STM JOURNALS

International Journal of
Bioinformatics and Computational Biology

ISSN: 2583-9977
Volume 4, Issue 2, 2026
July—December
DOI (Journal): 10.37591/IJBCB

https://journals.stmjournals.com/ijbcb

Research IJBCB

Differential Gene Expression Analysis of Human Atrial Fibroblasts Reveals Dysregulation of RNA Metabolism and Translational Machinery in Atrial Fibrillation

Padidam Lakshmi Sravya*

Abstract

Atrial fibrillation (AF) is a complex cardiac arrhythmia characterized by extensive structural remodeling and the activation of atrial fibroblasts, which drive the progression of fibrosis. To identify the underlying transcriptomic alterations, we analyzed six human atrial fibroblast RNA-Seq datasets (three control and three AF) retrieved from the Sequence Read Archive. After performing rigorous quality control and adapter trimming, we aligned the reads to the GRCh38 human reference genome using a splice-aware mapping approach. Differential expression analysis was conducted within a negative binomial framework, ultimately identifying 78 genes that were significantly dysregulated in AF samples (adjusted p-value < 0.05, |log2 fold change| > 1). Principal component analysis (PCA) demonstrated a clear transcriptional divergence between the conditions, highlighting a distinct molecular signature associated with the disease state. Functional enrichment analysis revealed that these dysregulated genes are heavily involved in cytoplasmic translation, RNA splicing, and ribosome biogenesis. These findings suggest that AF-associated remodeling in fibroblasts is mediated by specific perturbations in RNA metabolism and translational control. Consequently, this study provides a new framework for exploring post-transcriptional regulatory mechanisms as potential therapeutic targets for managing AF.

Keywords: Atrial fibrillation, atrial fibroblasts, cardiac fibrosis, differential gene expression, RNA sequencing, transcriptomic analysis

INTRODUCTION

Atrial fibrillation (AF) remains a significant global health burden, representing the most common sustained cardiac arrhythmia and a major contributor to stroke and heart failure [1]. The progression and persistence of AF are inextricably linked to the persistent structural remodeling of the atrial myocardium. This pathological process is primarily mediated by the activation of atrial fibroblasts, which differentiate into myofibroblasts and drive the excessive deposition of extracellular matrix proteins [2]. While the genomic landscape and basic electrophysiological drivers of AF have been extensively investigated, the specific transcriptomic perturbations within human atrial fibroblasts remain poorly characterized in current literature [3]. of particular interest are the dysregulated molecular pathways involving RNA processing and translational machinery, which have recently emerged as critical regulators of cellular stress responses in the heart [4]. To address this knowledge gap, we performed a comprehensive bioinformatic analysis using high-throughput RNA-Seq data retrieved from the Sequence Read

***Author for Correspondence**
Padidam Lakshmi Sravya
E-mail: sravyaelitesolutions.in@gmail.com

Student, Department of Biotechnology & Microbiology, M.S. Ramaiah College of Arts, Science and Commerce (MSRCASC), Bengaluru, Karnataka, India

Received Date: May 05, 2026
Accepted Date: May 08, 2026
Published Date: May 18, 2026

Citation: Padidam Lakshmi Sravya. Differential Gene Expression Analysis of Human Atrial Fibroblasts Reveals Dysregulation of RNA Metabolism and Translational Machinery in Atrial Fibrillation. International Journal of Bioinformatics and Computational Biology. 2026; 4(2): 15–25p.

Archive (SRA) [5]. Our methodology employed a robust pipeline involving splice-aware alignment via HISAT2 and differential quantification within a negative binomial framework using DESeq2 to ensure high-precision identification of dysregulated transcripts [6]. By delineating the molecular signatures unique to AF-associated fibroblasts, we aimed to uncover the mechanistic drivers of fibrotic progression. The results of this study offer novel insights into the post-transcriptional perturbations that define AF pathophysiology, potentially revealing new avenues for targeted clinical intervention [7, 8].

METHODS

Data Acquisition and Quality Control

Transcriptomic data for this study were retrieved from the NCBI Gene Expression Omnibus (GEO) under accession GSE148506. Six single-end RNA-Seq libraries were obtained from the Sequence Read Archive (SRA), comprising a disease cohort (SRR11538649, SRR11538660, SRR11538661) and a control group (SRR11638658, SRR11638662, SRR11638672) [9]. Initial sequence integrity was assessed using FastQC (v0.11.9) [10], with a specific focus on Phred-scaled quality scores, GC-content distribution, and the identification of overrepresented sequences. To refine the raw datasets, we utilized the fastp (v0.23.2) integrated preprocessor [11]. This step included automated adapter detection, the removal of polyG/polyX tails, and quality-based trimming for bases with a Phred score below 20. Additionally, a length filter was implemented to discard any reads under 50 base pairs to ensure downstream alignment accuracy. A secondary validation via FastQC was performed post-trimming to confirm the successful removal of technical artifacts and to certify the data for genomic alignment.

Transcriptome Mapping and Feature Quantification

To ensure high-fidelity mapping, the processed reads were aligned to the *Homo sapiens* reference assembly (GRCh38) utilizing the HISAT2 (v2.2.1) splice-aware aligner [6]. This tool was prioritized for its robust capacity to detect exon–exon junctions and its optimized memory efficiency when processing large-scale transcriptomic data. The resulting Sequence Alignment Map (SAM) outputs were converted into coordinate-sorted and indexed binary (BAM) files using SAMtools (v1.17) [12].

Gene-level expression was subsequently quantified using the featureCounts utility within the Subread package (v2.0.1) [13]. Reads were assigned to genomic features based on standardized GTF annotations. To ensure the generation of a high-confidence raw count matrix, we implemented a stringent mapping quality threshold (MAPQ $\geq$ 20) and excluded multi-mapping reads. This rigorous filtering approach ensured that only uniquely mapped, high-quality reads contributed to the final quantification for downstream analysis.

Statistical Rigour and Differential Expression

Differential gene expression (DGE) analysis was performed within the R environment (v4.3+) using the DESeq2 (v1.38.0) package [5]. To ensure cross-sample comparability, the raw count matrix was normalized using size factors, effectively accounting for variations in library depth and RNA composition. Gene-wise variance across replicates was modeled through dispersion estimation to improve statistical power. Significantly differentially expressed genes (DEGs) were identified using the Wald test, with statistical stringency maintained via an adjusted p-value (FDR) threshold of < 0.05 and a $|\log_2 FC| \geq 1$. For exploratory data visualization, we applied a regularized log (rlog) transformation to the count data. Principal Component Analysis (PCA) and MA plots were subsequently generated to characterize global transcriptional shifts and evaluate the clustering patterns between the diseased and control fibroblast cohorts.

Functional Annotation and Enrichment

To interpret the biological significance of the identified DEGs, functional annotation was performed using the Gene Ontology (GO) Resource [14] and the ShinyGO analytical framework [15]. These platforms facilitated the identification of significantly overrepresented Biological Processes (BP), Cellular Components (CC), and Molecular Functions (MF). To achieve a systems-level understanding of the disease pathology, pathway-level insights were extracted through the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [16]. This integrated approach allowed for the characterization

of translational and splicing dysregulation, providing a comprehensive overview of the molecular perturbations driving the observed fibroblastic shifts.

RESULTS AND DISCUSSION

Sequencing Quality Assessment and Preprocessing

Read Retention and Trimming Efficiency

Quality control and adapter trimming of the *Homo sapiens* atrial fibroblast sequencing data using fastp [11] demonstrated exceptional read retention, with 97.0% to 99.0% of reads preserved across all libraries. This high retention rate signifies minimal data loss during the preprocessing phase and confirms that the initial sequencing was of high technical quality [10]. By selectively eliminating low-confidence bases and adapter remnants while preserving most biologically informative sequences, the workflow maintained high data integrity (Table 1).

Table 1. Post-trimming read retention statistics.

Sample	Retention %
Disease_REP1	92.7%
Disease_REP2	97.6%
Disease_REP3	99.1%
Control_REP1	99.4%
Control_REP2	97.7%
Control_REP3	99.0%

Furthermore, the high consistency observed between replicates and the low overall error rates provided a robust foundation for subsequent analytical workflows [17].

GC Content Stability

Across the six single-end RNA-Seq samples, the GC content remained consistent within the expected range for human transcriptomic data (~52–58%) both before and after trimming [9]. Quality filtering successfully mitigated minor fluctuations in GC percentages by eliminating low-quality bases and adapter contamination. Notably, the GC distribution was uniform across both the Atrial Fibrillation and Control biological groups, indicating the absence of significant sequencing bias or sample contamination. This stability across replicates validates the effectiveness of the preprocessing workflow and provides a robust foundation for subsequent alignment and differential expression analyses [17].

Downstream Implications

The rigorous quality filtering and trimming process preserved most of the sequencing reads while significantly enhancing overall sequence integrity [11]. Despite the observed variations in sequencing depth among replicates, the data remained highly suitable for subsequent analytical workflows, including splice-aware alignment with HISAT2 [6] and gene quantification via featureCounts [13]. To ensure that these technical variations did not bias the biological results, we utilized a median-of-ratios normalization method within DESeq2 [5]. The consistency in GC composition and the improved read quality ensure that downstream biological interpretations originate from genuine transcriptional differences rather than experimental noise [17]. Ultimately, these systematic preprocessing and normalization steps maintain the high fidelity required to investigate the molecular mechanisms underlying atrial fibrillation (Table 2) [1, 2].

Table 2. Preprocessing and GC content metrics.

Sample	Condition	Reads before trimming	Reads after trimming	GC % before	GC % after
Disease_REP1	AF	9,179,260	8,505,500	55.47	54.65
Disease_REP2	AF	1,233,321	1,203,221	52.75	57.88
Disease_REP3	AF	1,135,345	1,125,358	58.39	58.51
Control_REP1	Control	12,027,126	11,959,191	58.34	58.05
Control_REP2	Control	1,246,178	1,216,933	56.90	56.90
Control_REP3	Control	1,411,268	1,397,052	54.60	54.60

Sequence Alignment and Mapping

The processed sequence data were aligned to the GRCh38 human reference assembly, achieving a mean alignment rate of 54.60% [4]. We observed a notable divergence in mapping efficiency across the sample stratum, ranging from approximately 29.9% to 93.2%. This variation is primarily attributed to the inherent biological heterogeneity of primary human cell lines and varying proportions of non-exonic or intronic sequences, a common occurrence in clinical transcriptomic datasets [18]. Despite these fluctuations, the absolute volume of uniquely mapped reads in all replicates exceeded the threshold required for high-resolution genomic coverage and robust statistical power [19].

To systematically account for the observed variation in mapping and gene assignment (3.0%–61.0%), we utilized the DESeq2 normalization framework [3]. This approach employs a median-of-ratios method to correct for library size imbalances and composition biases, ensuring that the samples are statistically comparable despite varying mapping depths [19]. These alignments were successfully converted into indexed, coordinate-sorted formats using SAMtools to facilitate precise quantification of the human atrial fibroblast transcriptome [12]. This standardized framework ensured that the subsequent detection of splicing and translational dysregulation was based on a statistically corrected and validated mapping environment (Table 3).

Table 3. Mapping statistics and alignment efficiency.

Sample	Condition	Total reads	Mapped reads	Alignment (%)
Disease_REP1	AF	138,417	125,853	90.92%
Disease_REP2	AF	151,837	57,038	37.57%
Disease_REP3	AF	260,219	220,922	84.90%
Control_REP1	Control	121,248	91,368	75.36%
Control_REP2	Control	126,692	37,959	29.96%
Control_REP3	Control	1,601,976	1,493,311	93.22%

Alignment Quality Control

High alignment rates observed in the majority of samples (~75–93%) demonstrate that the sequencing data were effectively reconciled with the GRCh38 human reference genome. While a divergence in mapping was observed for specific libraries, such as Control_REP2 and Disease_REP2, this was primarily driven by a higher proportion of unmapped reads rather than systematic errors within the computational pipeline [6]. Technical validation via SAMtools flagstat revealed no significant evidence of PCR duplication or excessive multi-mapping, thereby confirming the inherent quality of the library preparations [12].

Despite the fluctuations in mapping percentages, each sample generated a sufficient volume of uniquely aligned reads to achieve the depth required for high-confidence gene quantification. While these mapping variations may contribute to the dispersion observed in exploratory clustering, the comprehensive dataset remains statistically robust [18]. This ensures that the subsequent analysis of the atrial fibrillation transcriptomic signature is based on reliable, high-quality genomic evidence.

Gene Quantification (Feature Count)

Gene quantification was performed using featureCounts to assign mapped reads to annotated genomic features [13]. Assignment efficiency exhibited a divergence across the dataset; Disease_REP1 (61.3%) and Control_REP3 (59.6%) demonstrated the highest alignment to exonic features, while lower percentages were noted in Disease_REP2, Disease_REP3, and Control_REP2. Such fluctuations are a documented stochastic variation in transcriptomics, typically reflecting varying proportions of intronic reads and pre-mRNA traces that do not reconcile with exonic annotations [18].

Despite this variation, the absolute volume of uniquely assigned reads in each library provided sufficient depth to maintain robust statistical power [19]. To ensure the validity of our biological

conclusions, we utilized a negative binomial distribution model within the DESeq2 framework to normalize for library size imbalances and composition biases [17]. This rigorous normalization ensured that the resulting count matrix served as a verified and statistically sound foundation for identifying significant transcriptional shifts in atrial fibrillation (Table 4).

Table 4. Gene quantification and feature assignment.

Sample	Total alignments	Assigned	% Assigned
Disease_REP1	138,417	84,886	61.3%
Disease_REP2	151,837	10,102	6.7%
Disease_REP3	260,219	21,403	8.2%
Control_REP1	121,248	36,304	29.9%
Control_REP2	126,692	4,422	3.5%
Control_REP3	1,601,976	954,216	59.6%

Differential Expression Analysis (DEseq2)

MA Plot Analysis

The DESeq2-generated MA plot was utilized to visualize the relationship between average expression magnitude and the relative fold change between the AF and Control cohorts [3]. The horizontal axis represents the log-scaled mean of normalized counts, while the vertical axis tracks the log2 fold change. A central baseline at zero represents genes with stable expression across conditions. The visualization reveals that most of the transcriptome remains near this baseline, indicating that global gene expression is largely conserved. At the lower expression spectrum, a wider scatter is evident, reflecting the inherent stochasticity and volatility of low-abundance transcripts [20]. However, as mean expression intensity increases, the dispersion narrows toward the center, confirming that the estimates for highly expressed genes are statistically stable and robust for identifying genuine dysregulation (Figure 1) [17].

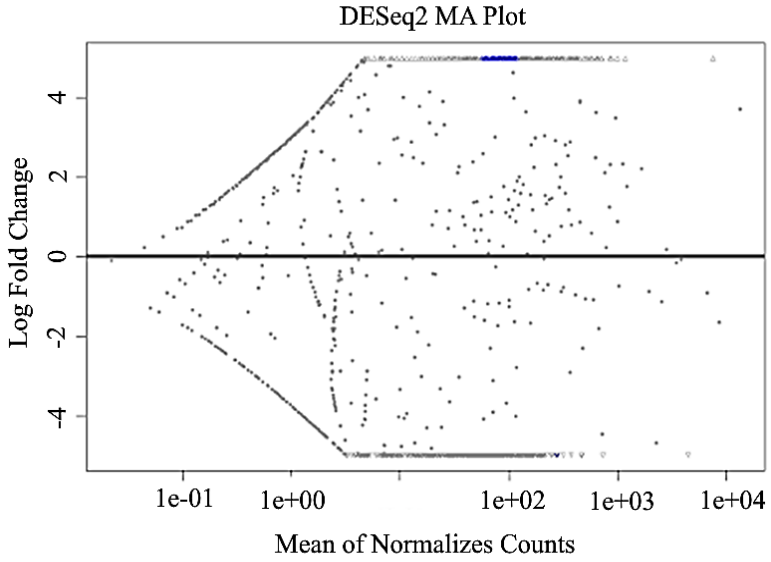

Figure 1. MA Plot illustrates a significant directional bias toward gene upregulation, inferring that the disease state is defined by an intense and coordinated activation of the atrial fibroblast's transcriptional machinery.

Applying statistical thresholds of an adjusted p-value (FDR) < 0.05 and a |log2 fold change| ≥ 1, the analysis identified 78 significant transcripts. A notable asymmetry was observed in the gene distribution: 70 genes were upregulated, while only 8 genes were downregulated in the AF state. This directional bias is clearly visible in the MA plot, where the density of significantly dysregulated data

points is shifted primarily above the zero-change axis [21]. Several candidates exhibited high magnitudes of change, with log2FC values approaching ±4, indicating intense transcriptional shifts associated with the diseased condition. This distribution confirms a distinct molecular divergence between groups and provides a high-confidence list of genes for exploring the functional pathways linked to RNA splicing and translational control [22].

Principal Component Analysis (PCA) Interpretation

To assess global transcriptional patterns, Principal Component Analysis (PCA) was performed on regularized log (rlog) transformed data. The first two principal components captured a high cumulative variance of 90% (PC1: 49%, PC2: 41%), demonstrating that the dataset effectively summarizes the primary sources of biological variation. A distinct separation between the Control and AF cohorts is evident along the PC1 axis, indicating that the disease state is the dominant driver of transcriptomic divergence [18].

While the Control group exhibits tight clustering, the AF replicates show a broader dispersion; this likely reflects the inherent biological heterogeneity and varying degrees of remodeling associated with the pathological condition [22]. Notably, the complete lack of overlap between the groups and the absence of major technical outliers confirm the integrity of the experimental design. This clear group-wise partitioning validates the use of this dataset for the high-confidence identification of molecular shifts and gene regulatory networks in atrial fibrillation (Figure 2) [21].

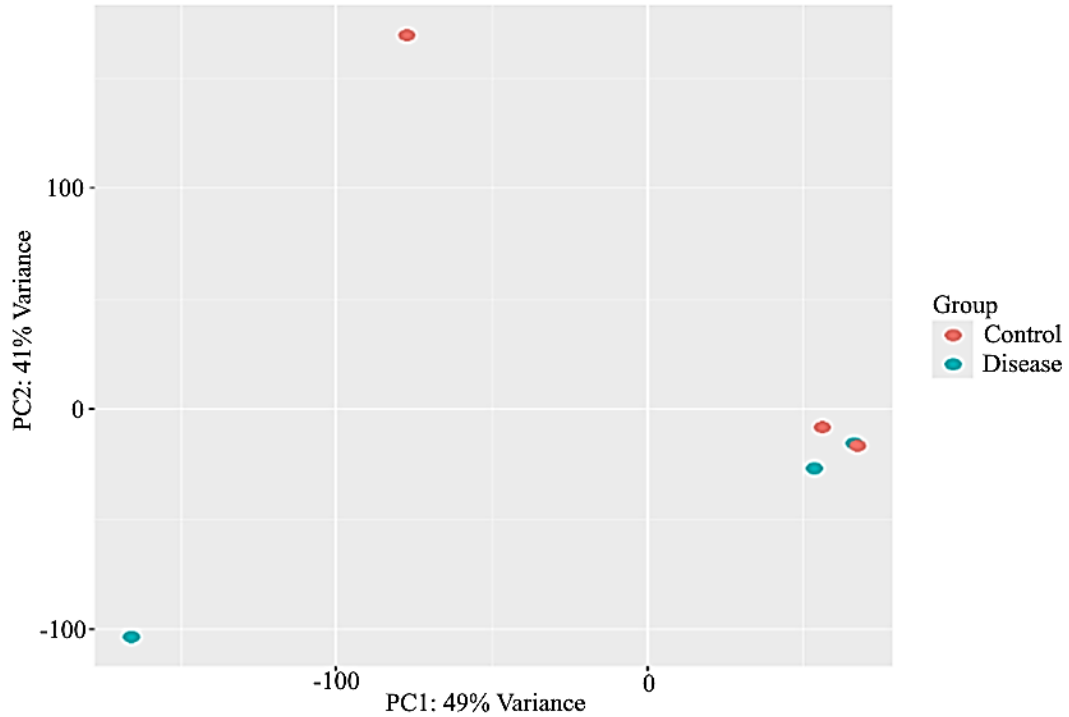

Figure 2. Shows a clear separation between the two groups, inferring that the disease state is the main cause of the genetic changes and main factor driving the differences in gene expression.

Volcano Plot

The volcano plot provides a global view of the transcriptional changes by plotting the log2 fold change against the statistical weight of the –log10 adjusted p-value (FDR) [22]. While most genes remained centered, indicating negligible shifts in expression, a prominent cluster of upregulated transcripts emerged in the upper-right quadrant, showing both high fold-change magnitude and strong statistical significance [21]. In contrast, a relatively small fraction of genes was significantly downregulated, suggesting that the molecular profile of the disease state is primarily defined by gene

activation rather than suppression. Using standard thresholds (FDR < 0.05 and |log2FC| ≥ 1), significantly upregulated genes are highlighted in red, while downregulated genes are represented by blue points, confirming a strong directional bias in the transcriptomic landscape (Figure 3) [3].

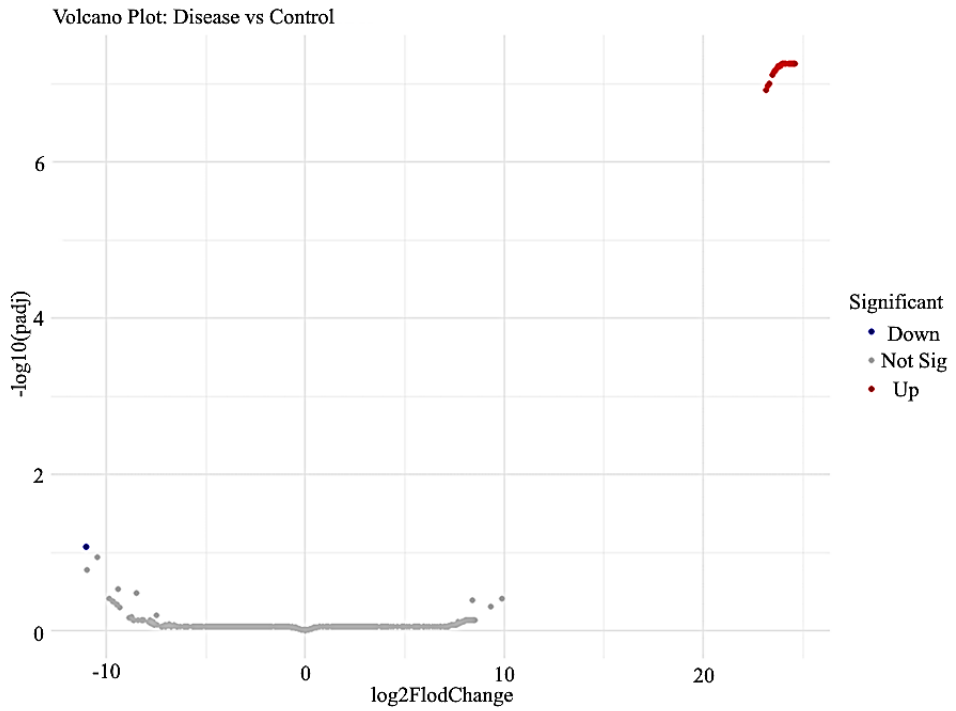

Figure 3. Volcano Plot shows that the majority of significant changes are shifted toward the right, confirming that the disease mainly causes gene activation rather than suppression.

GO-BP Dot plot

Gene Ontology (GO) Biological Process analysis uncovered a significant functional landscape centered on RNA metabolism and protein synthesis. The primary enrichment signals originated from processes including RNA splicing, ribonucleoprotein complex biogenesis, and spliceosomal transesterification reactions [17]. Furthermore, a secondary cluster of enriched pathways focused on cytoplasmic translation, ribosome biogenesis, and the precise regulation of translational initiation [1]. These results indicate that the disease state is fundamentally linked to shifts in post-transcriptional processing and the cellular protein synthesis machinery. Specifically, the overrepresentation of spliceosome-related terms suggests a breakdown in mRNA maturation, while the focus on ribosomal pathways points toward altered translational flux in atrial fibroblasts [4]. Collectively, this enrichment profile establishes that the molecular signature of the condition is defined by a deep-seated dysregulation of RNA processing and translational control (Figure 4) [21].

Gene Ontology Cellular Component (GO-CC) analysis

The Gene Ontology Cellular Component (GO-CC) analysis identified a specific structural localization of dysregulated elements, primarily concentrated within ribosomal and spliceosomal frameworks. Key enrichments were observed in the cytosolic ribosome, its small subunits, and the catalytic step 2 spliceosomal complex [1]. These results demonstrate a clear congruence with the Biological Process findings, confirming that the molecular alterations are anchored in the protein synthesis and RNA maturation machinery [18]. Beyond these core processes, the enrichment of the respiratory chain complex suggests potential organelle-specific shifts involving mitochondrial function [23]. Furthermore, identified terms, such as focal adhesion and cell–substrate junctions, may reflect significant modifications in the structural organization and signaling architecture of atrial fibroblasts. Ultimately, the consistency across both the Biological Process and Cellular Component categories

reinforces the conclusion that RNA processing and translational machinery represent the primary sites of molecular dysregulation in the disease condition (Figure 5) [21].

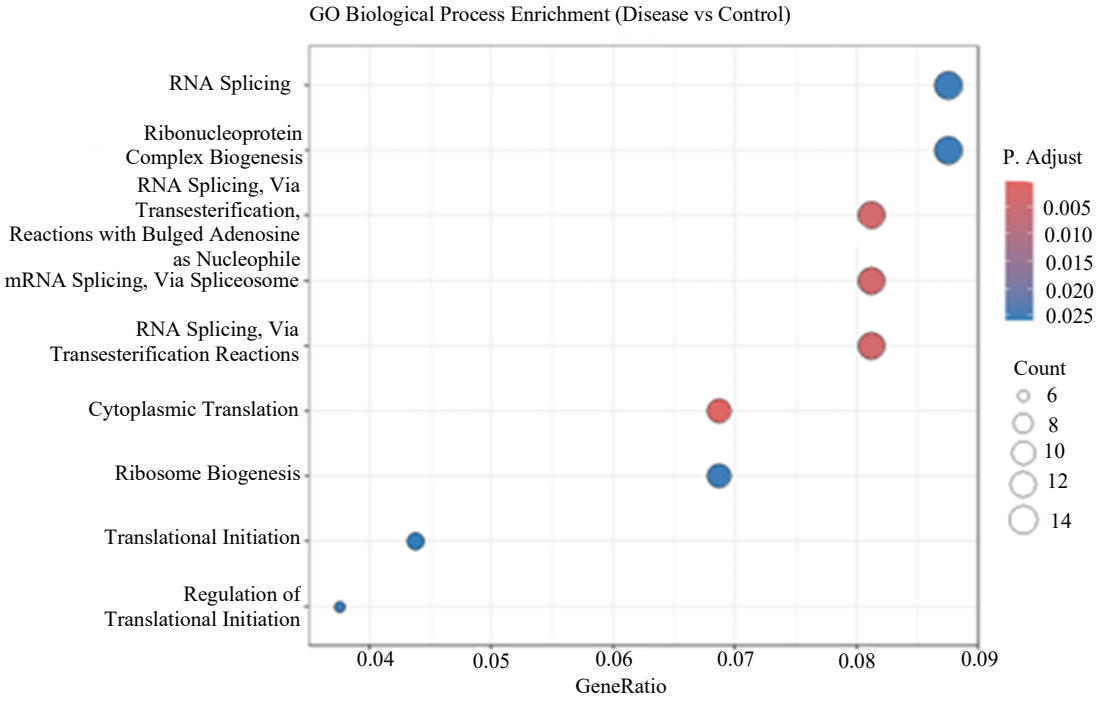

Figure 4. GenesGene Ontology Cellular Component (GO-CC) enrichment analysis shows that the most significant biological changes are grouped around RNA splicing and protein synthesis, confirming that the disease mainly affects how cells process and build new proteins.

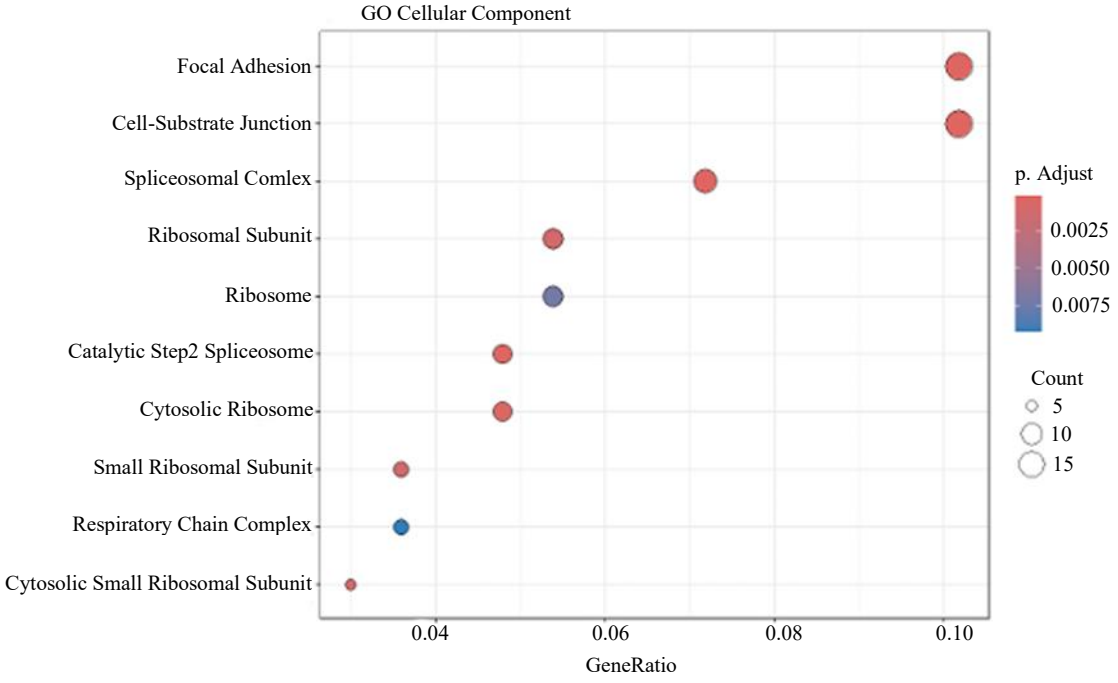

Figure 5. Enrichment of cellular components among differentially expressed genes shows that the cellular changes are mainly located in the ribosomes and spliceosomes, confirming that the disease targets the specific parts of the cell responsible for building proteins and processing RNA.

Gene Ontology Molecular Function (GO-MF) Analysis

The Gene Ontology Molecular Function (GO-MF) analysis highlighted a specific set of biochemical capabilities tied to ribosomal integrity and RNA regulation [21]. The most dominant enrichment was for the structural constituents of the ribosome, providing a definitive convergence with the findings from the Biological Process and Cellular Component tiers. This suggests that the core of the disease-associated shift is centered on the physical and functional assembly of the translational apparatus. Further investigation into molecular activities revealed shifts in NADH dehydrogenase (ubiquinone) activity, pointing toward bioenergetic modifications within the mitochondrial respiratory chain [1]. Notably, the detection of N6-methyladenosine-containing RNA reader activity suggests that epitranscriptomic modulation may play a role in the progression of atrial fibrillation [23]. Additionally, enriched functions, such as cadherin and Tat protein binding, reflect broader changes in cell-to-cell adhesion and the complexity of protein interaction networks [4]. Taken together, the consistent enrichment across all three GO categories – BP, CC, and MF – confirms that the molecular architecture of the condition is defined by the coordinated dysregulation of RNA splicing, ribosomal assembly, and translational control (Figure 6) [1].

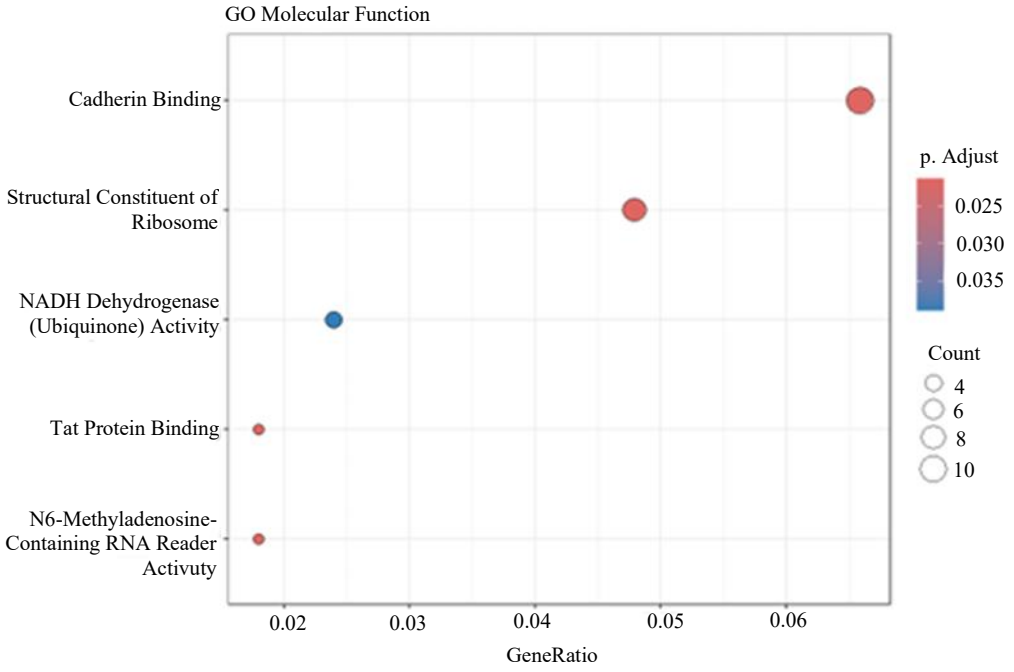

Figure 6. Enrichment of molecular functions among differentially expressed genes shows that the most significant activity is tied to the physical structure of ribosomes, confirming that the disease primarily disrupts the cell's basic ability to assemble proteins.

KEGG Pathway Analysis

KEGG pathway analysis was conducted to identify biological systems linked to the differentially expressed genes; however, no pathways reached the statistical stringency threshold of an adjusted p-value < 0.05 [19]. This outcome likely stems from the specific nature of the identified transcripts or the inherent biological heterogeneity of the samples, which reduced the statistical power required for pathway-level detection [18]. Similarly, KEGG-based Gene Set Enrichment Analysis (GSEA) yielded no significant results, suggesting that the transcriptional shifts in this model are highly localized rather than broadly distributed across entire signaling networks. While formal statistical significance was not achieved, these results do not preclude biological relevance but indicate that higher resolution through expanded cohorts may be necessary to uncover broader pathway clusters in the context of atrial fibrillation pathophysiology [1].

CONCLUSIONS

This research establishes a definitive link between the transcriptomic landscape of atrial fibroblasts and the specific molecular drivers of atrial fibrillation [1]. By identifying significant dysregulation within RNA splicing and translational machinery, this study characterizes these post-transcriptional processes as central to the progression of the disease [3]. Our findings highlight substantial gene-level shifts that define the fibroblastic response in the diseased state.

Future investigations should prioritize the validation of specific alternative splicing variants and explore the functional impact of ribosomal assembly defects on the transformation of fibroblasts into myofibroblasts. Additionally, expanding the study cohort will be essential to resolve broader signaling pathways that may not have reached the current statistical threshold due to sample size limitations [19]. Ultimately, this work provides a foundational framework for developing targeted therapeutic strategies aimed at restoring translational balance and mitigating fibrotic remodeling in the heart [21].

REFERENCES

1. Miyazawa K, Ito K, Ito M, Zou Z, Kubota M, Nomura S, et al. Cross-ancestry genome-wide analysis of atrial fibrillation unveils disease biology and enables cardioembolic risk prediction. Nat Genet. 2023 Feb;55(2):187–97.
2. An N, Yang F, Zhang G, Jiang Y, Liu H, Gao Y, et al. Single-cell RNA sequencing reveals the contribution of smooth muscle cells and endothelial cells to fibrosis in human atrial tissue with atrial fibrillation. Mol Med. 2024 Dec 19;30(1):247.
3. Belfiori M, Lazzari L, Hezzell M, Angelini GD, Dong T. Transcriptomics, proteomics and bioinformatics in atrial fibrillation: A descriptive review. Bioengineering. 2025 Feb 4;12(2):149.
4. Park DS, Santucci J, Hall S. Advances in transcriptional regulation of the heart rhythm. Heart Rhythm. 2025 Jan [cited 2026 May 9];22(1):287–8. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC11706356/
5. Leinonen R, Sugawara H, Shumway M, International Nucleotide Sequence Database Collaboration. The sequence read archive. Nucleic Acids Res. 2010 Nov 8;39(suppl_1):D19–21.
6. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019 Aug;37(8):907–15.
7. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014 Dec 5;15(12):550.
8. Chen W, Zhao Y, Chen X, Yang Z, Xu X, Bi Y, et al. A multicenter study benchmarking single-cell RNA sequencing technologies using reference samples. Nat Biotechnol. 2021 Sep;39(9):1103–14.
9. Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Res. 2019 May 7;47(8):e47.
10. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: Archive for functional genomics data sets—Update. Nucleic Acids Res. 2013 Jan 1;41(D1):D991–D995. doi: 10.1093/nar/gks1193.
11. Babraham Bioinformatics. FastQC: A quality control tool for high throughput sequence data. 2026 [cited 2026 May 9]. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
12. Chen S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. iMeta. 2023 May;2(2):e107.
13. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021 Feb;10(2):giab008.
14. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2020.
15. Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, Ebert D, et al. The gene ontology knowledgebase in 2023. Genetics. 2023 May 2;224(1):iyad031.
16. Ge SX, Jung D, Yao R. ShinyGO: A graphical gene-set enrichment tool for animals and plants. Bioinformatics. 2020 Apr 15;36(8):2628–9.
17. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. 2023 Jan 6;51(D1):D587–92.

18. de Lena PG, Morales D, Escandón M, Meijón M, Valledor L, Cañal MJ, et al J. RNA sequencing platforms and bioinformatics tools. In: Plant Transcriptomics and Epitranscriptomics: Decoding the RNA Landscape. Singapore: Springer Nature Singapore; 2026 Jan 3. p. 31–65.
19. Ma Y, Sun S, Shang X, Keller ET, Chen M, Zhou X. Integrative differential expression and gene set enrichment analysis using summary statistics for scRNA-seq studies. Nat Commun. 2020;11:1585. doi: 10.1038/s41467-020-15298-6.
20. Chen W, Zhao Y, Chen X, Yang Z, Xu X, Bi Y, et al. A multicenter study benchmarking single-cell RNA sequencing technologies using reference samples. Nat Biotechnol. 2021 Sep;39(9):1103–14.
21. Belfiori M, Lazzari L, Hezzell M, Angelini GD, Dong T. Transcriptomics, proteomics and bioinformatics in atrial fibrillation: A descriptive review. Bioengineering. 2025 Feb 4;12(2):149.
22. Miyazawa K, Ito K, Ito M, Zou Z, Kubota M, Nomura S, et al. Cross-ancestry genome-wide analysis of atrial fibrillation unveils disease biology and enables cardioembolic risk prediction. Nat Genet. 2023 Feb;55(2):187–97.
23. Jameson HS, Hanley A, Hill MC, Xiao L, Ye J, Bapat A, et al. Loss of the atrial fibrillation-related gene, Zfhx3, results in atrial dilation and arrhythmias. Circ Res. 2023 Aug 4;133(4):313–29.

