Published in final edited form as:

Curr Opin Pediatr. 2015 December ; 27(6): 659–664. doi:10.1097/MOP.0000000000000283.

Mutations in the non-coding genome

Cheryl A. Scacheri¹ and Peter C. Scacheri^{2,*}

¹GeneDx, Gaithersburg, MD

²Department of Genetics and Genome Sciences, Case Comprehensive Cancer Center, Case Western Reserve University School of Medicine; Cleveland, OH

Abstract

Purpose of review—Clinical diagnostic sequencing currently focuses on identifying causal mutations in the exome, where most disease-causing mutations are known to occur. The rest of the genome is mostly comprised of regulatory elements that control gene expression, but these have remained largely unexplored in clinical diagnostics due to the high cost of whole genome sequencing and interpretive challenges. The purpose of this review is to illustrate examples of diseases caused by mutations in regulatory elements and introduce the diagnostic potential for whole genome sequencing. Different classes of functional elements and chromatin structure are described to provide the clinician with a foundation for understanding the basis of these mutations.

Recent Findings—The utilization of whole genome sequence data, epigenomic maps and iPS cell technologies facilitated the discovery that mutations in the PTF1A enhancer can cause isolated pancreatic agenesis. High resolution array CGH, whole genome sequencing, maps of 3-D chromatin architecture, and mouse models generated using CRISPR-Cas were used to show that disruption of topological-associated domain (TAD) boundary elements cause limb defects. Structural variants that reposition enhancers in somatic cells have also been described in cancer.

Summary—While not ready for diagnostics, new technologies, epigenomic maps and improved knowledge of chromatin architecture will soon enable a better understanding and diagnostic solutions for currently unexplained genetic disorders.

Keywords

whole genome sequencing; enhancer elements; regulatory mutations; chromatin; epigenomics

INTRODUCTION

Since 1956, when Ingram, Hunt and Stretton described an association between an amino acid substitution in the beta-globin gene and sickle cell disease [1], mutations in the protein coding regions of genes have been associated with human genetic disease. To date, pathogenic variants have been described in association with monogenic disorders in over

*Corresponding Author: Peter C. Scacheri, PhD, Associate Professor, Case Western Reserve University School of Medicine, Department of Genetics and Genome Sciences, 10900 Euclid Ave, Cleveland, OH 44122, pxs183@case.edu.

Conflicts of interest

Cheryl Scacheri is a paid employee of GeneDx, a for-profit company that specializes in diagnostic sequencing. The views and opinions expressed in this article are those of the authors and do not necessarily reflect those of GeneDx.

4,500 genes. Sequencing the coding region, or exome, of individuals suspected of having a genetic disorder identifies 25–50% of disease-associated mutations [2, 3]. This leaves the genetic etiology of many cases undetermined. While some of the failure to identify mutations in the coding region may be due to technological, interpretive or other limitations of exome sequencing, some of the causative variants probably lie outside of the coding region and are within regulatory and other regions of the genome.

There is a growing body of literature indicating that functional regulatory elements, like enhancers and insulators, in non-coding regions of the genome are associated with congenital anomalies (Table).

Somatic disruption of these regulatory elements has also been described in cancer. Further illustrating their importance, it is clear that DNA variants within enhancer elements genetically predispose to a variety of common and complex traits, including heart disease, diabetes, cancer, obesity, hair color and hundreds of others [16, 17]. Whereas disease-causing mutations in genes often disrupt amino acids or gene splicing and alter the function or levels of the protein product, the predominant view is that variants in regulatory elements cause abnormal phenotypes through anomalous gene expression.

The exome makes up less than two percent of the human genome and the rest, formerly known as “junk DNA,” is largely comprised of elements that regulate the timing, cell type, and levels of gene expression. Until recently, limitations in technology and a lack of knowledge about the specific locations of regulatory elements in DNA caused the field to focus on coding regions of the genome, where 85% of known disease-causing variants have been described [18]. With projects like ENCODE [19] and the Epigenomics Roadmap [16] coupled with advances in next generation sequencing, we now have epigenomic reference maps that have helped to delineate the genome-wide locations of regulatory elements in hundreds of human cell types. These maps allow the interrogation of the other 98% of the genome, which is as essential to gene function as coding regions.

In this review we provide cursory background information on the basics of functional elements and gene regulation. More detailed reviews are provided in the following references [20–25]. We then cite specific examples of diseases caused by mutations in non-coding regions, and describe the proposed pathogenic mechanism for each disorder. Lastly we discuss whether the time is right for clinical diagnostic sequencing to transition from sequencing the exome to the whole genome.

Functional Elements

Non-coding functional elements include promoters, enhancers, insulators, operons, and silencers; each of which plays a different role in regulating gene expression. As much as 80% of the human genome is comprised of these elements [19]. DNA mutations that disrupt the function of regulatory elements cause gene dysregulation, and a corresponding abnormal phenotype. The DNA changes associated with disease are similar to what has been described previously: translocations, deletions, duplications, insertions and point mutations. The impact of pathogenic DNA variants in functional elements is highly complex and the full picture has not yet emerged. We provide examples here to summarize what we do know,

with the knowledge that we are in the early stages of understanding the complexities of transcriptional regulation and human development.

The Complex Dance of Enhancers

Gene enhancer elements dictate which genes are expressed in a particular cell type, the timing of their expression and the levels of their expression. Our work on comparing enhancer elements between two closely related pluripotent cell types (embryonic and epiblast stem cells) has shown that there are major differences in the locations of enhancers despite the fact that these two cell types express a similar set of genes [26]. This demonstrates a fundamental principle in enhancer biology; that enhancers are dynamic elements in the genome, in contrast to the unchanging positions of genes. Each cell type contains 20–50,000 potential gene enhancer elements functioning within the three-dimensional context of chromatin [19]. While it seems intuitive that enhancers, like promoters, would be located close to the genes they regulate, this is not often the case. Enhancers can be located as much as a megabase or more upstream or downstream from the genes they control, and may even be on different chromosomes [27]. This is easier to understand when one considers the dramatically different configurations between an unwound strand of DNA and the tightly packaged DNA found in the cell nucleus. The packaging of DNA around histones and chromatin change the proximity of genes to one another. Adding to the complexity, a single enhancer can control multiple genes and many individual genes are regulated by more than one enhancer [27–30]. During development, the locations of enhancers change with regularity, driving gene expression programs specific to cell types and determining cellular identity [31, 32]. As such, individual cell types of the human body have distinct epigenomic landscapes due to the specificity of enhancers required to determine and maintain cellular identity.

Preaxial polydactyly

Well before the advances in technology that have enabled genome-wide identification of enhancers, preaxial polydactyly 2 (PPD2) was described in association with a mutation in an enhancer element [7, 8]. The first case described a 3 year old with PPD2 and a balanced translocation, the breakpoint of which was within intron 5 of the *LMBR1* gene. Extensive investigations of this and other patients, as well as murine models, demonstrated that intron 5 of the *LMBR1* gene contains the ZRS (zone of polarizing activity regulatory sequence) which harbors an enhancer element that regulates expression of the *SHH* gene [4]. While the *LMBR1* gene is not impacted by mutations in the ZRS, disruption of the enhancer element, which is nearly 1MB from the sonic hedgehog (*SHH*) gene, causes *SHH* to be misexpressed in the developing limb bud. Of particular interest, mutations in the protein coding region of the *SHH* gene cause holoprosencephaly [33, 34], a congenital disorder of the forebrain that does not involve the limbs. Following the findings of Lettice, other groups also have shown variants in the ZRS in patients with PPD2 [5, 6, 9], verifying that these mutations are causal and suggesting the potential for diagnostic relevance.

Pancreatic agenesis

Pancreatic and cerebellar agenesis (PACA) is a neonatal lethal disorder that has been described in families with recessive mutations in the *PTF1A* (pancreas-specific transcription

factor 1a) gene [35–38]. Isolated pancreatic agenesis (PAGEN2) is a relatively recent example that points to a tissue-specific congenital defect stemming from enhancer mutations [10]. The investigators' success in finding this association is due in no small part to advances in technology since the work from Lettice [7, 8], as well as the databases and tools that have been made available through groups like ENCODE and the Epigenetics road map. Taking a two-pronged approach, the investigators performed whole genome sequencing in individuals with pancreatic agenesis from two families with consanguineous parents. The authors also differentiated normal human ES cells into pancreatic endoderm and identified the locations of active enhancers. By overlaying these two data sets, they were able to identify the presence of homozygous variants in a pancreatic developmental enhancer located 25 kb downstream of the PTF1A gene within a 400-bp evolutionarily conserved region. They subsequently found that, in an additional ten patients with isolated pancreatic agenesis, seven had mutations in the pancreatic embryonic progenitor enhancers. The authors further reported that this enhancer was specific to early pancreatic development and was not observed in other cell types. This is consistent with the phenotype and also speaks to the tissue-specificity of enhancers during development. We believe that this strategy of combining genome sequencing and epigenomic profiling of relevant cell types will help identify the molecular etiology of disorders that have so far remained unexplained.

Pierre-Robin sequence

Mutations in the coding region of the *SOX9* gene are associated with campomelic dysplasia (CD) [39, 40], an often-lethal congenital malformation syndrome characterized by severe bowing of the long bones, respiratory insufficiency and abnormal male sexual differentiation. Patients with CD may also have Pierre-Robin sequence (PRS), a malformation of the mandible that results in micrognathia or retrognathia, glossoptosis and cleft palate. The *SOX9* gene is an HMG-box transcription factor active during the embryologic development of many diverse progenitor cells [41]. Its regulatory circuitry is likely to be highly complex. Mutations both upstream and downstream of the *SOX9* gene have been associated with isolated cases of PRS and it is proposed that these mutations lie in *SOX9* enhancers and are required for appropriate *SOX9* expression during development [11]. Recently a large 1Mb-sized deletion upstream of *SOX9* was found in patients with either isolated PRS, isolated congenital heart defect (CHD), or both of these defects from two families [12]. This large deletion includes enhancers that regulate *NKX2.5* and *GATA4*, genes known to be associated with CHD [42, 43]. Further studies are required to better understand the association between this large deletion, the associated phenotypes, and the genes they may dysregulate.

Other Congenital Disorders

There are other examples of diseases that may be due to mutations in enhancers. For example, mutations in the coding region and an enhancer of the *RET* gene have been associated with Hirschsprung's disease [13]. Rare, homozygous mutations in *TBX5* enhancers have been reported in isolated congenital heart malformations (<0.5% in Brazilian patients) [14].

Mutations that disrupt chromatin boundary elements can cause developmental anomalies

In addition to enhancer elements, the 3D-structure of chromatin is a key contributor to gene expression. Recent studies have helped to delineate the three dimensional structure of chromatin using genome-wide chromosome conformation capture methods such as Hi-C [44–46]. These studies have shown that the genome is partitioned into megabase-size topological associated domains, or TADs. TADs are regulatory domains that constrain the interactions between enhancers and promoters. TADs are separated by CTCF-bound boundary elements that block physical interactions between neighboring TADs. The locations of TADs are relatively consistent between different cell types and are evolutionarily conserved between mammals.

Congenital limb malformations were described in association with disruptions of TADs [15]. The investigators used high-resolution array CGH and whole genome sequencing to identify megabase-sized structural variants in patients with brachydactyly, F syndrome (syndactyly) and severe polydactyly and craniofacial abnormalities. The large structural variants they observed disrupt TADs that encompass four genes: WNT6, IHH, EPHA4, PAX3. The variants, an inversion, a deletion excising the EPHA4 gene, and a tandem duplication, force new interactions between the EPHA4 enhancers and genes within the locus that they do not normally regulate. These mutations lead to misexpression of the genes in these TADs, causing the observed limb malformations. The loss of the EPHA4 gene itself is not associated with limb malformations, but rather the misexpression of genes adjacent to EPHA4 that are ectopically activated upon disruption of the TAD boundary elements. These examples show that the integrity of the 3D structure of chromatin can cause birth defects. This discovery is particularly exciting because we know that structural abnormalities are common in individuals with abnormal phenotypes, yet when these are identified in gene deserts their significance is uncertain. By integrating the Hi-C maps of TADs and boundary elements with the locations of genes and regulatory elements, new light can be shed on the potential pathogenicity of structural variants.

Cancer

There are parallels between germline mutations in regulatory elements that cause congenital disorders and somatic mutations in cancer. It is well known that most cancers have numerous and diverse somatic mutations, some of which are recurrent. It has long been observed that in lymphoid malignancies, there are translocations that result in repositioning of the MYC oncogene near enhancers of the immunoglobulin genes (IgH/IgL) [47]. A more recent study looked at functional outcomes of recurrent chromosome 3q inversions and translocations in patients with AML (acute myelogenous leukemia) [48]. These structural rearrangements move the GATA2 enhancer to a new location near the EVI1 gene, causing a decrease in GATA2 expression and a concomitant increase in EVI1 expression. Excitingly, the uncontrolled growth of AML cells was reduced with the excision of the misplaced enhancer and return to the appropriate silencing of EVI1 expression, suggesting that dysregulation of enhancers can be reversed and providing a basis for therapeutic possibilities.

As a second example, we highlight elegant studies in medulloblastoma [49]. Here, different types of structural variations (tandem duplications, deletions, inversions and more complex

rearrangements) were observed at chromosome 9q34 in approximately 7% of cases. The structural variations result in repositioning of the growth factor independent I (GFI) family proto-oncogenes GFI1 and GFI1B near a super enhancer located in the vicinity of the DDX31 gene, which drives ectopic activation of the GFI oncogenes. The authors describe this concept of merging oncogenes with active enhancer elements as “enhancer hijacking” and speculated that this may reflect a general mechanism of oncogene activation in other forms of cancer.

With newly available datasets from The Cancer Genome Atlas and publicly available epigenome maps, studies have found recurrent single nucleotide mutations in enhancers [50], for which *in vitro* assays show a reduction in enhancer activity. We expect an explosion of new information around the effects of these regulatory mutations on gene dysregulation in cancer.

CONCLUSIONS AND DISCUSSION

To date, there are only a handful of examples where mutations in noncoding elements have been shown to alter gene expression and cause disease. Yet sequencing of the coding region identifies a genetic diagnosis in only 25–50% of individuals [2, 3]. While the diagnostic limitations of the current state of whole exome sequencing (WES) may account for some of the missing diagnoses, we believe that based on the observations presented above, a considerable proportion of patients will likely have disease-causing mutations outside the exome. As sequencing costs come down, whole genome sequencing will become a clinical option for patients who have tested negative by WES and may eventually become a first line diagnostic test. Aside from cost, however, the clinical application of genome sequencing will require a host of resources to appropriately interpret and classify variants.

Various levels of data analysis are required to evaluate variants. As with exomes, genome sequencing will require population data to determine variant frequencies. Given that noncoding variants cannot be annotated for changing an amino acid or predicted effects on a protein product, the relevant analytic tools for genome sequencing will integrate variants with epigenomic maps that show the locations of regulatory elements from relevant human cell types. In combination with genomic sequence data from a patient, one can see if a putative mutation lies within a regulatory element in a cell type relevant to the patient’s phenotype. It will be important to functionally validate the candidate associations, and one of the most robust methods currently being used is the generation of iPS cell technologies. The beauty of iPS technology is that the pluripotent stem cells can be differentiated into progenitor cells representative of the cell types affected in the patient’s disorder. This methodology is still challenging, however, and requires expertise at a level not readily accessible to many investigators. As we’ve seen, CRISPR-Cas strategies have been used to model the mutations in mice [15], and this is a powerful method for recapitulating the effects of enhancer variants on gene expression. A major limitation of using mouse models is that enhancers are poorly conserved between mice and humans. CRISPR-cas editing to introduce noncoding mutations into wild type cells and study their consequences on cellular function is another approach that utilizes model systems to study gene dysregulation.

Given that specific noncoding variants are likely to be rare, crowdsourcing and data sharing will be essential to look for the hot spots associated with disease. While several pieces need to fall into place for genome sequencing to be a reality in rare disease diagnostics, the realization of its clinical application will benefit patients in whom a diagnosis has been elusive.

Acknowledgments

Financial support and sponsorship

This research was supported by a grant from the National Cancer Institute (R01CA160356 to PCS).

References

1. Ingram VM. A specific chemical difference between the globins of normal human and sickle-cell anaemia haemoglobin. *Nature*. 1956; 178(4537):792–4. [PubMed: 13369537]
2. Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, Braxton A, Beuten J, Xia F, Niu Z, Hardison M, Person R, Bekheirnia MR, Leduc MS, Kirby A, Pham P, Scull J, Wang M, Ding Y, Plon SE, Lupski JR, Beaudet AL, Gibbs RA, Eng CM. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med*. 2013; 369(16):1502–11. [PubMed: 24088041]
3. Beaulieu CL, Majewski J, Schwartzentruber J, Samuels ME, Fernandez BA, Bernier FP, Brudno M, Knoppers B, Marcadier J, Dymont D, Adam S, Bulman DE, Jones SJ, Avard D, Nguyen MT, Rousseau F, Marshall C, Wintle RF, Shen Y, Scherer SW, Consortium FC, Friedman JM, Michaud JL, Boycott KM. FORGE Canada Consortium: outcomes of a 2-year national rare-disease gene-discovery project. *Am J Hum Genet*. 2014; 94(6):809–17. [PubMed: 24906018]
4. Furniss D, Lettice LA, Taylor IB, Critchley PS, Giele H, Hill RE, Wilkie AO. A variant in the sonic hedgehog regulatory sequence (ZRS) is associated with triphalangeal thumb and deregulates expression in the developing limb. *Hum Mol Genet*. 2008; 17(16):2417–23. [PubMed: 18463159]
5. Klopocki E, Ott CE, Benatar N, Ullmann R, Mundlos S, Lehmann K. A microduplication of the long range SHH limb regulator (ZRS) is associated with triphalangeal thumb-polysyndactyly syndrome. *J Med Genet*. 2008; 45(6):370–5. [PubMed: 18178630]
6. Laurell T, Vandermeer JE, Wenger AM, Grigelioniene G, Nordenskjold A, Arner M, Ekblom AG, Bejerano G, Ahituv N, Nordgren A. A novel 13 base pair insertion in the sonic hedgehog ZRS limb enhancer (ZRS/LMBR1) causes preaxial polydactyly with triphalangeal thumb. *Hum Mutat*. 2012; 33(7):1063–6. [PubMed: 22495965]
7. 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(14):1725–35. [PubMed: 12837695]
8. Lettice LA, Horikoshi T, Heaney SJ, van Baren MJ, van der Linde HC, Breedveld GJ, Joosse M, Akarsu N, Oostra BA, Endo N, Shibata M, Suzuki M, Takahashi E, Shinka T, Nakahori Y, Ayusawa D, Nakabayashi K, Scherer SW, Heutink P, Hill RE, Noji S. Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. *Proc Natl Acad Sci U S A*. 2002; 99(11):7548–53. [PubMed: 12032320]
9. Sun M, Ma F, Zeng X, Liu Q, Zhao XL, Wu FX, Wu GP, Zhang ZF, Gu B, Zhao YF, Tian SH, Lin B, Kong XY, Zhang XL, Yang W, Lo WH, Zhang X. Triphalangeal thumb-polysyndactyly syndrome and syndactyly type IV are caused by genomic duplications involving the long range, limb-specific SHH enhancer. *J Med Genet*. 2008; 45(9):589–95. [PubMed: 18417549]
- 10**. Weedon MN, Cebola I, Patch AM, Flanagan SE, De Franco E, Caswell R, Rodriguez-Segui SA, Shaw-Smith C, Cho CH, Lango Allen H, Houghton JA, Roth CL, Chen R, Hussain K, Marsh P, Vallier L, Murray A, Ellard S, Ferrer J, Hattersley AT. C International Pancreatic Agenesis. Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. *Nat Genet*. 2014; 46(1):61–4. This paper is the first to combine whole genome sequencing with epigenomic maps and iPS-cell technologies to identify a disease-causing mutation in an enhancer element. [PubMed: 24212882]

- 11*. Gordon CT, Attanasio C, Bhatia S, Benko S, Ansari M, Tan TY, Munnich A, Pennacchio LA, Abadie V, Temple IK, Goldenberg A, van Heyningen V, Amiel J, FitzPatrick D, Kleinjan DA, Visel A, Lyonnet S. Identification of novel craniofacial regulatory domains located far upstream of SOX9 and disrupted in Pierre Robin sequence. *Hum Mutat.* 2014; 35(8):1011–20. This paper demonstrates that enhancer mutations affecting SOX9 expression may be the basis of the Pierre Robin sequence. [PubMed: 24934569]
12. Sanchez-Castro M, Gordon CT, Petit F, Nord AS, Callier P, Andrieux J, Guerin P, Pichon O, David A, Abadie V, Bonnet D, Visel A, Pennacchio LA, Amiel J, Lyonnet S, Le Caignec C. Congenital heart defects in patients with deletions upstream of SOX9. *Hum Mutat.* 2013; 34(12):1628–31. [PubMed: 24115316]
13. Emison ES, McCallion AS, Kashuk CS, Bush RT, Grice E, Lin S, Portnoy ME, Cutler DJ, Green ED, Chakravarti A. A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk. *Nature.* 2005; 434(7035):857–63. [PubMed: 15829955]
14. Smemo S, Campos LC, Moskowitz IP, Krieger JE, Pereira AC, Nobrega MA. Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. *Hum Mol Genet.* 2012; 21(14):3255–63. [PubMed: 22543974]
- 15**. Lupianez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, Santos-Simarro F, Gilbert-Dussardier B, Wittler L, Borschiwer M, Haas SA, Osterwalder M, Franke M, Timmermann B, Hecht J, Spielmann M, Visel A, Mundlos S. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell.* 2015; 161(5):1012–25. This paper is the first to demonstrate that mutations disrupting TAD-boundaries can lead to ectopic gene activation and cause limb malformations. [PubMed: 25959774]
- 16**. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, Sarkar A, Quon G, Sandstrom RS, Eaton ML, Wu YC, Pfennig AR, Wang X, Claussnitzer M, Liu Y, Coarfa C, Harris RA, Shores N, Epstein CB, Gjoneska E, Leung D, Xie W, Hawkins RD, Lister R, Hong C, Gascard P, Mungall AJ, Moore R, Chuah E, Tam A, Canfield TK, Hansen RS, Kaul R, Sabo PJ, Bansal MS, Carles A, Dixon JR, Farh KH, Feizi S, Karlic R, Kim AR, Kulkarni A, Li D, Lowdon R, Elliott G, Mercer TR, Neph SJ, Onuchic V, Polak P, Rajagopal N, Ray P, Sallari RC, Siebenthal KT, Sinnott-Armstrong NA, Stevens M, Thurman RE, Wu J, Zhang B, Zhou X, Beaudet AE, Boyer LA, De Jager PL, Farnham PJ, Fisher SJ, Haussler D, Jones SJ, Li W, Marra MA, McManus MT, Sunyaev S, Thomson JA, Tlsty TD, Tsai LH, Wang W, Waterland RA, Zhang MQ, Chadwick LH, Bernstein BE, Costello JF, Ecker JR, Hirst M, Meissner A, Milosavljevic A, Ren B, Stamatoyanopoulos JA, Wang T, Kellis M. Roadmap Epigenomics. C. Integrative analysis of 111 reference human epigenomes. *Nature.* 2015; 518(7539):317–30. This paper is from the Epigenomics Roadmap Consortium and maps the locations of regulatory elements in 111 human cell types. These references map will be useful for determining if candidate disease variants identified through whole genome sequencing studies potentially disrupt regulatory elements in disease-relevant cell types. [PubMed: 25693563]
17. Corradin O, Scacheri PC. Enhancer variants: evaluating functions in common disease. *Genome Med.* 2014; 6(10):85. [PubMed: 25473424]
18. Cooper, DN.; Krawczak, M.; Antonarakis, SE. The nature and mechanisms of human gene mutation. In: Scriver, CR., et al., editors. *The Metabolic and Molecular Bases of Inherited Disease.* McGraw-Hill; New York: 1995. p. 259-291.
19. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012; 489(7414):57–74. [PubMed: 22955616]
20. Cubenas-Potts C, Corces VG. Architectural proteins, transcription, and the three-dimensional organization of the genome. *FEBS Lett.* 2015
21. Bulger M, Groudine M. Functional and mechanistic diversity of distal transcription enhancers. *Cell.* 2011; 144(3):327–39. [PubMed: 21295696]
22. Zentner GE, Scacheri PC. The chromatin fingerprint of gene enhancer elements. *J Biol Chem.* 2012; 287(37):30888–96. [PubMed: 22952241]
23. Lee TI, Young RA. Transcriptional regulation and its misregulation in disease. *Cell.* 2013; 152(6):1237–51. [PubMed: 23498934]

24. Newman JJ, Young RA. Connecting transcriptional control to chromosome structure and human disease. *Cold Spring Harb Symp Quant Biol.* 2010; 75:227–35. [PubMed: 21209391]
25. Maston GA, van SKE, Green MR. Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet.* 2006; 7:29–59. [PubMed: 16719718]
26. Factor DC, Corradin O, Zentner GE, Saiakhova A, Song L, Chenoweth JG, McKay RD, Crawford GE, Scacheri PC, Tesar PJ. Epigenomic comparison reveals activation of “seed” enhancers during transition from naive to primed pluripotency. *Cell Stem Cell.* 2014; 14(6):854–63. [PubMed: 24905169]
27. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T, Ntini E, Arner E, Valen E, Li K, Schwarzfischer L, Glatz D, Raithel J, Lilje B, Rapin N, Bagger FO, Jorgensen M, Andersen PR, Bertin N, Rackham O, Burroughs AM, Baillie JK, Ishizu Y, Shimizu Y, Furuhashi E, Maeda S, Negishi Y, Mungall CJ, Meehan TF, Lassmann T, Itoh M, Kawaji H, Kondo N, Kawai J, Lennartsson A, Daub CO, Heutink P, Hume DA, Jensen TH, Suzuki H, Hayashizaki Y, Muller F, Consortium F, Forrest AR, Carninci P, Rehli M, Sandelin A. An atlas of active enhancers across human cell types and tissues. *Nature.* 2014; 507(7493):455–61. [PubMed: 24670763]
28. 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):1–13. [PubMed: 24196873]
29. Link N, Kurtz P, O’Neal M, Garcia-Hughes G, Abrams JM. A p53 enhancer region regulates target genes through chromatin conformations in cis and in trans. *Genes Dev.* 2013; 27(22):2433–8. [PubMed: 24240233]
30. He B, Chen C, Teng L, Tan K. Global view of enhancer-promoter interactome in human cells. *Proc Natl Acad Sci U S A.* 2014; 111(21):E2191–9. [PubMed: 24821768]
31. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, Ching KA, Antosiewicz-Bourget JE, Liu H, Zhang X, Green RD, Lobanov VV, Stewart R, Thomson JA, Crawford GE, Kellis M, Ren B. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature.* 2009; 459(7243):108–12. [PubMed: 19295514]
32. Gifford CA, Ziller MJ, Gu H, Trapnell C, Donaghey J, Tsankov A, Shalek AK, Kelley DR, Shishkin AA, Issner R, Zhang X, Coyne M, Fostel JL, Holmes L, Meldrum J, Guttman M, Epstein C, Park H, Kohlbacher O, Rinn J, Gnirke A, Lander ES, Bernstein BE, Meissner A. Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell.* 2013; 153(5):1149–63. [PubMed: 23664763]
33. Roessler E, Belloni E, Gaudenz K, Jay P, Berta P, Scherer SW, Tsui LC, Muenke M. Mutations in the human Sonic Hedgehog gene cause holoprosencephaly. *Nat Genet.* 1996; 14(3):357–60. [PubMed: 8896572]
34. Belloni E, Muenke M, Roessler E, Traverso G, Siegel-Bartelt J, Frumkin A, Mitchell HF, Donis-Keller H, Helms C, Hing AV, Heng HH, Koop B, Martindale D, Rommens JM, Tsui LC, Scherer SW. Identification of Sonic hedgehog as a candidate gene responsible for holoprosencephaly. *Nat Genet.* 1996; 14(3):353–6. [PubMed: 8896571]
35. Sellick GS, Barker KT, Stolte-Dijkstra I, Fleischmann C, Coleman RJ, Garrett C, Gloyn AL, Edghill EL, Hattersley AT, Wellauer PK, Goodwin G, Houlston RS. Mutations in PTF1A cause pancreatic and cerebellar agenesis. *Nat Genet.* 2004; 36(12):1301–5. [PubMed: 15543146]
36. Tutak E, Satar M, Yapicioglu H, Altintas A, Narli N, Herguner O, Bayram Y. A Turkish newborn infant with cerebellar agenesis/neonatal diabetes mellitus and PTF1A mutation. *Genet Couns.* 2009; 20(2):147–52. [PubMed: 19650412]
37. Al-Shammari M, Al-Husain M, Al-Kharfy T, Alkuraya FS. A novel PTF1A mutation in a patient with severe pancreatic and cerebellar involvement. *Clin Genet.* 2011; 80(2):196–8. [PubMed: 21749365]
38. Stoffers DA, Zinkin NT, Stanojevic V, Clarke WL, Habener JF. Pancreatic agenesis attributable to a single nucleotide deletion in the human IPF1 gene coding sequence. *Nat Genet.* 1997; 15(1):106–10. [PubMed: 8988180]

39. Wagner T, Wirth J, Meyer J, Zabel B, Held M, Zimmer J, Pasantes J, Bricarelli FD, Keutel J, Hustert E, Wolf U, Tommerup N, Schempp W, Scherer G. Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9. *Cell*. 1994; 79(6): 1111–20. [PubMed: 8001137]
40. Foster JW, Dominguez-Steglich MA, Guioli S, Kwok C, Weller PA, Stevanovic M, Weissenbach J, Mansour S, Young ID, Goodfellow PN, et al. Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. *Nature*. 1994; 372(6506):525–30. [PubMed: 7990924]
41. Pritchett J, Athwal V, Roberts N, Hanley NA, Hanley KP. Understanding the role of SOX9 in acquired diseases: lessons from development. *Trends Mol Med*. 2011; 17(3):166–74. [PubMed: 21237710]
42. McElhinney DB, Geiger E, Blinder J, Benson DW, Goldmuntz E. NKX2.5 mutations in patients with congenital heart disease. *J Am Coll Cardiol*. 2003; 42(9):1650–5. [PubMed: 14607454]
43. Garg V I, Kathirya S, Barnes R, Schluterman MK, King IN, Butler CA, Rothrock CR, Eapen RS, Hirayama-Yamada K, Joo K, Matsuoka R, Cohen JC, Srivastava D. GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature*. 2003; 424(6947): 443–7. [PubMed: 12845333]
44. 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
45. 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(7398): 376–80. [PubMed: 22495300]
46. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014; 159(7):1665–80. [PubMed: 25497547]
47. Adams JM, Harris AW, Pinkert CA, Corcoran LM, Alexander WS, Cory S, Palmiter RD, Brinster RL. The c-myc oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice. *Nature*. 1985; 318(6046):533–8. [PubMed: 3906410]
- 48*. Groschel S, Sanders MA, Hoogenboezem R, de Wit E, Bouwman BA, Erpelinck C, van der Velden VH, Havermans M, Avellino R, van Lom K, Rombouts EJ, van Duin M, Dohner K, Beverloo HB, Bradner JE, Dohner H, Lowenberg B, Valk PJ, Bindels EM, de Laat W, Delwel R. A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell*. 2014; 157(2):369–81. [PubMed: 24703711]
- 49*. Northcott PA, Lee C, Zichner T, Stutz AM, Erkek S, Kawauchi D, Shih DJ, Hovestadt V, Zapatka M, Sturm D, Jones DT, Kool M, Remke M, Cavalli FM, Zuyderduyn S, Bader GD, VandenBerg S, Esparza LA, Ryzhova M, Wang W, Wittmann A, Stark S, Sieber L, Seker-Cin H, Linke L, Kratochwil F, Jager N, Buchhalter I, Imbusch CD, Zipprich G, Raeder B, Schmidt S, Diessl N, Wolf S, Wiemann S, Brors B, Lawerenz C, Eils J, Warnatz HJ, Risch T, Yaspo ML, Weber UD, Bartholomae CC, von Kalle C, Turanyi E, Hauser P, Sanden E, Darabi A, Siesjo P, Sterba J, Zitterbart K, Sumerauer D, van Sluis P, Versteeg R, Volckmann R, Koster J, Schuhmann MU, Ebinger M, Grimes HL, Robinson GW, Gajjar A, Mynarek M, von Hoff K, Rutkowski S, Pietsch T, Scheurlen W, Felsberg J, Reifemberger G, Kulozik AE, von Deimling A, Witt O, Eils R, Gilbertson RJ, Korshunov A, Taylor MD, Lichter P, Korbel JO, Wechsler-Reya RJ, Pfister SM. Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature*. 2014; 511(7510):428–34. This paper, and the Groschel et al paper above, provide elegant illustrations of how somatic translocations in cancer can reposition enhancers near oncogenes that drive cancer. [PubMed: 25043047]
50. Melton C, Reuter JA, Spacek DV, Snyder M. Recurrent somatic mutations in regulatory regions of human cancer genomes. *Nat Genet*. 2015; 47(7):710–6. [PubMed: 26053494]

Key Points

- Exome sequencing identifies 25–50% of disease-associated mutations
- Mutations disrupting enhancer elements and chromatin boundary elements located outside the exome can cause congenital malformations
- Somatic mutations disrupting regulatory elements can activate oncogenes in cancer
- Studies that integrate whole genome sequencing with maps of chromatin landscapes can facilitate discovery of noncoding regulatory mutations.

Table

Examples of genetic conditions caused by mutations outside the exome

<u>Genetic Disease/Condition</u>	<u>Mutation</u>	<u>References</u>
Preaxial polydactyly 2 (PPD2)	SHH enhancer	[4–9]
Pancreatic agenesis	PTF1A enhancer	[10]
Pierre Robin Sequence (PRS)	SOX9 enhancer	[11, 12]
Hirschsprung disease	RET enhancer	[13]
Isolated congenital heart defect	TBX5 enhancer	[14]
F-syndrome, Polydactyly, Brachydactyly	Disruption of TAD boundary elements near WNT6, IHH, EPHA4, and PAX3	[15]