

TRANSCRIPTOME ANALYSIS OF LUNG TISSUES IN COPD PATIENTS

Identification of Dysregulated Genes Using RNA-Seq

Michelle Berendsen (S4459253), Lieke Boomstra (S5129656), Eda Ortaköse (S4505034), Laura Scholtens (S4973143), Emma Schut (S3971961), Delia Văărățiceanu (S5153379)

Master Biomedical Sciences Students, University of Groningen , Faculty of Science and Engineering

Introduction

Chronic obstructive pulmonary disease (COPD) is a progressive lung disorder characterised by persistent airflow limitation and abnormal inflammatory responses to harmful particles or gases, most commonly cigarette smoke. It affects over 250 million people worldwide and is a leading cause of morbidity and mortality.

While inflammation, oxidative stress, and tissue remodelling are known hallmarks of COPD, the precise molecular mechanisms that drive disease progression and irreversible airflow obstruction remain incompletely understood. Conventional studies have identified some key mediators, but large-scale molecular analyses are needed to uncover global gene expression changes associated with COPD.

Transcriptome profiling using RNA sequencing (RNA-Seq) provides a powerful tool to comprehensively examine these molecular alterations. In the study by Kim et al. (2015), lung tissue RNA from 98 COPD patients and 91 controls with normal lung function was sequenced using the Illumina HiSeq 2000 platform. Their analysis identified multiple dysregulated pathways, notably those involved in oxidative phosphorylation, protein degradation, and chromatin modification, suggesting mitochondrial dysfunction and altered gene regulation in COPD lungs.

In this case study, we re-analysed the Kim et al. dataset to identify differentially expressed genes (DEGs) between COPD and control lung tissues. Our goal was to highlight the most strongly upregulated and downregulated genes and visualise expression patterns that may underlie COPD pathophysiology.

Research Question: Which genes are differentially expressed in the lung tissue of COPD patients compared to healthy controls?

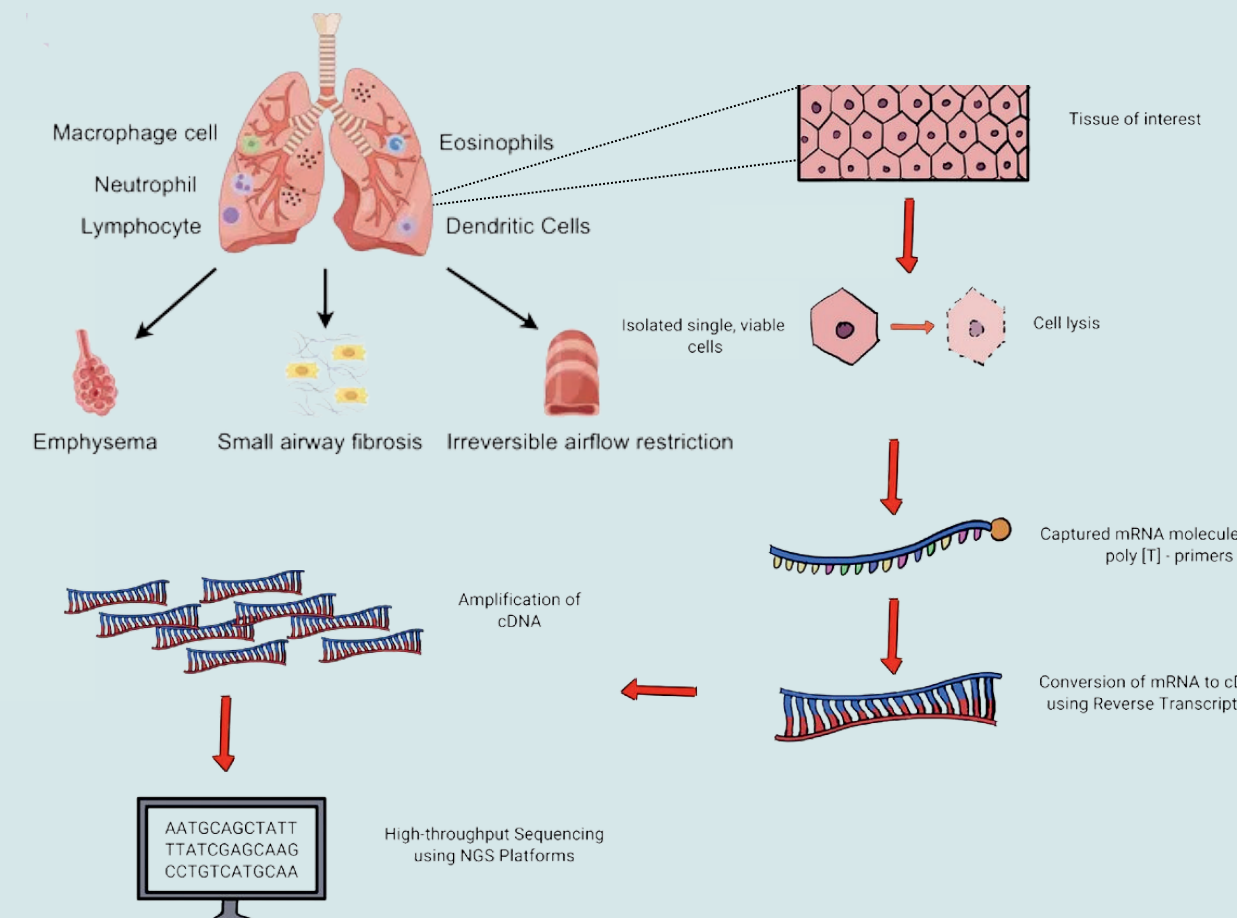


Figure 1: Experimental workflow for RNA-seq-based transcriptomic profiling of COPD and control lung tissues.

Methods

Bioinformatics Workflow

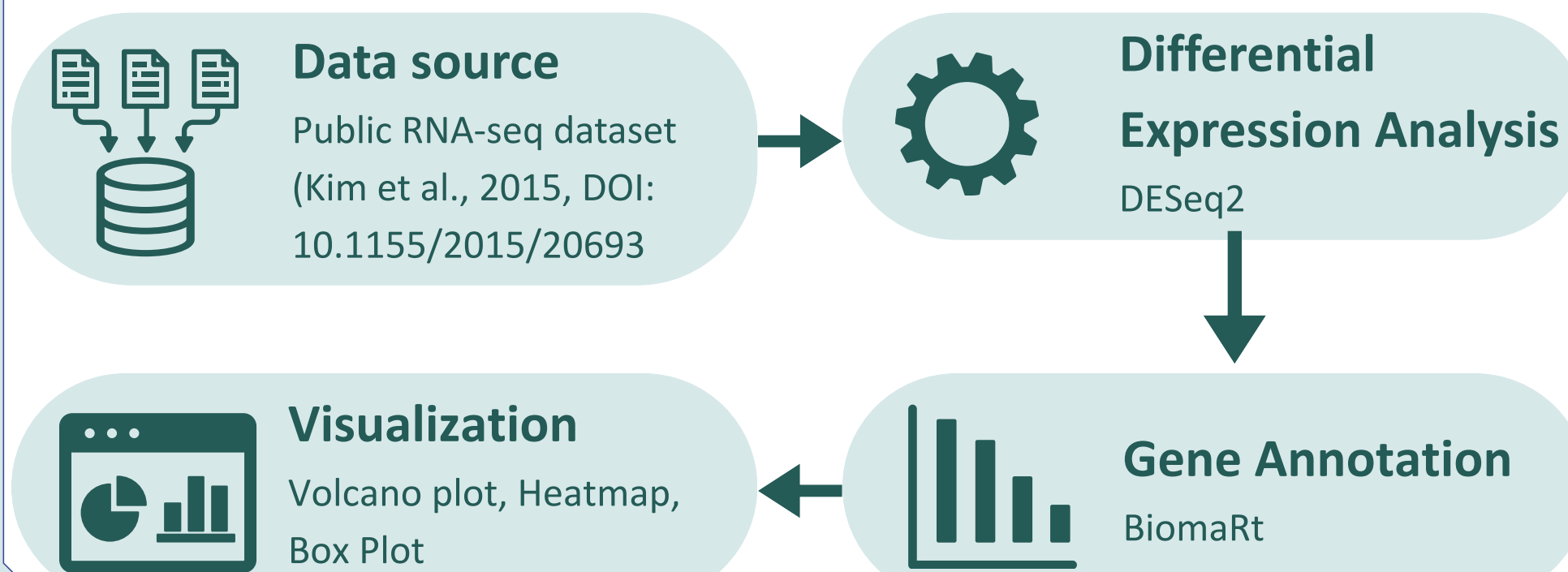


Figure 2: Workflow of RNA-seq data processing and analysis for differential gene expression in COPD. Schematic overview of the analytical pipeline used to identify differentially expressed genes between COPD and control lung tissues. RNA-seq data were obtained from a public dataset (Kim et al., 2015; DOI: 10.1155/2015/20693). Differential expression analysis was performed on 16,676 genes using DESeq2 in R. A total of 2,312 genes were identified as significantly differentially expressed (FDR < 0.01). Data visualisation was carried out using the ggplot2, pheatmap, biomart, reshape2, and ggrepel packages.

Differential Gene Expression in Lung Tissues

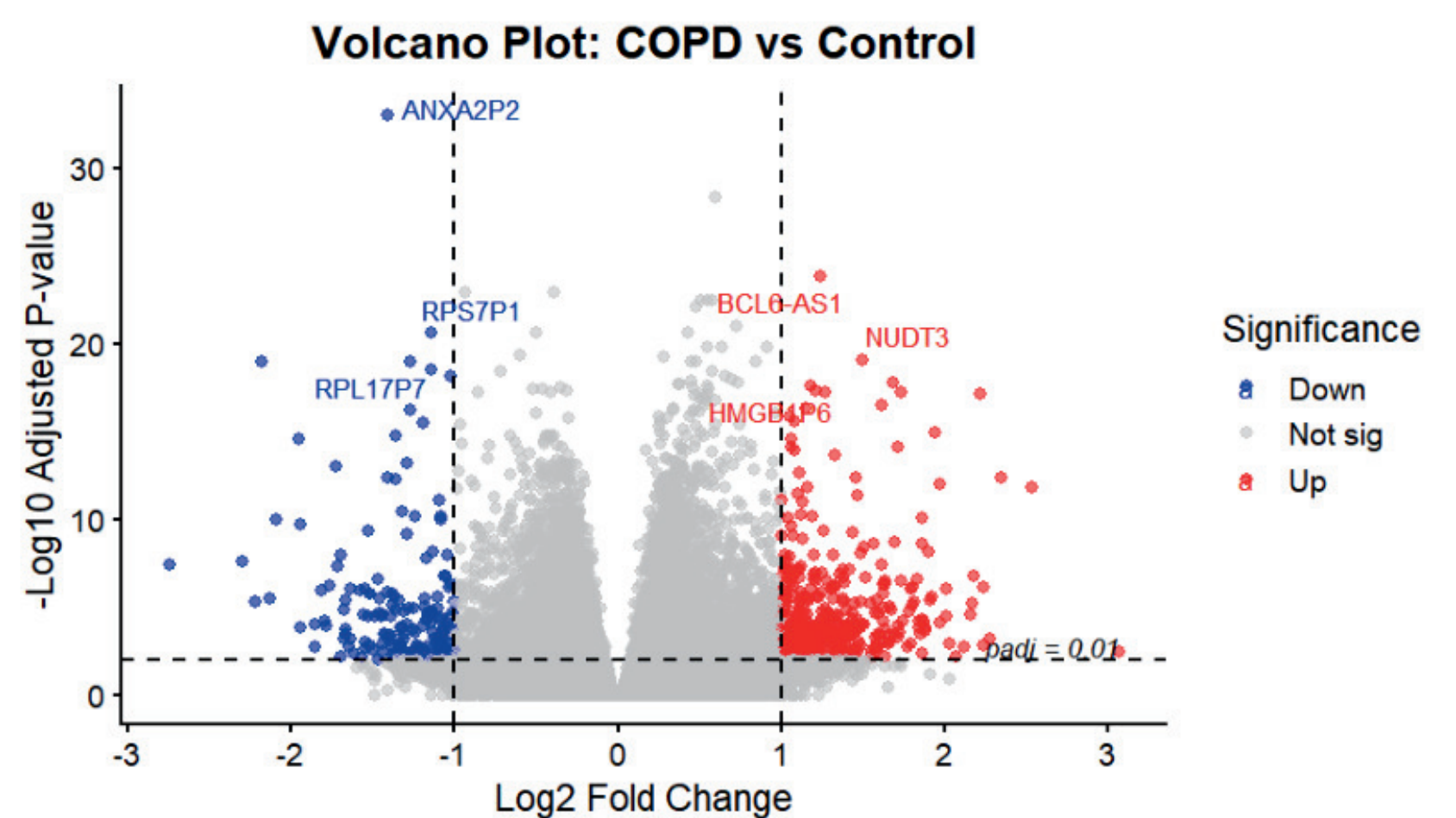


Figure 3: Volcano plot showing transcriptome-wide differential expression from RNA-seq analysis of lung tissues from subjects with COPD and matched healthy controls. Differential expression was assessed using the DESeq2 pipeline with the design formula ~ Disease after filtering for minimum read counts. The x-axis indicates $\log_2$ fold change (COPD vs. control), and the y-axis shows $-\log_{10}$ adjusted p-values (Benjamini-Hochberg correction). Genes with $|\log_2FC| > 1$ and adjusted $p < 0.01$ are highlighted in red (upregulated in COPD) or blue (downregulated), while nonsignificant genes appear in gray. Dashed lines mark statistical thresholds. The results reveal significant upregulation of oxidative stress-related and chromatin-modifying genes and downregulation of metabolic and structural genes, reflecting key molecular alterations associated with COPD progression.

Top 50 Differentially Expressed Genes

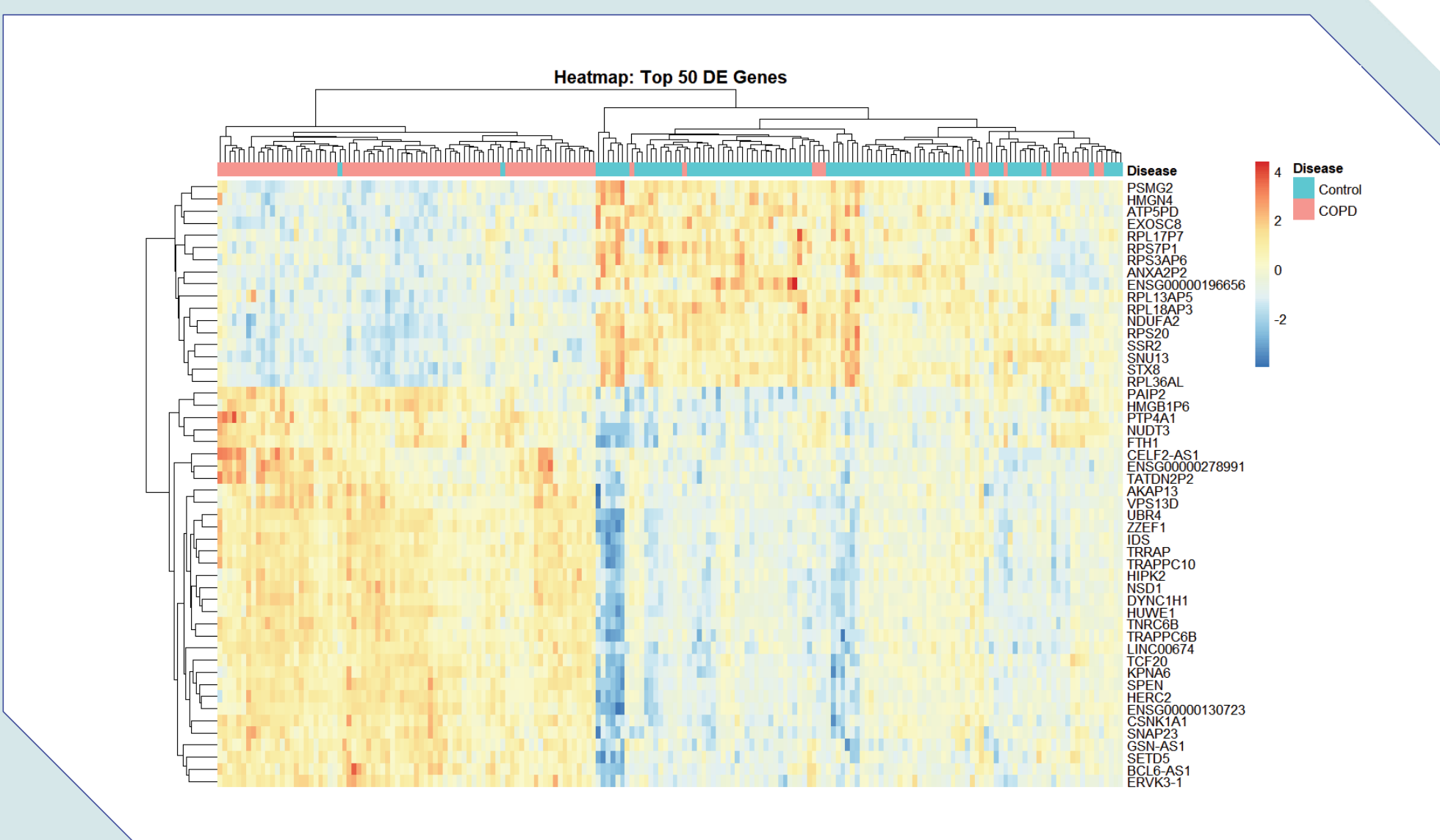


Figure 4: Heatmap showing relative expression patterns of the 50 most significantly differentially expressed genes identified by DESeq2 analysis of RNA-seq data from lung tissues of subjects with COPD and healthy controls. Genes were filtered for minimum expression and ranked by adjusted p-value. Variance-stabilizing transformation (VST) was applied to normalized counts, and expression values were mean-centered and scaled by row to emphasize relative differences. Columns represent individual samples (COPD or control), and rows correspond to genes ranked by significance. Hierarchical clustering using Euclidean distance and complete linkage revealed distinct sample groupings. Red indicates higher, and blue indicates lower relative expression. COPD samples clustered separately from controls, demonstrating robust transcriptional divergence. Upregulated genes (e.g., *NUDT3*, *BCL6-AS1*) were linked to oxidative stress and chromatin remodeling, while downregulated genes (*ANXA2P2*, *RPS7P1*) were associated with metabolic and structural pathways.

Differential Expression of ANXA2P2 and BCL6-AS

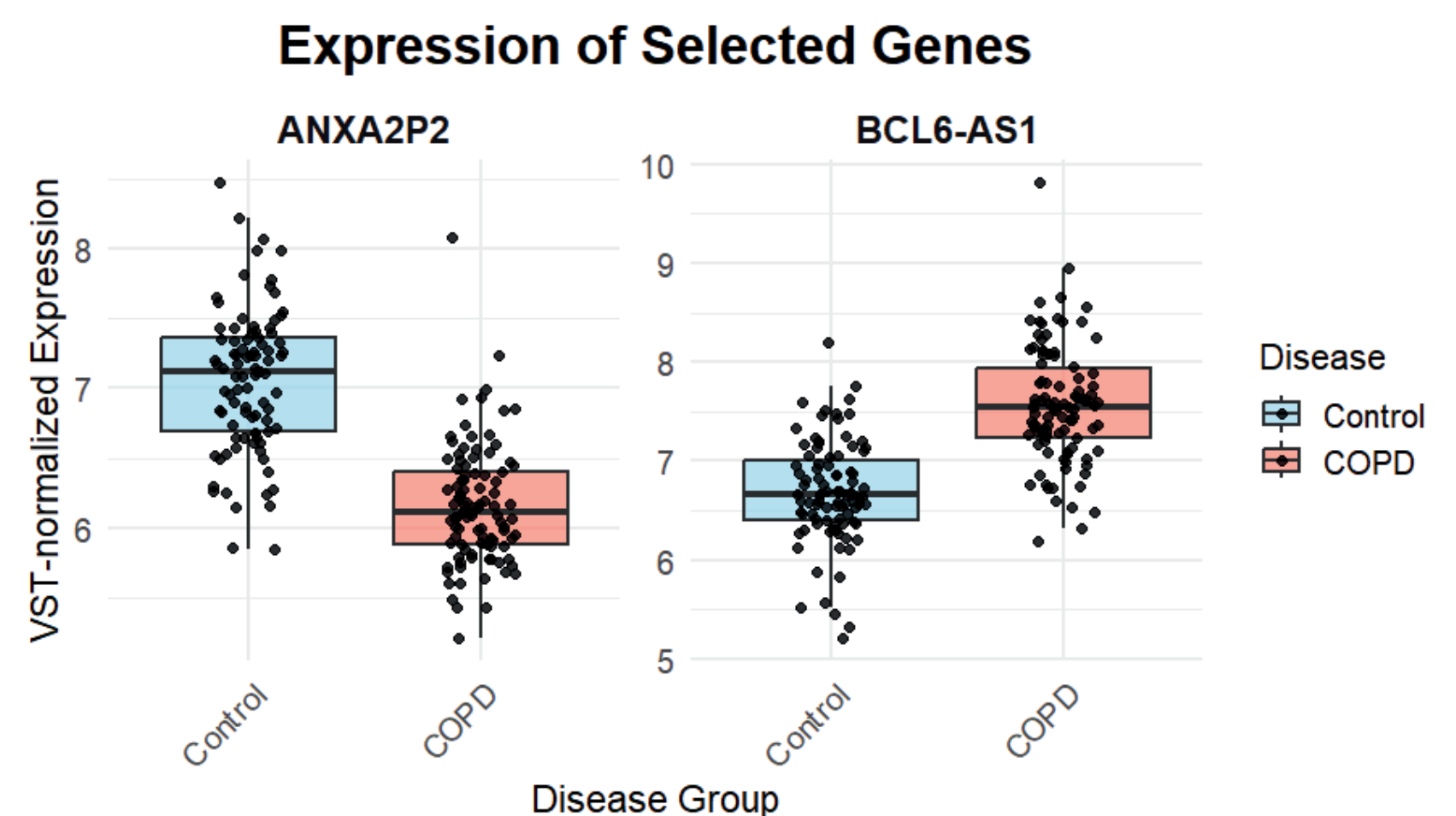


Figure 5: Boxplots show VST-normalized RNA-seq expression levels of ANXA2P2 and BCL6-AS1 in lung tissues from subjects with COPD (red) and healthy controls (blue). Expression values were obtained from DESeq2 analysis after variance-stabilizing transformation and normalization of raw counts. Each point represents an individual sample, and boxplots indicate the median and interquartile range. ANXA2P2 expression was significantly lower, whereas BCL6-AS1 expression was significantly higher in COPD compared with controls (adjusted $p < 0.05$, DESeq2 Wald test). These patterns highlight opposing regulation of genes linked to metabolic and chromatin remodeling pathways in COPD.

Conclusion

Analyses reveal clear transcriptional differences between COPD and control lung tissues, reflecting distinct molecular mechanisms. ANXA2P2 is downregulated, while NUDT3 and BCL6-AS1 are upregulated in COPD, indicating disrupted metabolic, oxidative stress, and chromatin remodeling pathways. Heatmaps and boxplots confirm these inverse expression patterns, with COPD samples clustering separately from controls.

Functionally, BCL6 regulates inflammation and neutrophil survival, NUDT3 protects against oxidative stress and supports DNA repair, and ANXA2P2/AnxA1 modulates inflammation and fibrosis. Dysregulation of these genes contributes to chronic inflammation, tissue remodeling, and disease progression in COPD, and also links to lung cancer development. These findings highlight key molecular drivers of COPD and potential therapeutic targets.

References