

# Uncovering hidden gene-trait patterns through biclustering analysis of the UK Biobank

This manuscript ([permalink](#)) was automatically generated from [pivlab/biclustering-twas-manuscript@46bcc1a](#) on November 6, 2024.

## Authors

---

- **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 · Funded by The National Human Genome Research Institute (K99/R00 HG011898); The Eunice Kennedy Shriver National Institute of Child Health and Human Development (R01 HD109765)

- **Suraju Sadeeq**

Department of Biomedical Informatics, University of Colorado School of Medicine, Aurora, CO 80045, USA

- **Arjun Krishnan**

 [0000-0002-7980-4110](#) ·  [compbiologist](#)



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

- **Barbara E. Stranger**

 [0000-0001-9021-7331](#) ·  [be\\_stranger](#) ·  [@be\\_stranger@genomic.social](#)

Department of Biomedical Informatics, University of Colorado School of Medicine, Aurora, CO 80045, USA

- **Christopher R. Gignoux**

 [0000-0001-9728-6567](#) ·  [popgenepi](#)

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 Milton Pividori <milton.pividori@cuanschutz.edu>.

# Abstract

---

The growing availability of genome-wide association studies (GWAS) and large-scale biobanks provides an unprecedented opportunity to explore the genetic basis of complex traits and diseases. However, with this vast amount of data comes the challenge of interpreting numerous associations across thousands of traits, especially given the high polygenicity and pleiotropy underlying complex phenotypes. Traditional clustering methods, which identify global patterns in data, lack the resolution to capture overlapping associations relevant to subsets of traits or genes. Consequently, there is a critical need for innovative analytic approaches capable of revealing local, biologically meaningful patterns that could advance our understanding of trait comorbidities and gene-trait interactions. Here, we applied BiBit, a biclustering algorithm, to transcriptome-wide association study (TWAS) results from PhenomeXcan, a large resource of gene-trait associations derived from the UK Biobank. BiBit allows simultaneous grouping of traits and genes, identifying biclusters that represent local, overlapping associations. Our analyses uncovered biologically interpretable patterns, including asthma-related biclusters enriched for immune-related gene sets, connections between eye traits and blood pressure, and associations between dietary traits, high cholesterol, and specific loci on chromosome 19. These biclusters highlight gene-trait relationships and patterns of trait co-occurrence that may otherwise be obscured by traditional methods. Our findings demonstrate that biclustering can provide a nuanced view of the genetic architecture of complex traits, offering insights into pleiotropy and disease mechanisms. By enabling the exploration of complex, overlapping patterns within biobank-scale datasets, this approach provides a valuable framework for advancing research on genetic associations, comorbidities, and polygenic traits.

## Introduction

---

The availability of genome-wide association studies (GWAS) and large-scale biobanks, which house extensive genomic and phenotypic data, has revolutionized our ability to investigate the biology underlying complex traits and diseases [1,2]. These resources allow researchers to identify genetic loci associated with various traits, offering a broad view of genetic influences on disease susceptibility and comorbidity. However, the immense scale of GWAS data from diverse populations and phenotypes presents challenges for interpretation and synthesis [3]. As more GWAS findings accumulate, understanding how these genetic variants contribute to the polygenic and pleiotropic nature of complex traits becomes increasingly critical [4]. This complexity necessitates innovative analytical approaches to uncover and interpret subtle, biologically meaningful relationships within large datasets.

Traditional clustering methods, such as  $k$ -means or hierarchical clustering, have been widely applied to biological data to group traits or genes based on global patterns [5,6,7]. These methods have been effective in identifying broad patterns across entire datasets, grouping either samples or genes according to shared characteristics. However, complex traits and diseases often exhibit polygenic architectures where specific genes interact with multiple traits, and certain traits share overlapping genetic influences. Traditional clustering fails to capture these local, overlapping patterns [8], especially in high-dimensional data, where subsets of genes might only be relevant to subsets of traits. In such cases, biclustering approaches [9] offer a solution by simultaneously grouping genes and traits, identifying local patterns that reflect biological specificity. While some studies have demonstrated the effectiveness of biclustering in gene expression analysis [10,11], its application to GWAS or TWAS data remains limited, leaving an unmet need for methods that can disentangle the intricate connections within these associations.

In the realm of complex traits, transcriptome-wide association studies (TWAS) offer an additional layer of insight by linking genetic variants to gene expression and, subsequently, to traits [12,13,14,15].

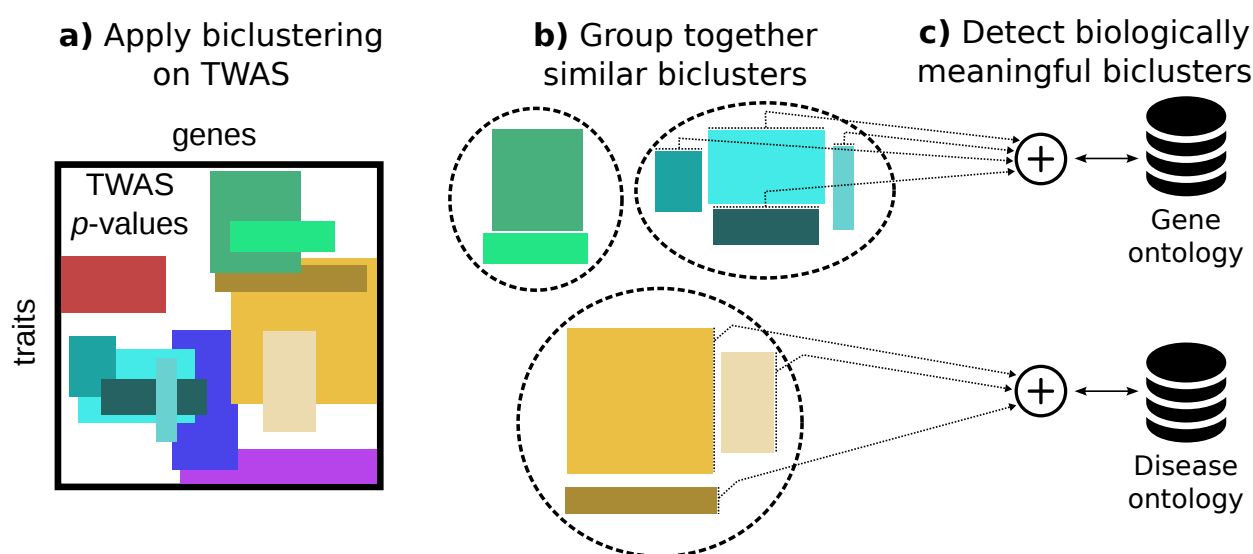
TWAS integrates gene-trait associations across tissues, enhancing our ability to infer potential causal genes and pathways underlying GWAS loci. However, even with TWAS, it remains challenging to interpret the vast array of gene-trait associations in a biologically meaningful way. Although TWAS has been instrumental in identifying genes involved in individual traits [16], understanding how these genes contribute to trait comorbidity and pleiotropy across many traits and tissues requires an advanced analytical approach that can manage high-dimensional, interdependent data relationships. Therefore, a key gap in the field is the need for methods that can capture these overlapping, complex associations in a way that supports the biological interpretation of TWAS results, particularly across the wide phenotypic spectrum represented in biobank-scale datasets.

In this study, we address this gap by applying a biclustering approach to TWAS results from PhenomeXcan [17], an extensive resource derived from the UK Biobank [2]. Our method uses BiBit [18], a biclustering algorithm that simultaneously groups traits and genes based on local patterns of association. This approach enables the detection of overlapping biclusters where a subset of genes may be linked to multiple traits and vice versa, offering a nuanced view of gene-trait interactions. Our analyses reveal several biologically relevant biclusters, such as those connecting immune-related gene sets to asthma diagnoses, linking eye traits with blood pressure, and associating dietary and cholesterol traits with genes on chromosome 19. These findings demonstrate the utility of biclustering for uncovering patterns that enhance our understanding of complex trait architecture and open new avenues for exploring disease mechanisms and comorbidities.

## Results

### Overview of the biclustering approach

To explore patterns within TWAS results from the UK Biobank, we utilized PhenomeXcan [17], a comprehensive resource of gene-trait associations based on 4,091 GWAS summary statistics from publicly available data [19] and the UK Biobank [2]. We applied the S-MultiXcan approach [14], a cross-tissue method that aggregates gene-trait associations across tissues from PrediXcan [12,13], enhancing statistical power for gene-trait associations. This yielded an association matrix containing  $p$ -values for 4,091 traits and 22,515 genes.

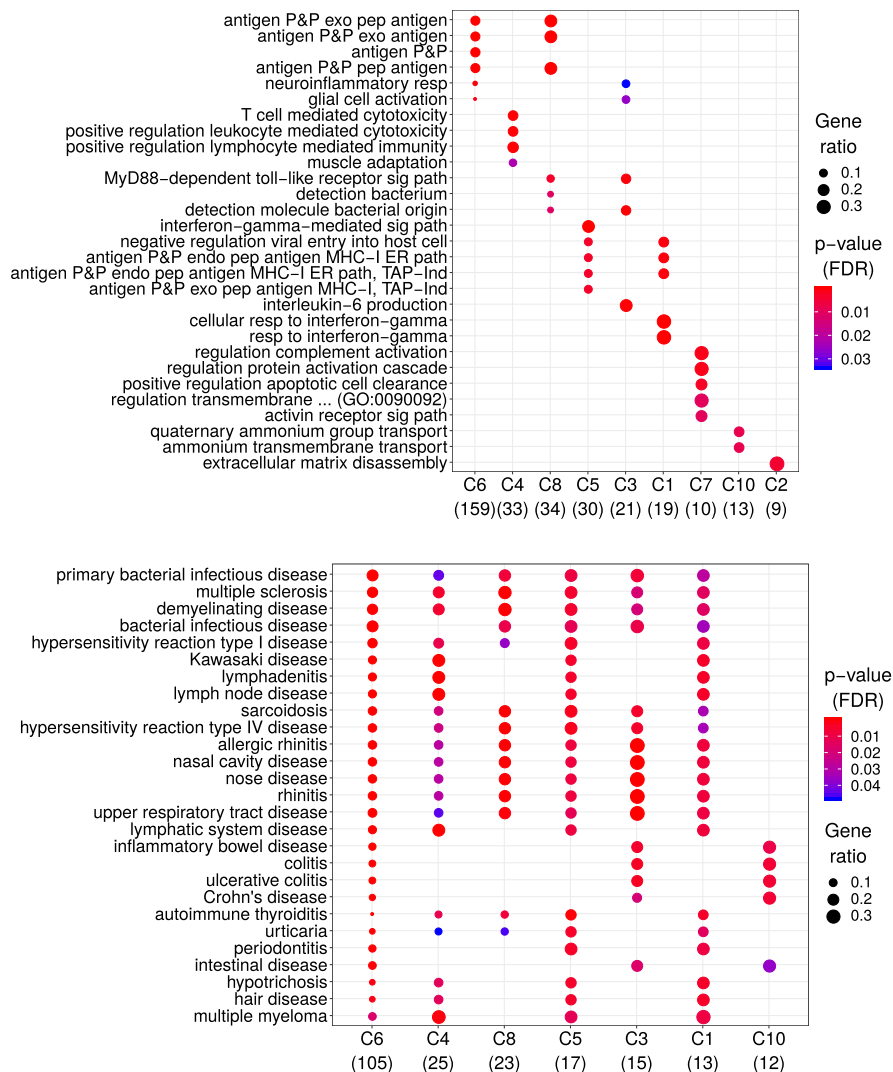


**Figure 1: Schematic of the computational analyses.** **a)** The BiBit biclustering algorithm was applied to the S-MultiXcan association matrix to identify biclusters—rectangles where subsets of traits (rows) are associated with subsets of genes (columns). **b)** Biclusters were grouped based on gene overlap using a clustering algorithm on the Jaccard similarity coefficient. **c)** Gene Ontology (GO) and Disease Ontology (DO) over-representation analysis was performed on grouped biclusters to assess biological pathway representation.

To capture local patterns of gene-trait associations, we applied the BiBit biclustering algorithm, which detects optimal biclusters in a binarized gene-trait association matrix (see Figure 1a). We binarized the matrix using a Bonferroni-corrected  $p$ -value threshold ( $5.49 \times 10^{-10}$ ), allowing BiBit to detect associations where all gene-trait pairs in each bicluster met this significance threshold. BiBit yielded 20,494 biclusters, with each bicluster representing gene and trait subsets that share a statistically significant association. The resulting data and interactive exploration tools are available at <https://pivlab.github.io/biclustering-twas/>.

## Asthma-related biclusters are enriched for immune-related gene sets

To determine if the biclustering approach could capture biologically meaningful patterns, we focused on biclusters containing asthma-related traits. Given the overlap in biclusters, we grouped similar biclusters into meta-biclusters based on gene overlap (Figure 1b). Our steps included: 1) selecting biclusters with asthma-related traits (self-reported asthma, ICD-10 codes J45/J46, and age of asthma onset) and at least 10 associated genes, 2) calculating the Jaccard similarity coefficient for pairwise bicluster comparisons, 3) clustering biclusters with high gene overlap, yielding 10 meta-biclusters, and 4) performing over-representation analyses for Gene Ontology (GO) terms (for genes) and Disease Ontology (DO) terms (for traits) to assess distinct biological mechanisms for asthma.



**Figure 2: Functional enrichment of asthma-related biclusters.** Enrichment analyses of 10 asthma-related meta-biclusters revealed distinct GO terms across genes (top) and DO terms across traits (bottom), illustrating pathway and disease enrichment for asthma susceptibility.

As shown in Figure 2, each meta-bicluster was enriched for unique GO and DO terms. Nine of the ten meta-biclusters demonstrated distinct GO enrichment, suggesting diverse biological pathways contributing to asthma. For example, meta-bicluster C6 included traits such as celiac disease, systemic lupus erythematosus, and type 1 diabetes, with genes localized to the HLA region on chromosome 6, underscoring its role in autoimmunity. Other meta-biclusters, such as C8, were associated with allergic diseases and the age of asthma onset, with strong links to the 17q12-21 locus, a well-established region for early-onset asthma [20,21,22]. Consistent with a large study on pleiotropy [4], these results suggest that a combination of genes seems to be specific to a particular combination of traits.

## Biclusters associated with eye measurements

We next investigated biclusters linked to eye-related traits to understand potential gene-trait connections in ocular health. Two notable biclusters involved traits like “Age started wearing glasses” and keratometry measurements alongside blood pressure.

|                              | <i>TXLNB</i> | <i>C2orf72</i> |                             | <i>SPPL3</i> | <i>RP11-328D5.4</i> |
|------------------------------|--------------|----------------|-----------------------------|--------------|---------------------|
| Age started wearing glasses  | 11.08        | 12.26          | Age started wearing glasses | 12.51        | 13.33               |
| 6mm cylindrical power*       | 11.23        | 10.33          | Spherical power*            | 12.59        | 16.57               |
| 6mm weak meridian angle*     | 10.75        | 10.29          | Systolic blood pressure     | 10.54        | 10.06               |
| 6mm cylindrical power angle* | 10.69        | 10.25          |                             |              |                     |

**Figure 3: Biclusters associated with eye measurements.** Two biclusters showing associations between “Age started wearing glasses” and keratometry measurements (left), and blood pressure (right). Eye traits are marked with an asterisk; cells contain  $z$ -scores derived from  $p$ -values.

In Figure 3, the left table shows the bicluster associated with “Age started wearing glasses” and keratometry measures, while the right table displays associations with blood pressure. The gene **TXLNB**, involved in muscle function [23], was notable in these associations, aligning with research linking blood pressure with eye health [24]. These findings indicate potential links between eye traits and blood pressure mediated by genes influencing muscular and vascular function.

## Biclusters associated with anthropometric traits

We also identified biclusters related to anthropometric traits, which included a diverse array of body composition metrics.

|                            | <i>TFAP2B</i> | <i>ZFP64</i> | <i>ZCCHC8</i> | ... | <i>USP36</i> | <i>SEMA4G</i> | <i>ZZZ3</i> |
|----------------------------|---------------|--------------|---------------|-----|--------------|---------------|-------------|
| Body mass index            | 13.40         | 11.90        | 10.36         | ... | 11.97        | 11.44         | 11.56       |
| Weight                     | 11.51         | 14.74        | 17.79         | ... | 27.98        | 19.88         | 14.59       |
| Hip circumference          | 11.96         | 11.93        | 15.48         | ... | 16.65        | 13.08         | 12.86       |
| Whole body fat mass        | 11.13         | 11.10        | 15.04         | ... | 19.05        | 13.58         | 12.15       |
| Trunk fat mass             | 10.02         | 10.81        | 15.67         | ... | 19.88        | 13.38         | 11.91       |
| Arm fat mass (right)       | 12.42         | 10.28        | 13.54         | ... | 15.24        | 12.85         | 11.77       |
| Arm fat mass (left)        | 12.61         | 10.08        | 13.55         | ... | 15.05        | 12.44         | 11.29       |
| Leg predicted mass (left)  | 11.07         | 15.33        | 17.81         | ... | 29.95        | 22.70         | 12.80       |
| Leg predicted mass (right) | 10.94         | 14.82        | 17.15         | ... | 27.98        | 22.42         | 12.50       |
| Leg fatfree mass (left)    | 11.10         | 15.25        | 17.80         | ... | 29.72        | 22.61         | 12.88       |
| Leg fatfree mass (right)   | 10.87         | 14.85        | 17.14         | ... | 28.06        | 22.32         | 12.55       |
| Leg fat mass (left)        | 11.22         | 10.82        | 12.87         | ... | 17.18        | 12.07         | 12.31       |
| Leg fat mass (right)       | 11.01         | 10.90        | 12.92         | ... | 17.68        | 12.38         | 12.23       |

**Figure 4: Bicluster associated with anthropometric traits.** A bicluster showing associations between 15 anthropometric traits, such as BMI and hip circumference, and 68 genes, a subset of which are shown. Cells contain  $z$ -scores derived from  $p$ -values.

As shown in Figure 4, this bicluster includes 15 traits related to body composition, including BMI, weight, and measurements of fat and lean mass. Key genes, such as *TFAP2B* and *USP36*, have roles in adipogenesis and metabolic regulation, suggesting involvement in body composition pathways. These

genes support the hypothesis that specific genetic networks influence distinct aspects of anthropometric traits, including fat and lean mass distribution.

## Biclusters associated with high cholesterol and dietary traits

Finally, we identified biclusters connecting high cholesterol with dietary and metabolic traits.

|                       | RASIP1 | FUT1  | FUT2  | IZUMO1 |
|-----------------------|--------|-------|-------|--------|
| Oily fish intake      | 13.92  | 14.91 | 14.81 | 14.67  |
| Sitting height        | 10.27  | 11.86 | 10.33 | 11.7   |
| Sodium in urine       | 12.54  | 16.2  | 12.12 | 13.67  |
| Cholelithiasis        | 10.82  | 11.68 | 12.37 | 10.99  |
| High cholesterol      | 12.58  | 14.83 | 14.1  | 12.34  |
| Mean platelet volume  | 19.44  | 23.59 | 18.48 | 21.58  |
| Processed meat intake | 10.4   | 11.83 | 10.56 | 14.51  |
| Nonoily fish intake   | 14.29  | 12.07 | 15.02 | 15.61  |

**Figure 5: Bicluster associated with high cholesterol and dietary traits.** This bicluster shows associations among high cholesterol, cholelithiasis, dietary intake, and specific genes located on chromosome 19. Cells are  $z$ -scores derived from  $p$ -values, with conditional formatting to highlight significance.

Figure 5 displays a bicluster associated with high cholesterol, cholelithiasis, and dietary traits, including processed meat intake, fish intake, and sodium levels. Genes in this cluster, such as *RASIP1*, *FUT1*, *FUT2*, and *IZUMO1*, are located within a gene-dense region on chromosome 19. This region, associated with macronutrient intake through genes like *FGF21* [25], suggests links between diet and chronic conditions such as obesity and diabetes. This bicluster highlights how dietary factors may interact with cholesterol metabolism, offering insights into gene-diet interactions relevant to metabolic health.

These results demonstrate that biclustering provides a framework to identify biologically relevant gene-trait relationships, supporting the exploration of complex traits and comorbidities within large biobank datasets.

## Discussion

This study leverages a biclustering approach, specifically the BiBit algorithm, to reveal complex associations within TWAS data, providing insights into the genetic architecture of multiple traits in the UK Biobank. By capturing both gene-trait and trait-trait patterns, we uncovered biologically relevant associations such as asthma-related immune gene clusters, eye trait associations with keratometry and blood pressure, and anthropometric traits connected to genes involved in adipogenesis and metabolic regulation. These findings underscore the potential of biclustering to enhance our understanding of pleiotropy and trait comorbidities, offering a novel framework for exploring the polygenic and interconnected nature of complex traits.

One strength of our study is the use of biclustering to detect local patterns that traditional clustering methods may overlook. While traditional clustering captures only global patterns, biclustering allows for the simultaneous grouping of genes and traits, revealing gene sets associated with specific trait subgroups. This flexibility is particularly beneficial in complex, polygenic traits like asthma, where subsets of genes and phenotypes contribute to distinct pathways. While this approach successfully identified asthma-related clusters enriched for immune-related pathways, the reliance on binary input and a Bonferroni-corrected  $p$ -value threshold may limit the detection of more subtle associations.

Nonetheless, the identified asthma biclusters are consistent with previous studies on autoimmune and allergic pathways, supporting the validity of our findings.

A limitation to consider is that our biclustering approach does not incorporate tissue specificity, which can affect gene-trait associations in TWAS data. The integration of tissue-specific data could refine the resolution of our biclusters and enhance the biological relevance of the findings. For instance, incorporating tissue information could help distinguish genes that influence systemic traits from those affecting tissue-specific phenotypes, such as eye or respiratory conditions. Future studies could address this limitation by using tissue-aware biclustering methods or incorporating multi-omic data to capture more nuanced layers of biological regulation.

Additionally, while we observed biclusters associating anthropometric traits with genes like *TFAP2B* linked to skeletal and adipogenic pathways, our findings rely on existing TWAS data that may reflect genetic correlations rather than causal relationships. Although TWAS improves on GWAS by prioritizing likely causal genes, causal inference remains a challenge due to potential confounders and linkage disequilibrium effects. Future work could employ Mendelian randomization or fine-mapping to validate the causal roles of identified genes within these biclusters, thereby strengthening the biological interpretations of our findings.

Despite these limitations, our study makes a significant contribution to the field by demonstrating how biclustering can uncover multi-dimensional patterns in TWAS data, an approach that is highly adaptable to other biobank-scale datasets. By revealing new connections between genes and phenotypes, this method provides a foundation for future research on genetic correlations, pleiotropy, and disease comorbidity, fostering new avenues for understanding the biological pathways underpinning complex traits. Moreover, the interactive online resource we provide allows other researchers to explore the identified biclusters, potentially catalyzing novel hypotheses and collaborative efforts to validate and expand upon our findings.

In conclusion, this work advances the field by presenting an innovative analytical framework that bridges gaps in current GWAS and TWAS studies, empowering researchers to interpret and synthesize the large volumes of data from biobanks like the UK Biobank. Our approach holds promise for accelerating biomedical discovery by facilitating the identification of complex trait architectures and may serve as a valuable tool for exploring the polygenic basis of traits and diseases in the post-GWAS era.

## Methods

---

### Gene-trait associations with TWAS

We used PhenomeXcan [17], a large resource of TWAS results with gene-tissue-trait associations across 4,091 diseases and traits from the UK Biobank [2] and other cohorts, 49 tissues from GTEx, and more than 22,000 genes. PhenomeXcan was built using publicly available GWAS summary statistics from Neale's lab (GWAS round 2 [19]) to compute 1) gene-based associations with the PrediXcan family of methods [12,13,14], including Summary-PrediXcan (S-PrediXcan) [13], which generates gene-tissue-trait associations, and Summary-MultiXcan (S-MultiXcan) [14], which combines tissue-specific results from S-PrediXcan to generate gene-trait associations; and 2) a locus regional colocalization probability (RCP) between GWAS loci and *cis*-eQTL (i.e., the overlapping of prioritized causal variants for a trait and for a gene's expression) with fastENLOC, a Bayesian hierarchical model for large-scale data [17,26]. The PrediXcan family of methods first builds prediction models using data from the Genotype-Tissue Expression project (GTEx v8) [27] for gene expression imputation and then correlates this predicted expression with the phenotype of interest.

## Biclustering analysis

We used a biclustering method that can jointly model pleiotropic genes and polygenic traits by allowing overlapping biclusters. For this, we applied the BiBit algorithm [18] on gene-trait associations from PhenomeXcan.

Since BiBit works on binary data, given a matrix  $M^{n \times m}$  ( $n = 4,091$  traits and  $m = 22,515$  genes) with  $p$ -values from S-MultiXcan, we defined the binary matrix  $E^{n \times m}$ , with  $E_{ij} = 1$  if  $M_{ij} < t$ , and  $E_{ij} = 0$  otherwise, where  $t$  was set to a Bonferroni significant  $p$ -value ( $5.49 \times 10^{-10}$ ). A bicluster  $\mathcal{B}_\ell = (\mathcal{T}_\ell, \mathcal{G}_\ell)$  corresponds to a subset of traits  $\mathcal{T}_\ell \subset \{1, \dots, n\}$  that are all jointly associated with a subset of genes  $\mathcal{G}_\ell \subset \{1, \dots, m\}$ , such that  $M_{ij} < t, \forall i \in \mathcal{T}_\ell, \forall j \in \mathcal{G}_\ell$ .

## Hierarchical clustering of biclusters and functional analyses

Biclusters across the entire set of results can have different degrees of overlap, which means that they can be grouped into several clusters. These clusters can be summarized by the frequency of phenotypes and genes across their biclusters. Thus, we proceeded as follows: 1) select all biclusters containing at least 10 genes and a disease of interest, such as an asthma diagnosis (self-reported asthma, ICD-10 codes J45/J46, and age of asthma onset); 2) perform a pairwise comparison of all these biclusters using the Jaccard similarity coefficient over their associated genes; 3) run a hierarchical clustering algorithm (average linkage) and obtain 10 groups of biclusters; 4) perform an over-representation analysis of Gene Ontology (GO) terms to compare those 10 groups of biclusters and assess whether their genes represent known and distinct biological mechanisms for the disease of interest; the same analysis can be done using the Disease Ontology (DO) terms on the traits.



## References

---

- 10 Years of GWAS Discovery: Biology, Function, and Translation**  
Peter M Visscher, Naomi R Wray, Qian Zhang, Pamela Sklar, Mark I McCarthy, Matthew A Brown, Jian Yang  
*The American Journal of Human Genetics* (2017-07) <https://doi.org/gcsmnm>  
DOI: [10.1016/j.ajhg.2017.06.005](https://doi.org/10.1016/j.ajhg.2017.06.005) · PMID: [28686856](https://pubmed.ncbi.nlm.nih.gov/28686856/) · PMCID: [PMC5501872](https://pubmed.ncbi.nlm.nih.gov/PMC5501872/)
- The UK Biobank resource with deep phenotyping and genomic data**  
Clare Bycroft, Colin Freeman, Desislava Petkova, Gavin Band, Lloyd T Elliott, Kevin Sharp, Allan Motyer, Damjan Vukcevic, Olivier Delaneau, Jared O'Connell, ... Jonathan Marchini  
*Nature* (2018-10) <https://doi.org/gfb7h2>  
DOI: [10.1038/s41586-018-0579-z](https://doi.org/10.1038/s41586-018-0579-z) · PMID: [30305743](https://pubmed.ncbi.nlm.nih.gov/30305743/) · PMCID: [PMC6786975](https://pubmed.ncbi.nlm.nih.gov/PMC6786975/)
- The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019**  
Annalisa Buniello, Jacqueline A L MacArthur, Maria Cerezo, Laura W Harris, James Hayhurst, Cinzia Malangone, Aoife McMahon, Joannella Morales, Edward Mountjoy, Elliot Sollis, ... Helen Parkinson  
*Nucleic Acids Research* (2018-11-16) <https://doi.org/gf5bz5>  
DOI: [10.1093/nar/gky1120](https://doi.org/10.1093/nar/gky1120) · PMID: [30445434](https://pubmed.ncbi.nlm.nih.gov/30445434/) · PMCID: [PMC6323933](https://pubmed.ncbi.nlm.nih.gov/PMC6323933/)
- A global overview of pleiotropy and genetic architecture in complex traits**  
Kyoko Watanabe, Sven Stringer, Oleksandr Frei, Maša Umičević Mirkov, Christiaan de Leeuw, Tinca JC Polderman, Sophie van der Sluis, Ole A Andreassen, Benjamin M Neale, Danielle Posthuma  
*Nature Genetics* (2019-08-19) <https://doi.org/ggr84r>  
DOI: [10.1038/s41588-019-0481-0](https://doi.org/10.1038/s41588-019-0481-0) · PMID: [31427789](https://pubmed.ncbi.nlm.nih.gov/31427789/)
- Dynamic regulatory network controlling TH17 cell differentiation**  
Nir Yosef, Alex K Shalek, Jellert T Gaublomme, Hulin Jin, Youjin Lee, Amit Awasthi, Chuan Wu, Katarzyna Karwacz, Sheng Xiao, Marsela Jorgolli, ... Aviv Regev  
*Nature* (2013-03-06) <https://doi.org/f4t3wq>  
DOI: [10.1038/nature11981](https://doi.org/10.1038/nature11981) · PMID: [23467089](https://pubmed.ncbi.nlm.nih.gov/23467089/) · PMCID: [PMC3637864](https://pubmed.ncbi.nlm.nih.gov/PMC3637864/)
- Identification of transcriptional regulators in the mouse immune system**  
, Vladimir Jojic, Tal Shay, Katelyn Sylvia, Or Zuk, Xin Sun, Joonsoo Kang, Aviv Regev, Daphne Koller  
*Nature Immunology* (2013-04-28) <https://doi.org/f4xz93>  
DOI: [10.1038/ni.2587](https://doi.org/10.1038/ni.2587) · PMID: [23624555](https://pubmed.ncbi.nlm.nih.gov/23624555/) · PMCID: [PMC3690947](https://pubmed.ncbi.nlm.nih.gov/PMC3690947/)
- Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors**  
Franziska Paul, Ya'ara Arkin, Amir Giladi, Diego Adhemar Jaitin, Ephraim Kenigsberg, Hadas Keren-Shaul, Deborah Winter, David Lara-Astiaso, Meital Gury, Assaf Weiner, ... Ido Amit  
*Cell* (2015-12) <https://doi.org/f74vmp>  
DOI: [10.1016/j.cell.2015.11.013](https://doi.org/10.1016/j.cell.2015.11.013) · PMID: [26627738](https://pubmed.ncbi.nlm.nih.gov/26627738/)
- A comprehensive evaluation of module detection methods for gene expression data**  
Wouter Saelens, Robrecht Cannoodt, Yvan Saeys  
*Nature Communications* (2018-03-15) <https://doi.org/gc9x36>  
DOI: [10.1038/s41467-018-03424-4](https://doi.org/10.1038/s41467-018-03424-4) · PMID: [29545622](https://pubmed.ncbi.nlm.nih.gov/29545622/) · PMCID: [PMC5854612](https://pubmed.ncbi.nlm.nih.gov/PMC5854612/)
- A comparative analysis of biclustering algorithms for gene expression data**  
K Eren, M Deveci, O Kucuktunc, UV Catalyurek

*Briefings in Bioinformatics* (2012-07-06) <https://doi.org/f4zcrm>  
DOI: [10.1093/bib/bbs032](https://doi.org/10.1093/bib/bbs032) · PMID: [22772837](https://pubmed.ncbi.nlm.nih.gov/22772837/) · PMCID: [PMC3659300](https://pubmed.ncbi.nlm.nih.gov/PMC3659300/)

10. **A systematic comparison and evaluation of biclustering methods for gene expression data**  
Amela Prelić, Stefan Bleuler, Philip Zimmermann, Anja Wille, Peter Bühlmann, Wilhelm Gruissem, Lars Hennig, Lothar Thiele, Eckart Zitzler  
*Bioinformatics* (2006-02-24) <https://doi.org/dw6sgp>  
DOI: [10.1093/bioinformatics/btl060](https://doi.org/10.1093/bioinformatics/btl060) · PMID: [16500941](https://pubmed.ncbi.nlm.nih.gov/16500941/)
11. **Biclustering Methods: Biological Relevance and Application in Gene Expression Analysis**  
Ali Oghabian, Sami Kilpinen, Sampsa Hautaniemi, Elena Czeizler  
*PLoS ONE* (2014-03-20) <https://doi.org/f53jzv>  
DOI: [10.1371/journal.pone.0090801](https://doi.org/10.1371/journal.pone.0090801) · PMID: [24651574](https://pubmed.ncbi.nlm.nih.gov/24651574/) · PMCID: [PMC3961251](https://pubmed.ncbi.nlm.nih.gov/PMC3961251/)
12. **A gene-based association method for mapping traits using reference transcriptome data**  
Eric R Gamazon, Heather E Wheeler, Kaanan P Shah, Sahar V Mozaffari, Keston Aquino-Michaels, Robert J Carroll, Anne E Eyler, Joshua C Denny, Dan L Nicolae, ... Hae Kyung Im  
*Nature Genetics* (2015-08-10) <https://doi.org/f7p9zv>  
DOI: [10.1038/ng.3367](https://doi.org/10.1038/ng.3367) · PMID: [26258848](https://pubmed.ncbi.nlm.nih.gov/26258848/) · PMCID: [PMC4552594](https://pubmed.ncbi.nlm.nih.gov/PMC4552594/)
13. **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](https://doi.org/10.1038/s41467-018-03621-1) · PMID: [29739930](https://pubmed.ncbi.nlm.nih.gov/29739930/) · PMCID: [PMC5940825](https://pubmed.ncbi.nlm.nih.gov/PMC5940825/)
14. **Integrating predicted transcriptome from multiple tissues improves association detection**  
Alvaro N Barbeira, Milton Pividori, Jiamao Zheng, Heather E Wheeler, Dan L Nicolae, Hae Kyung Im  
*PLOS Genetics* (2019-01-22) <https://doi.org/ghs8vx>  
DOI: [10.1371/journal.pgen.1007889](https://doi.org/10.1371/journal.pgen.1007889) · PMID: [30668570](https://pubmed.ncbi.nlm.nih.gov/30668570/) · PMCID: [PMC6358100](https://pubmed.ncbi.nlm.nih.gov/PMC6358100/)
15. **Integrative approaches for large-scale transcriptome-wide association studies**  
Alexander Gusev, Arthur Ko, Huwenbo Shi, Gaurav Bhatia, Wonil Chung, Brenda WJH Penninx, Rick Jansen, Eco JC de Geus, Dorret I Boomsma, Fred A Wright, ... Bogdan Pasaniuc  
*Nature Genetics* (2016-02-08) <https://doi.org/f3vf4p>  
DOI: [10.1038/ng.3506](https://doi.org/10.1038/ng.3506) · PMID: [26854917](https://pubmed.ncbi.nlm.nih.gov/26854917/) · PMCID: [PMC4767558](https://pubmed.ncbi.nlm.nih.gov/PMC4767558/)
16. **Opportunities and challenges for transcriptome-wide association studies**  
Michael Wainberg, Nasa Sinnott-Armstrong, Nicholas Mancuso, Alvaro N Barbeira, David A Knowles, David Golan, Raili Ermel, Arno Ruusalepp, Thomas Quertermous, Ke Hao, ... Anshul Kundaje  
*Nature Genetics* (2019-03-29) <https://doi.org/gf3hmr>  
DOI: [10.1038/s41588-019-0385-z](https://doi.org/10.1038/s41588-019-0385-z) · PMID: [30926968](https://pubmed.ncbi.nlm.nih.gov/30926968/) · PMCID: [PMC6777347](https://pubmed.ncbi.nlm.nih.gov/PMC6777347/)
17. **PhenomeXcan: Mapping the genome to the phenome through the transcriptome**  
Milton Pividori, Padma S Rajagopal, Alvaro Barbeira, Yanyu Liang, Owen Melia, Lisa Bastarache, YoSon Park, GTEx Consortium, Xiaoquan Wen, Hae K Im  
*Science Advances* (2020-09-11) <https://doi.org/ghbvbf>  
DOI: [10.1126/sciadv.aba2083](https://doi.org/10.1126/sciadv.aba2083) · PMID: [32917697](https://pubmed.ncbi.nlm.nih.gov/32917697/) · PMCID: [PMC11206444](https://pubmed.ncbi.nlm.nih.gov/PMC11206444/)

18. **A biclustering algorithm for extracting bit-patterns from binary datasets**  
Domingo S Rodriguez-Baena, Antonio J Perez-Pulido, Jesus S Aguilar-Ruiz  
*Bioinformatics* (2011-08-08) <https://doi.org/bm79fw>  
DOI: [10.1093/bioinformatics/btr464](https://doi.org/10.1093/bioinformatics/btr464) · PMID: [21824973](https://pubmed.ncbi.nlm.nih.gov/21824973/)
19. **UK Biobank**  
Neale lab  
<http://www.nealelab.is/uk-biobank>
20. **Shared and distinct genetic risk factors for childhood-onset and adult-onset asthma: genome-wide and transcriptome-wide studies**  
Milton Pividori, Nathan Schoettler, Dan L Nicolae, Carole Ober, Hae Kyung Im  
*The Lancet Respiratory Medicine* (2019-06) <https://doi.org/gmkmngk>  
DOI: [10.1016/s2213-2600\(19\)30055-4](https://doi.org/10.1016/s2213-2600(19)30055-4) · PMID: [31036433](https://pubmed.ncbi.nlm.nih.gov/31036433/) · PMCID: [PMC6534440](https://pubmed.ncbi.nlm.nih.gov/PMC6534440/)
21. **Effect of 17q21 Variants and Smoking Exposure in Early-Onset Asthma**  
Emmanuelle Bouzigon, Eve Corda, Hugues Aschard, Marie-Hélène Dizier, Anne Boland, Jean Bousquet, Nicolas Chateigner, Frédéric Gormand, Jocelyne Just, Nicole Le Moual, ... Florence Demenais  
*New England Journal of Medicine* (2008-11-06) <https://doi.org/df3bhg>  
DOI: [10.1056/nejmoa0806604](https://doi.org/10.1056/nejmoa0806604) · PMID: [18923164](https://pubmed.ncbi.nlm.nih.gov/18923164/)
22. **17q21 gene variation is not associated with asthma in adulthood**  
E Kreiner-Møller, DP Strachan, A Linneberg, LLN Husemoen, H Bisgaard, K Bønnelykke  
*Allergy* (2014-12-10) <https://doi.org/f6sh85>  
DOI: [10.1111/all.12537](https://doi.org/10.1111/all.12537) · PMID: [25331618](https://pubmed.ncbi.nlm.nih.gov/25331618/)
23. **RNA Interference Targeting CUG Repeats in a Mouse Model of Myotonic Dystrophy**  
Krzysztof Sobczak, Thurman M Wheeler, Wenli Wang, Charles A Thornton  
*Molecular Therapy* (2013-02) <https://doi.org/f4pd3m>  
DOI: [10.1038/mt.2012.222](https://doi.org/10.1038/mt.2012.222) · PMID: [23183533](https://pubmed.ncbi.nlm.nih.gov/23183533/) · PMCID: [PMC3594017](https://pubmed.ncbi.nlm.nih.gov/PMC3594017/)
24. **Review of hypertensive retinopathy**  
Amir Harjasouliha, Veena Raiji, Jose Maria Garcia Gonzalez  
*Disease-a-Month* (2017-03) <https://doi.org/gqhb4x>  
DOI: [10.1016/j.disamonth.2016.10.002](https://doi.org/10.1016/j.disamonth.2016.10.002) · PMID: [27931746](https://pubmed.ncbi.nlm.nih.gov/27931746/)
25. **Novel locus including FGF21 is associated with dietary macronutrient intake**  
Audrey Y Chu, Tsegaselassie Workalemahu, Nina P Paynter, Lynda M Rose, Franco Giulianini, Toshiko Tanaka, Julius S Ngwa, Qibin Qi, Gary C Curhan, Eric B Rimm, ... Lu Qi  
*Human Molecular Genetics* (2013-01-30) <https://doi.org/f4tv5t>  
DOI: [10.1093/hmg/ddt032](https://doi.org/10.1093/hmg/ddt032) · PMID: [23372041](https://pubmed.ncbi.nlm.nih.gov/23372041/) · PMCID: [PMC3612009](https://pubmed.ncbi.nlm.nih.gov/PMC3612009/)
26. **Probabilistic Colocalization of Genetic Variants from Complex and Molecular Traits: Promise and Limitations**  
Abhay Hukku, Milton Pividori, Francesca Luca, Roger Pique-Regi, Hae Kyung Im, Xiaoquan Wen  
*Cold Spring Harbor Laboratory* (2020-07-01) <https://doi.org/gjwh3p>  
DOI: [10.1101/2020.07.01.182097](https://doi.org/10.1101/2020.07.01.182097)
27. **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](https://doi.org/10.1126/science.aaz1776) · PMID: [32913098](https://pubmed.ncbi.nlm.nih.gov/32913098/) · PMCID: [PMC7737656](https://pubmed.ncbi.nlm.nih.gov/PMC7737656/)