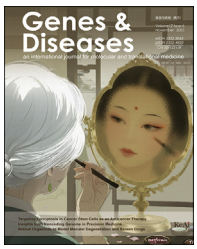
Genes & Diseases (2025) 12, 101652

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.keaipublishing.com/en/journals/genes-diseases



REVIEW ARTICLE

Unveiling the regulatory potential of the non-coding genome: Insights from the human genome project to precision medicine

Paola Ruffo [a,b,*], **Bryan J. Traynor** [a,c], **Francesca Luisa Conforti** [b]

[a] *Neuromuscular Diseases Research Section, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892, USA*
[b] *Medical Genetics Laboratory, Department of Pharmacy, Health and Nutritional Sciences, University of Calabria, Rende 87036, Italy*
[c] *Department of Neurology, Johns Hopkins University Medical Center, Baltimore, MD 21287, USA*

Received 22 July 2024; received in revised form 18 December 2024; accepted 29 March 2025
Available online 22 April 2025

KEYWORDS
Chromatin modifications;
Enhancers;
Gene expression;
Genomic technologies;
Next-generation sequencing (NGS);
Non-coding genome;
Precision medicine;
Regulatory mechanisms

Abstract The Human Genome Project marked a milestone in scientific exploration, unraveling the genetic blueprint of humanity. However, expectations of direct gene—disease associations gave way to realizing the complexity of genetic interactions, especially in polygenic diseases. This review explores the legacy of the HGP and subsequent advancements in genomic technologies, particularly next-generation sequencing, which have enabled more profound insights into the non-coding genome's role in gene regulation. While initially dismissed as "junk" DNA, non-coding regions are now officially approved as critical gene expression and genome organization regulators. Through integrative genomics approaches and advanced computational methods, researchers have unveiled the intricate network of enhancers, promoters, and chromatin modifications orchestrating gene expression. High-throughput sequencing techniques and functional assays have identified non-coding variants associated with numerous diseases, challenging the conventional focus on coding sequences in genomic studies. By elucidating the regulatory mechanisms governing gene expression, researchers can advance precision medicine approaches and develop novel diagnostic tools. As genomic research continues to evolve, a vast landscape is waiting to be explored, promising transformative insights into human health and disease. This review provides a comprehensive overview of the non-coding genome's role in gene regulation and its implications for understanding complex diseases and developing targeted therapeutic interventions.

* Corresponding author. Neuromuscular Diseases Research Section, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892, USA.
E-mail address: paola.ruffo@nih.gov (P. Ruffo).
Peer review under the responsibility of the Genes & Diseases Editorial Office, in alliance with the Association of Chinese Americans in Cancer Research (ACACR, Baltimore, MD, USA).

https://doi.org/10.1016/j.gendis.2025.101652

Introduction

The Human Genome Project (HGP) stands as one of the most monumental scientific undertakings in history. The project was a biological exploration led by an international team of researchers whose primary mission was deciphering the chemical sequence of the complete human genetic material, the entire genome.[1] The three billion nucleotides that comprise the textbook on human genetics were successfully sequenced in April 2003, fulfilling the project's main objective. The concept that emerged with the fulfillment of the HGP was the hope of comprehending the pathogenic factors underlying diseases and the potential to stratify individuals at risk through genetic testing. However, these expectations turned out to be overly optimistic. Instead of simple gene—disease associations, a complex network of interactions was uncovered within the genetic code concerning hundreds to millions of single-nucleotide polymorphisms, especially between the more common and complex diseases.[2] Such disorders have a polygenic basis involving tons of genetic variants, each of which has only a tiny effect on the disease process.[3] Moreover, the HGP's legacy went beyond the sequencing itself. The project laid the foundation for powerful bioinformatics tools and resources necessary to handle and analyze the enormous datasets generated during the sequencing process. Establishing comprehensive genomic databases and repositories facilitated data sharing and collaboration among scientists worldwide, accelerating discoveries and breakthroughs in genomics.

Developing high-throughput sequencing techniques, also known as next-generation sequencing, is particularly important in this scenario. A direct consequence of this evolution is the rapid update of computational methods and machine learning algorithms needed to understand and analyze the deepest and wealthiest datasets. Furthermore, integrative genomics approaches have emerged, combining data from various sources, such as genomics, transcriptomics, proteomics, and epigenomics, to better understand the molecular mechanisms governing cellular function and disease.[4] By correlating information from different levels of biological complexity, it is possible to unravel complex gene regulatory networks and identify potential therapeutic targets (Fig. 1).

The application of new DNA sequencing technologies has made it possible to read longer sequences without compromising accuracy. By enabling the generation of longer sequence reads while maintaining high accuracy, the PacBio HiFi and Oxford Nanopore DNA sequencing methods have greatly advanced genomic research. This has improved our ability to resolve complex genomic regions and detect structural variations with greater confidence and precision. The first can read about 20,000 nitrogenous bases with almost perfect precision. Meanwhile, the Oxford Nanopore DNA sequencing method analyzes long-read sequencing of up to one million nucleobases of DNA at a time. Both were used to generate the complete human genome sequence. Leveraging these technologies, researchers successfully sequenced the human genome, with the latest update of the HGP expanding the genetic code by almost 200 million letters. The final 8% of the genome features various genes and repetitive DNA sequences that can affect cell activity. Most of the new sequences are situated near the ends of each chromosome and in the centromeres, the densely packed central regions of chromosomes.

Beyond the code: unraveling non-coding regions

It is widely agreed[5–7] that approximately 90% of the human genome is not subject to purifying selection and hence—at least according to typical accounts of function in molecular biology—is not functional. Despite this, we argue that paradigmatically "non-functional" DNA is commonly referred to as "junk" DNA. This term was first used in the 1960s[8] but was formalized by Susumu Ohno in 1972.[9] Ohno's observation has revealed that the occurrence of deleterious mutations restricts the number of functional loci that can be expected under typical mutation rates. In a Nature review published in 1980,[10] Leslie Orgel and Francis Crick argued that junk DNA exhibited limited specificity and conferred minimal or no selective advantage to the organism.[11]

These non-protein coding regions were considered redundant and free from selective pressures, thus allowing the accumulation of mutations without damaging the organism.[9,12]

In light of this data, it is thought that some segments of what were formerly thought to be non-coding DNA (ncDNA) have undergone the exaptation process during evolution. This process describes how function is acquired by means other than natural selection.[13] Nonetheless, a number of ncDNA structural components that control gene expression have been identified, leading to the tridimensional genomic organization that is essential for appropriate gene regulation. Data from the ENCODE (Encyclopedia of DNA Elements) project suggests that more than 75% of the human genome is transcribed into RNAs, whereas only approximately 2% of these RNAs are protein-coding genes[13,14] (Fig. 2).

Many non-coding sequences are repeated transposable (*i.e.*, movable) elements that facilitate genomic rearrangements and are of evolutionary importance. Our genome is also packed with many types of tandemly repeated DNA sequences: about 45% consists of repetitive elements such as long terminal repeats and long/short interspersed nuclear elements (including as many as one million Alu repeats), and perhaps another 25% is made up of

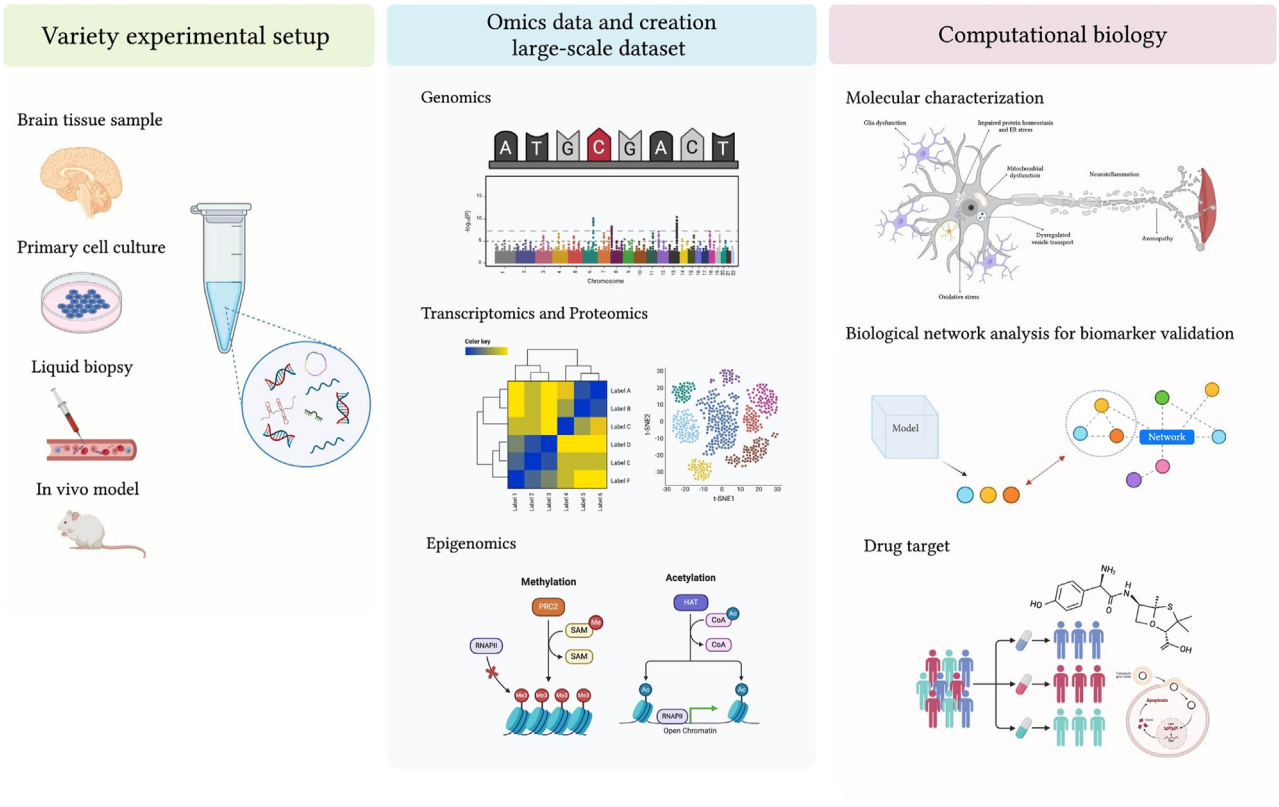


Figure 1 Technologies evolved related to the Human Genome Project. Next-generation sequencing data can come from various experimental and technological conditions. Depending on the experiment's aim, one or more of the represented omics types are analyzed (genomics, transcriptomics, epigenomics, or single-cell omics). These approaches have led to accumulating large-scale next-generation sequencing datasets to solve various challenges of heterogeneous disease research, molecular characterization, and drug target discovery. Statistical analysis of next-generation sequencing data allows the identification of candidate genes. Biological networks can validate these candidate genes and highlight underlying biological mechanisms. Furthermore, machine learning models that aim to reconstruct biological networks can incorporate prior knowledge from diverse omics data. Subsequently, the model will predict unknown interactions based on new omics information. When applied to diverse next-generation sequencing data, machine learning models integrate biological networks as prior knowledge to improve predictive performance. The goal is the identification of ideal drugs for personalized therapy. The figure was created with BioRender.com.

shorter tandem repeats such as satellites, minisatellites, and microsatellites.[15]

Although, as stated, 98% of the human genome consists of ncDNA sequences, most of the genome is transcribed into RNA if at a low level[16] (Fig. 2). Our genome also encodes a large number of non-coding RNAs (ncRNAs) that can be classified into infrastructure ncRNAs and regulatory ncRNAs. The former is constitutively expressed, including ribosomal, transfer, and small nuclear and nucleolar RNAs. While regulatory ncRNAs include long non-coding RNAs (lncRNAs), which are non-protein-coding transcripts of more than 200 nucleotides expressed at 10-fold lower abundance than protein-encoding mRNAs but appear to be functionally different from shorter RNA species such as microRNAs (miRNAs; thousands of these regulate the transcription of mRNAs), short interfering RNAs (siRNAs; 20–25 bases long and inhibit gene expression), Piwi-interacting RNAs (piRNAs; 26–31 bases long and probably involved in gene silencing), and small nucleolar RNAs (snoRNAs; interact with proteins and other RNAs but with incompletely understood functions).[17]

ncDNA repeats may appear unremarkable at first glance and resemble inherent noise. However, studies have shown that ncDNA repeats, such as satellites and interspersed elements, may play a role in guarding and protecting select important DNAs and proteins from suffering illegitimate rearrangements[18] as well as represent important functional regulators with biological relevance.[19] A recent study of the OCT4 gene (alias of POU5F1 (POU class 5 homeobox 1); OMIM* 164177) has highlighted that evolution may use ncDNA to change animal anatomy.[20] Coding DNA provides us with a limited understanding of genome functions. We now know that the genetic code cannot fully explain most genetic diseases. Thus, it might be more accommodating if we combine the genetic network, consisting of a complete set of genes, including coding DNAs, ncDNAs, and their topological interactions. Scientists have recently begun to study the underlying genetics of ncDNA to understand and possibly prevent cancer.[21] By silencing GNG12-AS1, a strand of lncRNA, a research team found that this strategy could enable the discrimination between functions related to its active transcription and that of the RNA products, which

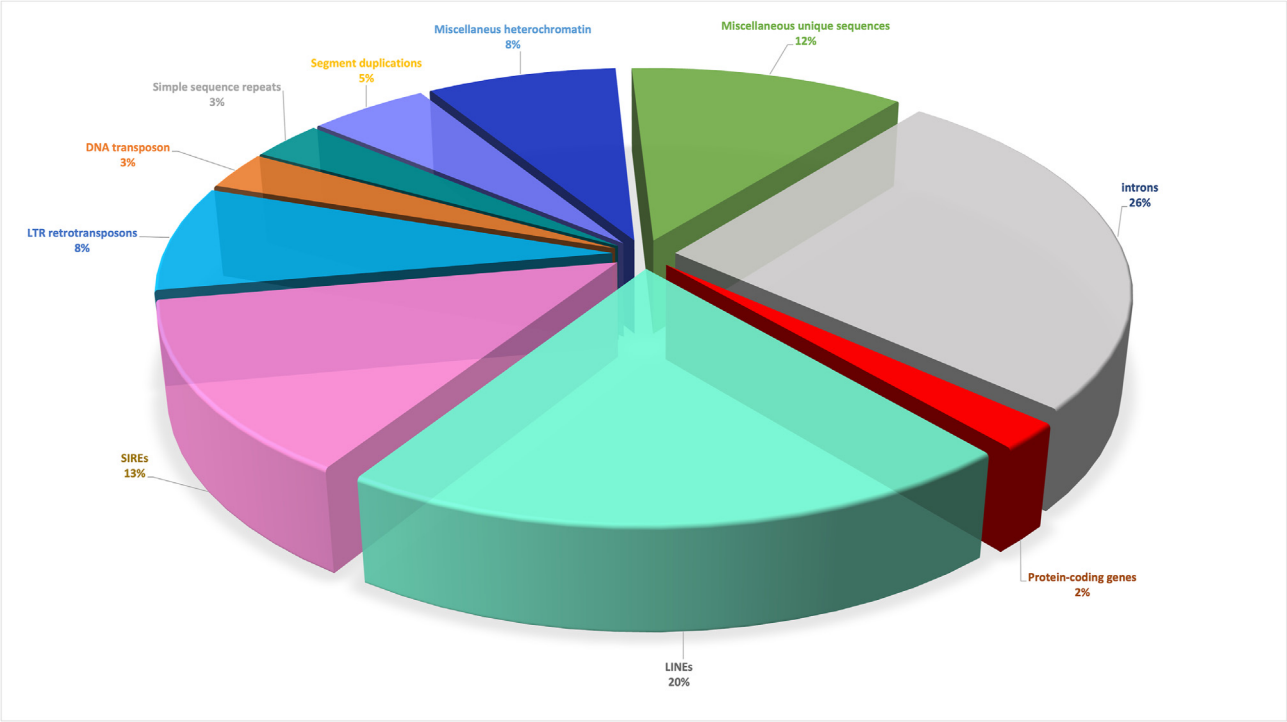


Figure 2 Composition of the human genome. The pie chart shows the distribution of the various components of the human genome (created with Biorender).

may be related to cancer metastasis. It is well known that trans-acting factors, customarily described as cis-acting elements and trans-acting factors, control gene expression through specific modules in a promoter. cis-elements are ncDNA sequences that control the transcription of a nearby gene, whereas trans-regulatory elements are ncDNAs that control the transcription of a distant gene. Both cis-elements and trans-regulatory elements are required for gene expression. Mutations that affect the function of phenotypic diversity have been reported and characterized in recent years, indicating that changes in non-coding trans-acting factors may result in organismal phenotype divergence.[22,23] Increasing evidence shows that ncDNA may cause diseases like cancer, genetic diseases, diabetes mellitus, and neurological diseases if these regulatory "junk" DNAs go wrong. In spinocerebellar ataxia type 10 (SCA10), a large repeat expansion within intron 9 of the ATAXIN10 gene is causative, with an interruption motif at the 5′ end of the expansion acting as a potential phenotypic modifier.[24] Non-coding repeat expansions also contribute to epilepsies, such as progressive myoclonus epilepsy of the Unverricht—Lundborg type (EPM1), which is associated with a dodecamer repeat expansion in the CSTB promoter. Benign adult familial myoclonic epilepsy is linked to the expansion of five-nucleotide sequences (TTTCA or TTTTA) within introns of different genes in various patients.[25]

In Parkinson's disease, variants within an enhancer of the SNCA gene influence susceptibility. One variant reduces SNCA expression, providing a protective effect, while another increases SNCA expression, elevating the risk of developing the disease.[26] Together, these examples underscore the critical impact of non-coding regions on gene regulation and the development of diverse neurological and skeletal disorders.

Moreover, variations in the telomerase reverse transcriptase gene (TERT) promoter are strongly associated with the development and progression of various cancers. These mutations, first identified in melanoma, create novel transcription factor binding sites, leading to increased telomerase activity.[27] Such alterations have since been linked to glioblastoma, hepatocellular carcinoma, bladder cancer, thyroid cancer, and breast cancer, where they are often correlated with enhanced tumor growth and poor clinical outcomes.[28—30]

Furthermore, studies have shown strong direct and indirect evidence that "junk" DNA prevents the production of "junk" protein. The increased understanding of the biological mechanisms of ncDNA will allow us to be more effective at designing targeted treatments and diagnosing diseases. Moreover, the crosstalk between epigenetics and "junk" DNA has been observed to play a crucial role in particular genetic activity. ncDNA may act as a regulator to affect epigenetic mechanisms.[31] In addition, epigenetic modification can lead to genetic alterations, which might go through non-coding DNA.[32]

The non-coding genome and its role in gene regulation

The genomic organization contains efficient DNA packaging in the restricted space of the nucleus while allowing for DNA replication and gene expression. First, nucleosomes are assembled, in which 147 base pairs (bp) of DNA are wrapped around eight histone proteins linked to each other

by DNA stretches of various lengths. This beads-on-a-string organization forms the basis of a 10-nm chromatin fiber typical of open chromatin, also known as euchromatin. This differs from tightly packaged heterochromatin, where multiple histones wrap into a 30-nm fiber consisting of nucleosome arrays in their most compact form. As a result, chromatin is organized into active and inactive compartments that are either open or condensed and vary in size between 1 and 10 megabases. Inactive compartments are often associated with the nuclear lamina, whereas active cases are more likely to be found in other nuclear regions.[33]

The epigenomic elements are essentially open areas of chromatin, encompassing several cis-regulatory elements like promoters, insulators, enhancers, and binding sites for transcription factors (TFs) (Fig. 3). Alongside these, histone modification marks also work together to regulate the transcriptome in a dynamic chromatin topology without changing the nucleotide sequences themselves. The chromatin topology is affected by chemical changes to DNA and histones, which are heritable and reversible. These modifications, known as epigenetic modifications, play a significant role in regulating the transcriptome and the chromatin structure.[34]

Various post-translational epigenetic modifications of histones put in place by chromatin-modifying enzymes can alter the accessibility of chromatin itself. They can thereby influence how chromatin is packaged and whether it is more or less likely to be active. For example, histone acetylation results in increased chromatin accessibility and makes chromatin more available for the binding of regulatory proteins, such as TFs.

Several studies focused on a wide variety of histone modifications and have led to a draft of a histone code, where different histone modifications reveal the functional role that the chromatin has at those modified places.[35,36] For example, putative enhancers are enriched in chromatin

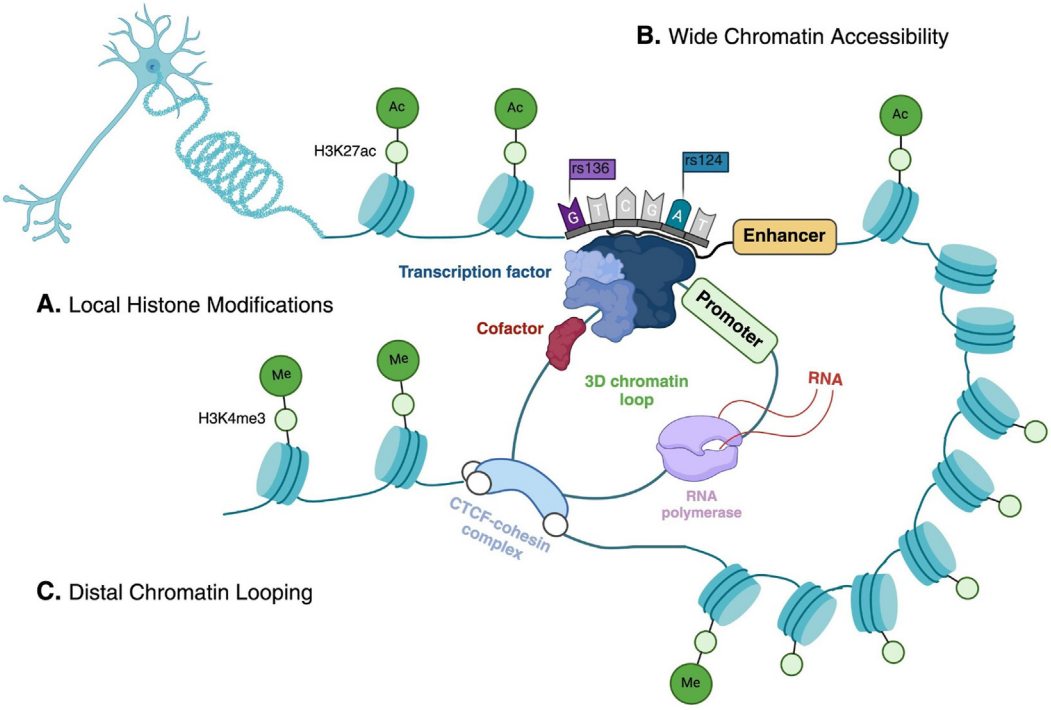


Figure 3 The chromatin environment. Open chromatin includes both coding and non-coding parts of the genome. The interactions among regulatory elements, genes, and non-coding regions can be local, such as through histone modifications, or distal, through three-dimensional chromatin interactions. Techniques for profiling chromatin can map the dynamic and functional aspects of the chromatin environment. **(A)** Local histone modifications: These include acetylation or methylation, which can alter chromatin accessibility and permit regulatory proteins like transcription factors (TFs) to bind to and influence the expression of nearby genes. Techniques like ChIP-seq, CUT&RUN, or CUT&TAG are used to assay these modifications and TF binding. **(B)** Chromatin accessibility: Major changes in chromatin structure and the redistribution of nucleosomes can affect gene expression either directly or indirectly. The chromatin environment, cis-regulatory elements, and nucleosome positioning can be studied using methods such as ATAC-seq, MNase-seq, DNase-seq, or FAIRE-seq. Genome-wide association studies (GWAS) often locate risk loci for complex traits within the open non-coding genome, where the most significant single-nucleotide polymorphism may not always be the disease-causing single-nucleotide polymorphism. Understanding the regulatory roles of epigenomic elements linked to risk variants helps elucidate the genetic mechanisms underlying complex traits. **(C)** Long-range gene regulation: This involves distal chromatin looping, which results in three-dimensional changes in chromatin structure. This occurs when DNA elements are situated more than 1–2 kb away from gene promoters. Techniques such as 3C, 4C, 5C, or Hi-C can map these spatial genomic interactions. Additionally, methods like ChIA-PET, 3C-ChIP, HiChIP, or PLAC-seq can assess genome-wide interactions involving regulatory proteins and chromatin looping. The figure was created with BioRender.com.

regions surrounded by histone-3 lysine-4 monomethylation (H3K4me1) and lysine-27 acetylation (H3K27ac), while H3K4me3 marks promoters. Insulators are responsible for organizing chromatin at a sub-compartment level. They are often bound by the TF CTCF (also known as 11-zinc finger protein or CCCTC-binding factor)[37] and establish the boundaries of so-called topologically associating domains (TADs). TADs are usually less than one megabase in size and trace those regions of our chromosomes in which sequences interact preferentially with each other instead of with elements in other regions in the genome. The prevalent model is that these TADs are formed by the dimerization of two CTCF molecules binding the boundaries of a TAD and are stabilized by the interaction with the ring-shaped cohesin complex through a process called loop extrusion.[38–40] Inside TADs, smaller DNA loops are assembled to allow enhancer–promoter interactions and, thus, regulation of transcription.[38,41] These enhancer–promoter loops, similar to the CTCF-mediated loops, are thought to be established by the binding and dimerization of the TF YY1 and its interaction with the cohesin complex[42,43] (Fig. 3).

Advancing enhancer identification: from chromatin features to high-throughput functional screening

Transcriptional enhancers were first described as DNA sequences that can enhance gene expression on an episomal plasmid (*e.g.*, a non-integrating, extrachromosomal circular DNA), irrespective of their location and orientation relative to the transcriptional start site (TSS)[44,45]; thus, enhancer identification was first confined to low-throughput reporter assays, where small fragments of DNA were tested for regulatory activity influencing reporter gene expression. Today's most widely applied experimental techniques for genome-wide identification of putative enhancers at the endogenous genomic locus do not rely directly on this functional property but on features that distinguish enhancers from non-regulatory regions at the chromatin level. Indeed, enhancers are bound by TFs and transcription coactivators and are located in open chromatin regions depleted of nucleosomes. The surrounding nucleosomes have specific histone tail modifications, such as H3K4me1 and H3K27ac. Moreover, some enhancers are bi-directionally transcribed in enhancer RNAs (eRNAs). However, even though these features correlate with enhancers, other genomic regions share the same chromatin characteristics, and more functional tests are required to prove that putative enhancers have a direct functional role in gene regulation.[46] This led to the development of high-throughput functional screenings known as massively parallel reporter assays (MPRAs) that quantify the enhancer activity of millions of sequences. The most widely used methods to identify putative regulatory regions are described in Table 1.

Enhancers, necessary cis-regulatory sequences, are typically found in chromatin regions that lack nucleosomes. DNase sequencing, FAIRE sequencing, and assay for transposase-accessible chromatin (ATAC) sequencing can be used to identify accessible DNA regions genome-wide. DNase sequencing uses nuclease digestion to identify

Table 1 Comparison of methods for identifying non-coding regulatory genome.

Method	Description	Advantages	Disadvantages	Non-coding elements and ncRNAs detected
ChiP sequencing	Captures DNA-protein interactions, revealing binding sites of transcription factors and histone modifications	Provides direct evidence of protein-DNA interactions and has high specificity in identifying regulatory elements	Requires specific antibodies; has potential bias due to antibody selection; may not capture transient interactions; cannot determine enhancer activity and identify target genes	Transcription factor binding sites, enhancers, silencers, promoters, lncRNAs
ATAC sequencing	Identifies open chromatin regions by the transposon Tn5, cuts the DNA, and inserts sequencing adapters	A rapid, low-input method that identifies open chromatin regions across the genome without prior knowledge	Limited to detecting chromatin accessibility; may miss specific regulatory elements; cannot determine enhancer activity and identify target genes	Open chromatin regions, enhancers, promoters, insulators, lncRNAs, eRNAs
DNase sequencing	Identifies open chromatin sites with high sensitivity and precision by employing the DNase I enzyme to cleave	High sensitivity in identifying open chromatin and precise mapping of regulatory regions	Requires large cell numbers; limited resolution	DNase I hypersensitive sites (DHSs), promoters, enhancers, lncRNAs

Table 1 (*continued*)

Method	Description	Advantages	Disadvantages	Non-coding elements and ncRNAs detected
	accessible DNA			
Hi-C and 3C	Captures long-range interactions between genomic regions, revealing three-dimensional genome architecture	Reveals long-range interactions and chromatin organization	Complicated data analysis may not directly identify specific regulatory elements	Enhancer—promoter loops, TADs, chromatin interactions
eRNA detection	Detects bidirectional transcription, indicating active enhancers through techniques such as GRO sequencing and CAGE	Identifies active enhancers with functional insights	Limited to detecting enhancers with bidirectional transcription	eRNAs
MPRAs	Quantifies the enhancer activity of thousands of sequences in parallel	A high-throughput screening method that identifies enhancers and regulatory potential	It may not fully recapitulate endogenous interactions; some constructs might behave differently	Enhancers, silencers, promoters
CRISPR-Cas9 screening	CRISPR technology perturbs non-coding regions to assess their impact on gene expression	Identifies functional enhancers and determines the endogenous effect of enhancer manipulation	Potential off-target effects; complicated data analysis may result in false negative results	Enhancers, promoters, silencers, regulatory sequences, lncRNAs, miRNAs
STARR sequencing	Identifies functional enhancers by a massive parallel reporter assay, where active enhancers drive their transcription	Identifies functional enhancers and quantitatively measures enhancer activity	Highly complicated plasmid libraries require a substantial number of cells for transfection; possible false negative results	Enhancers
Chromosome conformation capture	Detects topological interaction between two loci or genome wide	Identifies enhancer-target gene interaction	Cannot determine enhancer activity	Enhancer—promoter interactions, chromatin loops
Comparative genomics	Compares genomic sequences among species to identify evolutionarily conserved elements	Indicates functional significance and identifies conserved regulatory regions	Not all functional elements are conserved	Conserved non-coding elements, enhancers, promoters, lncRNAs, miRNAs, snoRNAs
Machine learning approaches	Predicts regulatory elements based on sequence motifs and chromatin features	A scalable, efficient method that provides predictions	Predictions may not always reflect functional elements accurately	Enhancers, promoters, silencers, insulators, lncRNAs, miRNAs, snoRNAs
Functional assays	Clones putative regulatory sequences into reporter constructs to test their impact on gene expression	Directly tests functionality and provides functional validation	Labor-intensive; may not capture complex *in vivo* interactions	Enhancers, promoters, silencers, insulators, lncRNAs, miRNAs, snoRNAs

Note: ChIP, chromatin immunoprecipitation; ATAC, assay for transposase-accessible chromatin; MPRAs, massively parallel reporter assays; STARR, self-transcribing active regulatory region; TADs, topologically associating domains; eRNA, enhancer RNA.

hypersensitive open chromatin regions, while FAIRE sequencing separates free and nucleosome-bound DNA. ATAC sequencing, the most recently developed method, uses a transposon to identify open chromatin regions.[47] However, these methods only identify putative enhancers and cannot quantify their activity. They should be combined with other techniques that are more selective for enhancers. A significant advantage of these techniques is that they screen for putative regulatory regions unbiasedly. Recent studies have used ATAC sequencing and RNA sequencing to determine open chromatin regions and gene expression during human fetal brain development.

Chromatin immunoprecipitation (ChIP) is a technique used to study protein-DNA interactions. Next-generation sequencing technologies have allowed genome-wide mapping of protein-DNA binding sites (ChIP sequencing), primarily used to identify putative enhancers across the entire genome. However, the binding of a TF or the presence of histone modifications does not provide definitive evidence that a sequence acts as a transcriptional enhancer. Several studies have used ChIP sequencing for histone modifications to predict enhancers during human brain development and in the adult brain, providing essential insights into the evolution of humans.[48]

Transcription of enhancer sequences into eRNAs has been validated genome-wide through sequencing, and eRNAs are generally bidirectionally transcribed and not polyadenylated. Enhancer transcription correlates with the presence of other enhancer marks, but whether their expression is a cause or consequence of gene transcription is still debated. The presence of eRNAs is not required or sufficient in all instances, and methods that only consider eRNA transcription may oversimplify the identification of putative enhancers. It may need to capture the complete regulatory landscape.[49]

Methods for identifying enhancers can be limited in determining which genes they regulate, as enhancers can be found long distances from their target gene. Chromatin conformation techniques can identify enhancer—promoter interactions but may not directly measure functional regulatory activity. More functional tests are needed to validate regulatory activity, especially in dynamic regulatory interactions. A recent study on human brain development has provided new insights into previously uncharacterized regulatory interactions relevant to neuropsychiatric disorders.[50]

The commonly used techniques to identify regulatory elements are predictive and do not directly measure enhancer activity. Direct high-throughput functional tests of enhancer activity, such as MPRAs and CRISPR-Cas9-based screening, have the potential to address these shortcomings. MPRAs are high-throughput reporter assays where DNA sequences are inserted before the minimal promoter of a vector with a specific barcode sequence downstream of the open reading frame, which allows the simultaneous assessment of 1000s of sequences for enhancer activity in parallel.[51] Self-transcribing active regulatory region (STARR) sequencing is an adapted approach that allows testing millions of sequences in a single experiment.[52] The CRISPR-Cas9 system can manipulate non-coding regulatory elements at the endogenous chromatin context, allowing for the ablation of an enhancer sequence or mutation of the sequence via non-homologous end joining. This approach can be used to study a selected enhancer of interest or in high-throughput screenings with extensive libraries of gRNAs introduced in cells expressing Cas9. CRISPR-Cas9 can also be applied to edit the epigenome, and functional domains can be used to alter the status of a non-coding regulatory element, forcing its activation or inactivation.[53]

Promoters encircle the TSS of genes, playing a critical role in instigating the transcription process. Conversely, enhancers act as promoters of transcriptional activity.[54] They display a diverse range of positions relative to the TSS of the controlled gene—spanning from immediate adjacency to the promoter, stretching across numerous kilobases in an upstream or downstream orientation (referred to as cis), and occasionally even nesting within the introns of different genes. Intriguingly, enhancers operate independently of their specific location, and their impact on transcription remains unaffected by their orientation. A classic example of a long-range regulatory element is the limb SHH enhancer located around one megabase away from its target gene.[55]

Furthermore, a solitary enhancer can orchestrate the expression of multiple genes, and in parallel, individual genes can be subject to control from multiple enhancers. This intricate network results in a surplus of regulatory mechanisms, contributing to the robustness of observable traits. This redundancy in the system offers evolutionary advantages, enhancing adaptability and ensuring stability across generations.[56] Therefore, cis-regulatory sequences' positions, identities, and arrangements ultimately determine the time and place that each gene is transcribed. On a mechanistic level, enhancers directly affect the recruitment of the transcriptional machinery to the TSS of genes.[57,58] Crucial for this long-range control of gene expression by enhancers is the formation of enhancer—promoter loops, which preferentially occur within the neighborhood of a TAD by DNA bending. The general TFs and the RNA polymerase II bind to the promoter sequence, whereas the distal cis-regulatory sequences are bound by TFs, which orchestrate the transcription initiation rate. Enhancers include TF binding sites that typically consist of DNA motifs found at multiple locations in the genome, but are not necessarily all equally likely to be bound by the recognizing TF.[59] To provide higher than background activity, homotypic or heterotypic dimerization of transcription regulators increases their DNA binding affinity and specificity.[60] TF binding itself can also be influenced by DNA methylation, which is established by DNA methyltransferases.[61]

Moreover, TF binding prevents the DNA from rewrapping around the nucleosome. In that case, it increases the likelihood that a second transcription regulator binds to the DNA, increasing the cooperative effect to the extent of displacing the histone core of the nucleosome.[62,63] Multiple TFs have been found to bind cooperatively in TF binding site hotspots,[64] later called stretch enhancers[65] or super-enhancers (SEs).[66] The latter is described as long regions with an increased density of enhancer elements characterized by a strong enrichment of H3K27ac and TFs and Mediator binding.[66,67] On the one hand, many studies suggest that SEs represent a novel class of non-coding regulatory elements that maintain, define, and control mammalian cell identity and whose transcriptional regulatory output is more significant than that of the individual

enhancer constituents.[66,68–70] On the other hand, an increasing number of studies have challenged this view and considered super-enhancers as a collection of typical enhancers that together do not have a more considerable activity than the sum of the individual parts.[71,72] Therefore, the debate on whether SEs are a new class of non-coding regulatory elements or whether they reflect a clustering of normal non-coding regulatory elements within proximity remains to be resolved.

Examining non-coding genome regulation: unlocking targeted therapeutic interventions

In this technology-driven era, computational data analysis involving various high-throughput sequencing technologies has shifted the genome paradigm from the presence of only known discrete hereditable coding entities to the presence of mysterious elements of the genome that might carry out some critical biological functions. Many repeated DNA segments like centromeres, transposable elements, satellite DNA, telomeres, and introns were present in the non-coding regions, revealing vital structural and functional roles at the chromosomal level. Besides, it was found that certain types of active ncRNA molecules have critical regulatory functions in controlling the expression of many genes in cells. The importance of these non-coding segments in the cutting and splicing of transposons for reassembly, genomic rearrangements, and the production of small RNAs that could serve as a source for new exons needs to be further investigated. Given this gap in knowledge and increasing evidence showing that alterations in these non-coding regions are linked to many diseases,[73] including ALS,[74] emphasis on expanding research in this focus area is warranted.

The use of next-generation sequencing technologies in genome-wide association studies (GWAS) has identified that nearly 90% of risk loci were located in the non-coding regions of the genome.[75,76] This is unsurprising given that targeted sequencing, whole-exome sequencing, and whole-genome sequencing analyses show a high percentage of variants/mutations are present in non-coding regions.[77] These non-coding regions are defined as those that are in the intronic or promoter regions, small non-coding RNAs such as miRNAs, lncRNAs, antisense and enhancer or insulator regions.[78]

The insights gained through the integration of advanced technological platforms and computational methodologies significantly contribute to the evolution of precision medicine. Specifically, these approaches will facilitate the development of targeted therapeutic interventions aimed at correcting the dysregulation of gene expression, which plays a critical role in perturbing diverse cellular pathways and biological processes.

At present, 11 RNA-based therapies have received approval from the US Food and Drug Administration (FDA) or European Medicines Agency (EMA) to target gene regulation in several organs, including the liver, muscles, and central nervous system (Table 2). These therapies include siRNAs and antisense oligonucleotides (ASOs) that either down-regulate specific genes or modulate pre-mRNA splicing by promoting the skipping or inclusion of exons. Additionally,

Table 2 RNA-based therapies in phase II or III clinical development.

Therapeutic agent	Category	Formulation and delivery	Delivery mode	Target site	Disease	Gene target & mechanism	Regulatory approval year	Development phase & approval
Fomivirsen (Vitravene)	21-mer ASO	1st gen; PT	Intravitreal	Eye	Cytomegalovirus (CMV) retinitis in immunocompromised patients	CMV IE-2 mRNA	1998 (FDA), 1999 (EMA)*	RNA therapeutics approved
Mipomersen (Kynamro)	20-mer ASO	2nd gen; 2′-MOE gapmer	Subcutaneous	Liver	Homozygous familial hypercholesterolaemia	Apolipoprotein B mRNA	2012 (EMA), 2013 (FDA)	RNA therapeutics approved
Eteplirsen (Exondys 51)	30-mer ASO	3rd gen; 2′-MOE PMO	Intravenous	Muscle	Duchenne muscular dystrophy	Dystrophin (DMD) pre-mRNA splicing (exon 51 skipping)	2016 (FDA)	RNA therapeutics approved
Nusinersen (Spinraza, ASO-10-27)	18-mer ASO	2nd gen; 2′-MOE	Intrathecal	Central nervous system	Spinal muscular atrophy	Survival of motor neuron 2 (SMN2) pre-mRNA splicing (exon 7 inclusion)	2017 (EMA), 2016 (FDA)	RNA therapeutics approved
Patisiran (Onpattro)	21 nt ds-siRNA	2nd gen; 2′-F/2′-O-Me; liposomal	Intravenous	Liver	Hereditary transthyretin amyloidosis	Transthyretin (TTR) mRNA	2018 (EMA), 2019 (FDA)	RNA therapeutics approved
Inotersen	20-mer ASO	2nd gen; 2′-MOE;	Subcutaneous	Liver	Hereditary	Transthyretin (TTR)	2018 (EMA), 2018	RNA therapeutics

(continued on next page)

Table 2 (*continued*)

Therapeutic agent	Category	Formulation and delivery	Delivery mode	Target site	Disease	Gene target & mechanism	Regulatory approval year	Development phase & approval
(Tegsedi, AKCEA-TTR-LRx)		GalNAc-conjugated			transthyretin amyloidosis	mRNA	(FDA)	approved
Volanesorsen (Waylivra)	20-mer ASO	2nd gen; 2′-MOE gapme	Subcutaneous	Liver	Familial chylomicronaemia syndrome	Apolipoprotein CIII (APOC3) mRNA	2019 (EMA)	RNA therapeutics approved
Golodirsen (Vyondys 53, SRP-4053)	25-mer ASO	3rd gen; 2′-MOE PMO	Intravenous	Muscle	Duchenne muscular dystrophy	DMDpre-mRNA splicing (exon 53 skipping)	2019 (FDA)	RNA therapeutics approved
Lumasiran (Oxlumo, ALN-GO1)	21 nt ds-siRNA	2nd gen; 2′-F/2′-O-Me; GalNAc-conjugated.	Subcutaneous	Liver	Primary hyperoxaluria type 1	Hydroxyacid oxidase-1 (HAO1) mRNA	2020 (EMA), 2020 (FDA)	RNA therapeutics approved
Inclisiran (Leqvio, ALN-PCSsc)	21 nt ds-siRNA	2nd gen; 2′-F/2′-O-Me; GalNAc-conjugated.	Subcutaneous	Liver	Atherosclerotic cardiovascular disease, elevated cholesterol, homozygous/ heterozygous familial hypercholesterolaemia	Proprotein convertase subtilisin/kexin type 9 (PCSK9) mRNA	2020 (EMA)	RNA therapeutics approved
Viltolarsen (Viltepso, NS-065, NCNP-01)	21-mer ASO	3rd gen; 2′-MOE PMO	Intravenous	Muscle	Duchenne muscular dystrophy	DMDpre-mRNA splicing (exon 53 skipping)	2020 (FDA)	RNA therapeutics approved
Givosiran (Givlaari)	21 nt ds-siRNA	2nd gen; 2′-F/2′-O-Me; GalNAc-conjugated	Subcutaneous	Liver	Acute hepatic porphyria	Delta aminolevulinic acid synthase 1 (ALAS1) mRNA	2020 (EMA), 2019 (FDA)	RNA therapeutics approved

Note: ASO, antisense oligonucleotide; GalNAc, N-acetylgalactosamine; LNA, locked nucleic acid; LODER, local drug eluter; NSCLC, non-small cell lung cancer; siRNA, small interfering RNA; SNP, single-nucleotide polymorphism.

many RNA therapeutics, including miRNA mimics and anti-miRs, are advancing through phase II or III clinical trials.[79]

Various therapeutic approaches harness the power of RNA and genetic materials to treat diseases, each utilizing different compounds and mechanisms. ASOs modulate gene expression through binding to specific RNA sequences, modulating gene expression. In the steric block mechanism, ASOs bind directly to target RNA, altering splicing patterns by either inducing exon skipping (*e.g.*, Eteplirsen[80]) or exon inclusion (*e.g.*, Nusinersen[81]) through interaction with splicing enhancers or silencers.[82] Specifically, Eteplirsen is a splice-modulating oligonucleotide used to treat Duchenne muscular dystrophy by binding to a splicing enhancer sequence in exon,[51] causing the spliceosome to skip exon[51] and read exon,[52] producing shorter but partially functional dystrophin proteins.[80] However, eteplirsen is effective only for 13%—14% of Duchenne muscular dystrophy patients with specific mutations. To address other mutations, drugs like golodirsen,[83] viltolarsen,[84] and casimersen[85] have been approved. These drugs promote dystrophin expression by inducing exon 53 or 45 skippings. Together with eteplirsen, these four drugs are third-generation ASOs with advanced chemical modifications. Given the urgent clinical needs, patient-specific oligonucleotide treatments have been developed, allowing for sequence-customized therapies.[82] ASOs can also trigger nonsense-mediated decay (NMD) by creating premature termination codons, leading to the degradation of faulty mRNAs.[86] Additionally, ASOs regulate translation by either down-regulating target RNA through translational arrest or up-regulating it by binding to translation-inhibitory elements. Lastly, ASOs can block miRNA-mediated repression by directly binding to miRNAs (as in miRNA inhibitors) or preventing miRNA—mRNA interactions (as in miRNA competitors), promoting target gene expression.[82]

Another method uses RNA interference (RNAi) to silence target genes by degrading their messenger RNA (mRNA). Patisiran is the first FDA-approved RNA interference (RNAi) therapy for treating hereditary transthyretin (hTTR) amyloidosis with polyneuropathy.[87] In a manner analogous to inotersen,[88] this siRNA (ALN-18328) silences all messenger RNAs with coding region mutations by specifically targeting the 3′ untranslated region of the TTR gene.[87] To improve the clinical effectiveness of siRNA therapies, Alnylam developed the GalNAc delivery platform, which has been used in about one-third of RNAi drugs currently in clinical trials. Revusiran, the first GalNAc-conjugated siRNA, demonstrated improved hepatic delivery by enhancing asialoglycoprotein receptor uptake but was discontinued due to an increased number of fatalities in the phase III "Endeavour" trial.[82] Lumasiran,[89] the second and third FDA-approved siRNA drugs, have shown that GalNAc-conjugated, subcutaneously delivered siRNAs are effective, well-tolerated, and reduce target mRNA levels with minimal side effects.[90]

Conclusion

The field of genomics has witnessed remarkable advancements over the past decades, leading to a deeper understanding of the human genome and its intricate regulatory mechanisms. Continuing advancements in genomic technologies and analytical methods will undoubtedly accelerate our understanding of the genetic basis of health and disease. By elucidating the complexities of the human genome, we pave the way for the development of targeted therapies, precision medicine interventions, and novel diagnostic tools. As we delve deeper into the genomic landscape, much remains to be explored, uncovered, and translated into tangible human health and well-being benefits.

CRediT authorship contribution statement

Paola Ruffo: Writing — review & editing, Writing — original draft, Methodology, Investigation, Conceptualization. **Bryan J. Traynor:** Writing — review & editing, Supervision, Resources. **Francesca Luisa Conforti:** Writing — review & editing, Supervision, Resources.

Conflict of interests

The authors declared no competing interests.

Acknowledgements

We would like to extend our deepest gratitude to the many researchers whose groundbreaking work and dedication have significantly advanced our understanding of genomic sciences. We also thank the members of the Department of Pharmacy of the University of Calabria (Italy) and the Neuromuscular Disease Research Section (NDRS) members at the National Institutes of Health for their support and collaboration. This research was supported entirely by the Intramural Research Program of the NIH, National Institute on Aging.

References

1. Collins FS, Patrinos A, Jordan E, Chakravarti A, Gesteland R, Walters L. New goals for the U.S. Human Genome Project: 1998-2003. *Science*. 1998;282(5389):682—689.
2. Moraes F, Góes A. A decade of human genome project conclusion: scientific diffusion about our genome knowledge. *Biochem Mol Biol Educ*. 2016;44(3):215—223.
3. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*. 2018;50(9):1219—1224.
4. Ruffo P, Cavallaro S, Conforti FL. The advent of omics sciences in clinical trials of motor neuron diseases. *J Personalized Med*. 2022;12(5):758.
5. Huber CD, Kim BY, Lohmueller KE. Population genetic models of GERP scores suggest pervasive turnover of constrained sites across mammalian evolution. *PLoS Genet*. 2020;16(5):e1008827.
6. Rands CM, Meader S, Ponting CP, Lunter G. 8.2% of the human genome is constrained: variation in rates of turnover across functional element classes in the human lineage. *PLoS Genet*. 2014;10(7):e1004525.
7. Ponting CP. Biological function in the twilight zone of sequence conservation. *BMC Biol*. 2017;15(1):71.

8. Ehret CF, De Haller G. Origin, development, and maturation of organelles and organelle systems of the cell surface in *Paramecium*. *J Ultrastruct Res*. 1963;9:1—42.
9. Ohno S. So much "junk" DNA in our genome. *Brookhaven Symp Biol*. 1972;23:366—370.
10. Orgel LE, Crick FH. Selfish DNA: the ultimate parasite. *Nature*. 1980;284(5757):604—607.
11. Graur D, Zheng Y, Azevedo RBR. An evolutionary classification of genomic function. *Genome Biol Evol*. 2015;7(3):642—645.
12. Kuska B. Should scientists scrap the notion of junk DNA? *J Natl Cancer Inst*. 1998;90(14):1032—1033.
13. Brosius J. Exaptation at the molecular genetic level. *Sci China Life Sci*. 2019;62(4):437—452.
14. Pennisi E. Genomics. ENCODE project writes eulogy for junk DNA. *Science*. 2012;337(6099):1159—1161.
15. Casa V, Gabellini D. A repetitive elements perspective in polycomb epigenetics. *Front Genet*. 2012;3:199.
16. Project Consortium ENCOEE, Birney E, Stamatoyannopoulos JA, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. 2007;447(7146):799—816.
17. Ruffo P, Strafella C, Cascella R, et al. Deregulation of ncRNA in neurodegenerative disease: focus on circRNA, lncRNA and miRNA in amyotrophic lateral sclerosis. *Front Genet*. 2021;12:784996.
18. Giunta S, Funabiki H. Integrity of the human centromere DNA repeats is protected by CENP-A, CENP-C, and CENP-T. *Proc Natl Acad Sci USA*. 2017;114(8):1928—1933.
19. Wu X, Brewer G. The regulation of mRNA stability in mammalian cells: 2.0. *Gene*. 2012;500(1):10—21.
20. Aires R, Jurberg AD, Leal F, Nóvoa A, Cohn MJ, Mallo M. Oct4 is a key regulator of vertebrate trunk length diversity. *Dev Cell*. 2016;38(3):262—274.
21. Stojic L, Niemczyk M, Orjalo A, et al. Transcriptional silencing of long noncoding RNA GNG12-AS1 uncouples its transcriptional and product-related functions. *Nat Commun*. 2016;7:10406.
22. Wittkopp PJ, Kalay G. *Cis*-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet*. 2011;13(1):59—69.
23. Lin A, Du Y, Xiao W. Yeast chromatin remodeling complexes and their roles in transcription. *Curr Genet*. 2020;66(4):657—670.
24. McFarland KN, Liu J, Landrian I, et al. SMRT sequencing of long tandem nucleotide repeats in SCA10 reveals unique insight of repeat expansion structure. *PLoS One*. 2015;10(8):e0135906.
25. Pagni S, Mills JD, Frankish A, Mudge JM, Sisodiya SM. Non-coding regulatory elements: potential roles in disease and the case of epilepsy. *Neuropathol Appl Neurobiol*. 2022;48(3):e12775.
26. Soldner F, Stelzer Y, Shivalila CS, et al. Parkinson-associated risk variant in distal enhancer of α-synuclein modulates target gene expression. *Nature*. 2016;533(7601):95—99.
27. Huang FW, Hodis E, Xu MJ, Kryukov GV, Chin L, Garraway LA. Highly recurrent TERT promoter mutations in human melanoma. *Science*. 2013;339(6122):957—959.
28. Killela PJ, Reitman ZJ, Jiao Y, et al. TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal. *Proc Natl Acad Sci USA*. 2013;110(15):6021—6026.
29. Pierini T, Nardelli C, Lema Fernandez AG, et al. New somatic TERT promoter variants enhance the telomerase activity in glioblastoma. *Acta Neuropathol Commun*. 2020;8(1):145.
30. Rachakonda PS, Hosen I, de Verdier PJ, et al. TERT promoter mutations in bladder cancer affect patient survival and disease recurrence through modification by a common polymorphism. *Proc Natl Acad Sci USA*. 2013;110(43):17426—17431.
31. Mattick JS, Amaral PP, Carninci P, et al. Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nat Rev Mol Cell Biol*. 2023;24(6):430—447.
32. Skinner MK, Nilsson EE. Role of environmentally induced epigenetic transgenerational inheritance in evolutionary biology: unified evolution theory. *Environ Epigenet*. 2021;7(1):dvab012.
33. van Steensel B, Belmont AS. *Lamina*-associated domains: links with chromosome architecture, heterochromatin, and gene repression. *Cell*. 2017;169(5):780—791.
34. Caputo V, Strafella C, Termine A, et al. Epigenomic signatures in age-related macular degeneration: focus on their role as disease modifiers and therapeutic targets. *Eur J Ophthalmol*. 2021;31(6):2856—2867.
35. Villaseñor R, Baubec T. Regulatory mechanisms governing chromatin organization and function. *Curr Opin Cell Biol*. 2021;70:10—17.
36. Hyun K, Jeon J, Park K, Kim J. Writing, erasing and reading histone lysine methylations. *Exp Mol Med*. 2017;49(4):e324.
37. Bell AC, West AG, Felsenfeld G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell*. 1999;98(3):387—396.
38. Dowen JM, Fan ZP, Hnisz D, et al. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell*. 2014;159(2):374—387.
39. Ong CT, Corces VG. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet*. 2014;15(4):234—246.
40. Jordan Rowley M, Corces VG. Organizational principles of 3D genome architecture. *Nat Rev Genet*. 2018;19(12):789—800.
41. Hnisz D, Day DS, Young RA. Insulated neighborhoods: structural and functional units of mammalian gene control. *Cell*. 2016;167(5):1188—1200.
42. Beagan JA, Duong MT, Titus KR, et al. YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. *Genome Res*. 2017;27(7):1139—1152.
43. Weintraub AS, Li CH, Zamudio AV, et al. YY1 is a structural regulator of enhancer-promoter loops. *Cell*. 2017;171(7):1573—1588.e28.
44. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell Res*. 2011;21(3):381—395.
45. Moreau P, Hen R, Wasylyk B, Everett R, Gaub MP, Chambon P. The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Res*. 1981;9(22):6047—6068.
46. Catarino RR, Stark A. Assessing sufficiency and necessity of enhancer activities for gene expression and the mechanisms of transcription activation. *Genes Dev*. 2018;32(3—4):202—223.
47. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol*. 2015;109:21.29.1—21.2921.29.9.
48. Markenscoff-Papadimitriou E, Whalen S, Przytycki P, et al. A chromatin accessibility atlas of the developing human telencephalon. *Cell*. 2020;182(3):754—769.e18.
49. Carullo NVN, Phillips Iii RA, Simon RC, et al. Enhancer RNAs predict enhancer-gene regulatory links and are critical for enhancer function in neuronal systems. *Nucleic Acids Res*. 2020;48(17):9550—9570.
50. Spiess K, Won H. Regulatory landscape in brain development and disease. *Curr Opin Genet Dev*. 2020;65:53—60.
51. Santiago-Algarra D, Dao LTM, Pradel L, España A, Spicuglia S. Recent advances in high-throughput approaches to dissect enhancer function. *F1000Res*. 2017;6:939.
52. Hansen TJ, Hodges E. ATAC-STARR-seq reveals transcription factor-bound activators and silencers within chromatin-accessible regions of the human genome. *Genome Res*. 2022;32(8):1529—1541.
53. Klann TS, Black JB, Chellappan M, et al. CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. *Nat Biotechnol*. 2017;35(6):561—568.

54. Banerji J, Rusconi S, Schaffner W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell*. 1981; 27(2 Pt 1):299—308.
55. Lettice LA, Heaney SJH, Purdie LA, et al. 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—1735.
56. Osterwalder M, Barozzi I, Tissières V, et al. Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature*. 2018;554(7691):239—243.
57. Stadhouders R, van den Heuvel A, Kolovos P, et al. Transcription regulation by distal enhancers: who's in the loop? *Transcription*. 2012;3(4):181—186.
58. Coulon A, Chow CC, Singer RH, Larson DR. Eukaryotic transcriptional dynamics: from single molecules to cell populations. *Nat Rev Genet*. 2013;14(8):572—584.
59. Lambert SA, Jolma A, Campitelli LF, et al. The human transcription factors. *Cell*. 2018;175(2):598—599.
60. Funnell APW, Crossley M. *Homo-* and heterodimerization in transcriptional regulation. *Adv Exp Med Biol*. 2012;747: 105—121.
61. Yin Y, Morgunova E, Jolma A, et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*. 2017;356(6337):eaaj2239.
62. Iwafuchi-Doi M, Donahue G, Kakumanu A, et al. The pioneer transcription factor FoxA maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. *Mol Cell*. 2016;62(1):79—91.
63. Iwafuchi-Doi M. The mechanistic basis for chromatin regulation by pioneer transcription factors. *WIREs Mechanisms Disease*. 2019;11(1):e1427.
64. Siersbæk R, Nielsen R, John S, et al. Extensive chromatin remodelling and establishment of transcription factor 'hotspots' during early adipogenesis. *EMBO J*. 2011;30(8):1459—1472.
65. Parker SCJ, Stitzel ML, Taylor DL, et al. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc Natl Acad Sci U S A*. 2013; 110(44):17921—17926.
66. Whyte WA, Orlando DA, Hnisz D, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*. 2013;153(2):307—319.
67. Hnisz D, Abraham BJ, Lee TI, et al. Super-enhancers in the control of cell identity and disease. *Cell*. 2013;155(4):934—947.
68. Young RA. Control of the embryonic stem cell state. *Cell*. 2011; 144(6):940—954.
69. Whyte WA, Bilodeau S, Orlando DA, et al. Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature*. 2012;482(7384):221—225.
70. Lovén J, Hoke HA, Lin CY, et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*. 2013;153(2): 320—334.
71. Shin HY, Willi M, HyunYoo K, et al. Hierarchy within the mammary STAT5-driven Wap super-enhancer. *Nat Genet*. 2016; 48(8):904—911.
72. Hay D, Hughes JR, Babbs C, et al. Genetic dissection of the α-globin super-enhancer *in vivo*. *Nat Genet*. 2016;48(8): 895—903.
73. Scacheri CA, Scacheri PC. Mutations in the noncoding genome. *Curr Opin Pediatr*. 2015;27(6):659—664.
74. Yousefian-Jazi A, Sung MK, Lee T, Hong YH, Choi JK, Choi J. Functional fine-mapping of noncoding risk variants in amyotrophic lateral sclerosis utilizing convolutional neural network. *Sci Rep*. 2020;10(1):12872.
75. Edwards SL, Beesley J, French JD, Dunning AM. Beyond GWASs: illuminating the dark road from association to function. *Am J Hum Genet*. 2013;93(5):779—797.
76. Mirza AH, Kaur S, Brorsson CA, Pociot F. Effects of GWAS-associated genetic variants on lncRNAs within IBD and T1D candidate loci. *PLoS One*. 2014;9(8):e105723.
77. Nakagawa H, Fujita M. Whole genome sequencing analysis for cancer genomics and precision medicine. *Cancer Sci*. 2018; 109(3):513—522.
78. Giral H, Landmesser U, Kratzer A. Into the wild: GWAS exploration of non-coding RNAs. *Front Cardiovasc Med*. 2018;5:181.
79. Winkle M, El-Daly SM, Fabbri M, Calin GA. Noncoding RNA therapeutics - challenges and potential solutions. *Nat Rev Drug Discov*. 2021;20(8):629—651.
80. Nakamura A, Takeda S. Exon-skipping therapy for Duchenne muscular dystrophy. *Lancet*. 2011;378(9791):546—547.
81. Corey DR. Nusinersen, an antisense oligonucleotide drug for spinal muscular atrophy. *Nat Neurosci*. 2017;20(4):497—499.
82. Zhu Y, Zhu L, Wang X, Jin H. RNA-based therapeutics: an overview and prospectus. *Cell Death Dis*. 2022;13(7):644.
83. Heo YA. Golodirsen: first approval. *Drugs*. 2020;80(3):329—333.
84. Dhillon S. Viltolarsen: first approval. *Drugs*. 2020;80(10): 1027—1031.
85. Shirley M. Casimersen: first approval. *Drugs*. 2021;81(7): 875—879.
86. Ward AJ, Norrbom M, Chun S, Frank Bennett C, Rigo F. Nonsense-mediated decay as a terminating mechanism for antisense oligonucleotides. *Nucleic Acids Res*. 2014;42(9): 5871—5879.
87. Adams D, Gonzalez-Duarte A, O'Riordan WD, et al. Patisiran, an RNAi therapeutic, for hereditary transthyretin amyloidosis. *N Engl J Med*. 2018;379(1):11—21.
88. Benson MD, Waddington-Cruz M, Berk JL, et al. Inotersen treatment for patients with hereditary transthyretin amyloidosis. *N Engl J Med*. 2018;379(1):22—31.
89. Scott LJ, Keam SJ. Lumasiran: first approval. *Drugs*. 2021; 81(2):277—282.
90. Kang C. Lumasiran: a review in primary hyperoxaluria type 1. *Drugs*. 2024;84(2):219—226.