Published in final edited form as:

Diabetes Obes Metab. 2012 October ; 14(Suppl 3): 1–11. doi:10.1111/j.1463-1326.2012.01645.x.

Chromatin structure, epigenetic mechanisms and long-range interactions in the human insulin locus

Z. Xu, G. M. Lefevre, and G. Felsenfeld

Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA

Abstract

Regulation of gene expression in eukaryotes is largely dependent on variations in chromatin structure. More recently, it has become clear that this may involve not only local chromatin organization but also distant regulatory elements that participate in large-scale chromatin architecture within the nucleus. We describe recent methods that make possible the detection of such structures and apply them to analysis of the human insulin (*INS*) locus in pancreatic islets. We show that the *INS* gene is part of an extended ‘open’ chromatin domain that includes adjacent genes as well. We also find that in islets, the *INS* promoter is in physical contact with distant sites on the same human chromosome and notably, with the *SYT8* gene, located nearly 300 kb away. The strength of the contact between *INS* and *SYT8* is increased by glucose, and this results in stimulation of *SYT8* expression. Inhibition of *INS* transcription decreases *SYT8* expression. Furthermore, downregulation of *SYT8* results in decreased secretion of insulin. Our results thus establish the existence of a regulatory network between the *INS* gene and other distant genes through long-range physical interactions, and suggest that such networks may have general importance for insulin biology and diabetes.

Keywords

chromatin conformation capture (3C); histones; synaptotagmin 8 (SYT8)

Introduction

It is widely understood that the organization of the eukaryotic genome within the nucleus plays a critical role in control of development and the patterns of gene expression characteristic of individual cell types. At the simplest level of chromatin organization DNA is packaged in nucleosomes, in which it is wrapped around octamers of basic proteins, the histones. Nucleosomes are themselves arranged in strings that are further folded into more compact structures, which ultimately become part of highly condensed chromosomal domains. Within the domains, however, there is evidence for specific architecture associated with mechanisms that are essential for appropriate patterns of gene expression.

Correspondence to: Gary Felsenfeld, Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA., gary.felsenfeld@nih.gov.

Conflict of Interest

The authors declare no conflict of interest related to this article.

During the past several years, studies of the control of gene expression have largely focused on the way in which DNA sequence-specific transcription factors act on local chromatin structure, through rearrangement of nucleosomes to allow access to regulatory sequences on DNA, or through modification of specific amino acids on the histones within the nucleosome. Among others, such modifications include lysine acetylation, methylation or ubiquitylation, arginine methylation, and serine or threonine phosphorylation or modification by N-acetylglucosamine. Most of these modifications, which reside on the relatively accessible N-or C-terminal tails of the histones, serve as signals for the recruitment of specific transcription factors or co-factors, as well as enzymes that mobilize nucleosomes or add or remove further modifications. Modifications are found not only at promoters and enhancers but also over gene bodies where they may affect the function of a transcribing RNA polymerase. They can also be found over transcriptionally silent, heterochromatic regions of the genome, where, often in cooperation with DNA methylation, they help to limit unwanted expression. Although the term ‘histone code’ has been used to describe the relationship between specific patterns of modification and particular functional states, the great variety and perhaps degeneracy of patterns often makes it difficult to establish a one-to-one correspondence between a given function and a unique pattern. Such an assignment is made more difficult by the observation that in many cases the activation of a gene, for example, requires a sequential order of modifications in which each one acts as a signal for recruitment of the next; the chromatin structure is a device for integrating signals over time by establishing a memory of preceding events, which may themselves be only transient [1]. Histone modifications are often referred to as ‘epigenetic marks’, and there is evidence that particular patterns of modification may in some cases be copied during cell division, transmitting a regulatory pattern to daughter cells. Whether or not that strict criterion for use of the term ‘epigenetic’ is met in most cases, it is clear that histone modifications play a dominant role in establishing or maintaining the distinguishing patterns of gene expression in different cell types.

Genome-wide analysis of the distribution of the various histone modifications has been helpful in identifying ‘signatures’ characteristic for example of enhancer elements [2]. That work has focused renewed attention on the importance of distant regulatory elements in regulation of gene expression, already made clear for example in early work on the human *β-globin* locus [3,4], and now a widely documented phenomenon. New techniques support the earlier proposal [5] that distant regulatory elements affect target genes by physical contacts with the genes’ promoters. Recent methods, however, have greatly expanded the scale of distances over which such interactions can be detected, so that we now have to contemplate the possibility of contacts between sites a megabase or more apart, or even on different chromosomes.

Long-range Organization of the Genome

The fact that distant genomic regions interact has led to realization that the three dimensional (3D) architecture of the nucleus has functional importance. As a result, there has been great interest in understanding what DNA sequence elements and associated proteins might play a role in constraining this architecture and organizing genomic regions into discrete domains. DNA elements called ‘insulators’ have this function [6]. The best-

documented vertebrate insulators were discovered in a compound element 5' hypersensitive site 4 (5'HS4) located upstream of the chicken β -globin locus [7]. This element combines two activities that serve to organize the nearby genome: (i) It can block the spread of silent, compact heterochromatin, marked by 'silencing' histone modifications, into adjacent transcriptionally active regions (figure 1A) and (ii) It can protect a gene from being activated inappropriately by a distal enhancer belonging to another gene system (figure 1B, C). These two activities are separable in the 5'HS4 element. We showed some years ago that the 'enhancer blocking' activity of 5'HS4 is attributable to the binding of the protein CCCTC-binding factor (CTCF) to a site within it [8]. The enhancer activity is blocked when a CTCF binding site is placed between an enhancer and promoter, but not when the site is positioned elsewhere in a test construct (figure 1C).

On the basis of earlier work in *Drosophila* [9], we suggested that CTCF insulator function reflected an ability of bound CTCF to stabilize the formation of large chromatin loop domains within the nucleus [10]. In this model, an enhancer and a promoter on different loops are restricted in their ability to interact, whereas there is no restriction when they are situated in the same loop (figure 2A). Many recent studies ([9] and see below) confirm this idea. It is apparent that such CTCF-stabilized interactions could in other situations result in increased expression of a target gene, for example when loop formation results in bringing an enhancer and promoter into close contact (figure 2B). The genome-wide distribution of sites occupied by CTCF has been determined for many cell types. Typically, there are 25 000 or more such sites, many or most of which could serve to stabilize or block long-range interactions. The human *insulin-like growth factor 2 (IGF2)/H19* locus provides examples of both kinds of processes.

IGF2/H19: Architecture of an Imprinted Locus

The *IGF2/H19* imprinted locus (figure 2C, D) provided the earliest demonstration of the functional importance of CTCF. Situated immediately downstream of the Insulin (*INS*) gene, *IGF2* is expressed only from the allele of paternal origin. The allele-specific expression of *IGF2* is mainly controlled by the methylation status of a nearby DNA sequence called the imprinted control region (ICR). This DNA element, which contains seven CTCF binding sites in man and four in mouse [11,12], can only be bound by CTCF when unmethylated, that is, on the maternal allele. CTCF then establishes an enhancer-blocking insulator that prevents an enhancer located further downstream from activating *IGF2* expression on the maternal allele (figure 2C). In contrast, on the paternal allele where the ICR is methylated, CTCF cannot bind, the insulator does not function and *IGF2* is expressed (figure 2D). Although there are other mechanisms that contribute to *IGF2* allele-specific expression, this appears to be the major contributor.

Subsequent studies have addressed the structural role of CTCF at the *IGF2/H19* locus. By making use of techniques for measuring long-range contacts within the nucleus (discussed below), it has been possible in mice to map independently the chromatin conformation in the neighbourhood of the locus for each allele. The two alleles show conformations that reflect the presence or absence of bound CTCF and are consistent with its ability to stabilize long-

range interactions [13]. The structures that are generated by these interactions may, as discussed above, prevent contacts between some regulatory elements but encourage others.

Chromatin Structure of the Human Insulin Locus

The human *INS* gene can reasonably be considered part of the *IGF2/H19* locus. It is only 1 kb upstream of *IGF2*, thus much closer than *H19*, which is over 160 kb away. Read-through RNAs transcribed from the *INS* gene and extending into the *IGF2* gene (*INS-IGF2*) have also been observed in human pancreatic islets [14]. In addition, there is evidence that *INS* is imprinted just like *IGF2* in the human yolk sac, although this feature is essentially lost later in development [14]. When analyzing the chromatin structure of the locus spanning from *Achaete-scute complex homolog 2 (ASCL2)* (figure 3), we found that in human islet cells, unlike what is observed in non-insulin producing cells [15], the region where *INS* and its neighbouring genes – *tyrosine hydroxylase (TH)*, *IGF2 antisense (IGF2AS)* and *IGF2* – lie constitutes an active chromatin domain, distinct from the flanking gene-poor regions (figure 3). Elevated levels of active histone marks, such as histone 3 lysine 4 dimethylation (H3K4me2) and H3 and H4 acetylation, and low levels of the inactive histone mark H3K27me3 were detected across the entire domain, indicating that in human islet cells, chromatin at the *INS* locus is in an ‘open’, transcriptionally permissive conformation. Interestingly, the promoter region of *INS* showed a lower level of the H3K4me3 histone mark associated with active promoters, despite a high transcription rate [15,16]. This suggests that *INS* transcription in islets is regulated through an unconventional mechanism, which based on genome-wide studies [16], seems to be common to many islet hormone-coding genes.

Further characterization of the open chromatin domain revealed features unique to islet cells [15]. First, the domain contains two CTCF-bound sites, and expression levels of the genes within the domain are correlated with each other, as well as with the expression level of PDX1 (figure 4A, B). Furthermore, levels of DNA methylation are lower than in control HeLa cells at all gene promoters and along gene bodies [15]. Second, non-coding RNAs are generated from intergenic regions across the 80-kb domain and the abundance of these non-coding transcripts is correlated with *INS* expression levels and the amount of active histone marks across the domain. In similar cases, such transcripts have been shown to help keep the chromatin in an open conformation [17]. Third, the zinc-finger protein (ZFP) *Myc-associated zinc-finger protein (MAZ)* is bound to the variable number of tandem repeats (VNTR) region present near the *INS* promoter [15]. A study showed that binding of *INS* VNTRs by an engineered ZFP in cells that normally do not express *INS* results not only in the transcription of *INS* but also in that of *IGF2* [18]. Although the reasons for the correlated gene expression within this open chromatin domain in islets from different donors are unclear, it seems plausible that the presence of MAZ, PDX1 and other nuclear factors at the *INS* promoter region of islet cells may have regulatory effects that extend over considerable distances within and beyond the domain (see below).

The islet-specificity of such an open chromatin domain at the *INS* locus strongly suggests a connection between the chromatin state and the unique propensity of islet β -cells to express high levels of insulin. In fact, in an attempt to find a mechanistic explanation for the latter,

efforts had been made to identify β -cell-specific transcription factors that bind to the *INS* promoter (reviewed in Ref. [19]). However, PDX-1, MafA and NeuroD – the three main β -cell factors activating *INS* transcription – were all found to be dispensable and replaceable by more ubiquitous transcription factors. The fact that only islet cells present a chromatin conformation favourable for gene expression to occur at the *INS* locus thus suggests that the physical state of the chromatin plays a key role in ensuring that *INS* is expressed at high levels only in islet cells. Understanding how such an open chromatin domain is established and maintained, and identifying the factors and DNA elements critical to ensure *INS* expression, should provide critical insights in terms of diabetes treatment and prevention.

Recent Technological Advances in the Study of the 3D Structure of Chromatin

The discovery of long-range interactions within the nucleus, and of their important regulatory role, has come as the result of the development of powerful new methods for detecting such interactions. A set of molecular techniques has recently been developed to investigate the nuclear topology of chromatin. Of necessity, chromatin is not linear in the nucleus, but compacted to fit into the small nuclear volume. However, packing is not random; rather, it is dynamically arranged into functional loops. These chromatin loops are usually established in a cell type-specific manner, through the formation of long-range chromatin interactions mediated by proteins bound at specific DNA regulatory elements (e.g. enhancers, promoters, insulators). The chromosome conformation capture (3C) technique and its derived genomics methods (figure 5) allow a snapshot to be obtained of the nuclear organization of the chromatin, by identifying the regions of the genome that are physically in close contact due to the formation of chromatin loops.

The 3C technique (figure 5A) was originally developed by Dekker and colleagues to determine whether two given regions of the genome interact [20]. For this purpose, the chromatin is chemically fixed (crosslinked) in a given conformation and subsequently digested by a restriction enzyme that cuts the DNA of the genome at sites of defined sequence, typically to produce small (3–4 kb) fragments. These fragments are then ligated with DNA ligase, the prior fixation ensuring that regions of the genome that are in proximity are preferentially ligated in comparison to those that are more distant. Crosslinks are then removed and PCR amplification is carried out using primers aligning to sequences within each surveyed region of the genome. Based on the idea that regions engaged in long-range interactions should be detected at a higher frequency than those that are not, the comparison of their actual and expected frequencies of contacts (based on their genomic distance on the linear chromatin) thus provides information about how physically close these two regions are *in vivo*. Importantly, such interactions probably do not establish a ‘frozen’ set of loop conformations; contacts are likely only transient and continuously broken and formed again, meaning that in many cases, only a small fraction of sites are in contact at any given moment. This may be sufficient, however, to establish long-term patterns of histone modifications and gene expression, which would have a much longer lifetime.

To investigate chromatin contacts on a larger scale, several techniques were later derived from the 3C technology, based on the same basic principle of fixing, digesting, ligating and identifying the ligation partners (figure 5B). Two distinct methods, 3C-on-chip and circular 3C (commonly referred to as 4C) both allow probing one genomic site against several others at once, the former relying on microarrays (known sites) and the latter, on cloning and sequencing (known and unknown sites) to identify the variable half of the ligation product [21,22]. Nowadays, however, both techniques have evolved to 4C-seq by making use of next-generation sequencing (NGS) to survey the entire genome at once [21,23].

Developed around the same time as 4C, the 3C carbon-copy (5C) method involves a slightly different modification of the initial 3C technology [24]. By adding a multiplexed amplification step at the end of the original 3C protocol and using microarrays or NGS to identify the carbon-copy amplified ligation products, multiple ‘anchor’ and ‘bait’ sites can be surveyed simultaneously. Unlike the 3C and 4C methods, 5C can be used to determine 3D chromatin structure in a single experiment. However, since primers need to be designed in every chromatin fragment tested, this technique can only be used to study relatively small regions (up to several megabases) of the genome. To address this limitation and allow a ‘genome versus genome’ approach, the Hi-C technique was recently developed [25]. Its major improvement lies in the introduction of a step for biotin labelling of the ligation product, which eliminates the need to design fragment-specific PCR primers and allows the ligation products to be selectively purified before being sequenced. In both cases, mathematical modelling of the generated matrix of interactions is subsequently required to establish the locus-specific or genome-wide 3D structure of chromatin.

Finally, the ChIP-loop method and its genome-wide application – ChIA-PET, solve the topology of chromatin from a different angle [26,27]. By combining chromatin immunoprecipitation with 3C or its derived methods, these techniques first focus on a chromatin-bound protein, and select in an additional step only interacting sites on DNA that involve loci occupied by this protein.

Long-range Gene Regulation at the *INS* Gene Locus by Chromatin Interactions

In the case of the *INS* locus, given what we already know about the chromatin structure, 3C-related technologies such as 4C, 5C, Hi-C or ChIA-PET are likely to provide important insights about the higher order structure of the open chromatin domain and how it integrates with the rest of the genome. Although its proximal promoter is sufficient to direct high-level and β -cell-specific *INS* expression in transgenic mice [28,29], several lines of evidence have also suggested the existence of long-range gene regulation at the insulin gene locus: first, *Ins2* (the mouse ortholog of *INS*) gene expression is positively regulated in yolk sac by the distal enhancers located over 100 kb away [30,31]; second, genetic studies have shown that allelic interaction at the insulin loci influences type 1 diabetes (T1D) susceptibility in humans [32] and regulates transcription of the *Ins2* gene in mice [33]. This phenomenon is reminiscent of paramutations originally found in plants, and similar allelic interactions have

also been reported later in animals [34]. The molecular bases for such interactions are still unknown.

Ten years of studies of spatial nuclear architecture using 3C and 3C-related technologies of the kind described above have revealed that direct physical contact is an important mechanism by which regulatory DNA elements influence transcription of a distal gene. As a continuation of our earlier studies, we were interested in determining the higher order chromatin structure at the human *INS* locus and its potential regulatory effects on β -cell gene expression. To that end, we carried out 4C experiments to map *INS* interactions genome wide in intact human pancreatic islets [23]. A site within the *INS* promoter was chosen as the bait (anchor); the 4C analysis revealed multiple contacts, mainly between this site and a 1-Mb region surrounding the *INS* gene on human chromosome 11 (figure 6A, B). The DNA sequence and gene organization in this gene-dense region are evolutionarily conserved [35]. Within this region, we identified genes interacting with *INS*, such as *IGF2* and *CDKN1C*, that are known to be co-expressed with the *INS* gene in human islet β -cells [36,37] and to be important for β -cell proliferation and survival. As mentioned above, overexpression of an engineered ZFP, targeted to the VNTR region located 365 bp upstream of the *INS* transcription start site, activates both *INS* and *IGF2* gene transcription in HEK-293 cells [18]. This work thus suggests a potential role for the *INS* gene locus in regulating transcription of a second gene *in vitro*.

These observations prompted us to investigate the regulatory effects of the *INS* gene locus on *in vivo* gene expression in human pancreatic islets [23]. Among the *INS* interactions uncovered by our 4C studies, we chose to study further the interaction of *INS* with *Synaptotagmin 8* (*SYT8*), in part because sites occupied by CTCF are present adjacent to both genes, and might in principle help stabilize long-range interactions between them (figure 6C). Furthermore, the *SYT8* homologues *Syt7* and *Syt9* have been shown to regulate insulin secretion in mouse and rat islets [38,39]. We used the quantitative 3C method to determine the ‘strength’ of the *INS*/*SYT8* interaction. Measurements revealed that in human islets, both *SYT8* expression (figure 7A) and its interaction with *INS* were elevated by glucose. The activity of the *INS* promoter was important for transactivation of *SYT8*: *INS* promoter-targeting siRNA markedly decreased *SYT8* gene expression (figure 7B). Consistent with a role in stabilizing the interaction between the two genes, depletion of CTCF resulted in decreased contact between them, and more significantly it attenuated *SYT8* expression, without affecting *INS* expression (figure 8A). Finally, we asked how *SYT8* functions in β -cells. As shown in figure 8B, RNAi-mediated depletion of *SYT8* markedly reduced insulin secretion [23].

Taken together, these results establish the existence of a physical regulatory network in which the *INS* promoter acts from a distance to stimulate *SYT8* expression, most likely through a direct, CTCF-mediated physical contact between the promoters of the two genes, which are separated by over 300 kb of DNA (figure 8C). The extent of contact and stimulation depends on the level of glucose or the activity of the *INS* promoter. The regulatory loop is closed by the action of *SYT8*, by mechanisms not yet determined, in stimulating insulin secretion. The *INS* promoter appears to link *INS* gene transcription to insulin secretion through regulation of *SYT8* gene expression in human islets [23]. Based on

these results, we speculate that food intake that increases glucose levels could stimulate *SYT8* gene transcription in islets *in vivo*. As the SYT8 protein has been reported to be lost after acrosomal exocytosis in sperm cells, the cytosolic SYT8 pool might be similarly depleted after insulin exocytosis. Glucose-stimulated *SYT8* expression could help to refill the depleted cytosolic SYT8 pool for the next round of insulin secretion. We propose that, through an effect on *SYT8* gene expression, the *INS* promoter may act to coordinate insulin transcription and secretion in human pancreatic islets. This model predicts that a severe defect in insulin transcription or physical separation of *INS* from *SYT8* and other nearby genes could lead to deregulation of insulin secretion. In support, inactivating mutations in the *INS* promoter regulator genes *PDX1* [40], *HNF1A* [41], *HNF1B* [42], *HNF4A* [43] and *NEUROD1* [40] all lead to defective insulin secretion observed in patients with the major monogenic form of diabetes, maturity-onset diabetes of the young (MODY) [44]. Impaired insulin secretion and diabetes have also been reported in an individual with a balanced chromosomal translocation t(1; 11)(p36.22; p15.5) [45], as well as in transgenic mice expressing the human *INS* gene [28,29]. Transcription of some of the other *INS*-interacting genes, like *Troponin I*, *fast skeletal muscle (TNNI2)* [23], and probably other genes as well, is also regulated in human islets by a mechanism similar to that regulating *SYT8* [23]. Thus, glucose-regulated interactions with the *INS* gene might be an important general molecular mechanism for the regulation of β -cell gene expression and function in human pancreatic islets.

These results also suggest that *INS* expression may be reciprocally modulated by its interactions with other genes or regulatory elements. The distal *H19* enhancer is known to enhance mouse *Ins2* gene transcription in the yolk sac through long-range chromatin interaction [30,31,46]. However, distal enhancers or other regulatory elements that can stimulate *INS* gene transcription in pancreatic islets are yet to be identified. Because of the physiological significance, it seems warranted to determine whether one or more regulatory elements within these newly identified *INS*-interacting loci could stimulate *INS* gene transcription in human islets.

Long-range Gene Regulation by Long Non-coding RNA

Long non-coding RNA (lncRNA) has also been shown to be important for long-range gene expression in diverse developmental programs [17]. LncRNAs have been found to exert long-range effects on either gene silencing or gene activation [17,47–49]. The lncRNA *KCNQ1OT1*, for example, is transcribed from a promoter overlapping with a differentially methylated region (DMR) located in an intron of the *KCNQ1* gene located about 500 kb away from the *INS* gene [50]. This paternally expressed lncRNA directly recruits the chromatin silencing complexes that spread *in cis* and bidirectionally along the chromosome; and by doing so, it silences the paternally imprinted genes such as *KCNQ1* and *CDKN1C* at the *KCNQ1* imprinted domain [51,52]. More recently, Jones and colleagues have shown that in mouse embryos, *Kcnq1ot1*-mediated gene silencing extends into the *Ins2* gene locus and contributes to the silencing of an *Ins2* transcript variant [53]. It is unknown if similar regulation exists in human pancreatic islets.

Genetic studies have suggested that lncRNAs may play a role in the regulation of β -cell function. First, increased dosage of the paternally expressed lncRNA *HYMAI* and its neighbouring gene *PLAGL1* secondary to a paternal chromosome duplication or imprinting defect causes a deficiency in insulin secretion in early infancy and leads to transient neonatal diabetes mellitus (TNDM) [54]. Second, partial loss of DNA methylation at the maternal DMR of the *KCNQ1OT1* gene, which presumably leads to activation of the maternal *KCNQ1OT1* allele, was observed in some patients with TNDM [55 – 57]. Third, lncRNA ANRIL, which is transcribed from a type 2 diabetes (T2D) associated locus at 9p21 [58], has been shown to regulate expression of the distal *Cdkn2a* and *Cdkn2b* genes [59], both genes known to be important for β -cell growth and apoptosis. Finally, a number of ncRNAs are transcribed from or near the *INS* locus. As mentioned previously, the levels of ncRNAs transcribed from the *INS* promoter region in human islets are positively correlated with *INS* mRNA levels [15]. Studies of the *Kit* locus in mice have demonstrated that RNA components play a fundamental role in the establishment of paramutations [60]. It remains to be determined whether this or other ncRNAs transcribed from the insulin gene locus might regulate the allelic interactions at the insulin gene loci observed in mice [33] and in humans as well [32].

Conclusions

We have been accustomed to studying regulation of gene expression by searching for regulatory elements either at the promoter immediately adjacent to the gene or somewhat further away at an enhancer. Recent results in many laboratories show that this is an oversimplification: regulatory elements situated in distant regions of the genome, not necessarily even on the same chromosome, can make physical contact with a target gene and significantly affect its expression. In the case of the human *INS* gene, we have found that the region near its promoter makes numerous physical contacts with other sites on chromosome 11 located hundreds of kilobases away. We have shown that the expression of two of these genes, *SYT8* and *TNNI2*, as well as their strength of contact with the *INS* locus, are affected by the levels of glucose or *INS* gene transcription. It seems possible that many other of these sites are near genes whose regulation is affected by the activity of the *INS* promoter. Similarly, a more recent genome-wide chromatin interaction study using ChIA-PET also showed that physical interactions of active promoters are widespread and some of these promoter interactions could influence each other's transcriptional activity in human cells [61].

Equally important, we were able to show that *SYT8* plays a role in secretion of insulin, establishing a physical feedback loop in which physical contacts between *INS* and *SYT8*, responding to glucose levels, can affect expression of *SYT8* and increase secretion of insulin. We think it likely that many of the other genes contacted by *INS* will turn out to be involved in regulatory processes associated with insulin biology. It is of particular interest that two major T2D susceptibility loci on chromosome 11 lie within regions that physically interact with the *INS* promoter in human islets [62,63].

Altogether, the molecular techniques described here may provide important clues as to the topological changes occurring within the *INS* open chromatin domain in health and disease.

For example, polymorphisms in the *INS* VNTR region have been associated with diabetes (for a review, see Ref. [64]). Class I repeats (30–60 repeats) seem to predispose to T1D, but could protect against T2D. In contrast, class III repeats (120–170 repeats) appear protective against T1D, but predispose to T2D, at least in Caucasians. Both types of repeats have also been associated with changes in *INS* expression, albeit in opposite directions, and only when paternally inherited. Using 3C or its derived technologies, one could assess how changes in distance between the *TH* and *INS* genes and/or in the secondary structures formed by the repeats influence local chromatin organization. Such changes could affect *INS* and other neighbouring genes' expression in human islets, as well as in other tissues such as the thymus, at different stages of development. An analysis of this kind could also provide some mechanistic explanations regarding the parent-of-origin effect, and its possible link to the initial imprinting of the locus.

More generally, these methods suggest ways in which to approach the problem of determining functions of risk SNP-containing regions that are located at considerable distances from coding regions. Often these regions contain local sites marked by histone modifications characteristic of an enhancer or a promoter, but there are no obvious gene targets nearby. As described above, there are at least two possibilities: either such a site is involved in long range contacts with a distant gene, or it is acting as a promoter or enhancer for expression of an ncRNA, which in turn may function to silence or activate a distant gene. Methods now exist to address both these possibilities, and can be applied to problems of β -cell function. Ultimately, gaining a better understanding of the molecular mechanisms at play behind T1D/T2D risks is likely to impact both preventive diagnosis and counseling and perhaps, in the longer term, treatment options for these two diseases as well.

Acknowledgements

This work was supported by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases, NIH.

References

1. Voss TC, Schiltz RL, Sung MH et al. Dynamic exchange at regulatory elements during chromatin remodeling underlies assisted loading mechanism. *Cell* 2011; 146: 544–554. [PubMed: 21835447]
2. Heintzman ND, Stuart RK, Hon G et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 2007; 39: 311–318. [PubMed: 17277777]
3. Talbot D, Collis P, Antoniou M et al. A dominant control region from the human beta-globin locus conferring integration site-independent gene expression. *Nature* 1989; 338: 352–355. [PubMed: 2922063]
4. Forrester WC, Takegawa S, Papayannopoulou T, Stamatoyannopoulos G, Groudine M. Evidence for a locus activation region: the formation of developmentally stable hypersensitive sites in globin-expressing hybrids. *Nucleic Acids Res* 1987; 15: 10159–10177. [PubMed: 3480506]
5. Dillon N, Trimborn T, Strouboulis J, Fraser P, Grosveld F. The effect of distance on long-range chromatin interactions. *Mol Cell* 1997; 1: 131–139. [PubMed: 9659910]
6. Felsenfeld G, Burgess-Beusse B, Farrell C et al. Chromatin boundaries and chromatin domains. *Cold Spring Harb Symp Quant Biol* 2004; 69: 245–250. [PubMed: 16117655]
7. Chung JH, Whiteley M, Felsenfeld G. A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in *Drosophila*. *Cell* 1993; 74: 505–514. [PubMed: 8348617]

8. 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]
9. Phillips JE, Corces VG. Ctfc: master weaver of the genome. *Cell* 2009; 137: 1194–1211. [PubMed: 19563753]
10. Yusufzai TM, Tagami H, Nakatani Y, Felsenfeld G. CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Mol Cell* 2004; 13: 291–298. [PubMed: 14759373]
11. Bell AC, Felsenfeld G. Methylation of a CTCF-dependent boundary controls imprinted expression of the *igf2* gene. *Nature* 2000; 405: 482–485. [PubMed: 10839546]
12. Kanduri C, Pant V, Loukinov D et al. Functional association of CTCF with the insulator upstream of the *h19* gene is parent of origin-specific and methylation-sensitive. *Curr Biol* 2000; 10: 853–856. [PubMed: 10899010]
13. Kurukuti S, Tiwari VK, Tavoosidana G et al. CTCF binding at the *h19* imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to *igf2*. *Proc Natl Acad Sci USA* 2006; 103: 10684–10689. [PubMed: 16815976]
14. Monk D, Sanches R, Arnaud P et al. Imprinting of *IGF2* p0 transcript and novel alternatively spliced *INS-IGF2* isoforms show differences between mouse and human. *Hum Mol Genet* 2006; 15: 1259–1269. [PubMed: 16531418]
15. Mutskov V, Felsenfeld G. The human insulin gene is part of a large open chromatin domain specific for human islets. *Proc Natl Acad Sci USA* 2009; 106: 17419–17424. [PubMed: 19805079]
16. Stitzel ML, Sethupathy P, Pearson DS et al. Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci. *Cell Metab* 2010; 12: 443–455. [PubMed: 21035756]
17. Wapinski O, Chang HY. Long noncoding RNAs and human disease. *Trends Cell Biol* 2011; 21: 354–361. [PubMed: 21550244]
18. Ferguson LA, Docherty HM, MacKenzie AE, Docherty K. An engineered zinc finger protein reveals a role for the insulin VNTR in the regulation of the insulin and adjacent *IGF2* genes. *FEBS Lett* 2009; 583: 3181–3186. [PubMed: 19733567]
19. Ohneda K, Ee H, German M. Regulation of insulin gene transcription. *Semin Cell Dev Biol* 2000; 11: 227–233. [PubMed: 10966856]
20. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science* 2002; 295: 1306–1311. [PubMed: 11847345]
21. Simonis M, Klous P, Splinter E et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4c). *Nat Genet* 2006; 38: 1348–1354. [PubMed: 17033623]
22. Zhao Z, Tavoosidana G, Sjolinder M et al. Circular chromosome conformation capture (4c) uncovers extensive networks of epigenetically regulated intra-and interchromosomal interactions. *Nat Genet* 2006; 38: 1341–1347. [PubMed: 17033624]
23. Xu Z, Wei G, Chepelev I, Zhao K, Felsenfeld G. Mapping of ins promoter interactions reveals its role in long-range regulation of *sy8* transcription. *Nat Struct Mol Biol* 2011; 18: 372–378. [PubMed: 21336277]
24. Dostie J, Richmond TA, Arnaout RA et al. Chromosome conformation capture carbon copy (5c): a massively parallel solution for mapping interactions between genomic elements. *Genome Res* 2006; 16: 1299–1309. [PubMed: 16954542]
25. Lieberman-Aiden E, van Berkum NL, Williams L et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 2009; 326: 289–293. [PubMed: 19815776]
26. Horike S, Cai S, Miyano M, Cheng JF, Kohwi-Shigematsu T. Loss of silent-chromatin looping and impaired imprinting of *dlx5* in rett syndrome. *Nat Genet* 2005; 37: 31–40. [PubMed: 15608638]
27. Fullwood MJ, Liu MH, Pan YF et al. An oestrogen-receptor- α -bound human chromatin interactome. *Nature* 2009; 462: 58–64. [PubMed: 19890323]
28. Marban SL, DeLoia JA, Gearhart JD. Hyperinsulinemia in transgenic mice carrying multiple copies of the human insulin gene. *Dev Genet* 1989; 10: 356–364. [PubMed: 2689020]

29. Karaca M, Durel B, Languille L et al. Transgenic expression of human ins gene in ins1/ins2 double knockout mice leads to insulin underproduction and diabetes in some male mice. *Front Biosci* 2007; 12: 1586–1593. [PubMed: 17127405]
30. Leighton PA, Saam JR, Ingram RS, Stewart CL, Tilghman SM. An enhancer deletion affects both h19 and IGF2 expression. *Genes Dev* 1995; 9: 2079–2089. [PubMed: 7544754]
31. Leighton PA, Ingram RS, Eggenschwiler J, Efstratiadis A, Tilghman SM. Disruption of imprinting caused by deletion of the h19 gene region in mice. *Nature* 1995; 375: 34–39. [PubMed: 7536897]
32. Bennett ST, Wilson AJ, Esposito L et al. Insulin VNTR allele-specific effect in type 1 diabetes depends on identity of untransmitted paternal allele. The IMDIAB Group. *Nat Genet* 1997; 17: 350–352. [PubMed: 9354805]
33. Duvillie B, Bucchini D, Tang T, Jami J, Paldi A. Imprinting at the mouse ins2 locus: evidence for cis-and trans-allelic interactions. *Genomics* 1998; 47: 52–57. [PubMed: 9465295]
34. Rassoulzadegan M, Grandjean V, Gounon P et al. RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature* 2006; 441: 469–474. [PubMed: 16724059]
35. Onyango P, Miller W, Lehoczyk J et al. Sequence and comparative analysis of the mouse 1-megabase region orthologous to the human 11p15 imprinted domain. *Genome Res* 2000; 10: 1697–1710. [PubMed: 11076855]
36. Kassem SA, Ariel I, Thornton PS et al. P57(kip2) expression in normal islet cells and in hyperinsulinism of infancy. *Diabetes* 2001; 50: 2763–2769. [PubMed: 11723059]
37. Sempoux C, Guiot Y, Dahan K et al. The focal form of persistent hyperinsulinemic hypoglycemia of infancy: morphological and molecular studies show structural and functional differences with insulinoma. *Diabetes* 2003; 52: 784–794. [PubMed: 12606521]
38. Gustavsson N, Lao Y, Maximov A et al. Impaired insulin secretion and glucose intolerance in synaptotagmin-7 null mutant mice. *Proc Natl Acad Sci USA* 2008; 105: 3992–3997. [PubMed: 18308938]
39. Iezzi M, Eliasson L, Fukuda M, Wollheim CB. Adenovirus-mediated silencing of synaptotagmin 9 inhibits Ca^{2+} -dependent insulin secretion in islets. *FEBS Lett* 2005; 579: 5241–5246. [PubMed: 16165130]
40. Babu DA, Chakrabarti SK, Garmey JC, Mirmira RG. Pdx1 and beta2/neurod1 participate in a transcriptional complex that mediates short-range DNA looping at the insulin gene. *J Biol Chem* 2008; 283: 8164–8172. [PubMed: 18252719]
41. Emens LA, Landers DW, Moss LG. Hepatocyte nuclear factor 1 alpha is expressed in a hamster insulinoma line and transactivates the rat insulin i gene. *Proc Natl Acad Sci USA* 1992; 89: 7300–7304. [PubMed: 1380153]
42. Kitanaka S, Sato U, Igarashi T. Regulation of human insulin, IGF-i, and multidrug resistance protein 2 promoter activity by hepatocyte nuclear factor (hnf)-1beta and hnf-1alpha and the abnormality of HNF-1beta mutants. *J Endocrinol* 2007; 192: 141–147. [PubMed: 17210751]
43. Bartoov-Shifman R, Hertz R, Wang H et al. Activation of the insulin gene promoter through a direct effect of hepatocyte nuclear factor 4 alpha. *J Biol Chem* 2002; 277: 25914–25919. [PubMed: 11994285]
44. O’Rahilly S Human genetics illuminates the paths to metabolic disease. *Nature* 2009; 462: 307–314. [PubMed: 19924209]
45. Murphy R, Baptista J, Holly J et al. Severe intrauterine growth retardation and atypical diabetes associated with a translocation breakpoint disrupting regulation of the insulin-like growth factor 2 gene. *J Clin Endocrinol Metab* 2008; 93: 4373–4380. [PubMed: 18728168]
46. Vu TH, Nguyen AH, Hoffman AR. Loss of IGF2 imprinting is associated with abrogation of long-range intrachromosomal interactions in human cancer cells. *Hum Mol Genet* 2010; 19: 901–919. [PubMed: 20015958]
47. Guttman M, Rinn JL. Modular regulatory principles of large non-coding RNAs. *Nature* 2010; 482: 339–346.
48. Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. *Mol Cell* 2011; 43: 904–914. [PubMed: 21925379]
49. Orom UA, Shiekhattar R. Noncoding RNAs and enhancers: complications of a long-distance relationship. *Trends Genet* 2011; 27: 433–439. [PubMed: 21831473]

50. Kanduri C Kcnq1ot1: a chromatin regulatory RNA. *Semin Cell Dev Biol* 2011; 22: 343–350. [PubMed: 21345374]
51. Mohammad F, Mondal T, Guseva N, Pandey GK, Kanduri C. Kcnq1ot1 noncoding RNA mediates transcriptional gene silencing by interacting with Dnmt1. *Development* 2010; 137: 2493–2499. [PubMed: 20573698]
52. Pandey RR, Mondal T, Mohammad F et al. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell* 2008; 32: 232–246. [PubMed: 18951091]
53. Jones MJ, Bogutz AB, Lefebvre L. An extended domain of kcnq1ot1 silencing revealed by an imprinted fluorescent reporter. *Mol Cell Biol* 2011; 31: 2827–2837. [PubMed: 21576366]
54. Temple IK, Shield JP. 6q24 transient neonatal diabetes. *Rev Endocr Metab Disord* 2010; 11: 199–204. [PubMed: 20922569]
55. Mackay DJ, Boonen SE, Clayton-Smith J et al. A maternal hypomethylation syndrome presenting as transient neonatal diabetes mellitus. *Hum Genet* 2006; 120: 262–269. [PubMed: 16816970]
56. Arima T, Kamikihara T, Hayashida T et al. Zac, lit1 (kcnq1ot1) and p57kip2 (cdkn1c) are in an imprinted gene network that may play a role in Beckwith-Wiedemann syndrome. *Nucleic Acids Res* 2005; 33: 2650–2660. [PubMed: 15888726]
57. Mackay DJ, Hahnemann JM, Boonen SE et al. Epimutation of the TNDM locus and the bekwit-wiedemann syndrome centromeric locus in individuals with transient neonatal diabetes mellitus. *Hum Genet* 2006; 119: 179–184. [PubMed: 16402210]
58. Scott LJ, Mohlke KL, Bonnycastle LL et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 2007; 316: 1341–1345. [PubMed: 17463248]
59. Visel A, Zhu Y, May D et al. Targeted deletion of the 9p21 non-coding coronary artery disease risk interval in mice. *Nature* 2010; 464: 409–412. [PubMed: 20173736]
60. Tomikawa J, Shimokawa H, Uesaka M et al. Single-stranded noncoding rnas mediate local epigenetic alterations at gene promoters in rat cell lines. *J Biol Chem* 2011; 286: 34788–34799. [PubMed: 21844201]
61. Li G, Ruan X, Auerbach RK et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* 2012; 148: 84–98. [PubMed: 22265404]
62. Kong A, Steinthorsdottir V, Masson G et al. Parental origin of sequence variants associated with complex diseases. *Nature* 2009; 462: 868–874. [PubMed: 20016592]
63. Yasuda K, Miyake K, Horikawa Y et al. Variants in kcnq1 are associated with susceptibility to type 2 diabetes mellitus. *Nat Genet* 2008; 40: 1092–1097. [PubMed: 18711367]
64. Pugliese A, Miceli D. The insulin gene in diabetes. *Diabetes Metab Res Rev* 2002; 18: 13–25. [PubMed: 11921414]

Figure 1.

Mode of insulator action. (A) One kind of insulator (I) can block the propagation of a heterochromatic, silencing domain into a region containing transcriptionally active genes. (B) Other ‘enhancer blocking’ insulator elements can prevent the action of an enhancer (Enh1) in gene system 1 from activating a promoter (Pro2) in gene system 2, without inhibiting the ability of Enh1 to activate Pro1. (C) A test for the enhancer blocking activity of CTCF (I) is to place it between an enhancer and promoter, which blocks enhancer action (top) or outside the enhancer-promoter domain, where it has no inhibitory action (bottom).

Figure 2.

CTCF helps bring together distant elements in the genome, with varying consequences. (A) When this results in exclusion of the enhancer (E) from the loop containing the promoter (P), silencing can result. (B) When it results in bringing enhancer and promoter closer together, the gene may be activated. (C) The paternally transmitted allele for *Igf2* is silenced in large part because CTCF molecules bound to the imprinted control region (ICR) create an enhancer-blocking element that prevents downstream enhancers from activating the gene. (D) On the maternal allele the DNA of the ICR is methylated, inhibiting CTCF binding and inactivating the insulator, allowing the enhancers to activate *Igf2* [11,12].

Figure 3.
Map of histone modifications and DNA methylation across the open chromatin domain at the human insulin locus.

Figure 4.

Chromatin structure and gene expression in the human insulin locus. (A) Quantitative RT-PCR measurements of genes expressed in human islets, in samples from six donors. (B) Correlation between levels of *PDX-1* RNA (*x*-axis) and *INS*, *TH*, *IGF2AS*, *IGF2* and *Actin* (control) RNA (*y*-axis), showing the linear dependence of the genes within the *INS* domain on *PDX1* expression [15].

Figure 5. Summary of 3C and its derived technologies. (A) Molecular principles of the 3C methodology. (B) Diagrams presenting the possibilities of each 3C or derived technique depending on the cellular context.

Figure 6.

4C study of the human *INS* locus in human pancreatic islets [23]. (A) Points of contact within the nucleus between a site within the *INS* promoter and various sites on human chromosome 11. (B) Higher resolution map of contacts near *INS*. (C) 4C contacts near *SYT8*, compared with known CTCF binding sites.

Figure 7. Glucose and the *INS* promoter positively regulate *SYT8* gene expression in human pancreatic islets. (A) Specific effect on *SYT8* expression of treating pancreatic islets with glucose for three donors [23]. (B) Effect of siRNA targeting of the *INS* promoter, which silences *INS* gene transcription, on *SYT8* and *TNNT2* expression [23].

Figure 8.

SYT8 gene expression is regulated by *CTCF* and is important for insulin secretion in human pancreatic islets. (A) Effect of siRNA-induced downregulation of *CTCF* expression on *SYT8* expression in islets from two donors. *INS* expression as well as *SYT7* is unaffected. (B) Effect of siRNA-induced downregulation of *SYT8* expression on *INS* secretion following glucose addition. (C) Schematic representation of the effect of glucose on *INS*–*SYT8* contact and *SYT8* expression (Adapted with permission from Ref. [23]).