Mechanistic insights into Down syndrome comorbidities via convergent RNA-seq and TWAS signals

This manuscript was automatically generated on June 5, 2025.

Authors

- **Marc Subirana-Granés** ✉
  0000-0003-3934-839X · msubirana · msubirana20
  Department of Biomedical Informatics, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

- **Haoyu Zhang**
  0009-0005-6025-0217 · haoyu-zc
  Department of Biomedical Informatics, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

- **Prashant Gupta**
  0000-0003-3660-4415 · pragup
  Department of Biomedical Informatics, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

- **Milton Pividori** ✉
  0000-0002-3035-4403 · miltondp · miltondp · @miltondp@genomic.social
  Department of Biomedical Informatics, University of Colorado School of Medicine, Aurora, CO 80045, USA; Colorado Center for Personalized Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

✉ — Correspondence possible via GitHub Issues or email to Marc Subirana-Granés <marc.subiranagranes@cuanschutz.edu>, Milton Pividori <milton.pividori@cuanschutz.edu>.

Abstract

Down syndrome (DS) is caused by trisomy of chromosome 21 and is associated with diverse clinical manifestations, yet the molecular pathways linking chromosome-21 dosage effects to DS comorbidities remain poorly defined. Here we address this gap by applying a network-based, integrative framework that combines whole-blood transcriptomic data with gene-trait associations to uncover mechanistic insights into DS-associated conditions. First, we performed matrix factorization using PLIER on Human Trisome Project (HTP) RNA-Seq profiles from 304 trisomy-21 (T21) and 95 euploid (D21) individuals, deriving 156 biologically interpretable gene modules. We then identified 92 modules whose activity differed significantly between T21 and D21 and annotated these with prior-knowledge and KEGG pathways. To connect modules to clinical traits, we integrated PrediXcan-derived TWAS results from the UK Biobank, revealing 25 T21-specific modules with significant gene-trait associations (FDR < 0.1), including modules linked to cardiovascular, hematological, immune, metabolic, and neurological phenotypes relevant to DS. Using HTP clinical records as a replication cohort, 13 of these modules reliably predicted comorbidity status (AUC > 0.65, mAPS > 0.65). Most notably module 37, an interferon-stimulated gene network, whose elevated expression robustly distinguished DS individuals with pulmonary hypertension (AUC = 0.76, mAPS = 0.73). Overall, our study demonstrates that integrating blood-derived gene modules with population-scale genetic data uncovers coherent molecular signatures underlying DS comorbidities, identifies candidate biomarkers and therapeutic targets (e.g., *ISG15*, *IFITs*, *MX1*), and highlights the power of combining transcriptomic and genetic evidence to elucidate complex disease mechanisms.

Introduction

Down syndrome (DS) is a complex genetic disorder resulting from trisomy of human chromosome 21 (T21), characterized by a broad spectrum of clinical manifestations such as intellectual disability, congenital heart defects, pulmonary complications, metabolic abnormalities, and early-onset Alzheimer's disease [1]. Genes on chromosome 21 typically exhibit an average 1.5-fold increase in expression [2], and thousands of genes elsewhere in the genome are also differentially expressed [3]; yet there is considerable inter-individual variability in these transcriptional perturbations [4]. Despite these widespread cis- and trans-acting dosage effects, the molecular mechanisms by which the extra chromosome orchestrates these effects remain largely undefined.

Significant advances in human genetics over the past decades have markedly improved our understanding of complex traits and diseases. Genome-wide association studies (GWAS) have identified thousands of genetic variants associated with diverse human phenotypes, underscoring the polygenic nature of most complex traits [5]. However, translating these genetic associations into biological insights has proven challenging due to the intricate interplay among multiple genes, and the fact that most GWAS variants reside within non-coding regions of the genome, complicating functional interpretations [6].

To bridge the gap between genetic associations and biological mechanisms, transcriptome-wide association studies (TWAS) integrate expression quantitative trait loci (eQTL) data to nominate candidate genes whose expression mediates trait variation [7]. In addition to revealing gene-level mechanisms, TWAS has demonstrated that many regulatory variants exhibit concordant effects across tissues, reflecting core pathways that govern transcription and splicing in diverse cell types [8,9]. Because eQTLs with broad, cross-tissue activity are more likely to impact systemic phenotypes, expression profiles from accessible tissues (for example, blood or skin) can serve as proxies for regulatory processes in less accessible organs. However, by design, TWAS typically evaluates one gene at a time, potentially overlooking critical interactions among genes within complex regulatory networks.

Recent conceptual models of complex trait genetics, particularly the omnigenic model, address these limitations by conceptualizing the genetic architecture of complex traits through highly interconnected gene

regulatory networks [10]. According to this model, peripheral genes exert indirect effects by modulating core genes, which directly influence biological mechanisms. In response, machine learning-based approaches, particularly matrix factorization methods like the Pathway-Level Information ExtractoR (PLIER), have emerged as powerful tools to elucidate these regulatory networks. PLIER effectively addresses this complexity by identifying gene modules or groups of genes exhibiting coordinated expression patterns linked to defined molecular functions or pathways [11]. By decomposing high-dimensional transcriptomic data into biologically interpretable latent variables aligned with known pathways, PLIER facilitates an integrated understanding of gene regulatory mechanisms at the module level.

Building upon these methodological advancements, recent studies underscore the benefits of integrating genetic associations from GWAS and TWAS into these gene modules. Frameworks like PhenoPLIER and multimodal RNA sequencing analyses exemplified by Pantry demonstrate that combining TWAS data with module-level insights markedly enhances the interpretability of complex genetic data. This integrative strategy not only elucidates underlying disease mechanisms more effectively but also reveals novel regulatory pathways and candidate genes often overlooked by conventional analyses [12,13,14].

We framed this study within the omnigenic model, hypothesizing that the molecular mechanisms underlying DS-associated comorbidities are parallel to those observed in the general population but become altered due to regulatory cascades initiated by the extra copy of chromosome 21. This modification subsequently gives rise to the extensive clinical manifestations characteristic of DS. To test this hypothesis, we reconstructed co-expression networks from transcriptomic profiles of individuals with T21 and D21, aligning them with molecular mechanisms derived from the general population. Each resulting gene module was then systematically integrated with clinical phenotype associations obtained from large-scale population studies. By employing this integrative, network-based approach, our study aims to uncover novel insights into the precise molecular mechanisms driving phenotypic differences between individuals with T21 and euploid controls (D21).

Results

Overview of approach

We employed an integrative computational framework combining gene-expression and genetic-association data to elucidate molecular mechanisms underlying DS comorbidities. We first applied the PLIER matrix-decomposition method to Human Trisome Project (HTP) whole-blood RNA-Seq normalized dataset [15,16] of 304 trisomy 21 (T21) and 95 disomy 21 (D21) individuals, generating 156 biologically interpretable patterns of correlated genes termed gene modules or latent variables (LVs). Briefly, PLIER decomposed the expression matrix X into a matrix Z (gene×LV), which has gene weights for each gene module or LV, and matrix B, which quantifies the contribution of each sample to each gene module or LV. Simultaneously, the learned gene modules are aligned to a prior-knowledge matrix C of curated gene–pathway annotations, improving interpretability while reducing technical noise [11] (**Figure 1*A***)

We identified those gene modules that were differentially expressed between T21 and D21 and focused our analyses on them. For this, we then compared gene module scores derived from PLIER between T21 and D21 samples using a Wilcoxon test. Modules exhibiting statistically significant differences after multiple-testing correction were designated as T21-Specific Gene Modules (T21-SGMs) (**Figure 1*B***). To validate the molecular mechanisms inferred by PLIER, we (i) performed pathway overrepresentation analysis using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (**Figure 1*C***) and (ii) assessed tissue-specific expression patterns via the FUMA platform [17] (**Figure 1*D***).

Next, we integrate genetic data into transcriptional gene modules by connecting these gene modules with clinically relevant traits. For this, we integrated PrediXcan-derived TWAS gene-trait associations from the UK Biobank (discovery cohort). In our regression framework, each gene module is tested to determine

whether genes with high loadings are strongly enriched with genes associated with a given trait [13] (**Figure 1*E***).

Finally, we leveraged clinical variables curated from the HTP dataset's medical-record annotations (replication cohort) to assess each DS-specific module's ability to classify comorbidity status, quantifying performance via the area under the receiver-operator characteristic curve (AUC) and the mean average precision score (mAPS). This integrated framework enabled us to systematically pinpoint gene modules with both mechanistic insight and clinical applicability in Down syndrome pathogenesis (**Figure 1*F***).

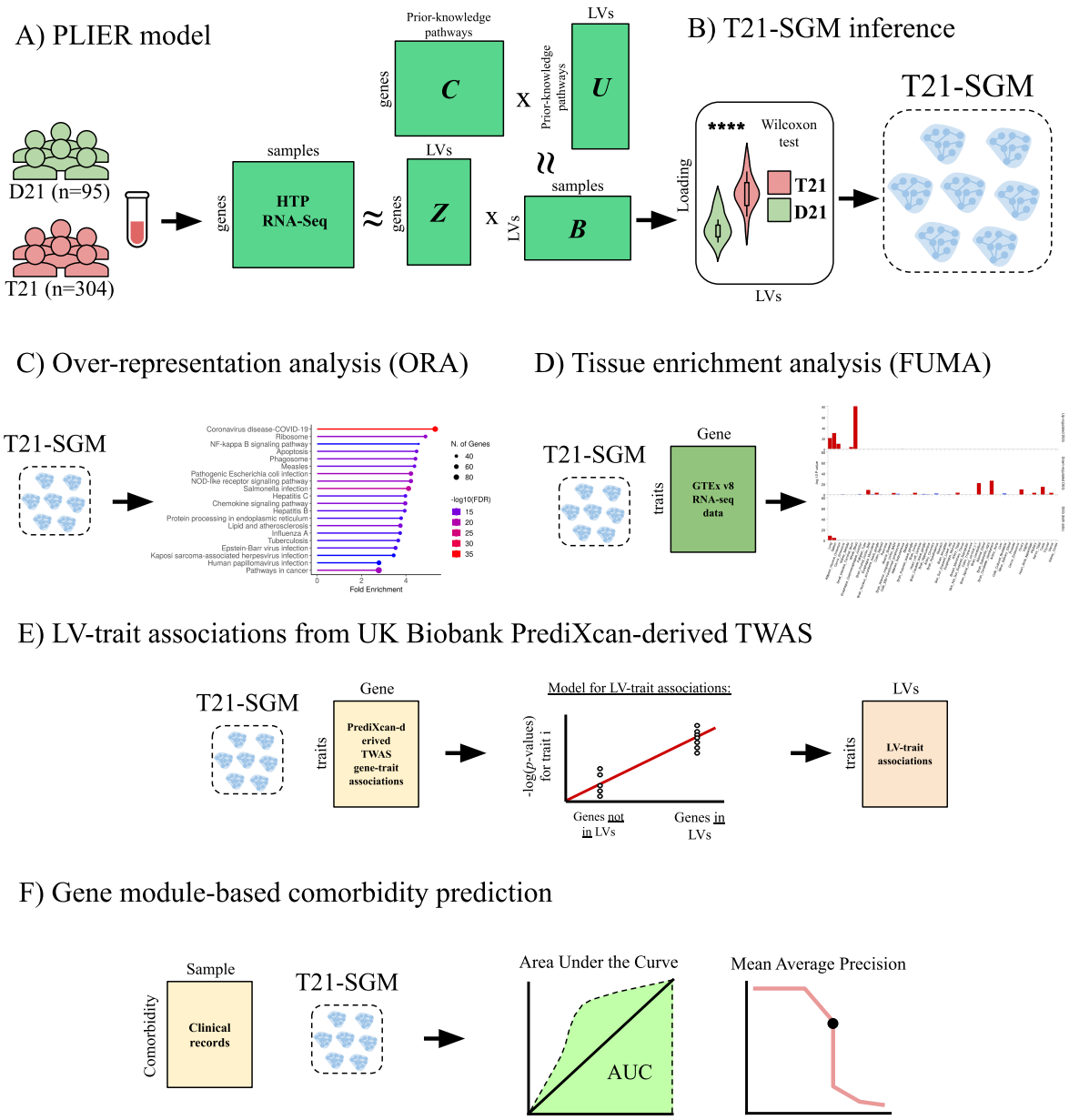

Figure 1: Overview of the methodology. (A) PLIER model. High-throughput HTP RNA-Seq data from euploid (D21, n = 95) and trisomy 21 (T21, n = 304) samples are decomposed by PLIER using prior-knowledge pathways (C) to generate gene modules/latent variables (LVs) (Z), pathway loadings (U), and sample-specific gene module loadings (B). **(B)** T21-Specific Gene Modules (T21-SGMs) inference. Wilcoxon rank-sum tests identify latent variables whose sample loadings differ significantly between T21 and D21, defining T21-SGMs. **(C)** Over-representation analysis (ORA). The top 5% of genes in each module are tested for enrichment against the KEGG pathway database. **(D)** Tissue enrichment analysis. The top 5% of genes in each module are submitted to FUMA for gene-to-tissue expression analysis using GTEx v8, identifying tissues in which module genes are significantly over- or under-expressed. **(E)** Gene module-trait association inference. For each T21-SGM, gene loadings are regressed against p-values from PrediXcan-derived TWAS in the UK Biobank (discovery cohort) to link modules with clinical phenotypes. **(F)** Module-based comorbidity prediction. Curated HTP comorbidity annotations (replication cohort) are used to evaluate each T21-SGM's predictive performance via area under the ROC curve (AUC) and mean average precision score (mAPS).

Trisomy 21 exhibits divergent gene module profiles

To assess whether chromosome-21 triplication globally alters these transcriptomic programs, we compared module loadings between T21 and D21 groups using Wilcoxon tests, followed by Benjamini–Hochberg correction for multiple testing. Of the 156 inferred gene modules, 92 (59%) showed significantly different transcriptomic gene module loadings between karyotype groups (FDR < 0.05) (**Figure 2**). We designated these modules as T21-specific gene modules (T21-SGM). Among T21-SGM, 37 exhibited increased module activity in T21 compared to D21, whereas 55 displayed decreased activity. Furthermore, 69 out of the 92 T21-SGM (75%) were significantly aligned (FDR < 0.05, AUC > 0.6) to canonical pathways derived from prior knowledge of the general population.

We next examined whether the observed differences in module activity could be attributed to chromosome-21 gene dosage effects. To test this, we compared chromosome-21 gene representation between T21-SGM and non-T21-SGM. Both sets contained a comparable proportion of chromosome-21 genes (Fisher's exact test, p-value = 0.30).

These results suggest an alteration of the transcriptional landscape in T21 rather than a shift to entirely new molecular mechanisms. These alterations occur predominantly via chromosome-21 gene overexpression rather than by overrepresentation of these genes within the affected modules.

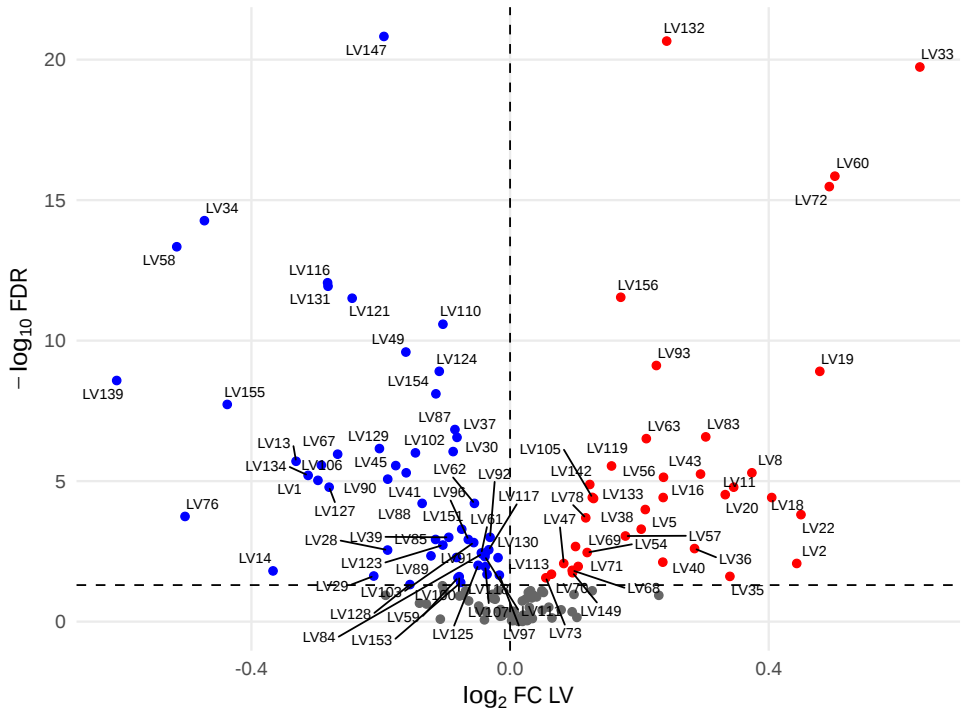

Figure 2: Differential activity of PLIER-derived gene modules between trisomy-21 and euploid samples. ach point represents one gene module, $\log_2$ fold-change in module activity (T21 vs. D21) on the x-axis and by $-\log_{10}$ FDR (Benjamini–Hochberg) on the y-axis. LVs significantly upregulated in T21 (FDR < 0.05, $\log_2$FC > 0) are shown in red; those significantly downregulated (FDR < 0.05, $\log_2$FC < 0) in blue; non-significant LVs (FDR ≥ 0.05) in gray. The vertical dashed line marks zero fold-change ($\log_2$FC = 0), and the horizontal dashed line indicates the FDR significance threshold (FDR = 0.05). A subset of the most highly perturbed T21-specific modules (e.g., LV33, LV60, LV72, LV132, LV147) is labeled.

T21-SGMs represent tissue-specific molecular mechanisms related to DS comorbidities

Then, we sought to determine whether the molecular mechanisms driving differential gene modules in T21 align with pathways implicated in well-characterized DS comorbidities.

To this end, we identified 28 T21-SGMs exhibiting the most differential expression between T21 and euploid controls (FDR < 0.05; $|\log_2 FC| > 0.25$). To provide a comprehensive overview of the molecular mechanisms related to these gene modules, each module was annotated through two different pathway databases: 1) PLIER's *internal* alignment to curated gene sets during model training (FDR < 0.05; AUC > 0.6) and 2) KEGG *external* database (never used for model training) by performing pathway over-representation analysis (ORA) on the top 5 percent of module genes (FDR < 0.05).

The resulting annotations suggest a link between T21-SGMs and pathways related to DS comorbidities. Six modules were enriched for signatures of neurodegeneration, including Alzheimer's and Parkinson's disease pathways and broader neurodegenerative gene sets. Two converged on programmes related to acute myeloid leukemia. Nine modules were enriched for pathways associated with congenital heart defects, chronic respiratory dysfunction, adipogenesis, atherosclerosis and non-alcoholic fatty liver disease. Eleven modules indicative of immune dysregulation were enriched for interferon-mediated signaling and dendritic cell activation, while other modules reflected hematological abnormalities through signatures of myeloid and lymphoid differentiation. Some of these gene modules exhibited tissue specificity consistent with their associated pathways; for example, module 14, enriched for Alzheimer's pathways, was overexpressed in brain tissue.

In summary, we observed that gene modules most strongly perturbed by trisomy 21 altered molecular mechanisms in the general population that can recapitulate tissue-specific transcriptional programs related to DS comorbidities.

Integration of genetic associations from the general population provides mechanistic insights into DS comorbidities

Motivated by the observation that T21-SGMs recapitulate molecular pathways underlying well-characterized DS comorbidities, we tested whether these transcriptomic signatures also capture phenotypic associations of DS-relevant conditions. However, a major limitation of our approach arises from using peripheral whole-blood samples to infer molecular mechanisms operative in tissues directly affected by DS comorbidities. Although whole blood is readily accessible, it may not fully represent tissue-specific gene regulation dynamics. Previous studies using TWAS integrating PrediXcan and GTEx data have demonstrated that many regulatory variants exhibit conserved effects across tissues, thereby enabling the inference of biological mechanisms in inaccessible tissues using gene expression profiles from accessible tissues like blood [8,9]. Thus, integrating TWAS-derived gene-trait associations partially mitigates this limitation by enabling the identification of tissue-specific regulatory effects relevant to DS comorbidities.

To this end, we integrated PrediXcan-derived transcriptome-wide association study (TWAS) results from the UK Biobank (using diseases with ≥1,000 cases and quantitative traits with ≥5,000 samples) with our 28 T21-SGMs (discovery cohort). For each trait, we regressed its TWAS p-values against a module's gene loadings, testing whether genes that strongly belong to the gene module are also more strongly associated with a given trait than the rest of the genes [13].

This analysis identified 25 gene modules exhibiting significant trait associations, yielding 109 module–trait links at an FDR < 0.1, of which 63 correspond to clinical phenotypes with established relevance to DS [1]. These associations map onto eleven core comorbidity categories: two modules associated with psychiatric traits (general happiness; methods of self-harm; self-harm in the past year), three modules associated with cardiovascular traits (atherosclerotic heart disease; doxazosin, an α-blocker used to treat hypertension; Adipine MR 10 m/r tablet, a calcium-channel blocker for hypertension), eleven modules associated with hematological disorders (lymphocyte count; eosinophil count; monocyte count; neutrophil count; basophil count; erythrocyte count; leukocyte count; reticulocyte count; platelet count), two modules associated with immune dysfunction (sulfasalazine used in autoimmune and inflammatory conditions; Mucodyne, a mucolytic agent often used during infection or inflammation), one module associated with obesity (weight),

three modules associated with bowel dysfunction and gastrointestinal structural defects (ulcerative colitis; peritonitis; Colofac-100 tablet, prescribed for irritable bowel syndrome), one module associated with refractive errors (astigmatism), three modules associated with respiratory diseases (asthma; Symbicort 100/6 Turbohaler, a combination inhaler used in asthma management; prednisolone, a systemic corticosteroid frequently used in asthma and other respiratory conditions), two modules associated with type 1 diabetes (type 1 diabetes mellitus; diabetic ketoacidosis; insulin), six modules associated with autoimmune skin conditions (psoriasis; Dovobet ointment, a topical treatment for psoriasis; Dovonex cream, a vitamin D analog for psoriasis; eczema/dermatitis; hay fever; allergic rhinitis or eczema), and five modules associated with other autoimmune disorders (rheumatoid arthritis; polyarthropathies; methotrexate used in autoimmune diseases; folic acid co-administered with methotrexate; sulfasalazine used in autoimmune gastrointestinal and joint diseases).

Collectively, these results demonstrate that T21-specific gene modules not only capture transcriptomic molecular mechanisms related to T21 via pathway analyses but also recapitulate DS clinical phenotypes via genetic studies and trait associations, indicating that this methodology can infer mechanistic insights into DS co-occurring conditions.

Gene modules predict comorbidity status in Down syndrome

To evaluate whether T21-SGMs recapitulate discovery cohort clinical comorbidities and thereby infer mechanistic insights into DS co-occurring conditions, we used the HTP dataset's curated medical-record annotations as a replication cohort and quantified each module's ability to predict comorbidity status. For every T21-SGM, we used matrix B (samples x LVs) to assess whether a gene module was overrepresented with samples from a particular DS comorbidity. For this, we computed the area under the receiver operating characteristic curve (AUC) and the mean average precision score (mAPS). We identified 13 T21-SGMs capable of predicting comorbidity status with high performance (AUC > 0.65, mAPS > 0.65) (**Figure 3**). Several of these gene modules demonstrate significant trait associations identified in the discovery UK Biobank cohort through integrated genetic analyses, uncover underlying molecular mechanisms via pathway enrichment, and display tissue-specific enrichment concordant with predicted comorbidities based on discovery HTP metadata. This convergence of different sources of evidence prioritizes potential mechanistic links between the molecular features captured by the gene modules and the manifestation of co-occurring conditions in individuals with DS.

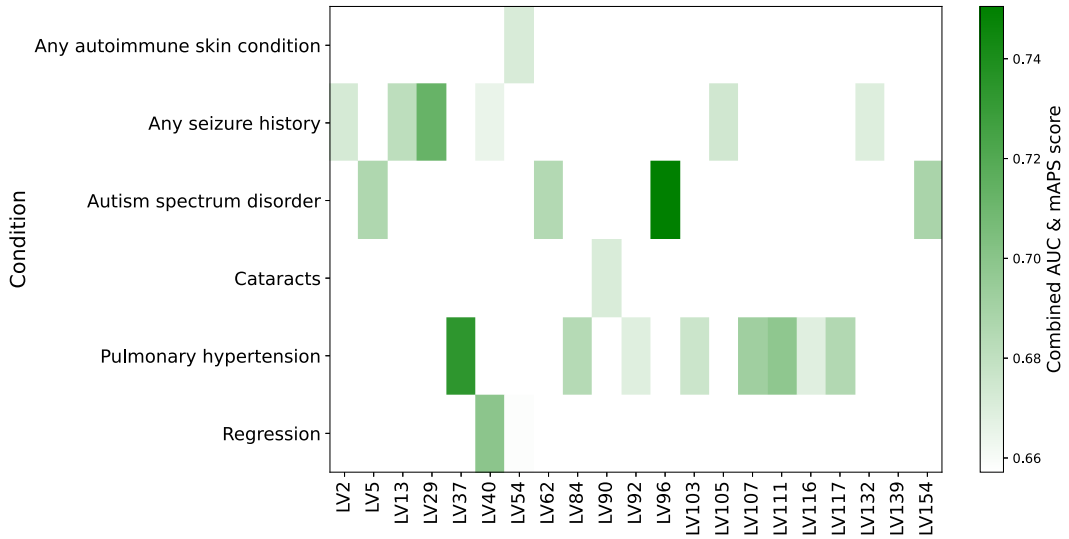

Figure 3: Predictive performance of T21-specific gene modules for Down syndrome comorbidities. The heatmap describes the combined predictive score (the average of AUC and mAPS) for each gene module that exceeded both AUC > 0.65 and mAPS > 0.65 when discriminating comorbidity status among T21 individuals. Columns represent individual gene modules and rows correspond to DS comorbidities from HTP replication clinical records. The color scale reflects the combined-score values, indicating joint predictive accuracy.

Gene module 37 captures an interferon-stimulated gene signature driving pulmonary hypertension in Down Syndrome

We systematically examined sources of evidence (enriched pathways, tissue specificity, module-trait associations, and clinical metadata) to establish associations between DS comorbidities and gene modules.

All sources of evidence consistently indicate that gene module 37 is robustly associated with pulmonary hypertension (PH) in individuals with DS. Gene module 37 expression levels are significantly higher in T21 subjects with PH compared to those without PH (Student's t-test, p-value < 0.01) (**Figure 4*A***). AUC and mAPS methodologies indicated that gene module 37 significantly differentiated between samples with and without pulmonary hypertension (AUC=0.76 and mAPS=0.73) (**Figure 4*B***), and samples within the top quartile (≥75th percentile) of expression for these gene modules showed a significant overrepresentation of pulmonary hypertension cases (p-value < 0.01).

Pathway enrichment analyses strongly support gene module 37's relevance to PH. PLIER identified significant enrichment (FDR < 0.05; AUC > 0.6) in pathways such as STAT3 and JNK signaling, both implicated in pulmonary arterial hypertension (PAH) pathogenesis [18,19]. Similarly, gene expression patterns in heart failure associated with PH were enriched [20]. Independently, overrepresentation analysis (ORA) of gene module 37's top genes revealed significant enrichment (FDR < 0.05) in NF-κB and NOD-like receptor signaling pathways, both directly linked to PAH [21,22,23]. Additional enrichment for COVID-19 and lipid/atherosclerosis pathways highlights potential inflammatory and metabolic contributions to PH [24] (**Figure 4*C***).

Tissue-specific expression analysis confirmed that lung, spleen, and whole blood are the primary sites of module activity (FDR < 0.05; $|\log_2 \text{fold change}| \geq 0.58$), aligning with known tissues affected by PH (**Figure 4*D***).

Module-trait association analyses from the UK Biobank (discovery cohort) further corroborate gene module 37's role in PH by revealing significant links (FDR < 0.1) to clinical traits, including tiotropium treatment for chronic obstructive pulmonary disease (COPD), elevated blood glucose, recent inhaler use, and cigarette smoke exposure [25,26,27,28] (**Figure 4*E***). These traits highlight clinically relevant factors and environmental risk contributors to PH.

At the individual gene level, gene module 37 prominently features interferon-stimulated genes (ISGs), including *ISG15*, *IFIT1-3*, *RSAD2*, and *IRF7* (**Figure 4*F***). Many of these genes have established roles in PH pathology. *ISG15* is identified as both a biomarker and mediator of pulmonary vascular remodeling [29,30]. *IFIT* family genes form part of a characteristic transcriptomic signature in PAH patients [31,32], while *IRF7* is upregulated in hypoxia-induced rodent models of PH [33]. *RSAD2* expression is elevated in peripheral neutrophils of idiopathic PAH patients, indicative of enhanced interferon-mediated inflammation [30].

Gene module 37 notably includes *MX1*, ranking as the 14th top gene. *MX1* is located on chromosome 21 and encodes a dynamin-like GTPase with antiviral functions, although its role in DS pathogenesis remains unclear. Recent studies implicate *MX1* in AP-1 transcriptional regulation in DS [34]. AP-1 itself plays a crucial role in pulmonary hypertension, driving aberrant pulmonary artery smooth muscle cell proliferation via endothelin-1 signaling in idiopathic pulmonary arterial hypertension (IPAH) patients [35]. These findings suggest that *MX1* may interact with AP–1–mediated pathways in pulmonary vasculature, highlighting *MX1* as a potential therapeutic target for pulmonary hypertension in DS.

Collectively, these multiple lines of evidence converge to support gene module 37's critical role in PH within DS, identifying specific genes and pathways as promising biomarkers and therapeutic targets.

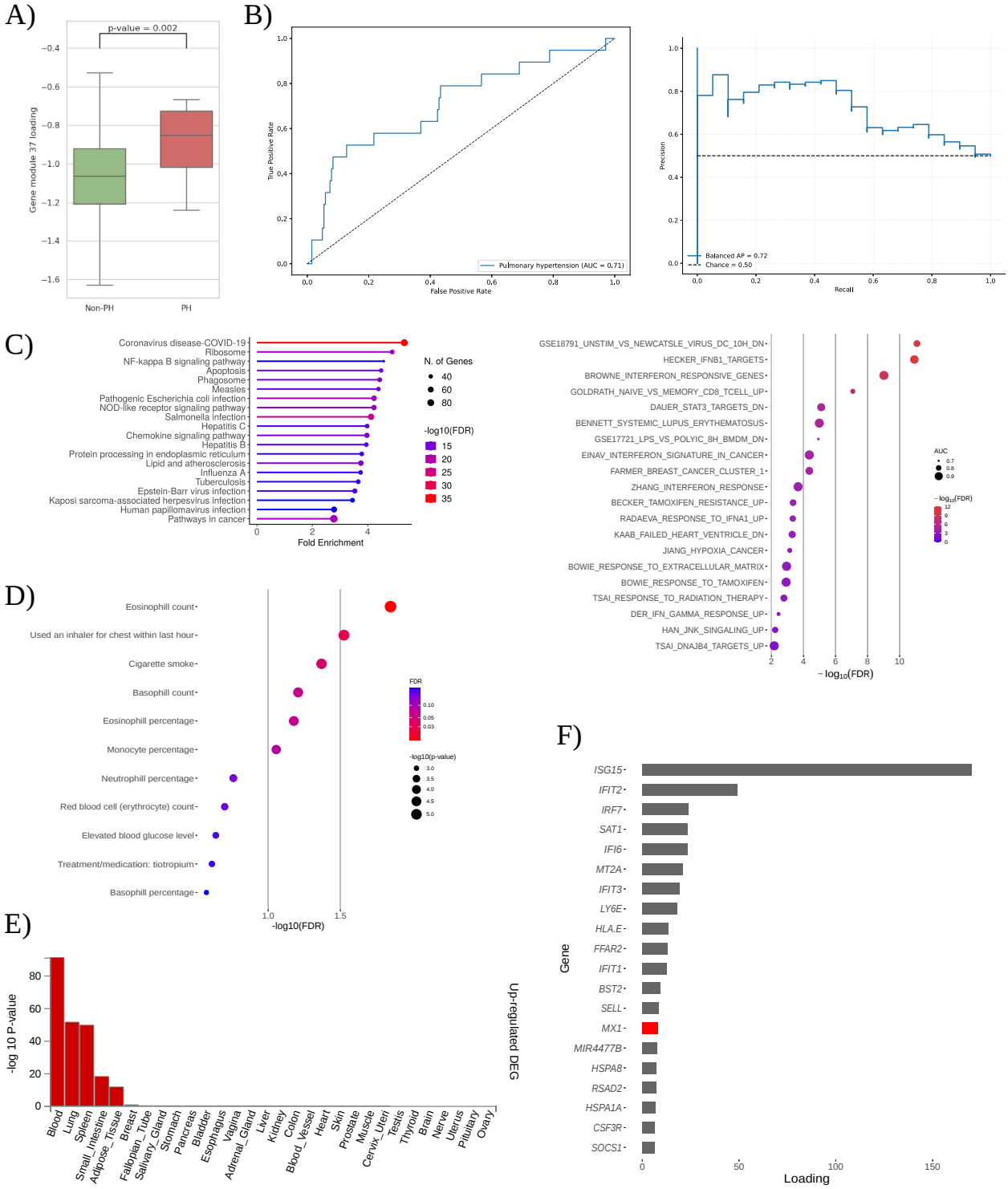

Figure 4: Comprehensive characterization of gene module 37 in Down syndrome–associated pulmonary hypertension. (A) Gene module 37 expression in T21 individuals without pulmonary hypertension (Non-PH) versus with pulmonary hypertension (PH); Student's t-test p-value = 0.002 is indicated. **(B)** Classification performance of gene module 37 for PH: receiver-operating-characteristic curve (left) and precision-recall curve (right). **(C)** Molecular mechanisms inference for gene model 37. Upper: KEGG over-representation analysis of gene module 37's top 5 % genes, showing fold enrichment (x-axis), $-\log_{10}$ FDR (color), and gene set size (point size). Lower: PLIER prior-knowledge alignment of gene module 37, plotting curated gene sets by AUC (point size) and $-\log_{10}$ FDR (color). **(D)** Gene module-trait association from PrediXcan-derived TWAS associations for gene module 37. Each bubble represents a clinical trait, sized by number of cases and colored by $-\log_{10}$ FDR. **(E)** FUMA tissue specificity results for gene module 37 genes, showing $-\log_{10}$ p-values across GTEx tissues. **(F)** Gene-loading profile for gene module 37 of the top 20 genes. Those encoded on chromosome 21 are highlighted in red.

Discussion

Despite substantial progress in understanding the genetic basis of DS, elucidating how trisomy of chromosome 21 broadly disrupts molecular networks, leading to its diverse co-ocurrent conditions, remains

challenging. Our study addresses this crucial gap by integrating transcriptomic and genetic data to unravel the molecular underpinnings of DS comorbidities. Based on the omnigenic model, we proposed that chromosome 21 trisomy initiates regulatory cascades affecting gene networks analogous to those observed in the general population. Leveraging PLIER-derived gene modules, TWAS gene-trait associations, and clinical data from the HTP records, we systematically connected transcriptomic disruptions with DS-associated phenotypes. The identification of trisomy 21-specific gene modules, significantly differentially expressed compared to euploid controls, highlights potential causal pathways driving DS comorbidities. Notably, these modules predominantly align with canonical pathways rather than entirely novel mechanisms, implicated in DS comorbidities, including neurodegeneration, cardiovascular dysfunction, hematological abnormalities, and immune perturbations.

A significant strength of our analysis lies in overcoming the limitation of using whole blood RNA-seq data by integrating diverse datasets. Previous research supports that many cis-regulatory genetic variants have conserved effects across multiple tissues, reinforcing the existence of shared regulatory pathways [8,9,36]. Consistent with these findings, our integrative methodology demonstrates that transcriptomic blood-derived gene modules significantly reflect mechanisms in tissues directly implicated in DS comorbidities. T21-SGM significantly overlap GTEx signatures in lung, brain, and immune tissues implicated in DS comorbidities. TWAS-derived gene-trait associations link these modules to eleven core DS comorbidity categories (e.g., psychiatric traits, cardiovascular conditions, hematological measures, immune dysfunction, obesity, respiratory diseases, autoimmune disorders), capturing broad regulatory dynamics across different biological contexts.

Ultimately, our integrative framework bridges transcriptomic signatures with genetic associations, providing deeper insights into DS-related complex diseases. While recognizing that future studies using single-cell RNA-seq from affected tissues and rigorous experimental validations in cellular or animal models are essential, our findings substantially advance the understanding of DS comorbidities. The identification of clinically relevant gene modules, exemplified by module 37's robust association with PH, underscores their potential utility as biomarkers and therapeutic targets. Specifically, the enrichment of interferon-stimulated genes in this module identifies and prioritize promising therapeutic candidates, such as *ISG15*, *IFIT* family genes, and chromosome-21 gene *MX1*. Framed within the omnigenic model, our results suggest that chromosomal dosage imbalance initiates regulatory cascades, highlighting novel intervention points for treating DS comorbidities.

Methods

Cohort and data preprocessing

This study was conducted under a protocol approved by the Colorado Multiple Institutional Review Board (COMIRB #15–2170). Participants were enrolled through the Crnic Institute's Human Trisome Project (HTP; www.trisome.org). Demographic and clinical metadata were collected via structured questionnaires administered to participants and caregivers and supplemented by abstraction of the medical record.

Peripheral whole-blood samples were drawn into PAXgene RNA Tubes (Qiagen) from 304 individuals with trisomy 21 (T21) and 95 euploid controls (D21). Total RNA was extracted and assessed for integrity according to established HTP protocols [2,15]. Sequencing libraries were generated and sequenced on an Illumina NovaSeq 6000 instrument (Novogene) with paired-end reads. Raw sequence data were quality-filtered and adapter-trimmed, then aligned to the GRCh38 human reference genome using STAR [37]. Gene-level expression values were computed as transcripts per million (TPM).

PLIER HTP model generation

For the HTP model generation, we applied the PLIER algorithm (PLIER R package v0.1.6; https://github.com/wgmao/PLIER) to the z-score-normalized TPM HTP RNA-Seq expression matrix (36,362 genes x 399 samples) [11]. Briefly, the gene-by-sample TPM expression matrix ($g \times s$) was first z-score normalized on a per-gene basis using the rowNorm function. PLIER then factorizes this matrix into a gene loadings matrix Z ($g \times k$) and a latent activity matrix B ($k \times s$), where k denotes the number of latent variables. To guide this decomposition, we supplied prior knowledge in the form of a binary gene-by-geneset membership matrix C ($g \times p$), comprising MSigDB v4.0 C2 (curated gene sets), C6 (oncogenic signatures), C7 (immunologic signatures), bloodCellMarkersIRISDMAP, and svmMarkers. During optimization, PLIER enforces correspondence between Z and its pathway-based prediction ($C \cdot U$, with U a $p \times k$ coefficient matrix) by penalizing the distance between these matrices, and applies an elastic-net penalty on U to ensure that each latent factor represents a sparse subset of pathways. The joint objective is solved end-to-end via block coordinate minimization. We determined k by applying the num.pcs function to the normalized TPM matrix, yielding 156 significant components. Model fitting used the k=156 and frac=0.7 options. L_1 and L_2 regularization parameters were inferred automatically. Upon convergence, both the Z and B matrices were mean-centered and scaled to unit variance to facilitate downstream comparisons and integrative analyses.

Regression model for latent variable-trait associations

We inferred gene-trait associations from the PrediXcan methods with PLIER-derived latent variables using generalized least squares (GLS) regression, described in detail in [13]. Briefly, PrediXcan methods included S-PrediXcan (gene–tissue associations) and S-MultiXcan (cross-tissue meta-analysis). For each latent variable l, we constructed a binary indicator vector s_l that flags genes within the top 1% of absolute loadings within that LV. In the GLS framework, the vector of S-MultiXcan p-values for a given trait was regressed on this indicator, together with an intercept and gene-specific covariates, while modeling residuals with a covariance structure given by the gene–gene correlation matrix. This approach tests whether genes most heavily weighted in a latent variable show stronger trait associations than other genes.

Pathway over-Representation analysis (ORA)

For each latent variable, we extracted the top 5 % of genes ranked by absolute loading values and performed over-representation analysis against the Kyoto Encyclopedia of Genes and Genomes (KEGG) database using the clusterProfiler R package (v4.0) [38]. The universe for enrichment was defined as all genes quantified in the HTP RNA-seq dataset. Resulting p-values were adjusted for multiple testing via the Benjamini–Hochberg procedure, and pathways with an adjusted p-value < 0.05 were considered significantly enriched.

Tissue specificity annotation

Tissue-specific expression of LV-associated genes was annotated using the FUMA web platform (v1.8.0; accessed May 26, 2025) [17]. The goal of this analysis was to test whether LVs found in our whole blood RNA-seq dataset from HTP also resembled mechanisms preserved in other tissues. For each latent variable, its constituent genes were submitted to FUMA's Tissue Expression Analysis module against GTEx v8. Default parameters were applied to test for significant over-expression of each gene set across 54 GTEx tissues; tissues meeting FUMA's built-in significance threshold were reported as showing LV-specific expression (Bonferroni-corrected $P < 0.05$ and $|\log_2 \text{fold change}| \geq 0.58$).

Comorbidity classification performance

We assessed each trisomy-21-specific LV's ability to predict between T21 individuals with and without a given comorbidity using two complementary metrics. First, for each annotated comorbidity, we retrieved the subset of T21 participants classified as cases or controls and extracted their corresponding LV activity

scores. Discriminative performance was quantified by the area under the receiver operating characteristic curve (ROC AUC), computed with scikit-learn's (v1.6.1) [39] roc_auc_score function, and by the mean average precision score (mAPS). Second, the mAPS was calculated using the average_precision_score function and further balanced by repeatedly subsampling the negative class to match the number of cases (50 iterations) and averaging the resulting precision scores. Only LVs meeting all three criteria, ROC AUC > 0.65 and mAPS > 0.65 were included.

Data and code availability

All original code has been deposited on GitHub (https://github.com/pivlab/htp_plier). Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Acknowledgements

This work is supported by the National Human Genome Research Institute (R00 HG011898 to M.P.), and The Eunice Kennedy Shriver National Institute of Child Health and Human Development (R01 HD109765 to M.P.).

References

1. **Down syndrome**
   Stylianos E Antonarakis, Brian G Skotko, Michael S Rafii, Andre Strydom, Sarah E Pape, Diana W Bianchi, Stephanie L Sherman, Roger H Reeves
   Nature Reviews Disease Primers (2020-02-06) https://doi.org/ghn9hj
   DOI: 10.1038/s41572-019-0143-7 · PMID: 32029743 · PMCID: PMC8428796

2. **Triplication of the interferon receptor locus contributes to hallmarks of Down syndrome in a mouse model**
   Katherine A Waugh, Ross Minter, Jessica Baxter, Congwu Chi, Matthew D Galbraith, Kathryn D Tuttle, Neetha P Eduthan, Kohl T Kinning, Zdenek Andrysik, Paula Araya, … Joaquin M Espinosa
   Nature Genetics (2023-06) https://doi.org/g9ndkj
   DOI: 10.1038/s41588-023-01399-7 · PMID: 37277650 · PMCID: PMC10260402

3. **Domains of genome-wide gene expression dysregulation in Down's syndrome**
   Audrey Letourneau, Federico A Santoni, Ximena Bonilla, MReza Sailani, David Gonzalez, Jop Kind, Claire Chevalier, Robert Thurman, Richard S Sandstrom, Youssef Hibaoui, … Stylianos E Antonarakis
   Nature (2014-04-16) https://doi.org/sf9
   DOI: 10.1038/nature13200 · PMID: 24740065

4. **Variegated overexpression of chromosome 21 genes reveals molecular and immune subtypes of Down syndrome**
   Micah G Donovan, Neetha P Eduthan, Keith P Smith, Eleanor C Britton, Hannah R Lyford, Paula Araya, Ross E Granrath, Katherine A Waugh, Belinda Enriquez Estrada, Angela L Rachubinski, … Joaquin M Espinosa
   Nature Communications (2024-06-28) https://doi.org/g9ndh8
   DOI: 10.1038/s41467-024-49781-1 · PMID: 38942750 · PMCID: PMC11213896

5. **Genome-wide association studies**
   Emil Uffelmann, Qin Qin Huang, Nchangwi Syntia Munung, Jantina de Vries, Yukinori Okada, Alicia R Martin, Hilary C Martin, Tuuli Lappalainen, Danielle Posthuma
   Nature Reviews Methods Primers (2021-08-26) https://doi.org/gmk7p2
   DOI: 10.1038/s43586-021-00056-9

6. **Benefits and limitations of genome-wide association studies**
   Vivian Tam, Nikunj Patel, Michelle Turcotte, Yohan Bossé, Guillaume Paré, David Meyre
   Nature Reviews Genetics (2019-05-08) https://doi.org/ggcxxb
   DOI: 10.1038/s41576-019-0127-1 · PMID: 31068683

7. **Transcriptome-wide association studies: recent advances in methods, applications and available databases**
   Jialin Mai, Mingming Lu, Qianwen Gao, Jingyao Zeng, Jingfa Xiao
   Communications Biology (2023-09-01) https://doi.org/gs9f8r
   DOI: 10.1038/s42003-023-05279-y · PMID: 37658226 · PMCID: PMC10474133

8. **Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics**
   Alvaro N Barbeira, Scott P Dickinson, Rodrigo Bonazzola, Jiamao Zheng, Heather E Wheeler, Jason M Torres, Eric S Torstenson, Kaanan P Shah, Tzintzuni Garcia, Todd L Edwards, … Hae Kyung Im
   Nature Communications (2018-05-08) https://doi.org/gdjvp5
   DOI: 10.1038/s41467-018-03621-1 · PMID: 29739930 · PMCID: PMC5940825

9. **The GTEx Consortium atlas of genetic regulatory effects across human tissues**
, François Aguet, Shankara Anand, Kristin G Ardlie, Stacey Gabriel, Gad A Getz, Aaron Graubert, Kane Hadley, Robert E Handsaker, Katherine H Huang, … Simona Volpi
Science (2020-09-11) https://doi.org/ghbnhr
DOI: 10.1126/science.aaz1776 · PMID: 32913098 · PMCID: PMC7737656

10. **An Expanded View of Complex Traits: From Polygenic to Omnigenic**
Evan A Boyle, Yang I Li, Jonathan K Pritchard
Cell (2017-06) https://doi.org/gcpgdz
DOI: 10.1016/j.cell.2017.05.038 · PMID: 28622505 · PMCID: PMC5536862

11. **Pathway-level information extractor (PLIER) for gene expression data**
Weiguang Mao, Elena Zaslavsky, Boris M Hartmann, Stuart C Sealfon, Maria Chikina
Nature Methods (2019-06-27) https://doi.org/gf75g6
DOI: 10.1038/s41592-019-0456-1 · PMID: 31249421 · PMCID: PMC7262669

12. **Genetic Studies Through the Lens of Gene Networks**
Marc Subirana-Granés, Jill Hoffman, Haoyu Zhang, Christina Akirtava, Sutanu Nandi, Kevin Fotso, Milton Pividori
Annual Review of Biomedical Data Science (2025-02-20) https://doi.org/g9ksrb
DOI: 10.1146/annurev-biodatasci-103123-095355 · PMID: 39977605

13. **Projecting genetic associations through gene expression patterns highlights disease etiology and drug mechanisms**
Milton Pividori, Sumei Lu, Binglan Li, Chun Su, Matthew E Johnson, Wei-Qi Wei, Qiping Feng, Bahram Namjou, Krzysztof Kiryluk, Iftikhar J Kullo, … Casey S Greene
Nature Communications (2023-09-09) https://doi.org/gspsxr
DOI: 10.1038/s41467-023-41057-4 · PMID: 37689782 · PMCID: PMC10492839

14. **Multimodal analysis of RNA sequencing data powers discovery of complex trait genetics**
Daniel Munro, Nava Ehsan, Seyed Mehdi Esmaeili-Fard, Alexander Gusev, Abraham A Palmer, Pejman Mohammadi
Nature Communications (2024-11-29) https://doi.org/g9ndjn
DOI: 10.1038/s41467-024-54840-8 · PMID: 39613793 · PMCID: PMC11607376

15. **Multidimensional definition of the interferonopathy of Down syndrome and its response to JAK inhibition**
Matthew D Galbraith, Angela L Rachubinski, Keith P Smith, Paula Araya, Katherine A Waugh, Belinda Enriquez-Estrada, Kayleigh Worek, Ross E Granrath, Kohl T Kinning, Neetha Paul Eduthan, … Joaquin M Espinosa
Science Advances (2023-06-30) https://doi.org/gtt8b2
DOI: 10.1126/sciadv.adg6218 · PMID: 37379383 · PMCID: PMC10306300

16. **Crnic Institute Human Trisome Project™**
Crnic Institute Human Trisome Project™
https://www.trisome.org

17. **Functional mapping and annotation of genetic associations with FUMA**
Kyoko Watanabe, Erdogan Taskesen, Arjen van Bochoven, Danielle Posthuma
Nature Communications (2017-11-28) https://doi.org/gcnxmr
DOI: 10.1038/s41467-017-01261-5 · PMID: 29184056 · PMCID: PMC5705698

18. **STAT3 signaling in pulmonary arterial hypertension**
Roxane Paulin, Jolyane Meloche, Sébastien Bonnet
JAK-STAT (2012-10) https://doi.org/gsbf7k

DOI: 10.4161/jkst.22366 · PMID: 24058777 · PMCID: PMC3670278

19. **JNK2 regulates vascular remodeling in pulmonary hypertension**
Mita Das, WMichael Zawada, James West, Kurt R Stenmark
Pulmonary Circulation (2018-07) https://doi.org/gdszts
DOI: 10.1177/2045894018778156 · PMID: 29718758 · PMCID: PMC6055330

20. **Molecular Signature of a Right Heart Failure Program in Chronic Severe Pulmonary Hypertension**
Jennifer I Drake, Herman J Bogaard, Shiro Mizuno, Berrick Clifton, Bin Xie, Yuan Gao, Catherine I Dumur, Paul Fawcett, Norbert F Voelkel, Ramesh Natarajan
American Journal of Respiratory Cell and Molecular Biology (2011-12) https://doi.org/fwn7xn
DOI: 10.1165/rcmb.2010-0412oc · PMID: 21719795 · PMCID: PMC3361357

21. **Nuclear Factor κB Inhibition Reduces Lung Vascular Lumen Obliteration in Severe Pulmonary Hypertension in Rats**
Daniela Farkas, Aysar A Alhussaini, Donatas Kraskauskas, Vita Kraskauskiene, Carlyne D Cool, Mark R Nicolls, Ramesh Natarajan, Laszlo Farkas
American Journal of Respiratory Cell and Molecular Biology (2014-09) https://doi.org/f6vpdr
DOI: 10.1165/rcmb.2013-0355oc · PMID: 24684441 · PMCID: PMC4189489

22. **Meta-analysis of blood genome-wide expression profiling studies in pulmonary arterial hypertension**
Jason M Elinoff, Adrien J Mazer, Rongman Cai, Mengyun Lu, Grace Graninger, Bonnie Harper, Gabriela A Ferreyra, Junfeng Sun, Michael A Solomon, Robert L Danner
American Journal of Physiology-Lung Cellular and Molecular Physiology (2020-01-01) https://doi.org/g9ndx3
DOI: 10.1152/ajplung.00252.2019 · PMID: 31617731 · PMCID: PMC6985867

23. **Transcriptomic analysis identifies Toll-like and Nod-like pathways and necroptosis in pulmonary arterial hypertension**
Genfa Xiao, Wei Zhuang, Tingjun Wang, Guili Lian, Li Luo, Chaoyi Ye, Huajun Wang, Liangdi Xie
Journal of Cellular and Molecular Medicine (2020-08-29) https://doi.org/g9ndwv
DOI: 10.1111/jcmm.15745 · PMID: 32860486 · PMCID: PMC7576255

24. **Identification of metabolic biomarkers in idiopathic pulmonary arterial hypertension using targeted metabolomics and bioinformatics analysis**
Chuang Yang, Yi-Hang Liu, Hai-Kuo Zheng
Scientific Reports (2024-10-25) https://doi.org/g9ndn9
DOI: 10.1038/s41598-024-76514-7 · PMID: 39455660 · PMCID: PMC11511845

25. **Pulmonary hypertension in COPD**
A Chaouat, R Naeije, E Weitzenblum
European Respiratory Journal (2008-10-31) https://doi.org/dbv8bw
DOI: 10.1183/09031936.00015608 · PMID: 18978137

26. **Pulmonary Hypertension and Hyperglycemia—Not a Sweet Combination**
Or Bruck, L Pandit
Diagnostics (2024-05-28) https://doi.org/g9nd3q
DOI: 10.3390/diagnostics14111119 · PMID: 38893645 · PMCID: PMC11171670

27. **Pulmonary hypertension inhaled therapies: An updated review**
Karim El-Kersh, Bilal A Jalil
The American Journal of the Medical Sciences (2023-07) https://doi.org/g9ndgn
DOI: 10.1016/j.amjms.2023.03.002 · PMID: 36921672

28. **Tobacco Smoke: A Risk Factor for Pulmonary Arterial Hypertension?**
Regula Schiess, Oliver Senn, Manuel Fischler, Lars C Huber, Serap Vatandaslar, Rudolf Speich, Silvia Ulrich
Chest (2010-11) https://doi.org/fvxpc4
DOI: 10.1378/chest.09-2962 · PMID: 20472864

29. **Identification of the shared hub gene signatures and molecular mechanisms between HIV-1 and pulmonary arterial hypertension**
Huanzhuo Mai, Xing Yang, Yulan Xie, Jie Zhou, Yiru Wei, Tingyan Luo, Jing Yang, Ping Cui, Li Ye, Hao Liang, Jiegang Huang
Scientific Reports (2024-03-25) https://doi.org/g9ndn2
DOI: 10.1038/s41598-024-55645-x · PMID: 38528047 · PMCID: PMC10963360

30. **Single-Cell Transcriptome Analysis of Peripheral Neutrophils From Patients With Idiopathic Pulmonary Arterial Hypertension**
Rui Zhang, Ji Zhang, Yun-Long Zhang, Su-Gang Gong, Qin-Hua Zhao, Xiao-Juan Wang, Jia-Yu Zhao, Rong Jiang, Hong-Ling Qiu, Hui-Ting Li, … Lan Wang
Hypertension (2023-08) https://doi.org/g9ndzk
DOI: 10.1161/hypertensionaha.123.21142 · PMID: 37313754

31. **Identification and Validation of Key Genes Associated With Systemic Sclerosis-Related Pulmonary Hypertension**
Ji-Na Zheng, Yang Li, Yue-Mei Yan, Hui Shi, Tian-Tian Zou, Wen-Qi Shao, Qiang Wang
Frontiers in Genetics (2020-07-24) https://doi.org/gqg53c
DOI: 10.3389/fgene.2020.00816 · PMID: 32793290 · PMCID: PMC7393672

32. **Role of Rendezvous-Procedure in the Treatment of Complications after Laparoscopic Sleeve Gastrectomy**
Dörte Wichmann, Veit Scheble, Stefano Fusco, Ulrich Schweizer, Felix Hönes, Wilfried Klingert, Alfred Königsrainer, Rami Archid
Journal of Clinical Medicine (2021-11-30) https://doi.org/g9nd3z
DOI: 10.3390/jcm10235670 · PMID: 34884372 · PMCID: PMC8658356

33. **STING contributes to pulmonary hypertension by targeting the interferon and BMPR2 signaling through targeting F2RL3**
Lin Deng, Chengrui Cao, Zongye Cai, Ziping Wang, Bin Leng, Zhen Chen, Fanhao Kong, Zhiyue Zhou, Jun He, Xiaowei Nie, Jin-Song Bian
Cold Spring Harbor Laboratory (2024-02-21) https://doi.org/g9ndt5
DOI: 10.1101/2024.02.20.578386

34. **Integration of ATAC-seq and RNA-seq identifies MX1-mediated AP-1 transcriptional regulation as a therapeutic target for Down syndrome**
Zhenglong Guo, Yongchang Zhu, Hai Xiao, Ranran Dai, Wenke Yang, Wei Xue, Xueying Zhang, Bingtao Hao, Shixiu Liao
Biological Research (2023-12-09) https://doi.org/g9nd2h
DOI: 10.1186/s40659-023-00474-x · PMID: 38066591 · PMCID: PMC10709892

35. **Transcription factors in the pathogenesis of pulmonary arterial hypertension—Current knowledge and therapeutic potential**
Jakob Körbelin, Julius Klein, Christiane Matuszcak, Johannes Runge, Lars Harbaum, Hans Klose, Jan K Hennigs
Frontiers in Cardiovascular Medicine (2023-01-06) https://doi.org/g9nd29
DOI: 10.3389/fcvm.2022.1036096 · PMID: 36684555 · PMCID: PMC9853303

36. **Prediction of the gene expression in normal lung tissue by the gene expression in blood**

Justin W Halloran, Dakai Zhu, David C Qian, Jinyoung Byun, Olga Y Gorlova, Christopher I Amos, Ivan P Gorlov
BMC Medical Genomics (2015-11-17) https://doi.org/f77xgx
DOI: 10.1186/s12920-015-0152-7 · PMID: 26576671 · PMCID: PMC4650316

37. **STAR: ultrafast universal RNA-seq aligner**
Alexander Dobin, Carrie A Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, Thomas R Gingeras
Bioinformatics (2012-10-25) https://doi.org/f4h523
DOI: 10.1093/bioinformatics/bts635 · PMID: 23104886 · PMCID: PMC3530905

38. **clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters**
Guangchuang Yu, Li-Gen Wang, Yanyan Han, Qing-Yu He
OMICS: A Journal of Integrative Biology (2012-05) https://doi.org/gdf33f
DOI: 10.1089/omi.2011.0118 · PMID: 22455463 · PMCID: PMC3339379

39. **Scikit-learn: Machine Learning in Python**
Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Andreas Müller, Joel Nothman, Gilles Louppe, … Édouard Duchesnay
arXiv (2012) https://doi.org/gpptwn
DOI: 10.48550/arxiv.1201.0490