Published in final edited form as:

Semin Cell Dev Biol. 2009 September ; 20(7): 842–848. doi:10.1016/j.semcdb.2009.06.005.

Toward a Genome-Wide Reconstruction of Cis-Regulatory Networks in the Human Genome

Katharine R. Cecchini¹, A. Raja Banerjee¹, and Tae Hoon Kim^{1,2}

¹Department of Genetics, Yale University, School of Medicine, New Haven, CT 06510

Abstract

The vast amount of recent progress made on the sequence of the human genome has allowed an unprecedented examination of cis-regulatory networks. These networks consist of functional elements such as promoters, enhancers, silencers, and insulators, and their coordinated activity is responsible for regulation of gene expression. Recent studies surveyed the entire genome, identifying novel elements and evaluating functional differences in respect to development. These investigations present the first steps towards a global regulatory map for expression in the human genome.

Keywords

chromatin structure; cis-regulatory network; enhancers; silencers; promoters; insulators functional genomics; ChIP

1.1: Approaches

Upon the completion of the human genome sequence, one of the major frontiers to be investigated was the miles of non-coding DNA. Most of these sequences are believed to play some role in regulation of gene expression. The availability of this completed sequence has led to the development of a number of comprehensive approaches to mapping regulatory sequences related either to specific genes or proteins. In this review we will discuss the impact of these techniques on the development of genome-wide regulatory networks as a landscape of sequence-level mapping that organizes into distinct conformations and nuclear locations.

The distribution of a specific transcription factor across a representative portion of the genome has conventionally been examined through utilization of a microarray. A transcription factor can be isolated with bound DNA sequences by cross-linking cells and purifying the target with an antibody; the sequences can then be recovered and analyzed by hybridization to a microarray (ChIP-chip) [1]. Most notably, the ENCODE project had previously designed an array platform composed of a representative 1% of the human genome sequence which subsequently became a valuable tool for genomic investigations [2]. A low through-put cousin to ChIP-chip, DIP-chip, can also be implemented to investigate *in vitro* bound sequences if the purified protein of interest is available [3]. Similarly, a DNA adenine methyltransferase (Dam) fusion of the protein of interest can be generated; this technique allows mapping of site-specific DNA-protein interactions (DamID) [4]. Analysis of the output of these techniques has shifted of late, however, in that recovered sequences can also be analyzed by parallel high-throughput sequencing, (ChIP-Seq) such as that developed by Illumina/Solexa [5,⁶]. This approach lends robustness to the aforementioned techniques as well as conventional sequencing-based

²Corresponding Author Address to: Tae Hoon Kim, Department of Genetics, Yale University, School of Medicine, 333 Cedar St., P.O. Box 208005, New Haven, CT 06520-8005, Voice: 203-785-7501, Fax: 206-984-2338, taehoon.kim@yale.edu.

methods such as ChIP-PET, a modified SAGE (Serial Analysis of Gene Expression) protocol where immunoprecipitated DNA are cloned to allow sequencing of both 3' and 5' ends [7]. Development of high-throughput sequencing allows any of these techniques to yield specific genomic sequence that can then be mapped with a higher degree of confidence than that afforded by microarray analysis. Depending on the nature of the DNA-binding protein, these results can also convey information about the physical organization of the genome, as will be described later.

Once a global survey has been undertaken using one of the methods listed above, the interaction of the protein of interest and its target DNA can be further resolved using biochemical approaches. Electrophoretic Mobility Shift Assays (EMSA), Surface Plasmon Resonance (SPR), Proximity Ligation assays, and Binding Site Selection assays all provide confirmation of microarray or sequence data and adjustment of the consensus sequence derived from this global survey. In coordination these assays detail a reliable map for specific transcription factors and/or DNA-binding proteins that, with expression data, can be incorporated into the regulatory landscape. These techniques have produced a wealth of information about the functional organization of the genome on both a genetic and structural level, and have effectively revised models of regulatory elements.

2.1: Epigenetic Structures

In the broadest sense the genome can be divided into two categories: non-expressing, compacted heterochromatin, and expressing, non-compacted euchromatin. By their nature these categories suggest a relationship between physical availability and expression level of their resident genes. Like the states of sleep and wakefulness, the difference between heterochromatin and euchromatin domains is easily determined visibly but the underlying biochemical cause is layered and complex. The epigenetic environment is modifiable in several ways that can be coordinated with functional regulatory elements. Histones can be proportionally present or absent, and amino acids on those histones can be modified, most notably by acetylation, phosphorylation, or methylation [8]. These modifications, if their meaning can be correctly interpreted through observation of genome-wide patterns, succeed in breaking all chromatin (and intrinsic functional elements) into a general dichotomy of either active or silent. Several investigators have implemented microarrays or high-throughput sequencing in coordination with targeted analysis of specific histone marks or DNA-binding proteins to both enhance and adjust the classical definition of the functional elements of a transcriptional unit.

3.0: Genomic Organization

3.1: Promoters

With the implementation of this whole-genome approach, the traditional concept of regulatory elements has been challenged, none more so than promoters. Bioinformatic surveys have indicated that fewer than 25% of human promoters contain the canonical TATA element or a similar motif [9]. Accurate transcription start site (TSS) prediction is further confounded by the majority of human coding sequences possessing multiple TSSs within a relatively small 200 base pair region 5' of the gene [10]. Therefore, sequence-based elements will likely only provide a framework for identifying promoters, while epigenetic characteristics would better serve to characterize these elements, as well as differentiating between active and inactive promoters.

Promoters, unlike other regulatory elements, are by necessity proximal to their TSSs. TSSs were first mapped using sequence-aligned cDNAs and promoter functionality was tested by transfection of a reporter construct [11]. Bioinformatically, they were also determined to be

generally but not exclusively associated with CpG islands [12,¹³]; methylation of these CpGs is thought to result in repression. Using TSSs as a landmark they can be further broken down into active and silent subsets based primarily on detection of transcriptional activity or features of the surrounding chromatin.

Obviously, the principal feature of active promoters will be presence of a transcript. Genome-wide transcriptional activity has been surveyed in the past by either ChIP-chip of RNA Pol II [14], or examination of active transcripts by sequencing or microarray expression analysis [15,¹⁶]. In a recent study [17], high-throughput sequencing of nuclear run-on transcripts in human fibroblasts indicated that 68% of previously identified transcripts could be designated as active based on RNA Pol II engagement. However, further investigation revealed that only 42% of these genes (21% of all genes) continued to elongate downstream from the TSS. The authors interpreted this transcript distribution as classifying promoters into one of three categories relative to RNA Pol II: active and not paused, active and paused, and inactive. They maintain that all promoters where RNA Pol II is bound are in fact active promoters. Technological advances will likely increase sensitivity and provide the as-yet undetectable transcript. Their results already demonstrate this expansion of definition; presence of RNA Pol II is not sufficient to indicate productive transcription. Genome-wide transcript sequencing can shed light on the mechanisms of transcription initiation and regulation, particularly for non-coding RNAs, but this technique offers little insight into identifying novel promoters.

Like other types of engaged functional elements, promoters that are active are likely to be found in nucleosome-free regions [18,¹⁹] or areas of DNase sensitivity [20]. Accordingly, active promoters can be further discriminated by surveying the genome for the components of the Pre-Initiation Complex (PIC). RNA Pol II and TAF1 binding sites had been identified by ChIP-chip in the ENCODE regions in a number of different cell types [14] and active promoters of a human fibroblast line were mapped in approximate entirety [21]. Heintzman et al. [18] used the same approach to directly investigate promoters that were induced by treatment of HeLa cells with IFN- γ . They found, in agreement with other studies [6,¹⁶,²²], that active promoters can be distinguished from active enhancers by their specific trimethylated lysine 4 marks on histone 3 (H3K4me3); furthermore, the intensity of this mark is proportional to the intensity of expression of the particular transcript [23]. Previous studies [23–²⁶] had also implicated methylation marks on this residue as a tag for promoters, but had not differentiated trimethylation patterns from di- (H3K4me2) or mono-methylation (H3K4me). With high resolution mapping, Heintzman et al. were able to demarcate differential methylation states in relation to the TSS. H3K4me3 marks occur prominently at the ground zero of the TSS (also in [23]), while the H3K4me2 and H3K4me marks are depleted at the TSS but increase at the immediate periphery and radiate outward in both directions (see Figure 1). This distribution of signal potentially allows discrimination between promoters and enhancers, which also bear the H3K4me and H3K4me2 marks. They speculate that by favoring interpretation of the distribution of these marks as opposed to their mere presence or absence, regulatory features and their activity can be identified on the basis of these chromatin characteristics. Identification of promoters would allow identification of novel genes, a more detailed investigation of transcription machinery, and further exploration of gene regulation on a global or specific scale.

Human ES cell experiments have offered insight into characteristics of cell specificity for promoters, again based on chromatin modifications. ES cells offer a useful model as a blank state cell that has not undergone any differentiation. Guenther et al. [23] associated H3K4me3 marks with acetylation of lysine residues on H3 (H3K9Ac and H3K14Ac), which are redundantly associated with activated promoters [24,²⁵]. ~75% of ES cell promoters bore these marks (in addition to RNA Pol II) indicating activity, and this percentage carried over into differentiated cell types such as B cells and hepatocytes. This correlates with their observation that H3K36me3 [27], a histone mark associated with elongation of RNA Pol II, is present at

only a minority of promoters designated as “active” based on the conjunction of Pol II and H3K4me3 and acetylation marks. This result parallels the findings in normal cells (see above). The remaining 25% of H3K4me3 positive promoters was designated as cell-type specific. Mikkelsen et al. [6] reported similar findings in mouse ES cells as well, with a preponderance of expressing promoters (~80%) bearing H3K4me3 marks, and the remainder bearing both trimethylated K4 and K27. Through parallel analysis of differentiated mouse cells, these bivalent promoters were found to resolve to active or repressed depending on the cell type. This effect was especially noticeable in CpG-poor promoters. The overlap of these marks indicates, in ES cells, that these promoters are poised for either activity or repression depending on the differentiated lineage. Through global survey, these studies outline a method for categorizing promoters as active, repressed, or developmentally determined, allowing portraits of expression to be painted that would represent different developmental stages as the cell’s potentiality narrows.

3.2: Enhancers

Discriminating between promoters and enhancers has been difficult due to some overlap between the distinguishing features (nucleosome-free regions, DHSs, mono-, di- or tri-methylated H3K4) [28,²⁹], and the variable distance between enhancers and the promoter they regulate. Several attempts have been made to map generalized enhancers by sequence motif, but these have met with variable success[30,³¹]. Some progress has been made by limited ChIP-chip mapping accompanied by functional verification of known transcription factors such as NF- κ B [31], estrogen receptor α (ER) [32], tumor suppressor p53 [7], its homologue p63 [33], and cAMP response element binding protein (CREB) [34]. Recently several global studies have been performed using ChIP-chip to map putative enhancers in the ENCODE regions of the genome without prior knowledge of the transcription factors involved. The histone acetyltransferase p300/CBP was first identified as one of several factors recruited to the enhancer region upon gene activation [28,³⁵]. Heintzman et al. [18] focused on p300 first in HeLa cells where activation of promoters was induced by treatment with IFN- γ (see Section 3.1). They found that the p300 binding sites resembled the general features ascribed to enhancer elements. Concurrently, they were able to discriminate between enhancers and promoters due to a lack of H3K4me3 at p300 sites, but a strong presence of mono- and di-methylated H3K4 (see fig.1). Based on this global survey, the chromatin pattern, either in conjunction with sequence alignment or by itself, appears to be more useful and accurate at predicting potential enhancer elements.

It has long been thought that cell type-specific programming is a result of differential engagement of regulatory elements such as enhancers and silencers. To this effect, Visel et al. [36] mapped p300 binding sites in mouse embryo forebrain, midbrain, and limb using ChIP-Seq and evaluated their tissue-specific activity by reporter assay. They found the vast majority of identified enhancer sites were tissue-specific. This study demonstrates the feasibility of mapping regulatory controls on a genome level for particular tissues throughout multiple stages of development. While H3K4me1 regions were thought to be cell-type specific, Heintzman et al. [37] resolved this pattern further by performing ChIP-chip on chromatin marks in a different set of five cell lines; in their studies enhancers across the entire human genome were more easily defined by chromatin modifications than their location relative to TSSs. In concordance with the previously-cited studies, they also observed that most active enhancers were specific to one cell type. The enhancer maps generated in these studies could also be instrumental in identifying aberrant changes in programming in cancer cells as the disease state progresses. Furthermore, computational and bioinformatic analyses of the comprehensive enhancer maps will allow us to reconstruct cis-regulatory network across the entire genome, to classify and refine various cis-regulatory modules, and to investigate the developmental and evolutionary landscape of these control elements.

3.3: Silencers

Mass silencing of genes is likely to be based as much on positional effects as on a sequence-specific element. Large segments of the genome are compacted into transcriptionally silent heterochromatin, and the relative position of the factors responsible for demarcating and propagating chromatin boundaries will have a large impact on the expression status of a coding sequence or locus. On a global level, repressed genes have been associated with nucleosomes bearing silencing mark H3K27me3 using SAGE and ChIP-chip, and similar to enhancer mapping, these genes are usually cell-specific [38]. This trend was also observed in mouse ES cells (see Section 3.1)[6].

The H3K27me3 nucleosomal mark is usually utilized to genomically screen for areas of silencing rather than silencer elements. The binding sites of multiple members of the polycomb group proteins (PcG), the complex responsible for this nucleosome mark, have been mapped in mouse ES cells, and as mentioned previously, target genes were found to be both cell-type specific and repressed [39]. Interestingly, these repressed genes were primarily developmental factors as well, indicating that global patterns of PcG-mediated repression, when coupled with global patterns of enhancer activation, could determine the programming mechanism and destiny of a differentiating cell. The signature of tissue-specific PcG binding may also be useful in predicting disease states, as has been seen for prostate cancer progression [40].

The neural silencer REST has also been extensively studied because of its association with regulation of an exclusively neural phenotype [5,⁴¹,⁴²]. As might be expected, REST is also associated with a global increase in H3K27me3 marks, and frequently recruits the histone deacetylase Sin3 [43]. Histone deacetylases are likely mediators of repression states, so mapping of their efficacy sites in conjunction with other transcription factors would expand on library of genomic information available and allow dissection of the interplay of various factors and patterns that results in an expression profile and complementary phenotype. In conjunction with chromatin marks, this pattern might allow global mapping of silencing elements much in the same way as enhancers.

3.4: Insulators

The establishment of the dichotomy of repressive heterochromatin and active euchromatin domains is achieved through binding of insulator proteins to their target sites. Insulators also prevent undesirable interactions between other regulatory elements, such as inappropriate activation by enhancers. The only insulator identified in vertebrates is the CCCTC-binding protein (CTCF) [44]. CTCF binding sites have been mapped across the whole genome using both ChIP-chip in fibroblasts [45] and in several other cell types [37,⁴⁶]. All studies found that CTCF-binding sites were mostly found at intergenic sites, correlated with gene-dense regions, and binding sites were largely conserved (~50%) between cell types. Mapping of these sites produced a universal CTCF recognition sequence; occupancy of CTCF sites is known to be dependent on their DNA methylation status [47], which accounts for the difference between observed and predicted sites. This methylation difference also can account for differences in CTCF occupation between various cell types. Cuddapah et al. [46] also used ChIP-Seq to place CTCF binding sites as acting boundaries between H3K27me3 repressive domains and acetylation of lysine 5 on histone 2A (H2AK5Ac) domains, representing domains of active gene expression. The role of those binding sites as barriers was observed to also depend on cell type. These studies established the insulator's role in partitioning the genome into distinct chromatin domains and insulating of local cis-regulatory circuits. A detailed understanding of insulator function and regulation will provide critical insights into the properties and behaviors of cis-regulatory networks.

3.5: Locus Control Regions

Locus control regions (LCRs) provide long-range control over the expression of suites of cell-type specific genes. They confer high levels of activity upon linked genes that is cell-type specific [48] and independent of the character of the surrounding chromatin [49]. LCRs are composite elements that can include enhancer [50], insulator [51], and chromatin opening [52] activities, and demonstrate intricate chromatin looping patterns (see section 4.2)[53,⁵⁴].

Within the genome, LCRs play critical roles in the orchestrated expression of their locus of genes. The genes in the β -globin locus, for example, are expressed within specific developmental stages in the erythroid lineage. The β -globin LCR appears to affect the levels of transcription of the expressed globin genes in a developmental-stage-specific fashion, operating as a master control over other intra-locus enhancer elements [55,⁵⁶]. It has been posited that the LCR can act as a recruitment site for a chromatin remodeling complex which is then transferred to the genes themselves, presumably via chromosomal looping [57]. These findings help affirm the frequent postulations that LCRs act within tissue-specific regulatory modules to effect broad changes to the chromatin state, facilitating high levels of expression of genes within its target loci in a tissue-specific fashion.

Are LCRs a frequently-used regulatory tool within the genome? Their principles of action could certainly be adopted in a variety of contexts, given the significant tissue-specific-differences in expression profiles. However, besides those identified, the extent of LCR presence within the human genome remains to be elucidated. Some general features are known: LCRs seem to include clusters DNase-hypersensitivity sites [58] and can stimulate formations of open chromatin [52]. Chromosomal looping could potentially identify more LCRs. As more factors are genomically mapped in a variety of cell types and conditions, it is possible that, like enhancers, putative LCRs could be identified on the basis of a particular chromatin signature or confluence of some other epigenetic characteristics.

4.0: Physical Organization

4.1: Subnuclear Location

As mentioned in section 2.1, the genome is separated into two classes of packaged material: generally active heterochromatin and generally silent euchromatin. Residing within one of these domains not only determines the activity of a particular element, but also the location within the nucleus. Heterochromatin is often localized to the periphery of the nucleus, while euchromatin forms a complex with the transcriptional machinery in the interior [59,⁶⁰]; this state implicates the structural components responsible for this segregation as potential regulatory proteins. Indeed, DamID analysis of the nuclear membrane protein lamin B1 and its associate emerlin indicates the majority of this chromatin is gene-poor, repeat-rich, and enriched in silencing mark H3K27me3 [61]. Additionally, they found the genes in these lamina-associated domains showed decreased transcriptional activity by microarray analysis and RNA Pol II recruitment. This effect has been observed in other cases as well; genes proximal to the nuclear lamina are often, but not exclusively, transcriptionally repressed [62–⁶⁴]. Guelen et al. were able to localize CTCF to the borders between these peripheral domains, as appropriate to predicted insulator sites. Of course, exceptions to this model have been noted: in mouse rod cells this pattern is reversed, purportedly to aid in channeling the incoming light, a specialization that overrides nuclear organization [65]. In flies and yeast, association of genes with the nuclear pore is indicative of fine control over expression, possibly due to insulator activity, although this effect has not been reproduced in vertebrates [66]. Nuclear compartmentalization as it stands presents another level of regulatory control for the genome that awaits further interrogation from the functional genomics perspective.

4.2: Inter-element Looping

Regulatory elements such as enhancers or silencers are known to be able to influence their target genes over long genomic distances (>10kb), but the mechanism of action over such a distance is unclear. A potential method would have to bridge sequence specificity and structural organization without a dramatic protein or energy expenditure. As mentioned in section 3.5, both the mouse and human β -globin loci have been utilized as models for probing these long-range interactions, focusing on the composite LCR and a DHS 3' to the locus, as well as other more distal elements [67]. Tolhuis et al. [68] were able to physically associate the β -globin LCR with the promoter using formaldehyde cross-linking, allowing the LCR to aid in RNA Pol II recruitment and direction. Chromosome conformation capture (3C) is a commonly-used technique, first developed with yeast chromosomes, to investigate spatial DNA interactions such as that of the β -globin LCR and its promoter [69]. 3C assays at the β -globin locus have since implicated a number of factors in maintenance of the chromosomal loops and differential gene expression [70–72], including the insulator protein CTCF [73].

CTCF and/or its partner cohesin made likely candidates for mediators of inter-element looping, since as insulators they are responsible for maintaining architectural domains within chromatin. CTCF has been implicated as one of several factors in the establishment of transcription-modifying interchromosomal interactions, but its necessity has varied from case to case [74–76]. Formation of chromosomal loops also provides a physical explanation for the enhancer-blocking and barrier activities of CTCF. Its partner cohesin is a natural candidate for this role as well, since it is known to be involved in chromatid cohesion and segregation during mitosis [77]. In conjunction with other transcription factors, this looping model may explain not only the long range interactions observed at many loci, but also the disparate stimulatory or inhibitory effects on gene regulation attributed to CTCF. Loop formation accounts for the polar effects of other transcription factors as well, as the associated DNA is sequestered into local regions of heterochromatin or euchromatin based on the formation of this secondary structure.

More elaborate techniques based on the rudimentary 3C protocol have been developed to map interactions across a larger window of the locus of interest. The 5C protocol (3C-carbon copy) uses a ligation-mediated amplification step to generate fragments that can be analyzed by either high-throughput sequencing or microarray analysis. 4C (circular 3C) ligates the nearby fragments together to form a circle and allow capturing and sequencing of the interacting fragments. Both 4C and 5C allow extensive mapping of interactions across different chromosomes as well as across a great genomic distance on the same chromosome. However, there are two pitfalls of these genome-surveying techniques: 1) “background” from physically close but functionally unrelated gene fragments will prevail in either sequencing or microarray analysis, and 2) a predetermined fragment of interest (a “bait”) must be utilized as a reference for analysis [78]. Thus far these techniques cannot accommodate global mapping of inter-locus or interchromosomal interactions, but can enhance our understanding of the scaffold of physical interactions dictating regulation of transcription, one candidate at a time. We expect genome-wide interrogation of chromosomal interactions and looping will be greatly advanced with the coming increases in capacity and throughput of the next generation sequencing technologies.

5.0: Future Directions

We present here a survey of surveys; looking at the number and method of many of the genome-wide studies investigating transcriptional regulation that have been performed to date in mammals. A pioneer of studies of this type, the initial ENCODE pilot project devoted unprecedented attention to 1% of the genome but the other 99% remains. Besides producing a wealth of information, ENCODE has changed our approach to understanding gene control and promises to offer novel insights into the landscape of the genome as a whole.

The advent of high-throughput sequencing has enabled genome-wide studies to shift directions towards a more extensive and robust method of survey and analysis. As the technique is brought into more common use, cost will no doubt decrease as competition and throughput increase. Sequencing as an analysis method can be adapted to numerous protocols, as have been mentioned here; of course, identifying target DNA using ChIP or DamID, measuring gene expression using SAGE, and detecting inter-element interactions using 4C. Using a target protein, these techniques allow us to dive headfirst into clarifying the sequence, structure, and function of the genome. Furthermore, from the genome-wide survey of binding sites of transcription factors, consensus sequences for those binding sites are available. By investigating the local sequence at a gene or locus of interest, the positions of promoters, transcription factors, or insulators can be located, predicting their effect on the gene of interest and across the genome. In this fashion a model for cis-regulatory networks governing expression of the entire genome can be developed, creating a testable framework to further investigate molecular mechanisms of genome control.

Acknowledgments

We thank YoonJung Kim for her help preparing this review. We apologize to the numerous researchers whose work couldn't be mentioned in this short perspective due to space limitations. Work in the laboratory is supported by Rita Allen Foundation, Sidney Kimmel Foundation for Cancer Research, Yale Comprehensive Cancer Center, and Yale University School of Medicine.

References

1. Wacker DA, Kim TH. From sextant to GPS: twenty-five years of mapping the genome with ChIP. *J Cell Biochem* 2009;107:6–10. [PubMed: 19180576]
2. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* 2004;306:636–640. [PubMed: 15499007]
3. Liu X, Noll DM, Lieb JD, Clarke ND. DIP-chip: rapid and accurate determination of DNA-binding specificity. *Genome Res* 2005;15:421–427. [PubMed: 15710749]
4. Vogel MJ, Peric-Hupkes D, van Steensel B. Detection of in vivo protein-DNA interactions using DamID in mammalian cells. *Nat Protoc* 2007;2:1467–1478. [PubMed: 17545983]
5. Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* 2007;316:1497–1502. [PubMed: 17540862]
6. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 2007;448:553–560. [PubMed: 17603471]
7. Wei CL, Wu Q, Vega VB, Chiu KP, Ng P, Zhang T, et al. A global map of p53 transcription-factor binding sites in the human genome. *Cell* 2006;124:207–219. [PubMed: 16413492]
8. Strahl BD, Allis CD. The language of covalent histone modifications. *Nature* 2000;403:41–45. [PubMed: 10638745]
9. Yang C, Bolotin E, Jiang T, Sladek FM, Martinez E. Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene* 2007;389:52–65. [PubMed: 17123746]
10. Juven-Gershon T, Hsu JY, Theisen JW, Kadonaga JT. The RNA polymerase II core promoter - the gateway to transcription. *Curr Opin Cell Biol* 2008;20:253–259. [PubMed: 18436437]
11. Trinklein ND, Aldred SJ, Saldanha AJ, Myers RM. Identification and functional analysis of human transcriptional promoters. *Genome Res* 2003;13:308–312. [PubMed: 12566409]
12. Ioshikhes IP, Zhang MQ. Large-scale human promoter mapping using CpG islands. *Nat Genet* 2000;26:61–63. [PubMed: 10973249]
13. Davuluri RV, Grosse I, Zhang MQ. Computational identification of promoters and first exons in the human genome. *Nat Genet* 2001;29:412–417. [PubMed: 11726928]

14. Kim TH, Barrera LO, Qu C, Van Calcar S, Trinklein ND, Cooper SJ, et al. Direct isolation and identification of promoters in the human genome. *Genome Res* 2005;15:830–839. [PubMed: 15899964]
15. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 2008;5:621–628. [PubMed: 18516045]
16. Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 2007;447:799–816. [PubMed: 17571346]
17. Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 2008;322:1845–1848. [PubMed: 19056941]
18. 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;39:311–318. [PubMed: 17277777]
19. Felsenfeld G, Groudine M. Controlling the double helix. *Nature* 2003;421:448–453. [PubMed: 12540921]
20. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, et al. High-resolution mapping and characterization of open chromatin across the genome. *Cell* 2008;132:311–322. [PubMed: 18243105]
21. Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, et al. A high-resolution map of active promoters in the human genome. *Nature* 2005;436:876–880. [PubMed: 15988478]
22. Schneider R, Bannister AJ, Myers FA, Thorne AW, Crane-Robinson C, Kouzarides T. Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nat Cell Biol* 2004;6:73–77. [PubMed: 14661024]
23. Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* 2007;130:77–88. [PubMed: 17632057]
24. Liang G, Lin JC, Wei V, Yoo C, Cheng JC, Nguyen CT, et al. Distinct localization of histone H3 acetylation and H3-K4 methylation to the transcription start sites in the human genome. *Proc Natl Acad Sci U S A* 2004;101:7357–7362. [PubMed: 15123803]
25. Schubeler D, MacAlpine DM, Scalzo D, Wirbelauer C, Kooperberg C, van Leeuwen F, et al. The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes Dev* 2004;18:1263–1271. [PubMed: 15175259]
26. Rougeulle C, Navarro P, Avner P. Promoter-restricted H3 Lys 4 di-methylation is an epigenetic mark for monoallelic expression. *Hum Mol Genet* 2003;12:3343–3348. [PubMed: 14570711]
27. Bannister AJ, Schneider R, Myers FA, Thorne AW, Crane-Robinson C, Kouzarides T. Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *J Biol Chem* 2005;280:17732–17736. [PubMed: 15760899]
28. Hatzis P, Talianidis I. Dynamics of enhancer-promoter communication during differentiation-induced gene activation. *Mol Cell* 2002;10:1467–1477. [PubMed: 12504020]
29. Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, et al. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* 2005;120:169–181. [PubMed: 15680324]
30. Prabhakar S, Poulin F, Shoukry M, Afzal V, Rubin EM, Couronne O, et al. Close sequence comparisons are sufficient to identify human cis-regulatory elements. *Genome Res* 2006;16:855–863. [PubMed: 16769978]
31. Pennacchio LA, Loots GG, Nobrega MA, Ovcharenko I. Predicting tissue-specific enhancers in the human genome. *Genome Res* 2007;17:201–211. [PubMed: 17210927]
32. Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, et al. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* 2005;122:33–43. [PubMed: 16009131]
33. Yang A, Zhu Z, Kapranov P, McKeon F, Church GM, Gingeras TR, et al. Relationships between p63 binding, DNA sequence, transcription activity, and biological function in human cells. *Mol Cell* 2006;24:593–602. [PubMed: 17188034]

34. Zhang X, Odom DT, Koo SH, Conkright MD, Canettieri G, Best J, et al. Genome-wide analysis of cAMP-response element binding protein occupancy, phosphorylation, and target gene activation in human tissues. *Proc Natl Acad Sci U S A* 2005;102:4459–4464. [PubMed: 15753290]
35. Merika M, Williams AJ, Chen G, Collins T, Thanos D. Recruitment of CBP/p300 by the IFN beta enhanceosome is required for synergistic activation of transcription. *Mol Cell* 1998;1:277–287. [PubMed: 9659924]
36. Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 2009;457:854–858. [PubMed: 19212405]
37. 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;459:108–112. [PubMed: 19295514]
38. Roh TY, Cuddapah S, Cui K, Zhao K. The genomic landscape of histone modifications in human T cells. *Proc Natl Acad Sci U S A* 2006;103:15782–15787. [PubMed: 17043231]
39. Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, et al. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 2006;441:349–353. [PubMed: 16625203]
40. Yu J, Rhodes DR, Tomlins SA, Cao X, Chen G, Mehra R, et al. A polycomb repression signature in metastatic prostate cancer predicts cancer outcome. *Cancer Res* 2007;67:10657–10663. [PubMed: 18006806]
41. Bruce AW, Donaldson IJ, Wood IC, Yerbury SA, Sadowski MI, Chapman M, et al. Genome-wide analysis of repressor element 1 silencing transcription factor/neuron-restrictive silencing factor (REST/NRSF) target genes. *Proc Natl Acad Sci U S A* 2004;101:10458–10463. [PubMed: 15240883]
42. Patel PD, Bochar DA, Turner DL, Meng F, Mueller HM, Pontrello CG. Regulation of tryptophan hydroxylase-2 gene expression by a bipartite RE-1 silencer of transcription/neuron restrictive silencing factor (REST/NRSF) binding motif. *J Biol Chem* 2007;282:26717–26724. [PubMed: 17613521]
43. Zheng D, Zhao K, Mehler MF. Profiling RE1/REST-mediated histone modifications in the human genome. *Genome Biol* 2009;10:R9. [PubMed: 19173732]
44. Bell AC, West AG, Felsenfeld G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* 1999;98:387–396. [PubMed: 10458613]
45. Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, et al. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* 2007;128:1231–1245. [PubMed: 17382889]
46. Cuddapah S, Jothi R, Schones DE, Roh TY, Cui K, Zhao K. Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res* 2009;19:24–32. [PubMed: 19056695]
47. Wallace JA, Felsenfeld G. We gather together: insulators and genome organization. *Curr Opin Genet Dev* 2007;17:400–407. [PubMed: 17913488]
48. Blom van Assendelft G, Hanscombe O, Grosveld F, Greaves DR. The beta-globin dominant control region activates homologous and heterologous promoters in a tissue-specific manner. *Cell* 1989;56:969–977. [PubMed: 2924354]
49. Grosveld F, van Assendelft GB, Greaves DR, Kollias G. Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell* 1987;51:975–985. [PubMed: 3690667]
50. Tuan DY, Solomon WB, London IM, Lee DP. An erythroid-specific, developmental-stage-independent enhancer far upstream of the human "beta-like globin" genes. *Proc Natl Acad Sci U S A* 1989;86:2554–2558. [PubMed: 2704733]
51. Farrell CM, West AG, Felsenfeld G. Conserved CTCF insulator elements flank the mouse and human beta-globin loci. *Mol Cell Biol* 2002;22:3820–3831. [PubMed: 11997516]
52. Ellis J, Tan-Un KC, Harper A, Michalovich D, Yannoutsos N, Philipsen S, et al. A dominant chromatin-opening activity in 5' hypersensitive site 3 of the human beta-globin locus control region. *EMBO J* 1996;15:562–568. [PubMed: 8599939]
53. Palstra RJ, Tolhuis B, Splinter E, Nijmeijer R, Grosveld F, de Laat W. The beta-globin nuclear compartment in development and erythroid differentiation. *Nat Genet* 2003;35:190–194. [PubMed: 14517543]

54. Spilianakis CG, Flavell RA. Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. *Nat Immunol* 2004;5:1017–1027. [PubMed: 15378057]
55. Hug BA, Wesselschmidt RL, Fiering S, Bender MA, Epner E, Groudine M, et al. Analysis of mice containing a targeted deletion of beta-globin locus control region 5' hypersensitive site 3. *Mol Cell Biol* 1996;16:2906–2912. [PubMed: 8649401]
56. Bender MA, Roach JN, Halow J, Close J, Alami R, Bouhassira EE, et al. Targeted deletion of 5'HS1 and 5'HS4 of the beta-globin locus control region reveals additive activity of the DNaseI hypersensitive sites. *Blood* 2001;98:2022–2027. [PubMed: 11567985]
57. Mahajan MC, Narlikar GJ, Boyapaty G, Kingston RE, Weissman SM. Heterogeneous nuclear ribonucleoprotein C1/C2, MeCP1, and SWI/SNF form a chromatin remodeling complex at the beta-globin locus control region. *Proc Natl Acad Sci U S A* 2005;102:15012–15017. [PubMed: 16217013]
58. Stamatoyanopoulos JA, Goodwin A, Joyce T, Lowrey CH. NF-E2 and GATA binding motifs are required for the formation of DNase I hypersensitive site 4 of the human beta-globin locus control region. *EMBO J* 1995;14:106–116. [PubMed: 7828582]
59. Kosak ST, Scalzo D, Alworth SV, Li F, Palmer S, Enver T, et al. Coordinate gene regulation during hematopoiesis is related to genomic organization. *PLoS Biol* 2007;5:e309. [PubMed: 18031200]
60. Misteli T. Beyond the sequence: cellular organization of genome function. *Cell* 2007;128:787–800. [PubMed: 17320514]
61. Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* 2008;453:948–951. [PubMed: 18463634]
62. Reddy KL, Zullo JM, Bertolino E, Singh H. Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature* 2008;452:243–247. [PubMed: 18272965]
63. Finlan LE, Sproul D, Thomson I, Boyle S, Kerr E, Perry P, et al. Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genet* 2008;4:e1000039.
64. Kumaran RI, Spector DL. A genetic locus targeted to the nuclear periphery in living cells maintains its transcriptional competence. *J Cell Biol* 2008;180:51–65. [PubMed: 18195101]
65. Solovei I, Kreysing M, Lanctot C, Kosem S, Peichl L, Cremer T, et al. Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. *Cell* 2009;137:356–368. [PubMed: 19379699]
66. Akhtar A, Gasser SM. The nuclear envelope and transcriptional control. *Nat Rev Genet* 2007;8:507–517. [PubMed: 17549064]
67. Palstra RJ, de Laat W, Grosveld F. Beta-globin regulation and long-range interactions. *Adv Genet* 2008;61:107–142. [PubMed: 18282504]
68. Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell* 2002;10:1453–1465. [PubMed: 12504019]
69. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science* 2002;295:1306–1311. [PubMed: 11847345]
70. Keys JR, Tallack MR, Zhan Y, Papathanasiou P, Goodnow CC, Gaensler KM, et al. A mechanism for Ikaros regulation of human globin gene switching. *Br J Haematol* 2008;141:398–406. [PubMed: 18318763]
71. Du MJ, Lv X, Hao DL, Zhao GW, Wu XS, Wu F, et al. MafK/NF-E2 p18 is required for beta-globin genes activation by mediating the proximity of LCR and active beta-globin genes in MEL cell line. *Int J Biochem Cell Biol* 2008;40:1481–1493. [PubMed: 18308612]
72. Vakoc CR, Letting DL, Gheldof N, Sawado T, Bender MA, Groudine M, et al. Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol Cell* 2005;17:453–462. [PubMed: 15694345]
73. Splinter E, Heath H, Kooren J, Palstra RJ, Klous P, Grosveld F, et al. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev* 2006;20:2349–2354. [PubMed: 16951251]
74. Hou C, Zhao H, Tanimoto K, Dean A. CTCF-dependent enhancer-blocking by alternative chromatin loop formation. *Proc Natl Acad Sci U S A* 2008;105:20398–20403. [PubMed: 19074263]

75. Majumder P, Gomez JA, Chadwick BP, Boss JM. The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions. *J Exp Med* 2008;205:785–798. [PubMed: 18347100]
76. Ling JQ, Li T, Hu JF, Vu TH, Chen HL, Qiu XW, et al. CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. *Science* 2006;312:269–272. [PubMed: 16614224]
77. Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, et al. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 2008;451:796–801. [PubMed: 18235444]
78. Simonis M, Kooren J, de Laat W. An evaluation of 3C-based methods to capture DNA interactions. *Nat Methods* 2007;4:895–901. [PubMed: 17971780]

Figure 1.

Model of Cis-Regulation at an Activated Promoter. Top Row: Organization of functional elements for a gene. Enh: enhancer; Sil: silencer; Ins: insulator; Pro: promoter. The p300-engaged enhancer will activate transcription, unless CTCF binds to an insulator site and blocks this activity. Middle row: Areas of distribution of H3 histone marks at an activated promoter. Bottom row: Nucleosome (circles) distribution at an activated promoter.