Gene Regulatory Elements

Chapter · September 2012

DOI: 10.1007/978-1-4614-1683-8_1

1 author:

Nadav Ahituv
University of California, San Francisco
211 PUBLICATIONS **7,836** CITATIONS

SEE PROFILE

Chapter 1
Gene Regulatory Elements

Nadav Ahituv

Abstract While the annotation and functional characterization of the 2% of our genome that encodes for protein has been extremely successful, the remaining 98% still remains primarily uncharted territory. Within this territory reside gene regulatory sequences that instruct genes when, where, and at what levels to turn on or off. There is abundant evidence, as described in this book, that nucleotide and epigenetic changes in these gene regulatory sequences can lead to human disease. In this chapter, we will define the different types of gene regulatory elements (promoters, enhancers, silencers, and insulators) and how to identify and functionally characterize them.

Keywords Promoter • Enhancer • Silencer • Insulator • Locus control region • Transcription factors • Nucleosome positioning • DNase I hypersensitive sites • ChIP • 3C

1.1 Introduction

The human genome consists of ~3.2 billion base pairs and encompasses around 20,500 genes (Clamp et al. 2007) which make up only ~1.6% of its content. Repetitive sequences make up an additional ~50%, and the remaining 48% is composed of DNA sequences with primarily unknown function. One vital function that is clearly

N. Ahituv (✉)
Department of Bioengineering and Therapeutic Sciences, Institute for Human Genetics, University of California San Francisco, 513 Parnassus Ave., Health Sciences East building, Room HSE901H, Box 0794, San Francisco, CA 94143, USA
e-mail: nadav.ahituv@ucsf.edu

N. Ahituv (ed.), *Gene Regulatory Sequences and Human Disease*,
DOI 10.1007/978-1-4614-1683-8_1,

embedded in these sequences is gene regulation, instructing genes when and where to turn on or off at specific levels. There is a variety of clinical and molecular data, as described in this book, demonstrating that regulatory elements can harbor mutational events that lead to human disease. However, the identification of distinctive nucleotide or epigenetic changes within these elements that lead to human disease has been extremely limited due in part to the vast genomic noncoding space in which to search for them, their scattered distribution, the unavailability of an established regulatory code, and the difficulties in linking these elements to specific genes.

The identification and functional characterization of these gene regulatory elements is not only important for their association with human disease but also for several other biological disciplines. Biologists have long been in need of defined sequences that drive precise patterns of expression. Such sequences can be used to express recombinant proteins (e.g., *Cre* to generate tissue-specific knockouts) or to overexpress various proteins in precise locations. These sequences could also be used to target specific DNA sequences to certain tissues for gene therapy purposes. Gene regulatory elements are also important developmental regulators, and the understanding of the factors that regulate these developmental genes can increase our knowledge of development. In evolution, regulatory sequences are thought to be a major contributor to the evolution of form (Carroll 2005). More and more examples are now being reported in various organisms, including humans (Prabhakar et al. 2008; McLean et al. 2011), that highlight the effect these sequences can have on morphological differences between species. An increased understanding of these gene regulatory elements will vastly assist these disciplines.

1.2 The Different Kinds of Gene Regulatory Elements

There are several different types of gene regulatory elements. In this section, we will define the most commonly characterized elements: promoters, enhancers, silencers, and insulators. The general dogma is that these regulatory elements get activated by the binding of *transcription factors*, proteins that bind to specific DNA sequences, and control mRNA transcription. There could be several transcription factors that need to bind to one regulatory element in order to activate it. In addition, several other proteins, called *transcription cofactors*, bind to the transcription factors themselves to control transcription.

The binding of these sequences changes the *nucleosome positioning* in that region. The *nucleosome* consists of 147 bp of DNA wrapped around a histone core. The binding of transcription-related proteins repositions the nucleosome and changes it into a more open state. As later described for DNase hypersensitive sites (Sect. 1.3.1), this change in nucleosome state can be used for the discovery of active gene regulatory elements. The nucleosome core consists of histone proteins which can have various posttranslational modifications. These modifications affect the state of this genomic region and can also be used to detect various gene regulatory elements as described in detail in the chromatin immunoprecipitation section (Sect. 1.3.3).

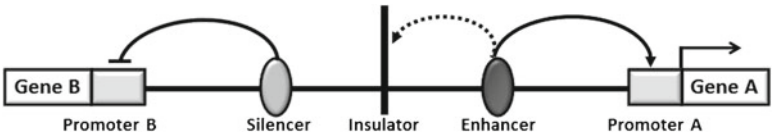

Fig. 1.1 *Schematic representation of the various gene regulatory elements described in this chapter.* Two different promoters are located next to the two different genes. By binding to *promoter A*, an enhancer actively regulates *Gene A* and leads to its transcription, as indicated by the *black arrow* above *Gene A*. An insulator prevents this enhancer from activating *Gene B*, which is not transcribed due to regulation by a silencer element

1.2.1 Promoters

The best characterized gene regulatory element is the promoter, in part due to its established location relative to the gene that it regulates. The *promoter* resides at the beginning of the gene and serves as the site where the transcription machinery assembles and transcription of the gene begins (Fig. 1.1). The *core promoter* is defined as the minimal stretch of DNA sequence that is sufficient to allow the RNA polymerase II machinery to initiate transcription (Butler and Kadonaga 2002; Smale and Kadonaga 2003). It is typically 35–40 base pairs (bp) long and encompasses several sequence motifs, such as the following: TATA box, TFIIB recognition element (BRE), initiator element (Inr), and the downstream core promoter element (DPE). It is important to note that not all of these elements need to be present in order to define a core promoter.

The *TATA box* was the first core promoter element to be discovered. Its consensus sequence is TATAAA, but this can be modified in many cases as long as the sequence remains A/T rich. In humans, it is thought to be bound predominantly by the TATA-binding protein (TBP) which is part of the transcription factor IID (TFIID) multiprotein complex that contributes to transcription initiation (Smale and Kadonaga 2003). 32% of all human promoters (Suzuki et al. 2001) are thought to contain a TATA box. The *TFIIB recognition element (BRE)* is a binding site for the transcription factor IIB (TFIIB) and is thought to facilitate the incorporation of this transcription factor to the transcription initiation complex (Lagrange et al. 1998). It has a consensus sequence of G/C-G/C-G/A-C-G-C-C and is usually located immediately upstream to the TATA box. The *initiator element (Inr)* is usually −3 to +5 bp from the transcription start site (TSS) with a typical consensus sequence of C/T-C/T-A-N-T/A-C/T-C/T. It is thought to be recognized by and interact with TFIID and RNA Polymerase II at different steps during the transcription process. The *downstream core promoter element (DPE)* is a binding site for TFIID that is usually found in promoters lacking a TATA box. It is located 28–32 bp upstream of the Inr, and the distance between both of these elements was shown to be important for proper TFIID binding and transcription (Burke and Kadonaga 1997; Kutach and Kadonaga 2000). Its consensus sequence is A/G-G-A/T-C/T-G/A/C.

CpG islands are defined as sequences that are at least 200 long with a G/C percentage that is greater than 50% (Gardiner-Garden and Frommer 1987). The human genome is estimated to have ~29,000 of these islands (Smale and Kadonaga 2003), half of which are near gene promoters (Suzuki et al. 2001). Promoters with CpG islands usually lack a TATA box and a DPE and have multiple binding sites for the transcription factor SP1 that is thought to direct the transcription machinery. CpG islands are usually unmethylated if the gene they regulate is expressed.

The *proximal promoter* is the region that is in the immediate vicinity (−250 to +250 bp) of the TSS of the gene. It can contain several transcription factor binding sites (TFBS) and is thought to serve as a tethering element for distant enhancers, enabling them to interact with the core promoter (Calhoun et al. 2002). Several additional promoter elements have been discovered. Since this book primarily deals with more distant regulatory elements, see Smale and Kadonaga (2003) and Riethoven (2010) for a more detailed review of various promoter elements.

1.2.2 Enhancers

Enhancers turn on the promoters at specific locations, times, and levels and can be simply defined as the "promoters of the promoter" (Fig. 1.1). They can be tissue specific or regulate gene expression in multiple tissues. They often have modular expression patterns, and a gene that is active in many tissues is likely influenced by multiple enhancers (Pennacchio et al. 2006; Visel et al. 2009a). They can also regulate gene activity at various time points. They can regulate in *cis*, meaning that they regulate a gene on the same chromosomal region as they are located, or in *trans*, regulating a gene that is located on a different chromosome as was shown for the H olfactory receptor enhancer (Lomvardas et al. 2006). *Cis* enhancers can be 5′ or 3′ to the regulated gene, in introns or even within the coding exon of the gene they regulate (Neznanov et al. 1997; Tumpel et al. 2008). These enhancers can be near the promoter or very far away, with some like the Sonic hedgehog (*SHH*) limb enhancer being as far as ~1,000,000 bp away from the gene it regulates (Lettice et al. 2003). Enhancer function is generally considered to be independent of location or orientation relative to the gene they regulate.

A given enhancer can have an additional enhancer or enhancers with overlapping activity, called a *shadow enhancer/s* (Hong et al. 2008). The enhancer closest to the gene is usually considered to be the primary enhancer while other, more distant, enhancer or enhancers with similar activity are the shadow enhancer/s. Shadow enhancers are thought to protect the essential activities of the primary enhancer in adverse genetic conditions or environmental pressure (Hobert 2010). In addition, they can also have their own unique regulatory activities. The existence of shadow enhancers might also explain why in certain cases the removal of potentially important enhancers in the mouse genome can lead to no apparent phenotype (Ahituv et al. 2007; Cretekos et al. 2008).

Enhancers are thought to function through the recruitment of transcription factors and subsequent physical interactions with the gene promoter. This physical interaction is thought to be carried out through DNA looping. The DNA looping is mediated by transcription cofactors which activate *cohesin*, a protein that links DNA sequences to one another (Wood et al. 2010; Dorsett 2011). Cohesin, along with Nipped-B homolog (NIPBL), its DNA loading factor, allows the binding of the enhancer to the promoter. In humans, mutations in *NIPBL* as described in Chap. 11 of this book, entitled "Cohesin and Human Diseases," can lead to a range of gene regulatory defects that cause Cornelia de Lange syndrome (MIM #122470). Following enhancer-promoter binding, it is then thought that a conformational change takes place in Mediator, a multiprotein complex consisting of over 30 proteins that subsequently activates transcription (Kagey et al. 2010; Malik and Roeder 2010). In addition to looping, other mechanisms such as enhancer transcription and chromatin modifications over large regions that encompass both enhancer and promoter have also been suggested for enhancer function (Bulger and Groudine 2011; Ong and Corces 2011).

1.2.3 Silencers

Opposite to enhancers, *silencers* are thought to turn off gene expression at specific time points and locations (Fig. 1.1). Not much is known about silencers in humans, primarily due to the lack of a good functional assay to characterize them. Similar to enhancers, they are also thought to be orientation independent and can be located almost anywhere with regard to the genes that they regulate. Transcription factors bind to the silencer sequences and along with transcription cofactors they act to repress the expression of the gene. These "negative" transcription factors are thus called *repressors*, and their cofactors are termed *corepressors* (Privalsky 2004).

Several different mechanisms have been proposed for silencer function. Recent work suggests that, similar to enhancers, they can interact with the promoter through DNA looping (Lanzuolo et al. 2007; Tiwari et al. 2008). Repressor and corepressor proteins can silence a gene by competing for the binding to a specific promoter (Li et al. 2004; Harris et al. 2005). They can also influence the local chromatin region by establishing repressive chromatin marks (Srinivasan and Atchison 2004).

1.2.4 Insulators

Insulators, also called boundary elements, are DNA sequences that create *cis*-regulatory boundaries that prevent the regulatory elements of one gene from

affecting neighboring genes (Fig. 1.1). Based on their function, they can be generally divided into two types: (1) *enhancer-blocking insulators* that obstruct enhancers by inhibiting their interaction with promoters and (2) *barrier insulators* that prevent the spread of heterochromatin.

Enhancer-blocking insulators have three different proposed models of function: (a) *promoter decoy*, where the insulator recruits the transcription machinery away from the promoter by having it bind to the insulator instead (Geyer 1997); (b) *physical barrier*, where the insulator sequence acts as a physical barrier that blocks the enhancer signal from reaching the promoter (Kong et al. 1997; Ling et al. 2004; Zhao and Dean 2004); and (c) *loop domain*, where the insulator sequences form loops with each other or other DNA sequences that interfere with enhancer function (Blanton et al. 2003; Byrd and Corces 2003; Ameres et al. 2005). In addition to binding to each other or other DNA sequences, it is thought that insulators could also form insulation loops by tethering the chromatin to various nuclear structures, such as the nucleolus (Yusufzai et al. 2004) and the nuclear lamina (Guelen et al. 2008). For a more detailed review of enhancer-blocking insulators, see Bushey et al. (2008).

Barrier insulators insulate genomic regions against silencing by *heterochromatin*, a chromatin state caused by tightly packing the DNA into a repressed/silenced form. Heterochromatin is marked by high levels of methylation in H3K9 and H3K27 histone residues (see Sect. 1.3.3 for more details) and CpG methylation along with a general lack of acetylation (a mark for active regions). Barrier insulators are thought to disrupt the spread of heterochromatin in the region they reside. They generally do this by modifying the substrates used to generate heterochromatin. This is done by recruiting histone acetyltransferases (HATs) or ATP-dependent nucleosome-remodeling complexes (Oki et al. 2004) or by removing nucleosomes (Bi et al. 2004). For a more detailed review of barrier insulators, see Gaszner and Felsenfeld (2006).

Probably the most widely studied insulator-associated protein is the *CCCTC-binding factor* (CTCF), which got its name due to its ability to recognize three regularly spaced repeats of CCCTC. CTCF is a ubiquitously expressed protein that has 11 zinc fingers and uses different combinations of them to identify and bind different DNA sequences. Binding of CTCF to DNA can protect that region from methylation, hence its ability to insulate (Filippova 2008). CTCF-binding sites, which serve as potential insulator regions, have been mapped throughout the human genome in several cell lines (Barski et al. 2007; Kim et al. 2007; Heintzman et al. 2009; Ernst et al. 2011) using chromatin immunoprecipitation (ChIP; technique described in detail in Sect. 1.3.3). Analysis of the location of these sites in various cell lines found that they remain largely unchanged, suggesting that insulator activity remains more or less constant in these different cell lines (Heintzman et al. 2009). CTCF-binding sites were also shown to overlap with cohesin-binding sites, suggesting that these two proteins interact together to achieve insulator function. In addition, CTCF depletion was shown to affect the genomic distribution of cohesin, but not the opposite, suggesting that cohesin localization is governed by CTCF (Parelho et al. 2008; Wendt et al. 2008).

1.2.5 Locus Control Region

A *locus control region* (LCR) is a region where various gene regulatory elements are clustered together to regulate a certain gene or several genes. One example is the well-characterized beta-globin LCR, described in Chap. 2 of this book entitled "The Hemoglobin Regulatory Regions." This region is composed of various regulatory elements that together regulate the expression of different globin genes during human development, a process known as hemoglobin switching (Sankaran et al. 2010).

1.3 Techniques to Identify Gene Regulatory Elements

Several techniques have been developed in order to identify gene regulatory elements. Building on advancements in molecular biology, primarily DNA sequencing technologies, these techniques are continuously being refined and made cost-efficient. These advancements are currently allowing the field to move from a "one-by-one" or "region-by-region" approach to a whole-genomic one.

1.3.1 DNase I Hypersensitive Sites

DNase I is an endonuclease that cleaves both single- and double-stranded DNA. Active or "open" chromatin regions are associated with nucleosome-free regions and hence known to be more attainable to DNase I. These regions were thus termed *DNase I hypersensitive sites* (DHSs). In the 1970s, it was already recognized that chromatin regions that contain active genes are twice as sensitive to DNase I digestion as regions where genes are inactive (Weintraub and Groudine 1976). In addition to active genes, DHSs can identify all types of active regulatory elements such as promoters, enhancers, silencers, and insulators. DHSs do not reveal the regulatory function of the sequence or the identity of the transcription factors that could be binding to it, but they can show whether a certain region in a specific cell type or tissue is active. With advancements in molecular biology, DHSs can now be discovered on a genomic scale using techniques such as DNase-chip (Crawford et al. 2006) which uses microarrays to hybridize captured DNase I hypersensitive sequences and DNase-Seq (Song and Crawford 2010) that uses massively parallel sequencing.

1.3.2 Comparative Genomics

The increasing availability of genomic data from multiple vertebrate genomes facilitated the ability to carry out comparative genomic studies on the human genome. These genomic comparisons allowed for the identification of noncoding regions in the human genome that have been conserved throughout evolution, suggesting that

this conservation is due to an important function (Boffelli et al. 2004; Dermitzakis et al. 2005). One such function could be gene regulation. The use of comparative genomic tools has been extremely successful in identifying regulatory elements in the human genome (Thomas et al. 2003; Woolfe et al. 2005; Pennacchio et al. 2006; Birney et al. 2007; Visel et al. 2008). These comparisons generally look at sequence conservation between two evolutionary distant species or between multiple closely related species. In general, genomic comparisons between species that are more distantly related through evolution yield fewer conserved sequences, but the sequences that are shared between them are more likely to be functional.

1.3.3 Chromatin Immunoprecipitation (ChIP)

ChIP is rapidly becoming the most effective and widely used tool to map putative regulatory sequences on a genomic scale. In ChIP, DNA-binding proteins are cross-linked to the DNA, and an antibody that is specific to a protein of interest is used to pull down the protein along with the DNA sequences that it is bound to. Following reversal of the cross-links, these DNA sequences can be identified either through quantitative PCR (ChIP-qPCR), by binding to a DNA microarray (ChIP-chip), or using massively parallel DNA sequencing technologies (ChIP-Seq) (Johnson et al. 2007; Visel et al. 2009b). Among these techniques, ChIP-Seq is rapidly becoming the most commonly used tool to map putative regulatory sequences on a genomic scale. As described later for the various regulatory elements, carrying out ChIP-Seq for various chromatin marks is rapidly becoming the gold standard in the identification of potential gene regulatory elements.

The most commonly used chromatin marks are histone modifications (listed in detail in Table 1.1). Each 147 bp of DNA is wrapped around eight histone proteins that are composed of the following protein pairs: H2A, H2B, H3, and H4, and sealed off by histone H1, making what is termed a nucleosome. These histones have protruding tails that have various modifications such as methylation, acetylation, phosphorylation, ubiquination, and sumoylation. These modifications can determine the status of the chromatin. For example, open chromatin which is indicative of enhancer activity can be identified based on one methyl group on the fourth lysine of histone H3. This is abbreviated as H3K4me1. Closed chromatin which is indicative of silencing can be identified by three methyl groups on the lysine in position 27 of H3 and is abbreviated as H3K27me3. Antibodies developed against these modifications can allow researchers to carry out ChIP and determine the chromatin state and regulatory potential (active, silenced, etc.) of specific sequences.

1.3.4 Chromatin Conformation Capture (3C)

In order to properly understand gene regulation, we must view it in three-dimensional space. As mentioned above, chromatin loops have been shown to be one of the

Table 1.1 Gene regulatory marks used for ChIP

Element	State	Mark	Selected references
Promoter	Active	H2A.Z histone variant	Barski et al. (2007)
		H2BK5me1	Barski et al. (2007)
		H3K4me2	Bernstein et al. (2005), Barski et al. (2007), Birney et al. (2007), Heintzman et al. (2007), Ernst et al. (2011)
		H3K4me3	Bernstein et al. (2005), Barski et al. (2007), Guenther et al. (2007), Mikkelsen et al. (2007), Heintzman et al. (2009), Ernst et al. (2011)
		H3K9me1	Barski et al. (2007)
		H3K9ac	Bernstein et al. (2005), Guenther et al. (2007), Ernst et al. (2011)
		H3K14ac	Guenther et al. (2007)
		H3K27me1	Barski et al. (2007)
		H4K20me1	Barski et al. (2007)
	Repressed	H3K27me3	Barski et al. (2007)
		H3K79me3	Barski et al. (2007)
Enhancer		p300 (EP300)	Heintzman et al. (2009)
		CBP	Kim et al. (2010)
		H2A.Z	Barski et al. (2007), Ernst et al. (2011)
		H3K4me1	Birney et al. (2007), Heintzman et al. (2007), Heintzman et al. (2009)
		H3K4me2	Bernstein et al. (2005), Barski et al. (2007), Birney et al. (2007), Heintzman et al. (2007), Ernst et al. (2011)
		H3K27ac	Heintzman et al. (2009), Creyghton et al. (2010), Ernst et al. (2011), Rada-Iglesias et al. (2011)
Silenced regions		DNA methylation	Tiwari et al. (2008)
		H3K9me2	Barski et al. (2007)
		H3K9me3	Barski et al. (2007), Ernst et al. (2011)
		H3K27me2	Barski et al. (2007)
		H3K27me3	Barski et al. (2007), Mikkelsen et al. (2007), Ernst et al. (2011)
Insulator		CTCF	Barski et al. (2007), Kim et al. (2007), Heintzman et al. (2009), Ernst et al. (2011)

major mechanisms by which the various regulatory elements (promoters, enhancers, silencers, and insulators) regulate. To unravel the physical interactions of the various regulatory elements in the nucleus, *chromatin conformation capture* (3C) and several derivatives of this technique have been developed. They are primarily based on cross-linking DNA-binding proteins with DNA (similar to ChIP) so that both the

regulatory element and the other region of DNA with which it interacts are bound together, bridged by the proteins that facilitate this interaction. The DNA is then cut randomly with restriction enzymes and ligated in conditions where the segments of DNA bridged by the protein cross-linked bundle will preferentially ligate to one another rather than to random free DNA. These newly ligated DNA segments are then analyzed in order to identify which regions of DNA have been joined, implying that they physically interact. The subsequent analysis of these sequences is what determines whether this technique is known as 3C, 4C, or 5C. 3C uses real-time PCR with primers matching two specific candidate regions in order to determine whether they interact with one another (Vassetzky et al. 2009). 4C generates circular DNA molecules following ligation, and the PCR primers are used in order to determine the identity of the DNA sequences that interacts with a specific sequence by having them faced outward on opposite ends of the restriction enzyme fragment (Vassetzky et al. 2009). For example, this technique can be used to determine the identity of the DNA sequences/regulatory elements that bind to a specific promoter by designing primers specific to that promoter. 5C can detect numerous chromatin interactions at the same time by using several primers (van Berkum and Dekker 2009). With the advent of massively parallel sequencing technologies, whole-genome adaptations of this technique have been introduced such as Hi-C (Lieberman-Aiden et al. 2009) and ChIA-PET (Fullwood et al. 2009).

1.4 Techniques to Functionally Characterize Gene Regulatory Elements

1.4.1 Promoters

The standard promoter assay is straightforward. The candidate promoter sequence is placed in front of a reporter gene (Fig. 1.2a), which is used as an indicator of where the promoter is active by inserting this construct into a cell culture or animal model. Various reporter genes are used for this assay, usually depending on the context of the assay. In cell culture, the most commonly used reporter gene is luciferase, in zebrafish fluorescent proteins such as GFP or mCherry, and in mice the frequently used reporter gene is LacZ. The DNA sequence that is typically assayed for promoter activity covers at least 250–500 bp upstream of the transcription start site, so as to include the proximal promoter and other potential promoter-associated sequences.

1.4.2 Enhancers

Enhancers can be functionally characterized by various methods: deletion series, enhancer traps, cell culture enhancer assays, and in vivo electroporation or transgenic

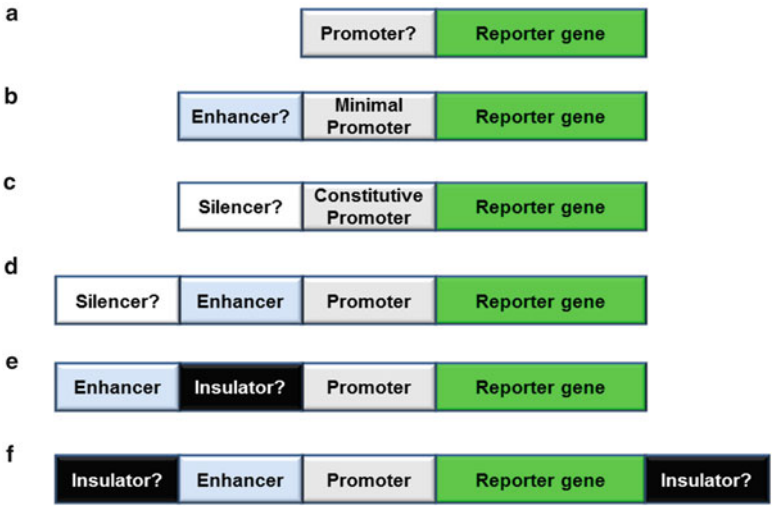

Fig. 1.2 *DNA construct designs that are commonly used to functionally characterize gene regulatory elements.* (**a**) A DNA sequence assayed for promoter activity is placed in front of a reporter gene. (**b**) A DNA sequence assayed for enhancer activity is placed in front of a minimal promoter (a promoter that is not sufficient to drive expression without the presence of a functional enhancer) and a reporter gene. (**c**) A DNA sequence assayed for silencer activity is placed in front of a constitutive promoter (a promoter that is always active) and a reporter gene or in front of a characterized enhancer and promoter followed by a reporter gene (**d**). (**e**) DNA sequences can be assayed for insulator activity by blocking the ability of a characterized enhancer to turn on a reporter gene via a characterized promoter. They can also be assayed for barrier insulator activity by placing them on both sides of a known enhancer and promoter that drive reporter gene expression and measuring the activity of this reporter gene for consistency and length of expression following stable integration (**f**)

enhancer assays. *Deletion series* uses a long DNA construct, such as a yeast artificial chromosome (YAC) or bacterial artificial chromosome (BAC), that encompasses the genomic region harboring a potential enhancer or enhancers and some kind of reporter gene that is inserted into this DNA fragment. These are then injected into an organism in order to observe the reporter gene expression pattern. Once an expression pattern of interest is observed, the YAC or BAC can be modified such that certain sequences are deleted and these new constructs are reinjected to compare for changes in reporter gene expression due to this manipulation. An analogous approach is the use of overlapping YACs or BACs that also encompass a reporter gene. Each construct is injected separately and observed for its respective reporter gene expression, allowing the ability to locate tissue-specific enhancers based on their shared genomic region (Mortlock et al. 2003). These overlapping constructs could also be deleted, as mentioned above, to further pinpoint the enhancer location. Combined, these assays are able to broadly identify the location of the enhancer/s, but usually fall short of determining its exact sequence.

The majority of enhancer assays are based on a common design: the assayed sequence is placed in front of a minimal promoter (a promoter that is not sufficient to

drive expression without the presence of a functional enhancer), followed by a reporter gene (Fig. 1.2b), similar to reporters used for promoter assays. If the assayed sequence is an enhancer, it will turn on the minimal promoter which in turn will drive the reporter gene expression. *Enhancer traps* use the random genomic integration of a DNA sequence containing a minimal promoter and the reporter gene to uncover enhancers around the region of integration (Parinov et al. 2004; Korzh 2007). Similar to the deletion series, the disadvantage of this method is that it does not pinpoint the exact location of the enhancer.

To assay specific sequences for enhancer activity, candidate sequences are typically cloned into a vector containing a minimal promoter and a reporter gene as outlined above. This vector is then introduced into a model system via different methods. *Cell culture enhancer assays* usually use transfection, electroporation, or virus integration to insert constructs into cells and assay for enhancer activity using luciferase as the reporter gene. The advantage of cell culture-based enhancer assays is that they are relatively cheap, can be done in a high-throughput manner, and can be quantitative. However, they are carried out in cell lines which can lose a lot of their tissue characteristics; they do not provide "whole-organism" properties, and the results can be highly variable due to factors such as different DNA preparation methods, number of cell passages, cell culture conditions, and others.

In vivo enhancer assays provide the advantage of being able to test the assayed sequence in the context of the whole organism. Due to the length of time and technical aspects of observing adult reporter gene expression in vertebrates, the majority of in vivo-based enhancer assays are carried out at developmental time points. Zebrafish transgenic enhancer assays often use the Tol2 transposon for genomic integration along with a fluorescent reporter gene for visualization (Fisher et al. 2006). In frogs, transgenic enhancer assays can be carried out by using standard sperm nuclear transplantation and can utilize transposons (Khokha and Loots 2005). Chicken enhancer assays are usually based on electroporation of the assayed construct into the embryonic tissue of interest, followed by visualization using fluorescence or LacZ (Uchikawa 2008). Mouse transgenic enhancer assays tend to use standard mouse transgenic techniques (Nagy et al. 2002) and LacZ as the reporter gene. These mouse assays are usually transient (embryos are removed at a certain time point for reporter gene detection), avoiding the costs associated with maintaining mouse lines (Pennacchio et al. 2006). The main caveat for the majority of these assays is that they are not able to quantitatively measure enhancer activity because there can be multiple integrations of the enhancer construct per animal, leading to variation in the reporter intensity.

1.4.3 Silencers

Silencers can be detected by placing the sequence to be assayed in front of a constitutive promoter (a promoter that is always active) and a reporter gene and comparing the activity of that reporter gene with and without the assayed sequence

(Fig. 1.2c; Petrykowska et al. 2008). If the sequence is a silencer, lower reporter gene expression levels due to the assayed sequence silencing the constitutive promoter would be observed. As previously described, silencers can also interfere with the binding of a nearby activating site. In order to detect these kinds of silencers, an enhancer blocker assay was developed where the assayed sequence is placed in front of a characterized enhancer, promoter, and a reporter gene (Fig. 1.2d; Petrykowska et al. 2008). If the sequence silences by enhancer blocking, there should be a reduction in reporter gene activity in this assay. Both these assays can work well in cell culture, using a luciferase reporter, because reporter expression can be quantified. However, these techniques are not straightforward for in vivo silencer assays. In both zebrafish and mouse transgenic assays mentioned above, there is a high degree of variability between embryos in the number of inserted transgenes. This variability does not allow quantitative measurements of reporter gene expression differences and has hampered the development of in vivo silencer assays.

1.4.4 Insulators

As mentioned previously, insulators can be divided into two types: enhancer blockers that obstruct enhancers by inhibiting their interaction with promoters and barrier insulators that prevent the spread of heterochromatin. To assay for enhancer blockers, the sequence being analyzed is placed in between a characterized enhancer and promoter, which is followed by a reporter gene (Fig. 1.2e). If the sequence is an enhancer blocker insulator, it should block the ability of the enhancer to activate the promoter and thus lead to reduced reporter gene expression versus a vector that does not contain this sequence. Barrier insulator assays consist of placing an enhancer, promoter, and a reporter gene in between two copies of the presumed barrier insulator (Fig. 1.2f) and having it stably integrate into the genome. The reporter gene expression is then monitored for consistency and length of expression. A sequence will be considered a barrier insulator if it provides for consistent reporter gene expression over a certain period of time (Recillas-Targa et al. 2002; Gaszner and Felsenfeld 2006).

1.5 Summary

We are in the midst of revolutionary times, with novel DNA sequencing technologies increasing our ability to sequence DNA at enormous rates. Due to these technological advances, individual whole-genome sequences are readily available at an affordable price. This availability will have the most immediate impact on two major fields: predicting human disease risk and pharmacogenomics. The pioneering work described in this book that led to the detection of various human disease-causing regulatory mutations and the techniques described in this chapter will

provide us with a starting foundation to analyze gene regulatory mutations in whole-genome data sets. As these whole-genome data sets expand and high-throughput functional gene regulatory assays develop, we will gain an increased understanding of how gene regulatory mutations lead to human phenotypes.

Abbreviations

BRE	TFIIB recognition element
Inr	Initiator element
DPE	Downstream core promoter element
TBP	TATA-binding protein
TFIID	Transcription factor IID
TFIIB	Transcription factor IIB
TSS	Transcription start site
TFBS	Transcription factor binding sites
SHH	Sonic hedgehog
NIPBL	Nipped-B homolog
CTCF	CCCTC-binding factor
LCR	Locus control region
DHSs	DNase I hypersensitive sites
ChIP	Chromatin immunoprecipitation
3C	Chromatin conformation capture
YAC	Yeast artificial chromosome
BAC	Bacterial artificial chromosome

References

Ahituv N, Zhu Y et al (2007) Deletion of ultraconserved elements yields viable mice. PLoS Biol 5(9):e234

Ameres SL, Drueppel L et al (2005) Inducible DNA-loop formation blocks transcriptional activation by an SV40 enhancer. EMBO J 24(2):358–367

Barski A, Cuddapah S et al (2007) High-resolution profiling of histone methylations in the human genome. Cell 129(4):823–837

Bernstein BE, Kamal M et al (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. Cell 120(2):169–181

Bi X, Yu Q et al (2004) Formation of boundaries of transcriptionally silent chromatin by nucleosome-excluding structures. Mol Cell Biol 24(5):2118–2131

Birney E, Stamatoyannopoulos JA et al (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447(7146):799–816

Blanton J, Gaszner M et al (2003) Protein:protein interactions and the pairing of boundary elements in vivo. Genes Dev 17(5):664–675

Boffelli D, Nobrega MA et al (2004) Comparative genomics at the vertebrate extremes. Nat Rev Genet 5(6):456–465

Bulger M, Groudine M (2011) Functional and mechanistic diversity of distal transcription enhancers. Cell 144(3):327–339

Burke TW, Kadonaga JT (1997) The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila. Genes Dev 11(22): 3020–3031
Bushey AM, Dorman ER et al (2008) Chromatin insulators: regulatory mechanisms and epigenetic inheritance. Mol Cell 32(1):1–9
Butler JE, Kadonaga JT (2002) The RNA polymerase II core promoter: a key component in the regulation of gene expression. Genes Dev 16(20):2583–2592
Byrd K, Corces VG (2003) Visualization of chromatin domains created by the gypsy insulator of Drosophila. J Cell Biol 162(4):565–574
Calhoun VC, Stathopoulos A et al (2002) Promoter-proximal tethering elements regulate enhancer-promoter specificity in the Drosophila Antennapedia complex. Proc Natl Acad Sci USA 99(14):9243–9247
Carroll SB (2005) Evolution at two levels: on genes and form. PLoS Biol 3(7):e245
Clamp M, Fry B et al (2007) Distinguishing protein-coding and noncoding genes in the human genome. Proc Natl Acad Sci USA 104(49):19428–19433
Crawford GE, Davis S et al (2006) DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. Nat Methods 3(7):503–509
Cretekos CJ, Wang Y et al (2008) Regulatory divergence modifies limb length between mammals. Genes Dev 22(2):141–151
Creyghton MP, Cheng AW et al (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci USA 107(50):21931–21936
Dermitzakis ET, Reymond A et al (2005) Conserved non-genic sequences – an unexpected feature of mammalian genomes. Nat Rev Genet 6(2):151–157
Dorsett D (2011) Cohesin: genomic insights into controlling gene transcription and development. Curr Opin Genet Dev 21(2):199–206
Ernst J, Kheradpour P et al (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473(7345):43–49
Filippova GN (2008) Genetics and epigenetics of the multifunctional protein CTCF. Curr Top Dev Biol 80:337–360
Fisher S, Grice EA et al (2006) Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. Nat Protoc 1(3):1297–1305
Fullwood MJ, Liu MH et al (2009) An oestrogen-receptor-alpha-bound human chromatin interactome. Nature 462(7269):58–64
Gardiner-Garden M, Frommer M (1987) CpG islands in vertebrate genomes. J Mol Biol 196(2):261–282
Gaszner M, Felsenfeld G (2006) Insulators: exploiting transcriptional and epigenetic mechanisms. Nat Rev Genet 7(9):703–713
Geyer PK (1997) The role of insulator elements in defining domains of gene expression. Curr Opin Genet Dev 7(2):242–248
Guelen L, Pagie L et al (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature 453(7197):948–951
Guenther MG, Levine SS et al (2007) A chromatin landmark and transcription initiation at most promoters in human cells. Cell 130(1):77–88
Harris MB, Mostecki J et al (2005) Repression of an interleukin-4-responsive promoter requires cooperative BCL-6 function. J Biol Chem 280(13):13114–13121
Heintzman ND, Stuart RK et al (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet 39(3):311–318
Heintzman ND, Hon GC et al (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 459(7243):108–112
Hobert O (2010) Gene regulation: enhancers stepping out of the shadow. Curr Biol 20(17):R697–R699
Hong JW, Hendrix DA et al (2008) Shadow enhancers as a source of evolutionary novelty. Science 321(5894):1314
Johnson DS, Mortazavi A et al (2007) Genome-wide mapping of in vivo protein-DNA interactions. Science 316(5830):1497–1502

Kagey MH, Newman JJ et al (2010) Mediator and cohesin connect gene expression and chromatin architecture. Nature 467(7314):430–435
Khokha MK, Loots GG (2005) Strategies for characterising cis-regulatory elements in Xenopus. Brief Funct Genomic Proteom 4(1):58–68
Kim TH, Abdullaev ZK et al (2007) Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell 128(6):1231–1245
Kim TK, Hemberg M et al (2010) Widespread transcription at neuronal activity-regulated enhancers. Nature 465(7295):182–187
Kong S, Bohl D et al (1997) Transcription of the HS2 enhancer toward a cis-linked gene is independent of the orientation, position, and distance of the enhancer relative to the gene. Mol Cell Biol 17(7):3955–3965
Korzh V (2007) Transposons as tools for enhancer trap screens in vertebrates. Genome Biol 8(Suppl 1):S8
Kutach AK, Kadonaga JT (2000) The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. Mol Cell Biol 20(13):4754–4764
Lagrange T, Kapanidis AN et al (1998) New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. Genes Dev 12(1):34–44
Lanzuolo C, Roure V et al (2007) Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. Nat Cell Biol 9(10):1167–1174
Lettice LA, Heaney SJ et al (2003) A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum Mol Genet 12(14):1725–1735
Li L, He S et al (2004) Gene regulation by Sp1 and Sp3. Biochem Cell Biol 82(4):460–471
Lieberman-Aiden E, van Berkum NL et al (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326(5950):289–293
Ling J, Ainol L et al (2004) HS2 enhancer function is blocked by a transcriptional terminator inserted between the enhancer and the promoter. J Biol Chem 279(49):51704–51713
Lomvardas S, Barnea G et al (2006) Interchromosomal interactions and olfactory receptor choice. Cell 126(2):403–413
Malik S, Roeder RG (2010) The metazoan mediator co-activator complex as an integrative hub for transcriptional regulation. Nat Rev Genet 11(11):761–772
McLean CY, Reno PL et al (2011) Human-specific loss of regulatory DNA and the evolution of human-specific traits. Nature 471(7337):216–219
Mikkelsen TS, Ku M et al (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448(7153):553–560
Mortlock DP, Guenther C et al (2003) A general approach for identifying distant regulatory elements applied to the Gdf6 gene. Genome Res 13(9):2069–2081
Nagy A, Gertsenstein M et al (2002) Manipulating the mouse embryo: a laboratory manual. Cold Spring Harbor, New York
Neznanov N, Umezawa A et al (1997) A regulatory element within a coding exon modulates keratin 18 gene expression in transgenic mice. J Biol Chem 272(44):27549–27557
Oki M, Valenzuela L et al (2004) Barrier proteins remodel and modify chromatin to restrict silenced domains. Mol Cell Biol 24(5):1956–1967
Ong CT, Corces VG (2011) Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat Rev Genet 12(4):283–293
Parelho V, Hadjur S et al (2008) Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell 132(3):422–433
Parinov S, Kondrichin I et al (2004) Tol2 transposon-mediated enhancer trap to identify developmentally regulated zebrafish genes in vivo. Dev Dyn 231(2):449–459
Pennacchio LA, Ahituv N et al (2006) In vivo enhancer analysis of human conserved non-coding sequences. Nature 444(7118):499–502
Petrykowska HM, Vockley CM et al (2008) Detection and characterization of silencers and enhancer-blockers in the greater CFTR locus. Genome Res 18(8):1238–1246

Prabhakar S, Visel A et al (2008) Human-specific gain of function in a developmental enhancer. Science 321(5894):1346–1350
Privalsky ML (2004) The role of corepressors in transcriptional regulation by nuclear hormone receptors. Annu Rev Physiol 66:315–360
Rada-Iglesias A, Bajpai R et al (2011) A unique chromatin signature uncovers early developmental enhancers in humans. Nature 470(7333):279–283
Recillas-Targa F, Pikaart MJ et al (2002) Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. Proc Natl Acad Sci USA 99(10):6883–6888
Riethoven JJ (2010) Regulatory regions in DNA: promoters, enhancers, silencers, and insulators. Methods 674:33–42
Sankaran VG, Xu J et al (2010) Advances in the understanding of haemoglobin switching. Br J Haematol 149(2):181–194
Smale ST, Kadonaga JT (2003) The RNA polymerase II core promoter. Annu Rev Biochem 72:449–479
Song L, Crawford GE (2010) DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. Cold 2010(2):pdb.prot5384
Srinivasan L, Atchison ML (2004) YY1 DNA binding and PcG recruitment requires CtBP. Genes Dev 18(21):2596–2601
Suzuki Y, Tsunoda T et al (2001) Identification and characterization of the potential promoter regions of 1031 kinds of human genes. Genome Res 11(5):677–684
Thomas JW, Touchman JW et al (2003) Comparative analyses of multi-species sequences from targeted genomic regions. Nature 424(6950):788–793
Tiwari VK, McGarvey KM et al (2008) PcG proteins, DNA methylation, and gene repression by chromatin looping. PLoS Biol 6(12):2911–2927
Tumpel S, Cambronero F et al (2008) A regulatory module embedded in the coding region of Hoxa2 controls expression in rhombomere 2. Proc Natl Acad Sci USA 105(51):20077–20082, Epub 2008 Dec 22
Uchikawa M (2008) Enhancer analysis by chicken embryo electroporation with aid of genome comparison. Dev Growth Differ 50(6):467–474
van Berkum NL, Dekker J (2009) Determining spatial chromatin organization of large genomic regions using 5C technology. Methods Mol Biol 567:189–213
Vassetzky Y, Gavrilov A et al (2009) Chromosome conformation capture (from 3C to 5C) and its ChIP-based modification. Methods Mol Biol 567:171–188
Visel A, Prabhakar S et al (2008) Ultraconservation identifies a small subset of extremely constrained developmental enhancers. Nat Genet 6:6
Visel A, Akiyama JA et al (2009a) Functional autonomy of distant-acting human enhancers. Genomics 93(6):509–513
Visel A, Blow MJ et al (2009b) ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 457(7231):854–858
Weintraub H, Groudine M (1976) Chromosomal subunits in active genes have an altered conformation. Science 193(4256):848–856
Wendt KS, Yoshida K et al (2008) Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature 451(7180):796–801
Wood AJ, Severson AF et al (2010) Condensin and cohesin complexity: the expanding repertoire of functions. Nat Rev Genet 11(6):391–404
Woolfe A, Goodson M et al (2005) Highly conserved non-coding sequences are associated with vertebrate development. PLoS Biol 3(1):e7
Yusufzai TM, Tagami H et al (2004) CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. Mol Cell 13(2):291–298
Zhao H, Dean A (2004) An insulator blocks spreading of histone acetylation and interferes with RNA polymerase II transfer between an enhancer and gene. Nucleic Acids Res 32(16): 4903–4919, Print 2004