bioRxiv, 2025, 1–17
doi:

From Understudied to Understood: The MiniENCODE Framework for Multi-Omics Analysis in Diverse Species

Hang Yang ,[1,2] Zhaojun Shan ,[1,2] Hanqiao Shang ,[1] Puxuan Jiang ,[1,2] Yingshu Li [1,*] and Qiang Tu [1,2,*]

[1] State Key Laboratory of Molecular Developmental Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, 100101, Beijing, China and [2] University of Chinese Academy of Sciences, 100049, Beijing, China

*Corresponding author. qtu@genetics.ac.cn, ysli@genetics.ac.cn

FOR PUBLISHER ONLY Received on Date Month Year; revised on Date Month Year; accepted on Date Month Year

Abstract

The ENCODE project has revolutionized our understanding of functional genomic elements, yet its exhaustive approach remains inaccessible to smaller communities studying non-model organisms. To address this challenge, we developed the miniENCODE framework - a resource-efficient strategy combining carefully selected experimental assays with an integrated computational platform. Using zebrafish as a model, we implemented this framework through three core assays: RNA-seq, ATAC-seq, and H3K27ac CUT&Tag, spanning six developmental stages and 11 adult tissues. Our analysis identified 52,350 candidate enhancer-like signatures, characterized their spatiotemporal activity patterns, and experimentally validated tissue-specific enhancers. We developed the mini Omics Data Portal (miniODP) to facilitate multi-omics data re-use, integration, visualization, and analysis. Through this platform, we characterized key transcription factors and their regulatory networks in various developmental stages and adult tissues. Extending this approach to additional model organisms (cattle, lancelet, Mexican tetra, and polyp), we demonstrated its broad applicability for understanding gene regulation in diverse species. The miniENCODE framework provides a practical solution for comprehensive regulatory element characterization, enabling smaller research communities to advance genomic studies in understudied organisms without overextending their resources. This approach not only provides a valuable resource for zebrafish but also establishes a scalable model for accelerating genomic research in diverse species.

Introduction

The Encyclopedia of DNA Elements (ENCODE) project has revolutionized our understanding of functional genomic elements through exhaustive mapping of regulatory landscapes in humans and mice (1, 2). While analogous efforts through modENCODE have extended these analyses to *C. elegans* and *D. melanogaster* (3, 4), these efforts represent only a fraction of model organisms used in biological research. A significant gap remains in understanding genomic regulation in most species. This gap is evident in the GEO database, where about 80% of samples derive from humans and mice, while only 9% come from the next eight most represented organisms, and a mere 11% from over 6,000 other species (Supplementary Figure 1). This imbalance creates substantial challenges for researchers working with non-ENCODE model organisms, who must navigate limited resources while attempting to obtain extensive information about genomic regulation. Moreover, experimental biologists face significant challenges in accessing and interpreting published omics data, as the lack of centralized data portals complicates data discovery, integration, and visualization for their model organisms.

To address these challenges, we propose the miniENCODE framework, comprising two integrated components: (1) a set of carefully selected experimental assays designed to capture the most essential regulatory information, and (2) the mini Omics Data Portal (miniODP), a computational platform for data integration and analysis. This framework focuses on generating core omics data through cost-effective technologies while establishing standardized analysis pipelines and intuitive visualization tools to facilitate data reuse and sharing. By balancing comprehensiveness with accessibility, this approach

© The Author 2025. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

enables effective genomic analysis for research communities with limited resources.

The zebrafish (*Danio rerio*) serves as an ideal test case for this approach. As a well-established model organism in developmental biology, toxicology, and disease research (5), zebrafish offers the additional advantage of being a vertebrate with significant evolutionary conservation of regulatory mechanisms while maintaining experimental tractability. Zebrafish benefits from a central data hub, the Zebrafish Information Network (ZFIN), which provides expertly curated genetic and genomic resources including genes, functions, phenotypes, lines, nomenclature, and reagents (6). Recent advances have significantly expanded these resources. Yang et al. generated a thorough epigenomic map of the zebrafish genome, identifying both conserved and species-specific regulatory sequences (7). The DANIO-CODE consortium established a data coordination center, identifying over 140,000 zebrafish *cis*-regulatory elements and enhancing functional annotation to link zebrafish and mammalian genomics (8, 9). Multiple single-cell atlases have emerged, including DanioCell's high-resolution transcriptional atlas (10), ZSCAPE's extensive dataset of 3.2 million cells (11), and ZebraHub's multimodal developmental atlas (12). Additionally, Sun et al. mapped chromatin accessibility throughout 20 developmental stages, identifying nearly one million candidate *cis*-regulatory elements (13). These comprehensive resources, coupled with the extensive existing knowledge base, make zebrafish an ideal system for evaluating whether a focused set of omics assays can provide meaningful insights into genome regulation.

In this study, we implemented the miniENCODE framework using zebrafish as a model system, integrating both published datasets and our newly generated data from three core assays: RNA-seq for gene expression profiling, ATAC-seq for chromatin accessibility mapping, and CUT&Tag for H3K27ac profiling. These datasets span six developmental stages and 11 adult tissues. We developed the mini Omics Data Portal (miniODP) to enable multi-omics data integration, analysis, and visualization. Through this integrated approach, we identified 52,350 candidate enhancer-like signatures (ELSs) and characterized key regulators and gene regulatory networks (GRNs) across developmental stages and adult tissues. Experimental validation and comparison with existing literature confirmed the miniENCODE approach as an efficient and practical framework for identifying regulatory elements and investigating gene regulatory mechanisms. To evaluate broader applicability, we extended our analysis to additional model organisms (polyp, lancelet, Mexican tetra, and cattle) using published datasets. These analyses demonstrate that the miniENCODE framework can generate valuable insights into genome regulation in diverse organisms, even with limited omics data. This approach provides a systematic method for understanding understudied genomes, equipping smaller research communities with tools to decode regulatory landscapes and contribute to fundamental discoveries in evolutionary biology, developmental genetics, and comparative genomics.

Results

Experimental Assays for the MiniENCODE Framework

The ENCODE project has established an all-encompassing suite of technologies to characterize functional DNA elements, encompassing DNA-protein interactions, transcription profiling, chromatin accessibility, DNA methylation, RNA binding, 3D chromatin architecture, and replication timing (2). While this extensive toolkit has proven invaluable for well-resourced projects, its implementation presents significant challenges for smaller research communities due to technical complexity, high costs, specialized equipment requirements, and substantial infrastructure needs. These factors make extensive ENCODE-style profiling impractical for most laboratories. Additionally, selecting appropriate technologies for initial genomic characterization remains a daunting task, especially for researchers with limited experience in omics technologies.

To address these issues, we propose using a carefully selected set of core and optional assays that maximize biological insights while minimizing resource requirements. We refer to this set of technologies as miniENCODE assays (Figure 1A), which comprises three core components: (1) Short-read RNA-seq for gene expression profiling; (2) ATAC-seq for genome-wide mapping of accessible chromatin regions; and (3) CUT&Tag targeting H3K27ac for precise identification of active regulatory elements. These core technologies were chosen based on their cost-effectiveness, technical accessibility, and complementary nature in capturing distinct aspects of genome activities. To address more specialized research needs, we also propose three optional assays: (4) Long-read RNA-seq for accurate transcriptome assembly; (5) Micro-C for high-resolution chromatin architecture mapping; and (6) Single-cell multi-omics (RNA-seq, ATAC-seq, and CUT&Tag) for cell-type-specific profiling. This tiered approach allows researchers to balance their experimental scope with available resources while maintaining flexibility for future expansion.

The core miniENCODE assays - short-read RNA-seq, ATAC-seq, and CUT&Tag - provide complementary insights into genome regulation. RNA-seq offers gene expression profiles, while ATAC-seq identifies chromatin-accessible regions that often contain functional elements like enhancers and promoters. We implemented CUT&Tag rather than traditional ChIP-seq due to its superior resolution, high signal-to-noise ratio, and reduced sequencing requirements (15). For histone modification profiling, we focused on H3K27ac, a well-established marker of active regulatory elements, particularly enhancers and promoters (2). Compared to H3K4me3, which primarily marks active promoters, H3K27ac offers greater utility in identifying both active enhancers and promoters, making it a top choice for regulatory element mapping. By integrating chromatin accessibility data from ATAC-seq with H3K27ac profiling via CUT&Tag, we achieve a robust approach for identifying candidate regulatory elements, which is particularly advantageous for resource-limited research settings.

For researchers requiring deeper insights, we propose three optional technologies that provide enhanced resolution, though with increased technical complexity: (1) Long-read RNA-seq for accurate transcriptome assembly, particularly valuable for newly sequenced genomes where precise gene models are essential for downstream analyses (16); (2) Micro-C, which offers superior resolution over traditional Hi-C methods for high-resolution chromatin architecture mapping and investigation of long-range regulatory interactions (17, 18); and (3) Single-cell multi-omics approaches (RNA-seq, ATAC-seq, and CUT&Tag) for cell-type-specific profiling, particularly valuable for understanding complex tissues (19–22). Emerging technologies such as spatial omics may further complement these approaches in the future.

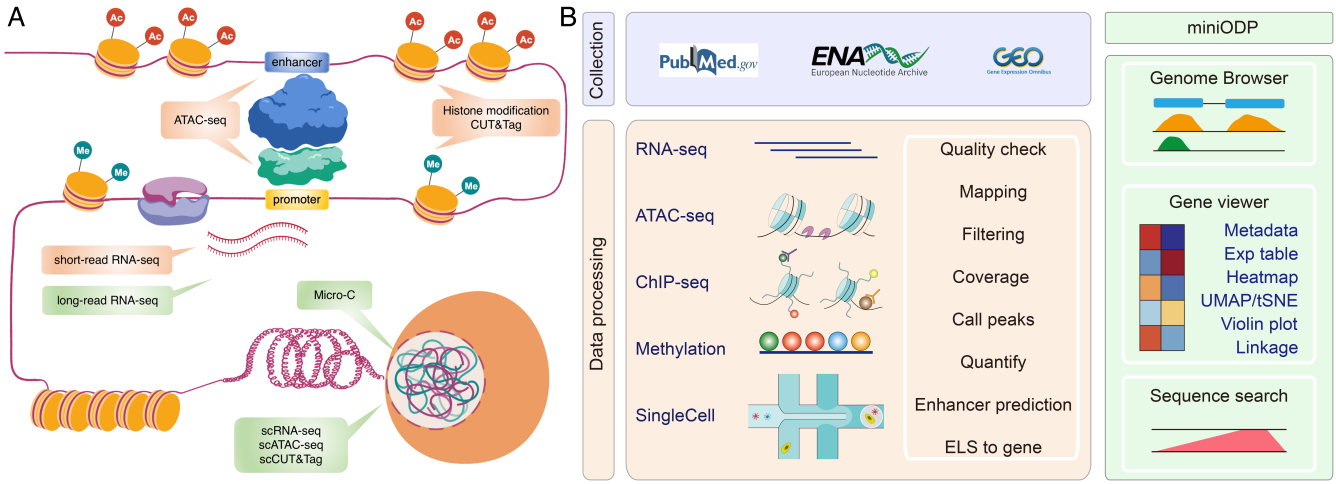

Figure 1: Overview of miniENCODE assays and the miniODP database. (A) Schema showing the technologies used in miniENCODE assays, with essential technologies in salmon boxes and optional technologies in green boxes. The illustration was partially generated using biogdp.com (14). (B) Architecture of miniODP database, which integrates multi-omics data from public sources, processes it through a standardized pipeline, and generates derived datasets including ELS predictions. The miniODP provides an intuitive web interface for data visualization and analysis, enabling comprehensive exploration of regulatory elements.

Computational Platform for the MiniENCODE Framework

The exponential growth of omics data offers unprecedented opportunities while posing substantial challenges. Although researchers working with established model organisms such as humans and mice can leverage specialized databases, investigators studying less common species encounter significant obstacles in data reuse and integration. To bridge this gap, we developed the mini Omics Data Portal (miniODP), an integrated platform for multi-omics data analysis, visualization, and interpretation (Figure 1B, `https://tulab.genetics.ac.cn/miniodp/`).

MiniODP provides three core modules for accessing and analyzing diverse data types. The first module features a JBrowse-based genome browser (23), offering an intuitive interface with advanced track selection capabilities for visualizing multi-omics data in various technologies and biological samples (Figure 2A-B). This module enables rapid navigation to genomic regions of interest through gene name- or coordinate-based searches, complemented by a track filtering system that allows users to organize and display hundreds of data tracks based on multiple parameters, including experimental category, investigator, sample type, data source, target molecule, technique, and data type.

The second module is a gene browser that provides detailed information about specific genes, including functional annotations, quantitative expression tables, and interactive visualizations. Built using R Shiny and ShinyProxy for the front-end with an SQLite database backend, this module enables users to search for genes using multiple criteria: Ensembl ID, gene name, GO annotations, InterPro protein domains, transcription factor (TF) families, and orthologs (Figure 2C). The results are seamlessly integrated with external resources through direct links to Ensembl (24), AnimalTFDB (25), and InterPro (26) databases. Users can explore quantitative expression data through multiple visualization formats, including heatmaps and line plots (Figure 2D). For single-cell datasets, the module supports advanced representations such as UMAP, t-SNE, violin plots, and heatmaps (Figure 2E). Additionally, it provides a detailed multi-omics view that integrates gene expression, chromatin accessibility, and ELS-to-gene linkages around the queried locus (Figure 2F).

The third module is a sequence search tool implemented using SequenceServer (27). This module enables users to perform sequence searches against multiple genomic resources, including whole genomes, cDNA collections, and reference proteomes (Figure 2G). A key feature of this module is the direct integration with the genome browser, where search hits are automatically linked to their corresponding genomic locations. This connection facilitates the biological interpretation of sequence matches within their proper genomic context, enabling researchers to quickly assess potential functional implications (Figure 2H).

Building the Zebrafish miniENCODE Resource

We implemented the miniENCODE framework in zebrafish as a proof of concept, demonstrating its utility for extensive regulatory element characterization through a curated set of omics technologies integrated with a unified database platform, thereby establishing a model for applying the miniENCODE approach to other model organisms.

To establish an exhaustive regulatory landscape, we systematically collected and curated zebrafish omics datasets from major public repositories, including GEO (28) and ENA (29). Each dataset underwent rigorous quality assessment and manual curation, with evaluation of metadata including experimental design, biological replicates, sequencing depth, and quality metrics. Our curation focused on six key data types: RNA-seq, ATAC-seq, ChIP-seq (including CUT&Tag), DNA methylation, scRNA-seq, and scATAC-seq. Selection criteria prioritized datasets with: (1) high-quality sequencing metrics; (2) biological replicates; (3) metadata; and (4) coverage of key developmental stages or tissues.

We organized these datasets into a data type matrix (Figure 3), classifying samples spanning eight developmental stages and 20 embryonic and adult tissues. For ChIP-seq/CUT&Tag datasets, we further classified them by target type, distinguishing between

Figure 2: Functional modules of the miniODP database. (A, B) The genome browser provides an interactive interface for visualizing multi-omics data tracks in various biological samples and technologies, enabling exploration of regulatory landscapes at specific genomic loci. (C-F) The gene browser combines diverse data types, including gene annotations, expression profiles, and regulatory element information, supporting in-depth gene regulation analysis through multiple visualization formats (heatmaps, line plots, UMAP, and t-SNE). (G, H) The BLAST server facilitates sequence similarity searches against genomic resources, with results directly linked to the genome browser for contextual interpretation of sequence matches within their genomic and regulatory context.

histone modifications (with primary focus on H3K27ac) and TFs. This structured organization enabled clear identification of both existing strengths and critical gaps in the current zebrafish regulatory data landscape.

For our final computational analysis, we prioritized achieving complete coverage of the three core miniENCODE assays (RNA-seq, ATAC-seq, and H3K27ac CUT&Tag) across six key developmental stages (from Blastula to Larval) and 11 major

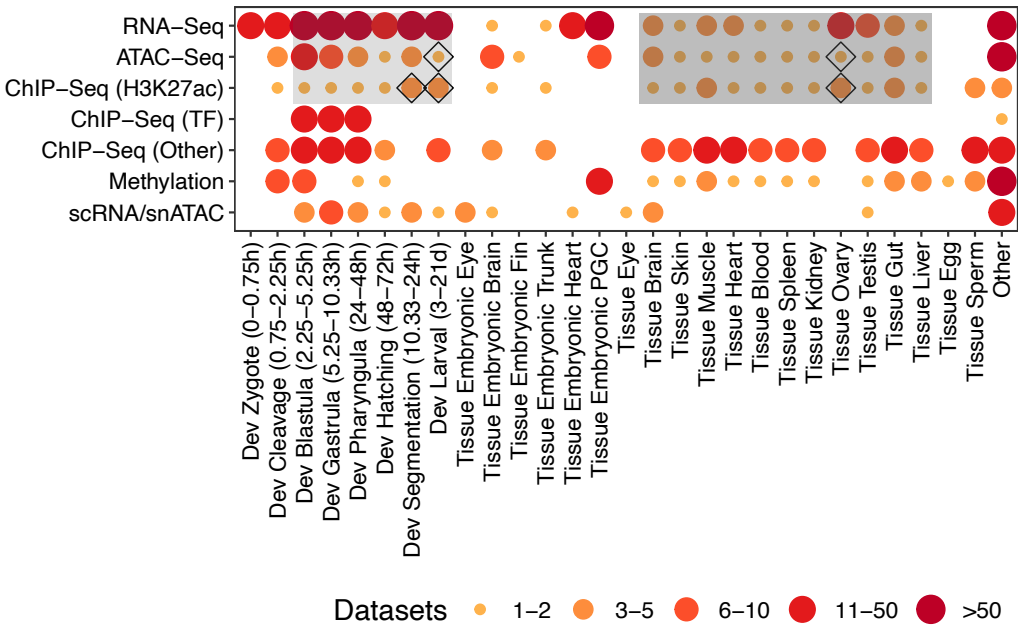

Figure 3: Data coverage of zebrafish in the miniODP database. The data type matrix illustrates the distribution of omics datasets in different biological samples and technologies. Each cell represents a specific combination of biosample category and technology, with both dot size and color intensity representing the number of datasets. Light and dark grey boxes highlight the essential miniENCODE assays (RNA-seq, ATAC-seq, and H3K27ac CUT&Tag) for developmental stages and adult tissues, respectively. ChIP-seq (Other) includes histone modifications beyond H3K27ac, such as H3K4me3, H3K27me3, H3K9me3, and H3K14ac. Diamond symbols indicate datasets generated specifically for this study, filling critical gaps in the existing data landscape.

adult tissues (including brain, heart, muscle, liver, gut, spleen, ovary, and testis). Although existing datasets offered significant coverage, we addressed critical gaps by generating new ATAC-seq and H3K27ac CUT&Tag datasets for the segmentation stage, larval stage, and ovary tissue (indicated by diamonds in Figure 3; see Supplementary Figures 2-3). This initiative successfully completed our primary objective matrix (3 assays × 17 biosamples), representing an extensive regulatory dataset currently available for zebrafish.

Notably, even in this extensively studied model organism, we identified substantial gaps in core regulatory data, particularly regarding H3K27ac modification profiles. Specifically, H3K27ac datasets were completely lacking for the segmentation stage and ovary tissue prior to this study. Our targeted data generation efforts filled these critical gaps, underscoring the persistent requirement for systematic regulatory element characterization in even well-established model systems.

The integrated multi-omics dataset facilitated systematic analysis through a three-step approach. First, we identified ELSs by combining ATAC-seq and H3K27ac ChIP-seq data. Next, we established ELS-gene regulatory relationships using RNA-seq expression profiles. Finally, we integrated TF binding motifs to identify key regulatory factors and their networks in various developmental stages and tissues.

Enhancer-like Signature Identification and Characterization

Enhancers are fundamental regulatory elements that precisely control the spatiotemporal expression of genes during development and tissue homeostasis. These elements are characterized by open chromatin architecture, specific histone modifications, and binding of cell type-specific TFs (30). These molecular signatures enable systematic genome-wide identification of enhancers. ATAC-seq, which employs a hyperactive Tn5 transposase to tag accessible chromatin regions, provides a sensitive method for detecting enhancer activity through their characteristic nuclease hypersensitivity (31). Furthermore, active enhancers are marked by specific histone modifications, particularly H3K27ac, which can be precisely mapped using ChIP-seq or the more recent CUT&Tag techniques (15, 32).

Using our collection of omics datasets covering three core miniENCODE assays in six developmental stages and 11 tissues, we predicted candidate enhancers following the computational framework established by the ENCODE project (2). Our analysis pipeline first identified representative DNA accessibility regions (rDARs) from ATAC-seq peaks, followed by calculation of Z-scores for both ATAC-seq and H3K27ac signals. We defined ELSs as rDARs exceeding the 95th percentile (Z-score 1.64) for both signals, classifying them as TSS-proximal ELSs (pELS, 0-2 kb) or TSS-distal ELSs (dELS, >2 kb). This approach identified 52,350 ELSs, representing 1.2% of the genome and 17% of ATAC-seq peaks, with 37% pELSs and 63% dELSs. Importantly, 20% of these ELSs were novel compared to previous studies (7, 9)(Supplementary Figure 4), and all predicted ELSs are accessible through our genome browser (Figure 4A).

Cell type-specific TFs bind enhancers to precisely regulate spatiotemporal gene expression (33), leading to cell-type-specific activity of many enhancers. Among the 52,350 identified ELSs, 51% (26,695) demonstrated stage- or tissue-specific activity.

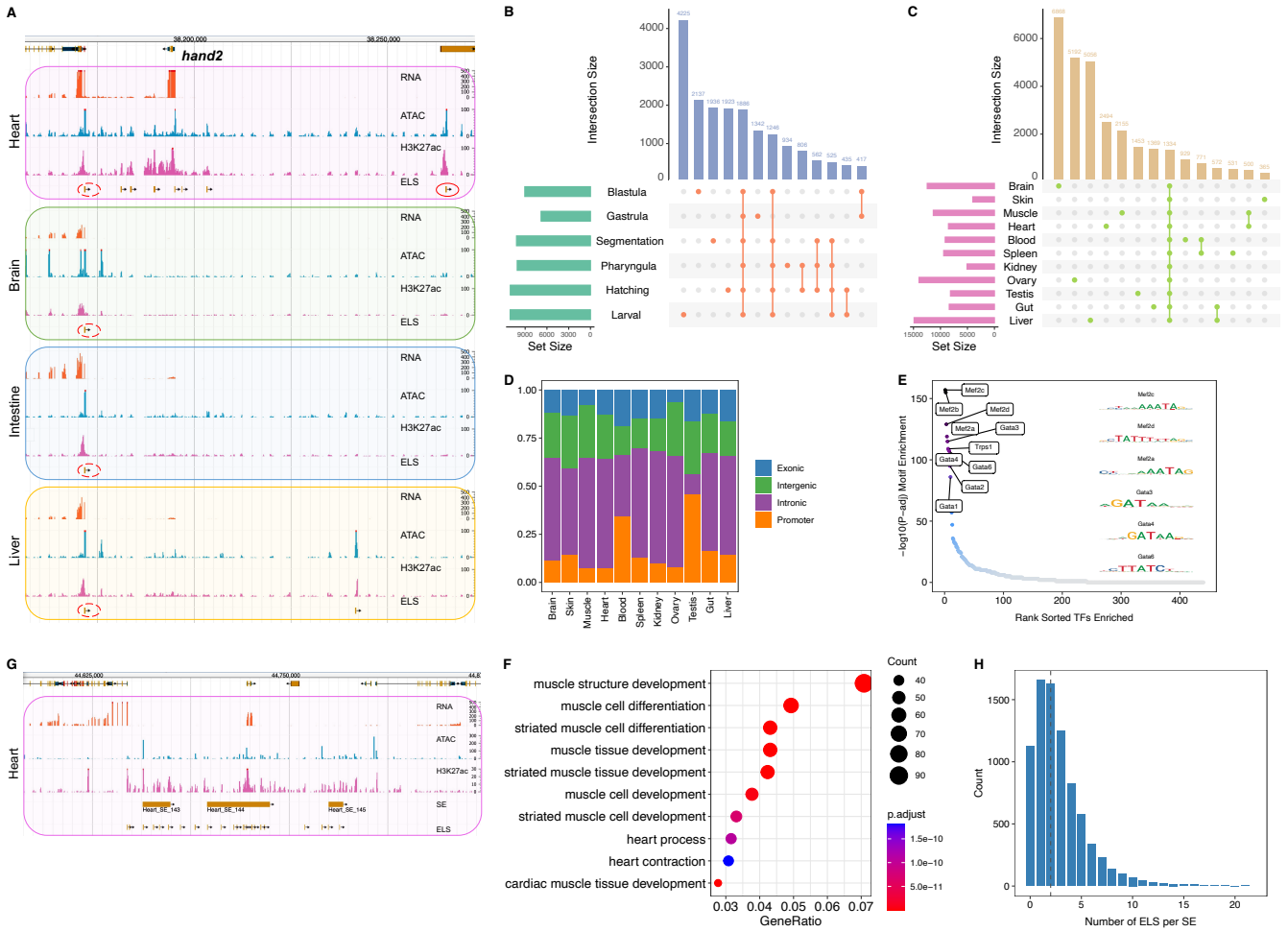

Figure 4: Genome-wide identification and characterization of ELSs in zebrafish. (A) Multi-omics visualization of ELSs in four representative tissues (heart, brain, intestine, and liver). The genomic view displays gene annotations followed by tissue-specific profiles. Each tissue panel shows four data tracks: RNA-seq (coral peaks), ATAC-seq (sky blue peaks), H3K27ac ChIP-seq (light pink), and predicted ELSs (mustard yellow blocks). ELS identification requires co-occurrence of both ATAC-seq and H3K27ac signals. The left ellipse highlights a ubiquitously active ELS present across all tissues, while the right ellipse shows a heart-specific ELS experimentally validated in this study. (B, C) Upset plots illustrating ELS activation patterns across developmental stages (B) and tissues (C). Horizontal bars represent individual stages/tissues, with bar length indicating the number of active ELSs. Vertical bars show intersection sizes between sets. The larval stage contains the highest number of stage-specific ELSs, while brain tissue shows the most tissue-specific ELSs. We identified 1,886 developmentally persistent and 1,334 tissue-ubiquitous ELSs. (D) Genomic distribution of ELSs, showing predominant localization in intronic (51%) and intergenic (23%) regions. (E) TF binding motif enrichment analysis for heart-specific ELSs. TF motifs are ranked by the significance of their enrichment. The top 10 TFs are labeled, with motifs of six known cardiac regulators highlighted. (F) Gene Ontology analysis of genes associated with heart-specific ELSs, revealing enrichment for cardiac-related biological processes. (G) Genomic architecture of super-enhancers, characterized by clusters of ELSs with strong ATAC-seq and H3K27ac signals. (H) Distribution of ELSs per super-enhancer, showing the number of constituent ELSs in each super-enhancer cluster.

Figure 4A illustrates both a ubiquitously active ELS and the experimentally validated heart-specific ELS22, both exhibiting strong ATAC-seq and H3K27ac signals. The larval stage contained the highest number of stage-specific ELSs (Figure 4B), consistent with its cellular diversity and chromatin dynamics, while the brain, ovary, and liver displayed the greatest numbers of tissue-specific ELSs (Figure 4C), reflecting their regulatory complexity.

Intriguingly, we observed substantial ELS overlap across developmental stages (evidenced by multiple vertical lines connecting dots in Figure 4B) but minimal overlap between distinct tissues (Figure 4C). This pattern likely reflects greater gene expression similarity in temporally contiguous stages, required for coordinated cell differentiation, compared to spatially discrete organs with specialized functions. Shared ELSs were found in organs with common developmental origins and related functions: mesoderm-derived contractile muscles and hearts; lateral plate mesoderm-derived blood cells and spleen involved in immunity (34); and endoderm-derived gut and liver essential for digestion and metabolism (35). These findings underscore the

critical role of tissue-specific gene regulation in maintaining organ functionality.

Genomic localization analysis revealed distinct patterns in the distribution of tissue-specific ELSs, with the majority residing in intronic (51%) or intergenic regions (23%) (Figure 4D and Supplementary Figure 5). This distribution aligns with findings from a recent mouse study (36), suggesting evolutionary conservation of enhancer genomic positioning. However, testis-specific ELSs displayed a unique distribution pattern, showing the lowest proportion of intronic locations and the highest enrichment in promoter regions, potentially reflecting the specialized transcriptional regulation required during spermatogenesis.

Furthermore, we identified 1886 developmentally persistent, 1334 tissue-ubiquitous, and 702 universally active ELSs (Supplementary Table 1). These ELSs constitute a valuable resource for zebrafish transgenesis. Stage- or tissue-specific ELSs enable precise spatiotemporal transgene expression, facilitating gene function studies. The constitutively active ELSs represent optimal targets for knock-in experiments, providing an advantage over the random integration of the *Tol2* transposon system (37), which can result in transgene silencing in inactive genomic regions. Together, these findings provide an extensive catalog of zebrafish regulatory elements and their functional characteristics across development and adult tissues.

Functional Annotation of Zebrafish ELSs

To identify transcriptional regulators governing tissue-specific gene expression, we performed motif enrichment analysis on tissue-specific ELSs. By ranking TFs based on their -$\log_{10}$(P-value) of motif enrichment in different tissues, we identified key TFs associated with each tissue's regulatory landscape. In heart tissue (Figure 4E), six of the top ten enriched TFs—Mef2c, Mef2d, Mef2a, Gata3, Gata6, and Gata4—are well-established regulators of cardiac development and function. Similar patterns were observed in other tissues, where highly enriched TFs corresponded to known functional regulators (Supplementary Figure 6). These results demonstrate that TF binding motif analysis of tissue-specific ELSs effectively identifies principal transcriptional drivers of tissue-specific gene expression programs.

Gene ontology (GO) analysis revealed that genes adjacent to tissue-specific ELSs were enriched for biological processes relevant to their respective tissues. Heart-specific ELS-associated genes showed enrichment for cardiac-related processes, including heart contraction and muscle development (Figure 4F). Similarly, brain-specific ELSs were associated with neuronal differentiation and axon development, liver-specific ELSs with metabolic processes, and testis-specific ELSs with cilium organization and cell cycle regulation (Supplementary Figure 7). These results indicate that tissue-specific ELSs are often located near genes crucial for tissue function and development, and interact with corresponding tissue-specific TFs, underscoring the value of ELS analysis in understanding tissue-specific regulatory mechanisms.

Super-enhancers (SEs), defined as clusters of enhancers that interact with TFs, cofactors, and histone modifications, play crucial roles in regulating gene expression and influencing cell fate and disease (38, 39). Using H3K27ac data and the ROSE algorithm (38) (Supplementary Figure 8), we identified approximately 470 SEs per embryonic stage and adult tissue (Supplementary Figure 9A). These SEs span large genomic regions characterized by strong ATAC-seq and H3K27ac signals (Figure 4G), indicating high regulatory activity. SEs had a median length of 28,344 bp (Supplementary Figure 9B), compared to 297 bp for ELSs, with each SE containing an average of two ELSs; 86% included at least one ELS (Figure 4H). GO analysis confirmed tissue-specific functions, such as heart morphogenesis, vasculogenesis, and cardiocyte differentiation for heart SEs (Supplementary Figure 8B). All SEs are accessible through our genome browser.

Multi-omics Integrative Analysis of Regulatory Networks

Gene regulation involves coordinated interactions among chromatin structure, TFs, regulatory elements, and target genes. These interactions frequently depend on spatial proximity rather than linear genomic distance, reflecting the complex three-dimensional organization of the genome. To systematically map these regulatory relationships, we developed an integrated multi-omics approach combining ELS identification, gene linkage prediction, and GRN construction. This strategy enables us to bridge the gap between regulatory elements and their functional consequences in diverse biological contexts.

To establish ELS-gene regulatory relationships, we developed a correlation-based approach linking chromatin accessibility to gene expression. Building upon single-cell analysis methods (40), we associated 52,350 predicted ELSs (identified through ATAC-seq and H3K27ac ChIP-seq) with 26,354 target genes by correlating ELS accessibility profiles (derived from ATAC-seq peak matrices) with corresponding gene expression patterns (from RNA-seq matrices). These relationships were quantified using linkage strength scores ranging from -1 to 1, which are visualized in miniODP's multi-omics integration plots (Figure 5A).

Our in vivo validation (described in the next section) confirmed ELS22 as a heart-specific enhancer (Figure 6), validating our approach. Located 70 kb downstream of *hand2* and specifically accessible in heart tissue, ELS22 was associated with *hand2* based on their positive correlation in accessibility and expression (Figure 5A). *hand2* plays essential roles in heart development and regeneration (41). Interestingly, ELS22 is situated within the intron of *BX000460*, which shows weak expression in ovary and testis but not in heart. This observation emphasizes the necessity of multi-omics evidence, beyond mere genomic location, for accurately establishing ELS-gene regulatory relationships.

Building on these ELS-gene linkages, we constructed GRNs for developmental stages and tissues using the ANANSE algorithm (42). First, we built a zebrafish TF binding motif database using GimmeMotifs (43), then calculated TF binding probabilities based on motif scores, ATAC-seq, and H3K27ac signals in ELSs. Next, we predicted TF-gene binding scores by integrating enhancer intensity and motif scores within 100 kb of each TSS. Finally, we combined binding scores, TF activity, and expression levels (both TF and target genes) into interaction scores (0-1 scale) to form the GRNs.

To evaluate the quality of zebrafish GRNs, we ranked TFs in each network by outdegree (number of target genes) and identified the top 10 TFs per sample. The majority represented well-characterized regulators for their respective tissues. In the blood GRN, 80% (8/10) were established hematopoiesis factors, including *hif1al*, *gata1a*, *hif1ab*, *tal1*, *crema*, *nr1d2a*, *nr1d2b*, and *nr3c1* (Figure 5B). In the brain GRN, 60% (6/10) were recognized neural regulators, such as *npas4a*, *zic1*, *pbx3b*, *gata3*, *klf6a*, and *jun* (Figure 5C).

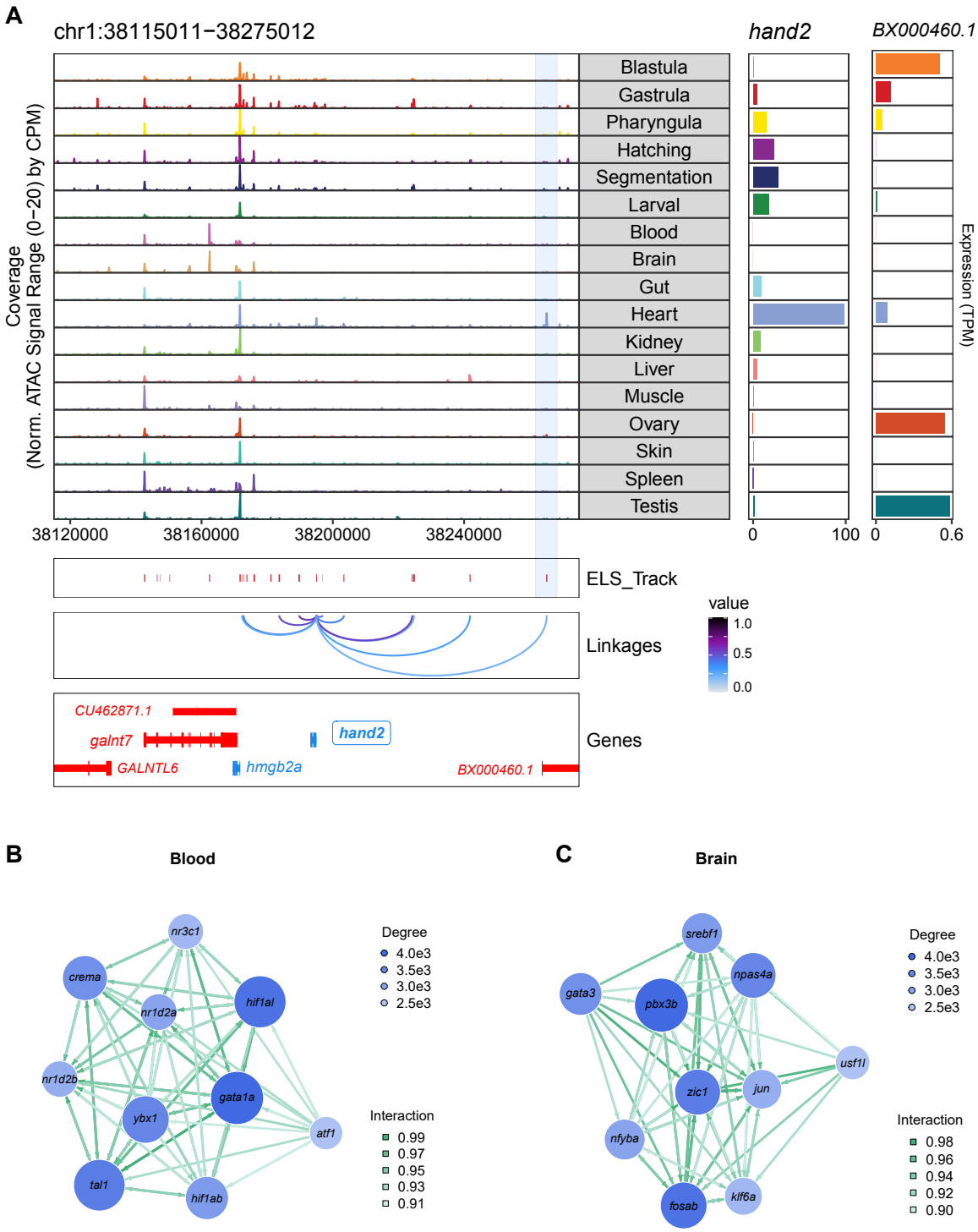

Figure 5: Multi-omics analysis of gene regulatory networks (GRNs) in zebrafish. (A) Multi-omics visualization of *hand2* regulatory elements. The central panel shows ATAC-seq tracks across samples, revealing chromatin accessibility patterns. The right panel displays RNA-seq-derived gene expression levels. The bottom panel depicts predicted ELSs, ELS-gene linkages, and associated genes. *hand2* shows heart-specific expression, with a corresponding ELS (ELS22, highlighted in blue) located 70 kb downstream. This ELS exhibits heart-specific chromatin accessibility, indicating a functional regulatory relationship. (B, C) Tissue-specific enhancer-gene regulatory sub-networks for blood (B) and brain (C) tissues. Blue circles denote the top 10 TFs in each tissue, ranked by outdegree (number of target genes) within the top 100,000 edges. Circle size and color intensity reflect the number of target genes regulated by each TF. Blue arrow shades represent interaction strength between TF pairs, with darker shades indicating stronger regulatory interactions. Most top-ranked TFs are known key regulators in their respective tissues, confirming our network analysis accuracy.

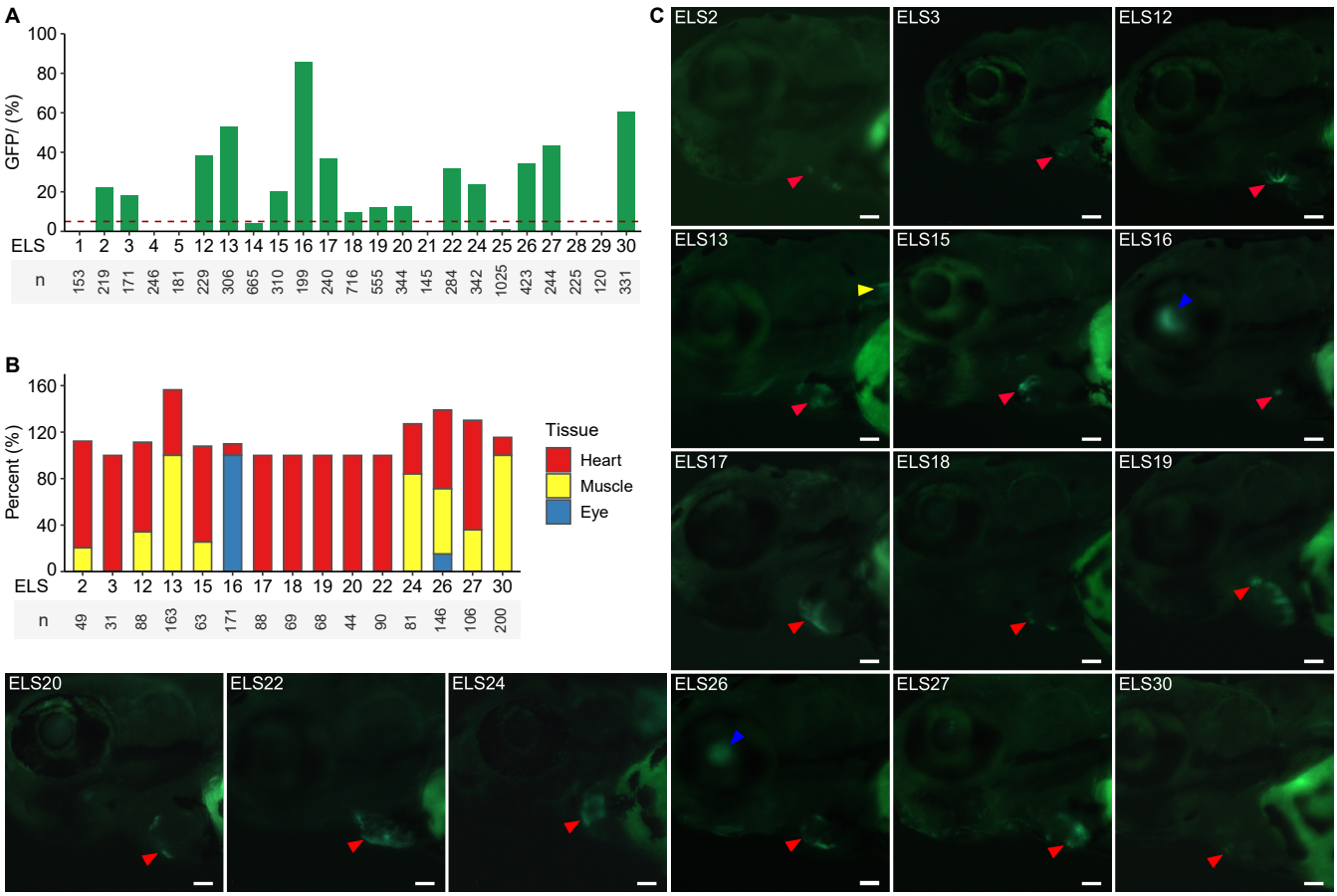

Figure 6: In vivo validation of heart-specific ELSs using GFP reporter assays in zebrafish. (A) Validation rate of ELSs, showing the percentage of GFP-positive embryos among total observed embryos. The dashed line indicates the 5% threshold, with 15 out of 23 tested ELSs driving GFP expression above this level. (B) Tissue specificity of GFP expression, showing the percentage of GFP signals observed in different organs among GFP-positive embryos. GFP expression was predominantly detected in the heart, with additional signals observed in muscles and eyes. (C) Representative images of GFP expression patterns driven by validated ELSs. Arrows indicate GFP signals corresponding to the color-coded tissues in panel B. Scale bar, 50 µm.

Notably, key TFs for blood development, including *hif1ab* (44) and *nr3c1* (45), exhibited broad expression across multiple tissues (Supplementary Figure 10) but displayed more extensive gene interactions specifically in blood (Figure 5B). Similarly, the brain development TF *jun* (46) was expressed in various tissues but showed stronger gene regulatory connections in the brain (Figure 5C). These observations highlight that TFs with extensive regulatory network interactions play pivotal roles in tissue-specific functions, underscoring the value of multi-omics integration over single-method approaches.

Our integrated approach, combining chromatin accessibility, histone modifications, and gene expression data, provides a powerful framework for unraveling complex regulatory networks. This multi-omics strategy offers significant advantages over single-method approaches, enabling more accurate identification of key regulatory elements and their functional relationships across different biological contexts.

In Vivo Functional Validation of Predicted ELSs

To validate the predicted ELSs in vivo, we performed reporter assays in zebrafish, focusing specifically on cardiac enhancers. From the 8,525 heart-activated ELSs identified through our multi-omics analysis, we selected 23 candidates from the 2,494 heart-specific ELSs (see Supplementary Table 2) for experimental validation. These ELSs were cloned into a GFP reporter vector and microinjected into zebrafish embryos. At 3 days post-fertilization (dpf), we observed GFP expression patterns, finding that 65% (15/23) of the tested ELSs—including ELS22, described in the previous section—drove specific GFP expression in the heart (Figure 6). This demonstrates their functional role as enhancers during zebrafish heart development. The remaining eight ELSs failed to show GFP activity, potentially due to temporal differences between our multi-omics data collection (spanning various developmental stages) and the single time point (3 dpf) used for reporter analysis. The high validation rate (65%) underscores the predictive power of our integrative omics approach in identifying functional regulatory elements and elucidating tissue-specific gene regulation mechanisms.

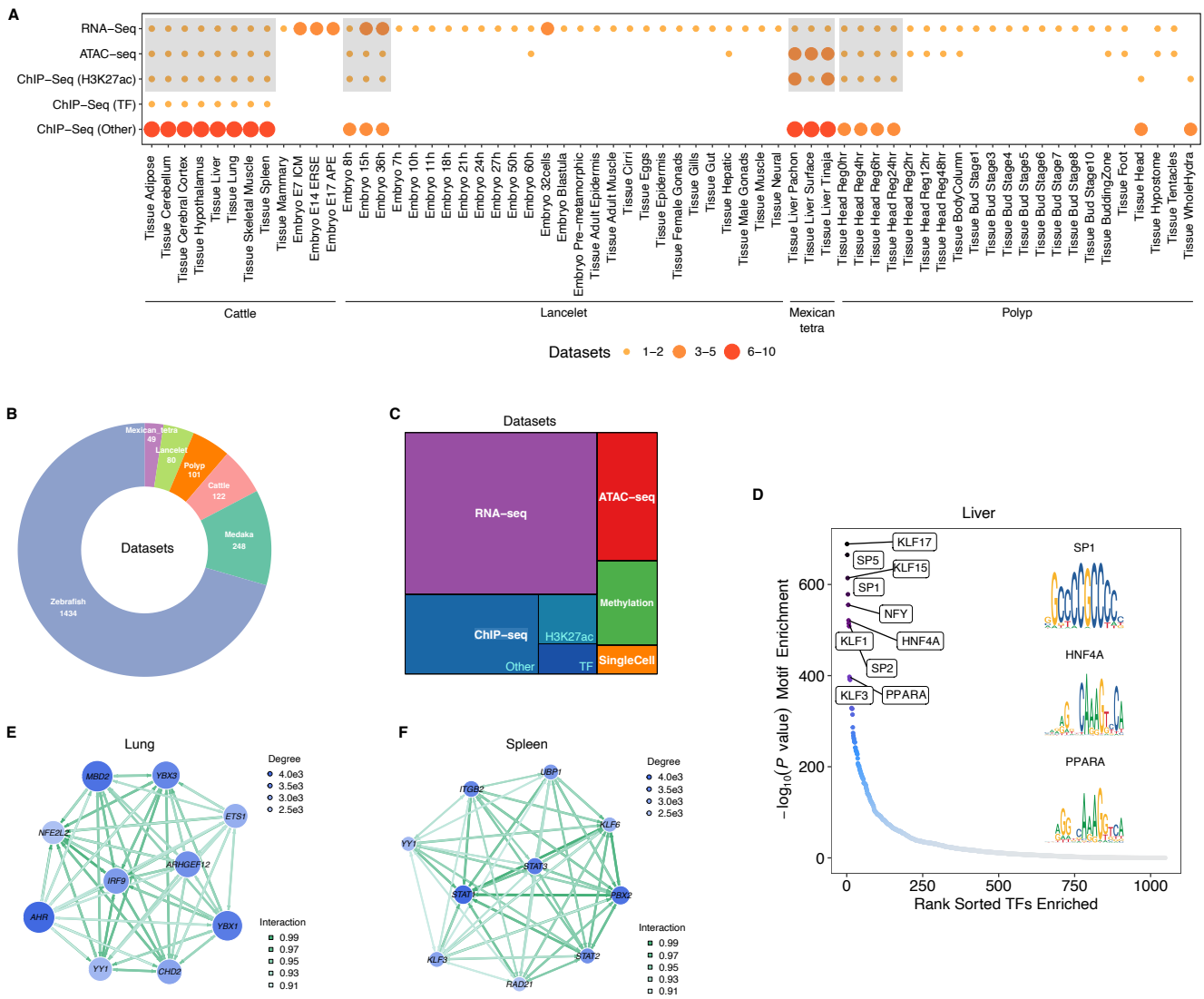

Figure 7: Multi-organism implementation in the miniODP database. (A) Data type matrix showing the distribution of omics datasets in four model organisms (cattle, lancelet, Mexican tetra, and polyp). Dot size and color intensity represent the number of datasets for each biosample-technology combination. Grey boxes highlight the essential miniENCODE assays (RNA-seq, ATAC-seq, and H3K27ac CUT&Tag) for various tissues. ChIP-seq (Other) includes histone modifications beyond H3K27ac, such as H3K4me3, H3K27me3, H3K9me3, and H3K14ac. (B) Summary of datasets by species, showing the total number of datasets collected for each organism. (C) Summary of datasets by technology, illustrating the distribution of different omics technologies in the study. (D) Ranked list of TF motifs enriched in ELSs of cattle liver. TFs are sorted by the significance of their motif enrichment, with color intensity indicating enrichment strength. This analysis identifies key transcriptional regulators associated with gene expression in cattle liver. (E-F) Enhancer-gene regulatory sub-networks for lung (E) and spleen (F) tissues in cattle. Blue circles represent the top 10 TFs in each tissue, ranked by their outdegree (number of target genes) within the top 100,000 edges. Circle size and color intensity correspond to the number of target genes regulated by each TF. The shade of the blue arrows indicates the interaction strength between TF pairs, with darker shades representing stronger regulatory interactions.

Multi-organism Implementation of miniENCODE

To demonstrate the broad applicability of the miniENCODE approach, we extended our analysis beyond zebrafish to include five additional model organisms with published datasets: medaka (*Oryzias latipes*), cattle (*Bos taurus*), polyp (*Hydra vulgaris*), lancelet (*Branchiostoma floridae*), and Mexican tetra (*Astyanax mexicanus*) (Figure 7A). This diverse set of organisms spans different evolutionary lineages and biological systems, providing a robust test of our methodology.

We collected a total of 1,976 bulk omics datasets (254 billion sequencing reads) and 58 single-cell datasets (385 thousand cells) from six model organisms. Dataset distribution by species was: zebrafish (1,434), medaka (248), cattle (122), polyp (101), lancelet (80), and Mexican tetra (49) (Figure 7B). By technology type, the

collection consisted of 1,037 RNA-seq, 511 ChIP-seq, 258 ATAC-seq, 170 methylation, and 58 single-cell datasets (Figure 7C). From these primary datasets, we generated 1,172 processed data tracks, distributed as zebrafish (728), medaka (130), cattle (123), polyp (90), lancelet (63), and Mexican tetra (38). As medaka lacked H3K27ac data, subsequent analyses focused on cattle, polyp, lancelet, and Mexican tetra.

Analysis of ELSs in these organisms revealed distinct regulatory patterns associated with tissue identity, developmental timing, and environmental adaptation (Supplementary Figure 11). In Mexican tetra, we observed extensive ELS sharing between liver tissues (Supplementary Figure 11A), reflecting their common organ identity and function (47). Notably, river-dwelling surface fish exhibited a substantial number of unique ELSs, likely reflecting regulatory adaptations to their specific aquatic environment. In lancelet development, we identified a peak in unique ELSs at 15 hpf (Supplementary Figure 11B), coinciding with the somite formation, a critical developmental event (48, 49). Analysis of polyp regeneration revealed 485 ELSs that were consistently active across multiple regeneration stages but absent in control conditions (Supplementary Figure 11C). This shared regulatory element set likely represents a core transcriptional program essential for tissue repair and regeneration processes (50). These findings highlight that ELS sharing is most prominent within the same tissue across different conditions and between adjacent developmental stages, while distinct tissues and specialized adaptations are associated with unique regulatory elements.

TF enrichment analysis in multiple organisms revealed tissue-specific regulatory patterns, with ELSs effectively capturing known key regulators in their respective tissues (Figure 7D, Supplementary Figure 12). In the hypothalamus, members of the SOX family emerged as crucial regulators, consistent with their established roles in neurodevelopment and hypothalamic function. The cerebral cortex showed prominent enrichment of KLF5, SOX10, SOX1, and RONIN, TFs known to regulate cortical development and neuronal differentiation (51–54). Adipose tissue analysis highlighted the importance of ATF3 and FOSL2, factors involved in adipocyte differentiation and metabolic regulation (55, 56). In the liver, we identified SP1, HNF4A, and PPARA as key regulators (Figure 7D), consistent with their well-characterized roles in hepatic metabolism and detoxification (57–59). These findings demonstrate that ELSs can effectively enrich for TFs that are functionally relevant to their respective tissues, providing valuable candidates for further mechanistic studies of gene regulation.

GO analysis of ELS-associated genes revealed tissue-specific functional enrichments (Supplementary Figure 13). In cattle, cerebellum-associated genes were significantly enriched for biological processes related to neural development, particularly neuron projection morphogenesis and brain development. Similarly, skeletal muscle-associated genes showed strong enrichment for muscle structure development and muscle cell differentiation, consistent with their tissue-specific functions.

Further analysis of tissue-specific GRNs identified key TFs associated with organ development. In the lung GRN, the majority of top-ranked TFs were known regulators of pulmonary development and function (Figure 7E), including *AHR*, *ETS1*, *IRF9*, *NFE2L2*, *YBX1*, and *YY1* (60–65). The spleen GRN contained multiple established regulators (Figure 7F), such as *PBX2*, *STAT1*, and *STAT3* (66–68). These findings demonstrate that ELSs effectively capture tissue-specific regulatory programs and their associated biological processes in diverse organisms.

Our multi-organism implementation of miniENCODE demonstrates broad applicability among species. Through analysis of 352 published datasets from four organisms, we identified tissue-specific regulatory elements, key TFs, and GRNs. The consistent identification of known tissue-specific regulators validates our approach, establishing miniENCODE as a valuable framework for studying gene regulation in diverse organisms with limited genomic resources.

Discussion

Understanding the regulatory landscape of genomes is fundamental to modern biology, yet our knowledge remains largely confined to a handful of model organisms. In this study, we demonstrate the effectiveness of the miniENCODE framework in deciphering gene regulatory mechanisms. By integrating multi-omics data from RNA-seq, ATAC-seq, and H3K27ac CUT&Tag assays spanning six developmental stages and 11 adult tissues in zebrafish, we identified 52,350 candidate ELSs and characterized their tissue- and stage-specific activities. Through experimental validation and computational analysis, we delineated key TFs and their regulatory networks, providing valuable insights into the spatiotemporal dynamics of gene regulation. The development of miniODP further enhanced our ability to integrate, visualize, and analyze complex multi-omics datasets. Extending this framework to additional model organisms—including cattle, lancelet, Mexican tetra, and polyp—we demonstrated its broad applicability for understanding gene regulation in diverse species. Collectively, these findings establish the miniENCODE approach as a resource-efficient tool for generating regulatory insights in understudied organisms, even with limited resources.

Design and Implementation of the miniENCODE Framework

The ENCODE project utilizes an all-encompassing suite of technologies, primarily relying on DNase signals and ChIP-seq data for multiple histone marks (H3K4me3, H3K27ac, or CTCF) to identify CREs (2). To streamline this approach while maintaining accuracy, we focused on two key signatures: DNA accessibility and H3K27ac histone modification. We selected H3K27ac because it robustly marks active enhancers and promoters, providing broader coverage than H3K4me3, which primarily labels active promoters. This strategy, validated through in vivo reporter assays, enables efficient ELS identification with reduced complexity. We further optimized our approach by implementing ATAC-seq for chromatin accessibility and CUT&Tag for histone modification profiling, taking advantage of their superior resolution and reduced sequencing requirements (15, 31).

While these two epigenomic methods effectively identify ELSs, they alone cannot fully elucidate complex regulatory mechanisms. Therefore, we incorporated RNA-seq as a core component, enabling correlation of ELS activity with gene expression. By combining these three orthogonal approaches, we can more accurately predict functional regulatory elements, their target genes, and GRNs. This integrated approach has proven effective, as demonstrated by our identification of key TFs whose roles and regulatory networks align with existing literature.

To address diverse research needs and provide deeper insights into gene regulation, we proposed several optional assays that complement the core miniENCODE framework. Long-read RNA-seq enhances transcriptome assembly by providing full-length transcript information, which is particularly important for newly sequenced genomes where accurate gene models are essential. Micro-C, a next-generation chromosome conformation capture technique, offers high-resolution chromatin architecture mapping, enabling the investigation of long-range regulatory interactions and 3D genome organization. Single-cell multi-omics technologies, including single-cell RNA-seq, ATAC-seq, and CUT&Tag, enable cell-type-specific profiling, providing unprecedented resolution in understanding regulatory mechanisms in complex tissues. While these optional assays are more resource-intensive in terms of cost, technical expertise, and computational requirements, they provide valuable additional perspectives on gene regulation that can significantly enhance our understanding of complex biological systems.

The rapid expansion of omics data presents both opportunities and challenges. The vast accumulation of multi-omics datasets can be overwhelming for experimental biologists searching for specific information about their genes or genomic regions of interest. To address this need, we developed the mini Omics Data Portal (miniODP). This platform integrates three key components: (1) a JBrowse-based genome browser for visualizing multi-omics data in genomic context; (2) a gene browser providing detailed functional annotations and expression profiles; and (3) a sequence search tool for identifying target sequences through BLAST searches, with results directly linked to their genomic context. Together, these features enhance data accessibility and interpretation, particularly for researchers working with understudied model organisms who may lack specialized bioinformatics expertise.

Advantages, Limitations, and Future Directions

The miniENCODE framework provides several important advantages for investigating understudied model organisms. First, its resource-efficient design allows for detailed regulatory analysis with limited resources. The computational component enables straightforward reuse of existing datasets, while the experimental assays are practical for individual laboratories. The core technologies - RNA-seq, ATAC-seq, and CUT&Tag - are technically feasible and cost-effective, with per-sample costs of approximately \$80-\$120, \$60-\$135, and \$100-\$190, respectively. This results in a total cost below \$600-\$1100 for one sample with two replicates, enabling individual laboratories to examine multiple biosamples of interest. Second, the framework's broad applicability in various organisms, as shown in our study, indicates its potential for characterizing regulatory landscapes in the expanding range of understudied species.

Compared to existing zebrafish data portals, miniODP offers several key advantages. First, it supports multiple species, currently integrating data from six model organisms with flexible expansion capabilities, enabling comparative analysis in diverse biological systems. Second, miniODP provides comprehensive data visualization through three interconnected modules: a genome browser with newly generated and reanalyzed tracks (including RNA-seq, ATAC-seq, and ChIP-seq data), a Shiny-based visualization platform for gene expression and functional annotation, and a BLAST search tool for sequence analysis. These features facilitate efficient exploration of multi-omics data, from raw data visualization to in-depth analysis. The platform's integration with external databases (Ensembl, AnimalTFDB, and InterPro) and support for single-cell RNA-seq and ATAC-seq data visualization further enhance its functionality. Collectively, these features make miniODP a versatile tool for multi-omics data reuse, particularly valuable for researchers working with non-model organisms. In fact, this platform has been operational for over a year, serving several thousand users during its testing phase.

However, several limitations should be acknowledged. Compared to the ENCODE project, which annotated 7.9% and 3.4% of the human and mouse genomes as candidate cis-regulatory elements (cCREs) (2), our zebrafish miniENCODE project identified ELSs in only 1.2% of the genome, covering 17% of ATAC-seq signal peaks. This limited coverage likely results from several factors: restricted sampling of developmental stages and tissues, lower sequencing depth, and the absence of additional technologies (e.g., ChIP-seq for CTCF and H3K4me3) that could identify other types of regulatory elements. These limitations highlight the need for more extensive data collection to fully characterize the zebrafish regulatory landscape.

Looking forward, the miniENCODE framework could catalyze community-driven projects. By coordinating efforts across multiple laboratories, research communities could implement scaled-down versions of the ENCODE project, focusing on critical stages or tissues of their chosen models, while sharing data through the miniODP database. This collaborative approach would enable rapid identification of key regulatory information without overextending individual laboratory resources. We invite fellow researchers interested in exploring potential joint ventures. Our team is prepared to assist in analyzing miniENCODE datasets from various model organisms and is committed to utilizing our existing miniODP database to serve a broader scientific community.

Our work provides a starting point for future studies aiming to unravel complex regulatory mechanisms in understudied model organisms. By advocating for resource-efficient genomic research, the miniENCODE framework has the potential to propel advancements in diverse fields, from developmental biology to disease modeling and evolutionary studies.

Methods

Zebrafish husbandry

The TU strain of zebrafish used in this study were raised and bred following standard fish-keeping procedures (69). The fish were maintained in a recirculating aquaculture system with freshwater at 28 °C and a 14-hour light/10-hour dark photoperiod cycle.

Sample collection

For ovary isolation from adult female zebrafish, we first anesthetized the fish with Tricaine methanesulfonate (MS222) and positioned them in a 2% agarose groove. Using ophthalmic scissors and fine forceps, we performed abdominal dissection and transferred the ovaries to a tube containing 2 mL of DPBS (Gibco, cat. C14190500BT). The ovarian tissue was mechanically dissociated into <1 mm fragments using microdissection scissors. After tissue sedimentation, we replaced the medium with 2 mL of digestive enzyme mixture I (3 mg/mL collagenase I (Worthington, cat. LS004196), 3 mg/mL collagenase II (Worthington, cat. LS004174), and 1.6 mg/mL hyaluronidase (Worthington, cat.

LS002594) in Leibovitz L-15 Medium (Gibco, cat. 21083027)). The mixture was incubated at 37 °C with gentle agitation at 15 rpm on an orbital rotator. We monitored the dissociation progress every 10 minutes until complete tissue dissociation was achieved. Following centrifugation at 300 *g* for 5 minutes, the cell pellet was resuspended in 2 mL of digestive enzyme mixture II (0.25% TRYPSIN-EDTA (Gibco, cat. 15400-054), 0.8 U/mL DNase I (Invitrogen, cat. M0303S) in L-15) and incubated for 10 minutes at 37 °C. The enzymatic reaction was terminated by adding 200 µL of FBS (Sigma-Aldrich, cat. F8318). The cells were then washed twice with 1 mL of DPBS containing 0.04% BSA, centrifuged at 500 *g* for 5 minutes, and filtered through a 100 µm cell strainer.

For embryo collection, we followed established staging criteria (70) and verified developmental stage through morphological examination. Chorion removal was achieved by incubating embryos in 1 mg/mL Pronase for 10 minutes followed by DPBS washing. Yolk removal was performed using 1 mL of deyolking buffer (55 mM NaCl, 1.8 mM KCl, and 1.25 mM $NaHCO_3$) with 20-30 pipetting cycles. Embryonic tissues were homogenized in 2 mL of FACSmax cell dissociation solution (Genlantis, cat. T200100) for 5 minutes at room temperature. The resulting cell suspension was filtered through a 40 µm cell strainer and centrifuged at 310 *g* for 5 minutes. Cell density and viability were quantified using the Countstar Rigel S2 system.

Library preparation

ATAC-seq library Freshly dissected samples were immediately processed for ATAC-seq. Tissues were resuspended in 1.5 mL of lysis buffer (10 mM Tris-HCl [pH 7.4], 10 mM NaCl, 3 mM $MgCl_2$, 0.10% Tween-20 [Bio-rad, cat. 1662404EDU], 0.10% NP-40 [Abcam, cat. ab142227], 0.01% Digitonin, 1% BSA in nuclease-free water). Following filtration through a 40 µm cell strainer, nuclei were collected by centrifugation at 500 g for 5 minutes. ATAC-seq libraries were prepared following established protocols (71). Nuclei were resuspended in 50 µL of transposition mix (10 µL 5x TTBL, 5 µL TTEmix V50 [Vazyme, cat. TD501], 0.5 µL 1% digitonin, 0.5 µL 10% Tween-20, 34 µL ddH_2O). Following PCR amplification, libraries were size-selected using 0.6x AMPure beads (Beckman, cat. A63880) to remove large fragments, followed by 1.2x AMPure beads for library recovery. All ATAC-seq libraries were sequenced on the Illumina NovaSeq platform, with two biological replicates per sample and a minimum of 60 million paired-end 150 bp reads.

CUT&Tag library CUT&Tag was performed following established protocols (15) using the Hyperactive Universal CUT&Tag Assay Kit for Illumina (Vazyme, cat. TD903). Cells were harvested and washed twice with 1.5 mL of Wash Buffer using gentle pipetting. 10 µL of activated concanavalin A-coated magnetic beads were added and incubated at room temperature for 10 minutes. The bead-bound cells were resuspended in 50 µL of chilled Antibody Buffer containing either 1 µL of H3K27ac antibody (Abcam, cat. ab4729) or Rabbit IgG (Abcam, cat. ab171870) as primary antibody, followed by overnight incubation at 4 °C.

After removing the primary antibody using a magnet stand, cells were incubated with Guinea Pig anti-Rabbit IgG (Heavy & Light Chain) antibody (Antibodies-Online, cat. ABIN101961) at room temperature for 60 minutes. The cells were then washed twice with 0.2 mL of Dig-Wash buffer for 5 minutes each using the magnet stand. For tagmentation, 50 µL of 0.04 µM pA/G-Tn5 adapter complex in Dig-300 Buffer was added with gentle vortexing and incubated at room temperature for 1 hour. Following additional washes with Dig-300 Buffer, cells were resuspended in 50 µL of Tagmentation buffer (10 µL of 5x TTBL in 40 µL of Dig-300 Buffer) and incubated at 37 °C for 60 minutes.

DNA extraction was performed by adding 5 µL of Proteinase K, 100 µL of Buffer L/B, and 20 µL of DNA Extract Beads to 50 µL of sample. After two washes with 200 µL of Buffer WA and WB, the beads were dried and resuspended in 22 µL of nuclease-free water. Library amplification was achieved by mixing 15 µL of DNA with 5 µL of universal i5 primer, 5 µL of uniquely barcoded i7 primer, and 25 µL of 2x CAM. Final library purification was performed using 100 µL of VAHTS DNA Clean Beads (Vazyme, cat. N411).

Enhancer reporter assays

To validate tissue-specific ELSs, we performed *Tol2* transposon-mediated zebrafish transgenesis following established protocols (72). Selected enhancers were PCR amplified from zebrafish genomic DNA using primers listed in Supplementary Table 2 and cloned into the pSTG vector upstream of the pig minimum EF1a promoter-EGFP cassette. The resulting reporter plasmids were co-injected with transposase mRNA into one-cell stage TU zebrafish embryos. Embryos were examined at 3 dpf using fluorescence microscopy (Leica) to assess reporter expression patterns, with particular attention to cardiac-specific regions. For each ELS construct, we analyzed at least 100 embryos to ensure robust validation of enhancer activity.

Data analysis

Data processing We constructed a multi-omics database incorporating genome and high-throughput omics data from six species: zebrafish, medaka, mexican tetra, lancelet, polyp, and cattle. Detailed data sources and references are provided in Supplementary Table 3. Raw data quality was evaluated using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), followed by quality filtering with Trim Galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/).

For RNA-seq analysis, we implemented a standardized bioinformatics pipeline. Reads were aligned to reference genomes using HISAT2 (version 2.2.1) (73), followed by gene quantification with Stringtie (version 2.1.2) (74). Gene expression levels were normalized and reported as transcripts per kilobase million (TPM).

For ChIP-seq and ATAC-seq analysis, reads were aligned to the zebrafish (GRCz11) or medaka reference genomes from Ensembl using Bowtie2 (version 2.4.2) (75). Duplicate fragments were removed using SAMtools (76) and Picard (http://broadinstitute.github.io/picard/). Peak calling was performed with MACS2 (version 2.2.7.1) (77), and resulting alignments were indexed using SAMtools. We implemented three quality control metrics for epigenomic samples: 1. Insert-size distribution. 2. TSS enrichment score (ratio of sequencing depth at transcription start sites versus flanking regions). 3. Pearson correlation between replicates.

For methylation analysis (WGBS and Bisulfite-Seq), we processed data using Bismark (version 0.22.3) (78) and deepTools (version 3.4.3) (79). Coverage files in Bigwig format are available through JBrowse.

Single-cell RNA-seq data were analyzed using Seurat (version 3.2.2) (80), with input from either pre-processed gene expression count tables or raw data processed through Cell Ranger (version 4.0.0). Data sources are listed in Supplementary Table 3. We applied standard quality filters based on: 1. Number of detected features (nFeature). 2. Total counts per cell (nCount). 3. Mitochondrial content. For snATAC-seq data, we constructed gene activity score matrices to estimate gene expression levels from chromatin accessibility patterns.

We performed dimensionality reduction using principal component analysis (PCA) on the top 2000 most variable genes. The optimal number of principal components was determined through ElbowPlot and JackStrawPlot analyses. Cell clustering was performed using both UMAP and t-SNE visualization methods. Differential gene expression analysis between clusters was conducted using the FindAllMarkers function with default parameters.

Interactive visualization tools The miniODP platform incorporates multiple visualization tools to support data exploration: 1. Genome browser: JBrowse (23) integration enables visualization of annotated genomic features, sequences, and multi-omics data across all reference genome assemblies. 2. Gene viewer: Built with R Shiny (https://shiny.rstudio.com/) and ShinyProxy (https://www.shinyproxy.io/) using an SQLite backend, this tool supports concurrent user access and interactive exploration of gene-specific data. 3. BLAST tool: SequenceServer (27) implementation provides a user-friendly interface with interactive visualization of sequence alignment results.

ELSs prediction We predicted potential enhancers using multi-omics data following ENCODE project methodology (2). The prediction process consisted of two main phases: representative DNA accessibility region (rDAR) identification and ELS scoring/classification.

For rDAR identification, we first called ATAC-seq peaks using MACS2 and merged adjacent regions with bedtools merge. We filtered peaks to retain only those between 50-500 bp in length, selecting the most significant peak in cases of overlap. This selection process was iterated until no additional peaks remained.

For ELS scoring and classification, we computed Z-scores for log10-transformed ATAC and H3K27ac signals. To account for histone modifications at flanking nucleosomes, we averaged H3K27ac signals across a region encompassing the rDAR plus 500 bp on either side. We defined ELS as rDARs meeting two criteria: (1) ATAC max-Z > 1.64 (corresponding to the 95th percentile of a one-sided Gaussian distribution) and (2) H3K27ac max-Z > 1.64.

We classified ELSs based on their genomic distance from transcription start sites (TSS): TSS-proximal ELS (pELS) were defined as 0-2000 bp from TSS, while TSS-distal ELS (dELS) were >2000 bp from TSS. Tissue-specific ELSs were identified using Bedtools (81) to analyze non-overlapping ELSs across tissues.

ELS-to-gene linkages We established ELS-to-gene linkages using a computational framework adapted from single-cell methodologies (40). The analysis pipeline comprised three key steps: (1) identification of ELS regions as described earlier; (2) generation of sample-specific matrices for ATAC-seq reads (sample $\times$ ELS matrix) and RNA-seq reads (sample $\times$ gene matrix); and (3) computation of Pearson correlation coefficients between log2-normalized values from both matrices for all potential ELS-gene pairs within a 250 kb window centered on each transcription start site (TSS). Significant linkages were then visualized and explored through the miniODP platform.

GO and motif analysis We performed functional annotation through two complementary approaches. First, Gene Ontology (GO) enrichment analysis was conducted using clusterProfiler (version 4.0.5) to identify significantly enriched biological processes and molecular functions. Second, we employed HOMER (82) for de novo motif discovery and enrichment analysis, enabling the identification of TF binding sites and regulatory motifs associated with specific genomic regions or biological conditions.

Identification of super-enhancers We identified super-enhancers through analysis of H3K27ac ChIP-seq data aligned to the GRCz11 reference genome. Using the ROSE algorithm (Rank Ordering of Super-Enhancers) (38), we distinguished super-enhancers from typical enhancers based on their extended H3K27ac signal domains and higher transcriptional regulatory potential. The ROSE algorithm was implemented with default parameters, identifying regions with significantly larger H3K27ac domains and higher signal intensity compared to typical enhancers.

GRN analysis We constructed GRNs using ANANSE (42), integrating gene expression, ATAC-seq, and ChIP-seq data. The analysis pipeline consisted of four main components: First, we established a comprehensive motif database for zebrafish by aggregating vertebrate motifs from the CIS-BP database (Build 2.00) using GimmeMotifs (43). These motifs were converted into a zebrafish-specific format using the motif2factors function. Second, we predicted TF binding sites by combining ATAC-seq and H3K27ac signals with motif scores through the ananse binding function. This approach allowed us to estimate TF binding probabilities across ELS regions. Third, we constructed sample-specific GRNs by integrating four key metrics: (1) predicted TF binding probability in enhancers, (2) genomic distance between enhancers and target genes, (3) expected TF activity, and (4) expression levels of both TFs and their target genes. The Network Influence Score was calculated to rank TFs based on their regulatory potential. Finally, we visualized the resulting networks using the igraph R package (version 1.3.1), enabling exploration of regulatory interactions and network topology.

Code and demonstration files To ensure reproducibility and enable application of our analytical framework, we developed a computational package containing: (1) A Docker image with all software dependencies and configurations, optimized for performance and memory efficiency across systems. (2) Complete source code for all analytical pipelines, including preprocessing, quality control, and downstream analysis modules, with inline comments and usage examples. (3) Detailed documentation covering installation, configuration, execution, troubleshooting, and frequently asked questions. (4) Pre-configured demonstration datasets and configuration files to validate installation and reproduce key results. All resources are available at https://github.com/QTuLab/miniENCODE, with step-by-step instructions for establishing the miniODP platform and conducting analyses.

Data and code availability

ATAC-seq and CUT&Tag data are deposited in the NCBI Gene Expression Omnibus (GEO) database under accession GSE207155 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE207155, reviewer token obmtikycxzunfgz). The miniODP database is accessible at: https://tulab.genetics.ac.cn/miniodp/. All code, documentation, pre-configured Docker environment, and

example files for data analysis and visualization are available at: https://github.com/QTuLab/miniENCODE with links to Figshare.

Acknowledgements

We thank Dr. Jing-Wei Xiong (Peking University) for providing the TU strain of zebrafish; Dr. Tong Chen (State Key Laboratory for Quality Ensurance and Sustainable Use of Dao-di Herbs) and Mr. Minghong Zhang (Institute of Genetics and Developmental Biology) for help with website development. This work was supported by the CAS Project for Young Scientists in Basic Research (YSBR073), CAS Key Research Program (ZDBS-ZRKJZ-TLC009), National Natural Science Foundation of China (32200439 to Y.L.and 32170859), and National Key Research and Development Program of China (2024YFA1803400) to Q.T.

Author contributions

Q.T. and Y.L. designed the project. H.Y. conducted the majority of experiments and computational analyses. Z.S. and H.S. provided experimental support. Y.L. and P.J. contributed to computational analysis. Q.T., H.Y., and Y.L. prepared the manuscript.

Additional information

Supplementary Information: accompanies this paper at Supplementary Figures and Supplementary Table 1, 2 and 3.
Competing interests: The authors declare no competing interests.

References

1. ENCODE Project Consortium (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636–640.
2. ENCODE Project Consortium, Moore, J. E., Purcaro, M. J., Pratt, H. E., Epstein, C. B., Shoresh, N., Adrian, J., Kawli, T., Davis, C. A., Dobin, A. *et al.* (2020) Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710.
3. modENCODE Consortium, Roy, S., Ernst, J., Kharchenko, P. V., Kheradpour, P., Negre, N., Eaton, M. L., Landolin, J. M., Bristow, C. A., Ma, L. *et al.* (2010) Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science* **330**, 1787–1797.
4. Gerstein, M. B., Lu, Z. J., Van Nostrand, E. L., Cheng, C., Arshinoff, B. I., Liu, T., Yip, K. Y., Robilotto, R., Rechtsteiner, A., Ikegami, K. *et al.* (2010) Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. *Science* **330**, 1775–1787.
5. Lieschke, G. J. and Currie, P. D. (2007) Animal models of human disease: Zebrafish swim into view. *Nat Rev Genet* **8**, 353–367.
6. Bradford, Y. M., Van Slyke, C. E., Ruzicka, L., Singer, A., Eagle, A., Fashena, D., Howe, D. G., Frazer, K., Martin, R., Paddock, H. *et al.* (2022) Zebrafish information network, the knowledgebase for Danio rerio research. *Genetics* **220**, iyac016.
7. Yang, H., Luan, Y., Liu, T., Lee, H. J., Fang, L., Wang, Y., Wang, X., Zhang, B., Jin, Q., Ang, K. C. *et al.* (2020) A map of cis-regulatory elements and 3D genome structures in zebrafish. *Nature* **588**, 337–343.
8. Tan, H., Onichtchouk, D. and Winata, C. (2016) DANIO-CODE: Toward an Encyclopedia of DNA Elements in Zebrafish. *Zebrafish* **13**, 54–60.
9. Baranasic, D., Hörtenhuber, M., Balwierz, P. J., Zehnder, T., Mukarram, A. K., Nepal, C., Várnai, C., Hadzhiev, Y., Jimenez-Gonzalez, A., Li, N. *et al.* (2022) Multiomic atlas with functional stratification and developmental dynamics of zebrafish cis-regulatory elements. *Nat Genet* **54**, 1037–1050.
10. Sur, A., Wang, Y., Capar, P., Margolin, G., Prochaska, M. K. and Farrell, J. A. (2023) Single-cell analysis of shared signatures and transcriptional diversity during zebrafish development. *Dev Cell* **58**, 3028–3047.e12.
11. Saunders, L. M., Srivatsan, S. R., Duran, M., Dorrity, M. W., Ewing, B., Linbo, T. H., Shendure, J., Raible, D. W., Moens, C. B., Kimelman, D. and Trapnell, C. (2023) Embryo-scale reverse genetics at single-cell resolution. *Nature* **623**, 782–791.
12. Lange, M., Granados, A., VijayKumar, S., Bragantini, J., Ancheta, S., Kim, Y.-J., Santhosh, S., Borja, M., Kobayashi, H., McGeever, E. *et al.* (2024) A multimodal zebrafish developmental atlas reveals the state-transition dynamics of late-vertebrate pluripotent axial progenitors. *Cell* S0092–8674(24)01147–4.
13. Sun, K., Liu, X., Xu, R., Liu, C., Meng, A. and Lan, X. (2024) Mapping the chromatin accessibility landscape of zebrafish embryogenesis at single-cell resolution by SPATAC-seq. *Nat Cell Biol* **26**, 1187–1199.
14. Jiang, S., Li, H., Zhang, L., Mu, W., Zhang, Y., Chen, T., Wu, J., Tang, H., Zheng, S., Liu, Y. *et al.* (2025) Generic Diagramming Platform (GDP): A comprehensive database of high-quality biomedical graphics. *Nucleic Acids Res* **53**, D1670–D1676.
15. Kaya-Okur, H. S., Wu, S. J., Codomo, C. A., Pledger, E. S., Bryson, T. D., Henikoff, J. G., Ahmad, K. and Henikoff, S. (2019) CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* **10**, 1930.
16. Sharon, D., Tilgner, H., Grubert, F. and Snyder, M. (2013) A single-molecule long-read survey of the human transcriptome. *Nat Biotechnol* **31**, 1009–1014.
17. Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293.
18. Hsieh, T.-H. S., Weiner, A., Lajoie, B., Dekker, J., Friedman, N. and Rando, O. J. (2015) Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell* **162**, 108–119.
19. Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B., Siddiqui, A. *et al.* (2009) mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* **6**, 377–382.
20. Buenrostro, J. D., Wu, B., Litzenburger, U. M., Ruff, D., Gonzales, M. L., Snyder, M. P., Chang, H. Y. and Greenleaf, W. J. (2015) Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490.
21. Wu, S. J., Furlan, S. N., Mihalas, A. B., Kaya-Okur, H. S., Feroze, A. H., Emerson, S. N., Zheng, Y., Carson, K., Cimino, P. J., Keene, C. D. *et al.* (2021) Single-cell CUT&Tag analysis of chromatin modifications in differentiation and tumor progression. *Nat Biotechnol* **39**, 819–824.
22. Bartosovic, M., Kabbe, M. and Castelo-Branco, G. (2021) Single-cell CUT&Tag profiles histone modifications and

transcription factors in complex tissues. *Nat Biotechnol* **39**, 825–835.

23. Buels, R., Yao, E., Diesh, C. M., Hayes, R. D., Munoz-Torres, M., Helt, G., Goodstein, D. M., Elsik, C. G., Lewis, S. E., Stein, L. and Holmes, I. H. (2016) JBrowse: A dynamic web platform for genome visualization and analysis. *Genome Biol* **17**, 66.
24. Cunningham, F., Allen, J. E., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Austine-Orimoloye, O., Azov, A. G., Barnes, I., Bennett, R. *et al.* (2022) Ensembl 2022. *Nucleic Acids Res* **50**, D988–D995.
25. Hu, H., Miao, Y.-R., Jia, L.-H., Yu, Q.-Y., Zhang, Q. and Guo, A.-Y. (2019) AnimalTFDB 3.0: A comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic Acids Res* **47**, D33–D38.
26. Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B. L., Salazar, G. A., Bileschi, M. L., Bork, P., Bridge, A., Colwell, L. *et al.* (2023) InterPro in 2022. *Nucleic Acids Res* **51**, D418–D427.
27. Priyam, A., Woodcroft, B. J., Rai, V., Moghul, I., Munagala, A., Ter, F., Chowdhary, H., Pieniak, I., Maynard, L. J., Gibbins, M. A. *et al.* (2019) Sequenceserver: A Modern Graphical User Interface for Custom BLAST Databases. *Mol Biol Evol* **36**, 2922–2924.
28. Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M. *et al.* (2013) NCBI GEO: Archive for functional genomics data sets–update. *Nucleic Acids Res* **41**, D991–995.
29. Leinonen, R., Akhtar, R., Birney, E., Bower, L., Cerdeno-Tárraga, A., Cheng, Y., Cleland, I., Faruque, N., Goodgame, N., Gibson, R. *et al.* (2011) The European Nucleotide Archive. *Nucleic Acids Res* **39**, D28–31.
30. Gasperini, M., Tome, J. M. and Shendure, J. (2020) Towards a comprehensive catalogue of validated and target-linked human enhancers. *Nat Rev Genet* **21**, 292–310.
31. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. and Greenleaf, W. J. (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213–1218.
32. Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A. *et al.* (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931–21936.
33. Heinz, S., Romanoski, C. E., Benner, C. and Glass, C. K. (2015) The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol* **16**, 144–154.
34. Prummel, K. D., Nieuwenhuize, S. and Mosimann, C. (2020) The lateral plate mesoderm. *Development* **147**, dev175059.
35. Zorn, A. M. and Wells, J. M. (2009) Vertebrate endoderm development and organ formation. *Annu Rev Cell Dev Biol* **25**, 221–251.
36. Sarropoulos, I., Sepp, M., Frömel, R., Leiss, K., Trost, N., Leushkin, E., Okonechnikov, K., Joshi, P., Giere, P., Kutscher, L. M. *et al.* (2021) Developmental and evolutionary dynamics of cis-regulatory elements in mouse cerebellar cells. *Science* **373**, eabg4696.
37. Suster, M. L., Kikuta, H., Urasaki, A., Asakawa, K. and Kawakami, K. (2009) Transgenesis in zebrafish with the tol2 transposon system. *Methods Mol Biol* **561**, 41–63.
38. Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., Rahl, P. B., Lee, T. I. and Young, R. A. (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319.
39. Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-André, V., Sigova, A. A., Hoke, H. A. and Young, R. A. (2013) Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947.
40. Granja, J. M., Corces, M. R., Pierce, S. E., Bagdatli, S. T., Choudhry, H., Chang, H. Y. and Greenleaf, W. J. (2021) ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet* **53**, 403–411.
41. Xin, M., Olson, E. N. and Bassel-Duby, R. (2013) Mending broken hearts: Cardiac development as a basis for adult heart regeneration and repair. *Nat Rev Mol Cell Biol* **14**, 529–541.
42. Xu, Q., Georgiou, G., Frölich, S., van der Sande, M., Veenstra, G. J. C., Zhou, H. and van Heeringen, S. J. (2021) ANANSE: An enhancer network-based computational approach for predicting key transcription factors in cell fate determination. *Nucleic Acids Res* **49**, 7966–7985.
43. van Heeringen, S. J. and Veenstra, G. J. C. (2011) GimmeMotifs: A de novo motif prediction pipeline for ChIP-sequencing experiments. *Bioinformatics* **27**, 270–271.
44. Gerri, C., Marass, M., Rossi, A. and Stainier, D. Y. R. (2018) Hif-1α and Hif-2α regulate hemogenic endothelium and hematopoietic stem cell formation in zebrafish. *Blood* **131**, 963–973.
45. Kwan, W., Cortes, M., Frost, I., Esain, V., Theodore, L. N., Liu, S. Y., Budrow, N., Goessling, W. and North, T. E. (2016) The Central Nervous System Regulates Embryonic HSPC Production via Stress-Responsive Glucocorticoid Receptor Signaling. *Cell Stem Cell* **19**, 370–382.
46. Han, R., Wang, R., Zhao, Q., Han, Y., Zong, S., Miao, S., Song, W. and Wang, L. (2016) Trim69 regulates zebrafish brain development by ap-1 pathway. *Sci Rep* **6**, 24034.
47. Xiong, S., Wang, W., Kenzior, A., Olsen, L., Krishnan, J., Persons, J., Medley, K., Peuß, R., Wang, Y., Chen, S. *et al.* (2022) Enhanced lipogenesis through Pparγ helps cavefish adapt to food scarcity. *Curr Biol* **32**, 2272–2280.e6.
48. Marlétaz, F., Firbas, P. N., Maeso, I., Tena, J. J., Bogdanovic, O., Perry, M., Wyatt, C. D. R., de la Calle-Mustienes, E., Bertrand, S., Burguera, D. *et al.* (2018) Amphioxus functional genomics and the origins of vertebrate gene regulation. *Nature* **564**, 64–70.
49. Mantica, F., Iñiguez, L. P., Marquez, Y., Permanyer, J., Torres-Mendez, A., Cruz, J., Franch-Marro, X., Tulenko, F., Burguera, D., Bertrand, S. *et al.* (2024) Evolution of tissue-specific expression of ancestral genes across vertebrates and insects. *Nat Ecol Evol* **8**, 1140–1153.
50. Murad, R., Macias-Muñoz, A., Wong, A., Ma, X. and Mortazavi, A. (2021) Coordinated Gene Expression and Chromatin Regulation during Hydra Head Regeneration. *Genome Biol Evol* **13**, evab221.
51. Wang, Y., Cui, Y., Liu, J., Song, Q., Cao, M., Hou, Y., Zhang, X. and Wang, P. (2022) Krüppel-like factor 5 accelerates the pathogenesis of Alzheimer's disease via BACE1-mediated APP processing. *Alzheimers Res Ther* **14**, 103.

52. Kim, J., Lo, L., Dormand, E. and Anderson, D. J. (2003) SOX10 maintains multipotency and inhibits neuronal differentiation of neural crest stem cells. *Neuron* **38**, 17–31.
53. Alcock, J. and Sottile, V. (2009) Dynamic distribution and stem cell characteristics of Sox1-expressing cells in the cerebellar cortex. *Cell Res* **19**, 1324–1333.
54. Zwaka, T. P., Skowronska, M., Richman, R. and Dejosez, M. (2021) Ronin overexpression induces cerebellar degeneration in a mouse model of ataxia. *Dis Model Mech* **14**, dmm044834.
55. Cheng, C.-F., Ku, H.-C., Cheng, J.-J., Chao, S.-W., Li, H.-F., Lai, P.-F., Chang, C.-C., Don, M.-J., Chen, H.-H. and Lin, H. (2019) Adipocyte browning and resistance to obesity in mice is induced by expression of ATF3. *Commun Biol* **2**, 389.
56. Wrann, C. D., Eguchi, J., Bozec, A., Xu, Z., Mikkelsen, T., Gimble, J., Nave, H., Wagner, E. F., Ong, S.-E. and Rosen, E. D. (2012) FOSL2 promotes leptin gene expression in human and mouse adipocytes. *J Clin Invest* **122**, 1010–1021.
57. D'Alessio, J. A., Ng, R., Willenbring, H. and Tjian, R. (2011) Core promoter recognition complex changes accompany liver development. *Proc Natl Acad Sci U S A* **108**, 3906–3911.
58. Xu, Y., Zhou, Z., Kang, X., Pan, L., Liu, C., Liang, X., Chu, J., Dong, S., Li, Y., Liu, Q. *et al.* (2022) Mettl3-mediated mRNA m6A modification controls postnatal liver development by modulating the transcription factor Hnf4a. *Nat Commun* **13**, 4555.
59. Wang, Y., Nakajima, T., Gonzalez, F. J. and Tanaka, N. (2020) PPARs as Metabolic Regulators in the Liver: Lessons from Liver-Specific PPAR-Null Mice. *Int J Mol Sci* **21**, 2061.
60. Major, J., Crotta, S., Finsterbusch, K., Chakravarty, P., Shah, K., Frederico, B., D'Antuono, R., Green, M., Meader, L., Suarez-Bonnet, A. *et al.* (2023) Endothelial AHR activity prevents lung barrier disruption in viral infection. *Nature* **621**, 813–820.
61. Ma, T., M, W., O, D., H, B. and Cm, C. (2016) Enhanced MAPK signaling drives ETS1-mediated induction of miR-29b leading to downregulation of TET1 and changes in epigenetic modifications in a subset of lung SCC. *Oncogene* **35**.
62. Gao, X., Huang, D., Liu, Y., Zhang, G., Zheng, X., Guan, B., Chen, A., Wu, J., Luo, S.-M., Liu, Z. *et al.* (2025) FAM210B activates STAT1/IRF9/IFIT3 axis by upregulating IFN-α/β expression to impede the progression of lung adenocarcinoma. *Cell Death Dis* **16**, 63.
63. Binkley, M. S., Jeon, Y.-J., Nesselbush, M., Moding, E. J., Nabet, B. Y., Almanza, D., Kunder, C., Stehr, H., Yoo, C. H., Rhee, S. *et al.* (2020) KEAP1/NFE2L2 Mutations Predict Lung Cancer Radiation Resistance That Can Be Targeted by Glutaminase Inhibition. *Cancer Discov* **10**, 1826–1841.
64. Yu, T., Zhang, Q., Yu, S.-K., Nie, F.-Q., Zhang, M.-L., Wang, Q. and Lu, K.-H. (2023) THOC3 interacts with YBX1 to promote lung squamous cell carcinoma progression through PFKFB4 mRNA modification. *Cell Death Dis* **14**, 475.
65. Boucherat, O., Landry-Truchon, K., Bérubé-Simard, F.-A., Houde, N., Beuret, L., Lezmi, G., Foulkes, W. D., Delacourt, C., Charron, J. and Jeannotte, L. (2015) Epithelial inactivation of Yy1 abrogates lung branching morphogenesis. *Development* **142**, 2981–2995.
66. Koss, M., Bolze, A., Brendolan, A., Saggese, M., Capellini, T. D., Bojilova, E., Boisson, B., Prall, O. W. J., Elliott, D. A., Solloway, M. *et al.* (2012) Congenital asplenia in mice and humans with mutations in a Pbx/Nkx2-5/p15 module. *Dev Cell* **22**, 913–926.
67. Solomon, A., Bossel Ben-Moshe, N., Hoffman, D., Trzebanski, S., Yehezkel, D., Vainman, L., Netea, M. G. and Avraham, R. (2025) Early and delayed STAT1-dependent responses drive local trained immunity of macrophages in the spleen. *Elife* **13**, RP100922.
68. Mello, A. S., de Oliveira, D. C., Bizzarro, B., Sá-Nunes, A., Hastreiter, A. A., Beltran, J. S. d. O., Xavier, J. G., Borelli, P. and Fock, R. A. (2014) Protein malnutrition alters spleen cell proliferation and IL-2 and IL-10 production by affecting the STAT-1 and STAT-3 balance. *Inflammation* **37**, 2125–2138.
69. Westerfield, M. *The Zebrafish Book. A Guide for the Laboratory Use of Zebrafish (Danio Rerio)* (Univ. of Oregon Press, Eugene, 2000), 4th ed edn.
70. Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B. and Schilling, T. F. (1995) Stages of embryonic development of the zebrafish. *Dev Dyn* **203**, 253–310.
71. Wu, J., Xu, J., Liu, B., Yao, G., Wang, P., Lin, Z., Huang, B., Wang, X., Li, T., Shi, S. *et al.* (2018) Chromatin analysis in human early development reveals epigenetic transition during ZGA. *Nature* **557**, 256–260.
72. Xie, W., Schultz, M. D., Lister, R., Hou, Z., Rajagopal, N., Ray, P., Whitaker, J. W., Tian, S., Hawkins, R. D., Leung, D. *et al.* (2013) Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134–1148.
73. Kim, D., Langmead, B. and Salzberg, S. L. (2015) HISAT: A fast spliced aligner with low memory requirements. *Nat Methods* **12**, 357–360.
74. Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T. and Salzberg, S. L. (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**, 290–295.
75. Langmead, B. and Salzberg, S. L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359.
76. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079.
77. Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W. and Liu, X. S. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137.
78. Krueger, F. and Andrews, S. R. (2011) Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572.
79. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. and Manke, T. (2014) deepTools: A flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, W187–191.
80. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P. and Satija, R. (2019) Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21.
81. Quinlan, A. R. and Hall, I. M. (2010) BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842.
82. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H. and Glass, C. K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576–589.