Published in final edited form as:

Trends Immunol. 2015 September ; 36(9): 519–526. doi:10.1016/j.it.2015.07.005.

Super-enhancers: Asset management in immune cell genomes

Steven Witte¹, John J. O'Shea¹, and Golnaz Vahedi^{2,3,*}

¹Lymphocyte Cell Biology Section, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, Maryland, 20892

²Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

³Institute for Immunology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

Abstract

Super-enhancers (SEs) are regions of the genome consisting of clusters of regulatory elements bound with very high amounts of transcription factors, and this architecture appears to be the hallmark of genes and noncoding RNAs linked with cell identity. Recent studies have identified SEs in CD4⁺ T cells and have further linked these regions to SNPs associated with immune-mediated disorders, pointing to an important role for these structures in the T cell differentiation and function. Here we review the features that define SEs, and discuss their function within the broader understanding of the mechanisms that define immune cell identity and function. We propose that SEs present crucial regulatory hubs, coordinating intrinsic and extrinsic differentiation signals, and argue that delineating these regions will provide important insight into the factors and mechanisms that define immune cell identity.

Genomic control of immune genes

The immune system consists of a diverse collection of cells that defend the host from pathogens and that participate in processes important for organismal homeostasis. Like other cells, the development of immune cells represents the integration of intrinsic signals provided by lineage-defining transcription factors (LDTFs) and environmental signals that act upon signal-dependent transcription factors (SDTFs), both of which work on the genomes of precursor cells to affect development and differentiation. Much effort has been devoted to understanding what this means at a molecular level (Box 1). Indeed, we have learned a great deal about the key transcription factors and soluble factors (e.g. cytokines) that regulate immune cell fate determination. Recent findings have ushered in a better understanding of how these factors modify the genome, and how these modifications are interpreted to guide immune cell development (reviewed in [1]). Often, intrinsic and

*Correspondence should be addressed to Golnaz Vahedi, mailto:vahedi@mail.med.upenn.edu, Phone: 215-898-8439.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

extrinsic signals work through transcription factors (TFs) to activate enhancer sites, which in turn control cell-specific gene expression.

Recent findings have identified a subclass of enhancers that have been termed super-enhancers (SEs) based on features related to the amount and nature of TFs associated with these loci as well, as associated chromatin modifications and genomic architecture [2-6]. These genomic structures have also been identified in CD4+ cells [7,8], raising the question of how they relate to our understanding of the mechanisms that guide immune cell differentiation and function, both in terms of responses to intrinsic signals and signals from the environment. This question is particularly relevant when considering that immune cells are plastic at a relatively mature state, when many features of the genomic landscape have already been defined by the action of LDTFs and SDTFs; naïve T cells response to TCR and cytokine signals to further differentiate into subsets with specific functions. What role do super-enhancers play in these processes, and how do these relate to the factors and mechanisms that have already shaped the immune cell genome?

Here, we discuss the concept of super-enhancers and review the features that define regions of the genome as such, outlining areas of consensus and points of uncertainty. We discuss recent studies that have examined the roles of enhancers and super-enhancers in T cell differentiation in light of the broader understanding of the mechanisms underlying immune cell differentiation, and in this context highlight gaps in understanding and important areas of future inquiry.

Finding cell-specific instructions embedded in the genome

Enhancers were first described in the 1980's as elements distal to genes, which, in contrast to promoters, affect transcription in an orientation independent manner [9]. Mammalian genomes encode regulatory elements in their linear sequence, but appropriate expression of their genes requires the formation of a loop between enhancers and promoters [10]. Delineating the exact location of enhancers in the mammalian genome has always been challenging. Until recently, the systematic characterization of enhancers almost exclusively relied on detection of sequences that are conserved across multiple species. Some of the earliest, and best characterized enhancers discovered based on sequence conservation are near the beta-globin and CD4 genes (CITE) [9,11,12]. Regulatory elements within cytokine genes such as *Il4* and *Ifng* have also been the focus of various studies [13-16]. Two regulatory elements for *Foxp3*, CNS1 and CNS2 play distinct roles in Tregs: the genomic region containing CNS1 is essential for peripheral Treg cells whereas CNS2 mediates stable inheritance of *Foxp3* expression and prevents their differentiation into inflammatory effector cells [17-19]. These striking findings emphasize the importance of noncoding genomic regions in regulating key genes with major roles in immune cells.

Pregenomic studies that relied on sequence conservation suggested an abundance of *cis*-regulatory sequences in non-coding DNA tightly regulating key cytokines and transcription factors. The recent genome-scale methods identifying genomic locations of putative enhancers complements the conservation-based approaches, as it is specific for cell type, condition, and activation states [20]. Noncoding functional elements are associated with

distinctive chromatin structures, which display signature patterns of histone modifications, DNase accessibility, and transcription factor occupancy [21]. Active enhancers are evidenced by histone acetylations and DNase-accessible chromatin and binding of sequence-specific transcription factors, coactivators such as EP300, and, often, RNA polymerase II. In contrast, poised enhancers are marked by histone methylation (H3K4me1) [22-25].

The genome-scale methodologies have also been applied to measure enhancer-promoter interactions in an unbiased manner. While chromosome conformation capture technology (3C) was used to define changes within the *Ifng* and beta-globin loci [26-30], organization of genomic loci are being explored at unprecedented resolution using newly developed genomic and computational approaches based on enhancements of chromosome conformation capture technology (such as 3C, 4C, 5C, ChIA-PET and Hi-C) [31-34]. Recent pioneering studies of mammalian chromosome structure have proposed that they are organized into a hierarchy of units, which include topologically associating domains (TADs) and gene loops: TADs are cell-type and species-invariant 1 Mb genomic regions of high local interaction frequency separated by non-interacting boundary elements [35,36]. In other words, there are portions of the genome that are highly interactive, which have boundaries that limit interactions outside this region.

The large-scale enhancer mapping relying on the chromatin signature of enhancers revealed unexpected complexity and dynamics of enhancer utilization patterns, with close to a million putative enhancers annotated by the ENCODE and NIH Epigenomics Roadmap projects in the human genome [22-25,37,38]. Studies in immune cells also added new dimensions to these international efforts. Activated B cells and embryonic stem cells both express *Myc*, but utilize distinct sets of enhancers in the *Myc* locus, suggesting a very different “logic” in regulation [33]. CD4⁺ T helper cell subsets (Th1, Th2, and Th17 cells) also employ very distinct sets of enhancers for their specialization. In fact, the extent of difference in enhancer elements is an order of magnitude larger than the number of differentially regulated genes [39-42]. To regulate a few hundred differentially transcribed genes between Th1 and Th2 differentiation, close to ten thousand subset-specific regulatory elements are differentially utilized across the genome [41].

Not only do enhancers vary greatly between related lineages, but the tissue microenvironment to which a cell is exposed can have profound effects on the enhancer repertoire. Two parallel studies profiled the expression and chromatin landscape of up to 9 populations of mouse macrophages isolated from distinct tissues [43,44]. Tissues included brain microglia, spleen red pulp macrophages, liver Kupffer cells, lung macrophages, large and small peritoneal macrophages, and colonic large and ileal small intestinal macrophages, in addition to thioglycollate-elicited peritoneal macrophages (TGEMs) and bone-marrow-derived macrophages (BMDMs). The regulatory elements of tissue-resident macrophages, including promoters, active enhancers, and poised enhancers were mapped using H3K27ac, H3K4me1 and ATAC-seq [45]. Comparison of these maps suggest that macrophages utilize distinct enhancer elements based on the tissue they reside in. These tissue-specific enhancers drive divergent programs of gene expression [43,44]. Some poised enhancers are shown to be inherited during development and thus provide ontogenic memory [43,44]. The highest percentages of poised enhancers are at regions present in only one or two macrophage

populations, demonstrating complex regulatory mechanisms, which can only be analyzed comprehensively on the level of the chromatin [44].

What sets up the enhancer landscape: intrinsic or extrinsic signals?

The discovery that enhancer landscapes are so distinctive raises the question as to what factors are responsible for creating these genomic elements. Studies in macrophages have argued that LDTFs exert a pervasive role in controlling the whole repertoire of active regulatory elements [46,47]. Approximately 80% of PU.1 binding sites were located distal to promoters and associated with p300 binding, the canonical chromatin signature of active enhancers [46]. Strikingly, most H3K4me1 regions had PU.1 binding in macrophages. The reconstitution of PU.1 expression in PU.1-deficient myeloid progenitors led to sequential nucleosome remodeling and increased H3K4me1 levels at PU.1 binding sites [47]. Moreover, PU.1 depletion results in the reduction of H3K4me1 amounts at several tested enhancers [46]. Collectively, the fact that PU.1 binding is necessary and sufficient to change the modification state of the neighboring chromatin has led to the conclusion that such factors actively shape the epigenome.

However, studies in T helper cells argue strongly against a simple model for the pervasive effect of LDTFs in shaping the enhancer landscape. LDTFs in T cells including T-bet, Gata3, Ror γ t, and Foxp3, also called *master regulators*, are both necessary and sufficient for differentiation of each corresponding T-cell subset. Multiple independent studies provided a new and very distinct perspective for the relative roles of various factors in T cell specialization. In Treg cells, Foxp3 deletion does not affect enhancers annotated by DNaseI hypersensitivity [40]. Using genomic DNaseI footprinting analysis across Tregs [23], it was shown that the pioneer factor Foxo1 seems to set the stage for Foxp3 binding [40]. The roles of T-bet and Ror γ t on the active enhancer landscape appear to be equally limited [39,41]. Enhancers including those in the *Ifng* locus were found to be T-bet-dependent. Overall though, deletion of T-bet and Ror γ t in Th1 and Th17 cells led to limited changes in enhancer landscapes annotated by binding of p300, suggesting different players working on the chromatin in these cell types. Rather, T-bet largely limits p300 recruitment on enhancers of opposite lineages as opposed to primarily establishing the enhancer elements of Th1 cells. Likewise, only a handful of loci including *Il17a*, *Il17f*, and *Il23r* are highly dependent on Ror γ t [39].

While LDTFs are undoubtedly important in regulating epigenomes, their action does not explain the *de novo* generation of enhancers in T cells. In this respect, there is another category of transcription factors – SDTFs – that play key roles on active enhancer landscapes. Earlier studies indicated a link between SDTFs and genomic enhancers. SDTFs can bind to open chromatin regions that are defined by combinations of LDTFs [48]. Smad3, mediating Transforming growth factor beta (TGF- β) signaling, co-occupies the genome with cell-type-specific master transcription factors: Oct4 in mouse embryonic stem cells (mESCs), Myod1 in myotubes, and PU.1 in pro-B cells.

The impact of SDTFs on the chromatin of T cells is not unexpected since a key aspect in the acquisition of distinct T helper cell phenotypes is the cytokine milieu. Upon encountering

diverse microbial pathogens, dendritic cells and other cells of the innate and adaptive immune system produce cytokines, which serve to instruct distinct T-cell fates. The major specifying cytokines exert their effect through signal transducer and activator of transcription (STAT) family transcription factors.

Multiple studies unraveled the pivotal role of STAT proteins in orchestrating the enhancer repertoire in immune cells [39-41,49]. Strikingly, the majority of differentially active enhancers in Th1 and Th2 cells were STAT4- or STAT6-dependent: specifically, in the absence of STAT4 and STAT6 proteins, the recruitment of p300 was disrupted at 50 and 75% of enhancers, respectively [41]. Based on ChIP-seq data, approximately 30% of these were direct targets of STATs. This was also the case in Th17 cells, in which STAT3 had a major role in p300 recruitment [39]. The interplay between SDTFs and LDTFs is clearly illustrated when the reconstitution of STAT4- and STAT6-deficient cells with LDTFs T-bet and GATA3 failed to recover the active enhancer landscapes annotated by p300 binding [41]. Mapping the chromatin signature of enhancers after the reconstitution of LDTFs in cells lacking SDTFs argue for a primary role of environmental sensors in dictating global landscapes. A similar study in macrophages also revealed that STAT1 and STAT6 play key roles in depositing H3K27Ac and H3K4me1 in response to IFNG and IL-4 [49]. These studies suggest a hierarchical model where STATs are the architects of epigenetic landscape and the lineage-specific transcription factors in directing global landscapes.

Collectively, enhancer utilization is distinct in every cell type, tissue and immune activation, and sensors of the environment thereby dramatically impact epigenomes in concert with intrinsic LDTFs. But what do these “big data” enhancer maps tell us about workings of the immune system? The key point is to remember what enhancers do. Enhancers control lineage identity by recruiting transcription factors. By performing motif and footprinting analyses on enhancer maps [37,50] and integrating those findings with transcriptome profiles, we can uncover the function of previously unrecognized factors and pathways in diverse types of immune cells.

Asymmetry in Genomic Distribution of Transcription Machinery

One of the insights obtained from genome-scale protein-DNA binding data (ChIP-seq) was the recognition that proteins within the transcriptional apparatus, such as Mediators, are not uniformly distributed across the genome [6]. Mediator is a transcriptional coactivator and has been shown to physically and functionally connect the enhancers and core promoters of active genes [51]. Recently, genomic regions associated with high density of Mediator or other proteins within transcription machinery have been given the term super-enhancers [3,5,6]. Super-enhancers were originally defined in mouse embryonic stem cells by first merging regions co-occupied by the “Yamanaka” factors Oct4, Sox2 and Nanog and then simply ranking the total background-normalized level of the Mediator signal at those co-occupied regions [6]. Less than 3% of these enhancers contained Mediator levels above a point where the slope of the plot was 1 [6]. These genomic domains representing the strongest Mediator-bound regions (above the slope) were thus dubbed ‘super-enhancers’ (SE), as compared with the remaining regions that were termed ‘typical’ enhancers (TE).

The relationship between super-enhancers and the structure of chromosomes in stem cells was assessed by mapping local chromosomal structure using chromatin interaction assay based on paired-end tag sequencing (ChIA-PET) [2,34]. Analysis of high-confidence cohesin ChIA-PET interaction data revealed a striking feature common to loci containing super-enhancers and their associated genes [2]. This feature consisted of a super-enhancer and its associated gene located within a loop connected by two interacting CTCF sites co-occupied by cohesin. The vast majority of ESC super-enhancers (84%) are shown to be contained within these structures. In contrast, only 48% of typical enhancers were found to occur within comparable loops between two CTCF sites. The interaction data in this study supported 83% of super-enhancer assignments to the proximal active gene with approximately half of the remainder assigned to the second most proximal gene. The loops associated with super-enhancers were also shown to be functionally linked to gene expression. CRISPR-mediated genome editing of CTCF sites at the boundaries of these loops results in the upregulation of genes immediately outside the super-enhancer domain. In this respect, CTCF seems to be functioning in its classic context as “insulating” surrounding regions from the effect of SEs [2].

Super-enhancer characterizations have gone beyond the embryonic stem cells. Binding of PU.1 in pro-B cells, MyoD in myotubes, T-bet in helper T cells and C/EBP α in macrophages was used to define these genomic regions [6]. The super-enhancer elements identified in ESCs, pro-B cells, myotubes, Th cells, and macrophages clearly spanned domains that were almost entirely cell-type specific, and the genes associated with these elements were highly cell-type specific relative to typical enhancer-associated genes [6]. In addition to LDTFs, various proteins within the transcription machinery and chromatin regulators have also been utilized to define super-enhancer regions. A catalogue of super-enhancers for 86 different human tissues and samples was created based on the genome-wide H3K27ac data [3]. Extending this definition to a broad range of human tissues emphasized the cell and tissue specificity of these genomic elements.

BRD4, a transcription coactivator containing two bromodomains that recognize acetylated lysine residues, was also found to co-occupy super-enhancers in multiple melanoma and diffuse large B cell lymphoma tumor cells [5,52]. Treatment of multiple melanoma tumor cells with the BET-bromodomain inhibitor JQ1 led to preferential loss of BRD4 at super-enhancers and consequent transcription elongation defects that preferentially impacted genes with super-enhancers [5]. For example, the IgH super-enhancer showed significantly greater reduction in BRD4 occupancy in cells treated with JQ1 than typical enhancer regions. Ultimately, virtually all BRD4 occupancy was lost at the IgH super-enhancer (97% reduction versus DMSO control) after treatment with 500 nM JQ1. It was postulated that super-enhancer associated genes confer disproportionate sensitivity to loss of the BRD4 coactivator and thus cause selective inhibition of transcription at oncogenes such as Myc [5,52].

Super-enhancers also underlie the identity, lineage commitment and plasticity of adult stem cell *in vivo* [53]. Using hair follicle as a model, SEs were mapped by H3K27ac modifications in hair follicle stem cells and their committed progenies in their native microenvironment. SEs were shown to be activated or silenced in a lineage-specific fashion.

Most SE genes involved in dictating stem cell fate were decommissioned in stem cell progeny by gaining H3K27me3, suggestive of ‘super-silencing’ [53]. Adam et al also rigorously addressed the functionality of SEs by testing the ability of SEs to drive reporter gene expression *in vivo*. By exploring the eGFP patterns in their reporter assays for hair follicle stem cells and their progenies, they provided compelling evidence for the functional importance of SEs *in vivo* [53].

Relying on an independent methodology, Collins and colleagues described genomic regions with similar properties and referred to them as ‘stretch enhancers’ [54]. This approach required the generation of the chromatin state maps of human pancreatic islets, defining regulatory elements such as promoters, enhancers, insulators, transcribed, and repressed regions based on distinct patterns of histone modifications including histone H3 lysine 27 acetylation (H3K27ac), H3K27 trimethylation (H3K27me3), H3K36me3, H3K4 monomethylation (H3K4me1), H3K4me3, and the CCCTC-binding factor (CTCF). Using these measurements in human pancreatic islets, they segmented the genome based on ChromHMM [55]. The idea behind ChromHMM is to define ‘chromatin states’ based on recurrent and spatially coherent combinations of chromatin marks, instead of linking individual modifications to the activity levels of various genetic functional elements [55]. Parker et al showed that certain parts of the human genome were continuously marked as “strong enhancers”, a chromatin state related to H3K27ac deposition. “Stretch-enhancers” were thus defined based on these enhancer marks that extended beyond a threshold, e.g. 3 kb. This study also demonstrated that locus-control-regions (LCR) regions are globally marked as stretch enhancer states in the relevant cell types exhibiting robust target gene expression [54]. In a recent review, Pott and Lieb [56] have noted that there appear to be an order of magnitude difference in the number of stretch enhancers versus super-enhancers, indicating substantial differences between the two features. However, the quantitation of these regions depends on the cut-offs defined in both methodologies, which to a certain degree are arbitrary. In this respect, the distinctions may be less sharp if different parameters are employed. This will be an important area to revisit.

An important discovery was the relationship between SEs and disease associated single-nucleotide polymorphisms (SNPs) (Box 2) [3,54]. SNPs associated with type 2 diabetes were shown to be enriched within islet specific stretch-enhancers [54]. This concept was also extended to a broad range of human diseases including autoimmune disorders and various cell types using the definition of super-enhancers [3,7,8]. For example, super-enhancers characterized in human brain tissue are enriched for SNPs associated with Alzheimer’s disease, in contrast variants associated with rheumatoid arthritis are enriched in super-enhancers identified in B and T cells [3].

Super-enhancers in immune cell development, differentiation and function

How does the concept of super-enhancers inform our understanding of immune cell development and function? Complex regulatory loci have been familiar to immunologists for a long time, notably the complex genomic landscape surrounding the *Ifng* and *Il4/Il13/Il15* loci [14,57]. We have known for some time that it is not possible to recapitulate the

true, tissue-specific expression of these cytokines using transgenic mice that expressed portions of these loci [58].

With genome-scale approaches, we can now identify SEs in immune cells under various conditions, and interrogate genomic regions with high concentration of factors relevant to transcription apparatus. In CD4⁺ Th1, Th2, and Th17 cells, cytokines such as *Ifng*, *Il4*, *Il2-Il21* and *Il10* and cytokine receptors such as *Il2ra* are enriched for the SE architecture [8]. Key transcription factors themselves including *Tbx21*, *Gata3*, *Stat4* and *Stat3* have SEs as well [8]. Interestingly, long noncoding RNAs (lncRNAs) preferentially exhibit the SE architecture and the SE feature distinguishes dynamically regulated ncRNAs from those that are constitutively expressed during T cell differentiation [8].

While many cytokine loci harbor SEs, the strongest SE in Th1, Th2 and Th17 cells with respect to the p300 recruitment belongs to a transcription factor called *Bach2* [8]. Earlier studies in *Bach2*-deficient mice have shown that this protein is a broad regulator of immune activation [59,60]. *BACH2* protein is required for efficient formation of Treg cells and consequently for suppression of lethal inflammation in a Treg-dependent manner [59]. Its absence during Treg polarization results in inappropriate diversion to effector lineages. In addition, *BACH2* constrains full effector differentiation within Th1, Th2 and Th17 cell lineages. It is therefore not surprising that the genomic region harboring *BACH2* scores as the strongest SE in Th1, Th2 and Th17 cells. Perhaps related, *BACH2* locus harbors HIV integration sites [61]. It remains to be determined whether HIV and other viruses preferentially integrate at SEs.

Relevant to the immune system, it was proposed that SEs may provide a platform for signaling pathways to regulate genes that control cell identity [8,43,62,63]. The study in tissue-resident macrophages suggested that environmental signals may play a role in the transition of collections of primed enhancers to genomic regions exhibiting features of super-enhancers [43]. The majority of SEs in tissue-resident macrophages were unique to one tissue but were located at regions marked by H3K4me2 in multiple macrophage subsets, indicating that these are “poised” enhancers [43]. The wide breadth of H3K27ac earmarked these regions as super-enhancer only in selected subsets, perhaps when signal transduction pathways relevant to those subsets were activated. In concert with these findings, the inhibitor of Jak/STAT pathway tofacitinib [64], also an effective drug used in the treatment of rheumatoid arthritis, preferentially acts on genes associated with the SE structure in T cells [8]. Moreover, superenhancers were shown to have a general role in integrating signaling inputs involving Notch and other pathways in T-lymphoblastic leukemia cells [62,63]. It was shown that 83% (843 out of 1,012) of dynamic NOTCH1 sites overlap with H3K27ac peaks, and that H3K27ac levels flanking dynamic NOTCH1 sites are reduced after the inhibition of NOTCH signaling [63]. Together, these studies suggest that signaling pathways may converge at these genomic regions during immune activation.

Concluding Remarks

The super- or stretch-enhancer concept is now more than a year old. Many genomic scientists recognize these exceptional regions and carried out studies to provide a framework

for us to understand them. Perhaps SEs can be likened to regions of exceptionally dense real estate. SEs represents “Manhattans” of the genome for any given cell. This is not to say that the other boroughs don’t count – after all, Yankee stadium is in the Bronx. Nonetheless super-enhancer domains may represent strategically placed discrete genomic regions in which tremendous resources seem to be invested - Location, location, location. Developmental biology might employ this strategy to accommodate highly cell-specific control of gene expression, but viruses and cancers also know this trick.

Nevertheless, debate continues in terms of the methodology used to characterize SE regions, and concerns have been raised regarding excessive emphasis in relevant publications [56]. Certainly, no part of the genome is super or typical and they are all employed to regulate gene expression. Given what is known about SEs at this time, we would like to argue that these regions are an important subset of enhancers and despite what we call them, they seem to have exceptional properties. Importantly, studies in immune cells show that the super-enhancer architecture is the hallmark of important genes and noncoding RNAs. Compared to gene expression analysis, a smaller subset of genes and noncoding RNAs seem to accommodate this exceptional architecture. Hence, discovering key players in newly identified cell types such as Innate lymphoid cells (ILCs) can be expedited once the SEs in these cell types are profiled.

Naturally, there are many remaining questions and concerns. Additional rigorous computational methods need to be developed to better characterize these genomic regions. The combinations of histone modifications and transcription factor binding reliably predicting SEs need to be fully characterized. It also remains unknown whether SEs work differently from any other enhancers in any qualitative sense. SEs assemble when multiple neighboring typical enhancers become simultaneously activated. To fully understand this process, it is important to consider what changes occur in the cell between a stage when only one or a few of the typical enhancers are active and another stage when these spread into a super-enhancer. Furthermore, it remains to be determined exactly which transcription factors regulate the formation and maintenance of these domains. The timing that these genomic regions are scored as super is also not known and the extent of their stability and relevance to gene expression needs to be explored. To prove their functional relevance, Crisper/Cas technology that edits elements within these regions needs to be employed more extensively. Whether super-enhancer delineation which relies on chromatin features can be a predictor of HIV integration sites requires further investigation. It is intriguing that drugs such as JQ1 and tofacitinib, Brd4 and Jak inhibitors, selectively target SEs, however, the reason for this selective impact is not fully identified. Despite these concerns and remaining issues (Box 3), focusing on SE regions and exploring the role of the environment on their formation is likely to provide an opportunity for immunologists to gain insights into potential regulatory hubs in the genomes of immune cells and their relevance to human disease.

Box 1

Prior to next-generation-sequencing (NGS)

To gain insights into immune cell identity and specialized function, immunologists initially focused on genes themselves. Microarray technology and RNA-sequencing has

been widely used to study the transcriptomes of immune cells [65-74]. These studies have provided a wealth of knowledge on cell-specific genes, which has aided our understanding of immune cell function. However, developments in 2007 largely supplanted these earlier techniques. The advent of massive parallel sequencing and “next generation sequencing (NGS)” permitted further advances in delineating cell-specific transcriptomes using RNA-seq [74]. Chromatin-immunoprecipitation was also coupled with NGS, ChIPseq, which enabled the analysis of genomewide binding of various proteins to DNA. For instance, the comprehensive binding of key transcription factors throughout the genome could suddenly be determined [75,76]. NGS technologies also allowed us to learn more about the packaging of DNA into chromatin. DNA is associated with histone octamers to form nucleosomes, an important determinant of whether genomic regions can be active or silent (euchromatin or heterochromatin respectively). That is, compact nucleosome structures restricts access of DNA-binding transcription regulators to regulatory elements whereas nucleosome-free regions permit binding of transcription factors and the transcriptional machinery. Initially, it was appreciated that acetylation of histones correlates with genomic accessibility. In early experiments, antibodies against acetylated histones were used to immunoprecipitate chromatin followed by PCR (ChIP-PCR) as a measure of accessibility of genomic regions (typically promoters). In contrast, trimethylation of histone 3 lysine 27 was suggested to correspond to inaccessible and repressed genomic regions. However, shortly thereafter, a more sophisticated understanding of the vast array of number of potential histone modifications along with their functional significance were obtained [77]. Using antibodies directed against a palette of histone modifications such as methylation and acetylation, ChIP-seq technology could be used to comprehensively visualize histone modifications annotating distinct types of accessible genomic elements [78]. While much effort had been exerted to understand the regulation of selected genes, NGS approaches proved to be unbiased and permitted assessment of the expression of all genes and the epigenetic correlates of regulation.

With the completion of the human genome project, it was quickly appreciated that the genome is mostly not genes – in fact, there were far fewer genes than had been initially predicted. While debate continues regarding the extent to which non-coding genome is functionally relevant or ‘junk’ [21,79], the revolution in sequencing called attention to activity outside protein coding regions. Since 2007, individual laboratories and international projects such as ENCODE and the NIH Roadmap Epigenomics Mapping Consortium have dug into the genomes of human and other mammals and revealed that an astonishing proportion of the genome is active in a tissue-specific manner [7,21,23,38,80-85]. The coordinated and context-specific regulation of genes is the focus of “epigenomics”. While the DNA sequence of every type of cell is the same, one can envision the epigenome as the genome’s second dimension, which contains key information specific to every type of cell [86]. An important appreciation was the utility of measuring histone modifications to identify enhancers on a genomic level.

Box 2**Disease associated mutations and SEs**

Advances in technology for finding diseases-associated variants led to a rapid increase in the number of known genetic variants that are associated with disease, largely through genome-wide association studies (GWAS). Many GWAS studies have focused on complex traits, those that are caused by the interaction of multiple genes, rather than monogenic diseases such as cystic fibrosis or Huntington's disease. These GWAS studies have identified dozens, and sometimes hundreds, of DNA variants that are associated with diseases such as Crohn's disease and colitis, multiple sclerosis, or traits like height. Unexpectedly, these newly identified disease-associated variants are located in non-coding regions of the genome, rather than within protein coding genes. This has led to a quest to try and find out exactly what genetic elements these variants are disrupting, to better understand the mechanisms underlying complex traits and diseases. The recent finding that many variants are located in SEs is intriguing and highlights the importance of SEs. It suggests that these variants may increase one's risk of disease by perturbing SEs, and also reveal how crucial SEs must be for cell function. Small perturbations such as a single nucleotide polymorphism (SNP) can lead to disrupted function and pathology.

Box 3**Outstanding Questions box**

- How stable SEs are and what factors control their stability?
- What combinations of histone modifications and transcription factor binding are more reliable predictors of genomic locations of SEs?
- What is the functional relevance of super-enhancers in regulation of immune cells?
- When and how do typical enhancers spread and become super-enhancers?
- Are integration sites for viruses other than HIV enriched within SEs of immune cells?
- Why SNPs associated with autoimmune diseases are enriched within SEs of T cells?
- Why Jak-inhibitor has a selective impact on genes associated with SEs in T cells?

References

1. Rothenberg EV. The chromatin landscape and transcription factors in T cell programming. *Trends Immunol.* 2014; 35:195–204. [PubMed: 24703587]
2. Downen JM, et al. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell.* 2014; 159:374–387. [PubMed: 25303531]

3. Hnisz D, et al. Super-enhancers in the control of cell identity and disease. *Cell*. 2013; 155:934–947. [PubMed: 24119843]
4. Hnisz D, et al. Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. *Mol Cell*. 2015; 58:362–370. [PubMed: 25801169]
5. Loven J, et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*. 2013; 153:320–334. [PubMed: 23582323]
6. Whyte WA, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*. 2013; 153:307–319. [PubMed: 23582322]
7. Farh KK, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*. 2015; 518:337–343. [PubMed: 25363779]
8. Vahedi G, et al. Super-enhancers delineate disease-associated regulatory nodes in T cells. *Nature*. 2015; 520:558–562. [PubMed: 25686607]
9. Blom van Assendelft G, et al. The beta-globin dominant control region activates homologous and heterologous promoters in a tissue-specific manner. *Cell*. 1989; 56:969–977. [PubMed: 2924354]
10. Gibcus JH, Dekker J. The hierarchy of the 3D genome. *Mol Cell*. 2013; 49:773–782. [PubMed: 23473598]
11. Chong MM, et al. Epigenetic propagation of CD4 expression is established by the Cd4 proximal enhancer in helper T cells. *Genes Dev*. 2010; 24:659–669. [PubMed: 20360383]
12. Sawada S, Littman DR. A heterodimer of HEB and an E12-related protein interacts with the CD4 enhancer and regulates its activity in T-cell lines. *Mol Cell Biol*. 1993; 13:5620–5628. [PubMed: 8355705]
13. Wilson CB, et al. Epigenetic control of T-helper-cell differentiation. *Nat Rev Immunol*. 2009; 9:91–105. [PubMed: 19151746]
14. Ansel KM, et al. Regulation of Th2 differentiation and Il4 locus accessibility. *Annual Review of Immunology*. 2006; 24:607–656.
15. Lee GR, et al. Regulation of IL-4 gene expression by distal regulatory elements and GATA-3 at the chromatin level. *Immunity*. 2001; 14:447–459. [PubMed: 11336690]
16. Spilianakis CG, Flavell RA. Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. *Nat Immunol*. 2004; 5:1017–1027. [PubMed: 15378057]
17. Josefowicz SZ, et al. Regulatory T cells: mechanisms of differentiation and function. *Annual review of immunology*. 2012; 30:531–564.
18. Josefowicz SZ, Rudensky A. Control of regulatory T cell lineage commitment and maintenance. *Immunity*. 2009; 30:616–625. [PubMed: 19464984]
19. Zheng Y, et al. Role of conserved non-coding DNA elements in the Foxp3 gene in regulatory T-cell fate. *Nature*. 2010; 463:808–812. [PubMed: 20072126]
20. Natoli G. Maintaining cell identity through global control of genomic organization. *Immunity*. 2012; 33:12–24. [PubMed: 20643336]
21. Kellis M, et al. Defining functional DNA elements in the human genome. *Proc Natl Acad Sci U S A*. 2014; 111:6131–6138. [PubMed: 24753594]
22. Creighton MP, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *PNAS*. 2010; 107:21931–21936. [PubMed: 21106759]
23. Neph S, et al. An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature*. 2012; 489:83–90. [PubMed: 22955618]
24. Rada-Iglesias A, et al. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*. 2010
25. Visel A, et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*. 2009; 457:854–858. [PubMed: 19212405]
26. Tolhuis B, et al. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell*. 2002; 10:1453–1465. [PubMed: 12504019]
27. Splinter E, et al. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev*. 2006; 20:2349–2354. [PubMed: 16951251]
28. Palstra RJ, et al. The beta-globin nuclear compartment in development and erythroid differentiation. *Nat Genet*. 2003; 35:190–194. [PubMed: 14517543]

29. Kooren J, et al. Beta-globin active chromatin Hub formation in differentiating erythroid cells and in p45 NF-E2 knock-out mice. *J Biol Chem.* 2007; 282:16544–16552. [PubMed: 17428799]
30. Drissen R, et al. The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes Dev.* 2004; 18:2485–2490. [PubMed: 15489291]
31. Lieberman-Aiden E, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science.* 2009; 326:289–293. [PubMed: 19815776]
32. Hadjur S, et al. Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature.* 2009; 460:410–U130. [PubMed: 19458616]
33. Kieffer-Kwon KR, et al. Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. *Cell.* 2013; 155:1507–1520. [PubMed: 24360274]
34. Handoko L, et al. CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat Genet.* 2011; 43:630–638. [PubMed: 21685913]
35. Dixon JR, et al. Chromatin architecture reorganization during stem cell differentiation. *Nature.* 2015; 518:331–336. [PubMed: 25693564]
36. Dixon JR, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature.* 2012; 485:376–380. [PubMed: 22495300]
37. Neph S, et al. Circuitry and dynamics of human transcription factor regulatory networks. *Cell.* 2012; 150:1274–1286. [PubMed: 22959076]
38. Roadmap Epigenomics C, et al. Integrative analysis of 111 reference human epigenomes. *Nature.* 2015; 518:317–330. [PubMed: 25693563]
39. Ciofani M, et al. A validated regulatory network for th17 cell specification. *Cell.* 2012; 151:289–303. [PubMed: 23021777]
40. Samstein RM, et al. Foxp3 exploits a pre-existent enhancer landscape for regulatory T cell lineage specification. *Cell.* 2012; 151:153–166. [PubMed: 23021222]
41. Vahedi G, et al. STATs shape the active enhancer landscape of T cell populations. *Cell.* 2012; 151:981–993. [PubMed: 23178119]
42. Hawkins RD, et al. Global chromatin state analysis reveals lineage-specific enhancers during the initiation of human T helper 1 and T helper 2 cell polarization. *Immunity.* 2013; 38:1271–1284. [PubMed: 23791644]
43. Gosselin D, et al. Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. *Cell.* 2015; 159:1327–1340. [PubMed: 25480297]
44. Lavin Y, et al. Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell.* 2015; 159:1312–1326. [PubMed: 25480296]
45. Buenrostro JD, et al. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods.* 2013; 10:1213–1218. [PubMed: 24097267]
46. Ghisletti S, et al. Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. *Immunity.* 2010; 32:317–328. [PubMed: 20206554]
47. Heinz S, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell.* 2010; 38:576–589. [PubMed: 20513432]
48. Mullen AC, et al. Master transcription factors determine cell-type-specific responses to TGF-beta signaling. *Cell.* 2011; 147:565–576. [PubMed: 22036565]
49. Ostuni R, et al. Latent enhancers activated by stimulation in differentiated cells. *Cell.* 2013; 152:157–171. [PubMed: 23332752]
50. Hesselberth JR, et al. Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nat Methods.* 2009; 6:283–289. [PubMed: 19305407]
51. Kagey MH, et al. Mediator and cohesin connect gene expression and chromatin architecture. *Nature.* 2010; 467:430–435. [PubMed: 20720539]
52. Chapuy B, et al. Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. *Cancer Cell.* 2013; 24:777–790. [PubMed: 24332044]
53. Adam RC, et al. Pioneer factors govern super-enhancer dynamics in stem cell plasticity and lineage choice. *Nature.* 2015; 521:366–370. [PubMed: 25799994]

54. Parker SC, et al. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc Natl Acad Sci U S A*. 2013; 110:17921–17926. [PubMed: 24127591]
55. Ernst J, Kellis M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol*. 2010; 28:817–825. [PubMed: 20657582]
56. Pott S, Lieb JD. What are super-enhancers? *Nat Genet*. 2015; 47:8–12. [PubMed: 25547603]
57. Zhu H, et al. Unexpected characteristics of the IFN-gamma reporters in nontransformed T cells. *J Immunol*. 2001; 167:855–865. [PubMed: 11441092]
58. Aune TM, et al. Epigenetics and T helper 1 differentiation. *Immunology*. 2009; 126:299–305. [PubMed: 19178593]
59. Roychoudhuri R, et al. BACH2 represses effector programs to stabilize T(reg)-mediated immune homeostasis. *Nature*. 2013; 498:506–510. [PubMed: 23728300]
60. Tsukumo S, et al. Bach2 maintains T cells in a naive state by suppressing effector memory-related genes. *Proc Natl Acad Sci U S A*. 2013; 110:10735–10740. [PubMed: 23754397]
61. Wagner TA, et al. HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science*. 2014; 345:570–573. [PubMed: 25011556]
62. Yashiro-Ohtani Y, et al. Long-range enhancer activity determines Myc sensitivity to Notch inhibitors in T cell leukemia. *Proc Natl Acad Sci U S A*. 2014; 111:E4946–4953. [PubMed: 25369933]
63. Wang H, et al. NOTCH1-RBPJ complexes drive target gene expression through dynamic interactions with superenhancers. *Proc Natl Acad Sci U S A*. 2014; 111:705–710. [PubMed: 24374627]
64. Ghoreschi K, et al. Modulation of innate and adaptive immune responses by tofacitinib (CP-690,550). *J Immunol*. 2011; 186:4234–4243. [PubMed: 21383241]
65. Heller RA, et al. Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. *Proc Natl Acad Sci U S A*. 1997; 94:2150–2155. [PubMed: 9122163]
66. Kelly K, Siebenlist U. Immediate-early genes induced by antigen receptor stimulation. *Curr Opin Immunol*. 1995; 7:327–332. [PubMed: 7546396]
67. Rogge L, et al. Transcript imaging of the development of human T helper cells using oligonucleotide arrays. *Nat Genet*. 2000; 25:96–101. [PubMed: 10802665]
68. Staudt LM, Brown PO. Genomic views of the immune system*. *Annu Rev Immunol*. 2000; 18:829–859. [PubMed: 10837077]
69. Kaech SM, et al. Molecular and functional profiling of memory CD8 T cell differentiation. *Cell*. 2002; 111:837–851. [PubMed: 12526810]
70. Wherry EJ, et al. Molecular signature of CD8+ T cell exhaustion during chronic viral infection. *Immunity*. 2007; 27:670–684. [PubMed: 17950003]
71. Hamalainen H, et al. Distinct gene expression profiles of human type 1 and type 2 T helper cells. *Genome Biol*. 2001; 2 RESEARCH0022.
72. Hoey T, et al. Distinct requirements for the naturally occurring splice forms Stat4alpha and Stat4beta in IL-12 responses. *EMBO J*. 2003; 22:4237–4248. [PubMed: 12912921]
73. Beima KM, et al. T-bet binding to newly identified target gene promoters is cell type-independent but results in variable context-dependent functional effects. *J Biol Chem*. 2006; 281:11992–12000. [PubMed: 16473879]
74. Mortazavi A, et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*. 2008; 5:621–628. [PubMed: 18516045]
75. Johnson DS, et al. Genome-wide mapping of in vivo protein-DNA interactions. *Science*. 2007; 316:1497–1502. [PubMed: 17540862]
76. Farnham PJ. Insights from genomic profiling of transcription factors. *Nat Rev Genet*. 2009; 10:605–616. [PubMed: 19668247]
77. Strahl BD, Allis CD. The language of covalent histone modifications. *Nature*. 2000; 403:41–45. [PubMed: 10638745]
78. Heintzman ND, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics*. 2007; 39:311–318. [PubMed: 17277777]

79. Doolittle WF. Is junk DNA bunk? A critique of ENCODE. *Proc Natl Acad Sci U S A*. 2013; 110:5294–5300. [PubMed: 23479647]
80. Djebali S, et al. Landscape of transcription in human cells. *Nature*. 2012; 489:101–108. [PubMed: 22955620]
81. Dunham I, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012; 489:57–74. [PubMed: 22955616]
82. Gjoneska E, et al. Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease. *Nature*. 2015; 518:365–369. [PubMed: 25693568]
83. Barski A, et al. High-resolution profiling of histone methylations in the human genome. *Cell*. 2007; 129:823–837. [PubMed: 17512414]
84. Schones DE, et al. Dynamic regulation of nucleosome positioning in the human genome. *Cell*. 2008; 132:887–898. [PubMed: 18329373]
85. Thurman RE, et al. The accessible chromatin landscape of the human genome. *Nature*. 2012; 489:75–82. [PubMed: 22955617]
86. Rivera CM, Ren B. Mapping human epigenomes. *Cell*. 2013; 155:39–55. [PubMed: 24074860]

Highlights

- Super-enhancers are defined by exceptionally strong binding of transcriptional regulators.
- These noncoding segments of the genome can bookmark important regulatory regions.
- Super-enhancers can reveal important pathways and regulatory nodes in immune cells.

Figure 1. Chromatin signature of typical and super/stretch enhancers

Various categories of enhancers can be differentiated from each other based on their chromatin feature. Latent enhancers are inactive, located in closed chromatin, and do not contain any distinguishing histone marks. However, these regions do have the potential to acquire the chromatin signature of active enhancers after stimulations. Poised enhancers are also located in closed chromatin, however these enhancer regions are characterized by histones bearing the H3K4me1 modification. Active enhancers are identified by histones bearing H3K27ac and H3K4me1 modifications. Super-enhancers are cluster of active enhancers (H3K27ac and H3K4me1) with exceptional level of modifications. These regions are also often enriched for master regulator transcription factors. Stretch-enhancers are defined based on a combination of histone modifications.

Figure 2. What's the difference between super-enhancers and stretch-enhancers?

Super-enhancers are defined by density of proteins or histone modifications found at enhancer sites (Med1, Brd4, etc.). Super-enhancers are differentiated from TEs because they contain exponentially higher levels of these factors. They occur near cell identity genes and are to a great extent cell specific. Super-enhancers contain disease-associated SNPs, drive cancer genes, and their number can be as high as 1000 in a given cell type. Stretch-enhancers are defined based on a machine-learning algorithm: ChromHMM. They require multiple histone modification datasets such as H3K27ac, H3K4me1, 2, and 3. The cutoff for the enhancer size is > 3kb. They are also near cell identity genes, are very cell/tissue specific, and contain disease-associated SNPs. Figure corresponding super vs typical enhancer is adopted from [6].