode of cis regulation of Sox2 by a distal enhancer (Li et al. 2014; Zhou et al. 2014).

With a growing number of enhancers having been investigated in their native contexts, new lessons of enhancer biology are emerging. Results showed that at the Sox2 gene, an enhancer cluster (referred to as a super-enhancer) plays a primary role for Sox2 expression in mouse embryonic stem cells (mESCs) (Li et al. 2014; Zhou et al. 2014). Interestingly, other annotated enhancers much closer to Sox2 did not appear to contribute to Sox2 expression, as no change in Sox2 transcription was observed on deletion of these elements (Zhou et al. 2014) individually or in combination. This surprising result may suggest that many biochemical signature–defined enhancers may not necessarily function as enhancers in the cells that they were defined in. This result highlights the challenges that face the community about understanding the in vivo function of enhancers. Clearly, much more needs to be done in this area to achieve a more comprehensive understanding of the enhancer function in vivo.

**SELECTION OF TARGET GENES FOR ACTIVATION BY ENHANCERS**

Genetic analysis of a family of β-thalassemia patients provided the first line of evidence that enhancers could function over a long distance on DNA (Kioussis et al. 1983). Now, it is generally accepted that enhancers can activate genes far away, and the targets are not necessarily their immediate neighbors (Levine et al. 2014). Research of the β-globin gene LCR further indicates that enhancer
targets may change during development—in fetal tissues, the LCR regulates $\gamma$-globin, but in adult, it controls $\beta$-globin expression (Hanscombe et al. 1991). The mechanisms by which enhancers select their target genes in specific cell type are still incompletely understood.

It is now generally agreed that enhancers achieve long-range regulation of target genes because of spatial proximity formed as a result of the three-dimensional architecture of the chromosome (Levine 2010). Currently, the exact structure of chromosomes, the factors that regulate the organization, and the functions between the chromatin structure and gene expression are being extensively investigated (Gorkin et al. 2014). Much progress has been made in experimentally mapping long-range looping interactions (Fig. 4). The experimental strategies mainly involve chromatin conformation capture (3C)-based methods, including 4C, ChIA-PET, 5C, Hi-C, Capture-C, T2C, DNase-HiC, and, more recently, in situ Hi-C (a version of Hi-C in which proximity ligation is performed in situ in permeabilized nuclei) (Dostie et al. 2006; Fullwood et al. 2009; Simonis et al. 2009; Duan et al. 2010; Kalhor et al. 2012; Nagano et al. 2013; Hughes et al. 2014; Kolovos et al. 2014; Rao et al. 2014; Ma et al. 2015). Collectively, data from these different 3C technologies have allowed researchers to identify general features of genome organization and determine chromatin interactions at increasing resolution (Rao et al. 2014; Hsieh et al. 2015; Ma et al. 2015; Sahlen et al. 2015).

Studies using 3C-based molecular analysis of the 3D genome organization have shown that chromosomes consist of relatively stable chromatin domains, referred to as

---

**Figure 3.** Identification the direct target genes of an enhancer using CRISPR (clustered regularly interspaced short palindromic repeat)/Cas9 genome editing and allelic gene expression analysis. (A) H3K27ac ChIP-seq data over the Sox2 locus. The targeted enhancer is shaded blue. (B) Derivation of hybrid the CAST/129 F123 mouse embryonic stem (ES) cell line for genome editing experiments. (C) Strategy for double CRISPR targeting of a 13-kb enhancer located 130 kb downstream from the Sox2 gene. (D) Ratio of Sox2 gene transcript levels from the CAST and 129 alleles in two experiments where the deletion is on the CAST allele ($y$-axis) or the 129 allele ($x$-axis). (Adapted from Li et al. 2014.)

**Figure 4.** Diagram of a genome-wide 3C (Hi-C) experiment. Cells are fixed with formaldehyde, digested with HindIII restriction enzyme, end labeled with biotin conjugated nucleotides, and releighted. Ligated DNA fragments are isolated by shearing DNA and isolating ligation products using streptavidin-coated beads. DNA is then PCR amplified, sequenced, and mapped to give a comprehensive picture of chromatin interactions.
topologically associating domains (TADs) (Fig. 5; Dixon et al. 2012; Nora et al. 2012; Sexton et al. 2012). Although the structure of the TADs can change and disappear during mitosis, they always reestablish the same patterns after each cell division (Naumova et al. 2013). More surprisingly, TADs are largely the same in different cell types and throughout development, and the TAD boundaries are highly conserved between human and mouse (Fig. 5; Dixon et al. 2012; Vietri Rudan et al. 2015). The strong conservation of TADs during evolution suggests that TADs serve important biological roles, and one of them is likely to physically constrain interactions between enhancers and promoters, limiting enhancers to target genes located in the TADs. This hypothesis has received multiple lines of evidence. First, it was shown that correlation of chromatin state at enhancers and promoters is most pronounced within the same TADs (Nora et al. 2012; Shen et al. 2012; Waszak et al. 2015). Second, deletion of TAD boundaries can lead to ectopic interactions between enhancers and noncognate promoters, resulting in misregulation of gene expression in various cell types (Nora et al. 2012; Lupiáñez et al. 2015). Third, TAD boundaries are enriched for binding sites of CTCF protein, a factor previously showed to be essential for function of insulators—a class of cis elements that act to block enhancer/promoter interactions (Dixon et al. 2012; Vietri Rudan et al. 2015). Mutating the CTCF-binding sites at the TAD boundaries or depleting CTCF protein levels both lead to slightly increased inter-TAD interactions, supporting a critical role for this DNA-binding protein in the formation of TADs (Zuin et al. 2014; Narendra et al. 2015).

A computational approach has also been developed to infer the target genes based on the chromatin state and chromatin accessibility in the genome (Ernst et al. 2011; Neph et al. 2012; Shen et al. 2012; Thurman et al. 2012). This strategy assumes that an enhancer and its cognate target(s) would show coordinated transcriptional activities across a large panel of cell types. Based on this assumption, one can examine the large data sets of chromatin accessibility, modifications, and transcription maps to identify the pairs of promoter/enhancer that show significantly correlated patterns of activities. This strategy has resulted in the prediction of enhancer/promoter targets in human and mouse genomes (Neph et al. 2012; Shen et al. 2012; Thurman et al. 2012; Vierstra et al. 2014). Such predictions were supported by chromatin loop data and are found to be well within TADs (Shen et al. 2012). These results support a strong correlation between chromatin interactions and regulatory activities of enhancers. However, more in-depth analyses will be necessary to inform on which strategy works more accurately.

UNDERSTANDING FUNCTIONAL CONSEQUENCES OF DNA VARIATIONS IN ENHANCERS

A major goal of the recent National Institutes of Health (NIH) Precision Medicine Initiative is to develop disease prevention and treatment strategies that are tailored to each patient’s own genetic makeup (Collins and Varmus 2015). To achieve this goal requires a better understand-
ing of how the unique set of DNA variants in each individual led to the specific phenotypic traits. Enhancers occupy a crucial junction that connects genetic sequences to cellular function and consequently phenotypic traits of each individual. Study of enhancers is an essential component of the efforts to dissect the effects of noncoding variants on gene expression patterns.

Genome-wide association studies (GWASs) have identified a large number of sequence variants in the human genome that are linked to various physiological traits and common diseases. On the one hand, these sequence variants provided clues to the mechanisms by which DNA sequences instruct the phenotypic traits and disease state. On the other hand, interpreting the functional consequences of these variants is still very challenging. This is because most sequence variants found in GWASs reside in noncoding regions of the human genome, where the function remains to be defined. And frequently, the GWAS hits are just in linkage disequilibrium with the true causal variants that remain hidden (Botstein and Risch 2003; Cardon and Abecasis 2003; Carlson et al. 2004; Altshuler et al. 2008; Frazer et al. 2009). Annotation of the large number of enhancers in the human genome has now paved the way for a better understanding the function of these DNA variants. Recent studies using data from ENCODE and Roadmap Epigenome projects have shown that variants in regulatory elements from specific tissues are enriched for related physiological traits and diseases at genome-wide significant association values (ENCODE Project Consortium 2012; Maurano et al. 2012; Roadmap Epigenomics Consortium et al. 2015). This strongly suggests that many sequence variants contribute to phenotype by affecting enhancer functions and likely gene expression.

Additional evidence supporting this hypothesis comes from analysis of allelic gene expression in the human genome. Because the two copies of DNA that each individual inherits from its parents are not identical but contain millions of sequence polymorphisms, it is conceivable that sequence variants that disrupt enhancer function could lead to allelic biases of gene expression. Indeed, widespread allelic biases in transcription have been reported (Dixon et al. 2015; Leung et al. 2015; Waszak et al. 2015). Such biases were recently shown to strongly correlate with allelic biases in histone acetylation at enhancers, supporting the hypothesis that sequence variants may disrupt enhancer function, which in turn affects transcription of target genes (Dixon et al. 2015; Grubert et al. 2015; Leung et al. 2015). Further supporting this hypothesis, allelic biases in histone acetylation at enhancers show a strong and significant correlation with sequence variants that disrupt transcription factor binding to DNA (Grubert et al. 2015; Leung et al. 2015).

Statistical analyses of gene expression patterns in various human cell and tissue types have shown that on average 20%–30% of the variations in gene expression are heritable (Grundberg et al. 2012). Of these, approximately one-third are accounted for by variants acting in cis, whereas the other approximately two-thirds come from variants acting in trans (Price et al. 2011; Grundberg et al. 2012). The observation that factors such as gene–gene interactions, gene–environmental interactions, and possibly epigenetic mechanisms play a more prominent role in gene expression than direct genetic mechanisms presents substantial challenges in predicting gene expression from sequences. Delineating the exact contribution of each component to gene expression patterns in different cell types is undoubtedly going to require a better understanding of the molecular mechanisms of enhancer function (Albert and Kruglyak 2015; Pai et al. 2015).

CLOSING REMARKS

Understanding the gene regulatory programs encoded by the human genome is a central objective of biomedical research. Recent advances in the annotation and characterization of enhancers in the human genomes represent a significant step toward this goal. The detailed maps of chromatin state and accessibility across hundreds of cell and tissue types make it possible to study the cell type-specific activities of cis-regulatory elements throughout the genome and generate hypotheses regarding the mechanisms controlling dynamic gene expression programs in different cell types. The catalogs of candidate enhancers and other regulatory elements have become a valuable resource for researchers to decipher the effects of noncoding variants in gene expression and human diseases. The high-throughput methodologies for cis elements mapping and characterization have also provided a powerful approach for investigating the pathogenesis of cancer and other human diseases.

More research is undoubtedly needed to better define the roles of enhancers in gene expression and cellular phenotypes. Specifically, it is necessary to achieve a more quantitative understanding of each enhancer regarding how extracellular and intracellular signaling leads to dynamic transcription factor binding and chromatin remodeling at the element, and how combinatorial binding of transcription factors in turn affects the transcription of the target gene. To this end, an integrative, systems biology approach is necessary. Quantitative and predictive models of gene regulatory networks will need to be developed and tested. Model organisms such as the laboratory mouse, coupled with the CRISPR/cas9 genome engineering, is proving invaluable in this effort because of the possibilities of sorting out causes and effects through genetic manipulations. Additionally, single-cell analyses of transcriptome, DNA methylation, chromatin accessibility, and higher-order organization are emerging as powerful tools to study gene regulation at unprecedented resolutions and details (Tang et al. 2010; Guo et al. 2013; Nagano et al. 2013; Deng et al. 2014; Buenrostro et al. 2015; Cusanovich et al. 2015). Application of these tools will yield more breakthroughs in coming years.

ACKNOWLEDGMENTS

We thank all the current and previous Ren laboratory members for their contributions to both the method-
ologies and discoveries described in the text. Research in the Ren laboratory has been generously supported by the Ludwig Institute for Cancer Research, the National Institutes of Health (U54HG006997, U01ES017166, and R01HG003991), and the California Institute of Regenerative Medicine (RN2-00905).

REFERENCES
Albert FW, Kruglyak L. 2015. The role of regulatory variation in complex traits and disease. *Nat Rev Genet* 16: 197–212.
Altschuler D, Daly MJ, Lander ES. 2008. Genetic mapping in human disease. *Science* 322: 881–888.
Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmid C, Suzuki T, et al. 2014. An atlas of active enhancers across human cell types and tissues. *Nature* 507: 455–461.
Arnold CD, Gerlach D, Stelzer C, Boryn LM, Rath M, Stark A. 2013. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* 339: 1074–1077.
Attanasio C, Nord AS, Zhu Y, Blow MJ, Li Z, Liberton DK, Morrison H, Platzer-Frick I, Holt A, Hosseini R, et al. 2013. Fine tuning of craniofacial morphology by distant-acting enhancers. *Science* 342: 1241006.
Banerji J, Rusconi S, Schaffner W. 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. *Cell* 27: 299–308.
Banks A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. 2007. High-resolution profiling of histone modifications in the human genome. *Cell* 129: 823–837.
Botstein D, Risch N. 2003. Discovering genotypes underlying human phenotypes: Past successes for Mendelian disease, future approaches for complex disease. *Nat Genet* 33(Suppl): 228–237.
Boyer LA, Lee TI, Levine SS, Zucker JB, Lee AY, Ye Z, Kim A, Rajagopal N, Xie W, et al. 2015. Chromatin architecture reorganization during stem cell differentiation. *Nature* 518: 331–336.
Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzalez ML, Morais DE, Blackburn KE, Shulha HP, Meltzer P, Margulies EH, Weng RK, Grubert F, Zaugg JB, Kasowski M, Ursu O, Spacek DV, Martin SA, Rubio ED, Krumm A, Lamb J, Nusbaum C, et al. 2015. Chromatin Conformation Capture Carbon Copy (SCC): A massively parallel solution for mapping interactions between genomic elements. *Genome Res* 16: 1299–1309.
Buying DW, Diffley JF. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122: 947–956.
Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng RK, Grubert F, Zaugg JB, Kasowski M, Ursu O, Spacek DV, Martin SA, Rubio ED, Krumm A, Lamb J, Nusbaum C, et al. 2015. Chromatin Conformation Capture Carbon Copy (SCC): A massively parallel solution for mapping interactions between genomic elements. *Genome Res* 16: 1299–1309.
Doudna JA, Charpentier E. 2014. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* 346: 1258096.
Duan Z, Andronescu M, Schutz K, McInwain S, Kim YJ, Lee C, Shendure J, Fields S, Blau CA, Noble WS. 2010. A three-dimensional model of the yeast genome. *Nature* 465: 363–367.
ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489: 57–74.
Ernst J, Kellis M. 2012. ChromHMM: Automating chromatin-state discovery and characterization. *Nat Methods* 9: 215–216.
Ernst J, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, et al. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473: 43–49.
Fraterrone K, Murray SS, Schork NJ, Topol EJ. 2009. Human genetic variation and its contribution to complex traits. *Nat Rev Genet* 10: 241–251.
Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, Orlov YL, Velkov S, Ho A, Meit PH, et al. 2009. An oestrogen-receptor-α-bound human chromatin interactome. *Nature* 462: 58–64.
Greshin DI, Leung D, Ren B. 2014. The 3D genome in transcriptional regulation and pluripotency. *Cell Stem Cell* 14: 762–775.
Grosfond F, Antoniou M, Berry M, de Boer E, Dillion N, Ellis J, Fraser KA, Murray SS, Schork NJ, Topol EJ. 1993. Regulation of human globin gene switching. *Cold Spring Harb Symp Quant Biol* 58: 7–13.
Grubert F, Zaug JB, Kasowski M, Urho O, Spacek DV, Martin AR, Greenside P, Srivast P, Phanstiel DH, Pekowska A, et al. 2015. Genetic control of chromatin states in humans involves local and distal chromosomal interactions. *Cell* 162: 1051–1065.
Guo H, Zhu P, Wu X, Li X, Wen L, Tang F. 2013. Single-cell profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature* 523: 486–490.
Bulger M, Groudine M. 2011. Functional and mechanistic diversity of distal transcription enhancers. *Cell* 144: 327–339.
Cano E, Wysocka J. 2013. Modification of enhancer chromatin: What, how, and why? *Mol Cell* 49: 825–837.
Cardon LR, Abecasis GR. 2003. Using haplotype blocks to map human complex trait loci. *Trends Genet* 19: 135–140.
Carlson CS, Eberle MA, Kruglyak L, Nickerson DA. 2004. Mapping complex disease loci in whole-genome association studies. *Nature* 429: 446–452.
Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wang X, Orlov YL, Zhang W, Jiang J, et al. 2008. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 133: 1106–1117.
Collins FS, Varma H. 2015. A new initiative on precision medicine. *N Engl J Med* 372: 793–795.
Creyghton MP, Chen AW, Welstead GG, Kooistra T, Carey BW, Steine DJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, et al. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci* 107: 21931–21936.

Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39:** 310–318.

Hermel E, Saragosti S, Blangy D, Yaniv M. 1981. Fine structure of the origin-proximal DNAse I-hypersensitive region in wild-type and EC mutant polyoma. *Cell* **25:** 651–658.

Hesselberth JR, Chen X, Zhang Z, Sabo PJ, Sandstrom R, Reynolds AP, Thurman RE, Neph S, Kuehn MS, Noble WS, et al. 2009. Global mapping of protein–DNA interactions in vivo by digital genomic footprinting. *Nat Methods* **6:** 283–297.

Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, Hoke HA, Young RA. 2013. Super-enhancers in the control of cell identity and disease. *Cell* **155:** 934–947.

Hnisz D, Schuiers J, Lin CY, Weintraub AS, Abraham BJ, Lee TI, Bradner JE, Young RA. 2015. Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. *Mol Cell* **58:** 362–370.

Hoffman MM, Buske OJ, Wang J, Weng Z, Bilmes JA, Noble WS. 2012. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nat Methods* **9:** 473–476.

Hon G, Ren B, Wang W. 2008. ChromaSig: A probabilistic approach to finding common chromatin signatures in the human genome. *PLoS Comput Biol* 4: e1000201.

Hon GC, Sung CY, Du T, Jin F, Selvaraj S, Lee AY, Yan CA, Ye Z, Mao SQ, Wang BA, et al. 2014. 5mc oxidation by Tet2 modulates enhancer activity and timing of transcription reprogramming during differentiation. *Nat Cell Biol* **16:** 488–495.

Hsieh TH, Weiner A, Lajoie B, Dekker J, Friedman N, Rando OJ. 2015. Mapping nucleosome resolution chromosomal folding in yeast by micro-C. *Cell* **162:** 108–119.

Hughes JR, Roberts N, McGowan S, Hay D, Giannoulatou E, Lynch M, De Gobbi M, Taylor S, Gibbons R, Higgs DR. 2014. Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat Genet* **46:** 205–212.

Johnson DS, Mortazavi A, Myers RM, Wold B. 2007. Genome-wide mapping of in vivo protein–DNA interactions. *Science* **316:** 1497–1502.

Kalhor R, Tjong H, Jayathilaka N, Alfer C, Chen L. 2012. Genome architectures revealed by tethered chromosome conformations and population-based modeling. *Nat Biotechnol* **30:** 90–98.

Kheradpour P, Ernst J, Melnikov A, Rogov P, Wang L, Zhang X, Alston J, Mikkelsen TS, Kellis M. 2013. Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res* 23: 800–811.

Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B. 2005. A high-resolution map of active promoters in the human genome. *Nature* **436:** 876–880.

Kim TH, Abdullaev ZK, Smith RP, May AE1004857.

Kheradpour P, Ernst J, Melnikov A, Rogov P, Wang L, Zhang X, Alston J, Mikkelsen TS, Kellis M. 2013. Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res* 23: 800–811.

Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B. 2005. A high-resolution map of active promoters in the human genome. *Nature* **436:** 876–880.

Kim TH, Abdullaev ZK, Smith RP, May AE1004857.

Kheradpour P, Ernst J, Melnikov A, Rogov P, Wang L, Zhang X, Alston J, Mikkelsen TS, Kellis M. 2013. Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res* 23: 800–811.

Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B. 2005. A high-resolution map of active promoters in the human genome. *Nature* **436:** 876–880.

Kim TH, Abdullaev ZK, Smith RP, May AE1004857.

Kheradpour P, Ernst J, Melnikov A, Rogov P, Wang L, Zhang X, Alston J, Mikkelsen TS, Kellis M. 2013. Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res* 23: 800–811.

Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B. 2005. A high-resolution map of active promoters in the human genome. *Nature* **436:** 876–880.

Kim TH, Abdullaev ZK, Smith RP, May AE1004857.

Kheradpour P, Ernst J, Melnikov A, Rogov P, Wang L, Zhang X, Alston J, Mikkelsen TS, Kellis M. 2013. Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res* 23: 800–811.

Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B. 2005. A high-resolution map of active promoters in the human genome. *Nature* **436:** 876–880.

Kim TH, Abdullaev ZK, Smith RP, May AE1004857.
Price AL, Helgason A, Thorleifsson G, McCarroll SA, Kong A, Stefansson K. 2011. Single-tissue and cross-tissue heritability of gene expression via identity-by-descent in related or unrelated individuals. *PLoS Genet* 7: e1001317.

Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470: 279–283.

Rajagopal N, Xie W, Li Y, Wagner U, Wang W, Stamatoyannopoulos J, Ernst J, Kellis M, Ren B. 2013. RFECs: A random-forest based algorithm for enhancer identification from chromatin state. *PLoS Comput Biol* 9: e1002968.

Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, et al. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159: 1665–1680.

Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* 518: 317–330.

Sabo PJ, Kuehn MS, Thurman R, Johnson BE, Johnson EM, Cao H, Yu M, Rosenzweig E, Goldly J, Haydock A, et al. 2006. Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. *Nat Methods* 3: 511–518.

Sahlén P, Abdullahiev I, Ramsköld D, Matskova L, Rilakovic N, Löštiedt B, Albert TJ, Lundeberg J, Sandberg R. 2015. Genome-wide map of promoter-anchored interactions with close to single-enhancer resolution. *Genome Biol* 16: 73.

Schultz MD, He Y, Whitaker JW, Harirahan M, Mukamel EA, Leung D, Rajagopal N, Nery JR, Urich MA, Chen H, et al. 2015. Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* 523: 212–216.

Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parinello H, Tanay A, Cavalli G. 2012. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* 148: 458–472.

Shalem O, Sanjana NE, Zhang F. 2015. High-throughput functional genomics using CRISPR-Cas9. *Nat Rev Genet* 16: 299–311.

Shen Y, Yue F, McClearay DF, Ye Z, Edsall L, Kuan S, Wagner U, Dixon J, Lee L, Lobanenkov VV, et al. 2012. A map of the cis-regulatory sequences in the mouse genome. *Nature* 488: 116–120.

Shen L, Wu H, Diep D, Yamaguchi S, D’Alessio AC, Fung HL, Zhang K, Zhang Y. 2013. Genome-wide analysis identifies coordinated and complex chromosomal rearrangements by 4C technology. *Nat Methods* 6: 837–842.

Smith E, Shilatifard A. 2014. Enhancer biology and enhanceropathies. *Nat Struct Mol Biol* 21: 210–219.

Smith RP, Taher L, Patwardhan RP, Kim MJ, Inoue F, Shendure J, Ovcharenko I, Ahtinen T. 2013. Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. *Nat Genet* 45: 1021–1028.

Song CX, Szulwach KE, Dai Q, Fu Y, Mao SQ, Lin L, Street C, Li Y, Poidevin M, Wu H, et al. 2013. Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* 153: 678–691.

Stadler MB, Murt R, Burger L, Ivanek R, Lienert F, Schöler A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D, et al. 2011. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* 480: 490–495.

Tang F, Barbacioru C, Bao S, Lee C, Nordman E, Wang X, Lao K, Surani MA. 2010. Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. *Cell Stem Cell* 6: 468–478.

Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, et al. 2012. The accessible chromatin landscape of the human genome. *Nature* 489: 75–82.

Tuan D, London IM. 1984. Mapping of DNase I-hypersensitive sites in the upstream DNA of human embryonic e-globin gene in K562 leukemia cells. *Proc Natl Acad Sci* 81: 2718–2722.

Vierstra J, Rynes E, Sandstrom R, Zhang M, Canfield T, Hansen RS, Stehling-Sun S, Sabo PJ, Byron R, Humbert R, et al. 2014. Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. *Science* 346: 1007–1012.

Vieira M, Jackson RD, D’Alessio AC, Lidstone JJ, Angers S, Duhart S, et al. 2014. Population variation and genetic control of modular chromatin architecture in humans. *Cell* 162: 1039–1050.

Weintraub H, Groudine M. 1976. Chromosomal subunits in active genes have an altered conformation. *Science* 193: 848–856.

Whitman MA, Myers CA, Corbo JC, Cohen BA. 2013. Massively parallel in vivo enhancer assay reveals that highly local features determine the cis-regulatory function of ChiP-seq peaks. *Proc Natl Acad Sci* 110: 11952–11957.

Wythe WA, Orlando DA, Hinisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA. 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153: 307–319.

Yu M, Hon GC, Szulwach KE, Song CX, Zhang L, Kim A, Li X, Dai Q, Shen Y, Park B, et al. 2012. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* 149: 1368–1380.

Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, Sandstrom R, Ma Z, Davis C, Pope BD, et al. 2014. A comparative encyclopedia of DNA elements in the mouse genome. *Nature* 515: 355–364.

Zhou HY, Katsman Y, Dhalwil NV, Davidson SM, Macpherson NN, Sakhidzevi M, Collura F, Mitchell JA. 2014. A Sox2 distal enhancer cluster regulates embryonic stem cell differentiation potential. *Genes Dev* 28: 2699–2711.

Ziller MJ, Giu H, Müller F, Donaghey J, Tsai LT, Kohlbacher O, De Jager PL, Rosen ED, Bennett DA, Bernstein BE, et al. 2013. Charting a dynamic DNA methylation landscape of the human genome. *Nature* 500: 477–481.

Zuin J, Dixon JR, van der Reijden MJJA, Ye Z, Kolovos P, Brouwer RWW, van de Corput MPC, van de Werken HJG, Knoch TA, van IJcken WJF, et al. 2014. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc Natl Acad Sci* 111: 996–1001.