



A comprehensive single-cell atlas of monoallelic expression across various tissues in zebrafish

Hui Zeng, Zhan Liu, Huimin Xiang, et al.

Genome Res. 2026 36: 1502-1513 originally published online May 28, 2026

Access the most recent version at doi:[10.1101/gr.281338.125](https://doi.org/10.1101/gr.281338.125)

References

This article cites 45 articles, 7 of which can be accessed free at:

<http://genome.cshlp.org/content/36/7/1502.full.html#ref-list-1>

Creative Commons License

This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see

<https://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

Comprehensive immune receptor profiling.
Discover the **DriverMap™ AIR Assay** difference.



To subscribe to *Genome Research* go to:

<https://genome.cshlp.org/subscriptions>

A comprehensive single-cell atlas of monoallelic expression across various tissues in zebrafish

Hui Zeng,^{1,4} Zhan Liu,^{1,4} Huimin Xiang,¹ Boliang Liu,¹ Di Zhang,^{1,2} Xionglei He,^{1,3} Shanjun Deng,¹ and Li Liu¹

¹MOE Key Laboratory of Gene Function and Regulation, State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-Sen University, Guangzhou 510275, China; ²School of Life Sciences, Jiaying University, Meizhou 514015, China; ³Innovation Center for Evolutionary Synthetic Biology, Sun Yat-Sen University, Guangzhou 510275, China

Monoallelic expression (MAE), the preferential transcription of one allele in diploid organisms, plays a crucial role in generating cellular heterogeneity and phenotypic diversity. Although MAE has been extensively characterized in humans and mice, previous studies have been limited by clonal analyses or insufficient cell-type resolution, leaving its dynamics during development largely unexplored. In this study, we present a comprehensive single-cell atlas of allelic expression in zebrafish, leveraging reciprocal crosses between divergent strains, whole-genome sequencing, and single-cell RNA sequencing at two developmental stages. Using maximum likelihood estimation, we profile allele-specific expression patterns across 50,819 cells and 37 distinct cell types, identifying 738 genes (16.5% of evaluable genes) with MAE. MAE SNVs exhibit chromosomal clustering and regional enrichment, and MAE genes show functional specialization, being enriched in chromatin remodeling and hematopoietic processes. Notably, MAE genes are more likely to be tissue-specific markers, particularly in the epidermis, periderm, and erythrocytes. Tissue context further shapes allele choice, with distinct allelic biases observed in specific cell types. Together, our findings provide a comprehensive single-cell atlas of allelic expression in zebrafish, revealing the widespread and dynamic nature of MAE during vertebrate development and its contribution to tissue specialization and regulatory diversity.

[Supplemental material is available for this article.]

In diploid eukaryotes, it is commonly posited that the two copies of most genes exhibit comparable transcriptional activity. However, emerging evidence challenges this paradigm, revealing that a subset of genes exhibit allele-specific expression biases, including monoallelic expression (MAE), the transcriptional dominance of one allele over the other (Gimelbrant et al. 2007; Zwemer et al. 2012; Deng et al. 2014; Eckersley-Maslin and Spector 2014; Reinius and Sandberg 2015; Chess 2016; Kravitz et al. 2023).

The conceptual foundation of MAE traces to the 1960s, when immunoglobulin genes in rabbit lymphoid tissues were first shown to exhibit allele-specific activation (Pernis et al. 1965). Subsequent studies in the 1990s identified analogous patterns in olfactory receptor genes (Chess et al. 1994), followed by discoveries of MAE in immune-related loci, including pheromone receptors, interleukins, and natural killer cell receptors (Held et al. 1995; Bix and Locksley 1998; Holländer et al. 1998; Rodriguez et al. 1999). Comparative genomic analyses have further revealed functional and spatial distinctions between MAE and biallelically expressed (BAE) genes (Savova et al. 2016; Xu et al. 2017; Kravitz et al. 2023). According to previous studies, MAE can be broadly classified into constitutive and random forms (Reinius and Sandberg 2015; Quan et al. 2024). Constitutive MAE represents stable, nonrandom expression from the same allele across most cells and includes parent-of-origin-dependent genomic imprint-

ing as well as genetically determined allele-preferential expression driven by *cis*-regulatory differences. In contrast, random MAE reflects stochastic allelic choice among individual cells and comprises clonally fixed and dynamic forms, of which the dynamic form largely arises from transcriptional bursting and leads to transient allelic fluctuations at the single-cell level. These findings firmly established MAE as a regulated and biologically meaningful process rather than as a rare or stochastic transcriptional anomaly.

Despite these, the reported prevalence of MAE varies substantially across species and tissues, ranging from 2% to 76%, in part reflecting differences in cell types, developmental context, and the relative contributions of burst-driven dynamic MAE versus regulated allele-preferential mechanisms. Microarray-based analyses in human B-lymphoblastoid clones demonstrated that >5% of genes exhibit random MAE (Gimelbrant et al. 2007), whereas mouse studies reported a prevalence of 10% (Zwemer et al. 2012). Neural stem cell models further revealed ~2% of autosomal genes with allele-specific biases in both humans and mice (Jeffries et al. 2012; Li et al. 2012). However, other research uncovered MAE in 12%–24% of genes in mouse embryos (Deng et al. 2014) and 76.4% of SNVs in human fibroblasts (Borel et al. 2015), respectively. Moreover, minimal overlap is observed among MAE genes identified across different studies even in the same organisms (Gimelbrant et al. 2007; Jeffries et al. 2012; Li et al. 2012; Zwemer et al. 2012), indicating that allelic biases are not static

⁴These authors contributed equally to this work.

Corresponding authors: liul47@mail.sysu.edu.cn, dengshj8@mail.sysu.edu.cn

Article published online before print. Article, supplemental material, and publication date are at <https://www.genome.org/cgi/doi/10.1101/gr.281338.125>.

© 2026 Zeng et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <https://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

but may arise dynamically during cellular differentiation and lineage specification. Supporting this, mouse models demonstrated a 5.6-fold increase in random MAE genes during the transition from embryonic stem cells to neural progenitors, highlighting the developmental regulation of allelic expression (Eckersley-Maslin et al. 2014).

However, most MAE studies have focused on clonal populations or limited tissues, failing to capture the landscape of MAE across the diverse cell populations within a whole organism (Fu et al. 2021; Liang et al. 2021; Kravitz et al. 2023). To bridge this gap, we leveraged the zebrafish (*Danio rerio*), a premier vertebrate model organism in developmental biology. Its reasonable embryo total cell number, coupled with rapid embryogenesis, genetic tractability, and high conservation of human genes (~70% orthology), uniquely enabled us to systematically profile MAE at single-cell resolution across all major cell types within the intact, developing vertebrate.

In this study, we performed reciprocal crosses between two genetically divergent zebrafish strains (TL and TU), collecting single-cell transcriptomic data from whole embryos at two developmental stages, 2 days postfertilization (dpf) and 7 dpf. Leveraging a maximum likelihood estimation (MLE) framework, we sought to quantitatively assess allele-specific expression across tens of thousands of individual cells, characterize the prevalence and tissue specificity of MAE, and underscore the complexity of allele-specific regulation during vertebrate development.

Results

Widespread MAE revealed at single-cell level across the whole body

To systematically characterize MAE pattern, we utilized single-nucleotide variants (SNVs) between the two genetically divergent zebrafish strains (TL and TU) as allelic markers, following established methodologies from previous studies in the mouse (Zwemer et al. 2012; Deng et al. 2014; Eckersley-Maslin et al. 2014; Reinius et al. 2016; Xu et al. 2017). We performed whole-genome sequencing (WGS) on homozygous individuals from both TL and TU strains (Fig. 1A; Supplemental Fig. S1). Across three independent TL–TU pairings, we identified 3.2 million to 6.9 million high-confidence SNVs that distinguished the parental genomes, representing ~0.4% of the zebrafish genome (Methods) (Supplemental Fig. S2), consistent in scale with previous findings (Howe et al. 2013). Of the total genomic SNVs, about 0.26 million to 0.29 million (4.0%–4.2% of total SNVs) fall within annotated gene bodies. Among these genic SNVs, ~40% are located in coding sequences (CDSs) and ~50% in 3' UTRs, the regions most accessible to our 3' single-cell RNA sequencing (scRNA-seq) platform.

Subsequently, we generated six F₁ hybrid offspring from these crosses, comprising three experiment groups based on cross direction and developmental stage: two offspring from TL female × TU male crosses at 2 dpf, two offspring from TL female × TU male crosses at 7 dpf, and two offspring from TU female × TL male crosses at 2 dpf. These individuals were lysed and subjected to scRNA-seq (Fig. 1A). After rigorous quality control, we retained a total of 53,926 high-quality cells, with individual samples contributing 4239–16,743 cells (Supplemental Fig. S3A–C). To accurately quantify allele-specific expression, we performed cell-level variant calling using matched SNV information derived from parental genomes. To assess potential reference-mapping bias, we remapped reads to a TL pseudoreference genome and found that

TL allele ratios estimated from the original reference and pseudoreference were highly correlated (Pearson's $R=0.81$, $P<2.2\times 10^{-16}$), with a median difference of only 0.01 (Supplemental Fig. S3D), indicating that reference-mapping bias was limited and unlikely to substantially affect the overall allelic expression patterns.

To minimize the influence of technical noise and stochastic variation (Prashant et al. 2020; Fu et al. 2021), we imposed filtering criteria: Only SNVs supported by at least five UMIs within a cell and expressed in at least five cells were retained (Supplemental Fig. S1). Simulations validated that this threshold provides robust MAE detection with low false-positive rates (Supplemental Fig. S1B,C). This yielded 15,849 high-confidence SNVs across 50,819 cells for downstream allelic expression analysis (Supplemental Table S1). The median number of cells per SNV was 15 (interquartile range, 8–44) (Supplemental Fig. S4A,B), and both alleles were detected in 99.6% of SNV-positive cells, indicating robust coverage and minimal allelic dropout.

These SNVs exhibited distinct allelic expression patterns. For example, Chr 7: 6,765,875 displayed symmetric allelic expression across most cells, with only a small fraction of cells (8/99, 8.1%) deviating from the expected allelic ratio (TL:TU=5:3). The majority of cells (91.9%) maintained allelic ratios within the 0.2–0.8 range (Fig. 1B). A similar pattern was observed for Chr 4: 15,996,066, in which only 10/58 (17.2%) cells exhibited allelic imbalance, with no significant bias toward either allele (TL:TU=5:5) (Fig. 1C; Supplemental Fig. S4C). In contrast, Chr 1: 8,650,290 showed strong TU allele preference, with a mean allelic frequency of 0.83. All outlier cells (20/37, 54.1%) exclusively expressed the TU allele (TL:TU=0:20) (Fig. 1D; Supplemental Fig. S4C). Additionally, some SNVs exhibited cell-to-cell heterogeneity in allelic usage. For instance, the TL allele frequency for Chr 16: 31,709,576 was approximately 0.34 overall, but individual cells displayed divergent expression patterns, with some expressing exclusively the TL allele and others expressing only the TU allele (32/83 outliers, 38.6%; TL:TU=7:25) (Fig. 1E; Supplemental Fig. S4D).

To explicitly capture allelic preference at single-cell resolution, we implemented an MLE framework comprising three hierarchical models (Methods) (Fig. 1A). Through likelihood ratio testing, we classified each SNV into distinct expression models, ultimately categorizing each SNV as exhibiting either BAE or MAE (Fig. 1A). To validate that our MLE framework reliably captures population-level allelic skew when present (i.e., model 1, directional MAE), we performed a pseudobulk analysis by aggregating UMI counts across all cells for each SNV and computing the pseudobulk allelic fraction. Applying the same directional monoallelic cutoff ($f\geq 0.8$ or $f\leq 0.2$) to the pseudobulk estimates, we identified 1871 bulk-M1 SNVs, of which 80.7% overlapped with the MLE-derived M1 SNVs (Supplemental Fig. S5). Across the six F₁ hybrid individuals, we detected between 429 and 1408 SNVs per individual that conformed to MAE, representing ~12%–17% of all high-confidence SNVs analyzed (Fig. 1F). This finding highlights the prevalence, supporting the utility of our MLE-based modeling approach for allele-specific expression analysis at single-cell resolution.

To exclude technical artifacts, we conducted targeted quality control analyses at multiple levels. First, MAE-associated SNVs exhibited higher median UMI coverage than BAE SNVs (12 vs. 8) (Fig. 1G), and the proportion of MAE genes increased with expression level (Supplemental Fig. S6A), ruling out low-expression bias. Second, to evaluate potential reference mapping bias, we assessed allelic balance both per sample and across all MAE sites. At the per-sample level, the distributions of allelic fractions for MAE SNVs

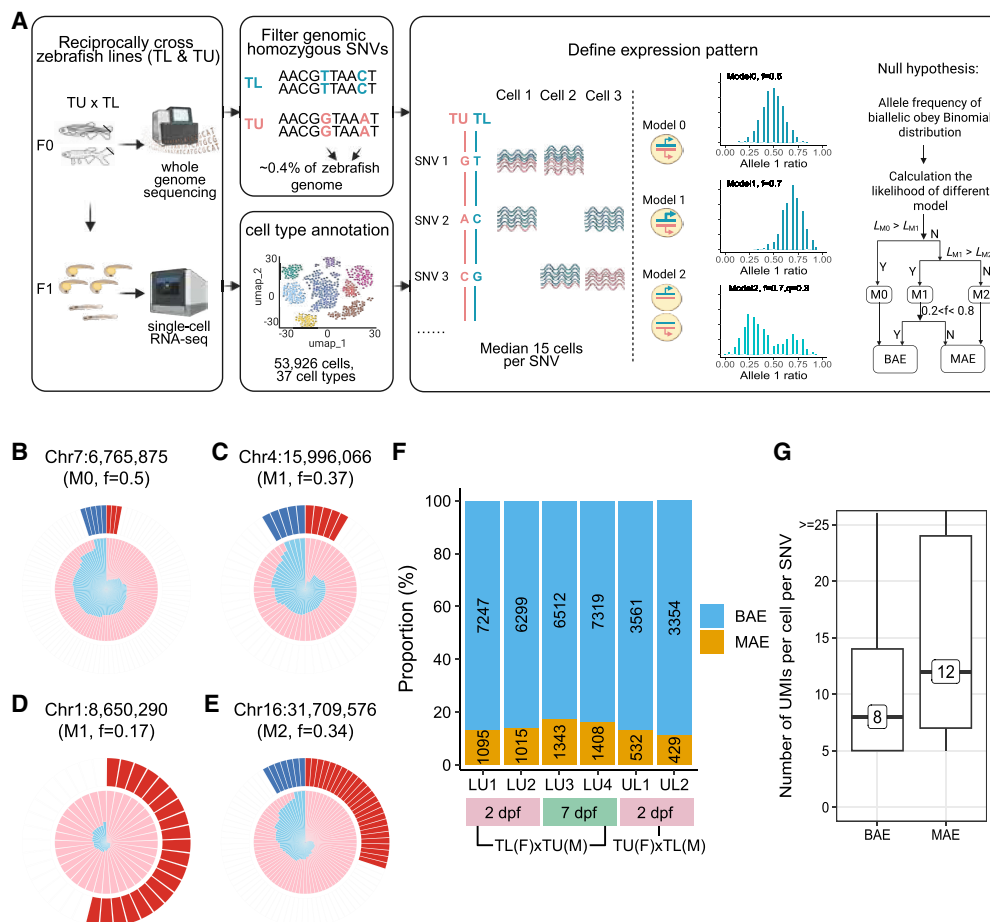


Figure 1. Experimental design and overview of allele-specific expression analysis. (A) Schematic diagram of the experimental design and analysis workflow for allelic expression analysis. Parental adult pairs are fin-clipped and subjected to whole-genome sequencing to identify SNVs distinguishing the TL and TU allele. Progeny from the sequenced adults (TL × TU) were collected at 2 dpf and 7 dpf, followed by single-cell RNA-seq to generate gene expression profiles and tissue annotations. The resulting data were integrated, and expression patterns were inferred using maximum likelihood estimation. (B) SNV Chr 7: 6,765,875 displays balanced biallelic expression (model 0, $f=0.5$). Circular bar plots illustrating representative SNVs with distinct allelic expression patterns. Each bar corresponds to a single cell, with the *inner* ring colors indicate the expressed allele (blue for TL, pink for TU). Outer ring colors highlight the cells with extreme TL allele ratios (blue for ratio > 0.8, red for ratio < 0.2). The greater the number of extreme cells observed for an SNV, the less likely the expression pattern is to represent balanced biallelic expression. A higher proportion of extreme cells for a given SNV indicates a lower likelihood of balanced biallelic expression. The proportion of extreme TL versus TU values further reflects potential allele-specific biases. (C) SNV Chr 4: 15,996,066 displays biased biallelic expression (model 1, $f=0.37$). (D) SNV Chr 1: 8,650,290 displays directional monoallelic (model 1, $f=0.17$). (E) SNV Chr 16: 31,709,576 displays random monoallelic (model 2, $f=0.34$). (F) Distribution of SNVs across allelic expression categories per sample. Blue denotes BAE SNVs, and yellow indicates MAE SNVs. (G) Box plot comparing the number of UMIs per cell per SNV between BAE (left) and MAE (right) SNVs. Median values are displayed above each distribution (BAE, eight; MAE, 12).

showed no systematic skew toward either the TL or TU allele (Supplemental Fig. S6B). When aggregating across individuals, the proportions of TL-biased and TU-biased MAE SNVs differed by only about three percentage points (11.2% vs. 14.6% among 5822 MAE SNVs) (Supplemental Fig. S6C), confirming limited bias. Together, these results support that the observed MAE patterns reflect genuine biological regulation rather than technical noise.

MAE SNVs exhibit chromosomal clustering and regional enrichment

To investigate whether MAE SNVs exhibit nonrandom spatial clustering along chromosomes, we compared their observed genomic distribution to a null distribution generated by within-individual permutations (Methods). To avoid artifacts from SNVs located on

the same sequencing reads, we excluded SNV pairs with intersite distances <150 bp. By shuffling MAE labels across all available informative SNV positions, we controlled for local SNV density and chromosomal architecture. We observed that MAE SNVs showed a higher tendency to occur within short genomic distances (<1 Mb) compared with the architecture-preserving shuffled backgrounds (Fig. 2A). This preferential chromosomal clustering indicates that MAE events are regionally concentrated. This pattern was robust across individuals at different developmental stages and from reciprocal crosses, suggesting a coordinated regional regulation of allelic expression.

To determine whether MAE SNVs are preferentially enriched in specific genomic regions, we calculated the proportion of MAE SNVs using a sliding window approach (1 Mb window with 10 kb step size) across all chromosomes for each individual (Methods). The spatial distribution patterns of MAE-enriched regions were

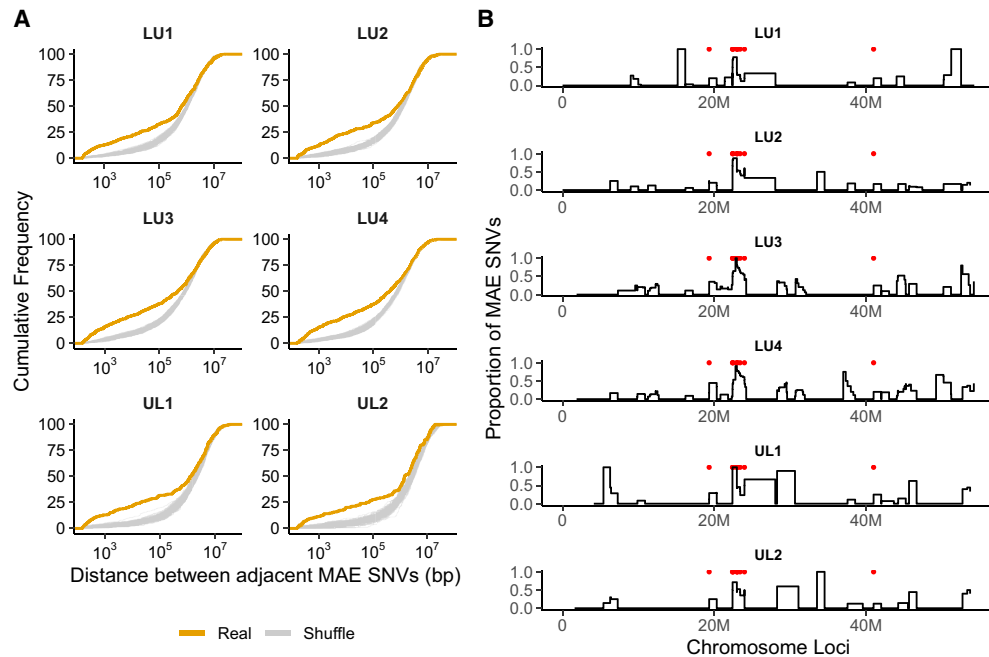


Figure 2. Chromosomal distribution features of MAE SNVs. (A) Empirical cumulative distribution function (ECDF) plot illustrating genomic distance between adjacent SNVs. Each panel represents one of the six F₁ hybrid samples. The yellow curve represents MAE SNVs, and the gray curve represents null distribution of shuffling. (B) Enrichment of MAE SNVs on Chromosome 16. Each panel represents one of the six F₁ hybrid samples. The x-axis indicates the genomic start position of each 1 Mb window; the y-axis represents the proportion of MAE SNVs (MAE count/total SNV count) within that window. Red dots highlight the regions with the strongest enrichment of MAE SNVs, defined as windows exhibiting the lowest coefficients of variation and the highest mean MAE SNV ratio.

notably similar among individuals from the same experimental group (Fig. 2B). Overall, although enrichment values varied because of differences in sequencing quality among individuals, the enriched regions across the genome exhibited similar patterns among different experimental groups.

This analysis revealed several distinct segments that consistently showed an elevated proportion of MAE SNVs across all individuals, for instance, the segment on Chromosome 16 (Chr 16: 22,000,000–23,970,000) (Fig. 2B; Supplemental Fig. S7). Together, these results demonstrate that MAE SNVs are not randomly distributed across the genome but tend to cluster within specific chromosomal regions, pointing to potential local regulatory mechanisms or chromatin-level influences on allelic expression.

Gene-level analysis reveals functional specialization of MAE

Following the classification of SNVs into two primary allelic expression categories, we extended our analysis to the gene level to characterize broader regulatory behaviors. Because SNV coverage and expression levels vary across biological samples, we analyzed each gene–sample pair independently, yielding a total of 13,912 evaluable gene observations across all individuals (rather than unique genes). The majority of genes (60%, 8355/13,912) contained multiple informative SNVs, enabling more reliable assessment of allelic expression consistency (Supplemental Fig. S8A). Most of these genes showed consistent expression patterns across all their SNVs, with only 18% (1504/8355) displaying mixed profiles containing both biallelic and monoallelic expression. Shuffling simulations further confirmed that the observed consistency is significantly higher than expected by chance

(Supplemental Fig. S8B,C), indicating that observed allelic expression patterns are more likely to reflect genuine biological regulation rather than random technical variation.

We applied a filtering approach to exclude ambiguous cases and obtained MAE and BAE genes for each individual (Methods). Among the 4480 genes analyzed, 738 (16.5%) were classified as MAE genes with the remainder as BAE genes (Fig. 3A), underscoring the substantial prevalence of MAE within the zebrafish transcriptome. Pairwise comparisons across all individuals showed significant enrichment of shared MAE genes (χ^2 test, $P < 0.001$, OR > 3.0) (Supplemental Fig. S9), confirming that MAE calls are reproducible and biologically regulated.

Gene Ontology (GO) enrichment analysis revealed distinct functional profiles between BAE and MAE genes (Fig. 3B). MAE genes were enriched in chromatin-related processes, nucleotide metabolism, and hematopoietic functions, whereas BAE genes were predominantly associated with structural development, RNA processing, and neurological functions including axon guidance and muscle differentiation. These results indicate that BAE and MAE genes are specialized for different biological roles. Although BAE genes appear more frequently in broad structural and developmental programs, MAE genes are notably associated with regulatory processes such as chromatin organization and specific lineage-defining pathways. This pattern is consistent with previous findings that MAE is frequently associated with genes that exhibit high epigenetic plasticity and are controlled by specific histone modifications or DNA methylation states (Wang et al. 2010; Jeffries et al. 2012; Eckersley-Maslin and Spector 2014; Gendrel et al. 2014). Such a pattern implies that MAE may play a role in fine-tuning the regulatory landscape of the cell rather than simply maintaining core cellular structures.

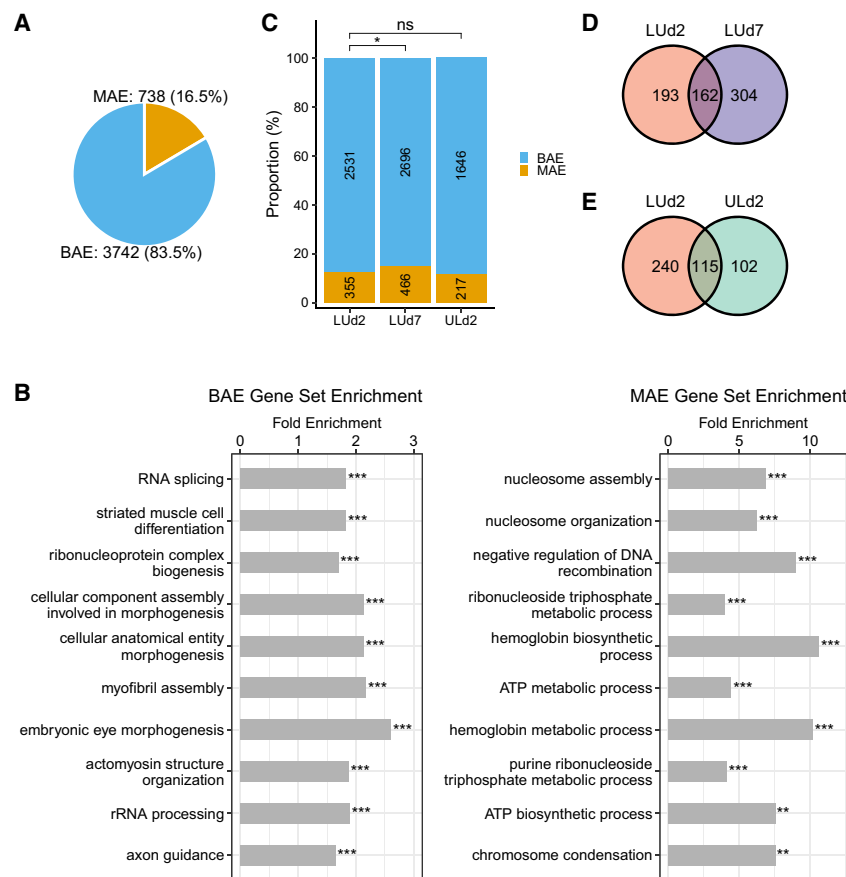


Figure 3. Overview of MAE genes. (A) Pie chart showing the number of genes classified into each allelic expression category per sample. Blue denotes BAE genes, and yellow denotes MAE genes. (B) Gene Ontology enrichment analysis of BAE and MAE genes. (C) Stacked bar plot illustrating the distribution of BAE and MAE genes across different experimental groups. The P -value indicates differences in the proportion of MAE genes between groups, calculated using the χ^2 test (LUD2 vs. LUD7: $P=0.019$; LUD2 vs. ULd2: $P=1$). (D,E) Venn diagrams showing the overlap of MAE genes between different experimental groups. The P -value indicates whether the observed overlap exceeds random expectation, assessed by the hypergeometric test (D, LUD2 vs. LUD7: $P=2.23 \times 10^{-68}$; E, LUD2 vs. ULd2: $P=3.12 \times 10^{-61}$).

We further investigated the MAE gene pattern in different experimental groups. Individuals from the same experimental group exhibited significantly higher overlap in MAE gene identity ($\sim 46.3\%$) compared with those from different groups ($\sim 19.5\%$), as expected (Supplemental Fig. S10). The proportion of MAE genes increased from 12.3% (355/2886) at 2 dpf to 14.7% (466/3162) at 7 dpf (χ^2 test, $P=0.019$) (Fig. 3C), showing a cross-specific developmental trend that is consistent with observations reported in the mouse (Eckersley-Maslin et al. 2014). The overlap between MAE genes at 2 dpf and 7 dpf comprised 162 genes with 4.0-fold enrichment (hypergeometric test, $P=2.23 \times 10^{-68}$), functionally enriched in structural and cytoskeletal functions, particularly epidermal integrity and intermediate filament organization (Fig. 3D; Supplemental Fig. S11A). To assess whether these shared MAE genes maintain consistent allelic preferences across developmental stages, we examined the concordance of allelic bias at the SNV level. Among MAE sites shared between LUD2 and LUD7, 88.1% (512/581) exhibited the same allelic bias (TL or TU), indicating that most MAE events preserve their allelic preference during development (Supplemental Fig. S12). We examined the unique MAE genes at each time point and found that 2 dpf-specific

MAE genes were enriched in oxygen binding and carrying, whereas 7 dpf-specific MAE genes were enriched in protoporphyrinogen biosynthetic and metabolic, heme biosynthesis, redox control, and ion transport, suggesting developmental stage-specific demands for heme-related metabolic process.

Although fewer genes were detected in the reciprocal cross owing to sequencing quality limitations, the proportion of MAE genes remained comparable between the two cross directions at 2 dpf (12.3% vs. 11.6%; χ^2 test, $P=1$). The overlapping MAE genes numbered 115 with 5.1-fold enrichment (hypergeometric test, $P=3.12 \times 10^{-61}$) (Fig. 3E). The genes were enriched in hemoglobin-related and regulation of lipid metabolism functions, suggesting that their allelic expression patterns were largely independent of parental genetic background (Supplemental Fig. S11B). Additionally, we assessed the concordance of allelic bias in these cross-direction-shared MAE genes. Among MAE sites shared between LUD2 and ULd2, 95.4% (350/367) showed the same allelic bias, indicating that MAE events shared across reciprocal crosses generally maintain consistent allelic preference regardless of cross direction (Supplemental Fig. S12). Seventy-four MAE genes were consistently detected across all three experimental groups, and these genes were enriched in cytoskeletal organization, protein synthesis machinery, lipid transport, chromatin packaging, and oxygen transport, suggesting that they exhibit constitutive MAE.

Together, these observations underscore that MAE is not randomly distributed across the transcriptome but instead reflects a developmental regulation that may contribute to tissue specialization and phenotypic diversity.

MAE genes are enriched for tissue-specific expression

To understand the biological implications of allelic imbalance, we asked whether MAE contributes to cell lineage identity and lineage commitment. We further investigated tissue-specific expression patterns of these MAE genes. Using single-cell transcriptomes, 50,819 cells were delineated into 37 distinct tissues through Uniform Manifold Approximation and Projection (UMAP [McInnes et al. 2018]; Methods) (Fig. 4A; Supplemental Table S2). For each cell type, we defined high-confidence tissue markers (top 50 genes per cluster), yielding 1627 unique marker genes. Tissue-type marker genes are operational proxies for lineage identity and often include master regulators and pathway gatekeepers whose small dosage shifts can trigger discrete state transitions. If MAE contributes to lineage commitment, we would expect an overrepresentation of MAE among markers. We then examined whether bulk-identified MAE genes align with tissue-specific expression

patterns. Among the 738 MAE genes, 138 (18.7%) overlapped with tissue marker genes compared with only 13.3% (497 out of 3742) for BAE genes (χ^2 test, $P=0.00015$) (Fig. 4B; Supplemental Table S3), indicating that MAE genes show a modest but statistically significant enrichment among tissue marker genes and that some MAE genes may be associated with lineage-specifying regulators.

We then examined the overlap between MAE genes and tissue marker genes for each tissue type (Fig. 4C). Several tissues showed a particularly high proportion of monoallelic expressed marker genes. In particular, the epidermis, periderm, and erythrocytes exhibited close to 10% of their marker genes as MAE, with five to seven MAE genes consistently detected across multiple individuals (Fig. 4C,D; Supplemental Fig. S13). Notably, members of the hemoglobin gene family displayed robust MAE in erythrocytes across at least four individuals, whereas *ecrg4b*, *pfn1*, and *krt91* showed similar patterns in the epidermis (Fig. 4D). These genes are associated with key tissue functions, suggesting that MAE may contribute to the regulation of lineage-defining genes. These findings strongly support that MAE genes exhibit preferential tissue-specific expression patterns.

Tissue context shapes allele choice in MAE genes

Having observed that certain tissues show elevated MAE among their marker genes, we next asked whether the same MAE genes exhibit distinct allelic preferences across different tissues within the same individual. To address this, we compared allelic expression patterns between different tissues using statistical analysis (Methods). We applied stringent criteria to identify biologically meaningful differences, requiring both statistical significance ($P<0.05$) and substantial effect sizes (Fig. 5A). To ensure robustness, we tested multiple effect-size cutoffs (top 10%, 5%, 1%, and 0.5% of absolute TL ratio differences). The number of SNVs showing significant allelic preference across tissue pairs under these thresholds are highly correlated with those from the original top 1% cutoff (Spearman's ρ , top 1% vs. 10%, $\rho=0.61$, $P=1.4 \times 10^{-19}$; vs. 5%, $\rho=0.76$, $P=3.4 \times 10^{-34}$; vs. 0.5%, $\rho=0.82$, $P=7 \times 10^{-46}$), confirming that our findings are robust to parameter choices (Supplemental Fig. S14). Our analysis revealed that certain tissues exhibit distinct allelic expression patterns compared to others within the same individuals (Fig. 5B; Supplemental Fig. S15). Periderm tissue frequently exhibited distinct allelic expression patterns compared with other tissues within the same individual, particularly in comparisons with fin epithelia.

A striking example was observed in the *txnb* gene (ENSDARG0000044125), which encodes thioredoxin, a protein

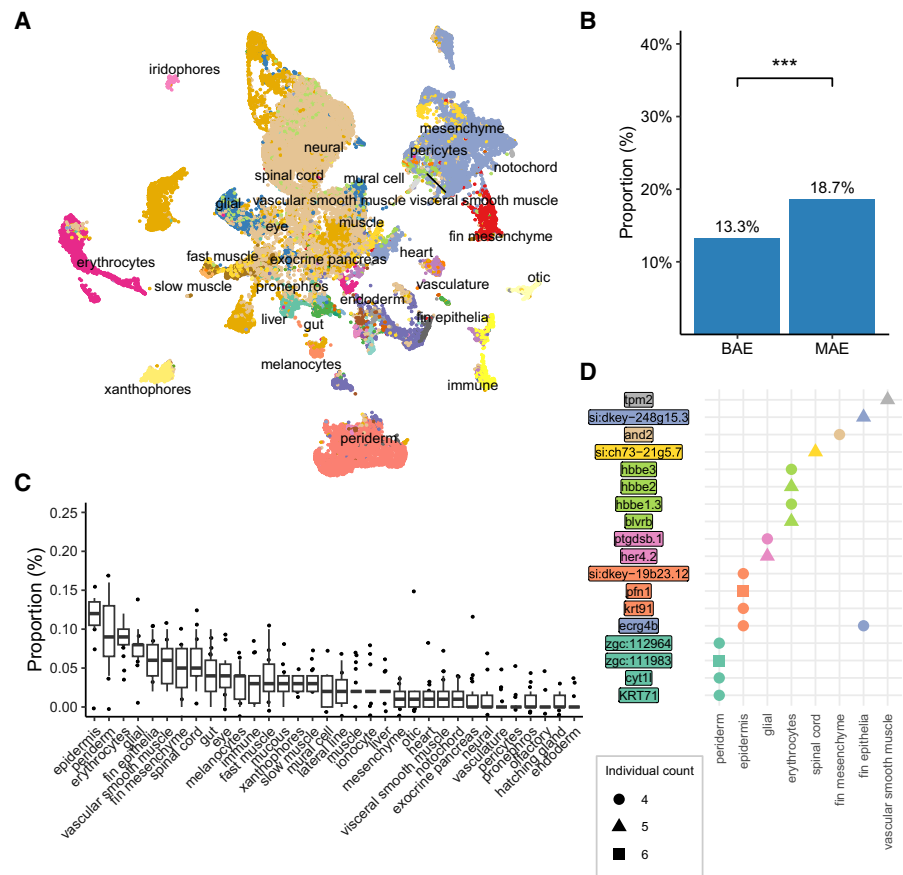


Figure 4. Tissue specificity of MAE genes. (A) UMAP visualization of single-cell transcriptomes. Colors correspond to cell-type annotations. (B) Bar plot showing the proportions of marker genes in BAE and MAE genes. The P -value indicates differences in proportions, calculated using the χ^2 test ($P=0.00015$). (C) Box plot showing the number of MAE marker genes per tissue across samples. Each dot represents an individual. (D) Dot plot of MAE marker genes detected in at least four individuals. Dot shape represents the number of individuals in which the gene exhibits MAE, and color corresponds to the tissue counts that SNVs display expression. Only tissues with four or more individuals are shown; others are included in Supplemental Figure S13.

critical for cellular stress responses. Although most tissues showed balanced or slightly TL-preferential expression, glial tissues consistently showed TL allele ratios approaching zero, signifying a strong and reproducible TU allele bias (Fig. 5C; Supplemental Fig. S16). This pattern remained robust after downsampling to equalize the number of SNV-detected cells across tissues (Methods) (Supplemental Fig. S17). This glial-specific bias may reflect distinct regulatory mechanisms governing *txnb* expression. Given the critical role of *txnb* in redox homeostasis and cellular stress responses, particularly in the nervous system, such MAE may fulfill specialized functional requirements of glial cells. These observations point toward a unique transcriptional or epigenetic landscape in glia that influences allelic choice, potentially linked to the metabolic and oxidative demands of these cells during neural development and maintenance. A similar pattern was observed for the *and2* gene (ENSDARG0000079302) in periderm tissues, in which tissue-specific allelic bias was also detected (Supplemental Fig. S18).

We then examined allele preferences of MAE SNVs across multiple tissues (Methods). Among the analyzed SNVs, 827 displayed allelic preference in no specific tissue, whereas 1522 SNVs

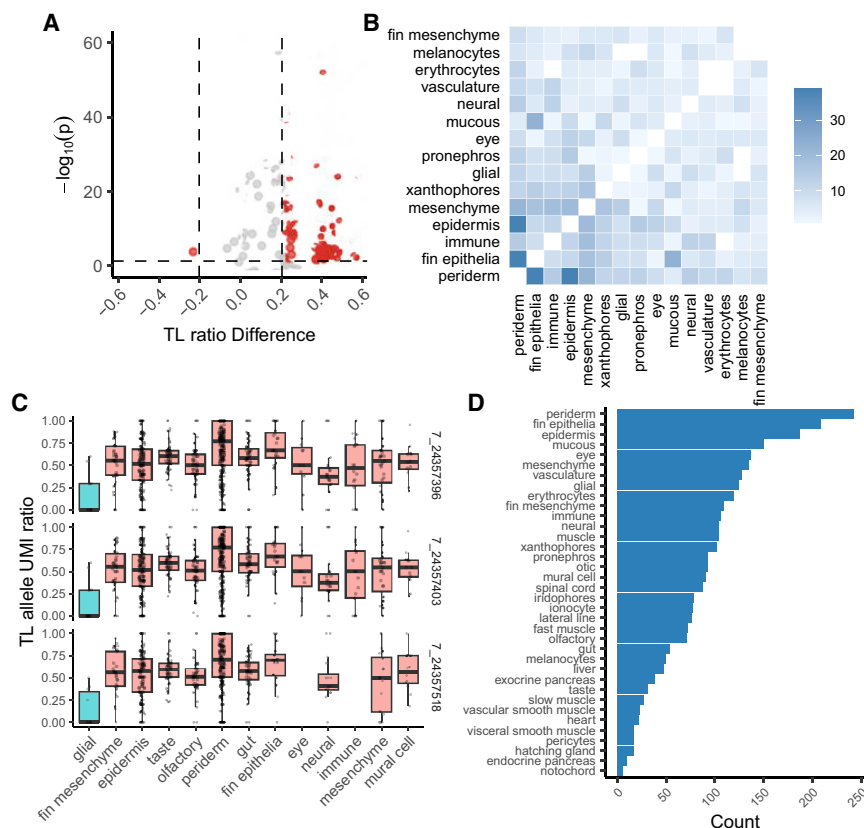


Figure 5. Tissue-specific allelic bias in gene expression. (A) Volcano plot illustrating differences in TL allele ratios between tissue pairs. Each point represents a gene-tissue comparison. The x-axis denotes the mean difference in TL ratio, and the y-axis represents the $-\log_{10}P$ from Wilcoxon rank-sum tests. Dashed lines indicate the significance thresholds: $|\text{mean TL ratio}| > 0.204$ and BH-adjusted P -values < 0.05 . (B) Heatmap summarizing the number of SNVs showing significant allelic bias across tissue pairs. Color intensity reflects the number of SNVs exceeding both statistical and biological thresholds. Only the top 15 tissue pair are displayed, and the complete results are provided in Supplemental Figure S14. (C) TL allele ratios of multiple SNVs within the *txnb* gene (ENSDARG00000044125) across tissues in sample LU4. Each point represents a single cell. Glial cells (light blue) exhibit reduced TL ratios relative to other tissues (pink), indicating glia-specific allelic preference for the TU allele. (D) Bar plot showing the number of genes for which each tissue displays a unique allelic bias.

exhibited allelic preference in at least one tissue. The majority of MAE SNVs exhibited allelic preference. Consistent with earlier observations, the periderm, fin epithelia, and epidermis were the most frequent tissue outliers, further emphasizing their distinct allelic regulatory profiles (Fig. 5D). These findings demonstrate that tissue-specific regulatory programs can influence which parental allele is preferentially expressed, adding another layer of complexity to gene regulation during development.

Discussion

This study presents a comprehensive single-cell atlas of MAE spanning 37 tissues throughout the zebrafish body, providing novel insights into the prevalence, functional roles, and tissue specificity of allelic biases across diverse cell types. Through systematic profiling of allele-specific expression patterns at genome-wide scale, we discovered that ~16.5% of evaluable genes exhibit MAE. These genes demonstrated nonrandom spatial distribution patterns and significant functional enrichment. Leveraging about 50,000 single-cell transcriptomes, we further uncovered pronounced tissue-specific allele preference, particularly in the periderm,

fin epithelia, and epidermis. Previous investigations in mammals, including humans and mice, have largely relied on clonal cell populations or bulk RNA-seq data, which limits resolution and fails to capture the true cellular heterogeneity of MAE (Gimelbrant et al. 2007; Zwemer et al. 2012; Eckersley-Maslin et al. 2014; Kravitz et al. 2023; Quan et al. 2024). Our research extends this work by enabling allele-specific expression analysis at single-cell resolution across the entire organism, thereby capturing spatially resolved MAE patterns that were previously inaccessible.

Traditional analytical approaches have operated under the assumption of biallelic gene expression; the relative abundance of transcripts from two alleles typically follows a binomial distribution with a fixed allelic probability $f=0.5$, reflecting equal expression from both alleles. In comparison, MAE genes are characterized by allelic imbalance that deviates from this expectation, which may be modeled by a binomial distribution with unequal probability ($f \neq 0.5$) (Eckersley-Maslin et al. 2014) or, more flexibly, by a beta-binomial distribution (Kravitz et al. 2023) to account for overdispersion and cellular heterogeneity. However, these approaches do not explicitly capture allelic preference across individual cells and are not well suited for modeling MAE at single-cell resolution. To address these limitations, we implemented an MLE framework specifically designed to capture allelic expression fluctuations at the single-cell level. Crucially, this framework operationalizes the distinction between constitutive

directional MAE (model 1) and dynamic random MAE (model 2), directly quantifying per-cell allelic heterogeneity. Our analysis revealed that 85.5% of SNVs exhibited balanced biallelic or biased biallelic expression patterns. Among the remaining SNVs, the majority (74.2%) displayed random MAE, whereas only 25.8% showed consistent directional MAE (Supplemental Fig. S6C). This demonstrates that transient, bursting-driven MAE is the prevalent form in vivo, which prior bulk-based models could not resolve.

We also identified some SNVs that displayed consistent allelic preferences, with these preferences being tissue-type dependent. Notable examples include three SNVs in the *txnb* gene specifically expressed in glial cells and three SNVs in the *and2* gene preferentially expressed in periderm tissue. These findings further support the notion that MAE contributes to fine-tuning gene expression in a cell-type-dependent manner, suggesting a more sophisticated regulatory mechanism than previously appreciated. These loci exemplify tissue-coupled constitutive MAE, indicating a stable allele preference imposed by local regulatory context.

Despite these insights, one limitation should be acknowledged. One notable constraint involves the detection of imprinted

genes, genes that represent an important class of MAE genes (Johnson 1974; Mann and Lovell-Badge 1984; McGrath and Solter 1984; Surani et al. 1984; Cattaneach and Kirk 1985; Peters 2014; Monk et al. 2019). Imprinted genes exhibit consistent parent-of-origin-dependent expression bias, persistently favoring either the paternal or maternal allele across individuals. Taking advantage of the reciprocal TL $\times$ TU and TU $\times$ TL crosses at 2 dpf, we explicitly searched for such parent-of-origin effects by identifying genes that fit the M1 model while exhibiting opposite allelic preference between reciprocal crosses. Under this operational definition, these genes would be strong candidates for genomic imprinting. However, we did not detect any genes satisfying this pattern, which we attribute primarily to the limited number of individuals and restricted allelic coverage available for robust cross-direction comparisons.

We further examined candidate imprinted genes curated from the literature. Although genomic imprinting has been extensively characterized in humans and mice, evidence in zebrafish remains scarce and fragmentary. In addition, only a small fraction of MAE genes is conserved between species (~22% overlap between human and mouse MAE genes) (Zwemer et al. 2012), indicating substantial evolutionary divergence and limiting the direct transferability of mammalian imprinting patterns to zebrafish. For example, DNMT3 family genes exhibit imprinting in humans but not in zebrafish (Smith et al. 2011), and in our data, these genes either showed biallelic expression or lacked sufficient coverage in reciprocal crosses to evaluate parent-of-origin effects. In addition, because our analyses were based on a standard reference-based alignment framework, subtle reference-mapping bias in allele-specific quantification cannot be fully excluded, although benchmarking with a TL pseudoreference genome suggested that its overall impact is limited. Given these limitations, the present study focuses on characterizing general random MAE patterns. A systematic analysis of imprinting and other MAE subtypes will require substantially larger sample sizes, higher allelic coverage, and further methodological refinements in future studies.

In conclusion, this study establishes a foundational resource for allele-specific expression in zebrafish at single-cell resolution and underscores the biological relevance of MAE in developmental- and tissue-specific contexts. By revealing the breadth and complexity of MAE across the zebrafish embryo, we provide a framework for future investigations into how allelic expression biases influence development, function, and disease.

Methods

Zebrafish husbandry

All the zebrafish were raised and maintained in standard zebrafish breeding guidelines of the Zebrafish Information Network (ZFIN) database. The adult zebrafish were kept on a 14 h light/10 h dark cycle at 28.5°C. The zebrafish research experiments conducted in this study were approved by the animal research and ethics committees of Sun Yat-Sen University.

Whole-genome sequencing and identification of genomic SNVs

Adult zebrafish were anesthetized by immersion in 0.02% MS-222 tricaine solution, waiting for ~10 min until fish gill movement was slowed. The anesthetized fish were then promptly transferred to a sterile 90 mm Petri dish, in which a small portion of the tail fin was excised using a sterile razor blade. Following fin clipping, each fish was swiftly placed in a container of fresh system water to

facilitate recovery. The excised fin clips were carefully collected using sterile tweezers and transferred to individual, clean microcentrifuge tubes. Genomic DNA (gDNA) was then extracted from the fin tissue using the QIAamp DNA micro kit (56304) according to the manufacturer's protocol. The extracted gDNA from each sample was sequenced on the Illumina NovaSeq platform at an average coverage depth of 30 $\times$ to ensure robust data quality for downstream analyses.

Reads from WGS data were preprocessed by trimming low-quality sequences and adapter contamination using Trim Galore! (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), and then the processed reads were aligned to the GRCz11 zebrafish reference genome using the BWA (Li and Durbin 2009). Duplicate reads, which could interfere with variant calling accuracy, were identified and marked using the GATK (McKenna et al. 2010) MarkDuplicates tool to minimize bias in variant frequency estimation. SNVs of each sample were called by Haplotypecaller, and joint genotyping was subsequently conducted on each parental pair using GATK GenotypeGVCFs. The low-quality SNVs (GQ < 20) was removed with SelectVariants tools. The resulting SNV set was rigorously filtered to retain only homozygous variants that were differentially expressed in paternal and maternal genomes, thus capturing variants informative for parental origin tracing. These filtered SNVs were then leveraged to assign allelic origin in the offspring, enabling precise determination of paternal and maternal contributions.

Preparation of single-cell suspensions and single-cell RNA sequencing

To investigate allele-specific gene expression and assess potential parental effects, adult zebrafish from two genetically distinct strains, TL and TU, were reciprocally crossed to introduce genome-wide heterozygosity in the offspring. Specifically, reciprocal crosses were performed between TL females $\times$ TU males (resulting in progeny LU1 to LU4) and TU females $\times$ TL males (resulting in progeny UL1 and UL2). Among these offspring, LU1, LU2, UL1, and UL2 were collected at 2 dpf, representing early embryonic stages, whereas LU3 and LU4 were collected at 7 dpf, representing later larval development.

The preparation of single-cell suspensions was conducted in accordance with established protocols described in the literature (Spanjaard et al. 2018). To improve data quality, the chorion and yolk sac of embryos must be removed, as these structures provide protection and nutrients that could affect downstream analysis. For chorion removal, embryos were incubated in a Petri dish containing 1 mg/mL Pronase solution for ~5–10 min at room temperature, with occasional gentle agitation until the chorion detached. Embryos were then transferred to a microcentrifuge tube, where the Pronase solution was removed and replaced with deysolking buffer (55 mM NaCl, 1.8 mM KCl, 1.25 mM NaHCO₃). Gentle pipetting with a pipette tip was applied until the yolk was fully detached. The tube was then centrifuged for 1 min 300g and at 4°C, and the deysolking buffer was carefully removed.

Next, 100 μ L of HBSS containing 1 $\times$ TrypLE (Thermo Fisher Scientific) was added, and the sample was incubated for ~20 min at room temperature with gentle pipette mixing every 5 min, until the embryo was fully dissociated. The cell state was examined under the microscope with 1 μ L cell suspension. When the cells were mostly singlet, 500 μ L cold HBSS (Thermo Fisher Scientific) supplemented with 0.05% BSA was added to end lysis. The suspension was then passed through a 45 μ m cell strainer and transferred into a fresh centrifuge tube. Cells were pelleted by centrifugation at 300g for 5 min at 4°C. The supernatant was carefully removed without disturbing the cell pellet, retaining ~45 μ L of residual liquid. The cell suspension was gently mixed by pipetting, and 5.5 μ L

of it was loaded onto a hemocytometer to calculate cell concentration, which was used to determine the appropriate cell loading volume.

The single-cell suspensions were processed into droplet emulsion using the Chromium platform (10x Genomics), and sequencing libraries were generated with the Chromium single-cell 3' reagent kits v2. Twelve or 13 cDNA amplification cycles were used to generate cDNA according to the loaded cell number. The indexed libraries were sequenced on the Illumina NovaSeq platform with 150 bp paired-end reads.

Single-cell RNA sequencing data analysis and cell-type annotation

The scRNA-seq raw data were processed using the Cell Ranger software (version 8.0.0) (Zheng et al. 2017), following the 10x Genomics pipeline, with reads aligned to the GRCz11 zebrafish reference genome using a standard reference-based pipeline. Filtered gene expression matrices generated by Cell Ranger were imported into R (version 4.3.0) and analyzed using the Seurat R package (Satija et al. 2015). Cells were retained for downstream analysis if they expressed a minimum of 200 genes and exhibited mitochondrial gene count below 20%.

To account for technical variability, such as differences in sequencing depth, the data were normalized using the SCTransform R package (Hafemeister and Satija 2019). Principal component analysis (PCA) and UMAP were performed with a resolution of 0.5. Potential technical artifacts, known as doublets, were identified and removed using the DoubletFinder package (McGinnis et al. 2019) to prevent confounding effects on subsequent analyses. The remaining singlet cells were renormalized using SCTransform. To address batch effects, data sets from six individuals were integrated using the IntegrateLayers function, employing a reciprocal PCA and UMAP-based integration pipeline to avoid overcorrection. The top 50 principal components were used for UMAP dimensionality reduction, and unsupervised clustering was performed using the Seurat function FindClusters with a resolution of 1.0. For each cluster, the top 50 differentially expressed genes were identified using the FindAllMarkers function, based on their elevated expression within the cluster relative to all other cells. Cell-type annotation for each cluster was determined by cross-referencing marker genes with established literature and the ZFIN database.

Detection of single-cell allele expression pattern

To obtain single-cell allele profiles, the BAM file output from Cell Ranger was further processed using the Cellsnp-lite tool (Huang and Huang 2021) incorporated a joint parental VCF file as a prior input, enabling accurate inference of genotypes in individual cells. The --genotype parameter was applied to generate an output VCF file that includes genotype calls for each single cell. The VCF file was trimmed, the allele-specific UMI counts extracted, and the data converted into a format compatible with R for downstream analysis. The resulting data were subsequently integrated with the scRNA-seq data set, facilitating simultaneous access to both gene expression profiles and allele-specific expression information for each cell. This combined data set allows for in-depth analysis of allelic variation and gene expression within individual cells.

Identification of MAE using MLE

To classify SNVs by their allelic expression pattern, we developed a hierarchical likelihood-based framework based on three biologically motivated models. Each model accounted for different degrees of allelic bias and variability across single cells and was evaluated using UMI counts assigned to the two alleles. Model 0

(M0) assumes strict biallelic expression with equal expression probability from both alleles. The likelihood is defined as

$$\mathcal{L}_0 = \sum_{i=1}^n \log[\text{Binom}(\text{Read1}_i | \text{Read1}_i + \text{Read2}_i, f = 0.5)],$$

where Read1_i and Read2_i denote UMI count for each allele in cell i , and n is the number of cells. Model 1 (M1) introduces a single free parameter $f \in (0, 1)$, representing a global allelic bias across all cells. The log-likelihood is

$$\mathcal{L}_1(f) = \sum_{i=1}^n \log[\text{Binom}(\text{Read1}_i | \text{Read1}_i + \text{Read2}_i, f)].$$

Model 2 (M2) accounts for cell-to-cell variation in allele preference by incorporating a binary direction vector $d \in \{-1, 1\}$. For each cell i , the effective allelic probability f_i is defined as

$$f_i = \begin{cases} f & \text{if } d_i = 1 (\text{TL allele favored}) \\ 1 - f & \text{if } d_i = -1 (\text{TU allele favored}) \end{cases},$$

yielding the overall log-likelihood,

$$\mathcal{L}_2(f, d) = \sum_{i=1}^n \log[\text{Binom}(\text{Read1}_i | \text{Read1}_i + \text{Read2}_i, f_i)].$$

Because the likelihood of a bimodal allelic ratio distribution is symmetric with respect to f and $1 - f$, the model without a direction vector suffers from a label-switching problem, in which MLE may arbitrarily converge to either f or $1 - f$. The latent direction parameter d simply determines which peak is treated as dominant, collapsing the symmetric optimization landscape into a single well-defined solution. As a discrete variable necessitated by this symmetry, parameter d is not estimated as an independent continuous parameter and does not increase the model's degrees of freedom or contribute to overfitting risk, even under modest UMI counts.

To initialize d , we ranked cell-specific allelic ratios ($\text{Read1}_i / (\text{Read1}_i + \text{Read2}_i)$) and assigned directions based on a quantile threshold q . Parameters f and q were optimized using the L-BFGS-B algorithm with constraints on $f \in (0, 1)$. The parameters were set as default convergence criteria: The control parameter factor was set to 1×10^7 (providing a convergence tolerance of approximately 1×10^{-8}), pgtol was set to zero, and the maximum number of iterations (maxit) was 100. To avoid local optima in the more complex model 2 (random MAE), we implemented a grid-search initialization (sampling f and q from 10^{-5} to $1 - 10^{-5}$) to provide the most robust starting parameters for the optimizer. Model comparison was performed using likelihood ratio tests, and the best-fitting model was selected for each SNV. To control for multiple hypothesis testing, Bonferroni correction was applied to adjust significance thresholds.

To validate that our MLE framework reliably captures population-level allelic skew when present (i.e., model 1; directional monoallelic expression, MAE), we performed a pseudobulk analysis. Briefly, for each SNV, we aggregated allele-specific UMI counts across all cells to obtain pseudobulk reference and alternative allele counts. We then computed the pseudobulk allelic fraction.

The null model (model 0) assumes perfect biallelic expression with a fixed allelic probability $f = 0.5$. Model 1 introduces a free parameter $f \in (0, 1)$, representing the optimized allelic bias across all cells. Deviation from 0.5 indicates imbalance, and based on this, we further classified model 1 results into biased biallelic ($0.2 < f < 0.8$) and directional monoallelic ($f \geq 0.8$ or $f \leq 0.2$) expression. The threshold was chosen based on established criteria reported in previous studies (Eckersley-Maslin et al. 2014; Xu et al. 2017). Model 2 extends this formulation by introducing a direction vector $d \in \{-1, 1\}$, which allows the preferred allele to vary across cells

and indicates the hallmark feature of MAE, stochastic allelic choice. In summary, we assigned each SNV to two expression patterns: biallelic (containing balanced biallelic [model 0], biased biallelic [model 1 with $0.2 < f < 0.8$]) and monoallelic (containing directional monoallelic [model 1 with $f \geq 0.8$ or $f \leq 0.2$] and random monoallelic [model 2]).

SNV distance and enrichment analysis

To assess the spatial distribution differences between MAE and BAE SNVs, we first calculated inter-SNV distances. Adjacent MAE sites located within 150 bp, which are likely originating from the same sequencing read or representing local transcriptional events, were excluded to avoid overestimating clustering owing to technical artifacts. Next, to evaluate whether MAE SNVs exhibit clustering beyond what would be expected from underlying genomic architecture, we employed a permutation-based null model. Specifically, MAE labels were randomly reassigned among all expressed SNVs while preserving chromosomal positions, individual identity, and overall SNV density. We then visualized the empirical cumulative distribution function (ECDF) of the inter-SNV distances for both categories to evaluate differences in genomic clustering.

To determine whether MAE SNVs exhibit regional enrichment across the genome, we implemented a sliding window approach. For each chromosome, we scanned the genome using 1 Mb windows with a 10 kb step size. Within each window, we quantified both the total number of SNVs and the number of MAE SNVs and subsequently calculated the proportion of MAE SNVs relative to total SNVs. Windows lacking any variants or exhibiting redundant MAE patterns across individuals were excluded from further analysis.

Gene annotation and GO enrichment analysis

To evaluate the expression type in gene level, we mapped the chromosomal positions of SNVs to their corresponding gene symbols using the GenomicRanges R package. Because some genes harbor more than one SNV, to assess intragene consistency in allelic expression patterns, we calculated the proportion of SNVs exhibiting the dominant expression model (MAE or BAE) relative to the total number of informative SNVs of these genes. To retain as much informative data as possible while minimizing ambiguity, we implemented a filtering strategy for genes exhibiting conflicting expression patterns. Specifically, for each such gene, we calculated the ratio of cells supporting the dominant expression patterns to those supporting the alternative. Genes with a support ratio exceeding a twofold threshold were assigned to the dominant model. Those failing below this threshold were excluded from downstream analyses owing to model ambiguity. This filtering retained ~70% of genes with mixed SNV patterns for further analysis.

To assess the biological significance of the identified gene sets, GO enrichment analysis was conducted using the clusterProfiler R package (Yu et al. 2012), with the default background implemented in the analysis pipeline. This analysis maps input gene lists to GO terms, which describe the biological processes, molecular functions, and cellular components associated with each gene. By identifying significantly enriched GO terms, the analysis provided insights into the functional categories and biological pathways influenced by the SNVs, enhancing our understanding of their roles within the data set.

Assessment of MAE enrichment among tissue-specific marker genes

To evaluate whether MAE genes are significantly enriched among tissue-specific marker genes, we intersected the MAE and BAE gene

sets with the top 50 marker genes for each tissue, as previously defined based on differential expression analysis from the single-cell RNA-seq atlas. We then constructed a 2×2 contingency table categorizing genes by MAE versus BAE status and marker versus non-marker designation. A χ^2 test of independence was applied to assess whether marker gene status is significantly associated with allelic expression mode (i.e., MAE or BAE).

Pairwise tissue comparisons of TL allele ratios for MAE SNVs

To investigate tissue-specific variation in TL allele ratios of MAE SNVs, we conducted a series of pairwise nonparametric statistical tests across tissues. For each SNV within each individual, we first retained only those with TL allele ratio measurements in at least two distinct tissues and detected in 10 or more cells to ensure sufficient coverage and statistical power. All possible pairwise tissue comparisons were then performed using the Wilcoxon rank-sum test (Mann–Whitney U test). To correct for multiple testing within each SNV, Benjamini–Hochberg (BH) adjustment was applied to the resulting P -values. To ensure both statistical significance and biological relevance, we implemented a dual-filtering criterion: Tissue comparisons were retained only if the adjusted P -value < 0.05 and if the absolute difference in mean TL ratios exceeded 0.204. This TL ratio threshold was empirically determined based on the top 1% of all observed intertissue TL ratio differences across SNVs, chosen to highlight the most pronounced allelic shifts.

To further identify global and tissue-specific trends in TL allele ratios, we employed a two-step analytical framework based on analysis of variance (ANOVA) and median absolute deviation (MAD) to the filtered SNVs. For each SNV, we first calculated the median TL allele ratio per tissue and counted the number of tissues with available measurements. SNVs detected in fewer than three tissues were excluded from ANOVA testing owing to insufficient statistical power. For SNVs with data across three or more tissues, we performed a one-way ANOVA on the median TL ratios to assess intertissue variability. We implemented a simplified linear model by randomly splitting the tissues into two halves and computing the ANOVA P -value. The resulting P -values were again adjusted using the BH method. If adjusted P -value < 0.05 and only three tissues were available, we applied a MAD-based outlier detection. Any tissue deviating by more than two times the MAD from the median was labeled as an outlier. If the adjusted P -value was significant and more than three tissues were available, the SNV was considered to exhibit allelic preference across two or more distinct tissue groups, indicating a systematic, group-specific tissue bias toward either the TL or TU allele. If ANOVA was not significant but at least one tissue exhibited a MAD-defined outlier, the SNV was still classified as having sporadic tissue-specific allelic deviation. All other SNVs that did not meet the criteria above were considered to exhibit no clear or uniform pattern of tissue-specific allelic expression.

To address the potential impact of unequal numbers of SNV-detected cells across tissues, we performed a capped cell-level downsampling analysis for tissues with large numbers of SNV-detected cells in *txnb* gene. Specifically, for each tissue, we retained all cells when the number of cells with detected SNVs was 20 or fewer and randomly sampled 20 cells when the number exceeded 20.

Statistical analysis

All statistical analyses in this study were performed using bash and R (R Core Team 2025). Student's t -test or Wilcoxon were utilized to investigate the difference between two groups. A significance threshold of 0.05 was applied to determine statistical

significance and the following statistical significance indicators are used: (*) $P < 0.05$, (**) $P < 0.01$, and (***) $P < 0.001$.

Data access

The raw data generated in this study have been submitted to the China National Center for Bioinformation (<https://ngdc.cncb.ac.cn/bioproject>) under accession number PRJCA044959. The related codes are available at GitHub (<https://github.com/zengh58/ZebrafishMonoallelic/>) and as [Supplemental Code](#).

Competing interest statement

The authors declare no competing interests.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (32293190, 32500524, 32470647, 32530042, 12526210), Guangdong Basic and Applied Basic Research Foundation (2025A1515011510), and the National Key R&D Program of China (2024YFC3407100).

Author contributions: X.H., L.L., and S.D. conceived the study. L.L. and S.D. supervised the study. H.Z. and Z.L. conducted the experiments and analyzed the data. H.X., D.Z., and B.L. performed data curation. L.L. and H.Z. wrote the manuscript with contributions from all authors.

References

- Bix M, Locksley RM. 1998. Independent and epigenetic regulation of the interleukin-4 alleles in CD4⁺ T cells. *Science* **281**: 1352–1354. doi:10.1126/science.281.5381.1352
- Borel C, Ferreira PG, Santoni F, Delaneau O, Fort A, Popadin KY, Garieri M, Falconnet E, Ribaux P, Guipponi M, et al. 2015. Biased allelic expression in human primary fibroblast single cells. *Am J Hum Genet* **96**: 70–80. doi:10.1016/j.ajhg.2014.12.001
- Cattanach BM, Kirk M. 1985. Differential activity of maternally and paternally derived chromosome regions in mice. *Nature* **315**: 496–498. doi:10.1038/315496a0
- Chess A. 2016. Monoallelic gene expression in mammals. *Annu Rev Genet* **50**: 317–327. doi:10.1146/annurev-genet-120215-035120
- Chess A, Simon I, Cedar H, Axel R. 1994. Allelic inactivation regulates olfactory receptor gene expression. *Cell* **78**: 823–834. doi:10.1016/S0092-8674(94)90562-2
- Deng Q, Ramsköld D, Reinius B, Sandberg R. 2014. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**: 193–196. doi:10.1126/science.1245316
- Eckersley-Maslin MA, Spector DL. 2014. Random monoallelic expression: regulating gene expression one allele at a time. *Trends Genet* **30**: 237–244. doi:10.1016/j.tig.2014.03.003
- Eckersley-Maslin MA, Thybert D, Bergmann JH, Marioni JC, Flicek P, Spector DL. 2014. Random monoallelic gene expression increases upon embryonic stem cell differentiation. *Dev Cell* **28**: 351–365. doi:10.1016/j.devcel.2014.01.017
- Fu R, Qin P, Zou X, Hu Z, Hong N, Wang Y, Jin W. 2021. A comprehensive characterization of monoallelic expression during hematopoiesis and leukemogenesis via single-cell RNA-sequencing. *Front Cell Dev Biol* **9**: 702897. doi:10.3389/fcell.2021.702897
- Gendrel A-V, Attia M, Chen C-J, Diabangouaya P, Servant N, Barillot E, Heard E. 2014. Developmental dynamics and disease potential of random monoallelic gene expression. *Dev Cell* **28**: 366–380. doi:10.1016/j.devcel.2014.01.016
- Gimelbrant A, Hutchinson JN, Thompson BR, Chess A. 2007. Widespread monoallelic expression on human autosomes. *Science* **318**: 1136–1140. doi:10.1126/science.1148910
- Hafemeister C, Satija R. 2019. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* **20**: 296. doi:10.1186/s13059-019-1874-1
- Held W, Roland J, Raulet DH. 1995. Allelic exclusion of Ly49-family genes encoding class I MHC-specific receptors on NK cells. *Nature* **376**: 355–358. doi:10.1038/376355a0
- Holländer GA, Zuklys S, Morel C, Mizoguchi E, Mobisson K, Simpson S, Terhorst C, Wishart W, Golan DE, Bhan AK, et al. 1998. Monoallelic expression of the interleukin-2 locus. *Science* **279**: 2118–2121. doi:10.1126/science.279.5359.2118
- Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L, et al. 2013. The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**: 498–503. doi:10.1038/nature12111
- Huang X, Huang Y. 2021. Cellsnp-lite: an efficient tool for genotyping single cells. *Bioinformatics* **37**: 4569–4571. doi:10.1093/bioinformatics/btab358
- Jeffries AR, Perfect LW, Ledderose J, Schalkwyk LC, Bray NJ, Mill J, Price J. 2012. Stochastic choice of allelic expression in human neural stem cells. *Stem Cells* **30**: 1938–1947. doi:10.1002/stem.1155
- Johnson DR. 1974. Hairpin-tail: a case of post-reductional gene action in the mouse egg. *Genetics* **76**: 795–805. doi:10.1093/genetics/76.4.795
- Kravitz SN, Ferris E, Love MI, Thomas A, Quinlan AR, Gregg C. 2023. Random allelic expression in the adult human body. *Cell Rep* **42**: 111945. doi:10.1016/j.celrep.2022.111945
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**: 1754–1760. doi:10.1093/bioinformatics/btp324
- Li SM, Valo Z, Wang J, Gao H, Bowers CW, Singer-Sam J. 2012. Transcriptome-wide survey of mouse CNS-derived cells reveals monoallelic expression within novel gene families. *PLoS One* **7**: e31751. doi:10.1371/journal.pone.0031751
- Liang W, Zou X, Li G, Zhou S, Tian C, Schaefer B. 2021. Systematic analysis of monoallelic gene expression and chromatin accessibility across multiple tissues in hybrid mice. *Front Cell Dev Biol* **9**: 717555. doi:10.3389/fcell.2021.717555
- Mann JR, Lovell-Badge RH. 1984. Inviability of parthenogenones is determined by pronuclei, not egg cytoplasm. *Nature* **310**: 66–67. doi:10.1038/310066a0
- McGinnis CS, Murrow LM, Gartner ZJ. 2019. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst* **8**: 329–337.e4. doi:10.1016/j.cels.2019.03.003
- McGrath J, Solter D. 1984. Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* **37**: 179–183. doi:10.1016/0092-8674(84)90313-1
- McInnes L, Healy J, Melville J. 2018. UMAP: Uniform Manifold Approximation and Projection for dimension reduction. *arXiv:1802.03426 [stat.ML]*. doi:10.48550/arXiv.1802.03426
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**: 1297–1303. doi:10.1101/gr.107524.110
- Monk D, Mackay DJG, Eggermann T, Maher ER, Riccio A. 2019. Genomic imprinting disorders: lessons on how genome, epigenome and environment interact. *Nat Rev Genet* **20**: 235–248. doi:10.1038/s41576-018-0092-0
- Pernis B, Chiappino G, Kelus AS, Gell PGH. 1965. Cellular localization of immunoglobulins with different allotypic specificities in rabbit lymphoid tissues. *J Exp Med* **122**: 853–876. doi:10.1084/jem.122.5.853
- Peters J. 2014. The role of genomic imprinting in biology and disease: an expanding view. *Nat Rev Genet* **15**: 517–530. doi:10.1038/nrg3766
- Prashant NM, Liu H, Bousounis P, Spurr L, Alomran N, Ibeawuchi H, Sein J, Reece-Stremtan D, Horvath A. 2020. Estimating the allele-specific expression of SNVs from 10× genomics single-cell RNA-sequencing data. *Genes (Basel)* **11**: 240. doi:10.3390/genes11030240
- Quan J, Yang M, Wang X, Cai G, Ding R, Zhuang Z, Zhou S, Tan S, Ruan D, Wu J, et al. 2024. Multi-omic characterization of allele-specific regulatory variation in hybrid pigs. *Nat Commun* **15**: 5587. doi:10.1038/s41467-024-49923-5
- R Core Team. 2025. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org/>.
- Reinius B, Sandberg R. 2015. Random monoallelic expression of autosomal genes: stochastic transcription and allele-level regulation. *Nat Rev Genet* **16**: 653–664. doi:10.1038/nrg3888
- Reinius B, Mold JE, Ramsköld D, Deng Q, Johnsson P, Michaëlsson J, Frisén J, Sandberg R. 2016. Analysis of allelic expression patterns in clonal somatic cells by single-cell RNA-seq. *Nat Genet* **48**: 1430–1435. doi:10.1038/ng.3678
- Rodriguez I, Feinstein P, Mombaerts P. 1999. Variable patterns of axonal projections of sensory neurons in the mouse vomeronasal system. *Cell* **97**: 199–208. doi:10.1016/S0092-8674(00)80730-8
- Satija R, Farrell JA, Gennert D, Schier AF, Regev A. 2015. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* **33**: 495–502. doi:10.1038/nbt.3192

- Savova V, Chun S, Sohail M, McCole RB, Witwicki R, Gai L, Lenz TL, Wu C, Sunyaev SR, Gimelbrant AA. 2016. Genes with monoallelic expression contribute disproportionately to genetic diversity in humans. *Nat Genet* **48**: 231–237. doi:10.1038/ng.3493
- Smith THL, Collins TM, McGowan RA. 2011. Expression of the *dnmt3* genes in zebrafish development: similarity to *Dnmt3a* and *Dnmt3b*. *Dev Genes Evol* **220**: 347–353. doi:10.1007/s00427-010-0347-z
- Spanjaard B, Hu B, Mitic N, Olivares-Chauvet P, Janjuha S, Ninov N, Junker JP. 2018. Simultaneous lineage tracing and cell-type identification using CRISPR/Cas9-induced genetic scars. *Nat Biotechnol* **36**: 469–473. doi:10.1038/nbt.4124
- Surani MA, Barton SC, Norris ML. 1984. Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature* **308**: 548–550. doi:10.1038/308548a0
- Wang J, Valo Z, Bowers CW, Smith DD, Liu Z, Singer-Sam J. 2010. Dual DNA methylation patterns in the CNS reveal developmentally poised chromatin and monoallelic expression of critical genes. *PLoS One* **5**: e13843. doi:10.1371/journal.pone.0013843
- Xu J, Carter AC, Gendrel A-V, Attia M, Loftus J, Greenleaf WJ, Tibshirani R, Heard E, Chang HY. 2017. Landscape of monoallelic DNA accessibility in mouse embryonic stem cells and neural progenitor cells. *Nat Genet* **49**: 377–386. doi:10.1038/ng.3769
- Yu G, Wang L-G, Han Y, He Q-Y. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**: 284–287. doi:10.1089/omi.2011.0118
- Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, et al. 2017. Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**: 14049. doi:10.1038/ncomms14049
- Zwemer LM, Zak A, Thompson BR, Kirby A, Daly MJ, Chess A, Gimelbrant AA. 2012. Autosomal monoallelic expression in the mouse. *Genome Biol* **13**: R10. doi:10.1186/gb-2012-13-2-r10

Received August 21, 2025; accepted in revised form May 15, 2026.