ARCHIVES OF BIOINFORMATICS AND COMPUTATIONAL BIOLOGY
Jul-Dec 2025, VOL. 1, ISSUE 2, pp. 1-10
https://doi.org/10.61577/abcb.2025.100006

RESEAPRO JOURNALS

ORIGINAL ARTICLE

OPEN ACCESS

Comparative genomics of non-coding DNA: Evolutionary insights into regulatory elements and lncRNAs in vertebrates

Nazish Parveen Quraishi

Department of Bioinformatics, University of Allahabad, Prayagraj, India

ABSTRACT

The completion of vertebrate genome projects revealed that only a small fraction of DNA encodes proteins, while most consists of non-coding regions containing essential regulatory elements. Comparative genomics offers a powerful framework to identify conserved non-coding elements (CNEs), enhancers, and long non-coding RNAs (lncRNAs) that govern gene expression and drive vertebrate diversity. This review summarizes advances in identifying and functionally annotating non-coding regulatory elements through comparative and integrative genomics. Using case studies such as the ZRS enhancer of Shh, Pitx1 pelvic enhancer, Human Accelerated Regions (HARs), and conserved lncRNAs like Fendrr and HOTTIP, we illustrate how subtle non-coding variations shape morphology and lineage-specific traits. Overall, the evolution of gene regulation rather than protein-coding changes alone emerges as a major force in vertebrate evolution. Integrating functional genomics, single-cell transcriptomics, and machine learning will further clarify how non-coding DNA contributes to adaptation and disease.

KEY WORDS

Non-coding DNA; Integrative genomics; Gene regulation; Single-cell transcriptomics; Vertebrate genomics; Single-cell transcriptomics

ARTICLE HISTORY

Received 12 July 2025;
Revised 01 August 2025;
Accepted 08 August 2025

Introduction

The sequencing of multiple vertebrate genomes over the past two decades has fundamentally transformed our understanding of genome organization, revealing that the vast majority of functional information resides outside of protein-coding regions. In humans, more than 98% of the genome is composed of non- coding DNA, encompassing a diverse repertoire of regulatory elements including enhancers, silencers, insulators, promoters, and long non-coding RNAs (lncRNAs) that together orchestrate precise spatial and temporal control of gene expression (Figure 1). Far from representing "junk DNA," these elements constitute the regulatory circuitry that underpins embryonic development, cell identity, and physiological homeostasis. Disruption of these elements is increasingly implicated in developmental disorders, cancer, and complex diseases, underscoring their biological and clinical significance [1-3].

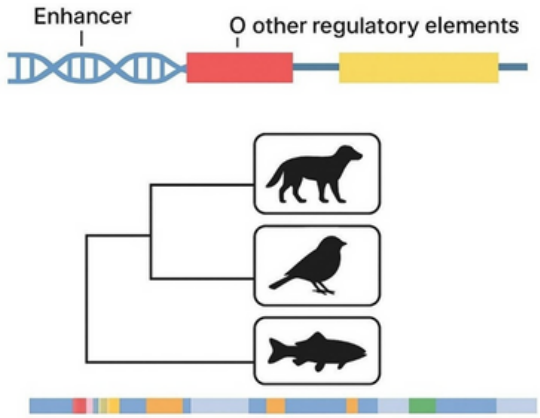

Figure 1. Comparative genomics identifies conserved non-coding elements such as enhancers and lncRNAs across vertebrates. Sequence conservation and divergence among species reveal how non-coding DNA regulates gene expression and drives evolutionary diversity.

Comparative genomics has emerged as a powerful framework for the systematic identification of functional non-coding elements. The approach leverages evolutionary constraint as a proxy for functionality: sequences that have been preserved across hundreds of millions of years are likely to be under purifying selection and thus essential for organismal fitness. Multi-species genome alignments and conservation scoring tools have identified thousands of conserved non-coding elements (CNEs), which are frequently enriched near key developmental regulators such as Hox, Sox, and Pax genes (Figure 2). Functional studies demonstrate that many CNEs act as enhancers with exquisitely specific spatiotemporal expression patterns during embryogenesis [4].

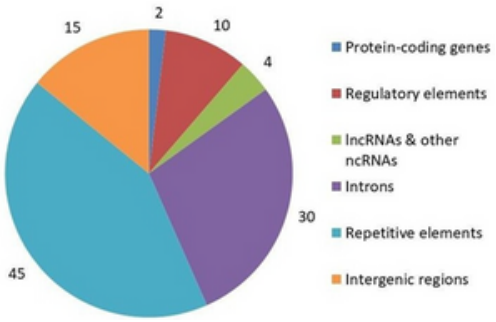

Figure 2. Compositionof the human genome.

The pie chart illustrates the proportional organization of major genomic components, emphasizing the predominance of non-coding regions. Protein-coding genes constitute only ~2% of the human genome, while the remaining ~98% comprises diverse non-coding sequences with regulatory, structural, and evolutionary functions. Repetitive elements (~45%) and introns (~30%) represent the largest genomic

***Correspondence:** Nazish Parveen Quraishi, Department of Bioinformatics, University of Allahabad, Prayagraj, India, e-mail: nazishparveenquraishi@gmail.com

fractions, followed by intergenic regions (~15%), regulatory elements (~10%), and long non-coding RNAs (lncRNAs) and other ncRNAs (~4%). Collectively, these non-coding components form an intricate regulatory landscape that governs gene expression, chromatin organization, and genome evolution.

Notably, comparative approaches also reveal lineage-specific regulatory innovations that underpin phenotypic diversity. Human accelerated regions (HARs), for example, are genomic segments that were highly conserved across vertebrates but underwent rapid evolution in the human lineage, implicating regulatory rewiring in the emergence of human-specific traits [5, 6]. Similarly, mutations in the deeply conserved ZRS enhancer of the Shh gene are associated with limb reduction and loss in snakes, illustrating how modification of a single cis-regulatory element can produce dramatic morphological changes [7, 8]. These findings highlight the dual nature of the non-coding genome: maintaining developmental stability while serving as a substrate for evolutionary innovation.

Recent advances in epigenomics have synergized with comparative approaches to refine the functional annotation of regulatory elements. Genome-wide profiling of chromatin states, histone modifications (e.g., H3K27ac, H3K4me1), and open chromatin landscapes using ATAC-seq and DNase-seq has enabled the identification of active regulatory regions across cell types and developmental stages. Integrating these epigenomic signatures with evolutionary conservation has revealed that regulatory programs active during the phylotypic stage of vertebrate development are among the most deeply conserved [3, 9, 10].

Types of Non-Coding Regulatory Elements in Vertebrate Genomes

The completion of vertebrate genome projects has revealed that only ~1–2% of the genome encodes proteins, while the vast majority comprises non-coding regions with diverse regulatory and structural functions [3, 11]. Understanding this "dark matter" of the genome remains a central challenge in functional genomics (Table 1). Comparative genomics has emerged as a powerful framework for classifying and functionally annotating these non-coding elements, providing insights into their roles in development, physiology, and evolution [12, 13].

Table 1. Types of non-coding regulatory elements in vertebrate genomes.

Element	Definition	Primary Functions	Case Studies	Evolutionary Insights	Key References
Enhancers	50–1500 bp cis-regulatory elements that activate transcription in position- and orientation-independent manner	Recruit TFs and cofactors; establish chromatin loops to activate distant promoters	ZRS enhancer of Shh; Sox9, Pax6 enhancers; VISTA-validated enhancers	Show deep conservation and rapid turnover; lineage-specific enhancers contribute to species-specific traits	[8, 14-16]
Human Accelerated Regions (HARs)	Non-coding elements conserved across vertebrates but with excess substitutions in human lineage	Many Developmental enhancers, especially in brain	HAR1 (neocortex), HAR2/HACNS1 (limb enhancer)	Indicate human-specific regulatory innovations linked to brain evolution and morphology	[17,18]
Conserved Non-Coding Elements (CNEs)	≥70–100 bp with high identity between mammals/fish; UCEs ≥200 bp conserved across mammals	Developmental Enhancers, silencers, splicing regulators, chromatin organizers	Pitx1 pelvic enhancer; CNE clusters near Hox, Sox, Otx	Retained >400 Myr; purifying selection maintains core body plans; lineage-specific loss linked to morphological changes	[19,20]
Long Non-Coding RNAs (lncRNAs)	Transcripts >200 nt, little/no coding potential, often syntenically conserved	Chromatin remodeling, transcriptional interference, scaffolding, RNA processing	Xist (X- inactivation), Malat1 (RNA Splicing), Terc (Telomerase RNA)	Rapid sequence evolution but conserved synteny/expression; essential for development and dosage compensation	[21,22,23]
Silencers	Negative regulatory elements that repress gene transcription	Recruit repressors and chromatin-modi fying complexes	Neural-restrictive silencer element (NRSE) controlling neural genes	Often under Strong selection to maintain spatiotemporal repression patterns	[24]
Insulators / Boundary Elements	DNA elements that block enhancer-promoter communication or define chromatin domains	Establish TAD boundaries; mediate chromatin architecture via CTCF/cohesin	CTCF binding sites at TAD boundaries	Evolutionarily conserved in mammals; maintain 3D genome integrity	[25, 26]

Enhancers- modular switches of gene expression

Enhancers are short cis-regulatory DNA elements, typically 50–1,500 bp in length, that increase transcription of target genes in a position- and orientation-independent manner [14]. They function by recruiting transcription factors and cofactors, promoting chromatin looping, and establishing spatial proximity between distal regulatory regions and promoters within the three-dimensional nuclear architecture [15, 17].

https://doi.org/10.61577/abcb.2025.100006

Comparative genomics and discovery

Multi-species genome alignments have identified thousands of evolutionarily conserved enhancer candidates, frequently clustered near key developmental genes [16, 18]. Resources such as the VISTA Enhancer Browser and ENCODE project have experimentally validated hundreds of enhancers, generating tissue-specific functional annotations across multiple species [19,20].

Evolutionary dynamics

Enhancers exhibit both deep conservation and rapid evolutionary turnover. Highly conserved enhancers regulate fundamental developmental pathways, whereas lineage-specific enhancers contribute to phenotypic innovation and species-specific traits [18, 21].

Case study: Shh limb enhancer (ZRS)

The limb-specific Shh enhancer ZRS is deeply conserved across vertebrates but contains snake-specific mutations linked to limb reduction [8].

Human accelerated regions (HARs)

Several HARs function as enhancers with human-specific activity in the developing neocortex, suggesting roles in the evolution of human cognition [5, 22].

Conserved non-coding elements (CNEs)- evolutionary anchors

CNEs are stretches of non-coding DNA showing exceptionally high conservation across distantly related vertebrates, typically ≥70–100 bp with ≥70% identity between mammals and fish [1, 4]. A notable subset, ultra conserved elements (UCEs), maintains perfect sequence conservation over ≥200 bp across human, mouse, and rat genomes [23].

Functional roles

CNEs are enriched near developmental transcription factor genes and often function as enhancers, silencers, or other cis-regulatory modules [16;24]. In vivo reporter assays demonstrate that many CNEs drive highly specific spatiotemporal expression patterns during embryogenesis [25, 26].

Evolutionary insights

CNEs have been preserved under strong purifying selection for over 400 million years, highlighting their critical roles in developmental robustness and body plan conservation [9, 13]. However, lineage-specific loss or modification of CNEs has contributed to morphological diversification, exemplified by pelvic reduction in sticklebacks caused by deletion of a Pitx1 enhancer [27].

Long non-coding RNAS (lncRNAS) – The regulatory transcriptome

LncRNAs are transcripts longer than 200 nucleotides with minimal protein-coding potential, yet they perform diverse regulatory functions, including chromatin remodeling, transcriptional interference, scaffolding of molecular complexes, and post-transcriptional regulation [28, 29].

Comparative genomics

Although lncRNAs exhibit lower primary sequence conservation than protein-coding genes, many retain conserved synteny and tissue-specific expression patterns across vertebrates [30, 31]. A subset including Xist (X-chromosome inactivation), Malat1 (RNA processing), and Terc (telomerase RNA component) is deeply conserved and essential for development [32].

Functional validation

CRISPR/Cas9-mediated loss-of-function studies demonstrate that numerous lncRNAs exert significant phenotypic effects, underscoring their roles as critical regulators of transcriptional and epigenetic programs [33, 34].

Additional regulatory elements

Beyond enhancers, CNEs, and lncRNAs, additional cis-regulatory sequences also contribute to transcriptional control:

Silencers

Repressive elements that downregulate gene expression by recruiting transcriptional repressors or chromatin-modifying complexes [35]. Genome-wide silencer maps reveal their abundance and context- dependent functionality, with disease-associated variants frequently enriched in these regions.

Insulators/boundary elements

These sequences define topologically associating domain (TAD) boundaries and restrict inappropriate enhancer–promoter interactions. Typically marked by CTCF binding and cohesin recruitment, they are essential for maintaining higher-order chromatin organization [36, 37].

Comparative Genomics Approaches and Tools

Comparative genomics has become an indispensable framework for uncovering and functionally characterizing non-coding regulatory elements in vertebrate genomes. By integrating high-quality genome assemblies, evolutionary conservation analyses, epigenomic profiling, and experimental perturbation studies, researchers can systematically resolve the regulatory architecture underlying development, physiology, and disease susceptibility. Figure 3 provides a schematic overview of the major steps in a comparative genomics workflow, and Table 2 summarizes the key approaches, computational tools, and resources discussed in this section.

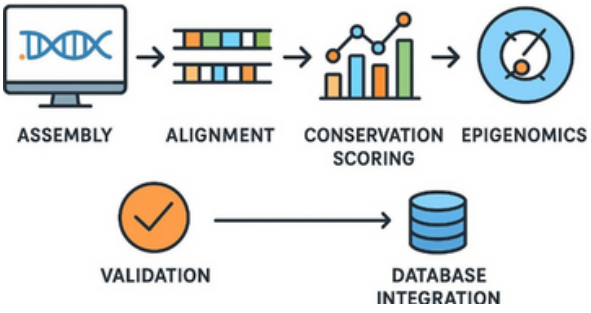

Figure 3. Comparative genomicsworkflow for identifying conserved regulatory elements.

https://doi.org/10.61577/abcb.2025.100006

Schematic overview of the major analytical steps in comparative genomics. Genome assemblies from multiple species are aligned to detect conserved regions, followed by conservation scoring to assess evolutionary constraint. Epigenomic data are integrated to refine functional annotation, and candidate elements are validated experimentally through reporter or CRISPR-based assays. Validated elements are then incorporated into public databases, enabling systematic exploration of non-coding regulatory sequences and their roles in vertebrate evolution and development.

The foundation of comparative genomics is the availability of high-quality reference genomes from a broad range of vertebrate lineages. Advances in long-read sequencing technologies such as PacBio HiFi and Oxford Nanopore together with Hi-C–based scaffolding have significantly improved genome assembly contiguity and completeness, enabling precise identification of conserved non-coding regions [38]. Multi-species alignments are then constructed using robust frameworks such as MULTIZ [39]. Progressive Cactus [40, 41], and LASTZ [42], which enable accurate orthology mapping across hundreds of species. These alignments form the essential substrate for identifying conserved non-coding elements (CNEs) and quantifying evolutionary turnover.

Conservation scoring and evolutionary constraint

Multi-species alignments are interrogated using phylogenetic models that detect signatures of purifying selection:

PhastCons applies a hidden Markov model to classify genomic regions as conserved or non-conserved based on substitution rates across the phylogeny [43].

GERP++ calculates "rejected substitutions," representing the number of mutations removed by purifying selection, thereby quantifying evolutionary constraint [44].

PhyloP provides both base-wise conservation and acceleration scores, enabling the detection of deeply conserved elements as well as lineage-specific regulatory innovations [45].

These conservation scores are widely used to prioritize candidate enhancers, silencers, and other cis- regulatory modules for experimental validation.

Epigenomic integration for functional annotation

Although sequence conservation provides a valuable indicator of evolutionary constraint, not all conserved elements perform regulatory roles, and conversely, many regulatory elements exhibit weak or no conservation. To address these limitations, conservation signals are increasingly integrated with epigenomic datasets that capture cell type–specific regulatory activity.

Histone modification profiles offer critical insights into enhancer states: H3K4me1 enrichment marks poised enhancers, while H3K27ac distinguishes active enhancers from poised ones. Similarly, chromatin accessibility assays such as DNase-seq and ATAC-seq identify regions of open chromatin that are accessible to transcription factors, thereby highlighting potential cis-regulatory modules.

Comprehensive initiatives such as the ENCODE Project and the Roadmap Epigenomics Program have generated large-scale maps of chromatin states, histone modifications, and accessibility landscapes across diverse tissues and developmental stages. By integrating these datasets with conservation analyses, researchers can prioritize and functionally annotate conserved elements with higher confidence, supporting the discovery of context-specific regulatory elements that underlie development, cellular identity, and disease [20, 46-49].

Functional validation of candidate elements

Computational predictions of regulatory elements must be corroborated through experimental validation to establish causal relationships with gene regulation. Traditional reporter assays such as luciferase or GFP-based constructs remain widely used to test enhancer activity in vitro or in transgenic organisms [50]. More recently, Massively Parallel Reporter Assays (MPRAs) have enabled the simultaneous functional interrogation of thousands of candidate elements, providing quantitative readouts of enhancer activity at scale [51, 52]. In addition, CRISPR-based perturbation strategies, including targeted deletions, CRISPR interference (CRISPRi), and CRISPR activation (CRISPRa), allow assessment of enhancer necessity and sufficiency in their native genomic context [53, 54]. These complementary approaches provide mechanistic insight into how specific regulatory elements orchestrate transcriptional programs in development and disease.

Databases and computational resources

A wide range of publicly accessible databases and computational resources facilitate the identification, annotation, and functional interpretation of conserved non-coding regulatory elements across vertebrate genomes. These repositories integrate comparative, epigenomic, and experimental datasets, enabling comprehensive analyses of genome regulation and evolution:

VISTA enhancer browser

A curated catalog of in vivo–validated enhancer elements tested in transgenic mice, providing insights into spatiotemporal enhancer activity during development [26].

UCNEbase

A specialized repository of ultra-conserved non-coding elements (UCNEs) and their associated genomic regulatory blocks, emphasizing their evolutionary conservation and developmental significance [55].

Ensembl compara

Provides multi-species genome alignments, orthology predictions, and conservation tracks across hundreds of species, forming a cornerstone for large-scale comparative genomics analyses [56].

FANTOM5

A comprehensive atlas of promoters and enhancers identified through Cap Analysis of Gene Expression (CAGE), offering a global view of transcriptional regulation across diverse human and mouse cell types [57].

ENCODE / roadmap epigenomics

Large-scale consortia providing extensive reference datasets of chromatin accessibility, histone modifications, and transcription factor binding across tissues, supporting high-resolution annotation of regulatory elements [10, 20].

These integrated resources collectively support the discovery and functional prioritization of regulatory elements, bridging computational prediction with experimental validation (Table 2).

https://doi.org/10.61577/abcb.2025.100006

Table 2. Summarizes the key approaches, computational tools, and resources.

Step	Approach / Tools	Key Contribution	References
Genome Assembly & Alignment	PacBio, ONT, Hi-C, MULTIZ, Progressive Cactus, LASTZ	High-contiguity assemblies; orthology mapping	[38,39]
Conservation Scoring	PhastCons, GERP++, PhyloP	Detect functional constraint across species	[43,44]
Epigenomic Integration	H3K27ac, H3K4me1, DNase-seq, ATAC-seq, ENCODE, Roadmap	Prioritization of active regulatory elements	[20]
Functional Validation	Reporter assays, MPRA, CRISPRi/a	Experimental confirmation of enhancer activity	[51]
Databases & Resources	VISTA, UCNEbase, Ensembl Compara, FANTOM5	Curated datasets of conserved/validated CREs	[19,55]

CaseStudies

Comparative genomics has been instrumental in linking conserved and divergent non-coding elements to phenotypic diversity across vertebrates. Below, we highlight four well-characterized examples where regulatory elements and lncRNAs have been functionally validated and connected to evolutionary innovations or disease phenotypes. These case studies collectively demonstrate that non-coding elements, including enhancers and lncRNAs, are powerful drivers of phenotypic diversity. They show that evolutionary changes in regulatory DNA, whether through point mutations, deletions, or accelerated substitution rates, can remodel gene expression programs and generate morphological and cognitive innovations.

Databases such as UCNEbase, which catalog ultraconserved non-coding elements (UCNEs) and their organization into genomic regulatory blocks, have further enhanced our ability to identify and functionally annotate conserved regulatory regions across species [55]. Integrating multi-species comparative genomics with experimental validation remains the gold standard for linking sequence evolution to phenotype (Figure 4).

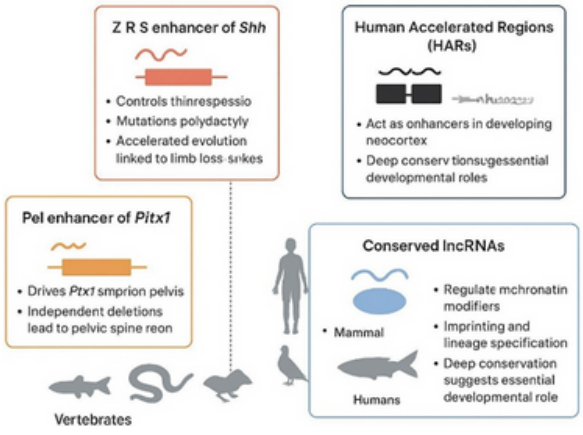

Figure 4. Comparative genomic case studieslinking non-coding elementsto vertebrate evolutionand development.

Representative examples showing how conserved and lineage-specific non-coding elements drive vertebrate diversity. ZRS enhancer of Shh controls limb patterning; mutations cause polydactyly and snake limb loss. Pel enhancer of Pitx1 regulates pelvic development in sticklebacks, with deletions leading to recurrent reduction. Human accelerated regions (HARs) act as enhancers in neocortical development, contributing to human-specific traits. Conserved lncRNAs such as Fendrr modulate chromatin and developmental gene networks, underscoring the evolutionary significance of non-coding DNA in vertebrate form and function.

ZRS enhancer of Shh and limb development

The zone of polarizing activity regulatory sequence (ZRS) is a long-range enhancer of the Sonic hedgehog (Shh) gene, located nearly 1 Mb upstream within intron 5 of Lmbr1. The ZRS orchestrates Shh expression in the posterior limb bud, specifying the anterior–posterior axis of the developing limb [57].

Comparative genomic analyses reveal that the ZRS is deeply conserved across vertebrates from fish to mammals underscoring its essential role in limb patterning [8]. Notably, accelerated substitution rates within the ZRS have been identified in limbless squamates such as snakes, correlating with reduced enhancer activity and the evolutionary loss of limbs [8].

Functional evidence from mouse transgenic assays and human genetic studies demonstrates that point mutations or duplications within key transcription factor binding motifs (e.g., ETS sites) can ectopically activate Shh in the anterior limb bud, resulting in preaxial polydactyly phenotypes [58, 59]. Thus, the ZRS exemplifies how subtle sequence perturbations within a highly conserved enhancer can drive significant morphological evolution.

Pel enhancer of Pitx1 and pelvic reduction in stickleback fish

The Pel enhancer regulates the expression of the hindlimb transcription factor Pitx1 during pelvic development in three spine sticklebacks (Gasterosteus aculeatus). Multiple freshwater stickleback populations exhibit recurrent pelvic reduction resulting from independent deletions or disruptions of this enhancer [27].

Comparative genomic analyses demonstrate that, while the Pitx1 coding sequence remains intact across populations, the loss of the Pel enhancer leads to complete downregulation of Pitx1 expression in the pelvic region producing a loss of pelvic structures. This represents a striking example of parallel evolution through cis-regulatory modification [27, 60].

Functional rescue experiments further validate this relationship: reintroduction of the Pel enhancer into pelvic-reduced fish restores Pitx1 expression and spine formation, confirming its causal role in morphological evolution [61].

This case exemplifies how morphological diversity often originates from mutations in regulatory elements rather

https://doi.org/10.61577/abcb.2025.100006

than protein-coding regions, emphasizing that cis-regulatory changes are powerful substrates for natural selection and phenotypic innovation.

Human accelerated regions (HARs) and brain development

Human Accelerated Regions (HARs) are short non-coding DNA elements that are deeply conserved across vertebrates yet exhibit an unusually high rate of substitution in the human lineage [22]. This rapid evolution suggests adaptive modifications potentially linked to human-specific traits.

Comparative and functional genomic analyses have revealed that many HARs overlap with enhancer elements active during cortical neurogenesis, particularly in the developing neocortex, implicating them in the evolution of human brain size and complexity [62].

A notable example, HAR1, is transcribed into a long non-coding RNA (HAR1A) expressed in Cajal– Retzius neurons during cortical development. Structural comparisons indicate significant secondary structure divergence between the human and chimpanzee HAR1A transcripts, suggesting potential changes in RNA function that may have contributed to human cortical evolution [63,5].

Functional assays further support this view: several HARs drive distinct spatiotemporal enhancer activity in human neural progenitor cells compared to other primates, indicating lineage-specific regulatory innovation [64].

Collectively, HARs represent a paradigm for understanding how subtle alterations in conserved regulatory elements can produce profound effects on gene expression, offering molecular insight into the genetic basis of human-specific cognitive and anatomical traits.

Conserved lncRNAs regulating developmental pathways

Although most long non-coding RNAs (lncRNAs) evolve rapidly, a subset exhibits remarkable conservation across vertebrates, suggesting critical regulatory functions. Comparative transcriptomic studies reveal that these conserved lncRNAs are often expressed during embryogenesis and participate in fine-tuning developmental gene networks [64, 65].

Prominent examples include Fendrr and HOTTIP, which orchestrate lineage-specific gene expression by recruiting chromatin-modifying complexes such as PRC2 and the Trithorax/MLL complex [66,67]. For broader context on lncRNA mechanisms and evolution, see 29 and 68.

Knockout studies in mice have shown that disruption of these lncRNAs leads to defects in organogenesis, underscoring their essential developmental roles. These findings highlight the dual nature of lncRNA evolution widespread turnover accompanied by pockets of deep conservation and reinforce the need for integrative comparative genomics and functional assays to uncover their biological significance (Table 3).

Table 3. A concise overviewof these case studies, summarizing regulatory elements, functions, evolutionary insights, and key references.

Case Study	Regulatory Element	Function	Evolutionary Insight / Phenotypic Outcome	Key References
ZRS Enhancer of Shh	ZRS (long-range enhancer of Shh)	Orchestrates Shh expression in posterior limb bud; specifies anterior–posterior limb axis	Deeply conserved across vertebrates; accelerated substitutions in snakes correlate with limb loss; mutations cause polydactyly	[8, 58]
Pel Enhancer of Pitx1	Pel enhancer (regulates Pitx1)	Controls Pitx1 expression during pelvic development	Independent loss-of-function deletions cause recurrent pelvic reduction in freshwater sticklebacks; reintroduction rescues phenotype	[27, 60, 61]
Human Accelerated Regions (HARs)	HAR1 (lncRNA), other HARs	Enhancer activity in developing neocortex; HAR1 transcribed in Cajal–Retzius neurons	Human-specific substitution bursts in conserved regions; likely contributed to cortical expansion and cognitive traits	[5, 22, 62, 63]
Conserved lncRNAs	Fendrr, HOTTIP, other conserved lncRNAs	Regulate chromatin-modifying complexes; fine-tune developmental gene networks	Despite rapid turnover of most lncRNAs, a conserved subset is essential for organogenesis; knockout causes developmental defects	[64, 67]

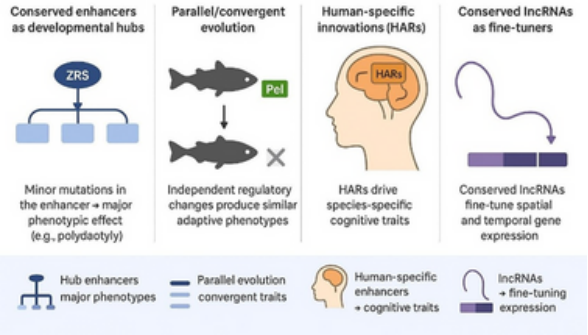

Figure 5. Principles of non-coding regulatory evolution in vertebrates.

Four major principles illustrated in figure 5: (1) conserved enhancers act as developmental hubs where small mutations can drive major phenotypic changes (e.g., ZRS and polydactyly); (2) parallel loss of regulatory elements yields convergent traits (e.g., Pel enhancer in sticklebacks); (3) human accelerated regions (HARs) near neurodevelopmental genes contribute to species-specific cognitive traits; and (4) conserved lncRNAsfine-tune gene expression. A summary ribbon highlights these unifying principles.

Synthesis of case study insights

Several unifying principles emerge from these examples.

Developmental hubs and sensitivity to change

Deeply conserved enhancers serve as critical regulatory "hubs" within developmental gene networks. Even subtle sequence alterations can produce profound morphological consequences, as exemplified by ZRS enhancer mutations that lead to polydactyly in vertebrates [68].

https://doi.org/10.61577/abcb.2025.100006

Parallel and convergent evolution

Regulatory evolution frequently proceeds in parallel across independent lineages. The repeated loss of the Pitx1 pelvic enhancer in freshwater stickleback populations illustrates how similar adaptive phenotypes can arise through convergent regulatory modifications rather than protein-coding changes [69].

Human-specific regulatory innovations

Human Accelerated Regions (HARs) exemplify the emergence of species-specific regulatory elements, particularly those associated with neurodevelopmental genes. These regions highlight how non-coding divergence contributes to cognitive and neurological traits unique to humans.

Fine-tuning by conserved lncRNAs

Conserved long non-coding RNAs (lncRNAs), such as Fendrr, modulate gene expression in subtle and context-dependent ways, integrating spatial and temporal developmental cues. Their deep conservation across vertebrates suggests a role in fine-tuning complex regulatory programs, complementing enhancer-mediated gene activation.

Collectively, these findings demonstrate that evolutionary innovation is not solely driven by changes in coding sequences but often emerges through the nuanced modulation of non-coding regulatory landscapes. Comparative genomics has been instrumental in linking such conserved and divergent regulatory elements to phenotypic diversity across vertebrates, revealing the molecular basis of morphological and functional evolution [70].

Limitations of current approaches

Despite major advances, several methodological and conceptual challenges continue to constrain the interpretation of regulatory evolution.

Conservation bias

Many comparative genomics pipelines emphasize deep conservation, which can obscure lineage-specific regulatory elements that play critical roles in adaptation and diversification.

Incomplete functional validation

Only a small subset of predicted enhancers and lncRNAs have undergone experimental validation. Epigenomic signatures such as histone modifications or chromatin accessibility provide valuable context but do not consistently confirm regulatory activity across conditions or species.

Regulatory redundancy

The existence of shadow enhancers and overlapping regulatory modules can buffer phenotypic effects, making single-element perturbations appear neutral and masking their full biological significance.

Variability in annotation frameworks

Differences in enhancer definitions, computational detection algorithms, and reference genome quality contribute to inconsistencies across datasets, complicating cross-species and meta-analytical comparisons.

Addressing these limitations requires integrating experimental and computational strategies that move beyond conservation-based inference and toward dynamic, functional characterization.

Integration with developmental biology

Achieving a systems-level understanding of regulatory evolution demands close integration with developmental biology. Temporal and spatial epigenomic profiling during critical morphogenetic windows can pinpoint active regulatory elements with greater resolution. Reconstructing gene regulatory networks (GRNs) situates enhancers and lncRNAs within broader developmental cascades, clarifying their mechanistic roles in phenotype formation.

Conclusions

Achieving a systems-level understanding of regulatory evolution demands close integration with developmental biology. Temporal and spatial epigenomic profiling during critical morphogenetic windows can pinpoint active regulatory elements with greater resolution. Reconstructing gene regulatory networks (GRNs) situates enhancers and lncRNAs within broader developmental cascades, clarifying their mechanistic roles in phenotype formation.

Comparative genomics, when coupled with experimental developmental biology, provides a robust framework for establishing causal links between regulatory sequence variation and phenotypic outcomes [11, 13, 64]. Cross-species transgenic reporter assays and CRISPR-based perturbations remain gold standards for testing enhancer function and causality. Furthermore, multi-omics approaches integrating comparative sequence analysis, chromatin conformation capture, transcriptomics, and 3D genome architecture promise to refine predictive models of regulatory evolution.

Ultimately, these integrative frameworks will deepen our understanding of how non-coding DNA orchestrates vertebrate development, drives morphological innovation, and underpins the emergence of species-specific traits.

Disclosure Statement

No potential conflict of interest was reported by the author.

References

1. Adam W, Martin G, Phil S, Tanya V, Phil N, Heather C, et al. Highly Conserved Non-Coding Sequences Are Associated with Vertebrate Development. https://doi.org/10.1371/journal.pbio.0030007
2. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012;337(6099):1190-1195. https://doi.org/10.1126/science.1222794
3. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57. https://doi.org/10.1038/nature11247
4. Vavouri T, Walter K, Gilks WR, Lehner B, Elgar G. Parallel evolution of conserved non-coding elements that target a common set of developmental regulatory genes from worms to humans. Genome Biol. 2007;8(2):R15. https://doi.org/10.1186/gb-2007-8-2-r15
5. Pollard KS, Salama SR, Lambert N, Lambot MA, Coppens S, Pedersen JS, et al. An RNA gene expressed during cortical development evolved rapidly in humans. Nature. 2006;443(7108):167-172. https://doi.org/10.1038/nature05113

https://doi.org/10.61577/abcb.2025.100006

6. Prabhakar S, Visel A, Akiyama JA, Shoukry M, Lewis KD, Holt A, et al. Human-specific gain of function in a developmental enhancer. Science. 2008;321(5894):1346-1350. https://doi.org/10.1126/science.1159974
7. Lettice LA, Heaney SJ, Purdie LA, Li L, De Beer P, Oostra BA, 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. https://doi.org/10.1093/hmg/ddg180
8. Kvon EZ, Kamneva OK, Melo US, Barozzi I, Osterwalder M, Mannion BJ, et al. Progressive loss of function in a limb enhancer during snake evolution. Cell. 2016;167(3):633-642. https://doi.org/10.1016/j.cell.2016.09.028
9. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Kheradpour P, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518(7539):317. https://doi.org/10.1038/nature14248
10. Kellis M, Wold B, Snyder MP, Bernstein BE, Kundaje A, Marinov GK, et al. Defining functional DNA elements in the human genome. Proc Natl Acad Sci. 2014;111(17):6131-6138. https://doi.org/10.1073/pnas.1318948111
11. Hardison RC. Conserved noncoding sequences are reliable guides to regulatory elements. Trends in Genetics. 2000;16(9):369-372. https://doi.org/10.1016/S0168-9525(00)02081-3
12. Lowe CB, Kellis M, Siepel A, Raney BJ, Clamp M, Salama SR, et al. Three periods of regulatory innovation during vertebrate evolution. science. 2011;333(6045):1019-1024. https://doi.org/10.1126/science.1202702
13. Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet. 2014;15(4):272-286. https://doi.org/10.1038/nrg3682
14. Bulger M, Groudine M. Functional and mechanistic diversity of distal transcription enhancers. Cell. 2011;144(3):327-339. https://doi.org/10.1016/j.cell.2011.01.024
15. Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, et al. In vivo enhancer analysis of human conserved non-coding sequences. Nature. 2006;444(7118):499-502. https://doi.org/10.1038/nature05295
16. Spitz F, Furlong EE. Transcription factors: from enhancer binding to developmental control. Nat Rev Genet. 2012;13(9):613-626. https://doi.org/10.1038/nrg3207
17. Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, et al. Enhancer evolution across 20 mammalian species. Cell. 2015;160(3):554-566. https://doi.org/10.1016/j.cell.2015.01.006
18. Visel A, Minovitsky S, Dubchak I, Pennacchio LA. VISTA Enhancer Browser—a database of tissue-specific human enhancers. Nucleic Acids Res. 2007;35(suppl_1):D88-92. https://doi.org/10.1093/nar/gkl822
19. ENCODE Project Consortium. Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature. 2020;583(7818):699-710. https://doi.org/10.1038/s41586-020-2493-4
20. Long HK, Prescott SL, Wysocka J. Ever-changing landscapes: transcriptional enhancers in development and evolution. Cell. 2016;167(5):1170-1187. https://doi.org/10.1016/j.cell.2016.09.018
21. Capra JA, Erwin GD, McKinsey G, Rubenstein JL, Pollard KS. Many human accelerated regions are developmental enhancers. Philos Trans R Soc B Biol Sci. 2013;368(1632):20130025. https://doi.org/10.1098/rstb.2013.0025
22. Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS. Ultraconserved elements in the human genome. Science. 2004;304(5675):1321-1325. https://doi.org/10.1126/science.1098119
23. Sandelin A, Bailey P, Bruce S, Engström PG, Klos JM, Wasserman WW, et al. Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. BMC genomics. 2004;5(1):99. https://doi.org/10.1186/1471-2164-5-99
24. De La Calle-Mustienes E, Feijóo CG, Manzanares M, Tena JJ, Rodríguez-Seguel E, Letizia A, et al. A functional survey of the enhancer activity of conserved non-coding sequences from vertebrate Iroquois cluster gene deserts. Genome Res. 2005;15(8):1061-1072. https://doi.org/10.1101/gr.4004805
25. Visel A, Prabhakar S, Akiyama JA, Shoukry M, Lewis KD, Holt A, et al. Ultraconservation identifies a small subset of extremely constrained developmental enhancers. Nat Genet. 2008;40(2):158-160. https://doi.org/10.1038/ng.2007.55
26. Chan YF, Marks ME, Jones FC, Villarreal Jr G, Shapiro MD, Brady SD, et al. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. science. 2010;327(5963):302-305. https://doi.org/10.1126/science.1182213
27. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. Annu Rev Biochem. 2012;81(1):145-166. https://doi.org/10.1146/annurev-biochem-051410092902
28. Statello L, Guo CJ, Chen LL, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol. 2021;22(2):96-118. https://doi.org/10.1038/s41580-021-00330-4
29. Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, et al. The evolution of lncRNA repertoires and expression patterns in tetrapods. Nature. 2014;505(7485):635-640. https://doi.org/10.1038/nature12943
30. Sarropoulos I, Marin R, Cardoso-Moreira M, Kaessmann H. Developmental dynamics of lncRNAs across mammalian organs and species. Nature. 2019;571(7766):510-514. https://doi.org/10.1038/s41586-019-1341-x
31. Marques AC, Ponting CP. Intergenic lncRNAs and the evolution of gene expression. Curr Opin Genet Dev. 2014;27:48-53. https://doi.org/10.1016/j.gde.2014.03.009
32. Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature. 2011;477(7364):295-300. https://doi.org/10.1038/nature10398
33. Luo S, Lu JY, Liu L, Yin Y, Chen C, Han X, et al. Divergent lncRNAs regulate gene expression and lineage differentiation in pluripotent cells. Cell stem cell. 2016;18(5):637-652. https://doi.org/10.1016/j.stem.2016.01.024
34. Ogbourne S, Antalis TM. Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. Biochem J. 1998;331(1):1-4. https://doi.org/10.1042/bj3310001
35. Phillips-Cremins JE, Corces VG. Chromatin insulators: linking genome organization to cellular function. Mol Cell. 2013;50(4):461-474. https://doi.org/10.1016/j.molcel.2013.04.018
36. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014;159(7):1665-1680. https://doi.org/10.1016/j.cell.2014.11.021

https://doi.org/10.61577/abcb.2025.100006

37. Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, Uliano-Silva M, et al. Towards complete and error-free genome assemblies of all vertebrate species. Nature. 2021;592(7856):737-746. https://doi.org/10.1038/s41586-021-03451-0
38. Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AF, Roskin KM, et al. Aligning multiple genomic sequences with the threaded blockset aligner. Genome Rese. 2004;14(4):708-715. https://doi.org/10.1101/gr.1933104
39. Paten B, Earl D, Nguyen N, Diekhans M, Zerbino D, Haussler D. Cactus: Algorithms for genome multiple sequence alignment. Genome Res. 2011;21(9):1512-1528. https://doi.org/10.1101/gr.123356.111
40. Armstrong J, Hickey G, Diekhans M, Fiddes IT, Novak AM, Deran A, et al. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. Nature. 2020;587(7833):246-251. https://doi.org/10.5281/zenodo.3873410
41. Harris RS. Improved pairwise alignment of genomic DNA. The Pennsylvania State University; 2007. Available at: https://www.bx.psu.edu/~rsharris/rsharris_phd_thesis_2007.pdf
42. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. 2005 15(8):1034-1050. https://doi.org/10.1101/gr.3715005
43. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Comput Biol. 2010;6(12):e1001025. https://doi.org/10.1371/journal.pcbi.1001025
44. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. 2010;20(1):110-121. https://doi.org/10.1101/gr.097857.109
45. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet. 2007;39(3):311-318. https://doi.org/10.1038/ng1966
46. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci. 2010;107(50):21931-21936. https://doi.org/10.1073/pnas.1016071107
47. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. The accessible chromatin landscape of the human genome. Nature. 2012;489(7414):75-82. https://doi.org/10.1038/nature11232
48. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013;10(12):1213-1218. https://doi.org/10.1038/nmeth.2688
49. Banerji J, Rusconi S, Schaffner W. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell. 1981;27(2):299-308. https://doi.org/10.1016/0092-8674(81)90413-X
50. Melnikov A, Murugan A, Zhang X, Tesileanu T, Wang L, Rogov P, et al. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. Nat Biotechnol. 2012;30(3):271-277. https://doi.org/10.1038/nbt.2137
51. Inoue F, Ahituv N. Decoding enhancers using massively parallel reporter assays. Genomics. 2015 106(3):159-164. https://doi.org/10.1016/j.ygeno.2015.06.005
52. Fulco CP, Munschauer M, Anyoha R, Munson G, Grossman SR, Perez EM, et al. Systematic mapping of functional enhancer–promoter connections with CRISPR interference. Science. 2016;354(6313):769-773. https://doi.org/10.1126/science.aag2445
53. Gasperini M, Hill AJ, McFaline-Figueroa JL, Martin B, Kim S, Zhang MD, et al. A genome-wide framework for mapping gene regulation via cellular genetic screens. Cell. 2019;176(1):377-390. https://doi.org/10.1016/j.cell.2018.11.029
54. Dimitrieva S, Bucher P. UCNEbase—a database of ultraconserved non-coding elements and genomic regulatory blocks. Nucleic Acids Res. 2013;41(D1):D101-109. https://doi.org/10.1093/nar/gks1092
55. Herrero J, Muffato M, Beal K, Fitzgerald S, Gordon L, Pignatelli M, et al. Ensembl comparative genomics resources. Database. 2016;2016:bav096. https://doi.org/10.1093/database/bav096
56. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. Nature. 2014;507(7493):455-461. https://doi.org/10.1038/nature12787
57. Wang B, Diao Y, Liu Q, An H, Ma R, Jiang G, et al. An increased duplication of ZRS region that caused more than one supernumerary digits preaxial polydactyly in a large Chinese family. Sci Rep. 2016;6(1):38500. https://doi.org/10.1038/srep38500
58. Xie KT, Wang G, Thompson AC, Wucherpfennig JI, Reimchen TE, MacColl AD, et al. DNA fragility in the parallel evolution of pelvic reduction in stickleback fish. Science. 2019;363(6422):81-84. https://doi.org/10.1126/science.aan1425
59. Levchenko A, Kanapin A, Samsonova A, Gainetdinov RR. Human accelerated regions and other human-specific sequence variations in the context of evolution and their relevance for brain development. Genome Biol Evol. 2018;10(1):166-188. https://doi.org/10.1093/gbe/evx240
60. Reilly SK, Yin J, Ayoub AE, Emera D, Leng J, Cotney J, et al. Evolutionary changes in promoter and enhancer activity during human corticogenesis. Science. 2015;347(6226):1155-1159. https://doi.org/10.1126/science.1260943
61. Johnsson P, Lipovich L, Grandér D, Morris KV. Evolutionary conservation of long non-coding RNAs; sequence, structure, function. BBA Gen Subj. 2014;1840(3):1063-1071. https://doi.org/10.1016/j.bbagen.2013.10.035
62. Grote P, Wittler L, Hendrix D, Koch F, Währisch S, Beisaw A, et al. The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. Developmental cell. 2013;24(2):206-214. https://doi.org/10.1016/j.devcel.2012.12.012
63. Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature. 2011;472(7341):120-124. https://doi.org/10.1038/nature09819
64. Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. Cell. 2013;154(1):26-46. https://doi.org/10.1016/j.cell.2013.06.020

https://doi.org/10.61577/abcb.2025.100006

65. Carroll SB. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell. 2008;134(1):25-36. https://doi.org/10.1016/j.cell.2008.06.030
66. Rebeiz M, Tsiantis M. Enhancer evolution and the origins of morphological novelty. Curr Opin Genet Dev. 2017;45:115-123. https://doi.org/10.1016/j.gde.2017.04.006
67. Spielmann M, Mundlos S. Structural variations, the regulatory landscape of the genome and their alteration in human disease. Bioessays. 2013;35(6):533-543. https://doi.org/10.1002/bies.201200178
68. Zhang F, Lupski JR. Non-coding genetic variants in human disease. Hum Mol Genet. 2015;24(R1):R102-110. https://doi.org/10.1093/hmg/ddv259
69. Fulco CP, Nasser J, Jones TR, Munson G, Bergman DT, Subramanian V, et al. Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. Nat Genet. 2019;51(12):1664-1669. https://doi.org/10.1038/s41588-019-0538-0
70. Kempfer R, Pombo A. Methods for mapping 3D chromosome architecture. Nat Rev Genet. 2020;21(4):207-226. https://doi.org/10.1038/s41576-019-0195-2

https://doi.org/10.61577/abcb.2025.100006