

Transcriptional enhancers: functional insights and role in human disease

Irene Miguel-Escalada^{1,2,3}, Lorenzo Pasquali^{3,4,5} and Jorge Ferrer^{1,2,3}



In recent years, studies of *cis*-regulatory mechanisms have evolved from a predominant focus on promoter regions to the realization that spatial and temporal gene regulation is frequently driven by long-range enhancer clusters that operate within chromosomal compartments. This increased understanding of genome function, together with the emergence of technologies that enable whole-genome sequencing of patients' DNAs, open the prospect of dissecting the role of *cis*-regulatory defects in human disease. In this review we discuss how recent epigenomic studies have provided insights into the function of transcriptional enhancers. We then present examples that illustrate how integrative genomics can help uncover enhancer sequence variants underlying Mendelian and common polygenic human disease.

Addresses

¹ Department of Medicine, Imperial College London, London W12 0NN, United Kingdom

² Genomic Programming of Beta-cells Laboratory, Institut d'Investigacions August Pi i Sunyer (IDIBAPS), 08036 Barcelona, Spain

³ CIBER de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), 08036 Barcelona, Spain

⁴ Division of Endocrinology, Germans Trias i Pujol University Hospital and Research Institute and Josep Carreras Leukaemia Research Institute, 08036 Barcelona, Spain

⁵ Josep Carreras Leukaemia Research Institute, 08036 Barcelona, Spain

Corresponding author: Ferrer, Jorge (j.ferrer@imperial.ac.uk)

Current Opinion in Genetics & Development 2015, 33:71–76

This review comes from a themed issue on **Molecular and genetic bases of disease**

Edited by **Dan E Arking** and **Johanna M Rommens**

<http://dx.doi.org/10.1016/j.gde.2015.08.009>

0959-437/© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Introduction

Massive sequencing technologies have demonstrated an extraordinary power to uncover disease-causing variants in protein-coding sequences. It is now necessary to ask whether similar technologies can be exploited to discover defects in the ~1 million transcriptional regulatory sequences that have been unearthed in the past few years. This challenge, however, is hindered by our incomplete understanding of the function of

transcriptional regulatory elements. This review will focus on recent advances in understanding the function of transcriptional enhancers, and present examples of how integrative genomics can help identify enhancer defects that underlie Mendelian and polygenic disease.

Clustering of active enhancers

Enhancers were first defined as DNA sequences that stimulate transcription from a minimal promoter, regardless of orientation or relative distance [1]. Subsequent studies showed that long-range enhancers are pivotal for spatial and temporal regulation of gene transcription in metazoan genomes. Most recently it has become possible to catalogue the entire genomic repertoire of active enhancers in any cellular population by exploiting distinctive enhancer features such as: increased accessibility to enzymes (DNase-seq, ATAC-seq), or relative nucleosome depletion (FAIRE-seq) [2–4]; enrichment of specific modified histones (H3K27Ac, H3K4me1) [3,5]; occupancy by co-regulatory factors (p300, BRD4, Mediator) [6,7]; and finally, RNA transcription from enhancer-flanking regions [8]. All such features can now be studied with high-throughput sequencing-based assays, which has enabled the generation of enhancer maps in numerous cell lines and primary tissues [8,9^{**},10,11].

A recurrent theme that emerged from recent enhancer maps is that most lineage-specific gene transcription occurs near clusters of active enhancers. This had been previously recognized in the form of clusters of evolutionary conserved sequences flanking lineage-specific regulatory genes [12], or from functional studies of numerous individual loci. However, enhancer maps now provide an unbiased perspective based on genome-scale experimental data. Regulatory clusters have thus been described as clusters of open regulatory elements (COREs) [2], superenhancers [13], stretch enhancers [14], or enhancer clusters [11]. One study mapped human pancreatic islet enhancers, and found that most islet-enriched genes are associated with three or more clustered enhancers, which tend to be co-occupied by multiple islet-specific TFs [11]. Chromatin conformation capture (3C) assays showed that clustered enhancers form higher order physical structures and establish physical interactions with target genes [11]. These enhancer clusters were consistent with earlier studies showing COREs (open chromatin clusters defined by FAIRE-seq) near

islet-specific genes [2], and with the more general observation that expression of genes across multiple tissues correlates with the activity of multiple local enhancers in the same locus [15]. Another study defined ‘stretch’ enhancers as H3K27Ac-rich chromatin regions >3 kb, and found them to be frequently located near cell-specific genes [14]. Another set of studies defined ‘superenhancers’ as extended enhancer regions that show unusually high occupancy by either Mediator, TFs, or H3K27Ac-modified nucleosomes [13,16**] (for an in-depth commentary on superenhancers see [17**]). Superenhancers have been linked to genes that are central for pluripotency or cell type identity as well as to oncogenes, and shown to be particularly sensitive to targeting by co-regulator inhibitors [13,16**,18].

Regardless of varying definitions and nomenclatures, recently described enhancer domains are, in essence, sets of adjacent active enhancers. This raises the question of why there is a need for multiple enhancers to create cell-specific transcription. Possible explanations include redundancy (‘shadow’ enhancers) and combinatorial or synergistic specificity, although recent genetic studies provide further explanations. Spitz and colleagues, for example, used a broad range of mouse genetic tools to dissect an enhancer cluster regulating *Fgf8* [19*]. This showed that the regulatory output of an enhancer cluster (in this case the cell types in which *Fgf8* is expressed) is not simply a summation of individual enhancer activities, but is instead dependent on a combined function of clustered enhancers, or ‘holo-enhancer units’ [19*]. On the other hand, multiple studies indicate that enhancer clusters form higher-order 3D structures [11,20–22,23*], suggesting that ‘holo-structures’ might be crucial for cell-specific transcription.

Enhancer function in the context of the 3D genome

3C studies have established general principles that underlie 3D genome organization, and promise to enlighten how enhancers interact with their functional targets. Hi-C sequencing has shown that the genome is packaged at multiple organizational levels, including so-called topologically associated domains (TADs) [24**]. TADs, which span on average ~0.8 Mb, are defined by a high number of intra-domain 3C interactions and rare interactions between adjacent domains. A recent study used random insertions of a reporter that acts as a sensor of endogenous enhancer activity, and showed that TADs provide a spatial compartment within which enhancers interact functionally (and not solely physically) with their target promoters [25]. Others have demonstrated coordinated gene regulation within the confines of TADs [26,27].

Increased resolution mapping using 5C or Hi-C libraries revealed further subdomains within TADs, including ‘loops’ that are bound at their stem by CTCF, as well

as cohesin and mediator-bound cell-specific ‘loops’ that link enhancers to promoters [28,29]. 4C-seq studies, a 3C variant that interrogates all genomic sites interacting with a viewpoint of interest at very high resolution, have shown that clusters of lineage-specific enhancers establish frequent interactions amongst themselves and with target gene promoters [11,20–22]. Interestingly, while TAD boundaries are typically invariant across cell types, they contain structures that are often cell-specific and dynamic [28,30].

Looping into promoters is thought to underlie enhancer function, and this was recently tested by artificial tethering of an enhancer to a promoter, leading to increased transcriptional activity [31]. It is nevertheless also true that each enhancer often shows 3C interaction signals with multiple nearby enhancers and promoters, and each promoter with multiple enhancers and promoters [32,33]. One theoretical implication of this observation is that if all such interactions are functional, then sequence variation in single enhancers could potentially impact multiple genes. However, while 3C assays most probably do capture regulatory interactions between enhancers and promoters, it is unclear if all 3C interactions are functional. In fact, studies have challenged the significance of 3C interactions, and questioned whether other variables apart from physical proximity affect ligation frequency in 3C experiments, and whether 3C interaction signals represent discrete loops [34]. This warrants a need for cross-link-independent methods for studying 3D structure. Interestingly, a recent study used high-resolution live cell imaging to show widespread Sox2-bound clustered enhancers in ESCs, providing further independent evidence that enhancer clusters form structural units [23*]. Diverse approaches are thus becoming available to probe the impact of enhancer mutations on higher order chromatin structures.

Taken together, recent studies provide an initial framework for understanding how long-range enhancers operate in the context of genome organization. Future studies that couple 3D interaction experiments with functional perturbations, including targeted mutations and eQTL studies, should provide further light on mechanistic and functional relationships between enhancers and target genes. This type of knowledge will be vital for understanding how enhancer variants could be deleterious in the context of 3D chromosomal structure, and to identify the genes that are affected by defective enhancers.

Mendelian regulatory defects

Notable examples of long-range enhancer mutations that cause monogenic disorders include those regulating *SHH* (preaxial polydactyly) [35], *SOX9* (Pierre Robin Syndrome) [36], and *TBX5* (congenital heart disease) [37]. These and other known enhancer mutations were identified after careful functional characterization of enhancers,

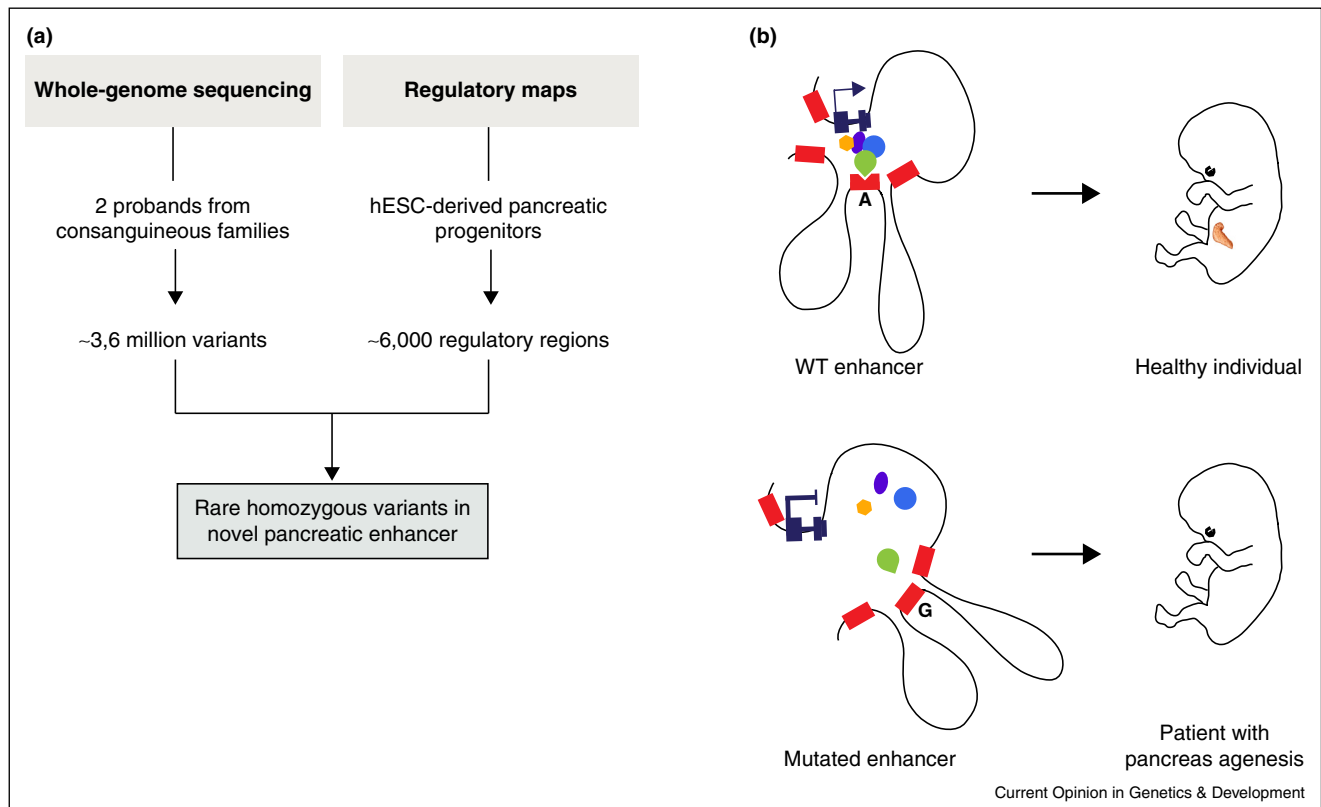
followed by targeted sequencing, or else by the discovery of large deletions or rearrangements that were subsequently shown to contain enhancers. This approach is relatively inefficient when compared with the success of whole-exome sequencing for detection of protein-coding mutations.

A recent study exemplifies a systematic approach to discover enhancer mutations (Figure 1). Hattersley and colleagues carried out whole-genome sequencing and homozygosity mapping of SNPs in two unrelated consanguineous probands with isolated pancreas agenesis and no causal protein-coding mutations [38**]. Integration of this data with enhancer charts from human embryonic pancreatic progenitors revealed homozygous point mutations in a single unannotated enhancer >25 kb from *PTF1A*, a known pancreatic regulatory gene. Subsequent analysis of 12 unrelated families with isolated pancreas agenesis showed that 10 had rare homozygous mutations in this

enhancer, including a large deletion and point mutations that disrupted functional binding sites of pancreatic developmental TFs [38**].

The analysis of isolated pancreas agenesis has noteworthy implications. One is that it illustrates how one can progress from a person's inventory of >3 million non-coding variants to the identification of a causal non-coding mutation. It was also an unbiased genome-scale analysis that showed that mutations that disrupt recognizable *cis*-regulatory sequences can be the most common cause of a discrete phenotype (in this case isolated pancreas agenesis). It is also noteworthy that the pancreatic progenitor enhancer that harbored mutations was inactive in a broad panel of tissues, which highlights that any search for non-coding defects needs to focus on disease-relevant epigenomic annotations. Finally, it is interesting that despite that there are multiple pancreatic progenitor enhancers near *PTF1A* [4,39], all mutations fell in a single enhancer.

Figure 1



Integrative genomics reveals that isolated pancreas agenesis is caused by enhancer mutations. **(a)** Integration of whole genome sequences from two patients with pancreatic agenesis with maps of active enhancers profiled in human embryonic pancreatic progenitors identified causal recessive mutations. These mutations map to a previously unannotated enhancer 25 kb away from *PTF1A*, a transcription factor that is known to be essential for the embryonic development of the pancreas. **(b)** Schematic representation of the *PTF1A* locus harboring wild-type (A) and mutated (G) enhancer sequences. The newly identified enhancer (indented red box) establishes a physical interaction with the *PTF1A* promoter and is bound by regulatory factors such as FOXA2 (green teardrop). The presence of a single-nucleotide enhancer variant in some patients with pancreas agenesis (g.23508437A > G) disrupts binding by FOXA2, abolishes enhancer activity and potentially alters the local chromatin structure of the enhancer cluster. A deletion of this enhancer region or other single base mutations that disrupt binding of FOXA2, PDX1 or an unidentified binding protein cause the same phenotype, thus highlighting a crucial role of this enhancer in the active conformation of the *PTF1A* locus.

Analogously, engineered mutations in enhancer clusters show that only some clustered enhancers in *Sox2* are essential in ESCs [40]. This suggests a functional hierarchy within enhancer clusters, perhaps due to a hub-like function of specific enhancers within 3D structures.

The pancreas agenesis studies support future efforts to integrate whole genome sequencing with regulatory annotations to discover Mendelian non-coding defects. Further discoveries of pathogenic enhancer mutations from screens of natural and engineered variants should inform computational algorithms that enable prediction of pathogenic regulatory variants. Despite the limited amount of data, several approaches have already been developed to predict which non-coding variants within regulatory elements are functional. Most have examined whether variants affect nucleotides in TF-binding motifs, are under evolutionary selective pressure, or show polymorphism in humans [41–43]. Additional factors, such as the position of variant enhancers in the context of regulatory domains, are also probably to affect pathogenicity. The availability of large numbers of regulatory mutations should thus facilitate future understanding of Mendelian and complex non-coding defects.

Common variation in enhancers and human disease

Most common diseases, including prevalent forms of cancer, Type 2 diabetes, or late-onset Alzheimer's disease, result from environmental factors interacting with genetic susceptibility variants. Genome-wide association studies (GWAS) have identified thousands of loci that affect the susceptibility to common diseases. Many risk loci do not contain causal protein-coding variants, suggesting a role for regulatory variation [44,45,46]. This entails major challenges for translating GWAS findings to molecular insights. Associated haplotype blocks include many variants, which means that it is necessary to identify the specific causal regulatory variants at each associated locus. Even after prioritizing functional variants, there is no straightforward approach to conclusively establish the genes that are affected by the variant, and the relevant biological context. Only when this information is available is it possible to study how inherited changes in gene regulation affect cellular pathways that underlie disease.

Recent studies have made considerable progress to address these challenges. A plethora of studies have now shown that variants associated with common diseases are enriched in enhancers, and this has sometimes led to identification of functional variants [9,11,16,44,45,47,48]. Studies have further revealed a specific enrichment in enhancer clusters or superenhancers that are active in cell types that match a coherent pathophysiological model of the disease [11,14,16,49]. Two studies, for example, showed that SNPs associated with Type 2 diabetes and fasting glycemia levels are enriched in pancreatic islet clustered enhancers

and stretch enhancers [11,14]. This showed that islet-specific regulatory variation is relevant to Type 2 diabetes pathophysiology, and enabled functional characterization of discrete risk variants that disrupt TF-binding motifs and impact the activity of islet enhancers [11]. Another example focused on 21 autoimmune disorders, and used dense genotyping of large patient cohorts to greatly reduce the number of candidate causal variants per locus [45]. These SNPs were enriched in non-synonymous protein-coding variants, but also in enhancer variants, with a notorious enrichment in dynamically stimulated and clustered T-cell enhancers [45]. Taken together, recent work has shown that it is possible to identify disease-associated functional variants in cell-specific enhancers, which represents a giant step towards understanding molecular mechanisms of common diseases.

In addition to identifying functional variants, it is challenging to define which genes are affected. A study that analyzed *FTO*, the major obesity susceptibility locus, provides a paradigm for how this problem can be tackled [50]. Risk variants in *FTO* intronic regions were presumed to affect *FTO*, whose mouse KO phenotype causes reduced body size [51,52]. Unexpectedly, *FTO* regions carrying risk variants form 3C interactions with *IRX3*, located >400 kb away, and confer enhancer activity in cellular domains that coincide with *IRX3* (rather than *FTO*) expression [50]. Furthermore, risk SNPs at *FTO* were associated with brain eQTLs that affected expression of *IRX3*, but not *FTO*. Interestingly *IRX3* KO mice show decreased lean body mass [50]. This landmark study therefore shows that contrary to previous assumptions *IRX3* is a functional target of regulatory variants that affect obesity susceptibility.

In summary, recent studies have started to deploy a broad range of genetic and functional tools that enable untangling the regulatory function of common disease-associated variants. Clearly, the level of evidence that is needed to conclusively implicate a specific non-coding variant in causality remains a challenge. Importantly, most common diseases are not easily modeled in an organism by introducing a single regulatory allele. However, as in classic protein-coding Mendelian diseases, it should be possible to implicate a genetic variant by combining multiple lines of evidence, including human genetics (fine mapping, transethnic studies) and functional studies (allelic expression, reporter and 3C assays, genome editing). Ultimately, the goal is not solely to determine which variants are causal, but to understand the genetic pathways they regulate and to harness this knowledge to develop more efficient therapies.

Until very recent times, the fields of gene regulation and human genetics were largely unlinked. The studies we have reviewed illustrate how recent advances in these two fields are rapidly converging to enlighten new genetic

mechanisms underlying Mendelian and polygenic human disease.

Acknowledgements

Irene Miguel is a Marie Skłodowska-Curie Fellow (658145), Lorenzo Pasquali is a Ramon y Cajal investigator (RYC 2014-0069) and Jorge Ferrer is a Wellcome Trust Senior Investigator (101033).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Banerji J, Rusconi S, Schaffner W: **Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences.** *Cell* 1981, **27**:299-308.
2. Gaulton KJ, Nammo T, Pasquali L, Simon JM, Giresi PG, Fogarty MP, Panhuis TM, Mieczkowski P, Secchi A, Bosco D *et al.*: **A map of open chromatin in human pancreatic islets.** *Nat Genet* 2010, **42**:255-259.
3. Song L, Zhang Z, Grasfeder LL, Boyle AP, Giresi PG, Lee BK, Sheffield NC, Graf S, Huss M, Keefe D *et al.*: **Open chromatin defined by DNase and FAIRE identifies regulatory elements that shape cell-type identity.** *Genome Res* 2011, **21**:1757-1767.
4. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ: **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.
5. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J: **A unique chromatin signature uncovers early developmental enhancers in humans.** *Nature* 2011, **470**:279-283.
6. Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F *et al.*: **ChIP-seq accurately predicts tissue-specific activity of enhancers.** *Nature* 2009, **457**:854-858.
7. Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS *et al.*: **Mediator and cohesin connect gene expression and chromatin architecture.** *Nature* 2010, **467**:430-435.
8. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T *et al.*: **An atlas of active enhancers across human cell types and tissues.** *Nature* 2014, **507**:455-461.
9. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J *et al.*: **Integrative analysis of 111 reference human epigenomes.** *Nature* 2015, **518**:317-330.
- One of the hallmark papers from roadmap epigenomics, describing 'reference epigenomes' based on 2805 datasets from a broad panel of adult, embryonic, and ESC-derived samples. They identify coordinated enhancer modules, and show enrichment of disease risk variants in tissue-specific enhancer sets.
10. Consortium EP: **An integrated encyclopedia of DNA elements in the human genome.** *Nature* 2012, **489**:57-74.
11. Pasquali L, Gaulton KJ, Rodriguez-Segui SA, Mularoni L, Miguel-Escalada I, Akerman I, Tena JJ, Moran I, Gomez-Marin C, van de Bunt M *et al.*: **Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants.** *Nat Genet* 2014, **46**:136-143.
- This study shows that human pancreatic beta cell-specific genes are almost invariably associated with enhancer clusters, some of which are shown to form 3D chromosomal structures. Type 2 diabetes risk variants are enriched in islet enhancer clusters, and some risk variants disrupt enhancer activity.
12. Akalin A, Fredman D, Arner E, Dong X, Bryne JC, Suzuki H, Daub CO, Hayashizaki Y, Lenhard B: **Transcriptional features of genomic regulatory blocks.** *Genome Biol* 2009, **10**:R38.
13. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA: **Master transcription factors and mediator establish super-enhancers at key cell identity genes.** *Cell* 2013, **153**:307-319.
14. Parker SC, Stitzel ML, Taylor DL, Orozco JM, Erdos MR, Akiyama JA, van Bueren KL, Chines PS, Narisu N, Program NCS *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.
15. Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, Wagner U, Dixon J, Lee L, Lobanenkov VV *et al.*: **A map of the cis-regulatory sequences in the mouse genome.** *Nature* 2012, **488**:116-120.
16. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, Hoke HA, Young RA: **Super-enhancers in the control of cell identity and disease.** *Cell* 2013, **155**:934-947.
- This article and Ref. [13] describe superenhancers after merging adjacent enhancers that showed particularly high Mediator occupancy or H3K27Ac enrichment. They show association with cell-specific regulators, and enrichment in disease-associated genes.
17. Pott S, Lieb JD: **What are super-enhancers?** *Nat Genet* 2015, **47**:8-12.
- A thoughtful perspective on the definitions and potential significance of superenhancers.
18. Loven J, Hoke HA, Lin CY, Lau A, Orlando DA, Vakoc CR, Bradner JE, Lee TI, Young RA: **Selective inhibition of tumor oncogenes by disruption of super-enhancers.** *Cell* 2013, **153**:320-334.
19. Marinic M, Aktas T, Ruf S, Spitz F: **An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape.** *Dev Cell* 2013, **24**:530-542.
- This study functionally dissects a cluster of enhancers that regulate *Fgf8*. They find that the *Fgf8* expression domain is best explained by an integrated function of multiple enhancers, rather than by a summation of individual enhancer activities, and speculate that this might result from an integrated 3D structure.
20. Montavon T, Soshnikova N, Mascrez B, Joye E, Thevenet L, Splinter E, de Laat W, Spitz F, Duboule D: **A regulatory archipelago controls Hox genes transcription in digits.** *Cell* 2011, **147**:1132-1145.
21. Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W: **Looping and interaction between hypersensitive sites in the active beta-globin locus.** *Mol Cell* 2002, **10**:1453-1465.
22. Tena JJ, Alonso ME, de la Calle-Mustienes E, Splinter E, de Laat W, Manzanares M, Gomez-Skarmeta JL: **An evolutionarily conserved three-dimensional structure in the vertebrate *Irx* clusters facilitates enhancer sharing and coregulation.** *Nat Commun* 2011, **2**:310.
23. Liu Z, Legant WR, Chen BC, Li L, Grimm JB, Lavis LD, Betzig E: **Tjian R3D imaging of Sox2 enhancer clusters in embryonic stem cells.** *Elife* 2014, **3**:e04236.
- In this study super resolution microscopy is used for live cell imaging of 3D enhancer cluster structures.
24. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B: **Topological domains in mammalian genomes identified by analysis of chromatin interactions.** *Nature* 2012, **485**:376-380.
- Hi-C, a genome-scale chromosome conformation capture assay, showed that the genome is organized in domains that show many intra-domain but few inter-domain interactions.
25. Symmons O, Uslu VV, Tsujimura T, Ruf S, Nassari S, Schwarzer W, Ettwiller L, Spitz F: **Functional and topological characteristics of mammalian regulatory domains.** *Genome Res* 2014, **24**:390-400.
26. Le Dily F, Bau D, Pohl A, Vicent GP, Serra F, Soronellas D, Castellano G, Wright RH, Ballare C, Filion G *et al.*: **Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation.** *Genes Dev* 2014, **28**:2151-2162.
27. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J *et al.*: **Spatial partitioning of the regulatory landscape of the X-inactivation centre.** *Nature* 2012, **485**:381-385.

28. Phillips-Cremins JE, Sauria ME, Sanyal A, Gerasimova TI, Lajoie BR, Bell JS, Ong CT, Hookway TA, Guo C, Sun Y *et al.*: **Architectural protein subclasses shape 3D organization of genomes during lineage commitment.** *Cell* 2013, **153**:1281-1295.
29. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES *et al.*: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.** *Cell* 2014, **159**:1665-1680.
30. Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, Ye Z, Kim A, Rajagopal N, Xie W *et al.*: **Chromatin architecture reorganization during stem cell differentiation.** *Nature* 2015, **518**:331-336.
31. Deng W, Lee J, Wang H, Miller J, Reik A, Gregory PD, Dean A, Blobel GA: **Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor.** *Cell* 2012, **149**:1233-1244.
32. Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J *et al.*: **Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation.** *Cell* 2012, **148**:84-98.
33. Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, Yen CA, Schmitt AD, Espinoza CA, Ren B: **A high-resolution map of the three-dimensional chromatin interactome in human cells.** *Nature* 2013, **503**:290-294.
34. Williamson I, Berlivet S, Eskeland R, Boyle S, Illingworth RS, Paquette D, Dostie J, Bickmore WA: **Spatial genome organization contrasting views from chromosome conformation capture and fluorescence in situ hybridization.** *Genes Dev* 2014, **28**:2778-2791.
35. Lettice LA, Heaney SJ, Purdie LA, Li L, de Beer P, Oostra BA, Goode D, Elgar G, Hill RE, de Graaff E: **A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly.** *Hum Mol Genet* 2003, **12**:1725-1735.
36. Benko S, Fantes JA, Amiel J, Kleinjan DJ, Thomas S, Ramsay J, Jamshidi N, Essafi A, Heaney S, Gordon CT *et al.*: **Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence.** *Nat Genet* 2009, **41**:359-364.
37. Smemo S, Campos LC, Moskowitz IP, Krieger JE, Pereira AC, Nobrega MA *et al.*: **Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease.** *Hum Mol Genet* 2012, **21**:3255-3263.
38. Weedon MN, Cebola I, Patch AM, Flanagan SE, De Franco E, Caswell R, Rodriguez-Segui SA, Shaw-Smith C, Cho CH, Lango Allen H *et al.*: **Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis.** *Nat Genet* 2014, **46**:61-64.
- Integration of whole-genome sequencing, homozygosity mapping, and epigenomic maps revealed that mutations in a developmental enhancer are the most common cause of isolated pancreas agenesis.
39. Pashos E, Park JT, Leach S, Fisher S: **Distinct enhancers of ptf1a mediate specification and expansion of ventral pancreas in zebrafish.** *Dev Biol* 2013, **381**:471-481.
40. Zhou HY, Katsman Y, Dhaliwal NK, Davidson S, Macpherson NN, Sakthidevi M, Collura F, Mitchell JA: **A Sox2 distal enhancer cluster regulates embryonic stem cell differentiation potential.** *Genes Dev* 2014, **28**:2699-2711.
41. Gronau I, Arbiza L, Mohammed J, Siepel A: **Inference of natural selection from interspersed genomic elements based on polymorphism and divergence.** *Mol Biol Evol* 2013, **30**:1159-1171.
42. Gulko B, Hubisz MJ, Gronau I, Siepel A: **A method for calculating probabilities of fitness consequences for point mutations across the human genome.** *Nat Genet* 2015, **47**:276-283.
43. Khurana E, Fu Y, Colonna V, Mu XJ, Kang HM, Lappalainen T, Sboner A, Lochovsky L, Chen J, Harmanci A *et al.*: **Integrative annotation of variants from 1092 humans: application to cancer genomics.** *Science* 2013, **342**:1235587.
44. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J *et al.*: **Systematic localization of common disease-associated variation in regulatory DNA.** *Science* 2012, **337**:1190-1195.
45. Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, Shores N, Whitton H, Ryan RJ, Shishkin AA *et al.*: **Genetic and epigenetic fine mapping of causal autoimmune disease variants.** *Nature* 2015, **518**:337-343.
- Fine mapping of variants associated with autoimmune disorders in a very large number of subjects allowed significant reduction of candidate causal variants, and showed enrichment in immune cell enhancers.
46. Ongen H, Andersen CL, Bramsen JB, Oster B, Rasmussen MH, Ferreira PG, Sandoval J, Vidal E, Whiffin N, Planchon A *et al.*: **Putative cis-regulatory drivers in colorectal cancer.** *Nature* 2014, **512**:87-90.
47. Gjoneska E, Pfenning AR, Mathys H, Quon G, Kundaje A, Tsai LH, Kellis M: **Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease.** *Nature* 2015, **518**:365-369.
- Gene expression in human and mice with Alzheimer's disease, as well as enrichment of risk variants in tissue-specific enhancers supports a role for immune cells in this disease.
48. Cowper-Sallari R, Zhang X, Wright JB, Bailey SD, Cole MD, Eeckhoutte J, Moore JH, Lupien M: **Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression.** *Nat Genet* 2012, **44**:1191-1198.
49. Corradin O, Saiakhova A, Akhtar-Zaidi B, Myeroff L, Willis J, Cowper-Sallari R, Lupien M, Markowitz S, Scacheri PC: **Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits.** *Genome Res* 2014, **24**:1-13.
50. Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, Gomez-Marín C, Aneas I, Credidio FL, Sobreira DR, Wasserman NF *et al.*: **Obesity-associated variants within FTO form long-range functional connections with IRX3.** *Nature* 2014, **507**:371-375.
- This study employed transgenic enhancer assays, 3C studies, and eQTLs, to show that FTO, the major obesity susceptibility locus, contains regulatory elements that control IRX3 located nearly 2 Mb away. Loss of function experiments showed that IRX3 controls body weight.
51. Fischer J, Koch L, Emmerling C, Vierkotten J, Peters T, Bruning JC, Ruther U: **Inactivation of the Fto gene protects from obesity.** *Nature* 2009, **458**:894-898.
52. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JR, Elliott KS, Lango H, Rayner NW *et al.*: **A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity.** *Science* 2007, **316**:889-894.