



HAL
open science

Comparative Genomics and Epigenomics of Transcriptional Regulation

Huaijun Zhou, Emily L. Clark, Dailu Guan, Sandrine Lagarrigue, Lingzhao Fang, Hao Cheng, Christopher K. Tuggle, Muskan Kapoor, Ying Wang, Elisabetta Giuffra, et al.

► **To cite this version:**

Huaijun Zhou, Emily L. Clark, Dailu Guan, Sandrine Lagarrigue, Lingzhao Fang, et al.. Comparative Genomics and Epigenomics of Transcriptional Regulation. Annual Review of Animal Biosciences, 2024, 14 (1), 10.1146/annurev-animal-111523- . hal-04802793

HAL Id: hal-04802793

<https://hal.science/hal-04802793v1>

Submitted on 3 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

Annual Review of Animal Biosciences

Comparative Genomics and Epigenomics of Transcriptional Regulation

Huaijun Zhou,¹ Emily Clark,² Dailu Guan,¹ Sandrine Lagarrigue,³ Lingzhao Fang,⁴ Hao Cheng,¹ Christopher K. Tuggle,⁵ Muskan Kapoor,⁵ Ying Wang,¹ Elisabetta Giuffra,⁶ and Giorgia Egidy⁶

¹Department of Animal Science, University of California, Davis, California, USA; email: hzhou@ucdavis.edu, dguan@ucdavis.edu, qtlcheng@ucdavis.edu, ucywang@ucdavis.edu

²The Roslin Institute, University of Edinburgh, Edinburgh, Midlothian, United Kingdom; email: Emily.Clark@roslin.ed.ac.uk

³PEGASE, INRAE, Institut Agro, Saint Gilles, France; email: sandrine.lagarrigue@institut-agro.fr

⁴Center for Quantitative Genetics and Genomics, Aarhus University, Aarhus, Denmark; email: lingzhao.fang@qgg.au.dk

⁵Department of Animal Science, Iowa State University, Iowa, USA; email: cktuggle@iastate.edu, muskan@iastate.edu

⁶GABI, AgroParisTech, INRAE, Jouy-en-Josas, France; email: elisabetta.giuffra@inrae.fr, giorgia.egidy-maskos@inrae.fr

Annu. Rev. Anim. Biosci. 2025. 13:23.1–23.26

The *Annual Review of Animal Biosciences* is online at animal.annualreviews.org

<https://doi.org/10.1146/annurev-animal-111523-102217>

Copyright © 2025 by the author(s).
All rights reserved

Keywords

transcriptional regulation, noncoding variants, regulatory elements, quantitative trait loci, molecule QTL, comparative genomics and epigenomics, deep learning modeling

Abstract

Transcriptional regulation in response to diverse physiological cues involves complicated biological processes. Recent initiatives that leverage whole genome sequencing and annotation of regulatory elements significantly contribute to our understanding of transcriptional gene regulation. Advances in the data sets available for comparative genomics and epigenomics can identify evolutionarily constrained regulatory variants and shed light on noncoding elements that influence transcription in different tissues and developmental stages across species. Most epigenomic data, however, are generated from healthy subjects at specific developmental stages. To bridge

the genotype–phenotype gap, future research should focus on generating multidimensional epigenomic data under diverse physiological conditions. Farm animal species offer advantages in terms of feasibility, cost, and experimental design for such integrative analyses in comparison to humans. Deep learning modeling and cutting-edge technologies in sequencing and functional screening and validation also provide great promise for better understanding transcriptional regulation in this dynamic field.

1. INTRODUCTION

Regulating gene expression at the transcriptional level involves intricate processes, including chromatin accessibility, transcription initiation, transcript elongation, and splicing (**Figure 1**). Gene expression variation contributes significantly to the phenotypic diversity of complex traits (1, 2). Notably, genome-wide association studies (GWAS) on complex traits have revealed that approximately 90% of GWAS hits are located in noncoding regions (3, 4). These trait-associated variants often are enriched in active regulatory elements (REs) of genes in biologically relevant tissues or cell types (3–6). These findings suggest a substantial role for noncoding regulatory variants in governing gene expression and affecting phenotypic variation. Consequently, identifying causative regulatory variants and understanding their functional roles in transcriptional regulation are essential steps in linking genotypes to phenotypes. Although variants in coding regions also influence gene expression, this review focuses primarily on noncoding regulatory sequences.

Noncoding REs, such as promoters and enhancers, play vital roles in transcriptional regulation (7). Variations in these regions, particularly at transcription factor binding sites (TFBSs), that have evolved under selection constraints contribute to differences in gene expression and, consequently, to phenotypic variations both within and between species (7, 8). Additionally, DNA sequences affecting chromatin accessibility dynamics can regulate gene expression (9). Variations in splice sites (10), polyadenylation signals (11), and 3' untranslated regions can also

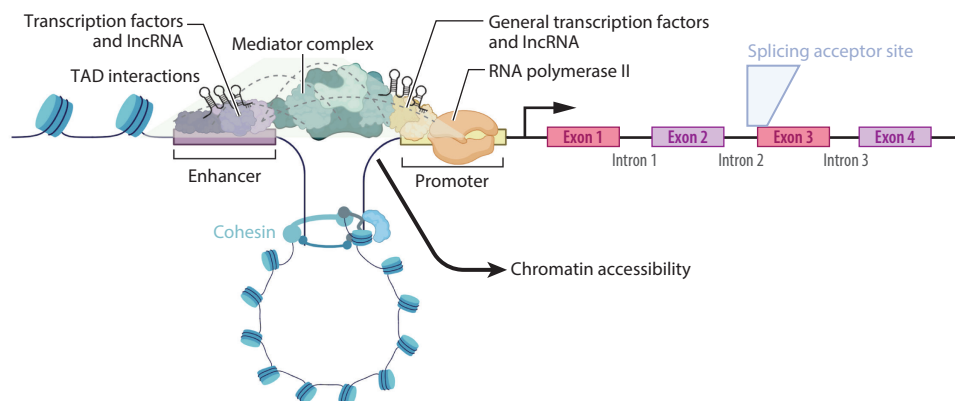


Figure 1

The schematic mechanisms of transcriptional regulation involve intricate processes, including chromatin accessibility, transcription initiation, transcript elongation, and splicing, as well as the involvement of the cohesin complex. Abbreviations: lncRNA, long noncoding RNA; TAD, topologically associated domain. Figure adapted with permission from artwork originally created in BioRender.

contribute to gene regulation (12). Mutations in regulatory genomic regions, driven by natural selection (humans and wild species) and artificial selection (such as selective breeding in farm animals), provide opportunities to investigate the underlying mechanisms driving transcriptional regulation mechanisms both within and between species.

The Zoonomia Consortium and Vertebrate Genomes Project have generated whole genome sequence data across hundreds of species, including mammalian and avian species (13, 14). In addition, completion of the recent telomere-to-telomere assembly of the human genome sequence (15) has triggered a paradigm shift, redirecting efforts toward investigating epigenetic and transcriptional gene regulation mediated by functional genomic elements. Recent progress in the functional annotation of REs in humans (16), mice (17), and farm animal species (5, 6, 18) has generated essential information to identify regulatory regions of the genome across multiple tissues, cell types, and developmental stages. Comparative genomics allows the investigation of potential functional constraints in regulatory DNA sequences by comparing orthologous sequences across different species (19). However, DNA-level sequence conservation does not always reflect functional conservation of orthologous regulatory sequences, especially in enhancers (18). For instance, a brain enhancer in mice, highly conserved at the DNA level across species, exhibited distinct enhancer activities in different species (20). Conversely, conserved enhancer sequences are not always required for functional conservation across species (21), because enhancer sequences generally experience less constraint compared to promoters and genic regions (22). Therefore, relying solely on sequence information at the DNA level is insufficient to fully understand transcriptional regulation, given its context-dependent nature. To bridge this knowledge gap, functional insights in chromatin accessibility, histone modification, and splicing across various contexts, such as developmental stage, tissue, and sex, are imperative. Comparative epigenomic analysis, leveraging multidimensional functional data, is a robust approach to elucidate the molecular and evolutionary mechanisms underlying gene regulation.

This review centers on recent advances in understanding transcriptional regulation through comparative genomics and epigenomics, with a major focus on noncoding variants. Initially, we describe studies identifying functional variants associated with gene expression through sequence conservation analysis. We then delve into the molecular phenotypes of regulatory variation, encompassing chromatin accessibility, histone modification, DNA methylation, expression quantitative trait loci (eQTL), splicing, and their contributions to transcriptional regulation differences within and between species. In addition, we explore the potential opportunities that a machine learning approach could offer for studying regulatory variation. We conclude by briefly discussing future perspectives on the impact of new technologies, including single-cell (SC) sequencing, long-read sequencing, massive parallel reporter assays (MPRAs), and clustered regularly interspaced short palindromic repeats (CRISPR) technologies, on understanding transcriptional regulation. **Figure 2** outlines the main approaches used to understand the underlying mechanisms of transcriptional regulation via comparative genomics and epigenomics.

2. COMPARATIVE GENOMICS RELATED TO TRANSCRIPTION FACTORS AND LONG NONCODING RNAs ON TRANSCRIPTIONAL REGULATION

2.1. Harnessing the Power of Comparative Analysis to Understand Regulatory Elements at the DNA Level

Comparative analysis of noncoding elements at the DNA level provides information to identify the genomic drivers of the fundamental differences between species and can reveal the functional basis of species-specific phenotypes (23, 24). By understanding the noncoding DNA elements driving



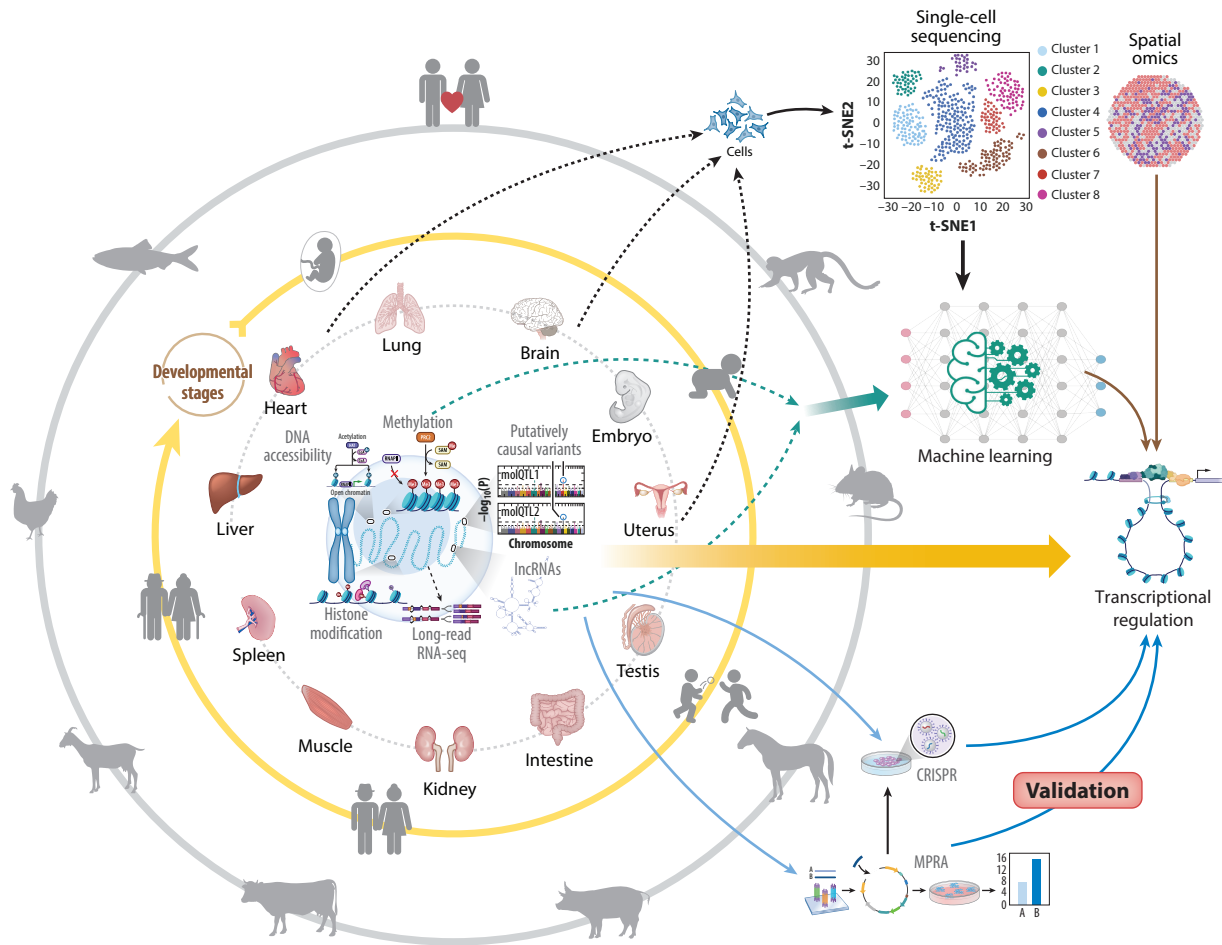


Figure 2

Key strategies for understanding transcriptional regulation through comparative genomics and epigenomics, with a focus on noncoding regulatory sequences. Orthologous regulatory sequences across species and epigenomic data in different tissues/cells across different developmental stages among species can be leveraged and integrated using deep learning modeling. Coupled with state-of-the-art technologies in sequencing and functional screening and validation, the underlying mechanisms of transcriptional regulation can be elucidated further. Abbreviations: CRISPR, clustered regularly interspaced short palindromic repeats; lncRNA, long noncoding RNA; MPRA, massive parallel reporter assay; QTL, quantitative trait loci; RNA-seq, RNA sequencing. Figure adapted with permission from artwork originally created in BioRender.

gene regulation that are unique to each species, we can infer their role or function in shaping species-specific phenotypes, e.g., divergent patterns in disease resistance among salmonids (25).

Nearly two decades ago, comparative analysis was used to establish the functional relevance of putative noncoding DNA elements in the genomes of humans and mice (e.g., 26). Since this seminal work, a vast quantity of additional genomic information has become available for other mammalian species. Kuderna et al. (27) recently constructed a whole-genome alignment of 239 primate species, representing almost half of all extant species in the primate order. Their analysis revealed hundreds of thousands of TFBSs constrained specifically in primates, distinguishing them from other placental mammals. Using MPRA, they successfully validated the *cis*-regulatory effects of these constrained TFBSs on gene expression (27). The findings highlight the significant role of

recent evolution in noncoding DNA elements influencing biological processes and contributing to the differentiation of primates, including humans, from other placental mammals (27). Similarly, the Zoonomia Consortium employed a comparable approach to generate a reference-free alignment across 241 mammalian genomes (28). Their analysis revealed that genes proximal to constrained elements are involved in fundamental cellular processes, whereas genes neighboring primate-specific elements play roles in environmental interaction, such as odor perception and immune response (28).

2.2. Understanding the Effects of Transcription Factor Binding on Transcriptional Regulation

The regulation of gene transcription hinges on the binding affinity of transcription factors (TFs) to specific DNA sequences known as TFBSs. Despite the high conservation of TF genes between humans and mice, only 22% of TFBSs exhibit conservation (29). Within the mammalian clade, only one-third of highly conserved TFBSs identified by UMAP (Uniform Manifold Approximation and Projection) for Dimension Reduction are evolutionarily constrained using a Gaussian mixture model, and they are found primarily within highly conserved REs (28). These constrained human TFBSs are bound by TFs in other mammals and display epigenetic signals suggestive of regulatory functions (28). A significant portion of conserved TFBSs are located within or near REs with tissue-specific functions. Notably, tissue-specific enhancers are enriched for specific TFs and motifs (28, 30). This pattern becomes more discernible in SC studies due to the homogeneity of epigenomic RE annotations (31, 32). Despite varying information content of motifs, cross-regulatory networks between TFs remain conserved between human and mouse, potentially facilitated by novel TF binding sites (29). Furthermore, different types of REs harbor distinct TFBSs (30), and the orientation and order of TFBS have a significant effect on gene regulatory activity (33). Interestingly, species-specific TFBSs often overlap with transposable elements, contributing to the emergence of lineage-specific regulatory regions that maintain functional conservation despite limited phylogenetic conservation (28, 29, 34).

2.3. Long Noncoding RNA

Long noncoding RNAs (lncRNAs) represent a sizable and heterogeneous class of genes defined by transcripts longer than 200 nucleotides without coding-potential capabilities (35). Although the exact number of lncRNAs per species remains a topic of debate, there are believed to be several thousand lncRNAs in different genomes (36). Despite this, the functions of most lncRNAs remain elusive, and they represent a diverse array of REs implicated in gene expression through diverse mechanisms involving DNA, RNA, or protein interactions (37). These mechanisms encompass nuclear organization, genome integrity, histone modifications, DNA loop formation connecting enhancers to promoters, RNA splicing, messenger RNA stabilization or degradation, and modulation of microRNA or protein activity. Due to their abundance and multifaceted roles in gene regulation, lncRNAs play roles in various biological and pathophysiological processes (37). Compared to protein-coding genes (PCGs), lncRNAs typically exhibit lower expression levels and a greater specificity to tissues, developmental stages, and conditions (38, 39). Consequently, the identification of these noncoding RNAs relies on RNA sequencing (RNA-seq) data from diverse resource samples.

The number of lncRNAs identified in animals is expanding continuously based on the genome assembly version, bioinformatic annotation pipelines, and RNA-seq data resources employed. For instance, as of 2024, the reference European Molecular Biology Laboratory European Bioinformatics Institute's (EMBL EBI's) Ensembl annotation database (v111) reports 10,972, 8,512, and



11,946 lncRNAs for pig, cow, and chicken, respectively, in contrast to 19,378 for humans. Furthermore, disparities exist between the two genome annotation databases, EMBL EBI's Ensembl and the US National Center for Biotechnology Information's RefSeq (e.g., 11,946 and 5,791 lncRNA genes, respectively, for chicken), and the gene loci display minimal overlap between both databases (40). Due to the complementary nature of these two databases in providing gene models, their combination enhances the lncRNA atlas, as Degalez et al. (39, 41) demonstrated for chicken. Of note, a reference annotation can be supplemented with lncRNA gene models from an independent RNA-seq data set using a recently developed pipeline called TAGADA (42). In the near future, lncRNA annotation is expected to benefit from new technologies, such as long-read RNA-seq, to enhance the accuracy of gene models.

Over the last decade, numerous studies have employed comparative genomic approaches to annotate novel lncRNAs across phylogenetically diverse species (40, 43). However, lncRNA sequences exhibit faster evolution compared to most PCGs, limiting their conservation across species even when synteny is evident; approximately 70% of lncRNAs lack sequence orthologs in species that have diverged for more than 50 million years (44). Functional elements of lncRNAs often consist of short sequences spanning a few nucleotides across the RNA sequence, enabling interactions with small RNAs or proteins (45, 46). Consequently, no lncRNA orthologs are reported in reference databases across different species.

In this context, two main approaches have been employed for comparative genomic analyses of lncRNAs. The first involves analyzing their positional conservation (synteny) within the genome of each species by identifying positionally conserved neighboring PCGs. Some studies have identified positionally conserved lncRNAs within more and less phylogenetically distant species (43, 44), including farm animal species (40, 47, 48). Importantly, comparative transcriptomic approaches can help refine orthologous status, given that lncRNAs exhibit greater tissue specificity than PCGs (49). The second approach combines synteny with the search for conserved short motifs between species, as demonstrated by recent computational tools. For instance, the algorithmic framework LncLOOM efficiently compares dozens of lncRNA sequences from different species and predicts positionally conserved short motifs (50). The computational pipeline lncHOME identifies, at the genome scale, lncRNAs with conserved genomic positions and short motifs of RNA-binding proteins, defined initially from experimental studies and databases (51). Another workflow combines synteny methods and analysis of short motifs using the Mercator–Pecan genome alignment method (52). Such approaches have been applied to a range of species (50–52), including farm animals (52), providing insights into the evolutionary conservation of lncRNAs and their functional elements. The tens of thousands of genes encoding lncRNAs, or at least a part of them, combined to the genes encoding TFs and microRNAs add complexity to the network of *trans*-type REs contained in genomes. More efforts are needed to explore their potential functions on transcriptional control.

3. COMPARATIVE EPIGENOMICS OF TRANSCRIPTIONAL REGULATION

Unlike DNA sequence, which remains consistent across tissues/organs within an individual (if there is no somatic mutation) due to the DNA replication mechanism, the functional regulation of gene expression by genomic elements exhibits significant temporal and spatial variation. Over the past almost two decades, consortia have been initiated to systematically profile multi-tissue epigenomic variations across species. Examples include the Encyclopedia of DNA Elements (ENCODE) (16, 17, 23, 29, 30) and the National Institutes of Health's Roadmap Epigenomics Program (53) for mouse and human, the Functional Annotation of Animal Genomes (FAANG)

for farm animals (5, 6, 18), modENCODE for model organisms including *Drosophila melanogaster* and *Caenorhabditis elegans* (54), and DANIO-CODE for zebrafish (55). Alongside community efforts such as the International Human Epigenome Consortium (56), as well as contributions from individual laboratories, a comprehensive curated multi-tissue atlas of functional annotations for humans has been created. This atlas provides an almost complete map of REs (e.g., promoters, enhancers, repressors, insulators, silencers) across various tissues, developmental stages, and species. Such efforts enable comparative epigenomic studies, significantly advancing our understanding of the functional and evolutionary roles of epigenomic REs in governing gene expression.

3.1. Identification and Characterization of Epigenomic Regulatory Elements

Epigenomic REs can be characterized through combinatorial histone modification profiles, such as H3K4me3 for promoters, H3K4me1 for enhancers, H3K36me3 for transcription, H3K27me3 for Polycomb repression, H3K9me3 for heterochromatin regions, and H3K27ac for active enhancers or promoters using chromatin immunoprecipitation sequencing (ChIP-seq) (17, 30, 57). Chromatin accessibility can be assessed through techniques like deoxyribonuclease sequencing (DNase-seq) and transposase-accessible chromatin using sequencing (ATAC-seq) (58, 59). Silencer elements can be identified by screening for genomic regions that repress caspase 9 transcription at a genome-wide level (60).

DNA chromatin accessibility, represented by open chromatin regions (OCRs), serves as a key indicator of regulatory DNA influencing gene expression, encompassing various epigenomic elements such as enhancers, promoters, and silencers (58, 61). The ENCODE project identified approximately three million high-confidence OCRs (58), demonstrating cell-specific function and evolutionary constraint (27, 61). These OCRs were enriched with human genetic variants impacting gene expression and complex traits (27, 62), underscoring their evolutionary contributions to speciation (27) and cross-species prediction of REs (63). Kuderna et al. (27), for instance, found that 42% of OCRs were under sequence constraint in species diverged more than 100 million years ago, with 11% specifically constrained in primates based on the alignment of 239 primate genomes. The primate-specific constrained OCRs exhibited higher and more consistent chromatin accessibility across all primate species, suggesting a correlation between sequence-level and functional constraint across species (27). Moreover, constrained OCRs might undergo purifying selection in human populations, indicating that variations in these REs likely have significant regulatory roles in transcriptional regulation (27).

3.2. Effects of Genomic and Epigenomic Regulatory Elements on Transcriptional Regulation

The key challenge of comparative epigenomics lies in the disparity between sequence alignability (i.e., sequence conservation) and biochemical signal variability (i.e., functional conservation) of REs, shaping the conserved and divergent patterns of gene transcription. Sequence analysis reveals that approximately 80% of REs exhibit at least 10% alignability between mouse and human, yet 15% of mouse REs lack human orthologs (29). Comparison of sequence conservation across species revealed a more pronounced divergence, with promoters being more evolutionarily constrained than enhancers (18, 22, 29).

The Zoonomia Consortium's (64) recent large-scale alignment across 241 mammalian genomes revealed that ~50% of the 0.92 million human epigenomic REs are under evolutionary constraint at the sequence level (28). Promoter-like sequences were found to be more conserved than enhancer sequences (28). Despite significant sequence conservation among REs across species, the functional conservation varies considerably; fewer than half of sequence-conserved



REs in mice retain their regulatory roles in humans (17, 29). Specifically, promoter activities exhibit notable consistency across species, reflecting tight evolutionary constraints, whereas enhancer regions, due to their varied regulatory roles (e.g., a high tissue specificity), evolve more rapidly (18, 22, 29).

Further analysis has revealed that functional conservation may be linked to some extent with sequence conservation. For instance, strong functional conservation is observed in both rapidly and slowly evolving DNA sequences but not in neutrally evolving sequences (65). However, evolutionary changes in the epigenome show a linear correlation with the transcriptome, suggesting the potential to predict regulatory sequences by conserved sequence colocalization of epigenomic marks (65). Epigenetically conserved REs are more likely to activate gene expression, whereas species-specific REs are more likely to contribute to divergent gene expression (66).

The latest phase III of the ENCODE project reported a comprehensive map of REs and confirmed the discrepancy between functional and sequence conservation. Whereas 56% of human and 72% of mouse epigenomic REs share orthologous sequences, functional conservation is significantly less common, with 46% for mouse REs compared to only 18% for human REs (17). Addressing inadequate annotations, the heterogeneity of bulk-tissue annotations, or the inherent complexity in correlating sequence conservation with functional outcomes across species might mitigate this discrepancy. Furthermore, ongoing technology advancement will help reduce the risk of false-positive and false-negative results caused by technical challenges or reproducibility issues.

3.3. The Effects of DNA Methylation on Transcriptional Regulation

DNA methylation typically is associated with repressed regulatory regions or active gene transcripts in a tissue-specific manner. High-throughput profiling of this modification can be achieved through methods such as whole-genome bisulfite sequencing and reduced-representation bisulfite sequencing, among others (57, 67). Moreover, a comprehensive characterization and cataloging of additional epigenomic REs, including chromatin-associated proteins such as TFs, transcriptional cofactors (68), and TFBSs across various cell types and tissues (59, 69), is essential. These annotation studies revealed that the individual and combined actions of epigenomic REs orchestrate the spatiotemporal pattern of gene expression across tissues, developmental stages, or environmental contexts, deepening our understanding of the cooperation of epigenomic REs and TFBSs in regulating gene expression and complex traits.

DNA methylation is a critical epigenomic factor that governs gene transcription, intricately intertwining with DNA accessibility, chromatin states, and conformation. Its significance lies in shaping and maintaining the cellular epigenomic landscape (70–72). Typically, promoter regions exhibit low DNA methylation, whereas transcribed regions are associated with elevated gene expression levels. Conversely, enhancer regions typically display intermediate levels of DNA methylation (57, 71). Furthermore, differentially methylated regions across cell types are often depleted in heterochromatic regions (H3K9me3) and enriched in regions containing active enhancers (H3K27ac and H3K4me1) and poised enhancers (H3K4me1 and H3K27me3) and promoters (H3K4me3) (71). These variations in methylation levels are more pronounced in highly expressed compared to lowly expressed genes (57).

This intricate interplay enables the simultaneous profiling of DNA methylation with other omics layers, such as histone modifications, using advanced methods like nanoHiMe-seq (72) and chromatin conformation with snm3C-seq (70, 71). These techniques operate at the SC or single-nucleus levels, revealing significant spatial diversity in global methylation levels across various cell types (70). Comparative analysis of the human and mouse genome revealed that approximately 40–60% of differential hypomethylation sites across cell types shared orthologous sequences, with

95% of them exhibiting consistent directions in methylation changes (71). Mapping DNA methylation across 535 vertebrates and 45 invertebrates indicates that the DNA methylation level tends to be lower in invertebrates than in vertebrates, lower in birds and marsupials than in other vertebrates, and higher in fish and amphibians than in other taxonomic groups (73). However, the evolutionary constraints of the methylome vary among tissues across species, irrespective of sequence conservation (74, 75). These conserved and divergent methylation levels are significantly correlated with cross-species traits, such as longevity (74, 76), and are particularly useful in annotating noncoding variants that affect transcriptional regulation and are associated with complex traits (75).

3.4. The Effects of Chromatin Conformation on Transcriptional Regulation

The linear DNA sequence, wrapped around associated proteins to form chromatin, is organized within the nucleus in a hierarchical structure closely tied to gene expression. Chromatin is systematically structured into compartments, chromatin domains, and loops, which can be investigated using microscopy-based imaging or chromosome conformation capture coupled sequencing methods (known as C-technologies) (77, 78). At the compartment level, chromosomes can be categorized into compartment A, associated with strong messenger RNA expression activities, and compartment B, linked to inactive gene activities (79). Approximately 40–60% of the human genome shows compartment A/B switching during stem cell differentiation, indicating that compartmentalization is often specific to cell type and developmental stage (80). A similar compartment partition has also been observed in other species, such as *Drosophila* (81), suggesting a conserved compartmental structure across different organisms.

At a finer scale, topologically associating domains (TADs) have been identified, displaying remarkable conservation across cell types and species (82, 83), seemingly unaffected by tissue-specific gene activity or histone modification patterns (82, 84). Despite the genome-wide switching between compartments A and B, the positioning of TADs remains unchanged (80). Comparison between human and mouse TADs revealed that 50–70% of TAD boundaries are shared between human and mouse embryonic stem cells (82, 84), suggesting evolutionarily conserved TAD structure. Moreover, cross-species comparison of TAD boundaries indicates that most human TAD boundaries possess conserved sequences in other species, and genes associated with divergent boundaries show increased divergent expression (66).

High-resolution techniques, such as Hi-C, enable the exploration of interactions between distant genomic locations, known as chromatin loops. Notably, one end of a chromatin loop often harbors a known promoter, whereas the opposite end aligns with a known enhancer (82). Although a smaller proportion of loops have been identified as conserved across mammals and primates compared to TAD boundary elements, conserved loops are more likely to contain promoters with conserved expression patterns. These suggested conserved 3D chromatin interactions maintain the conservation of gene regulatory functions (66). Moreover, loops with shorter distances are more likely to be conserved over evolutionary time frames (66).

3.5. The Effect of Transcriptional Regulation on Phenotypic Diversity

Transcriptional regulation plays a crucial role in shaping morphological and molecular diversity during development across species. This process is marked by a reduction in gene expression and the prevalence of purifying selection as development progresses (85). Evolutionary changes in cellular programs often stem from lineage- or species-specific differences in gene expression (86), highlighting the significance of unique epigenomic patterns in evolutionary biology. A reported 0.03–3% of chromatin states exhibit significant variation between adjacent developmental stages



of the same tissue, with enhancer states showing the most pronounced changes (87). Particularly noteworthy is the pervasive Polycomb-mediated repression observed during fetal development, where the expression of genes surrounded by bivalent states is repressed in a tissue- and time-point-specific manner (87, 88). In addition, dynamic OCRs during mouse development are enriched for TFs crucial for cell fate, suggesting conservation across species and developmental stages (87, 89). Notably, these OCR orthologs are often associated with genetic variations related to human diseases, displaying tissue-specific enrichment patterns (87, 89).

The functional and evolutionary characteristics of epigenomic REs across tissues and species elucidate the conserved and divergent regulation of gene transcription, contributing to the intricate regulatory architecture of complex traits. Typically, GWAS loci are located predominantly within various histone marks, REs, OCRs, DNA methylation sites, and TFBSs (18, 28, 29, 75). The partitioning of heritability by integrating large-scale epigenomic annotations reveals a complex multifactorial network underlying complex traits, featuring regulatory relationships within a tissue–regulator–trait triangle (53). The heritability enrichment of complex traits in REs, especially enhancers and TFBSs, often exhibits strong tissue- or cell-type specificity, shaped by tissue-specific gene regulation (6, 28). Notably, genetic variants associated with complex traits are most enriched in highly conserved REs and constrained TFBSs compared to the rest of REs and TFBSs (6, 66), underscoring the potential of cross-species epigenomic maps to enhance our understanding of gene transcriptional regulation and the genetics of complex traits (6, 29, 66). For instance, the heritability of Alzheimer’s disease was significantly enriched in tissue-specific orthologous enhancers between pigs and humans compared to those between mice and humans, highlighting the pig’s potential as a biomedical model for investigating Alzheimer’s disease (6). Furthermore, cross-species conservation analysis at the level of co-occurring TFBSs indicates the ability to delineate genetic association signals to disease-related molecular mechanisms (90).

4. MOLECULAR QTL AFFECTING GENE REGULATION

A portion of interindividual variation in gene expression is heritable and often subjected to negative or positive selection. This contributes to the phenotypic diversity observed within and across species (7). Systematic characterization of genetic control of gene expression, such as through eQTL, is essential for elucidating transcriptional regulation that contributes to complex traits. This is important because most genomic variants linked to complex traits by GWAS are noncoding and influence gene expression or other molecular phenotypes (91). Integrating eQTL resources with evolutionary biology can enhance our understanding of the evolutionary mechanisms that give rise to transcriptional variation and, consequently, phenotypic diversity across species (7). The insights and knowledge generated from these analyses can facilitate the translation of genetic findings between species. This not only benefits the development of animal models for understanding human biology and diseases but also contributes to advancements in animal genetic improvement in agriculture. In this context, we summarize the current resources for molecular QTL (molQTL) across species and review evolutionary features of regulatory variants.

4.1. Summary of molQTL Resources Across Species

Because it is regarded as the sole existing method to investigate the functional impacts of genomic variants in the native genomic and biological context (92), population-based molQTL mapping, particularly eQTL, has been studied extensively in humans. The Genotype-Tissue Expression (GTEx) project, a pioneering endeavor initiated more than a decade ago, identified eQTL and splicing QTL across 49 diverse tissues in healthy adults (93). This groundbreaking initiative delved deep into the molecular complexities underpinning both transcriptional and phenotypic variations

observed in humans, catalyzing larger community efforts on molQTL mapping in diverse human populations. For instance, the eQTLGen Consortium investigated the genetic regulation of blood gene expression in 31,684 individuals worldwide (94). Both the eQTL Catalogue and QTLbase2 have uniformly processed and curated various types of human molQTL, including eQTL, splicing QTL, and TF binding QTL (95–97).

In contrast to the abundant molQTL resources available for humans, resources for other species are notably limited. Various research initiatives on molQTL mapping have been pursued in fruit flies (98), rats (99), mice (100), and nonhuman primates (101–103). Since 2018, the FarmGTEEx project has built a comprehensive catalog of regulatory variants across diverse biological and environmental contexts in farm animals, including mammals (e.g., pigs and cattle), birds (e.g., chicken and duck), fish (e.g., Atlantic salmon and rainbow trout), and insects (e.g., bees) (104). In the pilot phase, the FarmGTEEx consortium successfully completed molQTL mapping in 22, 34, and 27 different tissues in cattle (105), pigs (106), and chickens (107), respectively, by uniformly processing all publicly available RNA-seq and whole genome sequencing data. The consortium is currently extending its efforts to include other farm animal species, such as sheep, goats, horses, rabbits, ducks, Atlantic salmon, and rainbow trout. In comparison to humans, other animals, particularly farm animals, have significant advantages in the collection of pan-tissue samples and biopsies across diverse conditions, such as developmental stages, *in vivo* pathogen infections, and dietary variations (103).

As the cost of SC sequencing has decreased dramatically, more projects have begun to map eQTL at the SC level (108–111), including the recently initiated SC eQTLGen and IGVF consortia (112, 113). Given that most disease-associated noncoding variants exert no detectable regulatory effects on gene expression in these healthy adults in the GTEEx population (110, 114), several context-specific GTEEx projects have been initiated to study the dynamic effects of regulatory variants across different biological and environmental conditions. Examples include the human developmental GTEEx (115) and pathogen-response eQTL mapping in immune cell types (109, 111).

4.2. Cross-Species Comparison of molQTL

Although the DNA sequences of many PCGs are evolutionarily conserved (i.e., orthologous genes), deciphering their tolerance to genetic perturbations during evolutionary adaption and speciation, and their functional contributions to convergent and divergent phenotypes between species, remains a formidable challenge. Multispecies molQTL integrative analysis offers an unprecedented opportunity for in-depth comparisons between species at multiple molecular and physiological levels. Tung et al. (116) conducted blood eQTL mapping in baboons, revealing significant overlap with human eQTLs. They found that eQTLs regulating lineage-specific and rapidly evolving genes were more likely to be detected, suggesting a degree of evolutionary conservation in eQTLs across primate species. Fair et al. (102) demonstrated the general conservation of genetic control of gene expression between humans and chimpanzees through a comparison of heart eQTL. Similar findings were further observed by Munro et al. (99) in RatGTEEx, Zhao et al. (117) in CattleGTEEx, and Teng et al. (106) in PigGTEEx, collectively suggesting that the genetic regulation of gene expression is evolutionarily conserved to a certain extent across species.

Integrating multispecies eQTL and GWAS results allows exploration of the evolutionary conservation of gene regulation underlying complex traits and diseases across species through the transcriptome-wide association study framework. The RatXcan, a cross-species polygenic translation framework based on RatGTEEx and human GTEEx, enables the prediction of complex phenotypes between species in the transcriptome-wide association study framework (118). These



findings promise to transform our understanding of evolutionary mechanisms underlying gene regulation variation and complex phenotype variation across species.

5. DEEP LEARNING MODELING

Deep learning, or deep neural networks, has demonstrated exceptional performance in various biological applications, particularly those involving extensive data sets, e.g., AlphaFold2 (119) for predicting the 3D structures of proteins and DeepVariant (120) for calling genetic variants in genome sequencing. Although classical statistical methods, like the linear mixed model, have been used widely in data-driven genomics research, neural networks can detect complex patterns in the genomic data without relying on many statistical assumptions. This flexibility allows for effective handling of nonlinearity and complex dependencies (121, 122). In genomics research, particularly in the study of transcriptional regulation, the complexity and high-dimensional nature of genomics and epigenomics involve the integration of diverse and extensive data sets. This includes not only DNA sequences but also a variety of heterogeneous information, such as epigenetic marks identified by ChIP-seq and ATAC-seq, as well as differential gene expression and alternative splicing patterns provided by RNA-seq across cells, tissues, and organisms. Deep learning is well-suited to address these challenges due to its state-of-the-art performance in pattern recognition, heterogeneous data integration, flexibility, and scalability.

5.1. Deep Learning Architectures

Major neural network architectures include, e.g., fully connected, convolutional, and recurrent architectures; autoencoders; and transformers (see 121–123 for a detailed review of neural network architectures in genomics). Deep learning's success relies on aligning the customized neural network architecture with the specific properties of the data. In understanding transcriptional regulation, convolutional or recurrent neural networks are the most commonly employed deep learning architectures (124). For example, when annotating the genome directly from the DNA sequence, a convolutional neural network architecture is frequently used, as seen in models like DeepSEA (125; refer to 126, 127 for a detailed review of neural network architectures in genome annotation). The DeepSEA model, equipped with tens of millions of parameters, excels at determining the presence or absence of nearly 1,000 chromatin features from a given 1,000-bp DNA sequence. These features include aspects such as TFBSs, chromatin accessibility, and histone modifications. DeepSEA's superior performance over nonneural network methods is achieved by detecting local patterns in sequential data through the application of convolutional layers that efficiently capture spatial and sequential dependencies. Importantly, this example differs from use of experimental measures such as ChIP-seq to predict functional annotations, for example, enhancers.

In the realm of comparative genomics and epigenomics, deep learning has streamlined the translation of genetic and biological findings across diverse species. A notable example is the transfer of annotations across species. Deep learning models, when trained on one or a few information-rich species, can effectively contribute to the genome annotation of less information-rich species. This approach proves invaluable, considering the formidable challenge of experimentally annotating every species. A few studies have validated this approach (20, 126), demonstrating its efficacy in transferring specific annotations, such as transcription start sites (128) and enhancer activity (20, 129). However, some models encounter challenges in generalizing across species, highlighting limitations in cross-species prediction linked to evolutionary conservation. Expanding the training data set to encompass a broader range of species holds promise for improving its cross-species applicability (129, 130).

Another illustrative example involves comparing functional annotations across multiple species to infer regulatory functional conservation (131, 132). This approach is particularly valuable, recognizing that DNA sequence conservation does not necessarily imply functional conservation across tissues between species (133, 134). For instance, the pseudo-Siamese network is adapted to represent each species as one branch, incorporating input data such as epigenome and gene expression profiles, as well as DNA sequences. This adaptation enables the network to learn genomic conservations at the functional level between multiple species (132).

5.2. Challenges and Opportunities for Deep Learning Applications in Genomics

Deep learning is often considered a black box with limited interpretability (26), even as its applications continue to expand rapidly. Therefore, a systematic evaluation of the growing number of applications for both robustness and interpretability (135), accompanied by validation through experimentation, becomes critical (135, 136).

Given deep learning's great flexibility, domain knowledge plays a crucial role in its successful application in genomics. As outlined in Reference 121, insights into REs inform neural network design, directing the use of a convolutional structure to capture spatial locality and selecting small filter sizes to accommodate the short length of regulatory motifs. The earlier example (132), which employs deep learning to investigate functional conservation across multiple species, further emphasizes the pivotal role of domain knowledge. The model's structure is tailored based on the number of species involved in the study and can be adjusted to incorporate various contexts, such as tissue types, developmental stages, and evolutionary periods. The model's input and output hinge on the assumption that orthologous regions are more likely to exhibit functional conservation than their nonorthologous counterparts. These examples underscore the critical importance of domain expertise in guiding the development and refinement of deep learning models.

We anticipate a growing prevalence of a synergistic approach that integrates flexible machine learning techniques, such as deep learning, with classical statistical methods in the fields of comparative genomics and epigenomics. A noteworthy illustration of this integrated approach is the TACIT toolbox (63). In its initial stages, TACIT employs deep learning to predict enhancer activities across more than 100 species. Subsequently, in exploring associations between phenotypes and enhancer activities across diverse species, TACIT incorporates linear mixed models. This dual strategy exemplifies the potential of combining classical statistical methods with flexible deep learning techniques to unravel complex biological phenomena in comparative genomics and epigenomics, particularly in the realm of transcriptional regulation.

6. APPLICATION OF CUTTING-EDGE TECHNOLOGIES FOR UNDERSTANDING TRANSCRIPTIONAL REGULATION

Cutting-edge technologies including long-read and SC sequencing, and functional screening tools like MPRA and CRISPR, have become instrumental in dissecting the complex regulatory networks that govern gene expression. This section delves into the transformative impact of these cutting-edge technologies, showcasing their pivotal role in advancing our understanding of transcriptional regulation.

6.1. Single-Cell Technology

SC genomics is revolutionizing molecular and cellular biology. The use of whole tissue-level or bulk analysis in molecular phenotypes involves data from a mixture of cell types. Bulk analyses thus cannot describe unique transcriptomic and epigenomic patterns of individual cells and especially overlook functions specific to rare cell types (137). To maximize our understanding of



RE function within and across species, we need to measure genomic function at the cellular level, leading to the discovery of new cell types and genome annotation (138). These technologies, combined with genotype data, facilitate linking genetic variation to molecular phenotypes with high precision, exemplified by the identification of cell-type-specific eQTLs in human blood cells (139). Such data can be used to construct genotype-specific regulatory networks and potentially predict physiological phenotypes (140). The development of improved methods for both the creation and analysis of SC data is extremely dynamic, and we do not intend to comprehensively describe these technical advances. Readers interested in exploring methods for creating single-cell RNA (scRNA)-seq, scATAC-seq, or other SC data sets, including spatial transcriptomics and multi-omic methods combining several modalities, are referred to recent reviews (141, 142). Another highly dynamic field that promises to add to the livestock genome-to-phenome toolbox is organoid modeling of tissues and organs (143). SC analysis has helped establish the validity of such organoid modeling (144), although SC genomic analysis of organoids has not yet been reported for farm animal species.

Several recent reports have integrated SC transcriptomics data across species, comparing livestock, human, and mouse cell type transcriptomes including gut, muscle, neural, and respiratory tissues to validate annotations and explore cell type similarities. Studies in pigs (145, 146), cattle (147, 148), and chickens (149, 150) have shown common immunological traits and genomic similarities, aiding translational research and model development. SC epigenomics methods have matured rapidly, offering the exciting potential to uncover master REs and factors within specific cell types, even in complex tissues. These methods can also be associated with noncoding genetic variation (151). Currently, experimental platforms for sc-epigenomics emphasize different aspects such as histone and DNA modification, chromatin accessibility, DNA–protein interactions, and 3D chromatin structure (152). For example, both scRNA-seq and scATAC-seq were employed to investigate cell type expression and regulation in porcine and cattle peripheral blood mononuclear cells (PBMCs), respectively (153, 154). Both papers reported the integration of scATAC-seq and scRNA-seq to determine which TF had TFBS enriched in regions with cell-type chromatin accessibility. Yang et al. (154) also used Cicero to identify regions of open chromatin near target genes that were correlated with the expression of nearby target genes and TF that had enriched TFBSs in those regions. Furthermore, Gao et al. (153) analyzed SC expression and chromatin information of immune cells during *in vitro* lipopolysaccharide stimulation as a model of mastitis. These multi-modal investigations demonstrate efficacy in finding important TFs and regulatory elements, providing new insights into mechanisms underlying gene expression in bovine and porcine PBMCs.

Deep learning methods are being developed to take advantage of rapidly growing SC genomics multi-omic data sets for advanced understanding of regulatory mechanisms (155). Furthermore, Kim et al. (156) reviewed gene regulatory network inference techniques ranging from correlation and regression methods to probabilistic models, dynamical system approaches, and deep learning algorithms. Each method is designed to resolve complex gene regulatory network relationships from integrated data. An example is Tangeman et al.'s (157) construction of a comprehensive SC multi-omic atlas for lens development in chicken. They used snRNA-seq, snATAC-seq, and CUT&RUN-seq for both histone modifications and regulatory proteins to reveal new mechanisms for determining cell destiny and regulatory networks linked to cataract formation. The work emphasizes mainly the significance of integrative analysis and multi-omic techniques in comprehending tissue development, with a focus on TF activity and dynamic chromatin reconfiguration that promote lens fiber cell differentiation.

Finally, scRNA-seq and snATAC-seq have been used to discover genetic effects on transcriptional regulation. Bao et al. (158) demonstrated SC transcriptome and chromatin accessibility analysis of PBMCs, revealing immune cell heterogeneity and breed differences in pigs. The

results reveal unique cell markers and TFs associated with various immune cell types. By analyzing transcriptomes and creating chromatin maps, this study highlights the role of innate immunity in breed-specific immune traits, shedding light on the relationship between gene expression in specific cell types and the impact of genetic variation (159).

These comparative and integrative approaches across modalities not only advance our understanding of diseases and tissue functions for advanced breeding applications but also highlight the utility of livestock models in biomedical research by uncovering key regulatory mechanisms underlying gene expression in livestock/domesticated species.

6.2. Long-Read Transcriptome Sequencing Technology

Long-read, full-length RNA sequencing technology, which can generate sequenced reads exceeding 10 kb, is employed widely to directly sequence RNA molecules without fragmentation or amplification. This approach enables the analysis of complete RNA transcripts, including full-length isoforms, transcriptional states, and splice variants (160). Compared to short-read RNA-seq, long-read RNA-seq offers advantages in characterizing complex RNA molecules by spanning intricate RNA structures (161), improving transcriptome assembly (162), and facilitating the identification of novel RNA species (163). The increased accuracy and comprehensive understanding of RNA transcripts provided by long-read RNA-seq offer significant potential for studying comparative transcriptome regulation.

The integration of long-read RNA-seq with the development of scRNA-seq and spatial transcriptomic analysis provides a comprehensive perspective on gene expression dynamics across various biological conditions, species, or developmental stages. High-resolution data from long-read full-length RNA-seq enables the discovery of species-specific transcripts (164), alternative splicing events (165), and regulatory networks (166) underlying phenotypic differences between organisms. Insights gained from long-read scRNA-seq further enhance understanding of evolutionary changes in gene expression between species (167). The latest development in spatial transcriptomics makes it feasible to profile cells within their tissue context (168), enabling better definition of cell types through spatially resolved cell-specific isoform expression data (169).

6.3. MPRA Technology

MPRA is a powerful tool for dissecting the functional significance of REs across species, biological conditions, or cellular contexts (170). Widely employed for enhancer analysis across species, MPRA also facilitates the identification of species-, cell type-, or tissue-specific REs (171). It has found extensive application in studying the functional impact of genetic variants on gene expression, particularly those located in noncoding regions (172). Comparative MPRA studies across different genetic backgrounds can pinpoint causal variants that alter RE activity, contributing to phenotypic differences and elucidating evolutionary conservation or divergent gene regulatory mechanisms (173). In addition, MPRA can also provide evidence for allele-specific regulatory activities across different cell types and species (27).

By integrating epigenomic data such as ChIP-seq or ATAC-seq, the relationships between chromatin accessibility, histone modifications, and RE activity can be studied extensively using the MPRA assay (174). When combined with scRNA-seq, SC MPRA enables the validation of REs at the SC level, providing insights into the heterogeneity of regulatory activity across individual cells (175). Integrating experimental data from MPRA with scRNA-seq and epigenetic profiling also allows for the validation and refinement of computational models predicting gene expression and regulatory interactions, which can improve the accuracy of predictive models and enhance our understanding of the regulatory code (176).



Furthermore, combining time-resolved MPRA with series scRNA-seq and epigenetic data facilitates dynamic profiling, which can capture the temporal dynamics of transcriptional regulation. This approach can identify key regulatory events during cellular transitions or responses across species, cell types, or disease contexts (175, 177).

6.4. CRISPR Technologies

Whereas MPRA identifies regulatory effects of the noncoding genome, CRISPR strategies pinpoint their transcriptional targets. Used in conjunction, these approaches converge on the regulatory mechanisms that lead to a specific phenotype (178). The CRISPR-Cas toolbox provides a versatile platform for tethering effectors capable of binding DNA within a limited range of the target, enabling transcriptional regulation or epigenetic modifications. Epigenomic perturbations use the nuclease-dead Cas9 protein tethered to specific activation/repression domains of TFs such as VP64/KRAB, to DNA methyltransferases such as DNMT3A, to TET1 for demethylation, or to methylation or acetylation histone-modifying enzymes, and their combinations (179). The single-guide RNA directs these effectors to the target sequence without altering the genome, thereby improving system accuracy. Additionally, barcoding methods for SC CRISPR screening, such as Perturb-seq (180) and CRISPR-seq (181), facilitate epigenome studies based on variable unsupervised functional perturbation screens.

The targeted nature and versatility of CRISPR approaches have led to major breakthroughs in understanding the role of noncoding regions in the epigenetic regulation of gene transcription. For instance, pooled CRISPR screens have identified REs influencing cell proliferation or cell cycle arrest in response to DNA damage (182). CRISPRoff and CRISPRon technologies allow rewiring of gene expression programs by writing and erasing epigenetic memories. Key questions pertain to the stability and heritability of these edits. CRISPRoff can silence multiple genes, including those without CpG islands and enhancers. This silencing persists through the differentiation of induced pluripotent stem cells into neurons and is erased by the CRISPRon counterpart without inducing a DNA damage response (183). Current efforts focus on elucidating the chromatin-related mechanisms mediating the initiation, spread, and maintenance of histone and DNA methylation marks and their causative or consequential effects on transcriptional regulation. An epigenetic editing tool kit, compatible with combinatorial chromatin marks targeting, has revealed the role of specific chromatin marks in the transcription of specific loci in mouse ES cells while also highlighting further layers of regulation (184).

Elucidating the regulatory mechanisms controlling tissue-specific gene expression requires cell systems that closely mimic the native genomic context. Advances in gene editing efficiency have facilitated its application in organoid models derived from pluripotent or adult stem cells. For example, combining SC CRISPR screening with flow cytometry in organoids has provided unprecedented insights into the regulation of human enteroendocrine cell differentiation (185) and, conversely, into factors important for maintaining an immature progenitor state (186). The OSCAR (organoid-based single-cell CRISPR screening analyzed with regulons) framework was developed to use changes in regulon activities as readouts to dissect liver lineage specifiers (187).

In animal science, organoids have been developed in major farm animal species (188) and will be valuable for identifying intrinsic and environmentally induced epigenetic changes affecting transcriptional regulation. However, genome editing efforts have so far relied on simple cell systems, albeit as physiologically relevant as possible to the traits being studied. Recently, altering the methylation status of the AKT1 promoter in a model of bovine mastitis mammary epithelial cells reactivated milk protein synthesis (189). In taxa where organoids are not yet fully developed, such as salmonids, research relies on the use of established cell lines suitable for CRISPR-Cas

engineering (190), but systems such as organotypic cell-based intestinal platforms may soon be employed (191).

Genome-wide CRISPR screens of farm animal species built on the improved annotations of these genomes (18, 192) and on the use of biologically relevant cell systems will follow the initial reports on mice and humans. These screens will inform the downstream transcriptomic effects of predicted—or MPRA-identified—REs and enable the validation of the variants pointed by fine genetic mapping and large-scale eQTL studies (105, 106). Development of more complex organoid systems for enhanced mimicry of organs will support further *in vitro* mechanistic studies. Moreover, the rapid evolution of CRISPR technology offers new technical possibilities and fields of investigation. CRISPR-associated transposases represent new approaches for programmable integration of large DNA sequences without double-strand breaks, and studies to understand and enhance the activity of CRISPR-associated transposases on chromatin will be pivotal for successful translational efforts (193).

7. FUTURE DIRECTIONS

Effective future studies will harness the power of comparative analysis of noncoding REs at both the DNA and functional (functional conservation of REs) level for diverse tissues across developmental stages under different physiological conditions across species and lineages. Comparative approaches that examine the evolutionary conservation and divergence of DNA elements across species can help to link regulatory divergence to phenotypic differences and species-specific adaptation (194). The scale and quality of new genomic and epigenomic resources for mammalian and avian species, including farmed animals, provide an unprecedented opportunity to investigate the evolutionary conservation of REs and the enriched genetic variants within these regions that control transcriptional regulation and contribute to complex traits.

One advantage that farmed animals have over other mammalian species is that access to multiple tissue types and developmental stages is comparatively easy (e.g., 22). Analysis of developmental stages and tissue types across several species provides further information to unravel complex shared and unique regulatory features of the genome and annotate the noncoding DNA elements, such as TFBSs, that regulate tissue-level gene expression. For example, the GENESWitCH project has provided a catalog of noncoding DNA elements in pig and chicken across multiple developmental stages, and the AQUA-FAANG project has produced a similar resource for five teleost fish species (195).

Despite significant progress at the bulk tissue level, the underlying molecular mechanism at the cellular level of gene regulation needs to be determined. In the next decade, enormous epigenomic data at the SC level will be generated. Therefore, improved data infrastructure and computing environments are crucial for continued advancement of SC genomics. A future vision for SC genomics in farm animals encompasses a platform offering secure, efficient cross-species and cross-modality research capabilities. One environment that already has much of this vision is the cloud-native computing platform Terra, used by the Human Cell Atlas community to facilitate collaborative biomedical research worldwide (196). Terra functionality serves as a secure platform for workflow development, data access, and analysis tools, providing infrastructure for cross-species analysis (197). Similarly, Galaxy, developed by Single Cell Expression Atlas (198), offers an open-source, web-based platform for reproducible research with a wide range of bioinformatic software, facilitating user-custom workflows. Adapting these frameworks for farm animal SC genomics data would significantly benefit the farm animal genomics community by enhancing the ability to use SC genomics data for understanding of genomic/genetic variation and its role in phenotype at the cellular level.



8. CONCLUDING REMARKS

Exploring transcriptional regulation through the lenses of comparative genomics and epigenomics has significantly enriched our comprehension of the complex orchestration of gene expression across diverse biological contexts. The identification of REs and genetic variants, particularly within noncoding regions, has provided crucial insights into the evolutionary dynamics that govern transcriptional regulations. As we move forward, future research endeavors should focus on expanding the breadth and depth of comparative epigenomic studies, encompassing a broader spectrum of physiological conditions, developmental stages, and environmental perturbations. This includes an emphasis on acquiring comprehensive epigenomic data from nonmodel organisms, such as farm animals, to enhance our understanding of transcriptional regulation beyond conventional model systems. The integration of high-throughput technologies, advanced sequencing methodologies, and sophisticated computational approaches, including deep learning models, will be pivotal for inferring meaningful patterns from the ever-growing omics data sets. Moreover, rapidly evolving functional screening tools like MPRA and CRISPR will continue to play a crucial role in validating REs and unraveling the intricacies of transcriptional networks. This interdisciplinary approach promises to discover new layers of regulatory complexity, shaping the future landscape of genomic and epigenomic research.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We apologize for not being able to cite many outstanding peer-reviewed publications on this topic due to limited citations allowed. D.G. and Y.W. were supported by Agriculture and Food Research Initiative Competitive grants nos. 2020-67015-31175, 2021-67015-33396, and 2022-67015-36215 (H.Z.) from the US Department of Agriculture National Institute of Food and Agriculture, Multistate Research Project NRSP8 and NC1170 (H.Z.), and the California Agricultural Experimental Station (H.Z.).

LITERATURE CITED

1. Hagai T, Chen X, Miragaia RJ, Rostom R, Gomes T, et al. 2018. Gene expression variability across cells and species shapes innate immunity. *Nature* 563(7730):197–202
2. Wolf S, Melo D, Garske KM, Pallares LF, Lea AJ, Ayroles JF. 2023. Characterizing the landscape of gene expression variance in humans. *PLoS Genet.* 19(7):e1010833
3. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, et al. 2012. Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337(6099):1190–95
4. Mostafavi H, Spence JP, Naqvi S, Pritchard JK. 2023. Systematic differences in discovery of genetic effects on gene expression and complex traits. *Nat. Genet.* 55(11):1866–75
5. Pan Z, Wang Y, Wang M, Wang Y, Zhu X, et al. 2023. An atlas of regulatory elements in chicken: a resource for chicken genetics and genomics. *Sci. Adv.* 9(18):eade1204
6. Pan Z, Yao Y, Yin H, Cai Z, Wang Y, et al. 2021. Pig genome functional annotation enhances the biological interpretation of complex traits and human disease. *Nat. Commun.* 12:5848
7. Hill MS, Vande Zande P, Wittkopp PJ. 2021. Molecular and evolutionary processes generating variation in gene expression. *Nat. Rev. Genet.* 22(4):203–15
8. Shih CH, Fay J. 2021. *Cis*-regulatory variants affect gene expression dynamics in yeast. *eLife* 10:e68469
9. Klemm SL, Shipony Z, Greenleaf WJ. 2019. Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* 20(4):207–20



10. Li YI, van de Geijn B, Raj A, Knowles DA, Petti AA, et al. 2016. RNA splicing is a primary link between genetic variation and disease. *Science* 352(6285):600–4
11. Xiao M-S, Zhang B, Li Y-S, Gao Q, Sun W, Chen W. 2016. Global analysis of regulatory divergence in the evolution of mouse alternative polyadenylation. *Mol. Syst. Biol.* 12(12):890
12. Pai AA, Cain CE, Mizrahi-Man O, De Leon S, Lewellen N, et al. 2012. The contribution of RNA decay quantitative trait loci to inter-individual variation in steady-state gene expression levels. *PLOS Genet.* 8(10):e1003000
13. Genereux DP, Serres A, Armstrong J, Johnson J, Marinescu VD, et al. 2020. A comparative genomics multitool for scientific discovery and conservation. *Nature* 587(7833):240–45
14. Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, et al. 2021. Towards complete and error-free genome assemblies of all vertebrate species. *Nature* 592(7856):737–46
15. Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, et al. 2022. The complete sequence of a human genome. *Science* 376(6588):44–53
16. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, et al. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414):57–74
17. Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shores N, et al. 2020. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* 583(7818):699–710
18. Kern C, Wang Y, Xu X, Pan Z, Halstead M, et al. 2021. Functional annotations of three domestic animal genomes provide vital resources for comparative and agricultural research. *Nat. Commun.* 12:1821
19. Hardison RC, Taylor J. 2012. Genomic approaches towards finding *cis*-regulatory modules in animals. *Nat. Rev. Genet.* 13(7):469–83
20. Kaplow IM, Schäffer DE, Wirthlin ME, Lawler AJ, Brown AR, et al. 2022. Inferring mammalian tissue-specific regulatory conservation by predicting tissue-specific differences in open chromatin. *BMC Genom.* 23(1):291
21. Snetkova V, Ypsilanti AR, Akiyama JA, Mannion BJ, Plajzer-Frick I, et al. 2021. Ultraconserved enhancer function does not require perfect sequence conservation. *Nat. Genet.* 53(4):521–28
22. Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, et al. 2015. Enhancer evolution across 20 mammalian species. *Cell* 160(3):554–66
23. Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447(7146):799–816
24. Dutta P, Talenti A, Young R, Jayaraman S, Callaby R, et al. 2020. Whole genome analysis of water buffalo and global cattle breeds highlights convergent signatures of domestication. *Nat. Commun.* 11:4739
25. Clark TC, Naseer S, Gundappa MK, Laurent A, Perquis A, et al. 2023. Conserved and divergent arms of the antiviral response in the duplicated genomes of salmonid fishes. *Genomics* 115(4):110663
26. Dermitzakis ET, Clark AG. 2002. Evolution of transcription factor binding sites in mammalian gene regulatory regions: conservation and turnover. *Mol. Biol. Evol.* 19(7):1114–21
27. Kuderna LFK, Ulirsch JC, Rashid S, Ameen M, Sundaram L, et al. 2024. Identification of constrained sequence elements across 239 primate genomes. *Nature* 625:735–42
28. Andrews G, Fan K, Pratt HE, Phalke N, Zoonomia Consort., et al. 2023. Mammalian evolution of human *cis*-regulatory elements and transcription factor binding sites. *Science* 380(6643):eabn7930
29. Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, et al. 2014. A comparative encyclopedia of DNA elements in the mouse genome. *Nature* 515(7527):355–64
30. Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, et al. 2012. A map of the *cis*-regulatory sequences in the mouse genome. *Nature* 488(7409):116–20
31. Zu S, Li YE, Wang K, Armand EJ, Mamde S, et al. 2023. Single-cell analysis of chromatin accessibility in the adult mouse brain. *Nature* 624(7991):378–89
32. Zhang K, Hocker JD, Miller M, Hou X, Chiou J, et al. A single-cell atlas of chromatin accessibility in the human genome. *Cell* 184(24):5985–6001.e19
33. Georgakopoulos-Soares I, Deng C, Agarwal V, Chan CSY, Zhao J, et al. 2023. Transcription factor binding site orientation and order are major drivers of gene regulatory activity. *Nat. Commun.* 14:2333
34. Du AY, Chobirko JD, Zhuo X, Feschotte C, Wang T. 2024. Regulatory transposable elements in the encyclopedia of DNA elements. *Nat. Commun.* 15:7594



35. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, et al. 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 22(9):1775–89
36. Uszczynska-Ratajczak B, Lagarde J, Frankish A, Guigó R, Johnson R. 2018. Towards a complete map of the human long non-coding RNA transcriptome. *Nat. Rev. Genet.* 19(9):535–48
37. Statello L, Guo CJ, Chen LL, Huarte M. 2021. Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* 22(2):96–118
38. Jehl F, Muret K, Bernard M, Boutin M, Lagoutte L, et al. 2020. An integrative atlas of chicken long non-coding genes and their annotations across 25 tissues. *Sci. Rep.* 10:20457
39. Degalez F, Charles M, Foissac S, Zhou H, Guan D, et al. 2024. Enriched atlas of lncRNA and protein-coding genes for the GRCg7b chicken assembly and its functional annotation across 47 tissues. *Genomics* 14:6588
40. Lagarrigue S, Lorthois M, Degalez F, Gilot D, Derrien T. 2022. LncRNAs in domesticated animals: from dog to livestock species. *Mamm. Genome* 33(2):248–70
41. Degalez F, Bardou P, Lagarrigue S. 2024. GEGA (Gallus Enriched Gene Annotation): an online tool gathering genomics and functional information across 47 tissues for protein-coding genes and lncRNA enriched atlas including Ensembl & Refseq genome annotations. *NAR Genom. Bioinform.* 6(3):lqae101
42. Kurylo C, Guyomar C, Foissac S, Djebali S. 2023. TAGADA: a scalable pipeline to improve genome annotations with RNA-seq data. *NAR Genom. Bioinform.* 5(4):lqad089
43. Sarropoulos I, Marin R, Cardoso-Moreira M, Kaessmann H. 2019. Developmental dynamics of lncRNAs across mammalian organs and species. *Nature* 571(7766):510–14
44. Hezroni H, Koppstein D, Schwartz MG, Avrutin A, Bartel DP, Ulitsky I. 2015. Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep.* 11(7):1110–22
45. Ulitsky I. 2016. Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nat. Rev. Genet.* 17(10):601–14
46. Quinn JJ, Chang HY. 2016. Unique features of long non-coding RNA biogenesis and function. *Nat. Rev. Genet.* 17(1):47–62
47. Muret K, Désert C, Lagoutte L, Boutin M, Gondret F, et al. 2019. Long noncoding RNAs in lipid metabolism: literature review and conservation analysis across species. *BMC Genom.* 20:882
48. Foissac S, Djebali S, Munyard K, Vialaneix N, Rau A, et al. 2019. Multi-species annotation of transcriptome and chromatin structure in domesticated animals. *BMC Biol.* 17:108
49. Smith J, Alfieri JM, Anthony N, Arensburger P, Athrey GN, et al. 2023. Fourth Report on Chicken Genes and Chromosomes 2022. *Cytogenet. Genome Res.* 162(8–9):405–528
50. Ross CJ, Rom A, Spinrad A, Gelbard-Solodkin D, Degani N, Ulitsky I. 2021. Uncovering deeply conserved motif combinations in rapidly evolving noncoding sequences. *Genome Biol.* 22:29
51. Huang W, Xiong T, Zhao Y, Heng J, Han G, et al. 2024. Computational prediction and experimental validation identify functionally conserved lncRNAs from zebrafish to human. *Nat. Genet.* 56:124–35
52. Degalez F, Allain C, Lagoutte L, Lagarrigue S. 2023. *Gene orthology detection for long non-coding RNA (LncRNA)*. Presented at the 39th International Society for Animal Genetics Conference, July 2–7, Cape Town, S. Afr.
53. Boix CA, James BT, Park YP, Meuleman W, Kellis M. 2021. Regulatory genomic circuitry of human disease loci by integrative epigenomics. *Nature* 590(7845):300–7
54. Modencode Consort., Roy S, Ernst J, Kharchenko PV, Kheradpour P, et al. 2010. Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* 330(6012):1787–97
55. Baranasic D, Hörtenhuber M, Balwiercz PJ, Zehnder T, Mukarram AK, et al. 2022. Multiomic atlas with functional stratification and developmental dynamics of zebrafish *cis*-regulatory elements. *Nat. Genet.* 54(7):1037–50
56. Stunnenberg HG, Int. Hum. Epigenome Consort., Hirst M. 2016. The International Human Epigenome Consortium: a blueprint for scientific collaboration and discovery. *Cell* 167(5):1145–49
57. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* 518(7539):317–30



58. Meuleman W, Muratov A, Rynes E, Halow J, Lee K, et al. 2020. Index and biological spectrum of human DNase I hypersensitive sites. *Nature* 584(7820):244–51
59. Vierstra J, Lazar J, Sandstrom R, Halow J, Lee K, et al. 2020. Global reference mapping of human transcription factor footprints. *Nature* 583(7818):729–36
60. Pang B, Snyder MP. 2020. Systematic identification of silencers in human cells. *Nat. Genet.* 52(3):254–63
61. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, et al. 2012. The accessible chromatin landscape of the human genome. *Nature* 489(7414):75–82
62. Sullivan PF, Meadows JRS, Gazal S, Phan BN, Li X, et al. 2023. Leveraging base-pair mammalian constraint to understand genetic variation and human disease. *Science* 380(6643):eabn2937
63. Kaplow IM, Lawler AJ, Schäffer DE, Srinivasan C, Sestili HH, et al. 2023. Relating enhancer genetic variation across mammals to complex phenotypes using machine learning. *Science* 380(6643):eabm7993
64. Christmas MJ, Kaplow IM, Genereux DP, Dong MX, Hughes GM, et al. 2023. Evolutionary constraint and innovation across hundreds of placental mammals. *Science* 380:eabn3943
65. Xiao S, Xie D, Cao X, Yu P, Xing X, et al. 2012. Comparative epigenomic annotation of regulatory DNA. *Cell* 149(6):1381–92
66. Zemke NR, Armand EJ, Wang W, Lee S, Zhou J, et al. 2023. Conserved and divergent gene regulatory programs of the mammalian neocortex. *Nature* 624(7991):390–402
67. Smith ZD, Chan MM, Humm KC, Karnik R, Mekhoubad S, et al. 2014. DNA methylation dynamics of the human preimplantation embryo. *Nature* 511(7511):611–15
68. Partridge EC, Chhetri SB, Prokop JW, Ramaker RC, Jansen CS, et al. 2020. Occupancy maps of 208 chromatin-associated proteins in one human cell type. *Nature* 583(7818):720–28
69. Fan K, Pfister E, Weng Z. 2023. Toward a comprehensive catalog of regulatory elements. *Hum. Genet.* 142(8):1091–111
70. Liu H, Zeng Q, Zhou J, Bartlett A, Wang B-A, et al. 2023. Single-cell DNA methylome and 3D multi-omic atlas of the adult mouse brain. *Nature* 624:366–77
71. Tian W, Zhou J, Bartlett A, Zeng Q, Liu H, et al. 2023. Single-cell DNA methylation and 3D genome architecture in the human brain. *Science* 382(6667):eadf5357
72. Yue X, Xie Z, Li M, Wang K, Li X, et al. 2022. Simultaneous profiling of histone modifications and DNA methylation via nanopore sequencing. *Nat. Commun.* 13:7939
73. Klughammer J, Romanovskaia D, Nemc A, Posautz A, Seid CA, et al. 2023. Comparative analysis of genome-scale, base-resolution DNA methylation profiles across 580 animal species. *Nat. Commun.* 14:232
74. Haghani A, Li CZ, Robeck TR, Zhang J, Lu AT, et al. 2023. DNA methylation networks underlying mammalian traits. *Science* 381(6658):eabq5693
75. Chen S, Liu S, Shi S, Yin H, Tang Y, et al. 2024. Cross-species comparative DNA methylation reveals novel insights into complex traits genetics among cattle, sheep and goats. *Mo. Biol. Evol.* 41(2):msae003
76. Lu AT, Fei Z, Haghani A, Robeck TR, Zoller JA, et al. 2023. Universal DNA methylation age across mammalian tissues. *Nat. Aging* 3(9):1144–66
77. Yu M, Ren B. 2017. The three-dimensional organization of mammalian genomes. *Annu. Rev. Cell Dev. Biol.* 33:265–89
78. Avdeyev P, Zhou J. 2022. Computational approaches for understanding sequence variation effects on the 3D genome architecture. *Annu. Rev. Biomed. Data Sci.* 5:183–204
79. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragozy T, et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326(5950):289–93
80. Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, et al. 2015. Chromatin architecture reorganization during stem cell differentiation. *Nature* 518(7539):331–36
81. Liao Y, Zhang X, Chakraborty M, Emerson JJ. 2021. Topologically associating domains and their role in the evolution of genome structure and function in *Drosophila*. *Genome Res.* 31(3):397–410
82. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, et al. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159(7):1665–80
83. Corbo M, Damas J, Bursell MG, Lewin HA. 2022. Conservation of chromatin conformation in carnivores. *PNAS* 119(9):e2120555119



84. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, et al. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485(7398):376–80
85. Cardoso-Moreira M, Halbert J, Vallotton D, Velten B, Chen C, et al. 2019. Gene expression across mammalian organ development. *Nature* 571(7766):505–9
86. Sepp M, Leiss K, Murat F, Okonechnikov K, Joshi P, et al. 2023. Cellular development and evolution of the mammalian cerebellum. *Nature* 625:788–96
87. Gorkin DU, Barozzi I, Zhao Y, Zhang Y, Huang H, et al. 2020. An atlas of dynamic chromatin landscapes in mouse fetal development. *Nature* 583(7818):744–51
88. van der Velde A, Fan K, Tsuji J, Moore JE, Purcaro MJ, et al. 2021. Annotation of chromatin states in 66 complete mouse epigenomes during development. *Commun. Biol.* 4:239
89. Domcke S, Hill AJ, Daza RM, Cao J, O'Day DR, et al. 2020. A human cell atlas of fetal chromatin accessibility. *Science* 370(6518):eaba7612
90. Claussnitzer M, Dankel SN, Klocke B, Grallert H, Glunk V, et al. 2014. Leveraging cross-species transcription factor binding site patterns: from diabetes risk loci to disease mechanisms. *Cell* 156(1):343–58
91. Albert FW, Kruglyak L. 2015. The role of regulatory variation in complex traits and disease. *Nat. Rev. Genet.* 16(4):197–212
92. Aguet F, Alasoo K, Li YI, Battle A, Im HK, et al. 2023. Molecular quantitative trait loci. *Nat. Rev. Methods Primers* 3:4
93. GTEx Consort. 2020. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369(6509):1318–30
94. Vösa U, Claringbould A, Westra HJ, Bonder MJ, Deelen P, et al. 2021. Large-scale *cis*- and *trans*-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet.* 53(9):1300–10
95. Kerimov N, Hayhurst JD, Peikova K, Manning JR, Walter P, et al. 2021. A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nat. Genet.* 53(9):1290–99
96. Huang D, Feng X, Yang H, Wang J, Zhang W, et al. 2023. QTLbase2: an enhanced catalog of human quantitative trait loci on extensive molecular phenotypes. *Nucleic Acids Res.* 51(D1):D1122–28
97. Kerimov N, Tambets R, Hayhurst JD, Rahu I, Kolberg P, et al. 2023. eQTL Catalogue 2023: new datasets, X chromosome QTLs, and improved detection and visualisation of transcript-level QTLs. *PLoS Genet.* 19(9):e1010932
98. Pallares LF, Melo D, Wolf S, Cofer EM, Abhyankar V, et al. 2023. Saturating the eQTL map in *Drosophila melanogaster*: genome-wide patterns of *cis* and *trans* regulation of transcriptional variation in outbred populations. bioRxiv 541576. <https://www.biorxiv.org/content/10.1101/2023.05.20.541576v3>
99. Munro D, Wang T, Chitre AS, Polesskaya O, Ehsan N, et al. 2022. The regulatory landscape of multiple brain regions in outbred heterogeneous stock rats. *Nucleic Acids Res.* 50(19):10882–95
100. Philip VM, He H, Saul MC, Dickson PE, Bubier JA, Chesler EJ. 2023. Gene expression genetics of the striatum of Diversity Outbred mice. *Sci. Data* 10:522
101. Jasinska AJ, Zelaya I, Service SK, Peterson CB, Cantor RM, et al. 2017. Genetic variation and gene expression across multiple tissues and developmental stages in a non-human primate. *Nat. Genet.* 49(12):1714–21
102. Fair BJ, Blake LE, Sarkar A, Pavlovic BJ, Cuevas C, Gilad Y. 2020. Gene expression variability in human and chimpanzee populations share common determinants. *eLife* 9:e59929
103. Lin W, Wall JD, Li G, Newman D, Yang Y, et al. 2023. Genetic regulatory effects in response to a high cholesterol, high fat diet in baboons. bioRxiv 551489. <https://doi.org/10.1101/2023.08.01.551489>
104. *Nat. Genet.* 2022. The CattleGTEx atlas reveals regulatory mechanisms underlying complex traits. 54(9):1273–74
105. Liu S, Gao Y, Canela-Xandri O, Wang S, Yu Y, et al. 2022. A multi-tissue atlas of regulatory variants in cattle. *Nat. Genet.* 54(9):1438–47
106. Teng J, Gao Y, Yin H, Bai Z, Liu S, et al. 2024. A compendium of genetic regulatory effects across pig tissues. *Nat. Genet.* 56(1):112–23



107. Guan D, Bai Z, Zhu X, Zhong C, Hou Y, et al. The ChickenGTEx pilot analysis: a reference of regulatory variants across 28 chicken tissues. bioRxiv 546670. <https://doi.org/10.1101/2023.06.27.546670v1>
108. Yazar S, Alquicira-Hernandez J, Wing K, Senabouth A, Gordon MG, et al. 2022. Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. *Science* 376(6589):eabf3041
109. Soskic B, Cano-Gamez E, Smyth DJ, Ambridge K, Ke Z, et al. 2022. Immune disease risk variants regulate gene expression dynamics during CD4⁺ T cell activation. *Nat. Genet.* 54(6):817–26
110. Schmiedel BJ, Gonzalez-Colin C, Fajardo V, Rocha J, Madrigal A, et al. 2022. Single-cell eQTL analysis of activated T cell subsets reveals activation and cell type-dependent effects of disease-risk variants. *Sci. Immunol.* 7(68):eabm2508
111. Kumasaka N, Rostom R, Huang N, Polanski K, Meyer KB, et al. 2023. Mapping interindividual dynamics of innate immune response at single-cell resolution. *Nat. Genet.* 55(6):1066–75
112. van der Wijst MGP, de Vries DH, Groot HE, Trynka G, Hon CC, et al. 2020. The single-cell eQTLGen consortium. *eLife* 9:e52155
113. IGVF Consort. 2023. The Impact of Genomic Variation on Function (IGVF) Consortium. ArXiv 13708. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10402186/>
114. Umans BD, Battle A, Gilad Y. 2021. Where are the disease-associated eQTLs? *Trends Genet.* 37(2):109–24
115. Natl. Advis. Child Health Hum. Dev. Counc. 2019. 201910 Developmental Genotype-Tissue Expression (dGTE_x) Project. <https://www.nichd.nih.gov/about/advisory/council/archive/201910/dGTEx-PPB-201910>
116. Tung J, Zhou X, Alberts SC, Stephens M, Gilad Y. 2015. The genetic architecture of gene expression levels in wild baboons. *eLife* 4:e04729
117. Zhao R, Talenti A, Fang L, Liu S, Liu G, et al. 2022. The conservation of human functional variants and their effects across livestock species. *Commun. Biol.* 5:1003
118. Santhanam N, Sanchez-Roige S, Liang Y, Chitre AS, Munro D, et al. 2023. RatXcan: framework for translating genetic results between species via transcriptome-wide association analyses. bioRxiv 494719. <https://www.biorxiv.org/content/10.1101/2022.06.03.494719v4>
119. Yang Z, Zeng X, Zhao Y, Chen R. 2023. AlphaFold2 and its applications in the fields of biology and medicine. *Signal Transduct. Target Ther.* 8:115
120. Poplin R, Chang PC, Alexander D, Schwartz S, Colthurst T, et al. 2018. A universal SNP and small-indel variant caller using deep neural networks. *Nat. Biotechnol.* 36(10):983–87
121. Zou J, Huss M, Abid A, Mohammadi P, Torkamani A, Talenti A. 2019. A primer on deep learning in genomics. *Nat. Genet.* 51:12–18
122. Eraslan G, Avsec Ž, Gagneur J, Theis FJ. 2019. Deep learning: new computational modelling techniques for genomics. *Nat. Rev. Genet.* 20(7):389–403
123. Sapoval N, Aghazadeh A, Nute MG, Antunes DA, Balaji A, et al. 2022. Current progress and open challenges for applying deep learning across the biosciences. *Nat. Commun.* 13:1728
124. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596(7873):583–89
125. Zhou J, Troyanskaya OG. 2015. Predicting effects of noncoding variants with deep learning-based sequence model. *Nat. Methods* 12(10):931–34
126. Routhier E, Mozziconacci J. 2022. Genomics enters the deep learning era. *PeerJ* 10:e13613
127. Alharbi WS, Rashid M. 2022. A review of deep learning applications in human genomics using next-generation sequencing data. *Hum. Genom.* 16:26
128. Khodabandelou G, Routhier E, Mozziconacci J. 2020. Genome annotation across species using deep convolutional neural networks. *PeerJ Comput. Sci.* 6:e278
129. Kelley DR. 2020. Cross-species regulatory sequence activity prediction. *PLOS Comput. Biol.* 16(7):e1008050
130. Cochran K, Srivastava D, Shrikumar A, Balsubramani A, Hardison RC, et al. 2022. Domain-adaptive neural networks improve cross-species prediction of transcription factor binding. *Genome Res.* 32(3):512–23
131. Kwon SB, Ernst J. 2021. Learning a genome-wide score of human-mouse conservation at the functional genomics level. *Nat. Commun.* 12:2495



132. Li J, Zhao T, Guan D, Pan Z, Bai Z, et al. 2023. Learning functional conservation between human and pig to decipher evolutionary mechanisms underlying gene expression and complex traits. *Cell Genom.* 3(10):100390
133. Bordeira-Carriço R, Teixeira J, Duque M, Galhardo M, Ribeiro D, et al. 2022. Multidimensional chromatin profiling of zebrafish pancreas to uncover and investigate disease-relevant enhancers. *Nat. Commun.* 13:1945
134. Pennacchio LA, Visel A. 2010. Limits of sequence and functional conservation. *Nat. Genet.* 42(7):557–58
135. Toneyan S, Tang Z, Koo PK. 2022. Evaluating deep learning for predicting epigenomic profiles. *Nat. Mach. Intell.* 4(12):1088–100
136. Novakovsky G, Dexter N, Libbrecht MW, Wasserman WW, Mostafavi S. 2023. Obtaining genetics insights from deep learning via explainable artificial intelligence. *Nat. Rev. Genet.* 24(2):125–37
137. Kolodziejczyk AA, Kim JK, Svensson V, Marioni JC, Teichmann SA. 2015. The technology and biology of single-cell RNA sequencing. *Mol. Cell* 58(4):610–20
138. Healey HM, Bassham S, Cresko WA. 2022. Single-cell Iso-Sequencing enables rapid genome annotation for scRNAseq analysis. *Genetics* 220(3):iyac017
139. van der Wijst MGP, Brugge H, de Vries DH, Deelen P, Swertz MA, et al. 2018. Single-cell RNA sequencing identifies celltype-specific *cis*-eQTLs and co-expression QTLs. *Nat. Genet.* 50(4):493–97
140. Cuomo ASE, Nathan A, Raychaudhuri S, MacArthur DG, Powell JE. 2023. Single-cell genomics meets human genetics. *Nat. Rev. Genet.* 24(8):535–49
141. Yarlagadda S, Giorgio TD. 2024. A guide to single-cell RNA sequencing analysis using web-based tools for non-bioinformatician. *FEBS J.* 291(12):2545–61
142. Zhang S, Li X, Lin J, Lin Q, Wong KC. 2023. Review of single-cell RNA-seq data clustering for cell-type identification and characterization. *RNA* 29(5):517–30
143. Ma P, Fang P, Ren T, Fang L, Xiao S. 2022. Porcine intestinal organoids: overview of the state of the art. *Viruses* 14(5):1110
144. Yin Y, Liu PY, Shi Y, Li P. 2021. Single-cell sequencing and organoids: a powerful combination for modelling organ development and diseases. *Rev. Physiol. Biochem. Pharmacol.* 179:189–210
145. Wang J, Gao M, Cheng M, Luo J, Lu M, et al. 2024. Single-cell transcriptional analysis of lamina propria lymphocytes in the jejunum reveals innate lymphoid cell-like cells in pigs. *J. Immunol.* 212(1):130–42
146. Li T, Morselli M, Su T, Million M, Larauche M, et al. 2023. Comparative transcriptomics reveals highly conserved regional programs between porcine and human colonic enteric nervous system. *Commun. Biol.* 6:98
147. Wang L, Gao P, Li C, Liu Q, Yao Z, et al. 2023. A single-cell atlas of bovine skeletal muscle reveals mechanisms regulating intramuscular adipogenesis and fibrogenesis. *J. Cachexia Sarcopenia Muscle* 14(5):2152–67
148. Cai C, Wan P, Wang H, Cai X, Wang J, et al. 2023. Transcriptional and open chromatin analysis of bovine skeletal muscle development by single-cell sequencing. *Cell Prolif.* 56(9):e13430
149. Maxwell M, Söderlund R, Härtle S, Wattrang E. 2024. Single-cell RNA-seq mapping of chicken peripheral blood leukocytes. *BMC Genom.* 25:124
150. Chen G, Chen J, Qi L, Yin Y, Lin Z, et al. 2024. Bulk and single-cell alternative splicing analyses reveal roles of TRA2B in myogenic differentiation. *Cell Prolif.* 57(2):e13545
151. Lyu P, Hoang T, Santiago CP, Thomas ED, Timms AE, et al. 2021. Gene regulatory networks controlling temporal patterning, neurogenesis, and cell-fate specification in mammalian retina. *Cell Rep.* 37(7):109994
152. Preissl S, Gaulton KJ, Ren B. 2023. Characterizing *cis*-regulatory elements using single-cell epigenomics. *Nat. Rev. Genet.* 24(1):21–43
153. Gao Y, Li J, Cai G, Wang Y, Yang W, et al. 2022. Single-cell transcriptomic and chromatin accessibility analyses of dairy cattle peripheral blood mononuclear cells and their responses to lipopolysaccharide. *BMC Genom.* 23:338
154. Yang P, Corbett R, Daharsh L, Uribe JH, Byrne KA, et al. 2024. Definition of regulatory elements and transcription factors controlling porcine immune cell gene expression at single cell resolution using single nucleus ATAC-seq. *Genomics* 116:110944



155. Forcato M, Romano O, Bicciato S. 2021. Computational methods for the integrative analysis of single-cell data. *Brief. Bioinform.* 22(3):bbaa042
156. Kim D, Tran A, Kim HJ, Lin Y, Yang JYH, Yang P. 2023. Gene regulatory network reconstruction: harnessing the power of single-cell multi-omic data. *npj Syst. Biol. Appl.* 9:51
157. Tangeman JA, Rebull SM, Grajales-Esquivel E, Weaver JM, Bendezu-Sayas S, et al. 2024. Integrated single-cell multiomics uncovers foundational regulatory mechanisms of lens development and pathology. *Development* 151(1):dev202249
158. Bao W, Fan H, Xu C, Du C, Wang H, et al. 2022. Single-cell transcriptomic and chromatin accessibility atlas of peripheral blood mononuclear cells reveal the immune cell heterogeneity of pigs. *Res. Square*. <https://www.researchsquare.com/article/rs-1887867/v1>
159. Sheng X, Guan Y, Ma Z, Wu J, Liu H, et al. 2021. Mapping the genetic architecture of human traits to cell types in the kidney identifies mechanisms of disease and potential treatments. *Nat. Genet.* 53(9):1322–33
160. Hu T, Chitnis N, Monos D, Dinh A. 2021. Next-generation sequencing technologies: an overview. *Hum. Immunol.* 82(11):801–11
161. Razaghi R. 2022. *Using nanopore sequencing to interrogate the genome and epigenome*. PhD Diss., Johns Hopkins Univ., Baltimore. <http://jhir.library.jhu.edu/handle/1774.2/68007>
162. Shumate A, Wong B, Pertea G, Pertea M. 2022. Improved transcriptome assembly using a hybrid of long and short reads with StringTie. *PLOS Comput. Biol.* 18(6):e1009730
163. Vlasova-St. Louis I. 2021. *Applications of RNA-Seq in Biology and Medicine*. London: InTechOpen
164. Oikonomopoulos S, Bayega A, Fahiminiya S, Djambazian H, Berube P, Ragoussis J. 2020. Methodologies for transcript profiling using long-read technologies. *Front. Genet.* 11:606
165. Benegas G, Fischer J, Song YS. 2022. Robust and annotation-free analysis of alternative splicing across diverse cell types in mice. *eLife* 11:e73520
166. Cardona-Alberich A, Tourbez M, Pearce SF, Sibley CR. 2021. Elucidating the cellular dynamics of the brain with single-cell RNA sequencing. *RNA Biol.* 18(7):1063–84
167. Wen L, Tang F. 2022. Recent advances in single-cell sequencing technologies. *Precis. Clin. Med.* 5(1):pbac002
168. Yuan CU, Quah FX, Hemberg M. 2024. Single-cell and spatial transcriptomics: bridging current technologies with long-read sequencing. *Mol. Aspects Med.* 96:101255
169. Lebrigand K, Bergensträhle J, Thrane K, Mollbrink A, Meletis K, et al. 2023. The spatial landscape of gene expression isoforms in tissue sections. *Nucleic Acids Res.* 51(8):e47
170. Inoue F, Kreimer A, Ashuach T, Ahituv N, Yosef N. 2019. Identification and massively parallel characterization of regulatory elements driving neural induction. *Cell Stem Cell* 25(5):713–27.e10
171. Dong S, Boyle AP. 2022. Prioritization of regulatory variants with tissue-specific function in the non-coding regions of human genome. *Nucleic Acids Res.* 50(1):e6
172. Findlay SD, Romo L, Burge CB. 2024. Quantifying negative selection in human 3' UTRs uncovers constrained targets of RNA-binding proteins. *Nat. Commun.* 15:85
173. Gallego Romero I, Lea AJ. 2023. Leveraging massively parallel reporter assays for evolutionary questions. *Genome Biol.* 24:26
174. Pang B, van Weerd JH, Hamoen FL, Snyder MP. 2022. Identification of non-coding silencer elements and their regulation of gene expression. *Nat. Rev. Mol. Cell Biol.* 24:383–95
175. Zhao S, Hong CKY, Myers CA, Granas DM, White MA, et al. 2023. A single-cell massively parallel reporter assay detects cell-type-specific gene regulation. *Nat. Genet.* 55(2):346–54
176. Koido M, Hon CC, Koyama S, Kawaji H, Murakawa Y, et al. 2023. Prediction of the cell-type-specific transcription of non-coding RNAs from genome sequences via machine learning. *Nat. Biomed. Eng.* 7(6):830–44
177. Bravo González-Blas C, Matetovici I, Hillen H, Taskiran II, Vandepoel R, et al. 2024. Single-cell spatial multi-omics and deep learning dissect enhancer-driven gene regulatory networks in liver zonation. *Nat. Cell Biol.* 26(1):153–67
178. Örd T, Örd D, Adler P, Örd T. 2023. Genome-wide census of ATF4 binding sites and functional profiling of trait-associated genetic variants overlapping ATF4 binding motifs. *PLOS Genet.* 19(10):e1011014



179. Martella A, Fisher DI. 2021. Regulation of gene expression and the elucidative role of CRISPR-based epigenetic modifiers and CRISPR-induced chromosome conformational changes. *CRISPR J.* 4(1):43–57
180. Dixit A, Parnas O, Li B, Chen J, Fulco CP, et al. 2016. Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* 167(7):1853–66.e17
181. Jaitin DA, Weiner A, Yofe I, Lara-Astiaso D, Keren-Shaul H, et al. 2016. Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-Seq. *Cell* 167(7):1883–96.e15
182. Borys SM, Younger ST. 2020. Identification of functional regulatory elements in the human genome using pooled CRISPR screens. *BMC Genom.* 21:107
183. Nuñez JK, Chen J, Pommier GC, Cogan JZ, Replogle JM, et al. 2021. Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing. *Cell* 184(9):2503–19.e17
184. Policarpi C, Munafo M, Tsagkris S, Carlini V, Hackett JA. 2024. Systematic epigenome editing captures the context-dependent instructive function of chromatin modifications. *Synth. Biol.* 56:1168–80
185. Lin L, DeMartino J, Wang D, Van Son GJF, Van Der Linden R, et al. 2023. Unbiased transcription factor CRISPR screen identifies ZNF800 as master repressor of enteroendocrine differentiation. *Science* 382(6669):451–58
186. Hansen SL, Larsen HL, Pikkupaura LM, Maciag G, Guiu J, et al. 2023. An organoid-based CRISPR-Cas9 screen for regulators of intestinal epithelial maturation and cell fate. *Sci. Adv.* 9(28):eadg4055
187. Liang J, Wei J, Cao J, Qian J, Gao R, et al. 2023. In-organoid single-cell CRISPR screening reveals determinants of hepatocyte differentiation and maturation. *Genome Biol.* 24:251
188. Kawasaki M, Goyama T, Tachibana Y, Nagao I, Ambrosini YM. 2022. Farm and companion animal organoid models in translational research: a powerful tool to bridge the gap between mice and humans. *Front. Med. Technol.* 4:895379
189. Liu J, Wei X, Zhang Y, Ran Y, Qu B, et al. 2024. dCas9-guided demethylation of the AKT1 promoter improves milk protein synthesis in a bovine mastitis mammary gland epithelial model induced by using *Staphylococcus aureus*. *Cell Biol. Int.* 48(3):300–10
190. Dehler CE, Lester K, Della Pelle G, Jouneau L, Houel A, et al. 2019. Viral resistance and IFN signaling in STAT2 knockout fish cells. *J. Immunol.* 203(2):465–75
191. Verdile N, Camin F, Pavlovic R, Pasquariello R, Stuknyte M, et al. 2023. Distinct organotypic platforms modulate rainbow trout (*Oncorhynchus mykiss*) intestinal cell differentiation in vitro. *Cells* 12(14):1843
192. Clark EL, Archibald AL, Daetwyler HD, Groenen MAM, Harrison PW, et al. 2020. From FAANG to fork: application of highly annotated genomes to improve farmed animal production. *Genome Biol.* 21:285
193. Lampe GD, King RT, Halpin-Healy TS, Klompe SE, Hogan MI, et al. 2024. Targeted DNA integration in human cells without double-strand breaks using CRISPR-associated transposases. *Nat. Biotechnol.* 42(1):87–98
194. Thompson D, Regev A, Roy S. 2015. Comparative analysis of gene regulatory networks: from network reconstruction to evolution. *Annu. Rev. Cell Dev. Biol.* 31:399–428
195. Johnston I, Kent M, Pierre B, Looseley M, Bargelloni L, et al. 2024. Advancing fish breeding in aquaculture through genome functional annotation. *Aquaculture* 583:740589
196. Hum. Cell Atlas Stand. Technol. Work. Group, Rozenblatt-Rosen O, Shin JW, Rood JE, Hupalowska A, et al. 2021. Building a high-quality Human Cell Atlas. *Nat. Biotechnol.* 39(2):149–53
197. Kapoor M, Tuggle CK, Burdett T, Tickle T, Harrison P, et al. 2023. PSII-6 computational tools and resources for analysis and exploration of single-cell Rnaseq data in agriculture. *J. Anim. Sci.* 101(Suppl. 2):267–68
198. Tekman M, Batut B, Ostrovsky A, Antoniewski C, Clements D, et al. 2020. A single-cell RNA-sequencing training and analysis suite using the Galaxy framework. *GigaScience* 9(10):giaa102

