Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Long-Read RNA Sequencing Reveals Tissue-Specific Expression and Transcript Boundary Variation of Mouse Prxl2b

Preeti Pilania
National Dairy Research Institute
Sikander Saini
National Dairy Research Institute
Diksha Sethi
National Dairy Research Institute
Sameer Kumar
National Dairy Research Institute
Prem Kumar
National Dairy Research Institute
Suman Chopra
National Dairy Research Institute
Satish Kumar
National Dairy Research Institute
Dhruba Malakar
dhrubamalakar@gmail.com
National Dairy Research Institute

Research Article

Keywords:

Posted Date: June 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9953807/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Alternative splicing and transcript boundary variation contribute substantially to transcriptome complexity in mammalian genomes. Prxl2b (Peroxiredoxin-like 2B), also known as prostaglandin F synthase (PGFS), encodes an enzyme with both antioxidant and prostaglandin-metabolizing activities. The protein catalyzes the conversion of prostaglandin H2 (PGH2) to prostaglandin F2α (PGF2α) and prostamide H2 to prostamide F2α, thereby contributing to prostaglandin and prostamide biosynthesis, inflammatory signaling, reproductive processes, and cellular redox homeostasis. Despite its important biological functions, the transcript architecture of Prxl2b across mouse tissues remains incompletely characterized. In this study, publicly available ENCODE long-read RNA-seq datasets generated using PacBio sequencing technology were analyzed to characterize Prxl2b expression patterns and transcript architecture.

The reads were aligned to the mouse reference genome (mm10), and transcript models were reconstructed using IsoQuant and validated using FLAIR-based transcript annotation. Gene expression was quantified using transcripts per million (TPM) across multiple mouse tissues and cell types, including cerebral cortex, hippocampus, gastrocnemius muscle, and the C2C12 immortalized mouse myoblast cell line. Genome browser visualization was performed using IGV to evaluate exon connectivity and transcript boundaries.

Expression analysis revealed marked tissue-specific variation in Prxl2b, abundance with expression detected predominantly in a subset of tissues and cell types. Transcript reconstruction consistently identified the canonical transcript ENSMUST00000030935.9 across independently analyzed datasets. Comparative analysis with GENCODE annotations demonstrated strong agreement in exon–intron organization and splice junction usage. Long-read evidence additionally suggested variation in transcript boundaries, particularly at the 5' region, while preserving the overall coding structure of the annotated transcript.

These findings demonstrate the utility of long-read RNA sequencing for accurate characterization of Prxl2b transcript architecture and provide a refined view of its expression landscape across mouse tissues.

1. Introduction

Identified over four decades ago, alternative splicing (AS) has emerged as a fundamental mechanism enabling extensive proteomic diversity from a limited set of genes (Alt 1980).

Most eukaryotic genes produce multiple transcriptional isoforms, with recent studies showing that a typical protein-coding gene harbors about 11 exons and yields an average of 5.4 distinct mRNA transcripts (Piovesan et al. 2016). Aberrant alternative splicing can activate oncogenes or inactivate tumor suppressors, thereby promoting tumorigenesis (Bashyam et al. 2019). One mechanism proposed to partially explain this apparently contradictory effect is the differential usage of transcript isoforms,

commonly referred to as isoform switching (IS). This regulatory phenomenon has been shown to exert substantial biological effects across diverse biological contexts, driven by the functional diversity of the different isoforms (Vitting-Seerup & Sandelin 2017).

PGFS enzymes such as AKR1C3 convert intermediates to PGF2α, promoting inflammation and cancer progression through FP receptor signaling. PTGDS-related synthases exhibit comparable roles in tumorigenesis, with genomic analyses identifying multiple copies that modulate stress responses (Bertrand et al. 2021). Studies document indirect effects of PG synthases on gene networks but underscore the need for systems biology approaches to comprehensively map downstream targets and pathways—aligning with the rationale that no direct genome-wide PGFS analyses exist. Closest approximations involve PTGS2 expression driving angiogenesis through PGE2/VEGF signaling, suggesting parallel unexplored networks for PGFS enzymes (Yang et al. 2022). Prxl2b catalyzes NADPH-dependent reduction of prostamide H2 to prostamide F2α and prostaglandin H2 to PGF2α, connecting it to prostaglandin biosynthesis and fatty acid metabolism. This peroxiredoxin-like FAM213 family member localizes to cytosol, extracellular, exosomes, and mitochondria (Yoshikawa et al. 2010). In mammals, splicing is mediated by the spliceosome, which recognizes consensus sequences at exon–intron boundaries, while cis-regulatory elements (such as exonic or intronic enhancers and silencers) and trans-acting factors, including SR proteins, modulate exon selection. Major alternative-splicing patterns include cassette exons (accounting for approximately 30% of events), mutually exclusive exons, intron retention (often occurring in untranslated regions), and alternative 5'/3' splice-site usage, which are regulated by chromatin modifications, DNA methylation, and diverse RNA-binding proteins (Chen & Manley 2009).

Splicing generates novel isoforms with altered coding potential—such as truncated proteins, gain or loss of functional domains, or nonsense-mediated decay (NMD)-sensitive transcripts—enabling tissue-specific expression and fine-tuned responses to cellular stimuli (Tao et al. 2024). Furthermore, this regulatory landscape supports the coordination of tissue-specific protein synthesis during development, as coordinated splicing networks facilitate organ maturation and tissue identity (Baralle & Giudice 2017).

Given the extensive isoform diversity generated by alternative splicing, accurately capturing and quantifying these transcript variants remains a major challenge in transcriptome studies. In transcriptome studies such as RNA-seq, alternative splicing complicates quantification, as isoform-specific alignment and differential-splicing tools (e.g., rMATS, MAJIQ) are required to distinguish genuine splicing events from technical noise, thereby influencing differential-expression calls and downstream functional annotation (Mazin et al. 2021).

Key platforms include Pacific Biosciences (PacBio) SMRT sequencing, which produces highly accurate long HiFi reads via circular consensus sequencing, and Oxford Nanopore Technologies (ONT), which enables direct RNA sequencing to detect RNA modifications such as m^6A without reverse transcription. Long-read RNA-seq (LR-RNA-seq) unambiguously assigns full-length isoforms, reveals novel transcripts, fusion genes, and alternative-splicing events (such as cassette exons) without de novo assembly,

thereby improving accuracy in repetitive regions and quantification of lowly expressed isoforms. Particularly well suited for mammalian splicing analysis, LR-RNA-seq captures end-to-end transcript structures for novel isoform discovery (for example, in genes such as Prxl2b), enhances cell-differentiation studies by profiling dynamic isoform switches, and supports disease research through precise mapping of transcriptome complexity (Ament et al. 2024; Chen et al. 2025).

The ENCODE (Encyclopedia of DNA Elements) Project is a comprehensive public resource that maps functional elements in the human and mouse genomes, including extensive transcriptomic data generated by RNA-seq. Launched in 2003 by the NHGRI, ENCODE integrates short- and long-read RNA-seq across diverse cell lines, tissues, and developmental stages to annotate genes, promoters, enhancers, and non-coding RNAs, revealing that over 80% of the genome exhibits detectable biochemical activity (Abascal et al. 2020). ENCODE4's long-read RNA-seq dataset (PacBio, > 1 billion reads from 81 samples) captures full-length isoforms, novel transcripts—including 40% new exon chains—and diversity-generating mechanisms such as promoter, splicing, and polyA-site shifts, complementing short-read data for precise isoform quantification using STAR/RSEM pipelines (Park et al. 2012).

In this study, we analyzed publicly available PacBio long-read RNA sequencing datasets from the ENCODE project to investigate the expression patterns and transcript architecture of the mouse Prxl2b gene across multiple tissues and cell types including cerebral cortex, hippocampus, gastrocnemius muscle, and the C2C12 immortalized mouse myoblast cell line. By integrating transcript quantification, long-read transcript reconstruction, and genome browser visualization, we aimed to characterize Prxl2b expression, evaluate transcript structure, and assess the extent to which long-read sequencing can refine current transcript annotations. Particular attention was given to transcript boundary variation and the validation of reconstructed transcript models using independent long-read datasets. This integrative approach effectively bridges the gap between fragmented short-read transcriptome assemblies and the actual physical isoforms present in diverse biological samples (Reese et al. 2023).

2. Materials and Methods

2.1 Data Acquisition and Reference Annotation

Publicly available mouse long-read RNA sequencing datasets were obtained from the Encyclopedia of DNA Elements (ENCODE) project (Luo et al. 2020). A total of 80 PacBio full-length transcriptome datasets were selected to investigate Prxl2b expression and transcript architecture across diverse biological contexts. The analyzed datasets included neural tissues (hippocampus and cerebral cortex), skeletal muscle tissues (gastrocnemius muscle and embryonic forelimb), cardiac tissue, adrenal gland, cultured C2C12 myoblasts and differentiated myotubes, embryonic stem cell-derived lines, and nuclear-fraction transcriptomes. The inclusion of both tissue-derived and cellular transcriptomic datasets enabled evaluation of Prxl2b transcription across multiple developmental stages, physiological states, and cellular environments. Detailed information regarding ENCODE project accessions, biosample descriptions, sequencing files, developmental stages, and tissue sources is provided in Supplementary

Table S1. The mouse reference genome assembly (mm10) and GENCODE release M25 annotation were used as reference resources for all downstream analyses.

2.2 Quality Assessment and Long-Read Alignment

Raw PacBio long-read RNA sequencing datasets were initially assessed using NanoPlot to evaluate sequencing characteristics, including read-length distribution and overall sequencing quality. Long-read sequences were subsequently aligned to the mouse reference genome (mm10) using Minimap2 (Li 2018) implemented through the Galaxy platform. Alignment was performed using parameters optimized for spliced alignment of PacBio Iso-Seq reads, enabling accurate mapping of full-length transcripts across exon–intron boundaries.

The resulting alignments were generated in BAM format and served as the basis for downstream transcript reconstruction, transcript quantification, and visualization analyses.

2.3 Transcript Reconstruction and Quantification Using IsoQuant

Transcript discovery, reconstruction, and abundance estimation were performed using IsoQuant (Prjibelski et al. 2023). Long-read alignments were analyzed to identify expressed transcript models and generate both gene-level and transcript-level expression estimates. Expression abundance was normalized as transcripts per million (TPM), allowing direct comparison of transcript expression across biological samples.

Gene-level TPM values corresponding to Prxl2b (ENSMUSG00000029059.9) were extracted from IsoQuant output files and used for comparative expression analysis. Transcript-level TPM values were additionally examined for the annotated Prxl2b transcript isoforms ENSMUST00000030935.9 (Prxl2b-201), ENSMUST00000123513.1 (Prxl2b-202), and ENSMUST00000132281.1 (Prxl2b-203) to evaluate their relative contribution to overall Prxl2b expression. Gene-level and transcript-level TPM values are provided in Supplementary Tables S2.

2.4 Independent Transcript Reconstruction Using FLAIR

To independently assess transcript architecture, Minimap2-generated BAM alignments were converted into 12-column BED (BED12) format using Galaxy. The resulting BED12 files, which preserve exon–intron structures and splice-junction information, were used as input for the FLAIR workflow.

Transcript correction was performed using FLAIR-correct with the mouse reference genome (mm10) and GENCODE release M25 annotation. This step refined splice-junction coordinates and improved transcript structure accuracy using reference-guided correction. Corrected reads were subsequently processed using FLAIR-collapse to generate high-confidence full-length transcript models representing consensus transcript structures supported by long-read sequencing evidence.

Transcript models reconstructed by FLAIR were compared with IsoQuant-derived transcript models and GENCODE reference annotations to evaluate exon–intron organization, splice-junction usage, transcript length, and transcript boundary positions. Transcript structures consistently identified by both reconstruction approaches were considered high-confidence transcript models.

2.5 Comparative Annotation Analysis

To investigate transcript architecture at the Prxl2b locus, reconstructed transcript models were systematically compared with GENCODE vM25 reference annotations. Particular attention was given to transcript boundary positions, exon composition, splice-junction usage, and transcript length. Differences between reconstructed transcript models and reference annotations were evaluated to identify potential annotation refinements supported by long-read sequencing evidence.

2.6 Validation of Transcript Structures

Transcript structures identified through computational reconstruction were validated through direct visualization of long-read alignments using the Integrative Genomics Viewer (IGV) (Robinson et al. 2011). Independent ENCODE datasets were examined to assess read coverage, splice-junction support, and transcript boundary consistency across the Prxl2b locus. Validation analyses focused particularly on transcript boundary variation observed at the 5' region of the gene. Read-level evidence from multiple independent datasets was used to evaluate the reproducibility of reconstructed transcript structures and transcript boundary extensions.

2.7 Expression Analysis

Prxl2b expression patterns were evaluated across all analyzed datasets using IsoQuant-derived TPM values. Samples exhibiting detectable expression were included in comparative analyses and graphical visualization. Gene-level TPM values were used to assess tissue-specific expression patterns, whereas transcript-level TPM values were examined to determine the relative abundance of annotated transcript isoforms and identify the predominant transcript contributing to overall Prxl2b expression.

2.8 Data Visualization

Genome-scale transcript architecture visualizations were generated using pyGenomeTracks (Lopez-Delisle et al. 2021). Reconstructed transcript models, long-read coverage profiles, and reference annotations were visualized simultaneously to facilitate comparison of transcript structures and transcript boundary positions. Manual inspection of genomic regions of interest was performed using IGV. Expression plots and summary visualizations were generated using R software (R Core Team 2024).

3. Results

3.1 Expression Profiling of Prxl2b Across Mouse Tissues

Gene expression analysis using IsoQuant-derived TPM values revealed tissue-specific expression of Prxl2b across the analyzed ENCODE long-read RNA sequencing datasets. Detectable expression was observed in a subset of tissues, whereas several samples showed little or no measurable transcription. Comparative analysis of expressing tissues demonstrated substantial variability in transcript abundance, indicating differential regulation of Prxl2b expression across biological contexts (Fig. 1).

The expression distribution suggests that Prxl2b is not constitutively expressed at equivalent levels across all tissues but instead exhibits tissue-dependent transcriptional activity. These findings are consistent with previous reports implicating Prxl2B in redox regulation and cellular stress-response pathways.

3.2 Reconstruction of Prxl2b Transcript Architecture

Transcript reconstruction performed using IsoQuant identified ENSMUST00000030935.9 as the predominant expressed Prxl2b transcript. Comparison with GENCODE vM25 annotations demonstrated extensive agreement in exon organization and splice-junction structure. Independent transcript reconstruction using FLAIR generated highly similar transcript models, providing additional support for the reconstructed exon–intron architecture. Comparative visualization of transcript models and long-read alignments showed that the major transcript structure detected across independent datasets closely matched the annotated reference transcript (Fig. 2). No evidence for major exon-skipping events or extensive alternative exon usage was observed. Instead, transcript variation was primarily associated with differences in transcript boundary positions.

3.3 Identification of Transcript Boundary Variation

Detailed examination of reconstructed transcript models revealed variation at the 5' boundary of the Prxl2b transcript relative to the current reference annotation. Long-read-supported transcript models extended beyond the annotated transcript start site while maintaining the canonical exon organization of ENSMUST00000030935.9.

The observed extension was supported by full-length transcript models generated independently by IsoQuant and FLAIR, suggesting that the extended transcript boundary reflects genuine transcriptional variation rather than an assembly artifact. Importantly, the coding structure and exon connectivity of the transcript remained largely unchanged, indicating that the variation is most likely associated with transcript boundary definition rather than the generation of a distinct protein-coding isoform.

Discussion

In this study, we performed a comprehensive analysis of the mouse Prxl2b gene using publicly available ENCODE long-read RNA sequencing datasets to investigate its expression profile and transcript architecture across multiple tissues and cell types. By integrating IsoQuant-based transcript quantification, FLAIR transcript reconstruction, comparative annotation analysis, and genome browser

validation, we provide a refined view of Prxl2b transcription and demonstrate the value of long-read sequencing for transcript annotation refinement.

Expression analysis revealed substantial variability in Prxl2b abundance across the analyzed datasets. While several tissues exhibited relatively high levels of expression, others showed little or no detectable transcription. This pattern suggests that Prxl2b is subject to tissue-dependent regulation and may participate in biological processes that vary according to cellular context. Such variability is consistent with the reported involvement of Prxl2b in oxidative stress responses, redox homeostasis, and prostaglandin-related pathways, which are known to exhibit context-specific activity.

Transcript reconstruction consistently identified ENSMUST00000030935.9 as the predominant expressed Prxl2b transcript across the analyzed datasets. Transcript-level quantification revealed little or no detectable expression of other annotated transcript models, indicating that the canonical transcript accounts for the majority of Prxl2b transcription in the examined samples. These findings suggest that although multiple transcript annotations exist within current reference databases, only a subset may contribute substantially to the expressed transcriptome under the analyzed conditions.

A major observation of this study was the identification of transcript boundary variation at the 5' end of Prxl2b. Comparative analysis of reconstructed transcript models and GENCODE annotations indicated the presence of an extended transcript boundary upstream of the annotated transcription start region. Importantly, this extension was supported by independent transcript reconstruction approaches and by direct visualization of long-read alignments in multiple ENCODE datasets. Despite this boundary variation, the overall exon–intron organization and splice-junction structure remained highly consistent with the canonical transcript model.

The observed transcript extension is most consistent with the presence of an alternative transcription start site or an extended 5' untranslated region rather than the generation of a distinct protein-coding isoform. Transcript boundary variation represents an important but often underappreciated source of transcript diversity and may influence transcript stability, translational efficiency, RNA secondary structure, or interactions with regulatory proteins. Long-read sequencing technologies are particularly well suited for identifying such transcript features because they capture full-length transcript structures without requiring transcript assembly from short sequencing fragments, enabling accurate transcript reconstruction and isoform characterization (Prjibelski et al. 2023).

Our results also highlight the importance of combining computational transcript reconstruction with manual inspection of long-read alignments. Initial transcript predictions may suggest the presence of transcript diversity; however, validation using genome browser visualization is essential to distinguish genuine transcript structures from reconstruction artifacts. The strong concordance observed between IsoQuant, FLAIR, and IGV-supported read alignments increases confidence in the transcript models reported in this study.

Several limitations should be acknowledged. First, the analysis relied exclusively on publicly available transcriptomic datasets and therefore reflects only the biological conditions represented within those experiments. Second, although long-read sequencing provides strong evidence for transcript structure, it does not directly establish the functional significance of transcript boundary variation. Experimental approaches such as 5' RACE, CAGE sequencing, promoter mapping, or functional assays would be required to precisely define transcription start sites and determine the biological consequences of the extended transcript region.

In conclusion, this study provides a comprehensive characterization of Prxl2b expression and transcript architecture using long-read RNA sequencing data. Our findings support the predominance of the canonical Prxl2b transcript while revealing evidence for transcript boundary variation at the 5' region. These results demonstrate the utility of long-read transcriptomics for refining gene annotations and contribute to a more accurate understanding of Prxl2b transcriptional organization in the mouse genome.

Conclusion

Using publicly available ENCODE long-read RNA sequencing datasets, we characterized the expression profile and transcript architecture of the mouse Prxl2b gene. Expression analysis revealed tissue-specific variation in Prxl2b abundance, while transcript reconstruction consistently identified ENSMUST00000030935.9 as the predominant expressed transcript. Comparative analysis of reconstructed transcripts and reference annotations revealed evidence for a 5' transcript boundary extension supported by independent long-read datasets and IGV validation. Collectively, these findings refine the current annotation of the Prxl2b locus and highlight the utility of long-read RNA sequencing for improving transcript structure characterization and gene annotation. Future investigations should prioritize the integration of orthogonal data and replicates to further validate these findings, as increasing read quantity alone does not guarantee the resolution of transcript complexity (Pardo-Palacios et al. 2024). Furthermore, the application of randomized primers in future studies may address potential limitations associated with internal oligo priming, thereby enhancing the detection of rare or structurally complex transcript variants (Uapinyoying et al. 2020). Ultimately, shifting toward such comprehensive annotation strategies is essential for addressing the persistent gaps in our existing reference models (Adams & Vollmers 2024; Lagarde et al. 2017). These efforts, combined with high-throughput polling of transcription start and termination sites, provide a robust framework for mapping the extensive regulatory variation inherent in mammalian genomes (Carninci et al. 2005).

Building upon this, the utilization of emerging long-read platforms like the PromethION will likely facilitate deeper sequencing coverage, further improving the identification and precise quantification of such unannotated transcript features (Sessegolo et al. 2019).

Declarations

Conflict of interest:

These authors declared no competing interests in

Ethical Approval:

Not applicable.

Informed consent:

Not applicable.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author Contribution

Preeti Pilania: Conceptualization, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. Sikander Saini: Data curation, Investigation. Diksha Sethi: Data curation, Investigation.Sameer Kumar Sahani: Data curation, Resources. Prem Kumar: Data curation, Resources. Suman Chopra: Supervision, Validation, Visualization. Satish Kumar: Supervision, Validation, Visualization.Dhruba Malakar: Conceptualization, Methodology, Software, Project administration, Writing – review & editing.

Data availability

The authors confirm that the data supporting the findings of this study are available within the article and supplementary materials.

References

1. Abascal F, Acosta R, Addleman NJ, Adrian J, Afzal V, Aken B, Akiyama JA, Jammal O, Al, Amrhein H, Anderson SM, Andrews GR, Antoshechkin I, Ardlie KG, Armstrong J, Astley M, Banerjee B, Barkal AA, Barnes IHA, Barozzi I, Zimmerman J (2020) Expanded encyclopedias of DNA elements in the human and mouse genomes. Nature 583(7818):699–710. https://doi.org/10.1038/s41586-020-2493-4
2. Adams M, Vollmers C (2024) Generation and analysis of a mouse multitissue genome annotation atlas. Genome Res 34(11):2108–2117. https://doi.org/10.1101/gr.279217.124

3. Ak FW, Bothwell ALM, Knapp M, Siden E, Mather E, Koshlandt M, Baltimore D (1980) Synthesis of Secreted and Membrane-Bound Immunoglobulin Mu Heavy Chains Is Directed by mRNAs That Differ at Their 3' Ends. In *Cell* 20
4. Ament IH, DeBruyne N, Wang F, Lin L (2025) Long-read RNA sequencing: A transformative technology for exploring transcriptome complexity in human diseases. In Molecular Therapy (Vol. 33, Number 3, pp. 883–894). Cell Press. https://doi.org/10.1016/j.ymthe.2024.11.025
5. Baralle FE, Giudice J (2017) Alternative splicing as a regulator of development and tissue identity. Nat Rev Mol Cell Biol 18(7):437–451. https://doi.org/10.1038/nrm.2017.27
6. Bashyam MD, Animireddy S, Bala P, Naz A, George SA (2019) The Yin and Yang of cancer genes. In Gene (Vol. 704, pp. 121–133). Elsevier B.V. https://doi.org/10.1016/j.gene.2019.04.025
7. Bertrand JA, Woodward DF, Sherwood JM, Spenlehauer A, Silvestri C, Piscitelli F, Marzo V, Di, Yamazaki M, Sakimura K, Inoue Y, Watanabe K, Overby DR (2021) Deletion of the gene encoding prostamide/prostaglandin F synthase reveals an important role in regulating intraocular pressure. *Prostaglandins Leukotrienes and Essential Fatty Acids*, *165*. https://doi.org/10.1016/j.plefa.2020.102235
8. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C et al (2005) The transcriptional landscape of the mammalian genome. Science 309(5740):1559–1563. https://doi.org/10.1126/science.1112014
9. Chen M, Manley JL (2009) Mechanisms of alternative splicing regulation: Insights from molecular and genomics approaches. Nat Reviews Mol Cell Biology (Vol 10(11):741–754. https://doi.org/10.1038/nrm2777
10. Chen Y, Davidson NM, Wan YK, Yao F, Su Y, Gamaarachchi H, Sim A, Patel H, Low HM, Hendra C, Wratten L, Hakkaart C, Sawyer C, Iakovleva V, Lee PL, Xin L, Ng HEV, Loo JM, Ong X, Göke J (2025) A systematic benchmark of Nanopore long-read RNA sequencing for transcript-level analysis in human cell lines. Nat Methods 22(4):801–812. https://doi.org/10.1038/s41592-025-02623-4
11. Lagarde J, Uszczynska-Ratajczak B, Carbonell S, Pérez-Lluch S, Abad A, Davis C, Gingeras TR, Frankish A, Harrow J, Guigo R, Johnson R (2017) High-throughput annotation of full-length long noncoding RNAs with capture long-read sequencing. Nat Genet 49(12):1731–1740. https://doi.org/10.1038/ng.3988
12. Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34(18):3094–3100. https://doi.org/10.1093/bioinformatics/bty191
13. Lopez-Delisle L, Rabbani L, Wolff J, Bhardwaj V, Backofen R, Grüning B, Ramírez F, Manke T (2021) pyGenomeTracks: reproducible plots for multivariate genomic datasets. Bioinformatics 37(3):422–423. https://doi.org/10.1093/bioinformatics/btaa692
14. Luo Y, Hitz BC, Gabdank I, Hilton JA, Kagda MS, Lam B, Myers Z, Sud P, Jou J, Lin K et al (2020) New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. Nucleic Acids Res 48(D1):D882–D889. https://doi.org/10.1093/nar/gkz1062

15. Mazin PV, Khaitovich P, Cardoso-Moreira M, Kaessmann H (2021) Alternative splicing during mammalian organ development. Nat Genet 53(6):925–934. https://doi.org/10.1038/s41588-021-00851-w
16. Park E, Williams B, Wold BJ, Mortazavi A (2012) RNA editing in the human ENCODE RNA-seq data. Genome Res 22(9):1626–1633. https://doi.org/10.1101/gr.134957.111
17. Piovesan A, Caracausi M, Antonaros F, Pelleri MC, Vitale L (2016) GeneBase 1.1: A tool to summarize data from NCBI gene datasets and its application to an update of human gene statistics. *Database*, *2016*. https://doi.org/10.1093/database/baw153
18. Prjibelski AD, Mikheenko A, Joglekar A, Smetanin A, Jarroux J, Lapidus AL, Tilgner HU (2023) Accurate isoform discovery with IsoQuant using long reads. Nat Biotechnol 41(7):915–918. https://doi.org/10.1038/s41587-022-01565-y
19. Reese F, Williams B, Balderrama-Gutierrez G, Wyman D, Çelik MH, Rebboah E, Rezaie N, Trout D, Razavi-Mohseni M, Jiang Y et al (2023) The ENCODE4 long-read RNA-seq collection reveals distinct classes of transcript structure diversity. https://doi.org/10.1101/2023.05.15.540865. bioRxiv
20. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative Genomics Viewer. Nat Biotechnol 29(1):24–26. https://doi.org/10.1038/nbt.1754
21. Sessegolo C, Cruaud C, Da Silva C, Cologne A, Dubarry M, Derrien T, Lacroix V, Aury JM (2019) Transcriptome profiling of mouse samples using nanopore sequencing of cDNA and RNA molecules. Sci Rep 9:14908. https://doi.org/10.1038/s41598-019-51470-9
22. Tao Y, Zhang Q, Wang H, Yang X, Mu H (2024) Alternative splicing and related RNA binding proteins in human health and disease. In *Signal Transduction and Targeted Therapy*. Springer Nat 9(1). https://doi.org/10.1038/s41392-024-01734-2
23. Vitting-Seerup K, Sandelin A (2017) The landscape of isoform switches in human cancers. Mol Cancer Res 15(9):1206–1220. https://doi.org/10.1158/1541-7786.MCR-16-0459
24. Yang H, Chen X, Li Z, Wu X, Zhou M, Zhang X, Liu Y, Sun Y, Zhu C, Guo Q, Chen T, Zhang J (2022) Genome-Wide Analysis Indicates a Complete Prostaglandin Pathway from Synthesis to Inactivation in Pacific White Shrimp, Litopenaeus vannamei. Int J Mol Sci 23(3). https://doi.org/10.3390/ijms23031654
25. Yoshikawa K, Takei S, Hasegawa-Ishii S, Chiba Y, Furukawa A, Kawamura N, Hosokawa M, Woodward DF, Watanabe K, Shimada A (2011) Preferential localization of prostamide/prostaglandin F synthase in myelin sheaths of the central nervous system. Brain Res 1367:22–32. https://doi.org/10.1016/j.brainres.2010.10.019

Figures

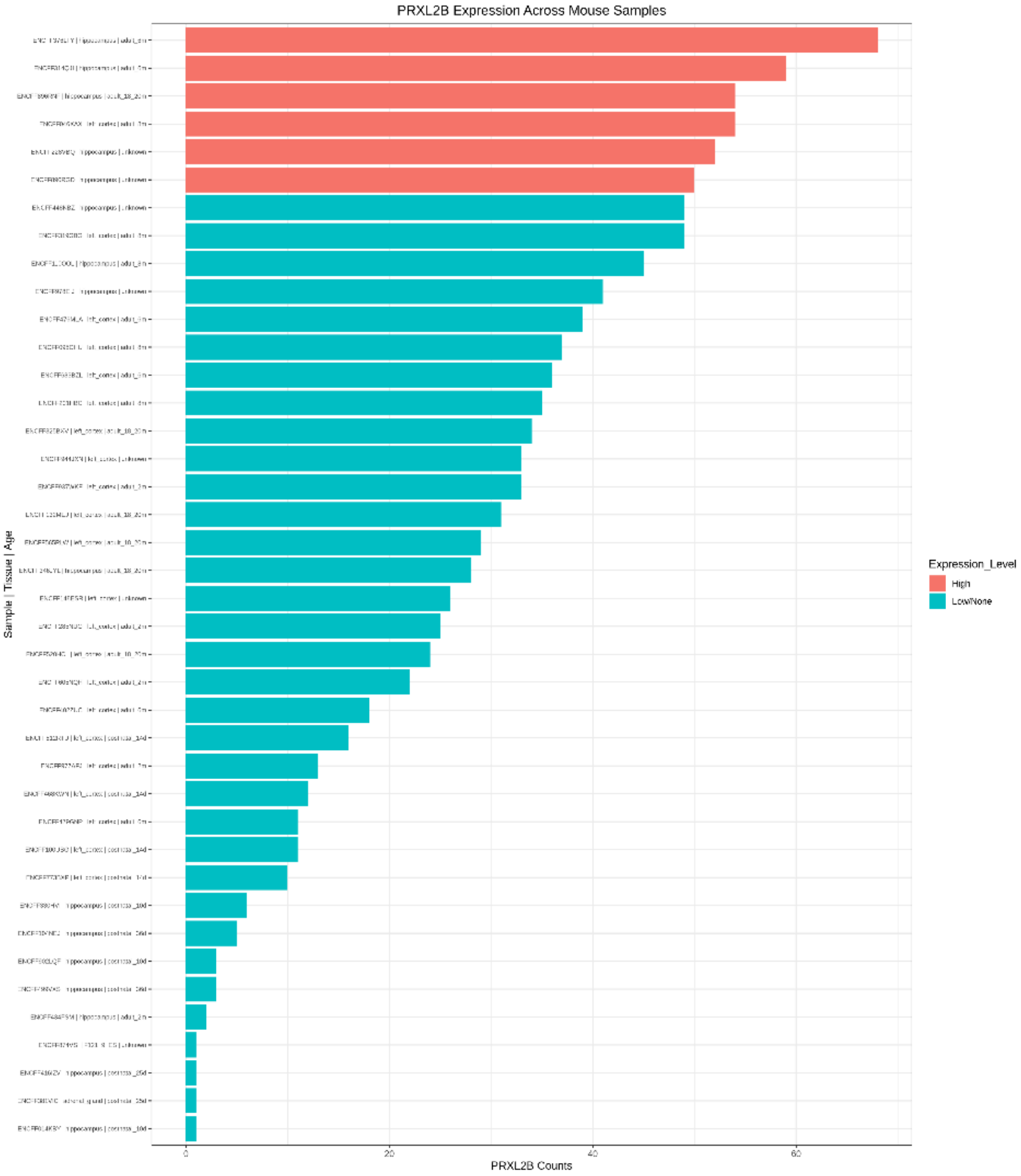


Figure 1

Expression profile of Prxl2b across ENCODE long-read RNA sequencing datasets. Gene-level expression of Prxl2b (ENSMUSG00000029059.9) was quantified using IsoQuant and normalized as transcripts per million (TPM). Only samples with detectable expression were included in the visualization. Expression levels varied substantially among tissues and cell types, indicating tissue-dependent regulation of Prxl2b

transcription. TPM values were derived from long-read RNA sequencing datasets aligned to the mouse reference genome (mm10) and quantified using IsoQuant.

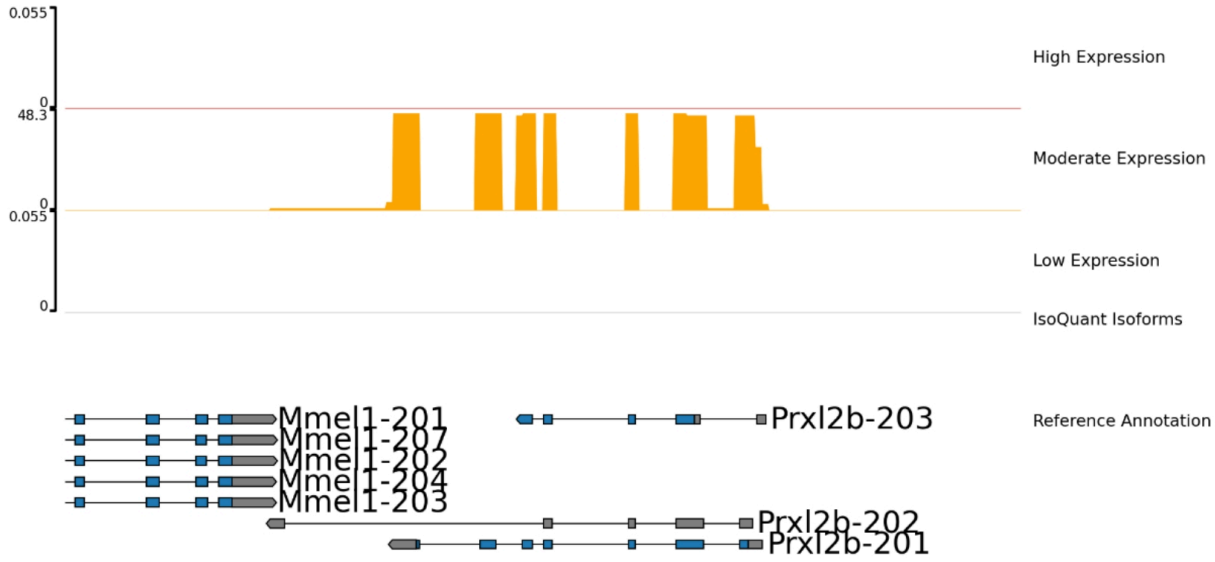


Figure 2

Long-read alignments and transcript models reconstructed from ENCODE datasets were compared with GENCODE vM25 and RefSeq annotations. The predominant transcript identified by IsoQuant and FLAIR corresponded closely to the annotated Prxl2b transcript ENSMUST00000030935.9, showing concordant exon organization and splice-junction usage. No major exon-skipping events or extensive alternative exon utilization were detected, with transcript variation primarily limited to transcript boundary differences.

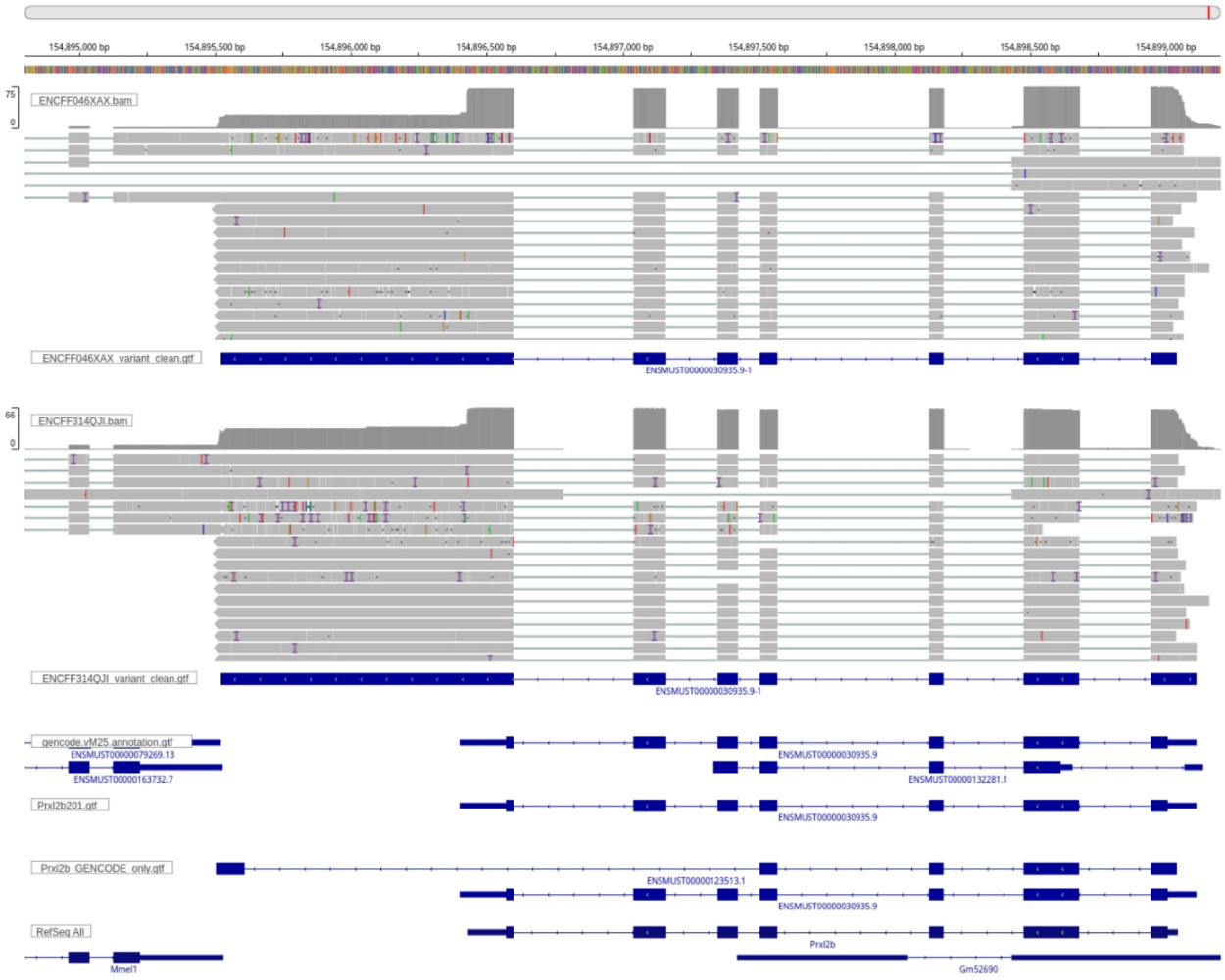


Figure 3

Zoomed visualization of the Prxl2b transcription start region showing long-read alignments and transcript models reconstructed by IsoQuant and FLAIR. Both reconstruction approaches identified transcript models extending upstream of the annotated GENCODE transcript ENSMUST00000030935.9 while preserving canonical exon connectivity and coding structure. The observed variation is consistent with alternative transcript boundary definition rather than the generation of a distinct protein-coding isoform.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementaryfile.docx