HHS Public Access

Author manuscript

Pac Symp Biocomput. Author manuscript; available in PMC 2025 January 01.

Published in final edited form as:

Pac Symp Biocomput. 2025 ; 30: 412–425.

A Pathway-Level Information ExtractoR (PLIER) framework to gain mechanistic insights into obesity in Down syndrome

Sutanu Nandi[1], **Yuehua Zhu**[2,3], **Lucas A Gillenwater**[1,4,5], **Marc Subirana-Granés**[6], **Haoyu Zhang**[6], **Negar Janani**[6], **Casey Greene**[4,6], **Milton Pividori**[5,6,*], **Maria Chikina**[2,*], **James C Costello**[1,4,5,6,*]

[1]Department of Pharmacology, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

[2]Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA, USA

[3]School of Medicine, Tsinghua University, Beijing, China

[4]Computational Bioscience Program, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

[5]Linda Crnic Institute for Down Syndrome, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

[6]Department of Biomedical Informatics, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

Abstract

Down syndrome (DS), caused by the triplication of chromosome 21 (T21), is a prevalent genetic disorder with a higher incidence of obesity. Traditional approaches have struggled to differentiate T21-specific molecular dysregulation from general obesity-related processes. This study introduces the omni-PLIER framework, combining the Pathway-Level Information ExtractoR (PLIER) with the omnigenic model, to uncover molecular mechanisms underlying obesity in DS. The PLIER framework aligns gene expression data with biological pathways, facilitating the identification of relevant molecular patterns. Using RNA sequencing data from the Human Trisome Project, omni-PLIER identified latent variables (LVs) significantly associated with both T21 and body mass index (BMI). Elastic net regression and causal mediation analysis revealed LVs mediating the effect of karyotype on BMI. Notably, LVs involving glutathione peroxidase-1 (GPX1) and MCL1 apoptosis regulator, BCL2 family members emerged as crucial mediators. These findings provide insights into the molecular interplay between DS and obesity. The omni-PLIER model offers a robust methodological advancement for dissecting complex genetic disorders, with implications for understanding obesity-related processes in both DS and the general population.

milton.pividori@cuanschutz.edu, mchikina@pitt.edu, james.costello@cuanschutz.edu.

*Co-Senior Authors

Keywords

Down syndrome; obesity; body mass index; matrix factorization; mediation analysis; RNA sequencing; mechanisms of disease; genetic/genomic studies; pathway analysis

1. Introduction

Down syndrome (DS), also known as trisomy 21 (T21), is the result of the triplication of chromosome 21 (chr21) and is the most frequent human aneuploidy[1]. Obesity, the result of disrupted metabolism leading to excessive adipose accumulation, is associated with increased comorbidities and decreased life expectancy[2]. Obesity, defined by a body mass index (BMI) ≥ 30, is more prevalent in individuals with DS than in the disomic population (D21)[3]. Multiple molecular profiling studies demonstrate systemic dysregulation of obesity-associated processes, including insulin resistance, oxidative phosphorylation, and lipid metabolism, in individuals with T21[4–7]. However, current approaches fail to adequately disentangle T21-specific from general molecular dysregulation in the pathogenesis of obesity. Thus, elucidating molecular mechanisms distinct to obesity in T21 will not only inform DS biology but also provide insights into obesity-related processes in the general population.

Mechanistic insights into DS are complicated by the simultaneous upregulation of most genes on chr21. While the mean overexpression of genes on chr21 is a 1.5X fold change, there is great variability in gene expression across people with DS[5]. Moreover, thousands of genes outside of chr21 are differentially expressed in people with DS. It is helpful to consider T21 in the context of the omnigenic model[8,9], which posits that gene regulatory networks are so highly interconnected that potentially all genes expressed in phenotype-relevant cell types have either a direct or indirect effect. Within this model, there are “core” genes that directly affect the phenotype and “peripheral” genes that indirectly affect the phenotype by regulating these core genes. Integrating gene co-expression modules and genome-wide association studies (GWAS) prioritizes genes missed by standard procedures while aiding interpretation[10]. Thus, the omnigenic framework can help understand the cascading effects in gene regulatory networks, which contribute to the co-occurring conditions in DS, such as obesity, through altered mechanisms compared to the D21.

The Pathway-Level Information ExtractoR (PLIER) is a semi-supervised matrix factorization framework[11]. It transforms an input matrix of high dimensional gene expression data into a relatively small number of latent variables (LVs) and then aligns these LVs with pre-defined pathway/geneset annotations. The LVs aim to maximize the variance within the data and the associated gene loadings are aligned with pathways/gene sets. By leveraging pathway/geneset annotations, PLIER achieves interpretable representations where LVs are more likely to align with independent measurements of biological pathways and processes. The LVs can be plugged into any supervised downstream analysis such as differential expression and eQTL discovery. The PLIER framework has been extensively adapted and reused in various applications[10,12–14].

Merging the omnigenic model with PLIER, we propose the omni-PLIER framework, a methodological advance to gain mechanistic insights into how complex genetic disorders drive the associated conditions. In the omnigenic model, gene co-expression modules impact downstream gene regulatory networks. Here, we use the PLIER model to define LVs as modules. Working with the hypothesis that a causal relationship between a genetic perturbation and a clinical phenotype must be mediated through molecular networks, we combine an elastic net model with causal inference methods to identify LVs derived from molecular data that are mediators in the formal statistical sense. We apply omni-PLIER to study the link between T21 and obesity and identify known and novel pathway associations providing a foundation for detailed follow-up studies. The omni-PLIER model is available at: https://github.com/CostelloLab/omni-PLIER

2. Methods

2.1. The omni-PLIER Framework

The omni-PLIER framework integrates gene expression data with clinical traits to identify latent variables (LVs) associated with both karyotype and BMI. This framework extends the PLIER model by applying elastic net regression and causal mediation analysis to uncover biological pathways mediating the relationship between DS and obesity. The framework allows for the discovery of mechanistic insights into how genetic perturbations drive phenotypic outcomes.

The omni-PLIER workflow, shown in Figure 1, proceeds as follows. 1) Input gene expression and sample annotation from the Human Trisome Project. 2) Apply the PLIER model to extract LVs aligned with known biological pathways. 3) Calculate elastic net regression and causal mediation analysis to identify significant LVs that mediate the relationship between T21 and BMI. 4) Output causal networks between LVs and phenotypes for further interpretation.

2.2. Human Trisome Project (HTP) RNA Sequencing Dataset

Under a study protocol approved by the Colorado Multiple Institutional Review Board (COMIRB #15–2170), the Crnic Institute enrolled participants as part of the Human Trisome Project (HTP; www.trisome.org). Demographic data for study participants were derived from participant and caregiver surveys and the annotation of medical records. Clinical variables relevant to this study include karyotype, age at visit, sex, and body mass index (BMI).

A detailed description of blood processing and molecular quantification for -omic profiling performed by the Human Trisome Project is described by Galbraith et al. and Waugh et al.[5,15]. Briefly, PAXgene RNA Tubes (Qiagen) were used to collect blood samples from 304 T21 and 95 D21 individuals. Whole-blood paired-end RNAseq was performed using Illumina NovaSeq 6000 instrument (Novogene). Reads were filtered for low quality, and adapters were trimmed. Reads were aligned to the human reference genome (assembly GRCh38) using STAR2 and quantified at the gene level to transcripts per million (TPM).

2.3. Gene Set Enrichment Analysis

The HTP RNA-seq dataset, along with the sample labels of karyotype or BMI ≥ 30, was input to the 'gseapy' python package (v 1.1.2) for gene set enrichment analysis. We utilized the same pathway information as in the PLIER model: Human Molecular Signatures Database (MSigDB v4.0) collections, C2 (curated gene sets), C6 (oncogenic signature gene sets), C7 (immunologic signature gene sets), bloodCellMarkersIRISDMAP and svmMarkers. These parameters were used: min_size = 5, max_size = 500, method = 'signal_to_noise', and permutation_num = 100,000.

2.4. PLIER model applied to HTP RNA-seq data

A gene-by-sample (g-by-s) matrix is factorized with k latent dimensions into $Z_{g \cdot k}\, B_{k \cdot s}$. In addition to the g-by-s matrix, PLIER considers an additional input of prior knowledge given by a gene-by-geneset binary matrix of pathway/geneset membership, C (g-by-p, where p is the number of pathways/genesets). PLIER enforces correspondence between the loadings Z and C by penalizing the distance between Z and its pathway-based prediction $C \cdot U$ (where U is a p-by-k matrix subject to optimization). An elastic-net penalty on the U coefficients ensures that each factor utilizes a small fraction of the pathways/genesets. The entire problem is optimized end-to-end using block coordinate minimization.

We determined the number of LVs (k parameter) by identifying the number of significant principal components using the *num.pcs* function in the PLIER R package (v0.1.6). The input expression matrix for the PLIER function was the HTP RNA-seq dataset, which was z-score transformed using the *rowNorm* function for genes. We incorporated prior information using the genesets defined in Section 2.2. The default settings of the PLIER function, which automatically configures the L1 and L2 parameters, were used to generate the LV (n=117) by sample matrix.

2.5. Regression model for latent variable-trait associations

We integrated gene-trait associations from the PrediXcan family of methods and PLIER LVs through generalized least squares (GLS) regression[10]. The PrediXcan family of methods was utilized for gene-based associations, including S-PrediXcan (for gene-tissue-trait associations) and S-MultiXcan (which combines S-PrediXcan results across tissues and computes gene-trait associations). Our GLS regression model computes an LV-trait association by fitting the model:

$$y = \beta_0 + s\beta_s + \sum_i x_i\beta_i + \in,$$

where y is a vector of S-MultiXcan gene p-values for a trait; s is a binary indicator vector $s_l = 1$ for the top 10% of genes with the largest weights for LV l and zero otherwise; x_i is a gene property used as a covariate (default covariates defined in Pividori et al.[10]); β are effect sizes (with β_0 as the intercept); and $\in \sim MVN(0, \sigma^2 \boldsymbol{R})$ are the error terms with a multivariate normal distribution (MVN) where $\boldsymbol{R}$ is the matrix of gene correlations. The model tests whether genes with high weights in an LV are more strongly associated with the phenotype than other genes with small or zero loadings. For more details, see Pividori

et al.[10] Consequently, we computed associations for five BMI traits in PhenomeXcan[16] (a large-scale resource with PrediXcan associations across the UK Biobank) across seven omni-PLIER LVs of interest. Due to the limited number of traits, we used nominal significance levels to assess the associations between traits and LVs.

2.6. Penalized Regression

The LV-by-sample matrix, *B*, generated from PLIER was batch corrected for the sample source variable using the Combat python package (v 0.20), and then LVs were z-score transformed. Karyotype and clinical variables (age, sex, BMI) were considered for downstream analysis.

We trained elastic net models for two prediction tasks, first to predict D21/T21 using 'LogisticRegressionCV', and second to predict BMI using 'ElasticNetCV', both from the sklearn (v 1.4.2) in python (v 3.11.0). The input dataset was split on the samples into 80% training and 20% testing sets. Using only the training dataset and 5-fold cross-validation, we tuned the α (ratio of L1 to L2 penalization) and λ (penalization weight) parameters using grid search across different ranges. We identified the optimal model parameters using balanced accuracy (T21/D21) or root mean squared error (BMI) over the 5-folds from the training dataset, trained the full model using the training dataset, and evaluated the performance on the testing dataset. Additionally, the contribution of each LV was calculated based on the best performing model coefficients. To establish the model's consistency and reproducibility, we repeated the procedure 1000 times. We then averaged the coefficients of all features from these 1000 models and ranked the features based on these averaged coefficients.

To establish a null model for comparison, we randomly shuffled the target labels (karyotype or BMI) and followed the elastic net model training above to evaluate model performance on the shuffled data. We repeated this procedure 1000 times to establish a null distribution of model performances, then the distribution of the model's performance on the original (unshuffled) data and the shuffled data were compared using a Kruskal-Wallis test.

2.7. Causal mediation analysis between molecular and clinical variables

The *mediation* R package (v4.5.0) was used to estimate the causal mediation effects of karyotype on BMI mediated through LVs, with sex and age as covariates. Outcome and mediator models were linear. We performed 100 Monte Carlo draws for quasi-Bayesian approximation (sims=100).

We performed causal analysis on selected LVs, karyotype, BMI, age, and sex. For causal discovery, we utilized the constrained continuous-optimization method PC-NOTEARS with a python implementation of the Peter-Clark algorithm for causal discovery (PC)[17] (causal-learn, v0.1.3.8) and a NOTEARS (Non-combinatorial Optimization via Trace Exponential and Augmented lagRangian for Structure learning)[18] implementation in the *bioCausal* R package (v0.1.0)[19], both with the edge constraints option available. The LVs were z-score transformed for input. PC is a conditional independence testing algorithm that has been extensively benchmarked and shows favorable performance. However, it may not orient all edges, and it does not estimate effects. NOTEARS continuous optimization is applied to

maximize the joint multivariate likelihood of the data under the constraint that the inferred relationships form a directed acyclic graph (DAG). Additionally, they provide causal effects. In recent benchmarking work we showed that a combination of PC and NOTEARS is optimal for biological network discovery[20]. In this setting, the continuous optimization is restricted to those edges returned by PC.

We manually set an edge constraint matrix to guarantee that karyotype, age, and sex do not have causal regulators. We ran PC with kernel-based conditional independence tests[21] and α=0.01, followed by NOTEARS with λ=0.01, for a total of 20 times with bootstrapped samples. The output adjacency matrices were averaged to increase the robustness of the output.

The top 10 ranked LVs from the elastic net model based on both predicting karyotype and BMI were used to filter the significant mediating LVs, which were then used for causal graph reconstruction and visualization. To validate the connection between karyotype, LV, and BMI, the correlation between karyotype and BMI was calculated before and after regressing out mediating LVs or non-mediating variables.

3. Results

3.1. Baseline gene expression and PLIER analysis

To establish a baseline comparison, we contrasted two alternative workflows for finding associations between phenotypes and pathways using identical input data. In the first workflow, we ran gene set enrichment analysis (GSEA) to identify genesets that were differentially regulated when comparing disomic (D21) to trisomic (T21) individuals. This comparison showed that there were no significant genesets (FDR < 0.25). We additionally compared individuals with a BMI ≥ 30 to those with a BMI < 30 and also found that there were no significant genesets identified.

In the second workflow, we applied PLIER to identify 117 LVs across the HTP cohort. We performed differential analysis using the LV-by-sample matrix and found LVs that were significantly up and down in both karyotype and BMI ≥ 30 (Figure 2A). Directly comparing the results from GSEA and PLIER showed that the PLIER LVs were highly enriched for phenotype associations, while the GSEA results were not, suggesting that the PLIER model identifies functionally relevant molecular patterns in the data that are differentially associated with the clinical variables of interest in this study, which are karyotype and BMI (Figure 2B).

3.2. BMI and karyotype associated latent variables

Given the strong signal found with the PLIER-identified LVs, we performed elastic net regression to first predict karyotype and second to predict BMI using the LVs. In both cases, the elastic net model showed robust predictive performance. We trained the model using 80% of the data and evaluated the held-out 20%. We performed this procedure over 1,000 iterations, randomly sampling to define training and test datasets. To establish the random model, we shuffled the sample labels and performed the same procedure. As shown in Figure 3A, the average balanced accuracy is 0.9 for predicting karyotype compared to the

expected 0.5 for the random model. For predicting BMI, the RMSE is 6.0 compared to 7.5 for the random model. A significant difference was seen for both prediction tasks (Kruskal–Wallis test, $p<0.001$). We next evaluated the LVs by comparing the average coefficient over the 1,000 iterations of the karyotype and BMI trained models. As shown in Figure 3B, 9 out of the top 10 LVs are unique in both conditions (BMI and karyotype), with LV56 being common to both. The top LV annotations are shown in Figure 3C.

3.3. Identifying causal mediators from highly ranked latent variables

Figure 4A illustrates the workflow for the causal mediation analysis, which was performed to estimate the average causal mediation effects (ACME) of karyotype on BMI mediated through each LV, considering age and sex as covariates. Average direct effect (ADE) and total effect of karyotype on BMI were also reported from each causal mediation test. LVs with p-values < 0.05 from the mediation test were defined as mediator LVs. We then intersected the elastic net model top 10 LVs for both karyotype and BMI models with the mediation analysis to determine which LVs were both causally mediating BMI though karyotype and were predictive of these conditions (Figure 4B). We identified 7 top-ranking LVs as causal mediators including LV37, LV76, and LV3. Figure 4C shows the causal mediation analysis result for LV37 and LV3.

3.4. Causal discovery prioritizes key causal mediators for BMI

To obtain a comprehensive understanding of the relationships between LVs and obesity in the HTP cohort, we performed causal analysis with all 16 mediating LVs using PCnt, a hybrid causal discovery method with no causal regulator constraints on karyotype, age, and sex (Figure 4A). A subgraph with seven selected LVs and clinical variables from bootstrapped output was visualized in Figure 4D. We found three direct causal regulators for BMI in the subgraph. LV3 and LV37 are the two mediating LVs in the directed path from karyotype to BMI, while age is an independent cause for BMI change. To validate the mediation effect of LV3 and LV37, we adjusted BMI for each of its direct regulators and calculated the correlation between karyotype and BMI (Figure 4E). A decrease in the karyotype-BMI correlation after regressing out mediating LVs, as opposed to regressing out age, adds to the confidence in prioritizing LV37 and LV3 as causal mediators.

3.5. Mechanisms underlying mediating latent variables

The omnigenic framework allows for the inference of both core genes that affect phenotypes and peripheral genes that propagate their effects across networks or LVs. In this study, we identified LV37 as the primary mediator of BMI through karyotype. Glutathione peroxidase-1 (GPX1) is the top gene in LV37. Interestingly, GPX1 has been previously implicated in obesity[22] and DS[23]. The relationship between karyotype and LV37 is mediated by LV43 (Figure 4D). MCL1 Apoptosis Regulator, BCL2 Family Member (MCL1) is the top gene in LV43. MCL1 has known involvement in apoptosis[24] and associations with acute myeloblastic leukemia (AML)[25]. Children with DS have a higher risk than the disomic population for AML[26], and MCL1 was previously identified as a target for treatment in leukemias in DS[27]. Furthermore, glutathione metabolism was previously implicated in modulating the efficacy of BCL2 inhibitors[28].

To interpret these results in terms of known DS biology, we identified which LVs in the causal network contained superoxide dismutase-1 (SOD1) with high loadings. SOD1 is located on chr21 and has been implicated in metabolic regulation of body weight and insulin levels[29]. Moreover, an altered SOD1/GPX1 ratio is observed in T21, contributing to the hallmark oxidative stress observed across DS phenotypes[23]. SOD1 had higher rankings in LV76 and LV3 compared to the other 117 LVs (ranking 5th and 13th, respectively). Interestingly, LV76 and LV3 were on causal pathways distinct from LV37. Furthermore, the position of SOD1 within these LVs suggests a peripheral role in the network, propagating effects that ultimately influence core genes.

LV3 was another mediator of BMI through karyotype (Figure 4C–E). The gene sets with the highest loadings for LV3 included those implicated in protein translation and a geneset of chr21 genes, supporting the mediation of karyotype (Figure 2C). The gene with the highest loading, *EEF1A1*, encodes an isoform of the alpha subunit of the elongation factor-1 complex responsible for the enzymatic delivery of aminoacyl tRNAs to the ribosome. Ribosomal dysfunction was previously implicated in a study on the impacts of high-fat diets on a DS mouse model (Ts65Dn)[30].

Notably, LV28, which mediates the LV34-LV3 relationship in a causal path between karyotype and BMI (Figure 4D), also has *EEF1A1* as the highest loading gene. Since the PLIER methodology optimizes for independent LVs, the high loading of *EEF1A1* implicates a distinct process related to protein translation as mediating the relationship between karyotype and BMI. By integrating PrediXcan gene-trait associations with LVs using a regression model, we identified a significant association between this LV and obesity ($p<0.05$) (Figure 5). This finding supports the involvement of LV28 with obesity in an independent analysis.

4. Discussion

The omni-PLIER framework presented here demonstrates how a complex network of phenotype-genotype/molecular trait interactions can be broken down into interpretable components, facilitating hypothesis formulation and validation. Disentangling disease co-occurrence in the context of T21 demonstrates the benefits of the omni-PLIER workflow.

A critical aspect of the omni-PLIER framework is leveraging the PLIER model to interpret gene expression through its latent components. Standard gene expression analysis performs statistical tests in gene space, such as differential expression (DE), and then conducts pathway level analysis, such as GSEA. This process requires defining groups for comparisons (contrasts) and uses predefined annotations to project gene level measurements into pathway/geneset space. In contrast, the PLIER model first performs sample annotation-agnostic latent variable extraction using pathways/genesets as prior information. Next, LVs are used to test the contrast groups and directly evaluate the LVs, which we have done here by treating the LVs as modules.

The PLIER approach offers several advantages. LVs capture multiple axes of variation while reducing redundant gene expression patterns. For example, a large group of highly

correlated genes can dominate the top of a DE list. In the case of PLIER, this group is reduced to a single variable, allowing other less dominant pathways to receive consideration. Additionally, the association between LVs and pathways/genesets is conducted through elastic net regression, so pathways "compete" to explain LVs. This reduces the number of redundant pathways/genesets, as multiple ones are included in the model only if they provide additional information.

The combination of these effects results in a considerable increase in contrast group-to-pathway associations inferred from the dataset, as we showed in Figure 2. We additionally combined PLIER LVs with causal modeling, which offers two advantages. First, we find that many LVs are associated with a clinical phenotype of interest. Here, we leverage this observation to gain mechanistic insights into how T21 can drive BMI. Second, causal modeling, a multivariate technique, optimizes a network representation of the data's conditional (in)dependent structure. PLIER analysis alone cannot address independence in the LVs. Similar to non-negative matrix factorization, PLIER LVs are not guaranteed to be orthogonal and are often associated with each other. Causal modeling provides a mathematical framework to organize these associations and infer directionality, though learning causal models from observational data remains challenging[20].

An important aspect of our study design is that one of the variables of interest is genotype, which provides a valuable constraint for learning directed causal models. Specifically, we look for molecular traits (PLIER LVs) that mediate the effect of genotype on BMI. The final result is a simplified network structure that lends itself to interpretation and candidate variable selection. While the number of LVs associated with clinical variables can be large, the combination of univariate mediation analysis and multivariate PC-NOTEARS analysis identified targeted LVs. In our case, three mediating LVs (two direct: LV37 and LV3, and one indirect: LV43).

From the limited number of mediating LVs, we provided evidence that they both support existing knowledge within DS and obesity, and find support for novel mechanisms. It is important to note that, within T21, the ground truth is unknown. Although many findings align with biological knowledge, we cannot directly verify if this approach yields actionable mechanistic insights. Defining suitable benchmarking scenarios that reflect the complexity of a real dataset while providing some notion of ground truth will be the focus of future work.

A particular challenge that is endemic to this area of research is that even well annotated, controlled, and deep molecular datasets are snapshots in time of a complex biological system. We face the same challenge with the HTP dataset. Additionally, the HTP study is based entirely on blood profiling, which lacks important molecular details from other tissues and cell types.

Despite these limitations, the omni-PLIER framework demonstrated a computationally efficient workflow that synthesizes a large number of observations, prior knowledge, and state-of-the-art algorithmic approaches into a unified analytic method.

Acknowledgments

This work is partly supported by HD109765, This work is partly supported by GM007635 and a Blumenthal Fellowship, This work is partly supported by HG011898

References

1. Mai CT, Isenburg JL, Canfield MA, et al. National population-based estimates for major birth defects, 2010–2014. Birth Defects Res. 2019;111(18):1420–1435. [PubMed: 31580536]
2. Ofei F. Obesity - a preventable disease. Ghana Med J. 2005;39(3):98–101. [PubMed: 17299552]
3. Bertapelli F, Pitetti K, Agiovlasitis S, Guerra-Junior G. Overweight and obesity in children and adolescents with Down syndrome-prevalence, determinants, consequences, and interventions: A literature review. Res Dev Disabil. 2016;57:181–192. [PubMed: 27448331]
4. Dierssen M, Fructuoso M, Martínez de Lagrán M, Perluigi M, Barone E. Down Syndrome Is a Metabolic Disease: Altered Insulin Signaling Mediates Peripheral and Brain Dysfunctions. Front Neurosci. 2020;14:670. [PubMed: 32733190]
5. Waugh KA, Minter R, Baxter J, et al. Triplication of the interferon receptor locus contributes to hallmarks of Down syndrome in a mouse model. Nat Genet. 2023;55(6):1034–1047. [PubMed: 37277650]
6. Bayona-Bafaluy MP, Garrido-Pérez N, Meade P, et al. Down syndrome is an oxidative phosphorylation disorder. Redox Biol. 2021;41:101871. [PubMed: 33540295]
7. Buonuomo PS, Bartuli A, Mastrogiorgio G, et al. Lipid profiles in a large cohort of Italian children with Down syndrome. Eur J Med Genet. 2016;59(8):392–395. [PubMed: 27343989]
8. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell. 2017;169(7):1177–1186. [PubMed: 28622505]
9. Liu X, Li YI, Pritchard JK. Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. Cell. 2019;177(4):1022–1034.e6. [PubMed: 31051098]
10. Pividori M, Lu S, Li B, et al. Projecting genetic associations through gene expression patterns highlights disease etiology and drug mechanisms. Nat Commun. 2023;14(1):5562. [PubMed: 37689782]
11. Mao W, Zaslavsky E, Hartmann BM, Sealfon SC, Chikina M. Pathway-level information extractor (PLIER) for gene expression data. Nat Methods. 2019;16(7):607–610. [PubMed: 31249421]
12. Ng B, Casazza W, Patrick E, et al. Using Transcriptomic Hidden Variables to Infer Context-Specific Genotype Effects in the Brain. Am J Hum Genet. 2019;105(3):562–572. [PubMed: 31447098]
13. Taroni JN, Grayson PC, Hu Q, et al. MultiPLIER: A Transfer Learning Framework for Transcriptomics Reveals Systemic Features of Rare Disease. Cell Syst. 2019;8(5):380–394.e4. [PubMed: 31121115]
14. Zhang S, Heil BJ, Mao W, Chikina M, Greene CS, Heller EA. MousiPLIER: A Mouse Pathway-Level Information Extractor Model. eNeuro. 2024;11(6). doi:10.1523/ENEURO.0313-23.2024
15. Galbraith MD, Rachubinski AL, Smith KP, et al. Multidimensional definition of the interferonopathy of Down syndrome and its response to JAK inhibition. Sci Adv. 2023;9(26):eadg6218. [PubMed: 37379383]
16. Pividori M, Rajagopal PS, Barbeira A, et al. PhenomeXcan: Mapping the genome to the phenome through the transcriptome. Sci Adv. 2020;6(37):eaba2083. [PubMed: 32917697]
17. Spirtes P, Glymour C, Scheines R. Causation, Prediction, and Search. MIT press; 2001.
18. Zheng X, Aragam B, Ravikumar P, Xing E. DAGs with NO TEARS: Continuous optimization for structure learning. Adv Neural Inf Process Syst. March 2018:9492–9503.
19. omni-PLIER. https://github.com/CostelloLab/omni-PLIER.
20. Zhu Yuehua, Benos Panayiotis V., and Chikina Maria. A Hybrid Constrained Continuous Optimization Approach for Optimal Causal Discovery from Biological Data. In: ECCB2024. Bioinformatics; 2024:bioinf - 2024–1389.

21. Zhang K, Peters J, Janzing D, Schölkopf B. Kernel-based Conditional Independence Test and Application in Causal Discovery. In: ; 01 2011:804–813.
22. Ahmad AA, Rahimi Z, Asadi S, Vaisi-Raygani A, Kohsari M. The GPx-1 Gene Variants (rs1050450) in Obesity: Association with the Risk of Obesity and the GPx Activity in Females. Rep Biochem Mol Biol. 2023;12(1):185–194. [PubMed: 37724151]
23. de Haan JB, Susil B, Pritchard M, Kola I. An altered antioxidant balance occurs in Down syndrome fetal organs: implications for the “gene dosage effect” hypothesis. J Neural Transm Suppl. 2003; (67):67–83. [PubMed: 15068240]
24. Widden H, Placzek WJ. The multiple mechanisms of MCL1 in the regulation of cell fate. Commun Biol. 2021;4(1):1029. [PubMed: 34475520]
25. Escudero S, Zaganjor E, Lee S, et al. Dynamic Regulation of Long-Chain Fatty Acid Oxidation by a Noncanonical Interaction between the MCL-1 BH3 Helix and VLCAD. Mol Cell. 2018;69(5):729–743.e7. [PubMed: 29499131]
26. Mateos MK, Barbaric D, Byatt SA, Sutton R, Marshall GM. Down syndrome and leukemia: insights into leukemogenesis and translational targets. Transl Pediatr. 2015;4(2):76–92. [PubMed: 26835364]
27. Toki T, Kanezaki R, Adachi S, et al. The key role of stem cell factor/KIT signaling in the proliferation of blast cells from Down syndrome-related leukemia. Leukemia. 2009;23(1):95–103. [PubMed: 18830255]
28. Ebner J, Schmoellerl J, Piontek M, et al. ABCC1 and glutathione metabolism limit the efficacy of BCL-2 inhibitors in acute myeloid leukemia. Nat Commun. 2023;14(1):5709. [PubMed: 37726279]
29. Sato A, Shiraishi Y, Kimura T, et al. Resistance to Obesity in SOD1 Deficient Mice with a High-Fat/High-Sucrose Diet. Antioxidants (Basel). 2022;11(7). doi:10.3390/antiox11071403
30. Sarver DC, Xu C, Velez LM, et al. Dysregulated systemic metabolism in a Down syndrome mouse model. Mol Metab. 2023;68:101666. [PubMed: 36587842]

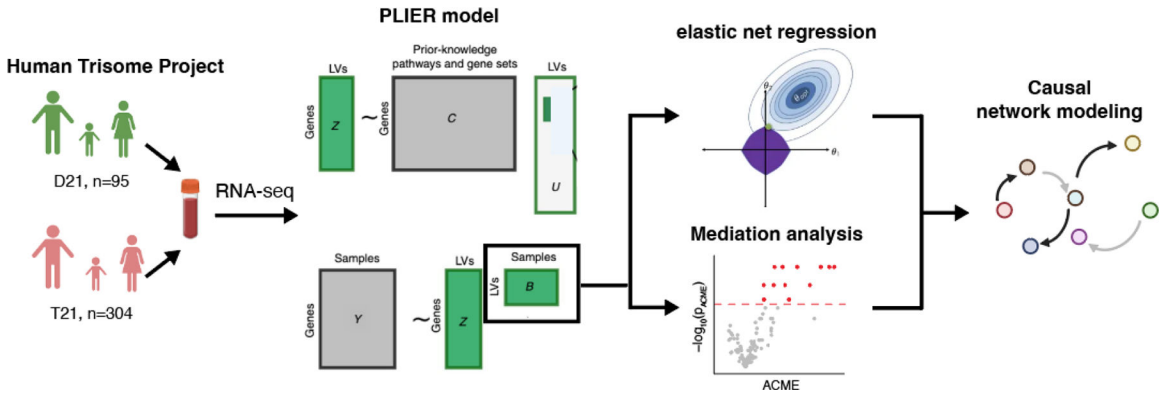

Fig 1.
Overview of the omni-PLIER framework

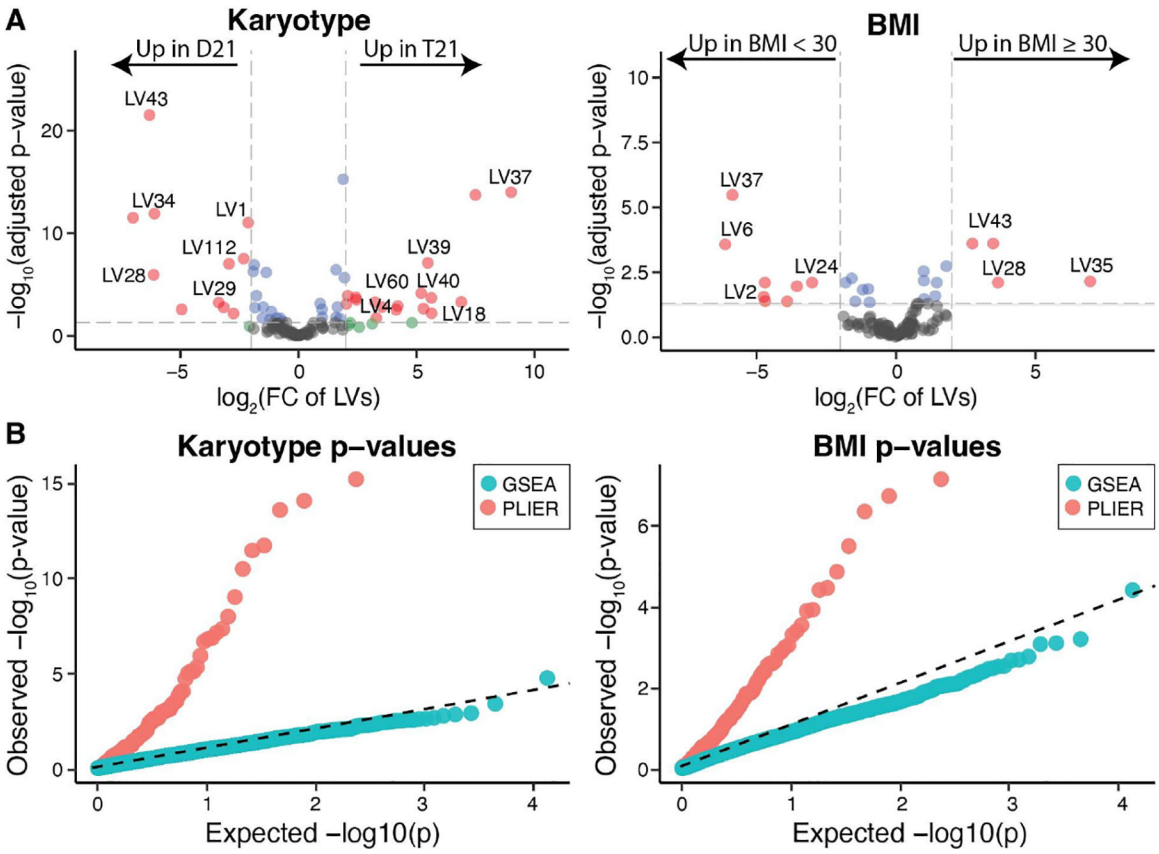

Fig 2.

(A) Differential latent variable analysis performed for karyotype (T21 vs. D21) and BMI (BMI ≥ 30 vs. BMI < 30). (B) Comparison between a GSEA pathway analysis p-values and p-values from the differential latent variable analysis. Each individual p-value represents a hypothesized pathway phenotype association.

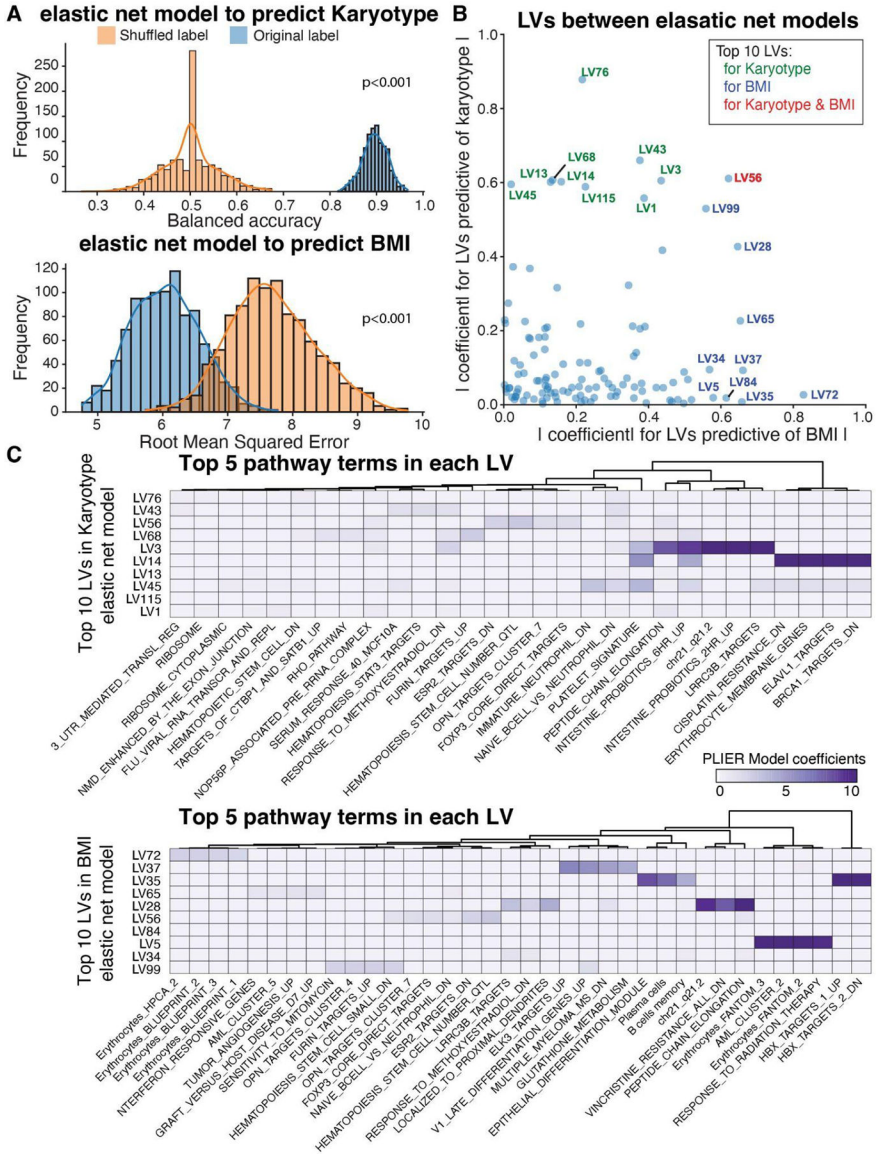

Fig 3.

(A) Comparison of model performance of 1000 elastic net models for karyotype and BMI prediction from the PLIER LVs in the HTP cohort. (B) Absolute value of LV coefficients between the karyotype and BMI models. The top 10 LVs are annotated by color. (C) The top 5 pathways associated with the top 10 LVs from the karyotype and BMI models are shown with the color representing model coefficients in the U matrix from PLIER.

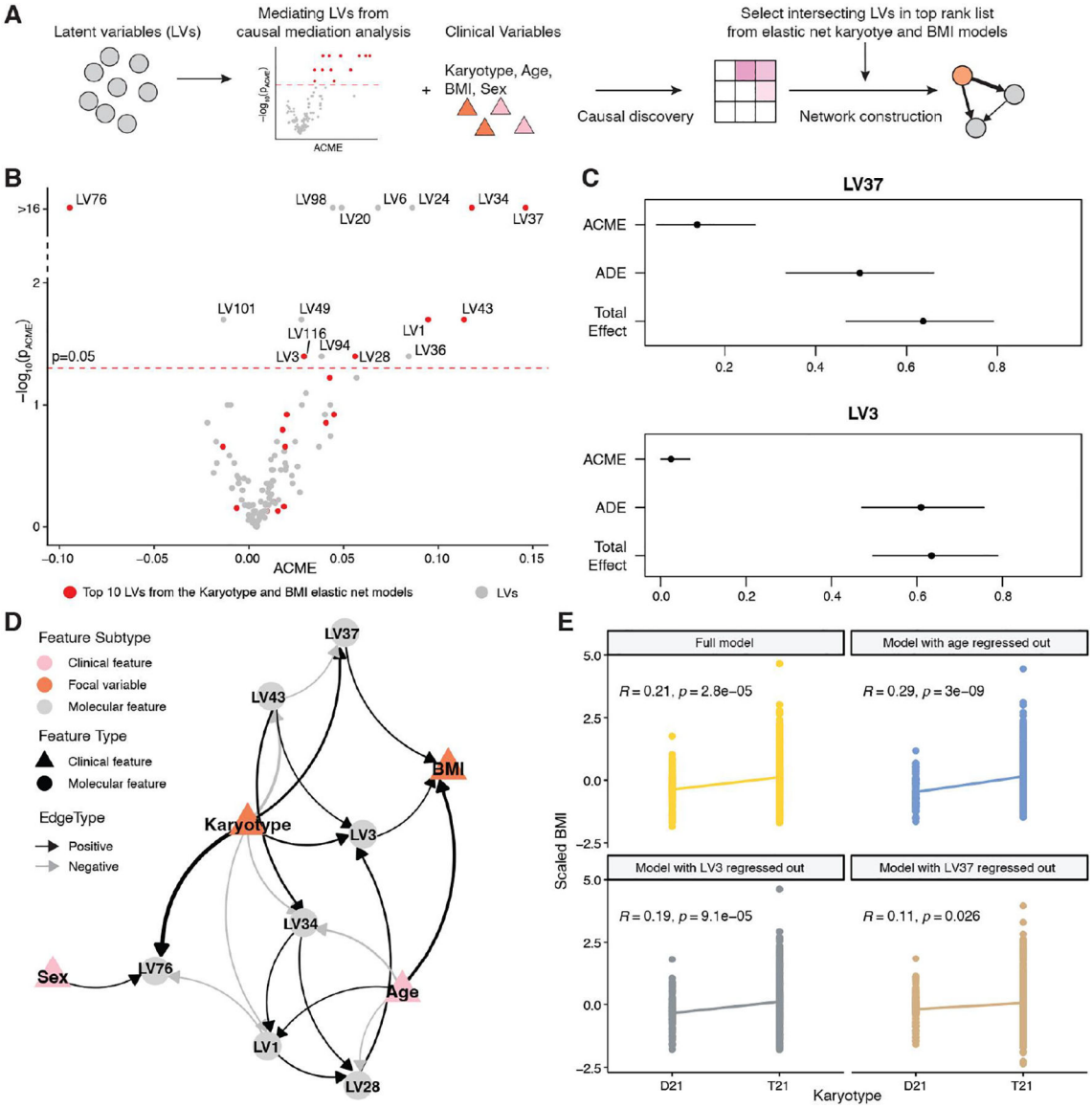

Fig 4.

Causal relationship between molecular and clinical variables. A) Causal analysis work flow. B) Volcano plot for average causal mediation effect (ACME) between karyotype and BMI for each latent variable (LV). C) ACME through LV37 and LV3, average direct effect (ADE) and total effect between karyotype and BMI. D) Causal sub-network with key LVs. Directed edges between nodes represent causal directions. Edge weight represents estimated causal effect size. E) Scatter plot for karyotype and BMI before and after regressing out key LVs from BMI.

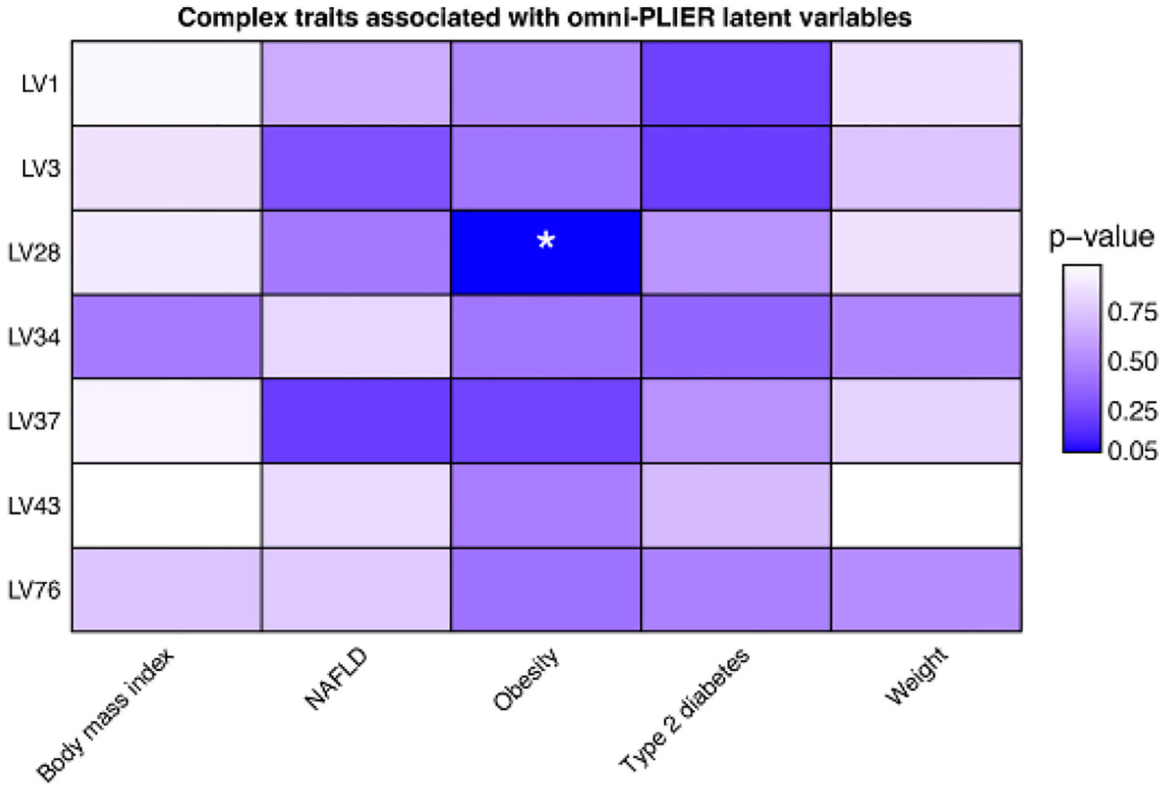

Fig 5.

The association between latent variables (LVs) and BMI PhenomeXcan selected traits regressed against the omni-PLIER LVs. The columns represent traits, including Body mass index, Non-Alcoholic Fatty Liver Disease (NAFLD), Obesity, Type 2 diabetes, and Weight. A significant association is marked by an asterisk (*).