Trends in Genetics

CellPress
REVIEWS

Review

The Role of Noncoding Variants in Heritable Disease

J.D. French[1,2,*,@] and S.L. Edwards[1,3,*,@]

The genetic basis of disease has largely focused on coding regions. However, it has become clear that a large proportion of the noncoding genome is functional and harbors genetic variants that contribute to disease etiology. Here, we review recent examples of inherited noncoding alterations that are responsible for Mendelian disorders or act to influence complex traits. We explore both rare and common genetic variants and discuss the wide range of mechanisms by which they affect gene regulation to promote disease. We also debate the challenges and progress associated with identifying and interpreting the functional and clinical significance of genetic variation in the context of the noncoding regulatory landscape.

Highlights

A large proportion of the noncoding genome is functional and can harbor genetic variants that are causally related to Mendelian disorders or act to influence complex traits.

Noncoding variants can influence gene expression or function through diverse functional units including untranslated regions, DNA regulatory elements, and noncoding RNAs.

The current challenges for the interpretation of noncoding genetic variants include identifying the disease-relevant cell type, predicting and evaluating the underlying mechanisms, and the lack of appropriate *in vivo* models.

A comprehensive understanding of the mechanisms and impact of noncoding genetic variation in human disease will assist future genetic diagnoses and selective targeting as a therapeutic strategy.

Functional Elements in Noncoding Regions

Until recently, the search for disease-causing **genetic variation** (see Glossary) has predominantly focused on protein-coding regions of the genome. However, a large proportion of the noncoding genome is functional and, similar to protein-coding regions, harbors genetic variants that are causally related to human disease. Substantial progress has been made toward defining the functional elements in the noncoding genome (Figure 1, Key Figure). The 5′ and 3′ **untranslated regions (UTRs)** of **mRNAs** together with **introns** account for up to ~35% of the human genome [1], with another ~50% comprising **transposable elements** and **tandem repeats** [2]. ***Cis*-regulatory elements** such as promoters and enhancers are also scattered across the noncoding genome and typically regulate gene transcription through the binding of **transcription factors (TFs)**. The noncoding genome is also extensively transcribed [3–5], producing thousands of **noncoding RNAs (ncRNAs)** including short ncRNAs (<200 nucleotides) and **long ncRNAs (lncRNAs)** (>200 nucleotides). On a larger scale, the human genome can be partitioned into megabase-sized **topologically associating domains (TADs)** that preferentially interact with each other rather than with regions outside (reviewed in [6]) and are critical for gene regulation [7–9]. They can contain part of one large gene or clusters of genes that are more likely to be coregulated than genes not clustered in TADs [10]. In this review, we focus on recent examples of genetic variation in these noncoding functional elements, their relevant underlying molecular mechanisms, and their contribution to Mendelian or complex disease.

Genetic Variants That Alter RNA Splicing

Pre-mRNA splicing is a complex process by which the noncoding introns are removed and the **exons** aligned and ligated to form mRNA. Several signals exist in introns that are critical for the pre-mRNA splicing process, including donor (5′) and acceptor (3′) splice sites, branch points, and **polypyrimidine tract** sequences. In addition to these canonical splice sites, sequences within exon and introns can act as splicing enhancer and silencer elements [11].These elements interact with splicing regulators in a tissue-specific and context-specific manner and contribute to the complex repertoire of alternative splicing in the human transcriptome [11]. Splicing abnormalities such as exon skipping and intron retention account for up to 15% of all inherited diseases

[1]QIMR Berghofer Medical Research Institute, Brisbane, QLD 4029, Australia
[2]https://www.qimrberghofer.edu.au/lab/functional-genetics/
[3]https://www.qimrberghofer.edu.au/lab/functional-cancer-genomics/

*Correspondence:
juliet.french@qimrberghofer.edu.au (J.D. French) and
stacey.edwards@qimrberghofer.edu.au (S.L. Edwards).
@Twitter: @JulietFrench01 (J.D. French) and @Edwards_Stace (S.L. Edwards).

 https://doi.org/10.1016/j.tig.2020.07.004

Key Figure

Overview of the Functional Elements in the Human Noncoding Genome

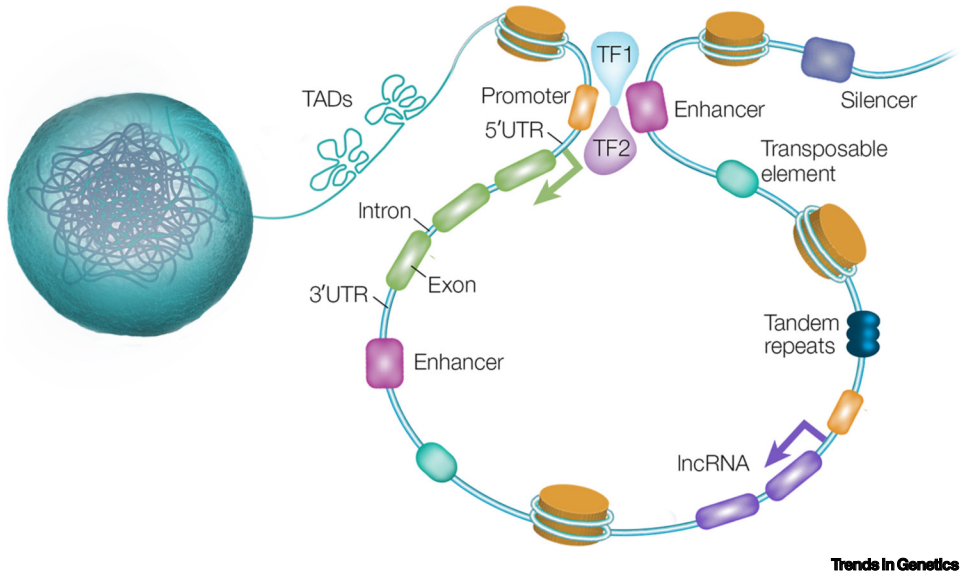

Figure 1. Chromosomes are partitioned into topologically associating domains (TADs) corresponding to domains of highly interacting chromatin. In TADs, DNA is complexed with histones to form nucleosomes (orange circles); active promoters (orange rectangles) and enhancer (pink rectangles) or silencer (dark blue rectangles) elements can form chromatin loops, mediated by protein effectors [transcription factor (TF)1, TF2]. Transcription of protein-coding genes (green blocks) or noncoding RNA genes (lncRNA; purple blocks) is shown as colored arrows. The generic protein-coding gene structure includes the flanking untranslated regions (UTRs) and alternating regions of introns (noncoding sequences) and exons (coding sequences). Transposable elements (aqua ovals) and tandem repeats (dark blue ovals) are repetitive DNA sequences that constitute ~50% of the human genome.

(reviewed in [12,13]). Recently, Jamshidi *et al.* identified seven families with inherited early-onset retinal degenerations (IRDs) with only one loss-of-function mutation in an IRD-associated gene, *RPGRIP1* [14]. Additional whole-genome sequencing (WGS) identified intronic mutations in six families as the second mutation in *RPGRIP1*, three of which resulted in intron retention and the creation of premature stop codons, which is likely to lead to nonsense-mediated decay of the aberrant transcripts [14] (Figure 2A). Transcriptome and genome *de novo* assembly has also identified a SINE-VNTR-Alu (SVA)-type retrotransposon insertion in intron 32 of *TAF1* causing aberrant splicing and reduced *TAF1* expression, leading to X-linked dystonia-Parkinsonism [15]. Furthermore, intronic deletions have been identified in two families with severe limb-girdle muscle weakness or neuromuscular disease. These deletions shorten the length between the 5′ splice site and branchpoint below a critical minimum length, leading to abnormal splicing [16]. Data mining identified a further 23 families with pathogenicity likely to be due to intronic deletions and abnormal splicing [16], suggesting that this mechanism maybe relevant for genetic diagnosis across a wide range of Mendelian disorders.

Variation in alternative spicing between individuals is also widespread and associated with common genetic variation [17]. **Splicing quantitative trait locus (sQTL)** analysis is now commonly used to identify these associations [18]. sQTLs are enriched in the canonical 5′ and 3′ splice sites [19], but are also found in nearby introns and exons and are likely to modify DNA elements that are required for splicing. Notably, sQTLs are also enriched for **genome-wide association study**

Glossary

Chromatin conformation capture (3C): a method that measures the frequency with which two DNA fragments interact within the 3D nuclear space.
***Cis*-regulatory elements:** noncoding DNA sequences that regulate the transcription of nearby genes.
ClinVar: a freely accessible public archive of reports that list information about genomic variation and human health.
CRISPR: (clustered regularly interspaced short palindromic repeats) a technology that allows precise editing of DNA in the genome.
Exons: nucleotide sequences that code information for protein synthesis.
Expression quantitative trait loci (eQTLs): associations between genetic variants and gene expression.
Genetic variation: includes SNPs, small indels (fewer than 50 base pairs in length), and large structural variants. Structural variants (greater than 50 base pairs) can be copy-number variations, such as deletions and duplications, copy-number neutral, such as inversions and translocations, or often a combination of these.
Genome-wide association studies (GWASs): used to identify common genetic variants in different individuals associated with a particular disease or trait.
Hi-C: a high-throughput 3C method that quantifies chromatin interactions between all possible pairs of DNA fragments simultaneously.
Introns: nucleotide sequences within a gene that do not code information for protein synthesis.
Long noncoding RNAs (lncRNAs): a class of regulatory ncRNAs (>200 nucleotides) that are involved in diverse cellular functions.
miRNAs: a class of small ncRNAs (20–25 nucleotides) that inhibit target gene expression by binding to the 3′UTRs of target mRNAs to induce mRNA degradation and translational repression.
mRNA: a single-stranded RNA molecule transcribed from the DNA of a gene that carries the genetic information needed to make proteins.
Noncoding RNAs (ncRNAs): RNA molecules that are transcribed from DNA but not translated into proteins.
Polygenic risk score (PRS): a computationally derived number that

(GWAS) variants [18–21] and have similar or even larger effect sizes than **expression quantitative trait loci (eQTLs)** [20], suggesting that variation in alternative splicing contributes substantially to complex disease traits. There are several studies that provide experimental evidence to support this premise. In one recent example, Li *et al.* identified schizophrenia risk variants associated with increased expression of a new *AS3MT* isoform missing exons 2 and 3 (named *AS3MT*^{d2d3}) [22]. *AS3MT*^{d2d3} was induced in early neuronal differentiation and significantly increased in brain tissue from individuals with schizophrenia or major depression compared with healthy controls. Moreover, Kim *et al.* used allele-specific expression analyses to detect an alternatively spliced *SLC22A1* isoform related to metabolic disease [23]. The SNP rs113569197 induced incorrect splicing of *SLC22A1*, which accelerated transcript degradation through nonsense-mediated decay, and impaired the SLC22A1-specific efflux activity. Together, these studies demonstrate that deregulated RNA splicing is a prominent mechanism underlying single-gene and complex disorders and contributes to both disease onset and severity.

Genetic Variants That Alter UTRs

The 5′ and 3′UTRs are the mRNA sequences flanking the beginning and end of the coding sequence. UTRs are transcribed into mRNA but not translated into proteins. UTRs play crucial roles in post-transcriptional gene regulatory processes including mRNA stability, secondary structure, localization, and translation through interactions with **RNA-binding proteins (RBPs)** and **miRNAs**. The importance of UTRs in regulating gene expression is highlighted by the discovery that 5′ and/or 3′UTR mutations can lead to disease (reviewed in [24]). A few recent examples include a 3′UTR mutation in *SLC4A4*, which contributes to a diverse array of ocular phenotypes [25]. This mutation creates a functional AU-rich element that leads to reduced *SLC4A4* mRNA levels, probably through binding of the ZFP36 RBP promoting RNA decay (Figure 2B). Furthermore, 5′ and 3′UTR mutations in *GJB1*, encoding the connexin 32 (Cx32) ion channel protein, are likely to cause X-linked Charcot–Marie–Tooth disease [26]. Approximately 3.7% of disease-associated variants in the GWAS catalog also localize to the UTRs of protein-coding genes [24], but few examples have experimentally confirmed a functional effect. In one example, a variant in the 3′UTR in *TNFSF13B* associated with risk of multiple sclerosis (MS) and systemic lupus erythematosus creates a premature polyadenylation signal leading to a shorter 3′UTR that is no longer regulated by miR-15a [27] (Figure 2B). This leads to reduced *TNFSF13B* mRNA translation resulting in lower levels of the encoded B cell-activating factor protein [27]. These examples provide evidence that UTR variants can impact both post-transcriptional and translational processes, resulting in altered mRNA and/or protein levels, which can drive or enhance diverse disease phenotypes.

Genetic Variants That Alter Promoters

Promoter sequences are typically located less than 0.5 kb from transcription start sites and recruit TFs and RNA polymerase II to initiate transcription (reviewed in [28]). Only a few pathogenic/likely pathogenic variants affecting promoter regions have been identified in Mendelian disorders, but, given that there is significant ascertainment bias, their prevalence is likely to be underestimated. Examples include four different mutations in the *OVOL2* proximal promoter that altered TF binding sites to increase *OVOL2* expression, causing a spectrum of phenotypes in over 100 affected individuals with corneal endothelial dystrophies [29] (Figure 2C). Promoter mutations have also been identified in cancer predisposition genes; recent examples include gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS), where mutations in an alternative promoter of *APC* disrupt the binding of YY1 and reduce *APC 1B* promoter activity in human gastric cell lines [30]. Notably, the canonical *APC 1A* promoter is methylated and not active in the normal stomach, suggesting that *APC 1B* is the dominant transcript in the gastric mucosa [30]. Mutations in the *TERT* promoter are also found in familial melanoma that alter TF binding leading to

reflects a person's inherited risk for a particular disease or trait.
Polypyrimidine tract: a pyrimidine-rich region of the pre-mRNA that directs intron removal in pre-mRNA splicing.
RNA-binding proteins (RBPs): proteins that bind to RNA in cells and change the fate or function of the bound RNAs.
Small nuclear RNAs (snRNAs): a class of small ncRNAs (150 nucleotides) involved in the splicing of introns from pre-mRNA in the nucleus.
Splicing quantitative trait loci (sQTLs): associations between genetic variants and alternative splicing of pre-mRNAs.
Tandem repeats: short DNA sequences (usually one to six nucleotides in length) that are repeated multiple times and lie adjacent to each other on the chromosome.
Topologically associating domains (TADs): defined as local domains within which genomic loci interact more frequently with one another than with surrounding regions.
Transcription factors (TFs): sequence-specific DNA-binding proteins that help to regulate gene transcription.
Transposable elements: DNA sequences that move from one location on the genome to another.
Untranslated regions (UTRs): regulatory regions of DNA located at the 5′ and 3′ ends of genes that are transcribed into mRNA but not translated into protein.

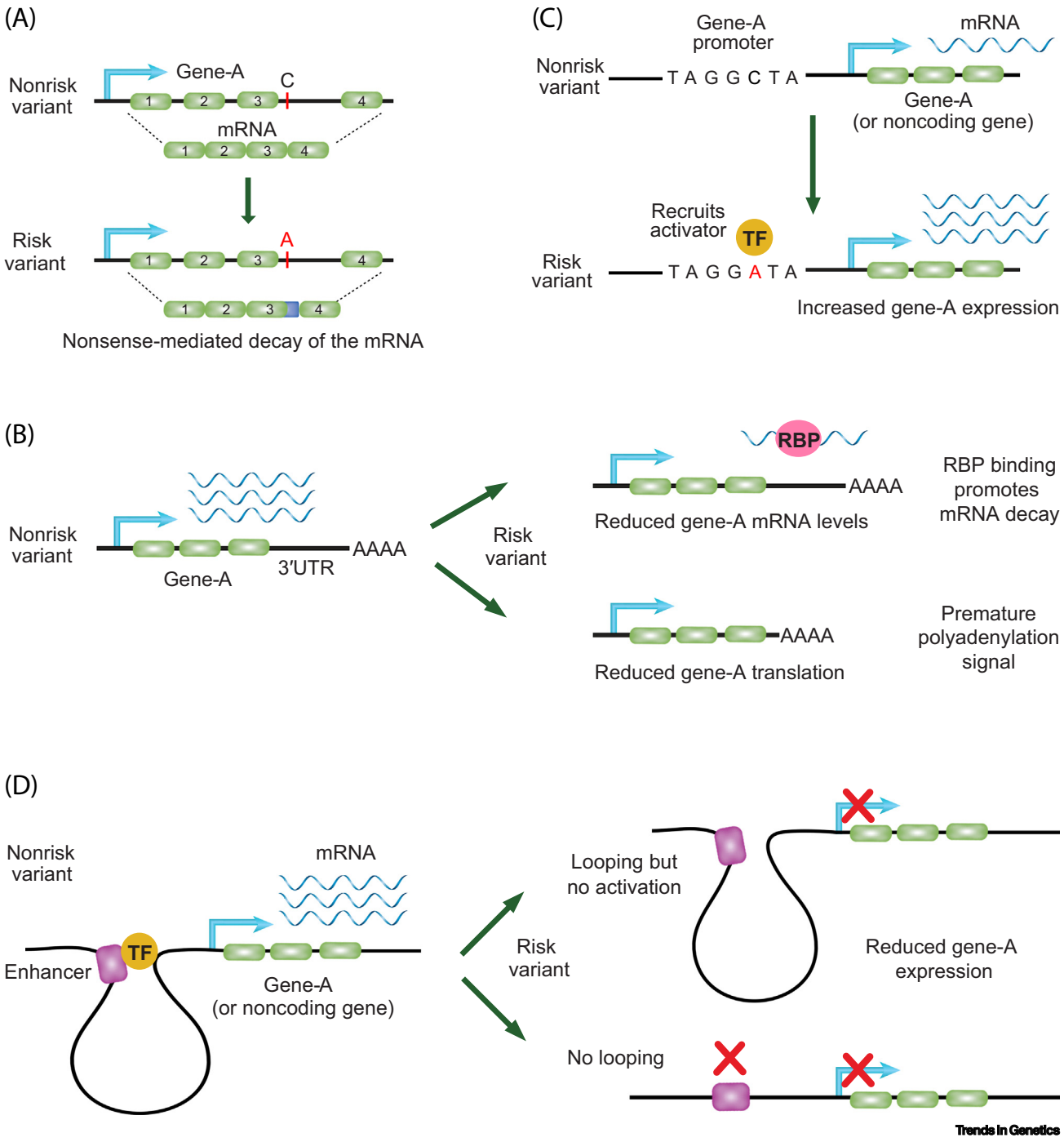

Figure 2. Examples of How Noncoding Variants Can Impact RNA Splicing, Untranslated Regions (UTRs), Promoters, and DNA Regulatory Elements. (A) RNA splicing. Risk variants can activate cryptic slice acceptor sites resulting in exon retention during pre-mRNA splicing (purple exon), creating premature stop codons leading to nonsense-mediated decay of the aberrant transcript. (B) UTRs. Risk variants in 3′UTRs can create an RNA-binding protein (RBP) recognition site that may promote mRNA degradation or produce an alternative polyadenylation signal leading to changes in mRNA stability or translation efficiency. (C) Promoters. Risk variants in protein-coding or noncoding gene promoter regions can lead to the recruitment of additional transcription factors (TFs) that activate transcription. (D) DNA regulatory elements. Risk variants can alter TF binding, reduce trans-activation of the target gene promoter, or reduce/ablate chromatin looping between the enhancer and the target gene promoter.

increased *TERT* promoter activity in human melanoma cell lines [31,32]. Common disease-associated variants are also found in gene promoter regions. For example, Tian *et al.* identified two SNPs located in the *ATF1* promoter and first intron associated with colorectal cancer (CRC) risk [33]. The SNPs facilitated a promoter–enhancer interaction, mediated by the SP1 and GATA3 TFs, that upregulated ATF1, which inhibited cell apoptosis and correlated with early onset of CRC. Moreover, breast cancer GWASs have identified SNPs in promoter regions of multiple genes, including *TERT*, *KLHDC7A*, *PIDD1*, and *ESR1*, with reporter assays showing that independent risk alleles alter target promoter activities [34–36]. Collectively, the majority of identified promoter variants mediate their regulatory effects through alteration of TF binding. However, it is established that coordinated transcription requires the interplay between gene promoter regions and more distal *cis*-regulatory DNA sequences.

Genetic Variants That Alter Distal DNA Regulatory Elements

Cis-regulatory elements such as enhancers and silencers are DNA sequences scattered across noncoding regions that bind proteins such as TFs and either enhance or silence gene transcription. These elements typically interact and regulate promoters (up to 1 Mb away) through long-range chromatin interactions (reviewed in [37]). For Mendelian disorders, only a few mutations have been identified in *cis*-regulatory elements, but this may reflect a historical focus on protein-coding genes and the difficulty in assessing the function of these variants. One recent study investigated the effects of copy number variation on enhancers at the *IHH* locus that caused syndactyly and craniosynostosis. They showed through transgenic reporter and **CRISPR** editing studies in mice that *Ihh* is regulated by at least nine tissue-specific enhancers, and duplications cause dose-dependent upregulation and misexpression of *Ihh*, leading to syndactyly [38]. Another study used high throughput *in vivo* transgenic LacZ reporter assays in mice to assess 21 reported mutations in a limb-specific enhancer linked to polydactyly [39]. Seventy-one percent (15/21) of the mutations showed increased enhancer activity in the anterior limb bud of the hind- or forelimbs, consistent with a pathogenic role in polydactyly, a limb-specific phenotype. Knock-in mice carrying the human-equivalent mutations confirmed that these mutations cause abnormal limb development [39]. Furthermore, Soukup *et al.* developed a mouse model to assess mutations in an enhancer of GATA2 located 9.5 kb downstream of the GATA2 transcription start site (+9.5 enhancer) that cause human GATA2 deficiency syndrome [40]. Notably, knock in of the most common noncoding *GATA2* mutation resulted in viable mice with normal steady-state hematopoiesis. However, myeloablation-induced stress caused by 5-fluorouracil (5-FU) led to abrogated stem/progenitor regeneration in mice, providing a new mechanism of hematopoietic failure caused by mutational sensitization to stress [40].

A prominent mechanism by which GWAS variants affect cell-type-specific enhancer function is through altered TF binding, which impacts target gene expression (reviewed in [41,42]) (Figure 2D). Genetic variation in enhancers often modulate the expression of the nearest gene [43–45], but there are many examples where enhancers bypass the nearest promoter to regulate a more distal gene [46,47]. In one recent example, Soldner *et al.* identified a Parkinson's disease-associated SNP in a distal enhancer that regulates *SNCA* (encoding α-synuclein) [43]. CRISPR editing of human embryonic stem cells showed that insertion of the SNP risk allele reduced brain-specific TF binding, leading to increased enhancer activity and higher *SNCA* expression [43]. Moreover, two elegant studies using a combination of human cell lines, primary human adipocytes, and mouse models showed that an enhancer containing obesity-associated variants located in the intron of *FTO* does not regulate *FTO* expression but loops to and regulates two distal genes, *IRX3* and *IRX5*, located ~490 kb and ~1.1 Mb away, respectively [48,49]. Increased expression of *IRX3* and *IRX5* during early adipocyte differentiation could lead to a cell-autonomous shift from thermogenic brown adipocytes to energy-storing white adipocytes, resulting in weight

gain [49]. As enhancers can physically interact with multiple target gene promoters [50,51], disease-associated variants in enhancers can also impact the expression of more than one gene. As an example, Kycia *et al.* recently identified a type 2 diabetes risk SNP that maps to an enhancer and increased the expression of *C2CD4B* and *C2CD4A,* both of which are induced by inflammatory cytokines in human pancreatic islet and β-cells [52].

While not as widely reported, some studies have demonstrated that common variants can modify chromatin contact frequencies between enhancers and their target gene promoters, thus impacting gene expression and disease risk (Figure 2D). **Chromatin conformation capture (3C)**-based methods are typically used to identify allele-specific interactions using cell lines or tissues heterozygous for the disease-associated haplotype [35,47,53–56]. For example, genotype-specific chromatin interaction assays showed preferential looping between the SNP rs6927172 risk allele and the *IL20RA* promoter in human lymphoblastoid cell lines, which correlated with NF-κB TF binding and increased gene expression [47]. Other studies have also demonstrated that risk SNPs affect chromatin looping and gene expression at breast cancer and asthma risk regions [35,53,54]. Notably, two recent studies have provided evidence that allelic imbalance in chromatin looping is likely to occur throughout the genome. Gorkin *et al.* performed **Hi-C** on human lymphoblastoid cell lines from 20 individuals and found that genetic variants influenced the loop strength and density of local *cis* contacts, which in some cases led to altered gene expression [55]. Similarly, Greenwald *et al.* combined Hi-C and phenotype analyses in induced pluripotent stem cells from seven individuals. They observed that genetic variation can exert subtle changes in looping frequency, leading to altered gene regulation [56]. Although the authors did not show this directly, it is likely that changes in looping frequency between enhancers and target gene promoters alter gene expression through reduced or increased transactivation mediated by specific TFs (Figure 2D). Together, these findings suggest that small modifications to local chromatin structure may be an important mechanism by which common variants confer disease risk.

Genetic Variants That Alter ncRNAs

The extent of the contribution of ncRNAs to the development of disease etiology remains unclear. The most widely studied are miRNAs that mediate post-transcriptional gene silencing via binding to complementary sites in their target mRNAs (reviewed in [57]). Mutations can occur in miRNA seed regions, which are 6–8-nucleotide-long sequences at the 5′ ends of mature miRNAs, or at miRNA binding sites in target gene 3′UTRs. For example, a heterozygous mutation in the miRNA seed region of *miR-204* segregated with a family with inherited retinal dystrophy and bilateral coloboma [58]. Functional studies showed that the mutation altered the *miR-204* mRNA targeting capabilities, leading to corresponding changes in mRNA expression and causing a retinal phenotype in medaka fish [58] (Figure 3A). For complex disease, *in silico* tools combined with eQTL data have increased the detection and characterization of SNP–miRNA pairs. Using this approach, de Almeida *et al.* identified 34 SNPs that affect miRNA binding sites associated with at least two autoimmune diseases, 30 of which were not reported by the original GWAS [59]. Furthermore, a recent genome-wide analysis integrated miRNA expression, mRNA expression, and genotype data derived from human tumors found in The Cancer Genome Atlas (TCGA) to identify SNPs that modulate miRNA regulation of gene targets (called regQTLs) [60]. They identified thousands of gene–miRNA–SNP trios across multiple cancer types, but additional functional experiments are still required to confirm the biological relevance of the results to complex disease. Mutations in core spliceosomal **small nuclear RNAs (snRNAs)** also underlie several human disorders. One recent study detected a single point mutation in the minor spliceosomal U12 snRNA (RNU12) associated with early onset cerebellar ataxia [61]. The mutation altered an essential

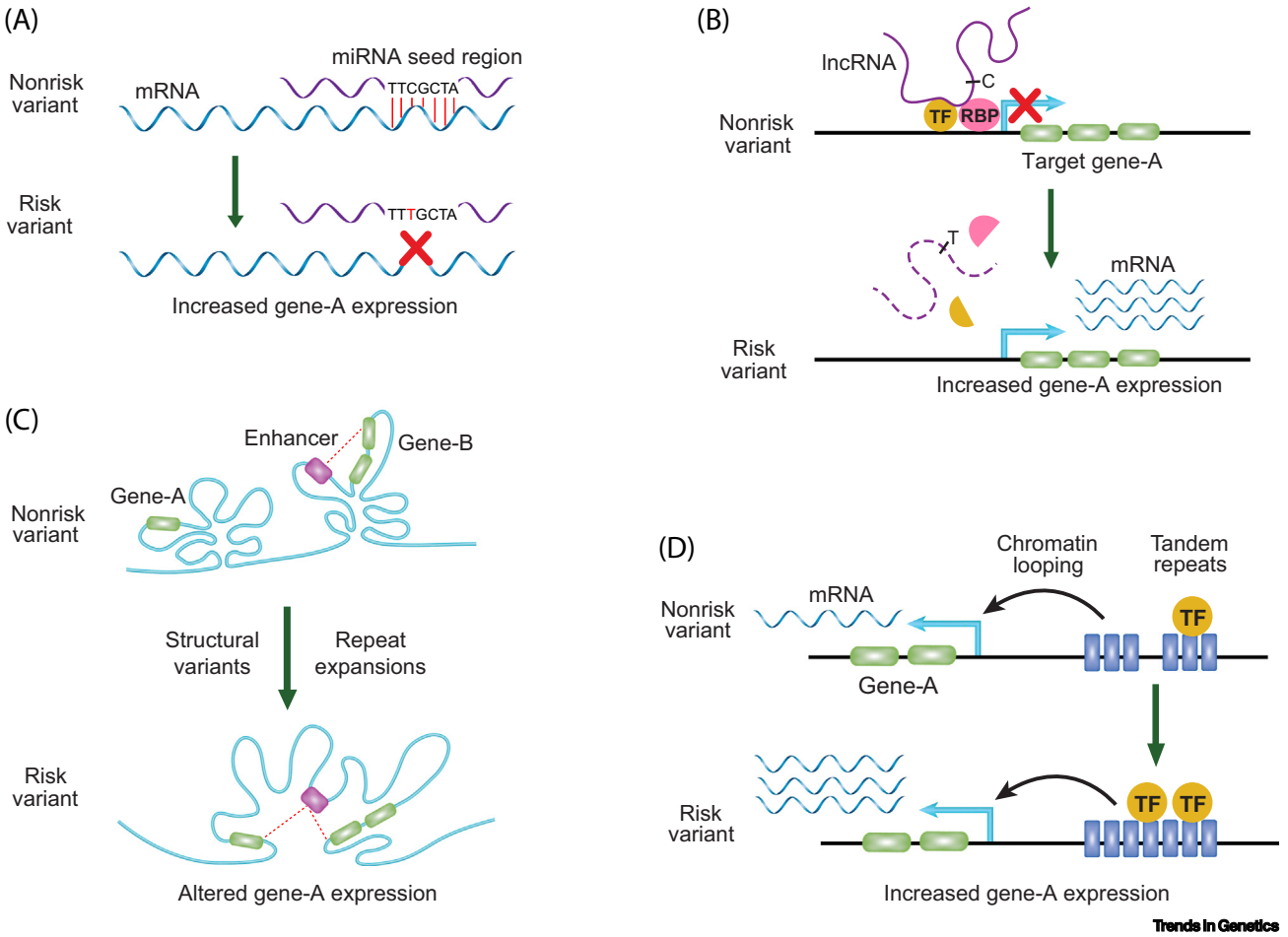

Figure 3. Examples of How Noncoding Variants Can Impact Noncoding RNAs (ncRNAs), Topologically Associating Domains (TADs), and Repeat Elements. (A) miRNAs. Risk variants in miRNA seed regions can prevent miRNA binding to target genes, leading to increased gene expression. (B) Long ncRNAs (lncRNAs). Risk variants can prevent protein binding [TF (transcription factor), RBP, (RNA-binding protein)] to the lncRNA, resulting in loss of lncRNA-mediated repression at target gene promoters and increased gene expression. (C) TADs. Structural variants or repeat expansions can disrupt TAD boundary elements, leading to inappropriate enhancer–promoter interactions and deregulated gene expression. (D) Repeat elements. Risk variants can connect adjacent tandem repeats, leading to increased TF binding and target gene expression.

stem-loop structure in RNU12, which resulted in significant minor intron retention in U12-type mRNA transcripts, 28 of which were differentially expressed between affected and unaffected individuals, including three genes previously associated with cerebellar ataxia [61].

lncRNAs are a large, heterogeneous group of ncRNAs longer than 200 nucleotides. They are involved in almost every cellular process including transcription, chromatin organization, and RNA processing (reviewed in [62]).Thousands of lncRNAs have been annotated in diverse cell types, but relatively few have been assigned biological functions. Genomics alterations have been reported to affect the expression of lncRNAs involved in disease phenotypes. In a study by Ang *et al.*, focal copy number variations were found to be enriched at several lncRNA loci in a large cohort of patients with autism spectrum disorder and intellectual disability (n = 29 085). Ones of these lncRNAs, *lnc-NR2F1*, is disrupted by a chromosomal translocation, t(5:12), in a family with neurodevelopmental symptoms [63]. Loss- and gain-of-function studies in mouse embryonic stem cells demonstrated that *lnc-NR2F1* but not the nearby protein-coding gene *NR2F1*

is important for neuronal cell maturation by regulating the transcription of various neuronal genes [63]. To date, only one family has been identified with *lnc-NR2F1* disruption; therefore, the identification of additional families will be required to confirm *lnc-NR2F2* as being responsible for the neurodevelopmental phenotypes [63].

GWAS SNPs are also enriched in lncRNA exons, some of which are lncRNA-eQTLs [64,65], suggesting that common variants alter lncRNA expression. Several studies have also demonstrated that GWAS SNPs can influence lncRNA expression by altering *cis*-regulatory elements that control lncRNA transcription. For example, Kulkarni *et al.* showed that an intronic SNP is associated with HIV outcome, increased HIV-1 viral load, and increased expression of the *CCR5AS* lncRNA in human peripheral blood lymphocytes [66]. Functional studies showed that the SNP affects the binding of ATF1 to the promoter of *CCR5AS*, leading to increased *CCR5AS* promoter activity. Importantly, knockdown of *CCR5A5* reduced the susceptibility of human $CD4^+$ T cells to HIV-1 infection by reducing the expression of the HIV coreceptor CCR5, a well-studied chemokine receptor that controls HIV entry, viral load, and progression to AIDS [66]. In another example, two prostate cancer risk SNPs modulated reciprocal expression of the *PCAT19* lncRNA short and long isoforms [67]. These SNPs act by decreasing NKX3.1 and YY1 TF binding at the *PCAT19*-short promoter resulting in weaker promoter but stronger enhancer activity that activates *PCAT19*-long. The authors then showed that *PCAT19*-long promotes prostate cancer tumor growth and metastasis by interacting with HNRNAPAB, leading to the activation of cell-cycle genes [67]. Breast cancer-associated SNPs have also been associated with reduced chromatin looping between an enhancer and the bidirectional promoter of two lncRNAs, *CUPID1* and *CUPID2*, in human breast cancer cell lines [53]. Both lncRNAs are implicated in the response to DNA damage, a pathway commonly disrupted in breast cancer, providing a plausible mechanism by which variants increase the risk of breast cancer [53]. In addition to altering lncRNA expression, genetic variants may alter lncRNA function without necessarily affecting expression. There are certainly precedents for this: in human U937 cells, a coeliac disease-associated SNP modified heterogeneous nuclear ribonucleoprotein D (hnRNPD) binding to *lnc13*, a lncRNA that represses the expression of proinflammatory genes via its interaction with hnRNPD and histone deacetylase 1 [68] (Figure 3B). These few examples are likely to represent only the tip of the iceberg, and with additional functional studies it is likely that lncRNA dysregulation will be a prominent mechanism driving human disease.

Genetic Variation within TADs

TADs are higher-order chromatin structures within which many gene regulatory interactions occur. The clustering of genes into TADs has been evolutionarily maintained and is a simple way to control the expression of genes that utilize related repertoires of TFs and splicing factors for gene expression. A structural rearrangement that disrupts this careful segregation of genes can have a substantial impact on gene expression. The extent of TAD disruption in Mendelian disease remains largely unknown, but deeper sequencing and greater precision in defining TAD boundaries has progressed their discovery. An elegant study by Lupianez *et al.* identified large structural variants at the *WNT6/IHH/EPHA4/PAX3* locus fused two TADs leading to inappropriate enhancer–promoter interactions, with consequent deregulated gene expression causing the observed limb malformations [69] (Figure 3C). Balanced chromosomal abnormalities have also been mapped to 5q14.3 associated with various neurological defects [70]. Intergenic breakpoints in eight subjects altered a single TAD containing *MEF2C*, an established driver of 5q14.3 microdeletion syndrome, resulting in decreased *MEF2C* expression. The power to detect common structural variation is currently limited. However, a recent study analyzed WGS data from 1162 Swedish schizophrenia cases and 936 ancestry-matched controls [71]. They identified several ultrarare structural variants that map to TAD boundaries associated with increased risk of

schizophrenia, but mechanistic studies are required to determine the precise functional impact of the variants in the human brain.

Genetic Variation within Tandem Repeats

Tandem repeats are copies of short DNA sequences positioned one after another in the genome in two possible orientations: head to tail (direct repeats) and head to head (inverted repeats). Tandem repeat expansions cause dozens of Mendelian diseases and are likely to contribute to complex disease. For many classic examples such as fragile X syndrome, Huntington disease, and spinocerebellar ataxia, repeat expansions are located in exons and disrupt or alter the protein product (reviewed in [72]). However, noncoding repeat expansions can also have significant effects, such as modulation of transcription and the sequestering of proteins involved in splicing and cell architecture. Sequencing and assembly difficulties have limited the detection of noncoding genetic variation within repeat regions; however, long-read sequencing technology is likely to overcome these issues. A few examples include a risk variant associated with Ewing sarcoma that connects adjacent GGAA repeats leading to increased EWSR1-FLI1-enhancer activity and *EGR2* expression [73] (Figure 3D). The spinocerebellar ataxia type 10 (SCA10) neurological disorder in humans is caused by a large noncoding repeat expansion located in intron 9 of *ATAXIN10*. McFarland *et al.* identified an interruption motif at the 5′ end of the expansion potentially acting as a phenotypic modifier [74]. Interestingly, Sun *et al.* recently reported that many disease-associated tandem repeats in humans are located at CpG island-rich TAD/sub-TAD boundaries [75]. They showed in four fragile X syndrome patients that boundary disruption correlated with loss of CTCF occupancy and *FMR1* transcriptional silencing, providing a new link between tandem repeat instability and 3D folding [75] (Figure 3C). To date, at least 30 inherited disorders are caused by coding and noncoding expanded repeats through diverse molecular mechanisms. New sequencing technologies are likely to increase this number as well as increase the detection of tandem repeat polymorphisms in polygenic disorders.

Challenges and Progress

A major challenge when assessing the impact of noncoding variants is that many functional elements are tissue- and cell-type specific. Assaying the correct cell types that are causally related to the phenotype is essential for the annotation of elements and follow-up studies. For GWASs, SNP or heritability enrichment in tissue-specific functional annotations is typically used to implicate a specific cell type in disease. For a large proportion of variants this will be helpful, but it is also likely that many traits are mediated through multiple cell types. Adding to the complexity, independent variants from the same trait can act through different cell types. For example, Factor *et al.* showed that independent SNP containing-enhancers associated with MS were active in human T/B cells, myeloid cells, and oligodendrocytes, implicating diverse cell types in MS pathogenesis [76]. Moreover, for many cancers the immune system plays a key role in tumor development [77,78]. Thus, it is possible that some cancer risk variants will act specifically in immune cell types rather than in the tissue or cell type from which the tumor was derived [79,80]. It is expected that emerging methods that allow increased resolution of cell types, such as single-cell approaches, may help to uncover the role of disease-associated variants in selected cell subpopulations.

Assessing the functional consequence of noncoding variants is also not straightforward. Prioritization of functional variants relies on functional annotation of the noncoding genome in the correct cell type, which remains far from complete. ncRNAs are also still being identified at exponential rates, and their diverse functions have hindered high-throughput studies. Moreover, given that noncoding regions were largely ignored for many years, it is likely that additional functional elements/units are yet to be discovered. Ultimately, the impact of genetic variants needs to be tested *in vivo*, which offers a new set of challenges. As noncoding

elements are generally not well conserved, it is often not possible to identify the syntenic region in model organisms. Thus, functional studies are frequently limited to human-derived models such as cultured cells and organoid models.

CRISPR-based technology may provide the answer to overcome some or all of these technical challenges. CRISPR/Cas9-mediated saturation genome editing is one strategy to edit enhancer sequences in their native context. This technology has successfully identified functional elements and candidate causal variants in the *HBS1L-MYB* interval associated with erythroid traits [81]. CRISPR/Cas13-mediated editing platforms are also being engineered to modify RNA transcripts. One of these tools, called RNA editing for specific C-to-U exchange (RESCUE), has been used to convert the Alzheimer's risk-related *APOE4* allele in the human kidney cell line HEK293FT [82]. Importantly, CRISPR technology is continuously evolving and new methods are proving more efficient with lower rates of insertion and deletion (indel) formation [83–85]. One new method used a catalytically inactive Cas9 fused to a reverse transcriptase to make a specific edit defined by a prime editing guide RNA [86]. The authors introduced 175 edits into four different human cell lines and primary neurons derived from mice, including small indels and point mutations, and correcting mutant alleles for sickle cell, Tay–Sachs, and prion disease [86]. It is anticipated that future studies will allow high-efficiency, targeted, simultaneous high-throughput interrogation of thousands of genetic variants.

Concluding Remarks

A driving force behind increased understanding of noncoding variation is the translation of this information to the clinic. As sequencing costs reduce it is likely that WGS will be the next clinical option for patients testing negative for coding mutations; however, a host of resources will still be required to interpret this information. Interestingly, it is predicted that CRISPR base editors could correct up to 89% of the pathogenic variants listed in **ClinVar** [87], but this is still many years away from clinical approval. Furthermore, noncoding variants are already included in **polygenic risk scores (PRSs)** for multiple complex diseases. For example, breast cancer polygenic risk testing is being used in specialist familial cancer clinics [88], but most PRS predictions are derived from GWASs of European ancestry and will need greater diversity to fully realize their potential (see Outstanding Questions).

Given the renewed interest in genetically supported drug mechanisms for Mendelian and complex traits [89], it is also tempting to speculate that noncoding variation could be leveraged to repurpose existing drug therapies. Drug repositioning on the basis of GWAS results has already been extremely successful for psoriasis, now treated with ixekizumab, an interleukin 17A antagonist [90]. Alternatively, the knowledge acquired may lead to entirely new modalities or therapeutic opportunities. For example, the identification of a *BCL11A* enhancer through a GWAS for fetal hemoglobin levels has led to a promising new therapy for sickle cell disease and β-thalassemia [91]. ncRNAs are also emerging as attractive targets for therapeutic interventions, particularly in cancer (reviewed in [92]). A few RNA-based therapeutics are already clinically approved [93,94] and many more are in clinical trials, suggesting that RNA-based drugs are gaining momentum in entering the clinic.

Looking forward, the accelerating pace of discovery of noncoding variation, fueled by sequencing projects such as the 100 000 Genomes [95] and Million Veteran [96], among others, will facilitate more accurate cataloguing of genetic variation in the human genome. Integrating this data with large-scale projects such as ENCODE [97] and FANTOM-CAT [98] plus individual functional studies should then progress the clinical interpretation of the noncoding genome. It is anticipated that increased linking of noncoding genetic variation to specific phenotypes will transform our ability to understand mechanisms, diagnose causes, translate findings, and better treat rare and common diseases.

Outstanding Questions

Will we solve the missing heritability problem by focusing efforts on noncoding regions of the genome?

To what extent do *in vitro*/model experimental systems reflect noncoding variation impact in humans. Can we find better ways to model noncoding variants?

Will CRISPR-based technologies be able to address current experimental deficiencies in assessing joint influences of risk alleles?

Will the inclusion of additional noncoding variants plus related transcriptomic data from diverse populations lead to better polygenic risk models and predictors?

To what extent will noncoding mechanisms be targetable?

Acknowledgments

We apologize to colleagues whose work could not be cited due to space limitations. The authors sincerely thank Madeleine Flynn for the figure designs and Georgia Chenevix-Trench and Jonathan Beesley for insightful comments on the manuscript. The authors' work related to the topic of this review is supported by grants from the National Health and Medical Research Council of Australia (NHMRC; 1120563 and 1122022). S.L.E. is an NHMRC Senior Research Fellow (1135932). J.D.F. is supported by a philanthropic donation from Isabel and Roderic Allpass.

References

1. Sakharkar, M.K. *et al.* (2004) Distributions of exons and introns in the human genome. *In Silico Biol.* 4, 387–393
2. de Koning, A.P. *et al.* (2011) Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* 7, e1002384
3. Carninci, P. *et al.* (2005) The transcriptional landscape of the mammalian genome. *Science* 309, 1559–1563
4. Djebali, S. *et al.* (2012) Landscape of transcription in human cells. *Nature* 489, 101–108
5. ENCODE Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74
6. Szabo, Q. *et al.* (2019) Principles of genome folding into topologically associating domains. *Sci. Adv.* 5, eaaw1668
7. Dixon, J.R. *et al.* (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380
8. Nora, E.P. *et al.* (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 381–385
9. Hsieh, T.S. *et al.* (2020) Resolving the 3D landscape of transcription-linked mammalian chromatin folding. *Mol. Cell* 78, 539–553
10. Bonev, B. *et al.* (2017) Multiscale 3D genome rewiring during mouse neural development. *Cell* 171, 557–572
11. Fu, X.D. and Ares Jr., M. (2014) Context-dependent control of alternative splicing by RNA-binding proteins. *Nat. Rev. Genet.* 15, 689–701
12. Scotti, M.M. and Swanson, M.S. (2016) RNA mis-splicing in disease. *Nat. Rev. Genet.* 17, 19–32
13. Abramowicz, A. and Gos, M. (2018) Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J. Appl. Genet.* 59, 253–268
14. Jamshidi, F. *et al.* (2019) Contribution of noncoding pathogenic variants to RPGRIP1-mediated inherited retinal degeneration. *Genet. Med.* 21, 694–704
15. Aneichyk, T. *et al.* (2018) Dissecting the causal mechanism of X-linked dystonia-Parkinsonism by integrating genome and transcriptome assembly. *Cell* 172, 897–909
16. Bryen, S.J. *et al.* (2019) Pathogenic abnormal splicing due to intronic deletions that induce biophysical space constraint for spliceosome assembly. *Am. J. Hum. Genet.* 105, 573–587
17. Lappalainen, T. *et al.* (2013) Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 501, 506–511
18. Takata, A. *et al.* (2017) Genome-wide identification of splicing QTLs in the human brain and their enrichment among schizophrenia-associated loci. *Nat. Commun.* 8, 14519
19. Pickrell, J.K. *et al.* (2010) Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* 464, 768–772
20. Li, Y.I. *et al.* (2016) RNA splicing is a primary link between genetic variation and disease. *Science* 352, 600–604
21. Zhang, X. *et al.* (2015) Identification of common genetic variants controlling transcript isoform variation in human whole blood. *Nat. Genet.* 47, 345–352
22. Li, M. *et al.* (2016) A human-specific AS3MT isoform and BORCS7 are molecular risk factors in the 10q24.32 schizophrenia-associated locus. *Nat. Med.* 22, 649–656
23. Kim, H.I. *et al.* (2017) Fine mapping and functional analysis reveal a role of SLC22A1 in acylcarnitine transport. *Am. J. Hum. Genet.* 101, 489–502
24. Steri, M. *et al.* (2018) Genetic variants in mRNA untranslated regions. *Wiley Interdiscip. Rev. RNA* 9, e1474
25. Patel, N. *et al.* (2017) A novel mechanism for variable phenotypic expressivity in Mendelian diseases uncovered by an AU-rich element (ARE)-creating mutation. *Genome Biol.* 18, 144
26. Tomaselli, P.J. *et al.* (2017) Mutations in noncoding regions of GJB1 are a major cause of X-linked CMT. *Neurology* 88, 1445–1453
27. Steri, M. *et al.* (2017) Overexpression of the cytokine BAFF and autoimmunity risk. *N. Engl. J. Med.* 376, 1615–1626
28. Haberle, V. and Stark, A. (2018) Eukaryotic core promoters and the functional basis of transcription initiation. *Nat. Rev. Mol. Cell Biol.* 19, 621–637
29. Davidson, A.E. *et al.* (2016) Autosomal-dominant corneal endothelial dystrophies CHED1 and PPCD1 are allelic disorders caused by non-coding mutations in the promoter of OVOL2. *Am. J. Hum. Genet.* 98, 75–89
30. Li, J. *et al.* (2016) Point mutations in exon 1B of APC reveal gastric adenocarcinoma and proximal polyposis of the stomach as a familial adenomatous polyposis variant. *Am. J. Hum. Genet.* 98, 830–842
31. Horn, S. *et al.* (2013) TERT promoter mutations in familial and sporadic melanoma. *Science* 339, 959–961
32. Harland, M. *et al.* (2016) Germline TERT promoter mutations are rare in familial melanoma. *Familial Cancer* 15, 139–144
33. Tian, J. *et al.* (2019) Systematic functional interrogation of genes in GWAS loci identified ATF1 as a key driver in colorectal cancer modulated by a promoter–enhancer interaction. *Am. J. Hum. Genet.* 105, 29–47
34. Bojesen, S.E. *et al.* (2013) Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat. Genet.* 45, 384e371–384e372
35. Michailidou, K. *et al.* (2017) Association analysis identifies 65 new breast cancer risk loci. *Nature* 551, 92–94
36. Fachal, L. *et al.* (2020) Fine-mapping of 150 breast cancer risk regions identifies 191 likely target genes. *Nat. Genet.* 52, 56–73
37. Bonev, B. and Cavalli, G. (2016) Organization and function of the 3D genome. *Nat. Rev. Genet.* 17, 772
38. Will, A.J. *et al.* (2017) Composition and dosage of a multipartite enhancer cluster control developmental expression of Ihh (Indian hedgehog). *Nat. Genet.* 49, 1539–1545
39. Kvon, E.Z. *et al.* (2020) Comprehensive *in vivo* interrogation reveals phenotypic impact of human enhancer variants. *Cell* 180, 1262–1271
40. Soukup, A.A. *et al.* (2019) Single-nucleotide human disease mutation inactivates a blood-regenerative GATA2 enhancer. *J. Clin. Invest.* 129, 1180–1192
41. Karnuta, J.M. and Scacheri, P.C. (2018) Enhancers: bridging the gap between gene control and human disease. *Hum. Mol. Genet.* 27, R219–R227
42. Chatterjee, S. and Ahituv, N. (2017) Gene regulatory elements, major drivers of human disease. *Annu. Rev. Genomics Hum. Genet.* 18, 45–63
43. Soldner, F. *et al.* (2016) Parkinson-associated risk variant in distal enhancer of α-synuclein modulates target gene expression. *Nature* 533, 95–99
44. French, J.D. *et al.* (2013) Functional variants at the 11q13 risk locus for breast cancer regulate cyclin D1 expression through long-range enhancers. *Am. J. Hum. Genet.* 92, 489–503
45. Benaglio, P. *et al.* (2019) Allele-specific NKX2-5 binding underlies multiple genetic associations with human electrocardiographic traits. *Nat. Genet.* 51, 1506–1517
46. Gupta, R.M. *et al.* (2017) A genetic variant associated with five vascular diseases is a distal regulator of endothelin-1 gene expression. *Cell* 170, 522–533 e515
47. McGovern, A. *et al.* (2016) Capture Hi-C identifies a novel causal gene, IL20RA, in the pan-autoimmune genetic susceptibility region 6q23. *Genome Biol.* 17, 212
48. Smemo, S. *et al.* (2014) Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* 507, 371–375
49. Claussnitzer, M. *et al.* (2015) FTO obesity variant circuitry and adipocyte browning in humans. *N. Engl. J. Med.* 373, 895–907

50. Mumbach, M.R. *et al.* (2017) Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat. Genet.* 49, 1602–1612
51. Beesley, J. *et al.* (2020) Chromatin interactome mapping at 139 independent breast cancer risk signals. *Genome Biol.* 21, 8
52. Kycia, I. *et al.* (2018) A common type 2 diabetes risk variant potentiates activity of an evolutionarily conserved islet stretch enhancer and increases C2CD4A and C2CD4B expression. *Am. J. Hum. Genet.* 102, 620–635
53. Betts, J.A. *et al.* (2017) Long noncoding RNAs CUPID1 and CUPID2 mediate breast cancer risk at 11q13 by modulating the response to DNA damage. *Am. J. Hum. Genet.* 101, 255–266
54. Vicente, C.T. *et al.* (2015) Long-range modulation of PAG1 expression by 8q21 allergy risk variants. *Am. J. Hum. Genet.* 97, 329–336
55. Gorkin, D.U. *et al.* (2019) Common DNA sequence variation influences 3-dimensional conformation of the human genome. *Genome Biol.* 20, 255
56. Greenwald, W.W. *et al.* (2019) Subtle changes in chromatin loop contact propensity are associated with differential gene regulation and expression. *Nat. Commun.* 10, 1054
57. Moszynska, A. *et al.* (2017) SNPs in microRNA target sites and their potential role in human disease. *Open Biol.* 7, 170019
58. Conte, I. *et al.* (2015) MiR-204 is responsible for inherited retinal dystrophy associated with ocular coloboma. *Proc. Natl. Acad. Sci. U. S. A.* 112, E3236–E3245
59. de Almeida, R.C. *et al.* (2018) Integrative analysis identifies genetic variants associated with autoimmune diseases affecting putative microRNA binding sites. *Front. Genet.* 9, 139
60. Wilk, G. and Braun, R. (2018) regQTLs: single nucleotide polymorphisms that modulate microRNA regulation of gene expression in tumors. *PLoS Genet.* 14, e1007837
61. Elsaid, M.F. *et al.* (2017) Mutation in noncoding RNA RNU12 causes early onset cerebellar ataxia. *Ann. Neurol.* 81, 68–78
62. Yao, R.W. *et al.* (2019) Cellular functions of long noncoding RNAs. *Nat. Cell Biol.* 21, 542–551
63. Ang, C.E. *et al.* (2019) The novel lncRNA lnc-NR2F1 is pro-neurogenic and mutated in human neurodevelopmental disorders. *eLife* 8, e41770
64. Moradi Marjaneh, M. *et al.* (2020) Non-coding RNAs underlie genetic predisposition to breast cancer. *Genome Biol.* 21, 7
65. Li, J. *et al.* (2020) ncRNA-eQTL: a database to systematically evaluate the effects of SNPs on non-coding RNA expression across cancer types. *Nucleic Acids Res.* 48, D956–D963
66. Kulkarni, S. *et al.* (2019) CCR5AS lncRNA variation differentially regulates CCR5, influencing HIV disease outcome. *Nat. Immunol.* 20, 824–834
67. Hua, J.T. *et al.* (2018) Risk SNP-mediated promoter–enhancer switching drives prostate cancer through lncRNA PCAT19. *Cell* 174, 564–575 e518
68. Castellanos-Rubio, A. *et al.* (2016) A long noncoding RNA associated with susceptibility to celiac disease. *Science* 352, 91–95
69. Lupianez, D.G. *et al.* (2015) Disruptions of topological chromatin domains cause pathogenic rewiring of gene–enhancer interactions. *Cell* 161, 1012–1025
70. Redin, C. *et al.* (2017) The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. *Nat. Genet.* 49, 36–45
71. Halvorsen, M. *et al.* (2020) Increased burden of ultra-rare structural variants localizing to boundaries of topologically associated domains in schizophrenia. *Nat. Commun.* 11, 1842
72. Hannan, A.J. (2018) Tandem repeats mediating genetic plasticity in health and disease. *Nat. Rev. Genet.* 19, 286–298
73. Grunewald, T.G. *et al.* (2015) Chimeric EWSR1-FLI1 regulates the Ewing sarcoma susceptibility gene EGR2 via a GGAA microsatellite. *Nat. Genet.* 47, 1073–1078
74. McFarland, K.N. *et al.* (2015) SMRT sequencing of long tandem nucleotide repeats in SCA10 reveals unique insight of repeat expansion structure. *PLoS One* 10, e0135906
75. Sun, J.H. *et al.* (2018) Disease-associated short tandem repeats co-localize with chromatin domain boundaries. *Cell* 175, 224–238
76. Factor, D.C. *et al.* (2020) Cell type-specific intralocus interactions reveal oligodendrocyte mechanisms in MS. *Cell* 181, 382–395
77. Gonzalez, H. *et al.* (2018) Roles of the immune system in cancer: from tumor initiation to metastatic progression. *Genes Dev.* 32, 1267–1284
78. Hanahan, D. and Weinberg, R.A. (2011) Hallmarks of cancer: the next generation. *Cell* 144, 646–674
79. Geeleher, P. *et al.* (2018) Cancer expression quantitative trait loci (eQTLs) can be determined from heterogeneous tumor gene expression data by modeling variation in tumor purity. *Genome Biol.* 19, 130
80. Lim, Y.W. *et al.* (2018) Germline genetic polymorphisms influence tumor gene expression and immune cell infiltration. *Proc. Natl. Acad. Sci. U. S. A.* 115, E11701–E11710
81. Canver, M.C. *et al.* (2017) Variant-aware saturating mutagenesis using multiple Cas9 nucleases identifies regulatory elements at trait-associated loci. *Nat. Genet.* 49, 625–634
82. Abudayyeh, O.O. *et al.* (2019) A cytosine deaminase for programmable single-base RNA editing. *Science* 365, 382–386
83. Komor, A.C. *et al.* (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420–424
84. Nishida, K. *et al.* (2016) Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* 353, aaf8729
85. Gaudelli, N.M. *et al.* (2017) Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature* 551, 464–471
86. Anzalone, A.V. *et al.* (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576, 149–157
87. Molla, K.A. and Yang, Y. (2019) CRISPR/Cas-mediated base editing: technical considerations and practical applications. *Trends Biotechnol.* 37, 1121–1142
88. Yanes, T. *et al.* (2020) Clinical applications of polygenic breast cancer risk: a critical review and perspectives of an emerging field. *Breast Cancer Res.* 22, 21
89. Nelson, M.R. *et al.* (2015) The support of human genetic evidence for approved drug indications. *Nat. Genet.* 47, 856–860
90. Ramaekers, B.L.T. *et al.* (2018) Ixekizumab for treating moderate-to-severe plaque psoriasis: an evidence review group perspective of a NICE single technology appraisal. *Pharmacoeconomics* 36, 917–927
91. Wu, Y. *et al.* (2019) Highly efficient therapeutic gene editing of human hematopoietic stem cells. *Nat. Med.* 25, 776–783
92. Arun, G. *et al.* (2018) Therapeutic targeting of long non-coding RNAs in cancer. *Trends Mol. Med.* 24, 257–277
93. Chiriboga, C.A. *et al.* (2016) Results from a Phase 1 study of nusinersen (ISIS-SMN$_{Rx}$) in children with spinal muscular atrophy. *Neurology* 86, 890–897
94. Kristen, A.V. *et al.* (2019) Patisiran, an RNAi therapeutic for the treatment of hereditary transthyretin-mediated amyloidosis. *Neurodegener. Dis. Manag.* 9, 5–23
95. Turnbull, C. *et al.* (2018) The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. *BMJ* 361, k1687
96. Gaziano, J.M. *et al.* (2016) Million Veteran Program: a mega-biobank to study genetic influences on health and disease. *J. Clin. Epidemiol.* 70, 214–223
97. Davis, C.A. *et al.* (2018) The Encyclopedia of DNA Elements (ENCODE): data portal update. *Nucleic Acids Res.* 46, D794–D801
98. Hon, C.C. *et al.* (2017) An atlas of human long non-coding RNAs with accurate 5′ ends. *Nature* 543, 199–204