

OPEN ACCESS

EDITED BY
Xiaoling Xuei,
Indiana University School of Medicine, United States

REVIEWED BY
Fazil Aliev,
Rutgers University, Newark, United States
Leah Wetherill,
Indiana University, United States

*CORRESPONDENCE
Erich Baker
✉ erich.baker@belmont.edu

RECEIVED 06 February 2025
ACCEPTED 02 July 2025
PUBLISHED 22 July 2025

CITATION

Castaneda E, Chesler E and Baker E (2025)
Spectral divergence prioritizes key classes,
genes, and pathways shared between
substance use disorders and cardiovascular
disease. *Front. Neurosci.* 19:1572243.
doi: 10.3389/fnins.2025.1572243

COPYRIGHT

© 2025 Castaneda, Chesler and Baker. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Spectral divergence prioritizes key classes, genes, and pathways shared between substance use disorders and cardiovascular disease

Everest Castaneda^{1,2}, Elissa Chesler³ and Erich Baker^{4*}

¹Department of Biology, Baylor University, Waco, TX, United States, ²School of Engineering and Computer Science, Baylor University, Waco, TX, United States, ³The Jackson Laboratory, Bar Harbor, ME, United States, ⁴Department of Mathematics and Computer Science, Belmont University, Nashville, TN, United States

Introduction: Substance use disorders (SUDs) are heterogeneous diseases with overlapping biological mechanisms and often present with co-occurring disease, such as cardiovascular disease (CVD). Gene networks associated with SUDs also implicate additional biological pathways and may be used to stratify disease subtypes. Node and edge arrangements within gene networks impact comparisons between classes of disease, and connectivity metrics, such as those focused on degrees, betweenness, and centrality, do not yield sufficient discernment of disease network classification. Comparatively, the graph spectrum's use of comprehensive information facilitates hypothesis testing and inter-disease clustering by using a larger range of graph characteristics. By adding a connectivity-based method, network rankings of similarity and relationships are explored between classes of SUDs and CVD.

Methods: Graph spectral clustering's utility is evaluated relative to commonly used network algorithms for discernment between two distinct co-occurring disorders and capacity to rank pathways based on their distinctiveness. A collection of graphs' structures and connectivity to functionally identify the relationship between CVD and each of four classes of SUDs, namely alcohol use disorder (AUD), cocaine use disorder (CUD), nicotine use disorder (NUD), and opioid use disorder (OUD) is evaluated. Moreover, a Kullback-Leibler (KL) divergence is implemented to identify maximally distinctive genes (D^g). The emphasis of genes with high D^g enables a Jaccard similarity ranking of pathway distinctiveness, creating a functional "network fingerprint".

Results: Spectral graph outperforms other connectivity-based approaches and reveals interesting observations about the relationship among SUDs. Between CUD and CVD, the gamma-aminobutyric acidergic and arginine metabolism pathways are distinctive. The neurodegenerative prion disease and tyrosine metabolism are emphasized between OUD and CVD. The graph spectrum between AUD and NUD to CVD is not significantly divergent.

Conclusion: Graph spectral clustering with KL divergence illustrates differences among SUDs with respect to their relationship to CVD, suggesting that despite a high-level co-occurring diagnosis or comorbidity, the nature of the relationship between SUD and CVD varies depending on the substance involved. The graph clustering method simultaneously provides insight into the specific biological

pathways underlying these distinctions and may reveal future basic and clinical research avenues into addressing the cardiovascular sequelae of SUD.

KEYWORDS

disease-associated prioritization, substance use disorder, cardiovascular disease, graph spectrum, functional fingerprint

1 Introduction

Complex diseases are caused by a variety of factors and include a range of psychological and physical disorders such as diabetes (Prasad and Groop, 2015), schizophrenia (Sullivan et al., 2003), substance use disorders (Hatoum et al., 2023) (SUDs), cardiovascular disease (CVD) (Musunuru and Kathiresan, 2019), and others (Thaker, 2017; Andrews et al., 2023). These are often co-occurring conditions (Hossain et al., 2020; Solovieff et al., 2013), and due to the many-to-many relationships among genes and disorders (Goh et al., 2007), identifying the specific biological basis of these relationships presents challenges (Wormington et al., 2024). Doing so would enable a refined classification of the particular subtypes of disease that exist and also provide a greater understanding of the nature and mechanism of comorbid disease (Chen et al., 2025; Sánchez-Valle and Valencia, 2023). For example, in various SUDs, there are associations to CVD, but for each drug, the nature of this relationship may differ (Havakuk et al., 2017; Toska and Mayrovitz, 2023; Pando-Naude et al., 2021). There are several approaches to comparing genetic studies to elucidate the nature and extent of these relationships among complex diseases (Gerring et al., 2024). For example, shared genetic liability between SUD and CVD has been found using polygenic risk scores and linkage disequilibrium, but even with shared multimorbid association (Zhou et al., 2017), the shared functional mechanisms are poorly understood (Morgan et al., 2022).

CVD is a leading cause of death and a common multimorbid and comorbid condition, with high prevalence in people with SUD (Chelikam et al., 2022). While the impacts of SUD and CVD concentrate in different tissues, they share similar genetic associations (Hatoum et al., 2023; Sanchez-Roige et al., 2022). Furthermore, the tremendous number of genetic variants impacting the function of the nervous system and heart (Jonker et al., 2023) presents challenges in prioritization of disease-associated genes (Zhukovsky et al., 2024; Guo et al., 2021). A functional enrichment provides foundational interpretation of variant effects at the level of cellular and metabolic processing underlying disease genes (Reimand et al., 2019). Furthermore, the shared risk factors between psychiatric disorders is such that a focus on specific disorders, independent of the context of comorbid and multimorbid conditions, is insufficient for classification (Chen et al., 2025). To compare disorders on a functional level, pathways have been assessed by the intersection of genes, such as network merging for “network fingerprinting”, which has shown that the arrangement of nodes and edges impacts comparisons on similarity scores (Cui et al., 2015). Moreover, the availability of data on a large number of SUDs (Bough and Pollock, 2018; Hatoum et al., 2023; Uffelmann et al., 2021) enables an assessment of the influence between SUD specific functional pathways and a common

multimorbid condition, CVD (Minhas et al., 2024). Comparisons between SUDs and cardiometabolic disease provides insight into shared genes, which are highly translatable to therapeutic potential (Sanchez-Roige et al., 2022; Peng et al., 2021).

Investigating and integrating genomic studies of disease can improve disease diagnosis and characterization (Wirka et al., 2018). From genome-wide associations (GWAS) (Uffelmann et al., 2021) to curated database mining (Piñero et al., 2020), discrete experimental investigations of disorders often converge to a functional classification (Reimand et al., 2019). Enrichment software then gauges biological pathways or functional terms that have, more than by random chance, a significant representation (Kuleshov et al., 2016; Wang et al., 2017; Raudvere et al., 2019; Reimand et al., 2019). A functional characterization may focus on a set of a gene’s medicinal, cellular, or biological significance (Wieder et al., 2021; Baltoumas et al., 2021). Several databases are used as a proxy for functional analyses, which include the following: the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000; Kanehisa, 2019; Kanehisa et al., 2023), WikiPath (Agrawal et al., 2023), Gene Ontology (GO) (Ashburner et al., 2000; Consortium et al., 2023), and Reactome (Milacic et al., 2023), among others (Geistlinger et al., 2021; Zhao and Rhee, 2023).

Spectral graph analysis presents a promising approach to simultaneously compare disease based on the various genomic data sources and to identify the biomolecular pathways that can be used to classify them. Graph spectrum has been used for hypothesis testing (Takahashi et al., 2012; Fujita et al., 2017), differentiating diseases and tissues (Santos et al., 2015; Jardim et al., 2019), identifying functional pathways (Fujita et al., 2017), and clustering (Sato et al., 2013) in neurological disorders. The spectrum of a graph contains information on several important dynamics such as number of walks, diameter, and cliques (Takahashi et al., 2012); therefore, the spectrum is more informative to characterizing complex networks than modern metrics (Fujita et al., 2017). Contemporary approaches have classified genes by importance through nodal connectivity evaluations and pinpointed dysfunction or distinguished biological conditions (Barabási and Oltvai, 2004; Gu et al., 2012; Rahmatallah et al., 2013; Santos et al., 2015).

Here, we present a functional analysis of five complex diseases describing the genes underlying 4 SUDs and CVD, a commonly co-occurring condition (Minhas et al., 2024). Moreover, we attempt to elucidate key differences in classes of SUDs and their associations to CVD, resulting in the prioritization of key disease-genes and pathways through graph spectral clustering, which relies on graph structure is and not limited to genomic intersections. We conduct a graph spectral analysis of SUD and CVD related genomic studies with functional KEGG pathways using the statGraph (Castro Guzman et al., 2024) package and posit a comparative insight

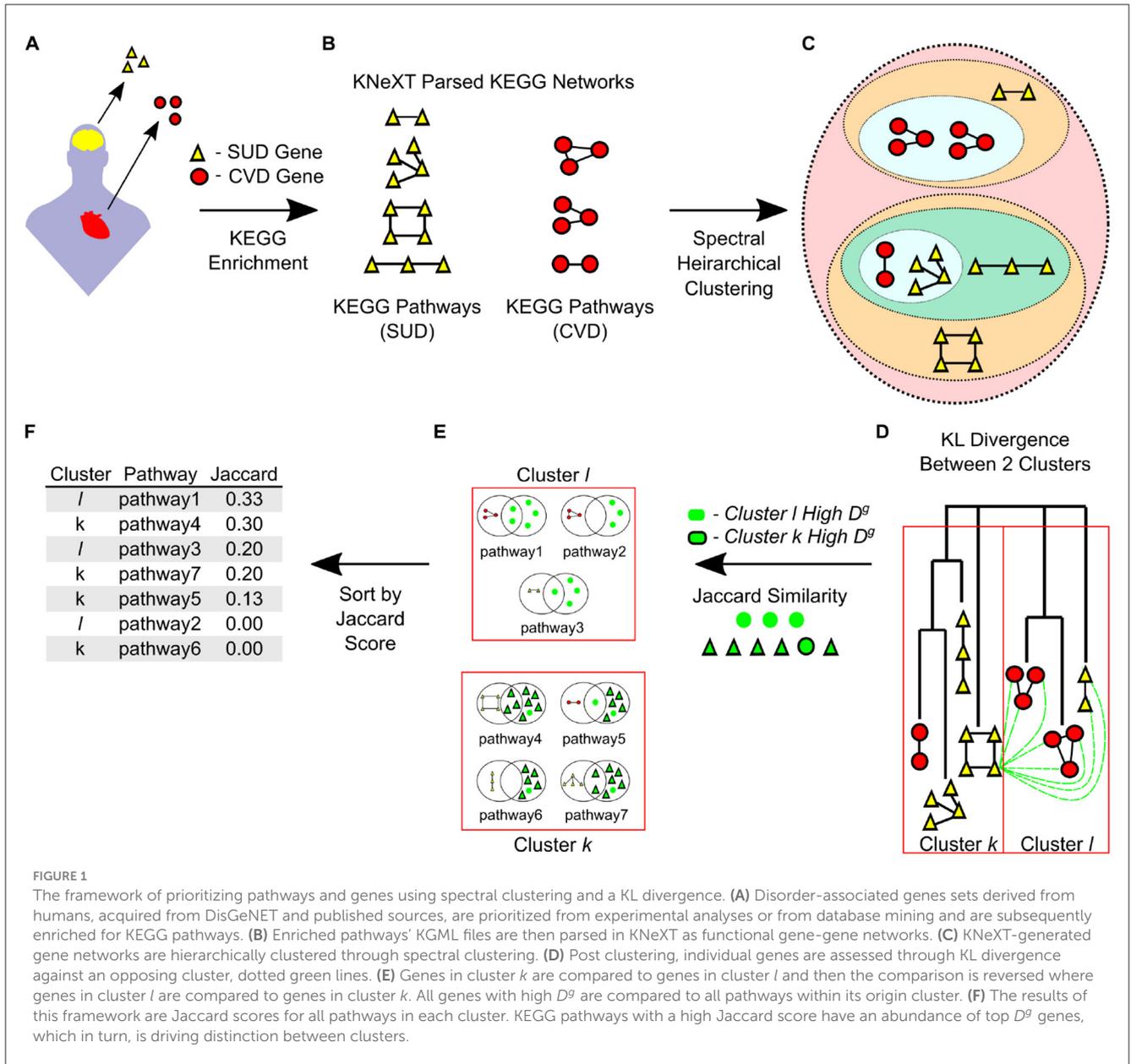


FIGURE 1

The framework of prioritizing pathways and genes using spectral clustering and a KL divergence. (A) Disorder-associated genes sets derived from humans, acquired from DisGeNET and published sources, are prioritized from experimental analyses or from database mining and are subsequently enriched for KEGG pathways. (B) Enriched pathways' KEGG files are then parsed in KNeXT as functional gene-gene networks. (C) KNeXT-generated gene networks are hierarchically clustered through spectral clustering. (D) Post clustering, individual genes are assessed through KL divergence against an opposing cluster, dotted green lines. (E) Genes in cluster *l* are compared to genes in cluster *k* and then the comparison is reversed where genes in cluster *k* are compared to genes in cluster *l*. All genes with high D^g are compared to all pathways within its origin cluster. (F) The results of this framework are Jaccard scores for all pathways in each cluster. KEGG pathways with a high Jaccard score have an abundance of top D^g genes, which in turn, is driving distinction between clusters.

against other connectivity indices (Fujita et al., 2017) using high fidelity pathways created by KNeXT (Castaneda and Baker, 2024). By leveraging the topological information, the arrangement of nodes and edges, offered by KEGG graphs, we demonstrate that the spectral distribution aids in defining key divergences, or its absence, between four of the following SUDs: AUD, CUD, OUD, and NUD with its comorbidity, CVD. Moreover, we elucidated the biological relevance of principle divergent-driving elements and outline functional class differences between CUD and OUD.

2 Materials and methods

2.1 Datasets

Homo sapiens gene sets were gathered from publicly available repositories and published sources (Figure 1A). Data were collected

from DisGeNET (Piñero et al., 2020) using the Harmonizome web service (Rouillard et al., 2016). DisGeNET has been widely used as a benchmark database (Barua et al., 2022; Gentili et al., 2022) due to its comprehensive and curated information on gene-disease associations (Piñero et al., 2020). For each SUD, we used affiliated search terms, noting that the current terminology for CUD has evolved but is not always used in existing data repositories (Deak and Johnson, 2021). These terms include the following: “alcohol use disorder” for AUD, “cocaine abuse” [sic] for CUD, “nicotine dependence” for NUD, and “cardiovascular pathology” for CVD. We combined “heroin dependence”, “opioid use disorder”, and “morphine dependence” for our OUD gene set, see Supplementary Table S1 for full list. Genes were then uploaded to g:Profiler, ignoring ambiguous gene queries, which was set to a g:SCS (Set Count and Sizes) threshold of 0.05. Published KEGG CVD pathways were gathered from (Barua et al., 2022), which used

Gene Expression Omnibus microarray datasets for assessing CVD and all its risk factors. Neurological KEGG pathways were acquired using the following Biological Relation Inference and Classification Engine (BRITE) terms: *Nervous system*, *Substance dependence*, and *Neurodegenerative diseases*. Additionally, we included the pathway *Neuroactive ligand-receptor interaction*. All neurological pathways derived from BRITE terms were used for systematically characterizing synapses across regions of the brain (Bar-Shira et al., 2015).

2.2 KEGG networks

Here, we focus on KEGG's database which hosts a series of biological systems maps that offer specific molecular pathways based on highly curated and experimentally verified gene-gene, gene-compound, and gene-pathway interactions (Kanehisa and Goto, 2000; Kanehisa, 2019; Kanehisa et al., 2023). KEGG is an important resource because all molecular associations are stored for secondary parsing and analyzing in a standard language: KEGG Markup Language (KGML). KGML files can be readily parsed and used with robust software packages including the following: KEGG NetworkX Topological (KNeXT) Parser (Castaneda and Baker, 2024), graphite (Sales et al., 2012), KEGGParser (Arakelyan and Nersisyan, 2013), among others (Nersisyan et al., 2014; Chanumolu et al., 2021). KNeXT, in particular, focuses on the spatiotemporal dynamics reflected in a KGML file to create high fidelity pathways (Castaneda and Baker, 2024). All KEGG pathways were parsed using the KNeXT parser, see Figure 1B. KNeXT creates high fidelity genes-only pathways (Castaneda and Baker, 2024). For simplicity, all pathways used NCBI gene identifiers, contained no compounds, and are unweighted and undirected. For full list of all KEGG pathways, see Supplementary Table S2.

2.3 Spectral discrimination

In this study, we used the statGraph (Castro Guzman et al., 2024) package in R version 4.3.1 (R Core Team, 2023). statGraph features several tests for conducting spectral analyses of graph lists. From this package, we used the Analysis Of Graph Variability (*anogva*), Takahashi Test (*takahashi.test*), and heirarchical clustering (*hclust*). *anogva* performs a statistical test on a set of two or more graphs to determine if they are generated by the same random process (Fujita et al., 2017). *takahashi.test* conducts a statistical test to determine if two sets of graphs are generated by the same random process (Takahashi et al., 2012; Fujita et al., 2017). All tests used a seed set at one. *hclust* conducts a hierarchical clustering of a list of graphs based on their spectral distribution (Figures 1C, D). We used default parameters, which include *complete* agglomerative clustering method with Silverman bandwidth and exact spectral density.

2.4 Algorithmic comparisons

For baseline comparisons, we used common indices, which included the following: degree, average betweenness centrality,

and closeness centrality (Fujita et al., 2017; Zito et al., 2021). Implementation of these metrics was through NetworkX (Hagberg et al., 2008). We used the Jensen-Shannon (JS) distance in the SciPy version 1.5.0 package (Virtanen et al., 2020) to create distance matrices for input into the *AgglomerativeClustering* function in the Scikit-learn package (Pedregosa et al., 2011). The same parameters to the *hclust* package were used with *complete* linkage.

2.5 Statistical comparisons

Statistical comparisons for graph performance was measured using the Adjusted Rand Index (ARI) (Warrens and van der Hoef, 2022). For R analyses, we used the *fossil* package version 0.4.0 (Vavrek, 2011), and for Python analyses, we used the *Scikit-learn* package (Pedregosa et al., 2011). Both metrics measure the accuracy of clustering with ARI being adjusted for randomness. An $ARI \leq 0$ is equivalent to random assignments (Yeung et al., 2003). ARI has been used for comparisons of clustering performance in previous works (Wu and Wu, 2020; Zelig et al., 2023).

2.6 Gene and pathway prioritization

To prioritize genes and pathways in biological clusters, we modified a method developed by Dey et al. (2017), which uses KL divergence to compare the distinctiveness of a gene, g , with respect to any cluster l see Equation 1. Here, we used the entropy function in the *Scikit-learn* package (Pedregosa et al., 2011).

$$KL^g[k, l] = \sum_{x \in X} p^{S_k}(x) \times \log \frac{p^{S_k}(x)}{p^{S_l}(x)} \quad (1)$$

Let $S = \{S_1, S_2, S_3, \dots, S_n\}$ be a collection of KEGG pathways in cluster k with vertex set, X . p is the degree distribution for any X in any pathway in S compared to any X in any pathway in cluster l . Thereby, for each cluster k , we measure the distinctiveness of each gene as the minimum divergence (Equation 2).

$$D^g[k] = \min_{l \neq k} KL^g[k, l] \quad (2)$$

Thereby, genes with a maximum distinctiveness (D^g) are the genes with the largest role in distinguishing cluster k from cluster l . After identifying the genes with the highest D^g we use a Jaccard index to determine the set similarity (Equation 3). D_k is the set of genes with the highest D^g in cluster k and S_k is some pathway in cluster k , see Figures 1D–F.

$$J(D_k, S_k) = \frac{|D_k \cap S_k|}{|D_k \cup S_k|} \quad (3)$$

3 Results

3.1 Comparisons of KEGG enriched pathway from DisGeNET derived gene sets

Our first analysis was to determine the significance of the divergence between the four SUDs and CVD KEGG pathway

TABLE 1 Results of test for the JS divergence between groups.

Source	Comparison	JS	<i>p</i>
DisGeNET	All ^a	< 0.001	< 0.001
	AUD vs CVD	0.016	0.39
	CUD vs CVD	0.027	0.005
	NUD vs CVD	0.030	0.12
	OUD vs CVD	0.021	0.036
Benchmarked	Brain ^b vs. CVD ^c	0.022	0.038

^aAll refers to AUD vs CUD vs CVD vs NUD vs OUD. ^bPublished in Bar-Shira et al. (2015). ^cPublished in Barua et al. (2022). Benchmarked: KEGG pathways used in published studies. Bolded values emphasize statistical significance.

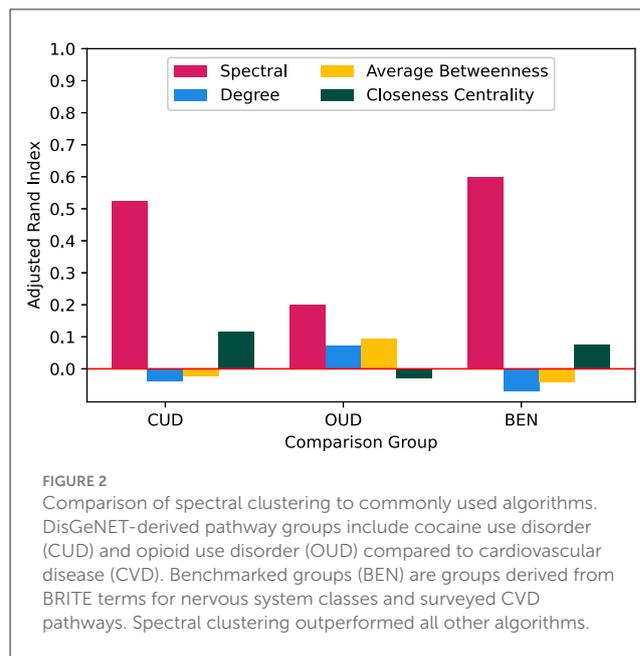
lists. Here, divergence refers to the disparate random processes underlying a collection of graphs as defined by Takahashi et al. (2012); Fujita et al. (2017). We compared each KEGG pathway list derived from highly supported DisGeNET (Piñero et al., 2020) gene sets. First, we used *anogva* to test the spectrum of all data. *anogva* controls for Type I errors and is robust for unbalanced data (Fujita et al., 2017). The results showed a significant JS divergence between all five sets of graphs, see Table 1. Moreover, Takahashi's Test revealed a significant JS divergence between CUD and OUD against CVD but not for AUD and NUD, see Table 1, which parallels past epidemiological studies signifying the strong relationship between AUD, NUD, and CVD (Yeates et al., 2015). Post Takahashi's Test, we then conducted a hierarchical agglomerative clustering on the significant SUDs. Clustering quality was determined by the ARI against commonly used algorithms (Fujita et al., 2017). CUD had the highest ARI and outperformed all other connectivity metrics, and OUD generated the highest ARI compared to baseline measures (Figure 2).

3.2 Comparisons of benchmarked KEGG pathways

From the perspective of KEGG pathway enrichment, there are several factors that may generate inconsistent pathway inclusions (Mubeen et al., 2022). In order to show the inherent functional divergence between neurobiologically-derived gene sets and CVD, we conducted an analysis on what we term "benchmarking" KEGG pathways. Pathways are retrieved from published surveys or extensive analyses which focused on KEGG function and require no enrichment profiling. Additionally, these pathways have been used or compared to benchmark data (Bar-Shira et al., 2015; Barua et al., 2022). Takahashi's test illustrates that KEGG pathways involved in the brain are significantly divergent from KEGG pathways involved in CVD, see Table 1. Furthermore, benchmarked KEGG pathways scored the highest ARI of all other pathways (Figure 2). Hence, evidence suggests that the sub-network of KEGG pathways involved in the brain are structurally different when compared to CVD, which is captured by the graph's spectrum.

3.3 Analysis of top driving genes

For OUD and CUD, the two SUDs that significantly differed in divergence, we used the highest $D^{\mathcal{S}}$ (Dey et al.,



2017) of each cluster to create gene sets which in turn were used to rank each pathway by their composition using Jaccard similarity. The results of an agglomerative clustering between CUD and CVD illustrate two pathways, hsa00220 (arginine biosynthesis) and hsa00330 (arginine and proline metabolism), are divergent from the rest of the maps, see Figure 3. The top driving genes created gene sets that had the highest similarity with arginine biosynthesis in cluster one and gamma aminobutyric acid (GABAergic) synapse in cluster two (Figure 3).

From the top driving genes in OUD vs CVD, the pathways with the highest $D^{\mathcal{S}}$ defined tyrosine metabolism and prion disease (Figure 4). In addition, the top driving genes belonging to the GABAergic synapse in CUD are similarly reflected in the benchmarked gene sets with Glutamatergic synapse, another organismal/nervous system pathway, being highly divergent (Supplementary Figures S1, S2). Given that NF-kappa B signaling pathway was a top driving pathway whose class is not similar to any SUD, this implies an inherent distinctiveness between classes of SUD and the brain in their comparison to CVD (Supplementary Figures S1, S2).

3.4 Functional fingerprint

We illustrate a "network fingerprint", as described in Cui et al. (2015) between SUDs and CVD (Figure 5). CUD and OUD, Figures 5A, B, respectively, differ based on the pathways included in each cluster. Furthermore, the division is solely in metabolic pathways in CUD, Figure 5A, while OUD differs in both human diseases and metabolism (Figure 5B). A comprehensive difference was generated between CUD and OUD (Figure 5C). An interesting aspect is the role metabolism plays in both CUD and OUD's largest magnitude of difference, see Figure 5C.

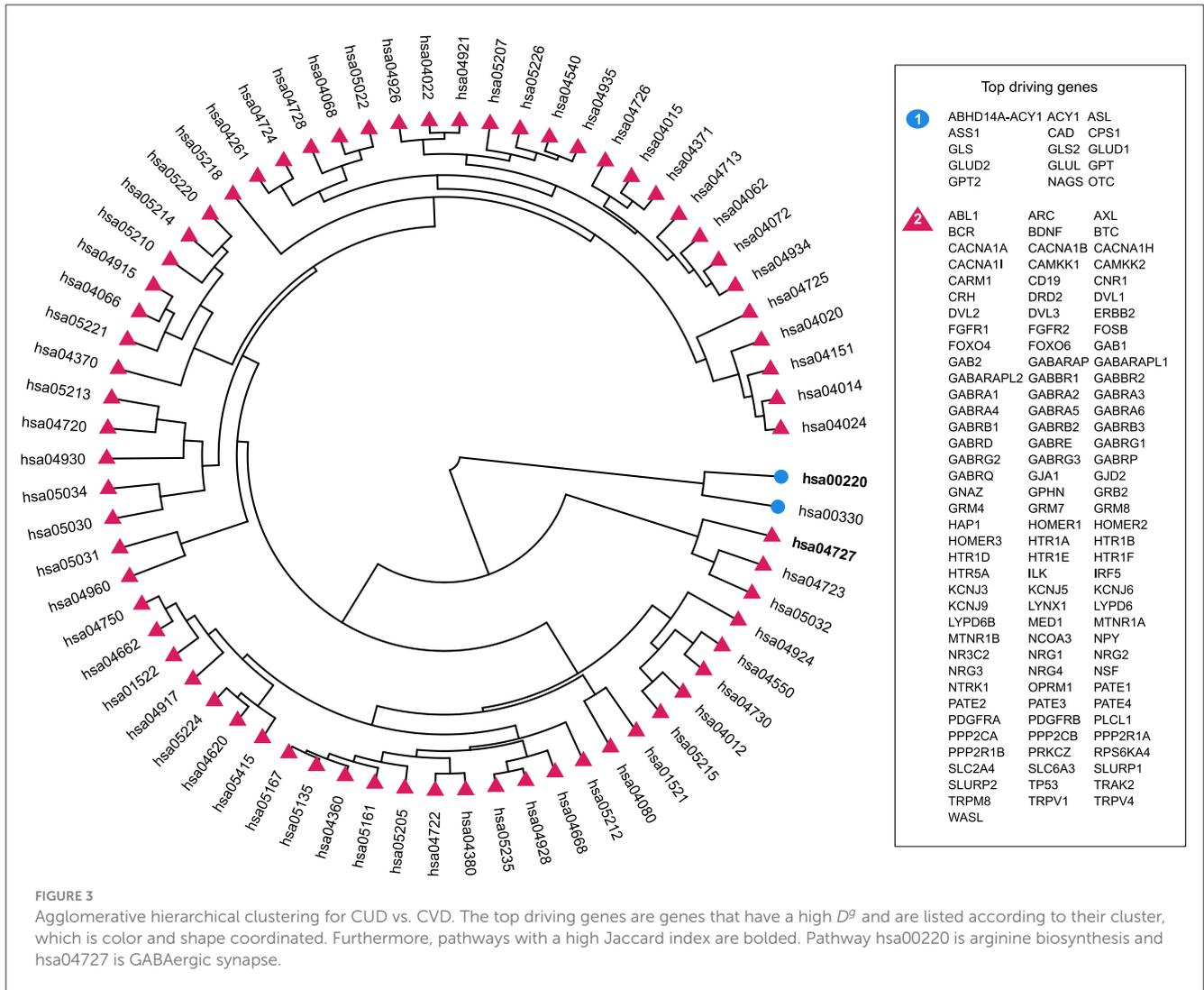


FIGURE 3

Agglomerative hierarchical clustering for CUD vs. CVD. The top driving genes are genes that have a high D^g and are listed according to their cluster, which is color and shape coordinated. Furthermore, pathways with a high Jaccard index are bolded. Pathway hsa00220 is arginine biosynthesis and hsa04727 is GABAergic synapse.

4 Discussion

Here, we demonstrated the utility of graph spectral clustering for differentiating between the bases of comorbidity of disease. The technique outlined reveals the effectiveness of assessing the distinctions between classes of SUDs and a commonly co-occurring disease, CVD. We have shown this method to outperform other commonly used algorithms in classifying KEGG pathways derived from SUDs and CVD gene sets. Furthermore, we leveraged spectral clustering to rank pathways according to their distinctiveness between the two conditions, revealing a “network fingerprint” comparison, similarity described in Cui et al. (2015). Our analysis pipeline finds differences among disorders and identifies key pathways, which may have therapeutic or diagnostic consequences. The facilitation of a “network fingerprint” diagram aids in hypothesis building and identifying key functional pathways. From these clusters, we can characterize the associations between two diseases, which are unmatched by KEGG pathway lists alone and other topology-based methods.

KEGG is an important tool for disease study from a functional perspective (Kanehisa and Goto, 2000; Kanehisa, 2019; Kanehisa

et al., 2023). Surveys of KEGG pathways and disease interactions infer novel association of overlapping risk factors (Barua et al., 2022) and common disorders, (Li et al., 2008) given sets of prioritized genes (Cirincione et al., 2018). Moreover, other analyses rank and prioritize pathways by significance levels (Chu et al., 2024). While these approaches are useful to generate new insights in highly related diseases that have strong pathway sharing within tissues, they do not provide adequate discernment for two convergent disorders involving different tissues, which share common and distinct biomarkers (Moon et al., 2025; Riley et al., 2022; Gu et al., 2021; Daneshafrooz et al., 2022). Here, we have emphasized pathways through graph connectivity, which does not rely merely on pathway member composition. Hence, we propose future usage for comparisons between disorders that exist in different tissues and experience limited functional overlap, such as certain comparisons of fibrotic disease (Gu et al., 2021) and frontotemporal dementia (FTP) and amyotrophic lateral sclerosis (ALS) (Daneshafrooz et al., 2022). While epidemiological studies demonstrate a strong comorbidity between SUDs, defined here as AUD, NUD, CUD, and OUD, and CVD (Gan et al., 2021; Chelikam et al., 2022), our approach distinguishes the more granular separation of SUDs and CVD based on their KEGG

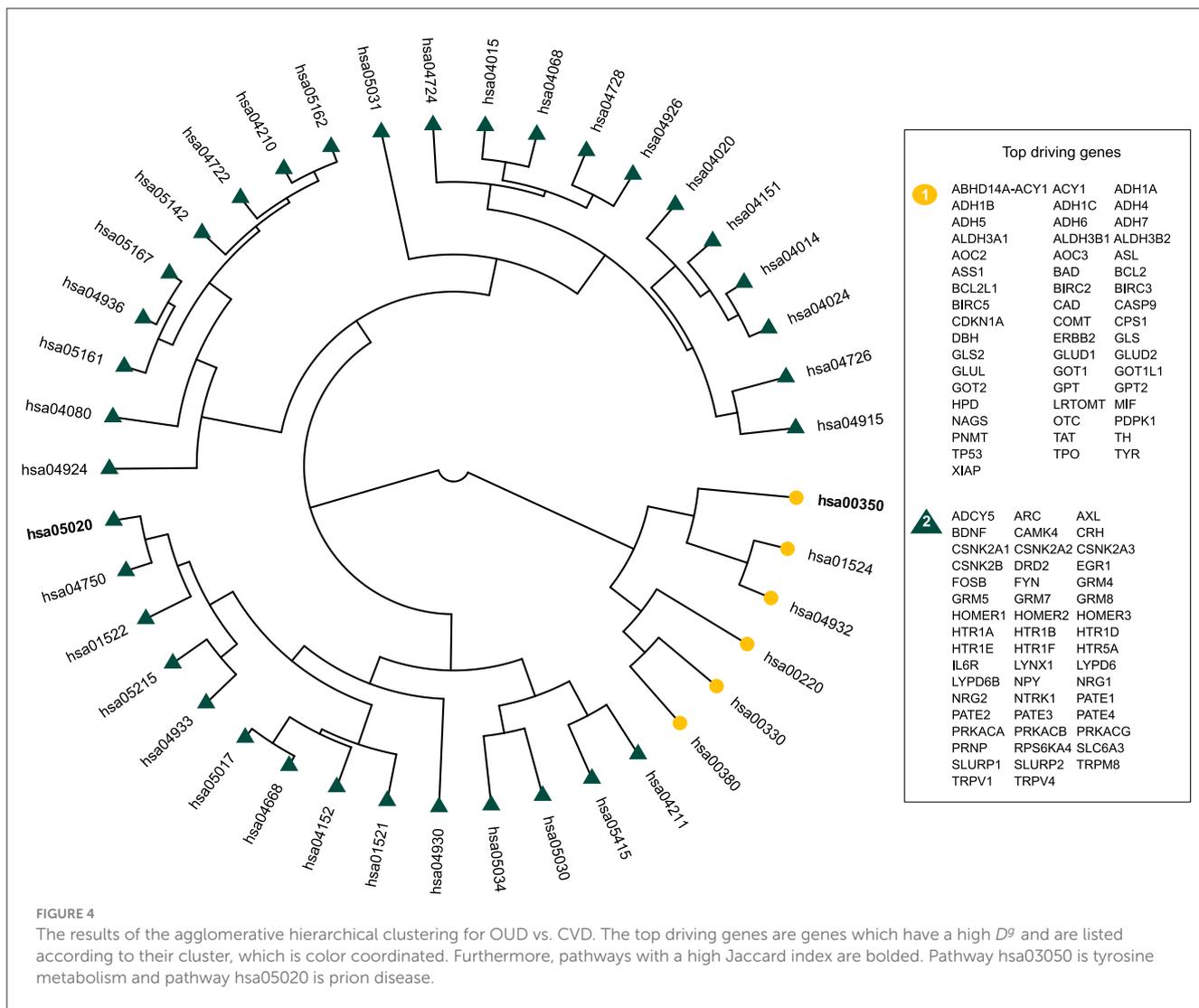


FIGURE 4

The results of the agglomerative hierarchical clustering for OUD vs. CVD. The top driving genes are genes which have a high D^g and are listed according to their cluster, which is color coordinated. Furthermore, pathways with a high Jaccard index are bolded. Pathway hsa03050 is tyrosine metabolism and pathway hsa05020 is prion disease.

pathway representations. We provide evidence to show that CUD and OUD are significantly divergent to CVD while AUD and NUD are not. This divergence may be explained by differences in how these substances interact with the cardiovascular system (Havakuk et al., 2017; Toska and Mayrovitz, 2023; Jalali et al., 2021) and areas of the brain (Pando-Naude et al., 2021). Additionally, we highlight the specific metabolic and neurological pathways and genes driving these distinctions. The profile of these clusters would be useful in disease state transition surveilling (Guo et al., 2021) and model organism testing. For example, knockouts of genes involved in these pathways may show insights for vulnerabilities to CVD for a given SUD (Cacheiro et al., 2023).

While the main focus of this work was the divergence created by each KEGG pathway, network merging is a crucial aspect of heterogeneous graph development where integrating and comparing graphs is essential (Chang et al., 2016; Zitnik et al., 2024). Moreover, existing software analyses have limited scalability on large data sets (Chang et al., 2016; Smedley et al., 2015). The technique examined here may be applicable for automated KEGG enrichment data set preprocessing, trimming, and curation (Orouji et al., 2024).

Biologically, a synergism exists between the representative genes from pathways with several high D^g genes. For example, tyrosine metabolism disruption (Rathor and Ch, 2023) and OUD is known to affect circadian rhythms (Puig et al., 2023). Neurodegeneration-related pathways are linked to OUD-mediated circadian rhythm disruption (Puig et al., 2023). The prioritized pathways might imply novel transitory genes that are implicated in circadian rhythm disruption as several glutaminergic synaptic signaling genes were prioritized alongside the implications of aromatic amino acid metabolism (Humer et al., 2020; Puig et al., 2023). Moreover, “network fingerprinting” (Cui et al., 2015) clustered prion disease with Type 2 diabetes and amyotrophic lateral sclerosis, which implicates neurodegeneration playing a role in complex diseases. In our CUD clusters, arginine has been studied for its role in CVD prevention and treatment (Tousoulis et al., 2007; Bahadoran et al., 2016). GABA plays a role in both CVD (Bu et al., 2021) and CUD (Wydra et al., 2024). GABA has shown promise as pharmacotherapy for addiction (Wydra et al., 2024), and accordingly, arginine has been shown to synaptically interact with GABA in the brain of rats (Shen et al., 1997). The synergism of these pathways and their divergent-driving genes might have

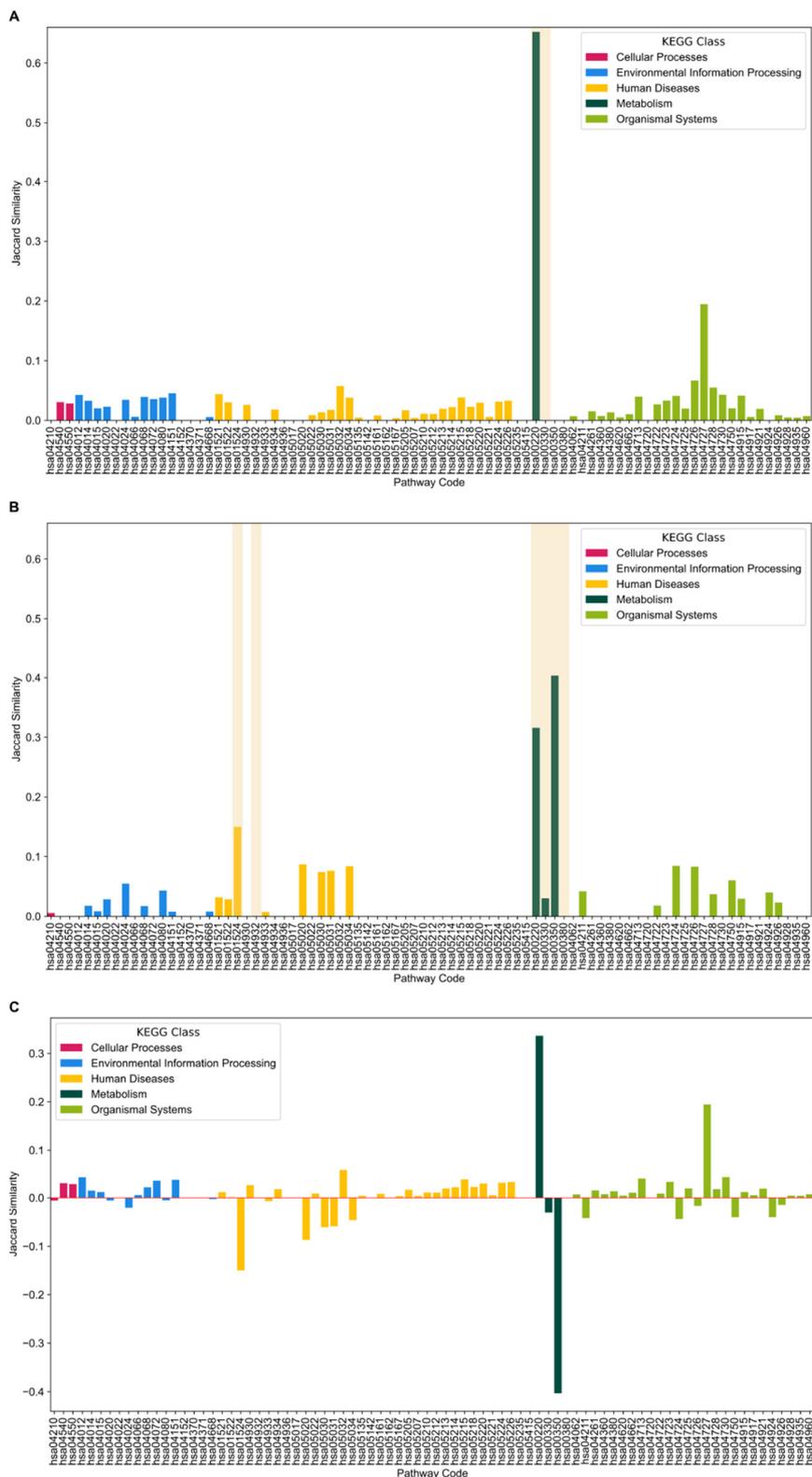


FIGURE 5

All similarity results sorted by KEGG class. Pathways that cluster separately are highlighted in yellow. **(A)** Similarity results for CUD. As shown, two metabolism pathways diverged compared to the rest of the results. **(B)** Similarity results for OUD. For OUD, metabolism and two human disease pathways drove the cluster separation. **(C)** The differences in magnitude of Jaccard similarity between CUD and OUD. As shown, metabolism plays the largest role in the differences between both CUD and OUD.

implications in co-occurring (Stoychev et al., 2021) CVD and CUD treatment and study (Wilson et al., 2001).

A limitation of this study is the statistical tests in statGraph do not account for multiple group memberships. Hence, diseases with high overlap of KEGG pathways will create difficulties in using the tools outlined in this analysis. An additional limitation of this study is the high redundancy of KEGG pathways (Karp et al., 2021), which creates issues in finding differences in topology, gene-gene connectivity, as suggested by low ARI scores. Moreover, several software exist for KEGG enrichment (Mubeen et al., 2022), and the pathway database itself may be biased to understudied genes, such as non-coding RNA genes (Li et al., 2022). Consequently, the topology and genomic composition of KEGG pathways are not comprehensive (Wilk and Braun, 2017; Gable et al., 2022). While a survey of all enrichment software and KEGG parsers is beyond the scope of this article, we note that use of different combinations of software and thresholds may produce varying results. Moreover, discrete combinations of search terms for disorders in DisGeNET may yield larger or smaller gene sets, which would render disparate amounts and combinations of KEGG pathways. Hence, the use of benchmarked data served solely to indicate an inherent divergence in pathways representing the brain (Bar-Shira et al., 2015) and pathways underlying a comprehensive study of CVD (Barua et al., 2022).

We have demonstrated how underexplored network features (Santos et al., 2015) may be employed to prioritize or differentiate disorders. In previous functional studies, SUDs are coalesced (Li et al., 2008), which overlooks underlying differences. We leveraged the divergence of the collection of KEGG graphs to prioritize genes that are implicated in driving the functional clustering between SUD and CVD. The pathway and genes prioritized are biologically relevant and might have implications for future studies in knockout or other experimental analyses. Additionally, the pathway and gene rankings could justify inclusion or exclusion in large-scale or heterogeneous network analyses of multiple disorder studies (Xiong et al., 2019; Gu et al., 2022). Furthermore, the magnitude of the pathway ranking differences decomposes the complexity of a collection of KEGG graphs, conferring critical visualization and processing where KEGG lists alone cannot provide.

The graph spectrum reveals a distinction among disorders that are co-occurring and can allow visualization of the relationships among multiple disorders simultaneously. Spectral clustering outperformed other commonly used algorithms in classifying clusters of a psychiatric disorder and a common multimorbidity or comorbidity in CVD, and thus its application to other comorbidities observed in SUDs, psychiatric disorders, and other complex disease is promising. Furthermore, the method can characterize the pathways that drive each cluster's distinction to reveal insights about their biological implications, potential diagnostic, and therapeutic targets. In contrast to many pathway overlap approaches that rely on data from disease that involve a limited tissue or cell population, the method introduced here has implications for identifying genes that drive co-morbid conditions in distinct diseases encompassing a diverse range of tissues and embodying systems networks that have little functional pathway overlap.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

ECa: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. EB: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing. ECh: Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2025.1572243/full#supplementary-material>

References

- Agrawal, A., Balc s, H., Hanspers, K., Coort, S. L., Martens, M., Slenter, D. N., et al. (2023). WikiPathways 2024: next generation pathway database. *Nucleic Acids Res.* 52, D679–D689. doi: 10.1093/nar/gkad960
- Andrews, S. J., Renton, A. E., Fulton-Howard, B., Podlesny-Drabiniok, A., Marcora, E., and Goate, A. M. (2023). The complex genetic architecture of Alzheimer's disease: novel insights and future directions. *eBioMedicine* 90:104511. doi: 10.1016/j.ebiom.2023.104511
- Arakelyan, A., and Nersisyan, L. (2013). KEGGParser: parsing and editing KEGG pathway maps in Matlab. *Bioinformatics* 29, 518–519. doi: 10.1093/bioinformatics/bts730
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556
- Bahadoran, Z., Mirmiran, P., Tahmasebinejad, Z., and Azizi, F. (2016). Dietary l-arginine intake and the incidence of coronary heart disease: Tehran lipid and glucose study. *Nutr Metab.* 13:23. doi: 10.1186/s12986-016-0084-z
- Baltoumas, F. A., Zafeiropoulou, S., Karatzas, E., Koutrouli, M., Thanati, F., Voutsadaki, K., et al. (2021). Biomolecule and bioentity interaction databases in systems biology: A comprehensive review. *Biomolecules* 11(8). doi: 10.3390/biom11081245
- Barab si, A.-L., and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5, 101–113. doi: 10.1038/nrg1272
- Bar-Shira, O., Maor, R., and Chechik, G. (2015). Gene expression switching of receptor subunits in human brain development. *PLoS Comput. Biol.* 11, 1–21. doi: 10.1371/journal.pcbi.1004559
- Barua, J. D., Omit, S. B. S., Rana, H. K., Podder, N. K., Chowdhury, U. N., and Rahman, M. H. (2022). Bioinformatics and system biological approaches for the identification of genetic risk factors in the progression of cardiovascular disease. *Cardiovasc. Ther.* 2022:9034996. doi: 10.1155/2022/9034996
- Bough, K. J., and Pollock, J. D. (2018). Defining substance use disorders: The need for peripheral biomarkers. *Trends Mol. Med.* 24, 109–120. doi: 10.1016/j.molmed.2017.12.009
- Bu, J., Huang, S., Wang, J., Xia, T., Liu, H., You, Y., et al. (2021). The gabaa receptor influences pressure overload-induced heart failure by modulating macrophages in mice. *Front. Immunol.* 12:670153. doi: 10.3389/fimmu.2021.670153
- Cacheiro, P., Spielmann, N., Mashhadi, H. H., Fuchs, H., Gailus-Durner, V., Smedley, D., et al. (2023). Knockout mice are an important tool for human monogenic heart disease studies. *Dis. Model. Mech.* 16:49770. doi: 10.1242/dmm.049770
- Castaneda, E. U., and Baker, E. J. (2024). Knext: a networkx-based topologically relevant kegg parser. *Front. Genet.* 15:1292394. doi: 10.3389/fgene.2024.1292394
- Castro Guzman, G. E., da Costa, D. R., Ramos, T. C., Santos, S. S., Lira, E. S., and Fujita, A. (2024). *statGraph: Statistical Methods for Graphs*. The Comprehensive R Archive Network; Institute of Mathematics and Statistics; University of S o Paulo.
- Chang, J., Cho, H., and Chou, H.-H. (2016). Mango: combining and analyzing heterogeneous biological networks. *BioData Min.* 9:25. doi: 10.1186/s13040-016-0105-5
- Chanumolu, S., Albahrani, M., Can, H., and Otu, H. (2021). KEGG2Net: Deducing gene interaction networks and acyclic graphs from KEGG pathways. *EMBNET J.* 26:949. doi: 10.14806/ej.26.0.949
- Chelikam, N., Vyas, V., Dondapati, L., Iskander, B., Patel, G., Jain, S., et al. (2022). Epidemiology, burden, and association of substance abuse amongst patients with cardiovascular disorders: national cross-sectional survey study. *Cureus* 14:e27016. doi: 10.7759/cureus.27016
- Chen, X., Lu, Y., Cue, J. M., Han, M. V., Nimgaonkar, V. L., Weinberger, D. R., et al. (2025). Classification of schizophrenia, bipolar disorder and major depressive disorder with comorbid traits and deep learning algorithms. *Schizophrenia* 11:14. doi: 10.1038/s41537-025-00564-7
- Chu, C., Liu, S., Nie, L., Hu, H., Liu, Y., and Yang, J. (2024). The interactions and biological pathways among metabolomics products of patients with coronary heart disease. *Biomed. Pharmacother.* 173:116305. doi: 10.1016/j.biopha.2024.116305
- Cirioncione, A. G., Clark, K. L., and Kann, M. G. (2018). Pathway networks generated from human disease phenome. *BMC Med. Genomics* 11:75. doi: 10.1186/s12920-018-0386-2
- Consortium, T. G. O., Aleksander, S. A., Balhoff, J., Carbon, S., Cherry, J. M., Drabkin, H. J., et al. (2023). The Gene Ontology knowledgebase in 2023. *Genetics* 224:iyad031. doi: 10.1093/genetics/iyad031
- Cui, X., He, H., He, F., Wang, S., Li, F., and Bo, X. (2015). Network fingerprint: a knowledge-based characterization of biomedical networks. *Sci. Rep.* 5:13286. doi: 10.1038/srep13286
- Daneshfarooz, N., Bagherzadeh Cham, M., Majidi, M., and Panahi, B. (2022). Identification of potentially functional modules and diagnostic genes related to amyotrophic lateral sclerosis based on the WGCNA and lasso algorithms. *Sci. Rep.* 12:20144. doi: 10.1038/s41598-022-24306-2
- Deak, J. D., and Johnson, E. C. (2021). Genetics of substance use disorders: a review. *Psychol. Med.* 51, 2189–2200. doi: 10.1017/S0033291721000969
- Dey, K. K., Hsiao, C. J., and Stephens, M. (2017). Visualizing the structure of rna-seq expression data using grade of membership models. *PLoS Genet.* 13, 1–23. doi: 10.1371/journal.pgen.1006759
- Fujita, A., Vidal, M. C., and Takahashi, D. Y. (2017). A statistical method to distinguish functional brain networks. *Front. Neurosci.* 11:66. doi: 10.3389/fnins.2017.00066
- Gable, A. L., Szklarczyk, D., Lyon, D., Matias Rodrigues, J. F., and von Mering, C. (2022). Systematic assessment of pathway databases, based on a diverse collection of user-submitted experiments. *Brief. Bioinformatics* 23:bbac355. doi: 10.1093/bib/bbac355
- Gan, W. Q., Buxton, J. A., Scheuermeyer, F. X., Palis, H., Zhao, B., Desai, R., et al. (2021). Risk of cardiovascular diseases in relation to substance use disorders. *Drug Alcohol Depend.* 229:109132. doi: 10.1016/j.drugalcdep.2021.109132
- Geistlinger, L., Csaba, G., Santarelli, M., Ramos, M., Schiffer, L., Turaga, N., et al. (2021). Toward a gold standard for benchmarking gene set enrichment analysis. *Brief. Bioinformatics* 22, 545–556. doi: 10.1093/bib/bbz158
- Genzile, M., Martini, L., Sponziello, M., and Becchetti, L. (2022). Biological Random Walks: multi-omics integration for disease gene prioritization. *Bioinformatics* 38, 4145–4152. doi: 10.1093/bioinformatics/btac446
- Gerring, Z. F., Thorp, J. G., Treur, J. L., Verweij, K. J. H., and Derks, E. M. (2024). The genetic landscape of substance use disorders. *Mol. Psychiatry* 29, 3694–3705. doi: 10.1038/s41380-024-02547-z
- Goh, K.-I., Cusick, M. E., Valle, D., Childs, B., Vidal, M., and Barab si, A.-L. (2007). The human disease network. *Proc. Nat. Acad. Sci.* 104, 8685–8690. doi: 10.1073/pnas.0701361104
- Gu, C., Shi, X., Dang, X., Chen, J., Chen, C., Chen, Y., et al. (2021). Identification of common genes and pathways in eight fibrosis diseases. *Front. Genet.* 11:627396. doi: 10.3389/fgene.2020.627396
- Gu, Y., Zheng, S., Yin, Q., Jiang, R., and Li, J. (2022). Redda: Integrating multiple biological relations to heterogeneous graph neural network for drug-disease association prediction. *Comput. Biol. Med.* 150:106127. doi: 10.1016/j.combiomed.2022.106127
- Gu, Z., Liu, J., Cao, K., Zhang, J., and Wang, J. (2012). Centrality-based pathway enrichment: a systematic approach for finding significant pathways dominated by key genes. *BMC Syst. Biol.* 6:56. doi: 10.1186/1752-0509-6-56
- Guo, Z., Fu, Y., Huang, C., Zheng, C., Wu, Z., Chen, X., et al. (2021). Noga: A network-oriented gene entropy approach for dissecting disease comorbidity and drug repositioning. *Genom. Proteom. Bioinform.* 19, 549–564. doi: 10.1016/j.gpb.2020.06.023
- Hagberg, A. A., Schult, D. A., and Swart, P. J. (2008). Exploring network structure, dynamics, and function using networkx, in *Proceedings of the 7th Python in Science Conference*, eds G. Varoquaux, T. Vaught, and J. Millman (Pasadena, CA: SciPy2008), 11–15.
- Hatoum, A. S., Colbert, S. M. C., Johnson, E. C., Huggett, S. B., Deak, J., Pathak, G. A., et al. (2023). Multivariate genome-wide association meta-analysis of over 1 million subjects identifies loci underlying multiple substance use disorders. *Nat. Mental Health* 1, 210–223. doi: 10.1101/2022.01.06.22268753
- Havakuk, O., Rezkalla, S. H., and Kloner, R. A. (2017). The cardiovascular effects of cocaine. *J. Am. Coll. Cardiol.* 70, 101–113. doi: 10.1016/j.jacc.2017.05.014
- Hossain, M. E., Uddin, S., Khan, A., and Moni, M. A. (2020). A framework to understand the progression of cardiovascular disease for type 2 diabetes mellitus patients using a network approach. *Int. J. Environ. Res. Public Health* 17:596. doi: 10.3390/ijerph17020596
- Humer, E., Pieh, C., and Brandmayr, G. (2020). Metabolomics in sleep, insomnia and sleep apnea. *Int. J. Mol. Sci.* 21:7244. doi: 10.3390/ijms21197244
- Jalali, Z., Khademalhosseini, M., Soltani, N., and Esmaili Nadimi, A. (2021). Smoking, alcohol and opioids effect on coronary microcirculation: an update overview. *BMC Cardiovasc. Disord.* 21:185. doi: 10.1186/s12872-021-01990-y
- Jardim, V. C., Santos, S. S., Fujita, A., and Buckeridge, M. S. (2019). Bionetstat: A tool for biological networks differential analysis. *Front. Genet.* 10:594. doi: 10.3389/fgene.2019.00594
- Jonker, T., Barnett, P., Boink, G. J. J., and Christoffels, V. M. (2023). Role of genetic variation in transcriptional regulatory elements in heart rhythm. *Cells* 13:4. doi: 10.3390/cells13010004
- Kanehisa, M. (2019). Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* 28, 1947–1951. doi: 10.1002/pro.3715

- Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., and Ishiguro-Watanabe, M. (2023). KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* 51, D587–D592. doi: 10.1093/nar/gkac963
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27
- Karp, P. D., Midford, P. E., Caspi, R., and Khodursky, A. (2021). Pathway size matters: the influence of pathway granularity on over-representation (enrichment analysis) statistics. *BMC Genomics* 22:191. doi: 10.1186/s12864-021-07502-8
- Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 44, W90–W97. doi: 10.1093/nar/gkw377
- Li, C.-Y., Mao, X., and Wei, L. (2008). Genes and (common) pathways underlying drug addiction. *PLoS Comput. Biol.* 4, 1–7. doi: 10.1371/journal.pcbi.0040002
- Li, Z., Zhang, Y., Fang, J., Xu, Z., Zhang, H., Mao, M., et al. (2022). Ncpath: a novel platform for visualization and enrichment analysis of human non-coding RNA and KEGG signaling pathways. *Bioinformatics* 39:494777. doi: 10.1101/2022.06.03.494777
- Milacic, M., Beavers, D., Conley, P., Gong, C., Gillespie, M., Griss, J., et al. (2023). The reactome pathway knowledgebase 2024. *Nucleic Acids Res.* 52, D672–D678. doi: 10.1093/nar/gkad1025
- Minhas, A. M. K., Kewcharoen, J., Hall, M. E., Warraich, H. J., Greene, S. J., Shapiro, M. D., et al. (2024). Temporal trends in substance use and cardiovascular disease-related mortality in the united states. *J. Am. Heart Assoc.* 13:e030969. doi: 10.1161/JAHA.123.030969
- Moon, K. Z., Rahman, M. H., Alam, M. J., Hossain, M. A., Hwang, S., Kang, S., et al. (2025). Unraveling the interplay between cardiovascular diseases and alcohol use disorder: a bioinformatics and network-based exploration of shared molecular pathways and key biomarkers validation via western blot analysis. *Comput. Biol. Chem.* 115:108338. doi: 10.1016/j.compbiolchem.2024.108338
- Morgan, S. L., Naderi, P., Koler, K., Pita-Juarez, Y., Prokopenko, D., Vlachos, I. S., et al. (2022). Most pathways can be related to the pathogenesis of alzheimer's disease. *Front. Aging Neurosci.* 14:846902. doi: 10.3389/fnagi.2022.846902
- Mubeen, S., Tom Kodamullil, A., Hofmann-Apitius, M., and Domingo-Fernández, D. (2022). On the influence of several factors on pathway enrichment analysis. *Brief. Bioinform.* 23:bbac143. doi: 10.1093/bib/bbac143
- Musunuru, K., and Kathiresan, S. (2019). Genetics of common, complex coronary artery disease. *Cell* 177, 132–145. doi: 10.1016/j.cell.2019.02.015
- Nersisyan, L., Samsonyan, R., and Arakelyan, A. (2014). Cykeggparser: tailoring kegg pathways to fit into systems biology workflows [version 2; peer review: 2 approved]. *F1000Research* 3:145. doi: 10.12688/f1000research.4410.2
- Orouji, S., Liu, M. C., Korem, T., and Peters, M. A. K. (2024). Domain adaptation in small-scale and heterogeneous biological datasets. *Sci. Adv.* 10:eadp6040. doi: 10.1126/sciadv.adp6040
- Pando-Naude, V., Toxto, S., Fernandez-Lozano, S., Parsons, C. E., Alcauter, S., and Garza-Villarreal, E. A. (2021). Gray and white matter morphology in substance use disorders: a neuroimaging systematic review and meta-analysis. *Transl. Psychiatry* 11:29. doi: 10.1038/s41398-020-01128-2
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830. doi: 10.5555/1953048.2078195
- Peng, Q., Wilhelmsen, K. C., and Ehlers, C. L. (2021). Common genetic substrates of alcohol and substance use disorder severity revealed by pleiotropy detection against gwas catalog in two populations. *Addict. Biol.* 26:e12877. doi: 10.1111/adb.12877
- Piñero, J., Ramírez-Anguita, J. M., Saüch-Pitarch, J., Ronzano, F., Centeno, E., Sanz, F., et al. (2020). The disgenet knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.* 48, D845–855. doi: 10.1093/nar/gkz1021
- Prasad, R. B., and Groop, L. (2015). Genetics of type 2 diabetes-pitfalls and possibilities. *Genes* 6, 87–123. doi: 10.3390/genes6010087
- Puig, S., Xue, X., Salisbury, R., Shelton, M. A., Kim, S.-M., Hildebrand, M. A., et al. (2023). Circadian rhythm disruptions associated with opioid use disorder in synaptic proteomes of human dorsolateral prefrontal cortex and nucleus accumbens. *Mol. Psychiatry* 28, 4777–4792. doi: 10.1038/s41380-023-02241-6
- R Core Team (2023). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rahmatallah, Y., Emmert-Streib, F., and Glazko, G. (2013). Gene sets net correlations analysis (gsnca): a multivariate differential coexpression test for gene sets. *Bioinformatics* 30, 360–368. doi: 10.1093/bioinformatics/btt687
- Rathor, P., and Ch, R. (2023). Metabolic basis of circadian dysfunction in parkinson's disease. *Biology* 12:1294. doi: 10.3390/biology12101294
- Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., et al. (2019). gprofiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 47, W191–198. doi: 10.1093/nar/gkz369
- Reimand, J., Isserlin, R., Voisin, V., Kucera, M., Tannus-Lopes, C., Rostamianfar, A., et al. (2019). Pathway enrichment analysis and visualization of omics data using gprofiler, GSEA, cytoscape and EnrichmentMap. *Nat. Protoc.* 14, 482–517. doi: 10.1038/s41596-018-0103-9
- Riley, E. D., Kazi, D. S., Coffin, P. O., Vittinghoff, E., Wade, A. N., Bulfone, T. C., et al. (2022). Impact of multiple substance use on circulating st2, a biomarker of adverse cardiac remodelling, in women. *Biomarkers* 27, 802–808. doi: 10.1080/1354750X.2022.2129451
- Rouillard, A. D., Gundersen, G. W., Fernandez, N. F., Wang, Z., Monteiro, C. D., McDermott, M. G., et al. (2016). The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database* 2016:baw100. doi: 10.1093/database/baw100
- Sales, G., Calura, E., Cavaliere, D., and Romualdi, C. (2012). graphite - a bioconductor package to convert pathway topology to gene network. *BMC Bioinformatics* 13:20. doi: 10.1186/1471-2105-13-20
- Sanchez-Roige, S., Kember, R. L., and Agrawal, A. (2022). Substance use and common contributors to morbidity: a genetics perspective. *eBioMedicine* 83:104212. doi: 10.1016/j.ebiom.2022.104212
- Sánchez-Valle, J., and Valencia, A. (2023). Molecular bases of comorbidities: present and future perspectives. *Trends Genet.* 39, 773–786. doi: 10.1016/j.tig.2023.06.003
- Santos, S., d., S., Galatro, T. F. A., Watanabe, R. A., Oba-Shinjo, S. M., Nagahashi Marie, S. K., et al. (2015). Coga: An r package to identify differentially co-expressed gene sets by analyzing the graph spectra. *PLoS ONE* 10:e0135831. doi: 10.1371/journal.pone.0135831
- Sato, J. R., Takahashi, D. Y., Hoexter, M. Q., Massier, K. B., and Fujita, A. (2013). Measuring network's entropy in adhd: a new approach to investigate neuropsychiatric disorders. *Neuroimage* 77, 44–51. doi: 10.1016/j.neuroimage.2013.03.035
- Shen, K. Z., Cox, B. A., and Johnson, S. W. (1997). L-arginine potentiates gaba-mediated synaptic transmission by a nitric oxide-independent mechanism in rat dopamine neurons. *Neuroscience* 79, 649–658. doi: 10.1016/S0306-4522(97)00024-9
- Smedley, D., Jacobsen, J. O., Jäger, M., Köhler, S., Holtgrewe, M., Schubach, M., et al. (2015). Next-generation diagnostics and disease-gene discovery with the exomiser. *Nat. Protoc.* 10, 2004–2015. doi: 10.1038/nprot.2015.124
- Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M., and Smoller, J. W. (2013). Pleiotropy in complex traits: challenges and strategies. *Nat. Rev. Genet.* 14, 483–495. doi: 10.1038/nrg3461
- Stoychev, K., Dilkov, D., Naghavi, E., and Kamburova, Z. (2021). Genetic basis of dual diagnosis: A review of genome-wide association studies (gwas) focusing on patients with mood or anxiety disorders and co-occurring alcohol-use disorders. *Diagnostics* 11:1055. doi: 10.3390/diagnostics11061055
- Sullivan, P. F., Kendler, K. S., and Neale, M. C. (2003). Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch. Gen. Psychiatry* 60, 1187–1192. doi: 10.1001/archpsyc.60.12.1187
- Takahashi, D. Y., Sato, J. R., Ferreira, C. E., and Fujita, A. (2012). Discriminating different classes of biological networks by analyzing the graphs spectra distribution. *PLoS ONE* 7:e0049949. doi: 10.1371/journal.pone.0049949
- Thaker, V. V. (2017). Genetic and epigenetic causes of obesity. *Adolesc. Med. State Art Rev.* 28, 379–405. doi: 10.1542/9781581109405-genetic
- Toska, E., and Mayrovitz, H. N. (2023). Opioid impacts on cardiovascular health. *Cureus* 15:e46224. doi: 10.7759/cureus.46224
- Tousoulis, D., Böger, R. H., Antoniadis, C., Siasos, G., Stefanadi, E., and Stefanadis, C. (2007). Mechanisms of disease: L-arginine in coronary atherosclerosis—a clinical perspective. *Nature Clin. Pract. Cardiovascul. Med.* 4, 274–283. doi: 10.1038/ncpcardio0878
- Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., et al. (2021). Genome-wide association studies. *Nat. Rev. Methods Primers* 1:59. doi: 10.1038/s43586-021-00056-9
- Vavrek, M. J. (2011). fossil: Palaeoecological and palaeogeographical analysis tools, in *Palaeontologia Electronica*, 14.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272. doi: 10.1038/s41592-020-0772-5
- Wang, J., Vasaikar, S., Shi, Z., Greer, M., and Zhang, B. (2017). WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.* 45, W130–W137. doi: 10.1093/nar/gkx356
- Warrens, M. J., and van der Hoef, H. (2022). Understanding the adjusted rand index and other partition comparison indices based on counting object pairs. *J. Classificat.* 39, 487–509. doi: 10.1007/s00357-022-09413-z
- Wieder, C., Frainay, C., Poupin, N., Rodríguez-Mier, P., Vinson, F., Cooke, J., et al. (2021). Pathway analysis in metabolomics: recommendations for the use of over-representation analysis. *PLoS Comput. Biol.* 17, 1–23. doi: 10.1371/journal.pcbi.1009105
- Wilk, G., and Braun, R. (2017). Integrative analysis reveals disrupted pathways regulated by microRNAs in cancer. *Nucleic Acids Res.* 46, 1089–1101. doi: 10.1093/nar/gkx1250

- Wilson, L. D., Jeromin, J., Garvey, L., and Dorbandt, A. (2001). Cocaine, ethanol, and cocaethylene cardiotoxicity in an animal model of cocaine and ethanol abuse. *Acad. Emerg. Med.* 8, 211–222. doi: 10.1111/j.1553-2712.2001.tb01296.x
- Wirka, R. C., Pjanic, M., and Quertermous, T. (2018). Advances in transcriptomics. *Circ. Res.* 122, 1200–1220. doi: 10.1161/CIRCRESAHA.117.310910
- Wormington, B., Thorp, J. G., Gerring, Z. F., Scott, J. G., Akosile, W., and Derks, E. M. (2024). The genetic architecture of substance use and its diverse correlations with mental health traits. *Psychiatry Res.* 342:116200. doi: 10.1016/j.psychres.2024.116200
- Wu, Z., and Wu, H. (2020). Accounting for cell type hierarchy in evaluating single cell rna-seq clustering. *Genome Biol.* 21:123. doi: 10.1186/s13059-020-02027-x
- Wydra, K., Frankowska, M., and Filip, M. (2024). *Recent Advances on GABAB Receptor Positive Allosteric Modulators as Potential Pharmacotherapies for Substance Use Disorder and Food Addiction*. (Cham: Springer International Publishing), 239–258.
- Xiong, Y., Guo, M., Ruan, L., Kong, X., Tang, C., Zhu, Y., et al. (2019). Heterogeneous network embedding enabling accurate disease association predictions. *BMC Med. Genomics* 12:186. doi: 10.1186/s12920-019-0623-3
- Yeates, K., Lohfeld, L., Sleeth, J., Morales, F., Rajkotia, Y., and Ogedegbe, O. (2015). A global perspective on cardiovascular disease in vulnerable populations. *Can. J. Cardiol.* 31, 1081–1093. doi: 10.1016/j.cjca.2015.06.035
- Yeung, K. Y., Medvedovic, M., and Bumgarner, R. E. (2003). Clustering gene-expression data with repeated measurements. *Genome Biol.* 4:R34. doi: 10.1186/gb-2003-4-5-r34
- Zelig, A., Kariti, H., and Kaplan, N. (2023). Kmd clustering: robust general-purpose clustering of biological data. *Commun. Biol.* 6:1110. doi: 10.1038/s42003-023-05480-z
- Zhao, K., and Rhee, S. Y. (2023). Interpreting omics data with pathway enrichment analysis. *Trends in Genetics* 39, 308–319. doi: 10.1016/j.tig.2023.01.003
- Zhou, H., Polimanti, R., Yang, B. Z., Wang, Q., Han, S., Sherva, R., et al. (2017). Genetic risk variants associated with comorbid alcohol dependence and major depression. *JAMA Psychiatry* 74, 1234–1241. doi: 10.1001/jamapsychiatry.2017.3275
- Zhukovsky, P., Tio, E. S., Coughlan, G., Bennett, D. A., Wang, Y., Hohman, T. J., et al. (2024). Genetic influences on brain and cognitive health and their interactions with cardiovascular conditions and depression. *Nat. Commun.* 15:5207. doi: 10.1038/s41467-024-49430-7
- Zitnik, M., Li, M. M., Wells, A., Glass, K., Morselli Gysi, D., Krishnan, A., et al. (2024). Current and future directions in network biology. *Bioinform. Adv.* 4:vbae099. doi: 10.1093/bioadv/vbae099
- Zito, A., Lualdi, M., Granata, P., Cocciadiferro, D., Novelli, A., Alberio, T., et al. (2021). Gene set enrichment analysis of interaction networks weighted by node centrality. *Front. Genet.* 12:577623. doi: 10.3389/fgene.2021.577623