Introduction

Isl1 (Islet-1) is a LIM-homeodomain transcription factor expressed in early cardiac and extraembryonic mesoderm progenitors, particularly marking the allantoic mesoderm lineage, which contributes to umbilical vasculature and placental development(Zhu et al., 2024). Loss of Isl1 disrupts allantois outgrowth, chorioallantoic fusion, and vascular formation during early embryogenesis(Zhu et al., 2024). While previous studies have implicated Isl1 in endothelial differentiation and suppression of mesenchymal gene programs, the full transcriptomic consequences of Isl1 deletion in the allantois are not fully characterized(Zhu et al., 2024).

In this analysis, we reanalyze publicly available RNA-seq data (GEO accession: GSE247182) comparing allantois tissue from Isl1 knockout (MUT) and wild-type (WT) mouse embryos at embryonic day 8.25 (E8.25). Our goal is to identify differentially expressed genes (DEGs) and enriched pathways that define Isl1-dependent regulatory programs during early mesodermal development.

We hypothesize that loss of Isl1 disrupts progenitor maintenance and promotes differentiation-related transcriptional programs. Through computational analysis of this RNA-seq dataset using the DESeq2 pipeline(Tomlinson, 2025), we aim to uncover key genes and signaling pathways altered in the absence of Isl1, offering insights into developmental defects and mesodermal lineage specification.

Methods

Data Acquisition, Quality Control, and Alignment:
We analyzed RNA-seq data from the GEO database (GSE247182), consisting of six mouse allantois samples (three wild-type and three Isl1 knockouts) at embryonic day 8.25 (E8.25). After assigning genotypes, we performed quality control using FastQC and MultiQC, which showed high Phred scores (>33), no adapter contamination, and minor sequence bias due to random hexamer priming—trimming was not required. Reads were aligned to the GRCm39 mouse genome (Ensembl release 113) using STAR in 2-pass mode, yielding high alignment rates (76–80%) with no GC or mapping biases. Gene-level quantification was performed using the featureCounts function from the Rsubread package, applied to STAR-aligned BAM files. Only properly paired reads with a mapping quality score ≥30 were counted. This generated a gene-by-sample count matrix, which retained Ensembl gene IDs for downstream normalization and differential expression analysis(Tomlinson, 2025).

Data Normalization and Filtering: We created a DESeqDataSet using the DESeq2 package, specifying genotype (WT vs MUT) as the design factor. Genes with extremely low expression (total count ≤1 across all samples) were filtered out. Library size normalization was performed using DESeq2's median-of-ratios method, and dispersion estimates were calculated to model count variability(Tomlinson, 2025).

Variance Stabilization and Visualization: To enable effective visualization and clustering, data were transformed using variance-stabilizing transformation (VST) and regularized log (rlog). Both approaches produced comparable distributions across samples. We primarily used rlog-transformed data for Principal Component Analysis (PCA) to visualize genotype separation and Sample-to-sample distance heatmaps to assess biological replicate clustering(Tomlinson, 2025).

Differential Expression Analysis: Differentially expressed genes (DEGs) were identified using Wald tests in DESeq2. Adjusted p-values were calculated using the Benjamini–Hochberg false discovery rate (FDR) method. Genes with FDR < 0.05 were considered significantly differentially expressed. We also reported results using relaxed thresholds (nominal $p < 0.05$) for comparative purposes(Tomlinson, 2025).

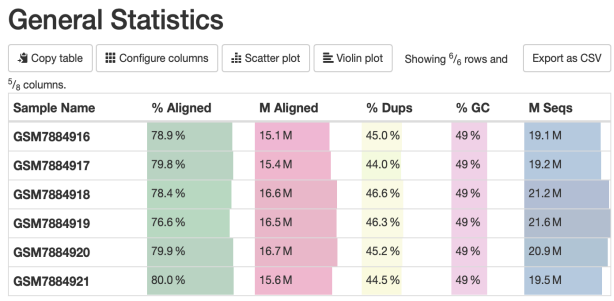
Gene Annotation and DEG Ranking: biomaRt was used to map Ensembl gene IDs to gene symbols, Entrez IDs, and functional descriptions. For ranking the top 10 DEGs, we prioritized genes with the

smallest FDR-adjusted p-values and biologically relevant fold-changes. This method emphasizes both statistical significance and effect size(Tomlinson, 2025).

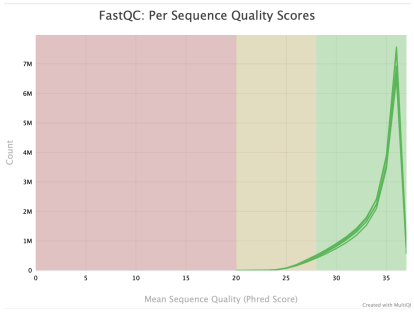
Enrichment Analysis: To explore broader biological processes, we performed Gene Set Enrichment Analysis (GSEA) using the fgsea package. Genes were ranked by DESeq2's Wald test statistic, maintaining directional information. We used MSigDB Hallmark gene sets (mouse-converted) as the reference database. Pathways with adjusted p-value < 0.05 were considered significantly enriched. Selected hallmark pathways were visualized using enrichment plots(Tomlinson, 2025).

Results:

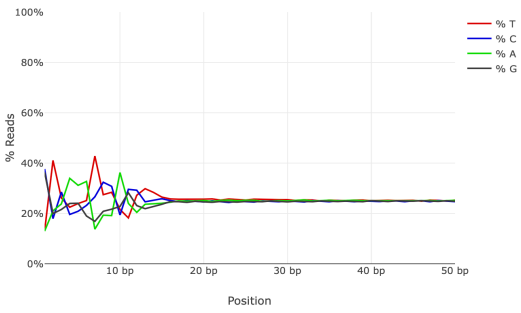
Sequencing Quality and Alignment:
Quality control using FastQC and MultiQC confirmed excellent sequencing across all six libraries, with per-base Phred scores >33 and GC content centered around 49% (Figure 1b, 1a). Minor sequence bias at the start of reads was attributed to random hexamer priming (Figure 1c) and did not warrant trimming. Adapter contamination was negligible (<0.1%), and sequence duplication levels (44–47%) were within expected ranges for RNA-seq (Figure 1d). STAR aligner produced high mapping rates (76.6–80.0%) with 15.1–16.7 million uniquely aligned reads per sample (Figure 1e). Post-alignment metrics showed uniform GC content and accurate mapping, supporting the decision to retain full read lengths(Tomlinson, 2025).



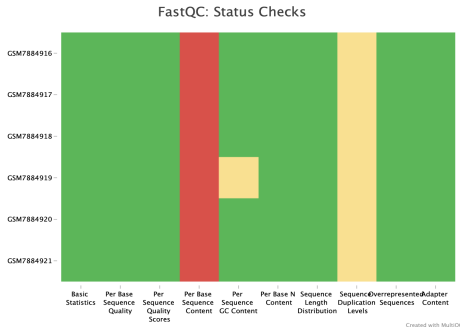
(a)



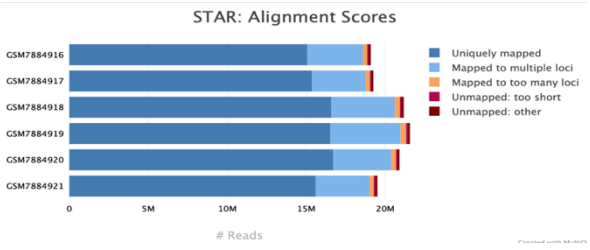
(b)



(c)



(d)



(e)

Figure 1. RNA-Seq Read Quality and Alignment Metrics Across WT and *Isl1*-KO Samples

a) General Statistics Table
b) Per Sequence Quality Scores
c) Per Base Sequence Content
d) FastQC Status Summary
e) STAR Alignment Scores

Data Normalization and Transformation: To enable robust downstream analyses such as principal component analysis (PCA) and differential expression testing, raw read counts were normalized using two approaches provided by the DESeq2 package: variance stabilizing transformation (VST) and regularized log transformation (rlog). These methods stabilize variance across genes with different expression levels and correct for library size differences. As illustrated in Figure 2, both VST- and rlog-transformed data showed highly comparable distributions across all samples. Median expression levels remained consistent between wild-type (WT) and Isl1 knockout (Isl1_KO) groups, with VST medians clustered around 8.7–8.9 and rlog values uniformly spread across the 0–8 range. This consistency confirms effective normalization, eliminating technical bias and enabling reliable comparison between experimental conditions(Tomlinson, 2025).

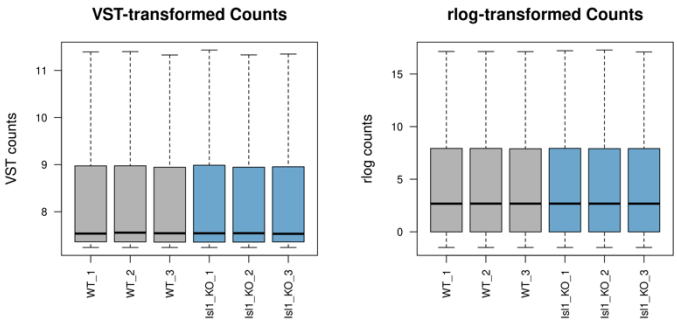


Figure 2. Distribution of Normalized Gene Expression Values

(a) Boxplot of VST-transformed counts for WT and Isl1_KO replicates.

(b) Boxplot of rlog-transformed counts for the same samples.

Principal Component Analysis (PCA): To explore global transcriptional variation and examine clustering among biological replicates, a principal component analysis (PCA) was performed on the rlog-transformed count matrix using DESeq2. The first principal component (PC1), which accounted for 41% of the variance, clearly separated the wild-type (WT) and Isl1 knockout (Isl1_KO) samples, demonstrating genotype-dependent transcriptomic shifts. The second component (PC2), explaining 26% of the variance, contributed to within-group dispersion, possibly reflecting subtle biological or technical differences. The PCA plot (Figure 3) shows well-clustered replicates within each genotype, indicating strong internal consistency and distinct expression profiles between conditions(Tomlinson, 2025).

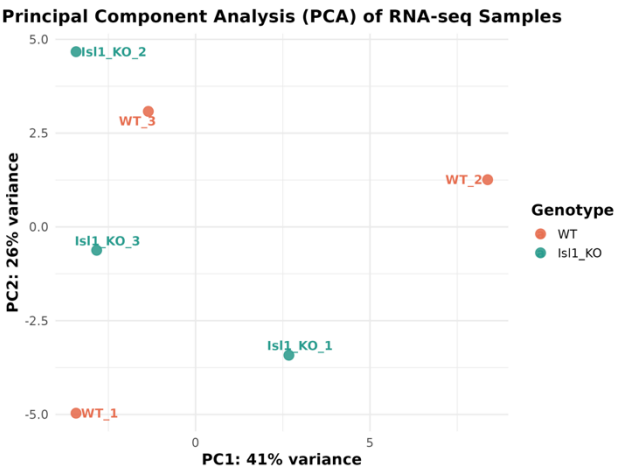


Figure 3. Principal Component Analysis of RNA-seq Samples

Principal component analysis (PCA) of rlog-transformed RNA-seq data across six samples (three WT and three Isl1_KO). PC1 explains 41% of the total variance and separates samples by genotype, while PC2 accounts for 26% of the variance and reflects within-group variability. Each point represents one biological replicate, colored by genotype. Sample names are labeled using ggrepel to prevent overlap.

MA Plot of Differential Expression: An MA plot was generated to visualize the relationship between gene expression magnitude and $\log_2$ fold change between Isl1_KO and wild-type (WT) samples (Figure 4). Most genes are distributed symmetrically around $\log_2FC = 0$, indicating minimal or no differential expression. A small number of transcripts are substantially upregulated or downregulated in mutants, appearing as outliers. In total, 9 genes were found to be significantly differentially expressed at FDR < 0.05, with 15 genes meeting the less stringent threshold of FDR < 0.1. Significant genes are depicted in orange, while those with extreme fold changes beyond ±3 are shown as open triangle symbols near the top and bottom plot boundaries. This distribution

suggests that differentially expressed genes (DEGs) tend to occur across a range of expression intensities, but with clearer fold-change separation at higher mean counts(Tomlinson, 2025).

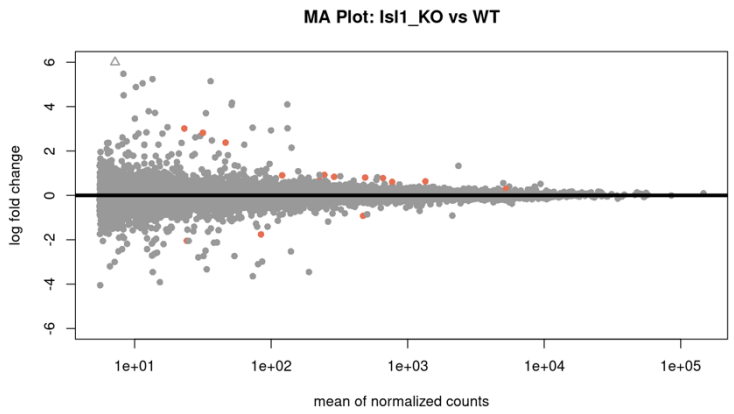


Figure 4. MA Plot of Differential Gene Expression Between Isl1_KO and WT Samples

The MA plot visualizes $\log_2$ fold change (y-axis) against the mean of normalized counts (x-axis, log scale) for each gene. Each dot represents a gene, with non-significant genes shown in gray and significant genes (adjusted p-value < 0.05) highlighted in orange. Open triangles indicate genes with $\log_2$ fold changes that exceed the plotting window (i.e., greater than ±3). The majority of genes center around $\log_2FC = 0$, indicating no differential expression. A small subset exhibits substantial changes in expression between genotypes, notably among genes with moderate to high expression levels.

Heatmap of Top Differentially Expressed Genes: A heatmap of the top 30 differentially expressed genes (ranked by adjusted p-value) was created using rlog-transformed, z-score-normalized expression data (Figure 5). The expression profiles exhibit clear genotype-specific clustering, with wild-type (WT) and Isl1 knockout (MUT) samples forming distinct groups. Genes downregulated in MUT (shown in blue) were more highly expressed in WT, while upregulated genes (red) showed the opposite trend. The strong intra-group consistency and separation highlight the transcriptional impact of Isl1 deletion(Tomlinson, 2025).

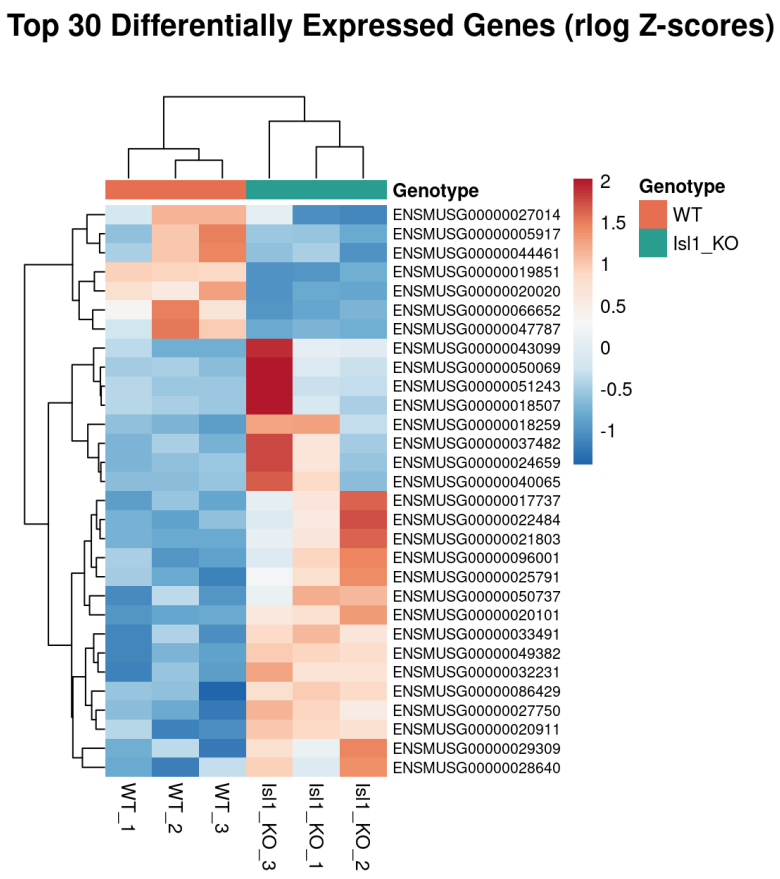


Figure 5. Heatmap of Top 30 Differentially Expressed Genes (rlog Z-scores)

This heatmap visualizes the top 30 differentially expressed genes identified by DESeq2, ranked by adjusted p-value. The data are rlog-transformed and z-score normalized across rows (genes). Color represents relative expression: red indicates higher expression, and blue indicates lower expression, relative to the gene's mean expression across all samples. Columns represent individual samples, annotated by genotype (WT or Isl1_KO). Hierarchical clustering was applied to both rows and columns.

Differential Expression Analysis: Differential expression analysis using DESeq2 identified 9 significant differentially expressed genes (DEGs) at a false discovery rate (FDR) < 0.05. Among the 15 significant genes when FDR < 0.1, 12 were upregulated and 3 were downregulated in Isl1 knockout (KO) samples compared to wild-type (WT). Approximately 48% of genes were excluded due to low average expression (mean count < 6)(Tomlinson, 2025).

Key DEGs included Prl8a2 (*$log_2FC = 2.82$*), a highly upregulated gene in mutants, and Lefty2 (*$log_2FC = -2.04$*), which was strongly downregulated. The top 10 DEGs based on adjusted p-values are summarized in Table 1, serving as high-confidence candidates for further investigation.

Gene Symbol	Ensembl ID	log_2 Fold Change	Adjusted p-value	Description
Perp	ENSMUSG00000019851	-0.921	5.04×10^{-10}	PERP, TP53 apoptosis effector
Postn	ENSMUSG00000027750	0.778	6.58×10^{-6}	periostin, osteoblast specific factor
Krt19	ENSMUSG00000020911	0.607	1.13×10^{-3}	keratin 19
Prl8a2	ENSMUSG00000018259	2.818	7.21×10^{-3}	prolactin family 8, subfamily a, member 2
Ptges	ENSMUSG00000050737	0.836	1.91×10^{-2}	prostaglandin E synthase
Sparcl1	ENSMUSG00000029309	0.633	2.55×10^{-2}	SPARC-like 1
Mmp9	ENSMUSG00000017737	0.919	2.89×10^{-2}	matrix metallopeptidase 9
Hoxc10	ENSMUSG00000022484	0.805	4.72×10^{-2}	homeobox C10
Vsir	ENSMUSG00000020101	0.686	4.82×10^{-2}	V-set immunoregulatory receptor
Lefty2	ENSMUSG00000066652	-2.044	5.79×10^{-2}	left-right determination factor 2

Table 1. Top 10 Differentially Expressed Genes (DEGs) in Isl1_KO vs WT. *Genes are ranked by adjusted p-value. Log_2 fold change indicates expression difference in mutants relative to wild-type.*

Gene Annotation and Biological Interpretation: To facilitate biological insight, the DEGs were annotated using the Ensembl BioMart database via the biomaRt R package. Each gene was matched to its official symbol, functional description, Entrez ID, and genomic coordinates. These annotations enabled downstream interpretation and candidate selection. Notably,

- Perp (ENSMUSG00000019851), a TP53-associated effector, was significantly downregulated in Isl1_KO samples.
- Pcmtd1 (ENSMUSG00000051285), involved in methyltransferase activity, showed modest upregulation.
- Predicted genes like Gm38372 (ENSMUSG00000103509) were also detected, though with low expression and fold change.

The full annotated dataset is provided in the supplementary file DESeq2_results_annotated.csv for use in enrichment analysis or target validation(Tomlinson, 2025).

Differential Expression Visualization Using Volcano Plots: To visualize the transcriptional changes between Isl1 knockout (MUT) and wild-type (WT) allantois samples, volcano plots were generated based on the DESeq2 output. These plots integrate both statistical significance and magnitude of gene expression change, offering a comprehensive overview of the differential expression landscape.

Two volcano plots were generated using distinct threshold criteria to balance discovery sensitivity and biological stringency:

1. Relaxed Criteria (padj < 0.05 and |log₂FC| > 0.38):
   This threshold corresponds to a minimum ~1.3-fold change and was adopted from the criteria reported in Zhu et al., 2024. The relaxed volcano plot (Figure 6A) highlights a broader set of differentially expressed genes (DEGs), capturing modest but reproducible expression shifts. Top DEGs included *Postn*, *Mmp9*, *Ptges*, *Perp*, and *Prl8a2*, while known Isl1 targets such as *Isl1*, *Sox2*, *Mesp1*, and *Procr* also emerged, confirming biological relevance(Tomlinson, 2025; Zhu et al., 2024).

2. Stringent Criteria (padj < 0.05 and |log₂FC| > 1):
   The strict volcano plot (Figure 6B) focuses on DEGs with robust fold changes (≥2-fold) and strong statistical support. This visualization underscores high-confidence candidates such as *Prl8a2*, which retained significance under this more rigorous threshold(Tomlinson, 2025).

Together, these complementary plots enable the identification of both subtle regulatory effects and core transcriptomic alterations driven by Isl1 deletion.

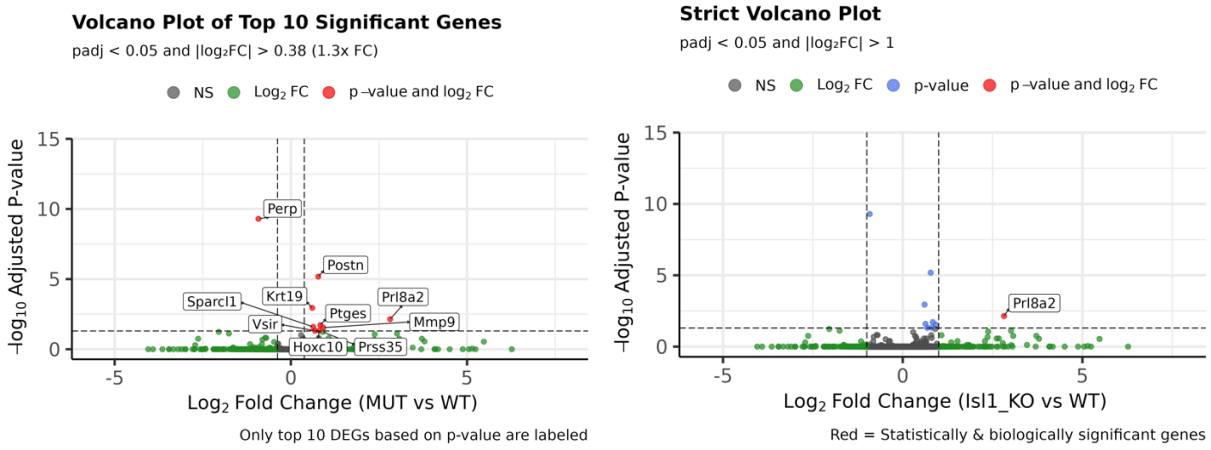


Figure 6. Volcano Plots of Differential Gene Expression Between Isl1 Mutant and Wild-Type Samples (A) Volcano Plot with Relaxed Thresholds (padj < 0.05 and |log₂FC| > 0.38) (B) Volcano Plot with Stringent Thresholds (padj < 0.05 and |log₂FC| > 1)

Gene Set Enrichment Analysis (GSEA): To gain insight into the biological pathways altered between Isl1 mutant (MUT) and wild-type (WT) samples, Gene Set Enrichment Analysis (GSEA) was conducted using the fgsea package and MSigDB Hallmark gene sets for *Mus musculus*. Genes were ranked by the Wald statistic from DESeq2, capturing both the magnitude and direction of differential expression(figure 7). Ensembl gene IDs were converted to Entrez IDs using biomaRt, with duplicates resolved by selecting the most significant gene per ID(Tomlinson, 2025).

A total of 28 hallmark gene sets were significantly enriched at adjusted p < 0.05, reflecting major biological programs affected by the loss of Isl1. The top enriched pathways, based on Normalized Enrichment Score (NES), included:

- HALLMARK_HYPOXIA (NES = 2.30, FDR ≈ 7.1e–13)
- HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION (EMT) (NES = 2.15, FDR ≈ 8.9e–10)
- HALLMARK_CHOLESTEROL_HOMEOSTASIS (NES = 2.27, FDR ≈ 1.3e–07)
- Additional top-ranked pathways: GLYCOLYSIS, APICAL_JUNCTION, ADIPOGENESIS

As shown in Figure 5, the EMT enrichment plot highlights a concentration of EMT-related genes near the top of the ranked list, indicating coordinated upregulation of this pathway in MUT samples relative to WT.

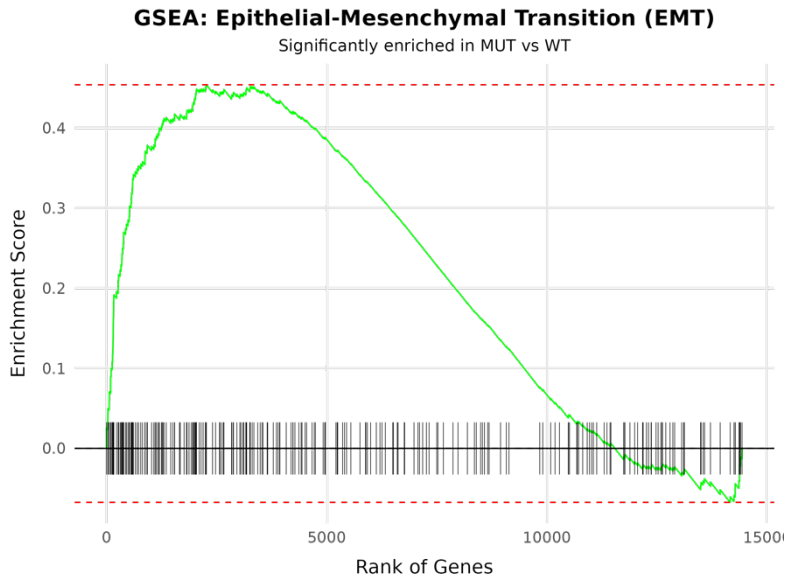


Figure 7. GSEA Enrichment Plot for EMT Hallmark Pathway

Enrichment Score Curve for EMT genes: The green line represents the running enrichment score for EMT-associated genes across the DESeq2-ranked list. The upward shift early in the curve indicates that EMT genes are strongly upregulated in Isl1 mutant samples. Black tick marks show the positions of EMT genes within the ranked list, and red dashed lines denote significance thresholds.

In addition to confirming known vascular and progenitor-related changes, we observed notable upregulation trends in genes such as *Thy1*, *Pdpn*, *Procr*, and *Lgals3* in the Isl1_KO allantois, consistent with enhanced mesenchymal, stromal, and immune-related transcriptional activity. Although these genes did not consistently meet the adjusted significance threshold (padj < 0.05), their directional changes suggest potentially meaningful roles that warrant further investigation. *Thy1* and *Pdpn* are stromal markers associated with fibroblast-like cells and extracellular matrix (ECM) remodelling(Griffin et al., 2020; Satija et al., 2015), while Procr marks endothelial progenitors(Critser & Yoder, 2010). Lgals3 (Galectin-3) is linked to placental immune regulation(Than et al., 2012). Together, these findings indicate that Isl1 loss shifts transcriptional programs toward stromal differentiation and immune signaling, in addition to impaired vascular development.

Discussion

This RNA-seq analysis of Isl1 knockout (KO) versus wild-type (WT) mouse allantois samples at embryonic day 8.25 revealed a clear genotype-driven shift in transcriptomic profiles. Principal component analysis demonstrated strong clustering by genotype, confirming the consistency of biological replicates. Differential expression analysis identified 9 genes significantly altered at FDR < 0.05, with additional candidates emerging under relaxed thresholds(Tomlinson, 2025; Zhu et al., 2024). Notable upregulated genes in Isl1_KO samples included *Prl8a2*, *Postn*, and *Ptges*, while key progenitor-associated genes such as *Lefty2* and *Isl1* were significantly downregulated(Tomlinson, 2025; Zhu et al., 2024).

To rank the top DEGs, we prioritized genes based on adjusted p-values and biologically relevant fold changes. This approach balanced statistical robustness with functional significance. Enrichment analysis using the fgsea package and MSigDB Hallmark gene sets highlighted upregulation of pathways including Epithelial-Mesenchymal Transition (EMT), Hypoxia, and Cholesterol Homeostasis in mutants, suggesting that loss of Isl1 promotes mesenchymal differentiation and stress-response programs(Tomlinson, 2025; Zhu et al., 2024).

Notably, we observed upregulation trends in several stromal and immune-associated genes, including *Thy1*, *Pdpn*, *Procr*, and *Lgals3*. Although these genes did not consistently meet the adjusted significance threshold, their expression patterns suggest enhanced fibroblast-like identity, extracellular matrix remodeling, endothelial progenitor presence, and immune signaling at the maternal–fetal interface. These findings imply that Isl1 may play a broader role beyond vascular and progenitor regulation—specifically in restraining stromal expansion and modulating immune-related transcriptional activity in the developing allantois.

Overall, the results suggest that Isl1 functions as a key regulator of mesodermal lineage balance, promoting endothelial identity while limiting mesenchymal and inflammatory transcriptional programs. These findings offer new insights into Isl1's role in early placental development and demonstrate the power of RNA-seq in uncovering regulatory networks underlying embryonic tissue specification.

Conclusion

Through a comprehensive RNA-seq analysis of Isl1 knockout versus wild-type allantois samples, we identified key transcriptional changes associated with early mesodermal development. Our results confirm the known role of Isl1 in regulating vascular and progenitor gene programs and further reveal upregulation of mesenchymal, stromal, and immune-related markers in its absence. Differential expression and enrichment analyses indicate that Isl1 acts as a transcriptional gatekeeper, promoting endothelial identity while restraining mesenchymal and immune-associated pathways. These findings extend our understanding of Isl1's developmental functions and underscore its broader role in coordinating cell fate decisions within extraembryonic tissues.

References:

Critser, P. J., & Yoder, M. C. (2010). Endothelial colony-forming cell role in neoangiogenesis and tissue repair. *Curr Opin Organ Transplant*, *15*(1), 68-72. https://doi.org/10.1097/MOT.0b013e32833454b5

Griffin, M. F., desJardins-Park, H. E., Mascharak, S., Borrelli, M. R., & Longaker, M. T. (2020). Understanding the impact of fibroblast heterogeneity on skin fibrosis. *Dis Model Mech*, *13*(6). https://doi.org/10.1242/dmm.044164

Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., & Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol*, *33*(5), 495-502. https://doi.org/10.1038/nbt.3192

Than, N. G., Romero, R., Kim, C. J., McGowen, M. R., Papp, Z., & Wildman, D. E. (2012). Galectins: guardians of eutherian pregnancy at the maternal-fetal interface. *Trends Endocrinol Metab*, *23*(1), 23-31. https://doi.org/10.1016/j.tem.2011.09.003

Tomlinson, S. (2025). Functional Genomics and Transciptomics Lecture Notes. In.

Zhu, Z., Zou, Q., Wang, C., Li, D., Yang, Y., Xiao, Y., Jin, Y., Yan, J., Luo, L., Sun, Y., & Liang, X. (2024). Isl Identifies the Extraembryonic Mesodermal/Allantois Progenitors and is Required for Placenta Morphogenesis and Vasculature Formation. *Adv Sci (Weinh)*, *11*(32), e2400238. https://doi.org/10.1002/advs.202400238