

<https://doi.org/10.1038/s43856-025-01329-0>

# Sex-specific transcriptome similarity networks elucidate comorbidity relationships



Jon Sánchez-Valle <sup>1</sup> ✉, María Flores-Rodero <sup>1,2</sup>, Felipe Xavier Costa <sup>3,4</sup>, Jose Carbonell-Caballero<sup>1</sup>, Iker Núñez-Carpintero <sup>1,5</sup>, Rafael Tabarés-Seisdedos <sup>2</sup>, Luis Mateus Rocha <sup>3,4</sup>, Davide Cirillo <sup>5</sup> & Alfonso Valencia <sup>1,6</sup> ✉

## Abstract

**Background** Biological differences between women and men lead to variations in the prevalence and progression of many diseases, influencing diagnosis, management, and treatment outcomes. However, the biological mechanisms that contribute to sex differences in disease co-occurrence remain largely unexplored. This study aims to uncover the molecular processes underlying sex-specific patterns of comorbidity.

**Methods** We analyze gene expression data from over 100 diseases, considering the biological sex of each sample (8906 samples, 43.06% women). For each sex, we construct disease similarity networks based on differential gene expression profiles and identify enriched biological processes. We then compare these networks with epidemiological data from population-level comorbidity studies to assess their concordance. Finally, we investigate drugs associated with sex-specific comorbidities to identify potential differences in therapeutic response.

**Results** We show that 13–16% of transcriptomically similar disease pairs are sex-specific. These similarities recover 53–60% of known comorbidities that differ between women and men. Diseases can co-occur through the differential alteration of biological processes, with immune and metabolic pathways playing a greater role in women, and extracellular matrix organization and signal transduction pathways in men. We also identify drugs differentially linked to comorbid diseases depending on sex, suggesting possible sex-dependent effects on disease co-occurrence.

**Conclusions** Our findings demonstrate that transcriptomic data can reveal sex-specific molecular links between diseases and suggest that biological sex should be considered in the design of therapeutic strategies and drug administration.

## Plain language summary

Men and women often develop different co-occurring health conditions, but the biological reasons are not well understood. Our study investigated differences in gene activity between men and women across more than 100 diseases. By comparing these patterns, we discovered distinct groups of diseases that are more strongly connected to one sex than the other. Our results suggest that the immune system and metabolism play a more prominent role in women's diseases, while tissue structure and cell communication are more dominant in men's. We also found evidence that some treatments may work differently in men and women, highlighting the critical need to consider sex in both medical research and treatment decisions.

Comorbidity and multimorbidity, defined as the presence of more than one medical condition in individuals<sup>1,2</sup>, have been investigated using cohort, case-control, and nested case-control study designs. These studies have shown, for example, that patients with type II diabetes (T2D) exhibit a higher prevalence of dementia and cancer, with men showing a particularly elevated risk of Alzheimer's disease<sup>3,4</sup>. The development of digital medical record systems<sup>5</sup> and the implementation of the Veterans Health

Information System and Technology Architecture<sup>6</sup> enabled large-scale collection of clinical information in the form of electronic health records (EHR). Since then, multiple studies have systematically analyzed comorbidity relations and constructed networks based on epidemiological evidence. For example, Hidalgo et al. generated a comorbidity network using Medicare data from more than 30 million patients in the United States<sup>7</sup>. They found insightful sex-specific differences in comorbidity patterns—for

<sup>1</sup>Computational Biology, Barcelona Supercomputing Center, Barcelona, Spain. <sup>2</sup>Department of Medicine, University of Valencia, CIBERSAM, INCLIVA, Valencia, Spain. <sup>3</sup>Universidade Católica Portuguesa, Católica Medical School, Católica Biomedical Research Centre, Lisbon, Portugal. <sup>4</sup>School of Systems Science and Industrial Engineering, Binghamton University (State University of New York), Binghamton, NY, USA. <sup>5</sup>Machine Learning for Biomedical Research, Barcelona Supercomputing Center, Barcelona, Spain. <sup>6</sup>ICREA, Barcelona, Spain. ✉e-mail: [jon.sanchez@bsc.es](mailto:jon.sanchez@bsc.es); [alfonso.valencia@bsc.es](mailto:alfonso.valencia@bsc.es)

example, a higher risk of nephropathies in women and acute myocardial infarction in men with T2D. Similarly, Jensen et al. used Danish EHRs to construct temporal disease trajectories by linking statistically significant disease co-occurrences, thereby identifying key conditions that drive disease progression and increase mortality<sup>8</sup>. In the same population, Westergaard et al. reported that women exhibit a greater number of comorbidities than men, as well as more frequent disease co-occurrences over longer time spans<sup>9</sup>. Together, these findings indicate that men and women differ substantially in their comorbidity profiles, underscoring the importance of understanding these differences from a molecular perspective.

Since the publication of the first human disease network in 2007 by Goh et al., in which diseases were connected if they shared at least one altered gene (the *diseasome*)<sup>10</sup>, numerous studies have explored molecular similarities between diseases, including efforts to better understand comorbidities. Lee et al. demonstrated that diseases involving coupled metabolic reactions co-occur three to seven times more frequently than those without such connections<sup>11</sup>. Likewise, studies measuring distances between disease modules in the protein interactome have shown that diseases with overlapping modules tend to co-occur more often than expected by chance<sup>12,13</sup>. Transcriptomic similarities between diseases have also been found to reflect epidemiologically observed co-occurrences<sup>14,15</sup>. More recently, Dong et al. integrated EHR and genome-wide association studies (GWAS) data from the UK Biobank to recapitulate 46% of observed multimorbidities<sup>16</sup>, while Murrin et al. found that most pairs of chronic conditions with shared genetic features co-occur in the primary care setting<sup>17</sup>.

We have previously highlighted the lack of studies examining biological differences between women and men to better understand sex-specific comorbidity patterns<sup>18</sup>. This gap is notable given that 37% of all genes exhibit sex-biased expression in at least one tissue<sup>19</sup>. Liu et al. found sex-specific disease-associated polymorphisms in GWAS<sup>20</sup>, and Lopes-Ramos et al. constructed sample-specific gene regulatory networks from healthy human tissues to reveal that many transcription factors have sex-dependent regulatory targets. Interestingly, these differentially targeted genes are enriched for tissue-related functions and diseases. For example, genes associated with Alzheimer's disease are regulated by distinct transcription factors depending on the sex of the sample<sup>21</sup>.

To address the knowledge gap in the molecular bases of sex-specific comorbidities, we have generated disease transcriptomic similarity networks separately for men and women. To maintain consistent terminology, we use the gender terms women and men to refer to the binary categories in both transcriptomic data (females/males) and epidemiological data (women/men), and refer to them collectively as sexes without implying alignment between sex and gender. The resulting networks recover a representative set of comorbidities previously described for women and men<sup>9</sup>. By analyzing pathways altered in the same direction in comorbid diseases, we propose hypotheses explaining differences in disease co-occurrence between sexes. Moreover, we find that disease pairs may co-occur more frequently than expected by chance in both sexes, but through distinct biological processes. Finally, we extend these findings to potentially related drugs, emphasizing the scientific and clinical relevance of studying sex-specific molecular differences in disease and comorbidity. In summary, this study provides molecular hypotheses to explain sex differences in comorbidity relationships and explores the potential roles of drugs within these relationships.

## Methods

### Gene expression analysis

Raw gene expression data were obtained from the Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>) and ArrayExpress repositories (<https://www.ebi.ac.uk/arrayexpress>). Studies conducted using the *Affymetrix HG U133Plus 2* microarray platform were selected for their cost-effectiveness, reproducibility, and translational potential compared to other assays. This selection also ensured sufficiently large sample sizes for robust disease-disease association analyses, while minimizing biases from

platform heterogeneity. After excluding low-quality samples (GNUSE median values  $>1.25^{22}$ ), we retained 128 diseases with at least three case and three control samples (Supplementary Data 1). Because sex information was available for only 52.23% (6685/12,797) of samples, we used the *massIR* package to infer sample sex, correctly recovering sex in 94.24% of annotated samples<sup>23</sup>. This method classifies samples as male or female by analyzing the expression levels of probes corresponding to Y-chromosome genes. To harmonize our transcriptomic similarity networks with previously published epidemiological networks<sup>9</sup>, disease names were mapped to three-digit codes of the World Health Organization's International Classification of Diseases, version 10 (ICD-10), grouping specific conditions under each code. This resulted in 76 ICD-10 codes in women (2465 cases and 1370 controls) and 77 in men (3200 cases and 1871 controls), with 59 codes common to both sexes (see Supplementary Fig. 1). Overall, women represented 43% of cases and 42% of controls, with case samples comprising ~63–64% of total samples in both sexes. Lowly expressed genes (detection  $p$ -value  $< 0.05$ ) were identified and removed using the MAS 5.0 algorithm<sup>24</sup>. Background correction, normalization, and summarization were performed using the frozen Robust Multiarray Analysis (fRMA) preprocessing algorithm with default parameters<sup>25</sup>. Differential expression analyses comparing case (disease) versus control (disease-free) samples were conducted using the modeling framework implemented in the LIMMA package<sup>26</sup>. Analyses were performed both separately by sex and jointly across all samples, adjusting for potential confounders such as study of origin and sex. Genes with a false discovery rate (FDR)  $\leq 0.05$  and log fold change (logFC)  $< 0$  or  $> 0$  were classified as significantly down- or up-regulated, respectively. All R packages used for analysis and visualization are listed at <https://github.com/jonsv89/SHDC><sup>27</sup>.

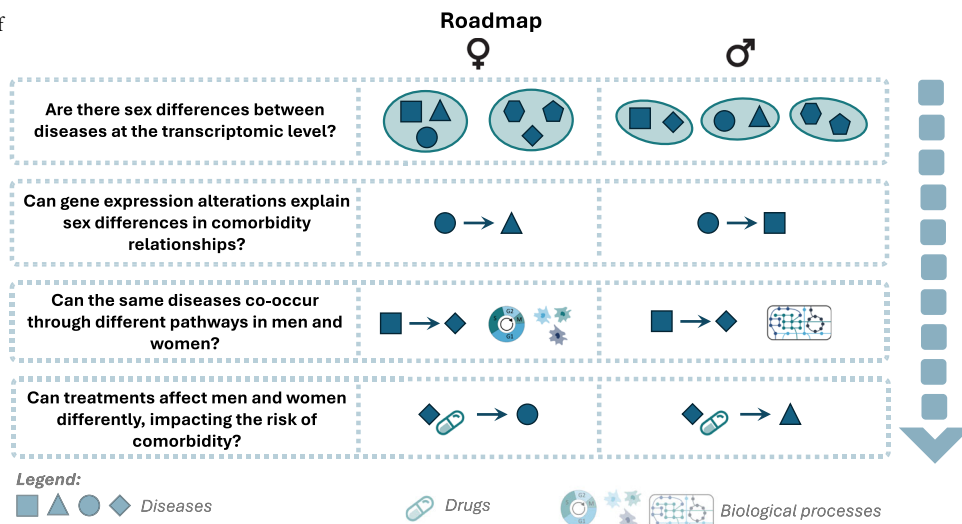
### Gene set enrichment analysis

Functional enrichment was performed using gene set enrichment analysis (GSEA)<sup>28</sup>, applied to the full list of genes ranked by logFC from the differential expression results. Gene sets from Reactome, Gene Ontology (GO), and KEGG databases were used for GSEA. Disease clustering was based specifically on Reactome pathways, whose hierarchical organization enabled the identification of 29 lowest-level pathway categories. Pairwise Euclidean distances were computed between diseases based on their Reactome pathway enrichment profiles (using normalized enrichment scores from GSEA). Hierarchical clustering was performed using the Ward2 linkage method<sup>29</sup>, and significant clusters ( $p$  value  $\leq 0.05$ ) were identified via bootstrap resampling using the *pvclust* R package<sup>30</sup>. Resulting dendrograms for women and men were compared using *tanglegrams* generated with the *dendextend* R package<sup>31</sup>.

### Network construction

Transcriptional similarities between diseases were calculated using three gene sets: (i) all annotated genes, (ii) the union of genes with significant differential expression (sDEGs), and (iii) the intersection of sDEGs, following previous studies on the molecular bases of comorbidities<sup>14,15</sup> (see Fig. 1). Six similarity metrics were computed: Pearson and Spearman correlation coefficients, cosine similarity, and Euclidean, Canberra, and Manhattan distances. Empirical  $p$  values were obtained from 10,000 random permutations for cosine, Euclidean, Canberra, and Manhattan metrics, with Bonferroni correction applied; similarities with FDR  $\leq 0.05$  were considered significant. For distance metrics (Euclidean, Canberra, Manhattan) observed distances were compared to the mean of random distances, yielding positive (greater) or negative (lesser) similarity values relative to expectation (see Supplementary Fig. 2). Similarity values were then binarized: coefficients  $> 0$  were set to  $+1$ , and those  $< 0$  to  $-1$ . The resulting disease transcriptomic similarity networks (DTSNs) generated from the different metrics were largely consistent, particularly those based on Pearson, Spearman, cosine, and Euclidean measures (see Supplementary Fig. 3). For clarity, results shown in the main text correspond to Euclidean-based DTSNs, which recovered the highest number of comorbidity relationships (see Supplementary Table 1). As the number of sDEGs strongly influences

**Fig. 1 | Study roadmap.** Schematic representation of the main questions raised in the study. Squares, diamonds, triangles and circles denote diseases. Arrows indicate a higher-than-expected risk of developing a disease when suffering from a different one.



similarity detection (correlations = 0.75–0.86 for sDEG sets vs. 0.34 using all genes; see Supplementary Fig. 4), analyses were focused on the complete list of genes. All DTSNs are available at the disease-perception portal (<https://disease-perception.bsc.es/shdc/>). The network backbone was extracted following the method described by ref. 32.

### Overlap with epidemiology

To evaluate the extent to which DTSNs capture known comorbidity patterns, we used the epidemiological network published by ref. 9 (Supplementary Note 1). Overlap analyses were conducted on the shared set of diseases present in both transcriptomic and epidemiological networks. Positive and negative transcriptomic similarities were compared separately with epidemiological associations. All comparisons were stratified by sex (women-women, men-men). Statistical significance was assessed using Fisher's exact test and a randomization approach, generating 10,000 degree-preserving random networks to establish null distributions.

### Disease–drug associations

To explore potential sex-specific drug influences on comorbidities, drug target data were retrieved from the DrugBank<sup>33</sup>. Because the number of direct targets per drug is relatively small for enrichment analyses, we expanded target sets using the experimentally validated human protein-protein interaction network from IID<sup>34</sup>. First-neighbor interactors of each drug's primary targets were included to increase pathway coverage. We then performed GSEA<sup>28</sup> to associate drugs targeting proteins encoded by up- or down-regulated genes with corresponding diseases, analyzed separately by sex. Disease–drug associations were obtained from the SIDER database, which compiles drug indications and adverse reactions extracted from the package inserts using natural language processing<sup>35</sup>. Disease names were converted to ICD-10 codes via the Unified Medical Language System<sup>36</sup>, and DrugBank identifiers were mapped to standardized drug names.

### Statistics and reproducibility

The five sections, “Gene expression analysis”, “Gene set enrichment analysis”, “Network construction”, “Overlap with epidemiology”, and “Disease–drug associations” describe how the statistical analyses of the data were conducted.

### Ethics

All analyses were performed on publicly available, de-identified gene expression datasets obtained from the Gene Expression Omnibus (GEO) and ArrayExpress repositories. According to GEO submission requirements, data submitters must ensure that deposited datasets comply with institutional and national ethical regulations, including that the information “does not compromise participant privacy and is in accord with the original

consent,” and that non-NIH-funded studies have “appropriate consent/permission to submit the data to a public database.” Similarly, ArrayExpress accepts high-throughput functional genomics data only when accompanied by appropriate study metadata, sample annotations, and protocols, in accordance with community guidelines for ethically conducted studies.

Ethical approval and informed consent for each original dataset were therefore obtained by the investigators who generated the data, as documented in the corresponding primary publications. No new data were collected for the present study, and all datasets were fully de-identified prior to public release. Because this work involves secondary analysis of publicly available, non-identifiable data, it does not constitute human subjects research, and no additional IRB/Ethics approval was required at our institution.

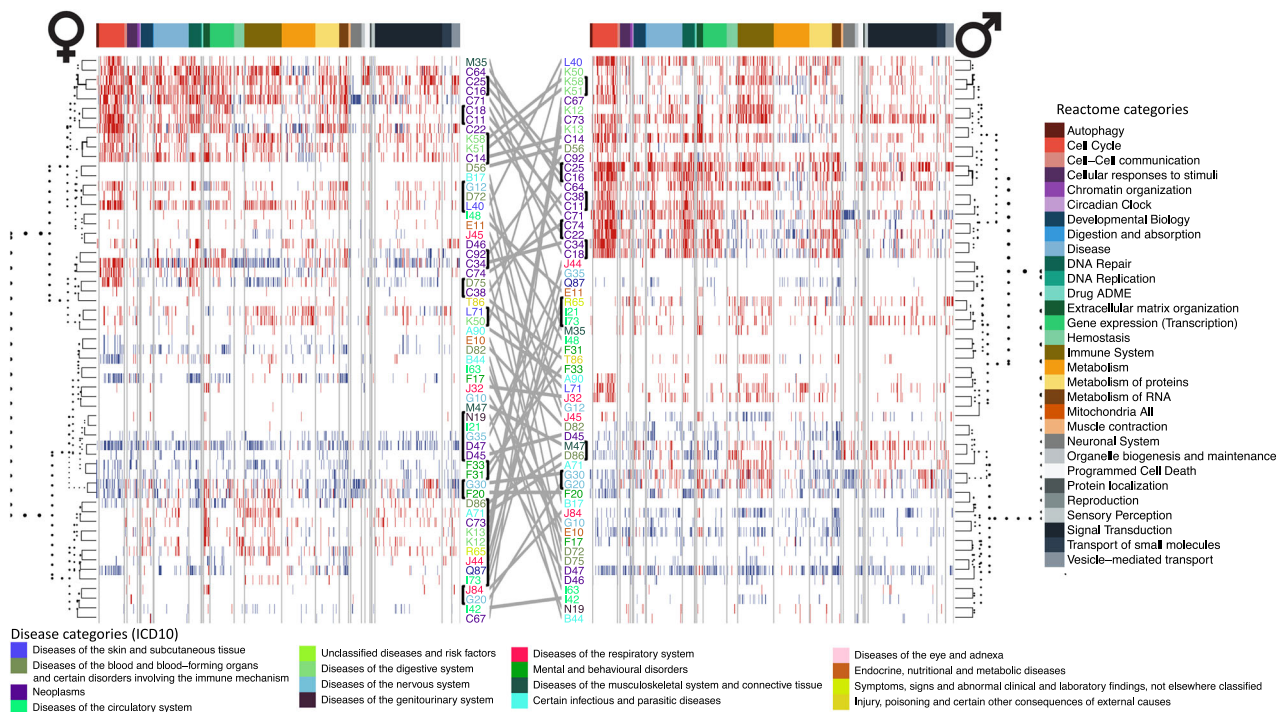
## Results

### Sex-associated differences in gene expression

To obtain an initial overview of disease-related differences between women and men, we performed hierarchical clustering on the normalized enrichment score of their enriched pathways based on GSEA (see methods). On average, clusters were larger and more heterogeneous in women than in men (66% of clusters in women and 25% in men included diseases from different categories, see Fig. 2). Overall, the overlap between men and women clusters was limited. Parkinson's and Alzheimer's diseases (G20 and G30), which have been previously linked at the molecular level<sup>37</sup>, clustered significantly in men. These two diseases shared 112 pathways enriched in differentially expressed genes, primarily related to signal transduction, the immune system, and neuronal function. In contrast, in women, Alzheimer's disease clustered with schizophrenia (F20), sharing 116 enriched pathways. These included metabolic pathways (glucose metabolism and respiratory electron transport)<sup>38,39</sup>, protein metabolism (protein ubiquitination)<sup>40,41</sup>, and neuronal system pathways (such as the serotonin neurotransmitter release cycle and GABA synthesis, release, reuptake, and degradation)<sup>42,43</sup>, all previously associated with both diseases. Interestingly, dementia (including Alzheimer's disease) has been reported to co-occur significantly in patients with schizophrenia, with a relative risk (RR) of 2.29, being the risk higher among women with schizophrenia<sup>44</sup>.

Only two pairs of diseases clustered together significantly in both sexes: pancreatic cancer—gastric cancer (C25 and C16) and irritable bowel syndrome (IBS)—ulcerative colitis (K58 and K51). Although both tumors clustered together in men and women, the dominant molecular similarities differed: cell cycle-related processes predominated in men, while immune system-related processes predominated in women. Notably, interleukin signaling pathways were overactivated in women, supporting their potential as candidates for developing immunomodulatory therapeutic strategies<sup>45</sup>. Considerable sex-specific divergence in tumor immune responses has been





**Fig. 2 | Reactome pathways are significantly altered in disease in women and men.** The diseases are represented in rows, and the pathways are in columns. The colors of the diseases refer to the category of the ICD-10 to which they belong. The color bar in the columns represents the Reactome parents to which each pathway belongs. The red (blue) lines represent pathways that are under- (over-) expressed in each disease.

The diseases have been clustered by calculating Euclidean distances between them and using the ward.D2 method<sup>29</sup> on the GSEA's normalized enrichment scores. The gray lines connect diseases in women and men. Black lines near ICD codes indicate significant clusters after bootstrapping.

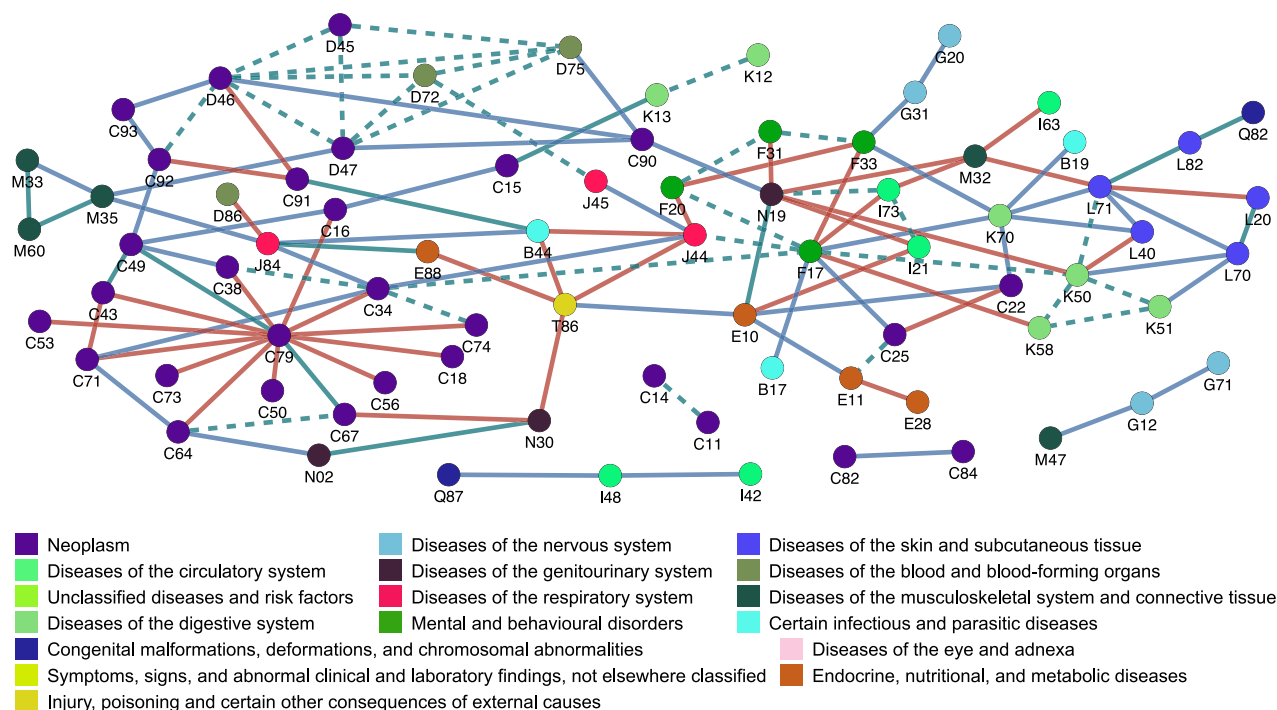
documented, which may have implications for sex differences in immunotherapy efficacy<sup>46</sup>. For the two digestive system diseases (K58 and K51), these also clustered with oral cavity cancer (C14) in women, a known comorbidity. The risk of oral cavity cancer is higher in women (standardized incidence ratio of 12.07 in women vs. 8.49 in men)<sup>47</sup>. The increasing prevalence of HPV may further amplify this risk in IBS patients, potentially enhanced by immunosuppressive therapies<sup>48</sup>, underscoring the influence of treatments in shaping comorbidity relationships. In this women-specific cluster, 37 pathways were significantly altered—28 upregulated (mainly signal transduction pathways) and nine downregulated (primarily metabolic pathways). Among them, NOTCH4 signaling was overactivated, a pathway proposed as an oral cancer marker<sup>49</sup>. NOTCH4 expression is induced in macrophages following activation by Toll-like receptors (TLRs) and interferon- $\gamma$  (IFN- $\gamma$ ), both extensively studied in the context of IBS<sup>50</sup>. Together, these findings reveal substantial molecular differences between women and men across diseases from distinct categories, raising the question of whether such molecular disparities contribute to differences in comorbidity patterns.

### Disease transcriptomic similarity networks

After observing sex-specific differences in disease clustering based on pathway enrichment, we constructed disease transcriptomic similarity networks (DTSNs) to evaluate pairwise disease relationships. Similarities between disease pairs were calculated adjusting for sex differences, and separately for women (wDTSN) and men (mDTSN, see Methods).

Between 45 and 48% of all edges in the wDTSN and mDTSN were positive, indicating smaller-than-expected distances between diseases. In contrast, 25–27% of the edges were negative, reflecting larger-than-expected distances. Consistent with epidemiological observations<sup>9</sup>, significantly more positive interactions (i.e., direct comorbidities) were detected between diseases belonging to the same category than between those from different categories in both women and men (odds ratio (OR) = 2.81 in women vs.

OR = 1.83 in men; Supplementary Table 2). This finding suggests that transcriptional similarity may provide insight into why diseases affecting the same system are more likely to co-occur than those affecting different systems. The ORs increased further when considering only the connections that preserved the most similar comorbidity trajectories—that is, the network backbone<sup>32</sup>, (4.4 in women and 3.2 in men)—highlighting the stronger relevance of intra-category connections in the flow of disease information. This pattern is also consistent in epidemiology (see Supplementary Note 2). When analyzing specific disease categories within the mDTSN and wDTSN, we found that digestive and musculoskeletal system diseases in men and skin diseases in women each exhibited a clustering coefficient of 1, indicating full interconnection within the category. Mental illnesses followed, with a network density of 0.83 in women compared with 0.4 in men, reflecting the higher comorbidity burden of mental disorders among women<sup>51</sup> (see Supplementary Fig. 5). When focusing on diseases common to both sexes, we observed significantly fewer positive interactions in women than in men between digestive system and nervous system diseases (OR = 0.21) and between digestive system and blood diseases (OR = 0.27, see Supplementary Table 3). These findings align with the higher risk of co-occurrence between IBS and neurological disorders such as multiple sclerosis<sup>52</sup>, Parkinson's disease<sup>53</sup>, or dementia<sup>54</sup>. Notably, no positive interactions were found between IBS and multiple sclerosis in women. Conversely, women displayed significantly more negative interactions between digestive and nervous system diseases (OR = 10.81, see Supplementary Table 4). Comparing the networks generated jointly (both sexes combined) with those generated separately, 16.22 and 13.49% of positive interactions were exclusive to women and men, respectively—proportions roughly consistent with epidemiological estimates (9% in women and 4% in men)<sup>9</sup> (see Supplementary Fig. 6A, B). These findings demonstrate that, as in epidemiology, sex-specific transcriptional relationships can be obscured when analyzing data from women and men together.



**Fig. 3 | Comorbidities explained by the disease transcriptomic similarity network.** Network of comorbidities recovered based on transcriptomic similarities. Each node represents a disease, colored based on the disease category it belongs to (ICD-10). Red, blue, and green edges represent comorbidities (retrieved by analyzing the epidemiology separately for women and men and adjusting for sex)

recovered by calculating similarities at the gene expression level between diseases in the same way (separately for women and men and adjusting for sex differences). The dashed edges represent comorbidities described in women and men that are recovered by analyzing transcriptomic similarities separately for women and men. The network's visualizations have been generated using Cytoscape software<sup>112</sup>.

### Biological clues to differences in disease co-occurrences in men and women

Previous studies have demonstrated that transcriptional similarity can significantly recover comorbidity relationships<sup>14,15</sup>, even across diverse populations, underscoring the robustness of this approach (see Supplementary Note 3). However, as discussed in the Introduction, this phenomenon has not yet been evaluated separately for women and men<sup>18</sup>. In our analysis, 53.12 and 60.68% of the disease co-occurrences reported in women and men by ref. 9 were also captured in the wDTSN and mDTSN, respectively (see Fig. 3, Supplementary Fig. 6C, and Supplementary Table 1).

The sex-specific DTSNs revealed altered biological processes in comorbid diseases within both the same and different disease categories. For example, schizophrenia (F20) and chronic obstructive pulmonary disease (COPD, J44) were connected in the wDTSN and shared enrichment for mitochondrial processes (mitochondrial translation, mitochondrial RNA metabolic process, and mitochondrial gene expression) and immune-related pathways (interleukin 10 signaling, neutrophil degranulation, macrophage cytokine production, and antimicrobial peptides) (see Supplementary Data 3). Interestingly, patients with schizophrenia have been reported to exhibit an elevated risk of COPD<sup>55</sup>, even when compared to smoking-matched controls<sup>56</sup>. Impaired lung function is often observed in schizophrenia<sup>57</sup> and shared eQTL variants affecting both lung function and neuropsychiatric traits<sup>58</sup> may contribute to this relationship. Clinical data also show sex differences in schizophrenia prevalence, symptomatology, and treatment response<sup>59</sup>, as well as in comorbidity patterns—for instance, the risk of developing COPD appears to be higher among women<sup>60</sup> and is statistically significant in women but not men in the Danish population<sup>9</sup>. The enrichment of mitochondrial and immune pathways is consistent with reports of stronger mitochondrial signatures in women with schizophrenia<sup>61</sup> and COPD<sup>62</sup> and with known sex-related immune differences<sup>63,64</sup>. These findings support the hypothesis that sex-specific molecular variations may contribute to distinct comorbidity patterns between women and men.

A comparable pattern emerged for smoking (F17) and irritable bowel syndrome (IBS, K58), which were linked in the wDTSN through shared enrichment of mitochondrial respiratory chain assembly, electron transport, complex I biogenesis, detoxification of reactive oxygen species, and immune processes such as neutrophil activation and immunoregulatory interactions between lymphoid and non-lymphoid cells. Epidemiological data indicate an elevated risk of IBS among smokers<sup>65</sup>, particularly in women, consistent with the greater physiological impact of smoking in women<sup>66</sup> and the higher prevalence of IBS in women<sup>67</sup>, possibly influenced by hormonal factors<sup>68</sup>. Prior molecular studies have independently reported mitochondrial and immune alterations in both conditions<sup>69–71</sup>. Additional disease pairs that co-occur more frequently in women—but not in men—and were uniquely connected in the wDTSN included type I diabetes (T1D, E10) with myocardial infarction (I21)<sup>72</sup>, bipolar disorder (F31) with uremia (N19), and COPD (J44) with chronic lung allograft dysfunction (B44, see Fig. 3).

In contrast, the mDTSN highlighted a men-specific association between T1D (E10) and liver cancer (C22), correlating with the higher-than-expected risk of developing liver cancer (C22) in T1D patients<sup>73</sup>. Notably, T1D prevalence is greater in prepubertal girls but becomes approximately twice as common in men after puberty<sup>74</sup>. In men, most biological processes altered in the same direction in both diseases were associated with metabolism (metabolism of amino acids and derivatives, metabolism of vitamins and cofactors, glutathione metabolic process, reactive oxygen species metabolic process, and response to starvation) and immune regulation (humoral immune response, positive regulation of immune effector process, and regulation of T helper 1 type immune response). Additional examples are detailed in Supplementary Note 4).

Finally, we calculated disease similarities separately within each Reactome parent category (see Supplementary Note 5). Gene expression and immune system pathways recovered the highest proportions of known comorbidities (41–47%), whereas drug ADME pathways recovered the fewest (4–5%) (see Supplementary Table 5). Despite variation in the number

of comorbidities captured across categories, all overlaps with epidemiological data were statistically significant (see Supplementary Fig. 7A, B and Supplementary Table 5). Collectively, these findings provide strong evidence for sex-dependent differences in the biological processes underlying diseases and suggest that such differences may contribute to the distinct patterns of disease co-occurrence observed between women and men. Among the most relevant processes are those related to the immune system, metabolism, and mitochondrial function—mechanisms previously reported to drive biological differences between sexes<sup>75,76</sup>. All generated networks are available for interactive visualization at <https://disease-perception.bsc.es/shdc/>.

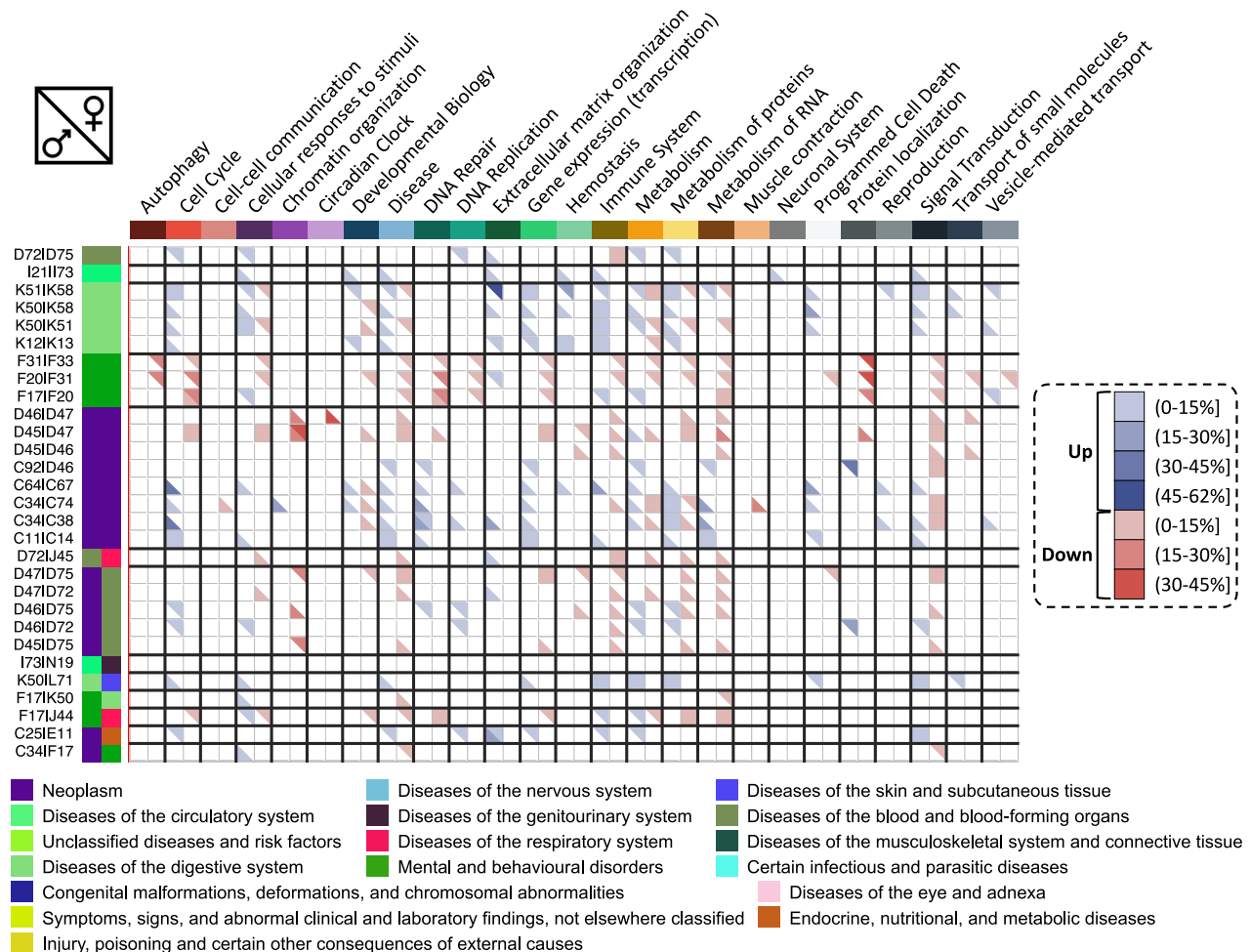
### Comorbidities occur through different mechanisms in women and men

After confirming that sex-specific differences in transcriptional similarities between diseases can explain differences in comorbidity patterns, we next investigated whether the underlying mechanisms driving disease co-occurrence might differ between women and men. Although mechanistic sex differences have been described for several biological processes—such as mechanical pain hypersensitivity, which is mediated by distinct immune cells in male (microglia) and female (T-cell) mice<sup>77</sup>—such distinctions have not been explored in the context of comorbidities. As illustrated in Fig. 4, 29 disease pairs were found to co-occur more often than expected by chance in both women and men (12 of which belong to different disease categories), and these comorbidities were consistently recovered in both the wDTSN

and mDTSN. Nevertheless, the associated biological alterations displayed clear sex-specific patterns (see Supplementary Data 4).

A well-established example is smoking (F17) and COPD (J44), which co-occur in both women and men<sup>78</sup>. In women, shared alterations between smokers and COPD patients involved pathways related to the cell cycle, response to stimuli, metabolism, immune system, and developmental biology, whereas these pathways were not observed in men (see Fig. 4). Conversely, in men, shared alterations were primarily found in DNA repair, protein metabolism, and RNA metabolism pathways, which were not observed in women. These differences align with prior studies suggesting that cigarette metabolism may differ between sexes due to variations in cytochrome P450 enzyme expression and activity<sup>79</sup>, a process found over-expressed in women with smoking exposure and COPD but not in men. Mitochondrial functional pathways were altered in women but not men, potentially representing key drivers of sex-specific lung disease pathophysiology<sup>62</sup>, whereas processes such as sumoylation of transcription cofactors, processing of capped intron-containing pre-mRNA, and base excision repair were downregulated in men but not women.

Another illustrative example is the significant co-occurrence between pancreatic cancer (C25) and T2D (E11)<sup>80</sup>. Interestingly, men and women with pancreatic cancer shared a high number of altered pathways in the same direction (Jaccard indices of 0.78 and 0.58 for up- and down-regulated pathways, respectively), whereas the overlaps for T2D were minimal (0.03 and 0, see Supplementary Table 6). These findings indicate that sex



**Fig. 4 | Pathways shared between comorbid diseases in women or men only.**

Heatmap with disease pairs in rows and Reactome categories in columns representing the percentages of pathways in each category that are up- (blue) or down-regulated (red) in both diseases. The color of the squares following the ICD-10 codes denotes the category of disease to which they belong (ICD-10), highlighting rows

with two different colors reflect comorbidities between diseases from different categories. Each square is divided into two triangles; the one on the top (bottom) refers to women (men). For specific pathways enriched in comorbid diseases in women and men, see <https://disease-perception.bsc.es/shdc/>.



differences are more pronounced in T2D than in pancreatic cancer, suggesting that the mechanisms leading to pancreatic cancer may differ between sexes<sup>81</sup>. In men, T2D and pancreatic cancer shared alterations in extracellular matrix organization pathways—including collagen biosynthesis and modifying enzymes, laminin interactions, and ECM proteoglycans—as well as signal transduction pathways altered in the same direction, none of which were shared in women. Conversely, in women, the two diseases shared immune system pathways (e.g., neutrophil degranulation, antiviral mechanism by IFN-stimulated genes) and metabolic pathways (including sphingolipid de novo biosynthesis and regulation of cholesterol biosynthesis by SREBP (SREBF)). These findings underscore the value of our approach in identifying biological processes that are differentially altered between women and men and may underlie sex-specific patterns of disease development and comorbidity. Nonetheless, substantial knowledge gaps remain regarding the biological differences between the sexes in disease pathogenesis—gaps that must be addressed to fully understand the mechanisms driving sex-specific comorbidity profiles.

### Sex differences in drug effects

In previous work, we identified drugs that may influence disease co-occurrence—either by increasing or decreasing comorbidity risk—highlighting this strategy as a potential avenue for drug repositioning<sup>82</sup>. Building on this, we previously identified patient subgroups in which specific drug associations might contribute to the elevated risk of developing secondary diseases<sup>14</sup>. In the present study, we investigated how drug-disease associations may differ between women and men, potentially explaining sex-specific comorbidity patterns. Given the scarcity of studies examining sex-specific gene expression changes following drug exposure, we extracted drug targets from DrugBank<sup>33</sup>, expanded these associations using a protein-protein interaction network<sup>34</sup>, and performed enrichment analyses<sup>28</sup> to identify drugs whose expanded targets were significantly over- or under-expressed across diseases in a sex-dependent manner. In total, 3878 DrugBank IDs were significantly enriched in at least one disease. Of these, 616 were linked through 3997 associations to 568 ICD-10 codes in the SIDER database<sup>35</sup>. Focusing on diseases with sufficient sample sizes in both sexes, 43 of 59 diseases had at least one enriched drug associated with them in SIDER. Among the top 10 drugs associated with the largest number of diseases in women and men, only three were shared—metformin, clofarabine and irinotecan—agents used to treat diabetes, acute lymphoblastic leukemia, and metastatic cancers. Other highly connected drugs included arginine, bortezomib, and carfilzomib in women, and dexrazoxane, etoposide, and idarubicin in men—all used to treat high blood pressure, cardiomyopathies, or cancer. Seven diseases showed no overlap in drug associations between sexes—*Aspergillus* colonization of lung allograft, T1D, amyotrophic lateral sclerosis, interstitial lung disease, rosacea, connective tissue disorders, and axial spondyloarthritis (see Supplementary Fig. 8). In contrast, eight diseases exhibited a Jaccard index >0.5: pancreatic, colon, and kidney cancers; neoplasms of uncertain behavior of lymphoid, hematopoietic and related tissues; oral dysplasia; Job's syndrome; thalassemia; and IBS. These findings highlight substantial heterogeneity in the extent of drug overlap between sexes across the diseases. Overall, women had more drug associations in 19 diseases, whereas men had more in 20. These differences were not explained by disparities in sample size between sexes (correlation  $p$  value = 0.64). For instance, in schizophrenia, 3.13 more drugs were identified in women than in men, despite there being 2.7 more men than women samples. Analysis of drugs enriched across diseases in women and men revealed notable sex-specific differences in drug-disease relationships (see Fig. 5 and Supplementary Fig. 9).

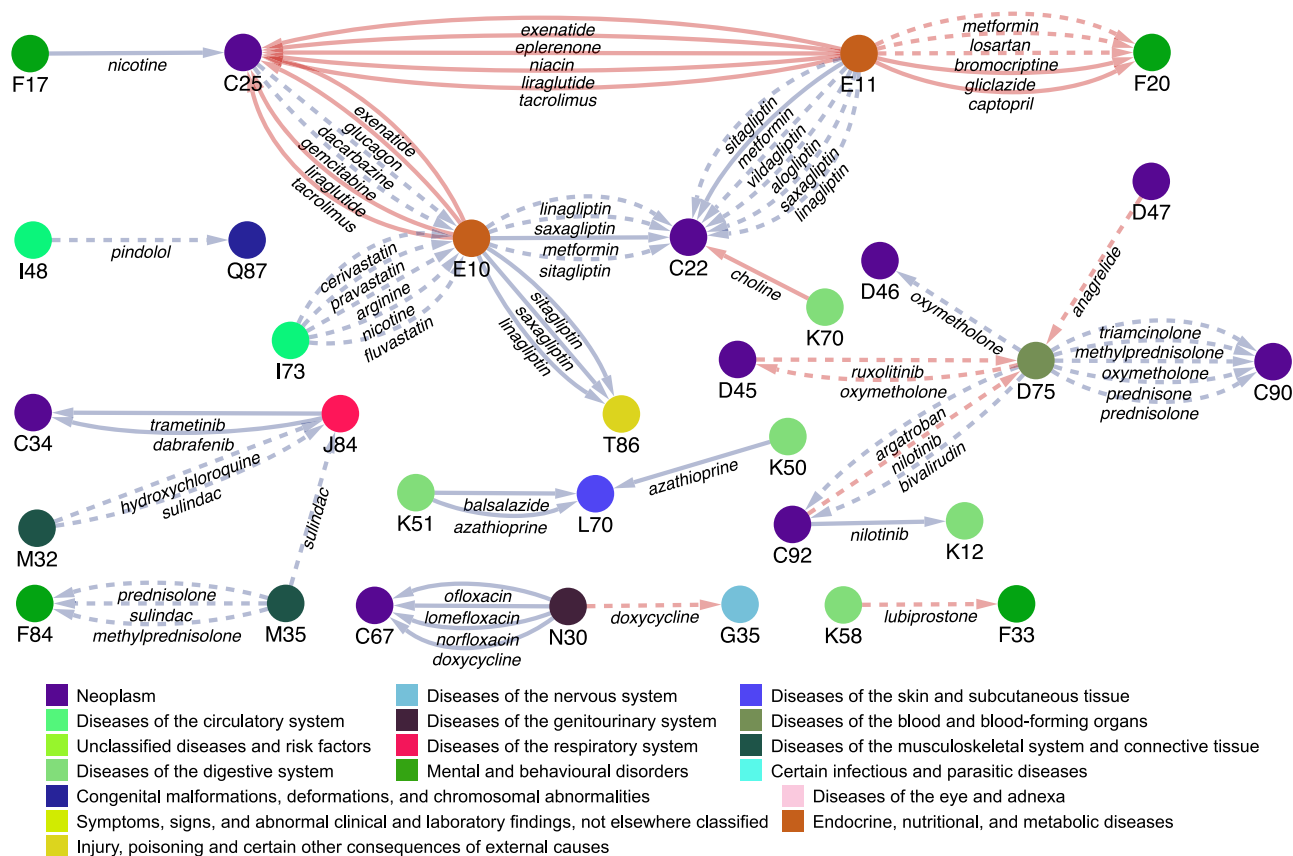
Drug-mediated disease associations may arise for several reasons: (i) the two diseases share symptoms, (ii) patients with one disease also have the other, (iii) treatment for one disease contributes to the onset of the other, or (iv) the drug used to treat one disease could be used to treat the other. In the first scenario, IBS (K58) and major depression (F33)—two comorbid conditions with shared pathophysiological mechanisms<sup>83</sup>—were linked by lubiprostone, a drug used to treat constipation, a common manifestation of

both diseases<sup>84</sup> (see Fig. 5 and Supplementary Data 5). Lubiprostone was indicated for IBS and was enriched in the differential expression profile of major depression in women but not men. Notably, constipation occurs more frequently in women with depression than in men<sup>85</sup>. In the second scenario, the association between T2D (E11) and schizophrenia (F20)—where diabetes medications such as metformin and gliclazide were enriched in women with schizophrenia—may suggest that these patients were taking these drugs or had coexisting T2D. This interpretation is supported by the high representation of elderly schizophrenia patients in our dataset, consistent with previous reports that late-life schizophrenia is associated with a higher prevalence in women than in men (35 vs. 21.53%)<sup>86</sup>. In the third scenario, patients with essential thrombocythemia are known to have an elevated risk of progression to myelofibrosis, particularly when treated with anagrelide<sup>87</sup>—an effect more pronounced in women. Consistent with this, anagrelide (indicated for essential thrombocythemia (D47)) was enriched in women with myelofibrosis (D75). However, the absence of detailed meta-data prevents further confirmation of this hypothesis. In the fourth scenario, patients with type 2 diabetes have been observed to be at greater risk of developing liver cancer, with the risk being higher in men<sup>88</sup>. One hypothesis to explain this sex difference involves higher circulating levels of adiponectin in women<sup>89</sup>. Notably, both diseases are linked through metformin, which is indicated for the treatment of type 2 diabetes and whose targets are significantly upregulated in liver cancer. Interestingly, metformin has been described as exerting a protective effect against the development of liver cancer, with this effect being stronger in men<sup>90–92</sup>. A possible explanation for the sex difference in metformin's protective effect could lie in the hormonal context: higher baseline adiponectin levels in women may already confer protection, whereas in men, metformin partly compensates for their lower adiponectin levels and higher baseline risk. These observations underscore the importance of considering hormonal context when evaluating potential drug repositioning strategies.

Additional pharmacologically supported comorbidity relationships not reported by ref. 9 included the association between T1D and asthma<sup>93</sup>, where the risk is higher in boys compared to girls<sup>94</sup>. We identified 15 T1D-enriched drugs indicated for asthma in men—including salbutamol and salmeterol—that were not observed in women. Notably, salbutamol use has been linked to elevated blood glucose levels<sup>95</sup>, emphasizing the importance of considering both sex and age when prescribing treatment. Similarly, two drugs—risperidone and citalopram—were significantly enriched in women with schizophrenia and have also been associated with Alzheimer's disease. Risperidone is an atypical antipsychotic used to treat schizophrenia<sup>96</sup> and behavioral symptoms in Alzheimer's disease patients<sup>97</sup>, whereas citalopram is an antidepressant prescribed for depression and negative symptoms in schizophrenia<sup>98</sup> as well as agitation in Alzheimer's disease<sup>99</sup>. Notably, as previously mentioned, schizophrenia patients face a higher risk of developing Alzheimer's disease—particularly among women<sup>44</sup>—and citalopram response has also been shown to differ by sex<sup>100</sup>. Together, these findings support the hypothesis that drug-disease associations vary between women and men and may contribute to sex-specific comorbidity patterns. Nevertheless, additional studies are needed to directly investigate molecular-level differences in drug effects as a function of patient sex.

### Discussion

Calculating similarities between diseases using molecular data is a well-established approach for understanding disease co-occurrence and for identifying opportunities for drug repositioning<sup>13,18,101</sup>. Among various molecular data types, transcriptomics has emerged as a particularly promising source, given its strong capacity to reveal biological mechanisms underlying comorbidity relationships. Transcriptomic analyses have proven useful for elucidating both direct and inverse comorbidity relationships, identifying candidate drugs for repurposing, and detecting disease subtype-specific molecular similarities<sup>14,15,82,102,103</sup>. However, despite the clear physical and physiological differences between women and men—and their evident impact on disease development and comorbidity—sex differences in transcriptional similarities between diseases have not been systematically



**Fig. 5 | Sex-specific disease–drug associations.** The nodes represent diseases, colored according to the disease category (ICD-10) to which they belong. An edge connects two diseases if the drug indicated to treat one disease (source node) is enriched in the differential expression profile of another disease (sink node). Only associations between comorbid diseases<sup>9</sup> from different ICD-10 categories are shown. The name of the drug connecting both diseases is indicated below the edge.

Only associations detected exclusively in one sex are shown, with red (blue) edges denoting associations detected only in females (males). Solid (dashed) lines indicate that the normalized enrichment score of the GSEA<sup>28</sup> is positive (negative), for all other associations see <https://disease-perception.bsc.es/shdc/> and Supplementary Fig. 9. The network's visualizations have been generated using Cytoscape software<sup>112</sup>.

investigated. Although public transcriptomic databases have historically lacked reliable sex annotations<sup>104</sup>, advances in computational methods now enable accurate inference of sample sex, making such analyses feasible.

In this study, we generated disease networks separately for women and men and observed that diseases cluster differently by sex based on their differential expression profiles—consistent with previous PheWAS-based findings<sup>105</sup>. Notably, the clusters in women were more heterogeneous across disease categories, suggesting differences in multimorbidity patterns between the sexes. The relevance of these observations is supported by the strong concordance between sex-specific transcriptional similarities and epidemiologically observed disease co-occurrences. For example, type 1 diabetes and liver cancer co-occur frequently in men, while schizophrenia and COPD do so in women. In both cases, these patterns correlate with sex-specific alterations in metabolic and immune system-related biological processes.

Collectively, our findings highlight the importance of analyzing comorbidity patterns separately in women and men and investigating their underlying molecular mechanisms—an approach that may ultimately inform more effective treatments.

Historically, biomedical research has predominantly focused on men, contributing to diagnostic biases and suboptimal therapeutic strategies for women. Moreover, women remain underrepresented in clinical trials, resulting in poorer therapeutic optimization and health outcomes<sup>106</sup>. Combined with the limited inclusion of comorbid and multimorbid patients in clinical trials<sup>107</sup>, this underrepresentation may contribute to future challenges in the safe and effective use of medications. Therefore, incorporating sex-based molecular and comorbidity differences into research and clinical

guidelines is essential for developing safer and more personalized medical practices<sup>108</sup>.

While our study is currently limited by the availability and quality of sex-specific information across public databases, ongoing efforts to improve data annotation and increase the representation of both sexes in transcriptomic studies will progressively strengthen future analyses. Likewise, as more population-level studies systematically investigate comorbidity relationships by sex, it will become increasingly feasible to validate and integrate our findings within DTSNs, enhancing their robustness and translational relevance.

Although a few studies have explored sex-based comorbidity differences<sup>7</sup>, many of these datasets—such as the extensive US medical claims-based network describing fourfold more comorbidities than ref. 9—are no longer publicly available (see Supplementary Note 5). Furthermore, most transcriptomic studies focus on individual diseases and lack key contextual metadata such as comorbidities, medication use, and patient age—all of which may influence molecular profiles. Other relevant variables, such as ethnicity or socioeconomic status—known to significantly affect disease development and comorbidity patterns<sup>109,110</sup>—could not be incorporated due to the absence of such metadata. Additionally, for comparison with epidemiological networks, we standardized disease names to three-digit ICD-10 codes, which was necessary for studying sex-related comorbidity patterns. Consequently, certain disease subtypes—such as lung cancer, non-small cell lung cancer, and basaloid lung cancer—were grouped under the same ICD-10 code (C34), potentially obscuring subtype-specific differences. Future research should therefore aim to investigate sex differences at the level of specific disease subtypes. Finally, greater availability of



molecular data would enhance the statistical power of such analyses. The growing number of biobanks integrating molecular profiles with electronic health records may, in the future, enable more comprehensive studies of this kind<sup>18</sup>.

In summary, this study reinforces the marked differences between the sexes in the development of diseases and comorbidities previously described at the epidemiological level. It also generates multiple molecular hypotheses regarding sex-specific differences in comorbidity relationships, paving the way for future experimental validation. Future work should also explore, from a molecular perspective, how aging influences the development of comorbidities and their sex-specific differences, while integrating additional data sources to refine and expand the map of disease interrelationship generated here.

## Data availability

All data needed to understand and assess the conclusions of this research are available in the main text, supplementary materials, the Disease-PERCEPTION portal (<https://disease-perception.bsc.es/shdc/>) and <https://github.com/jonsv89/SHDC/blob/main/Datasets.txt>. The raw gene expression datasets are publicly available and can be downloaded from GEO and ArrayExpress (<https://www.ebi.ac.uk/arrayexpress>). The identifiers for each of the studies are provided in Supplementary Data 1. The numerical values required to generate the main figures are provided in Supplementary Data 2. The source data for Figs. 2–5 can be found in Supplementary Data 2. Each sheet specifically indicates which figure has been generated using the data contained therein.

## Code availability

The code developed to perform the analyses described is available at <https://github.com/jonsv89/SHDC><sup>27</sup> and <https://github.com/bsc-life/SHDC><sup>11</sup>.

Received: 25 July 2025; Accepted: 10 December 2025;

Published online: 30 December 2025

## References

- Valderas, J. M., Starfield, B., Sibbald, B., Salisbury, C. & Roland, M. Defining comorbidity: implications for understanding health and health services. *Ann. Fam. Med.* **7**, 357–363 (2009).
- Feinstein, A. R. The pre-therapeutic classification of co-morbidity in chronic disease. *J. Chronic Dis.* **23**, 455–468 (1970).
- Leibson, C. L. et al. Risk of dementia among persons with diabetes mellitus: a population-based cohort study. *Am. J. Epidemiol.* **145**, 301–308 (1997).
- Pandey, A. et al. Diabetes mellitus and the risk of cancer: results from a large-scale population-based cohort study in Japan. *Arch. Intern. Med.* **166**, 187–209 (2006).
- McDonald, C. J. The barriers to electronic medical record systems and how to overcome them. *J. Am. Med. Inform. Assoc.* **4**, 213–221 (1997).
- Kolodner, R. M. *Computerizing Large Integrated Health Networks* (Springer, 1997).
- Hidalgo, C. A., Blumm, N., Barabási, A.-L. & Christakis, N. A. A dynamic network approach for the study of human phenotypes. *PLoS Comput. Biol.* **5**, e1000353 (2009).
- Jensen, A. B. et al. Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. *Nat. Commun.* **5**, 4022 (2014).
- Westergaard, D., Moseley, P., Sørup, F. K. H., Baldi, P. & Brunak, S. Population-wide analysis of differences in disease progression patterns in men and women. *Nat. Commun.* **10**, 666 (2019).
- Goh, K.-I. et al. The human disease network. *Proc. Natl. Acad. Sci. USA* **104**, 8685–8690 (2007).
- Lee, D. S. et al. The implications of human metabolic network topology for disease comorbidity. *Proc. Natl. Acad. Sci. USA* **105**, 9880–9885 (2008).
- Menche, J. et al. Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science* **347**, 841 (2015).
- Morselli Gysi, D. & Barabási, A.-L. Noncoding RNAs improve the predictive power of network medicine. *Proc. Natl. Acad. Sci. USA* **120**, e2301342120 (2023).
- Sánchez-Valle, J. et al. Interpreting molecular similarity between patients as a determinant of disease comorbidity relationships. *Nat. Commun.* **11**, 2854 (2020).
- Urda-García, B., Sánchez-Valle, J., Lepore, R. & Valencia, A. Patient stratification reveals the molecular basis of disease co-occurrences. *Proc. Natl. Acad. Sci. USA* **122**, e2421060122 (2025).
- Dong, G., Feng, J., Sun, F., Chen, J. & Zhao, X. M. A global overview of genetically interpretable multimorbidities among common diseases in the UK Biobank. *Genome Med.* **13**, 110 (2021).
- Murrin, O. et al. A systematic analysis of the contribution of genetics to multimorbidity and comparisons with primary care data. *eBioMedicine* **113**, 105584 (2025).
- Sánchez-Valle, J. & Valencia, A. Molecular bases of comorbidities: present and future perspectives. *Trends Genet.* **39**, 773–786 (2023).
- Aguet, F. et al. The impact of sex on gene expression across human tissues. *Science* **369**, eaba3066 (2020).
- Liu, L. Y., Schaub, M. A., Sirota, M. & Butte, A. J. Sex differences in disease risk from reported genome-wide association study findings. *Hum. Genet.* **131**, 353–364 (2012).
- Lopes-Ramos, C. M. et al. Sex differences in gene expression and regulatory networks across 29 human tissues. *Cell Rep.* **31**, 107795 (2020).
- McCall, M. N., Murakami, P. N., Lukk, M., Huber, W. & Irizarry, R. A. Assessing affymetrix GeneChip microarray quality. *BMC Bioinformatics* **12**, 137 (2011).
- Buckberry, S., Bent, S. J., Bianco-Miotto, T. & Roberts, C. T. massIR: a method for predicting the sex of samples in gene expression microarray datasets. *Bioinformatics* **30**, 2084–2085 (2014).
- Hubbell, E., Liu, W.-M. & Mei, R. Robust estimators for expression analysis. *Bioinformatics* **18**, 1585–1592 (2002).
- McCall, M. N., Jaffee, H. A. & Irizarry, R. A. fRMA ST: frozen robust multiarray analysis for Affymetrix Exon and Gene ST arrays. *Bioinformatics* **28**, 3153–3154 (2012).
- Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
- Jon & Carpintero, I. N. jonsv89/SHDC: v1.0.0. *Zenodo* <https://doi.org/10.5281/zenodo.17724410> (2025).
- Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
- Murtagh, F. & Legendre, P. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *J. Classif.* **31**, 274–295 (2014).
- Suzuki, R. & Shimodaira, H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540–1542 (2006).
- Galili, T. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* **31**, 3718–3720 (2015).
- Simas, T., Correia, R. B. & Rocha, L. M. The distance backbone of complex networks. *J. Complex Netw.* **9**, 1–35 (2021).
- Wishart, D. S. et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074–D1082 (2018).
- Kotlyar, M. et al. IID 2021: towards context-specific protein interaction analyses by increased coverage, enhanced annotation and enrichment analysis. *Nucleic Acids Res.* **50**, D640–D647 (2022).

35. Kuhn, M., Letunic, I., Jensen, L. J. & Bork, P. The SIDER database of drugs and side effects. *Nucleic Acids Res.* **44**, D1075–D1079 (2016).
36. Bodenreider, O. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic Acids Res.* **32**, D267–D270 (2004).
37. Kelly, J., Moyeed, R., Carroll, C., Luo, S. & Li, X. Genetic networks in Parkinson's and Alzheimer's disease. *Aging* **12**, 5221–5243 (2020).
38. Cunnane, S. et al. Brain fuel metabolism, aging, and Alzheimer's disease. *Nutrition* **27**, 3–20 (2011).
39. Kucukgoncu, S. et al. Glucose metabolism dysregulation at the onset of mental illness is not limited to first episode psychosis: a systematic review and meta-analysis. *Early Interv. Psychiatry* **13**, 1021–1031 (2019).
40. Harris, L. D., Jasem, S. & Licchesi, J. D. F. The ubiquitin system in Alzheimer's disease. *Adv. Exp. Med. Biol.* **1233**, 195–221 (2020).
41. Bousman, C. A. et al. Elevated ubiquitinated proteins in brain and blood of individuals with schizophrenia. *Sci. Rep.* **9**, 2307 (2019).
42. Eggers, A. E. A serotonin hypothesis of schizophrenia. *Med. Hypotheses* **80**, 791–794 (2013).
43. Fuhrer, T. E. et al. Impaired expression of GABA transporters in the human Alzheimer's disease hippocampus, subiculum, entorhinal cortex and superior temporal gyrus. *Neuroscience* **351**, 108–118 (2017).
44. Cai, L. & Huang, J. Schizophrenia and risk of dementia: a meta-analysis study. *Neuropsychiatr. Dis. Treat.* **14**, 2047–2055 (2018).
45. Mirlekar, B. & Pylayeva-Gupta, Y. IL-12 family cytokines in cancer and immunotherapy. *Cancers* **13**, 167 (2021).
46. Ye, Y. et al. Sex-associated molecular differences for cancer immunotherapy. *Nat. Commun.* **11**, 1779 (2020).
47. Katsanos, K. H., Roda, G., McBride, R. B., Cohen, B. & Colombel, J.-F. Increased risk of oral cancer in patients with inflammatory bowel diseases. *Clin. Gastroenterol. Hepatol.* **14**, 413–420 (2016).
48. Katsanos, K. H., Roda, G., Brygo, A., Delaporte, E. & Colombel, J.-F. Oral cancer and oral precancerous lesions in inflammatory bowel diseases: a systematic review. *J. Crohns Colitis* **9**, 1043–1052 (2015).
49. Harishankar, M. K., Mohan, A. M., Krishnan, A. V. & Devi, A. Downregulation of Notch4 - a prognostic marker in distinguishing oral verrucous carcinoma from oral squamous cell carcinoma. *Braz. J. Otorhinolaryngol.* **85**, 11–16 (2019).
50. Lu, Y., Li, X., Liu, S., Zhang, Y. & Zhang, D. Toll-like receptors and inflammatory bowel disease. *Front. Immunol.* **9**, 72 (2018).
51. Astbury, J. Gender disparities in mental health. In: *Mental health. Ministerial Round Tables 2001, 54th World Health Assembly 73–92* (WHO, 2001).
52. Yaqubi, K. et al. Inflammatory bowel disease is associated with an increase in the incidence of multiple sclerosis: a retrospective cohort study of 24,934 patients. *Eur. J. Med. Res.* **29**, 186 (2024).
53. Lin, J.-C., Lin, C.-S., Hsu, C.-W., Lin, C.-L. & Kao, C.-H. Association between Parkinson's disease and inflammatory bowel disease: a nationwide Taiwanese retrospective cohort study. *Inflamm. Bowel Dis.* **22**, 1049–1055 (2016).
54. Liu, N. et al. Inflammatory bowel disease and risk of dementia: an updated meta-analysis. *Front. Aging Neurosci.* **14**, 962681 (2022).
55. Hsu, J.-H., Chien, I.-C., Lin, C.-H., Chou, Y.-J. & Chou, P. Increased risk of chronic obstructive pulmonary disease in patients with schizophrenia: a population-based study. *Psychosomatics* **54**, 345–351 (2013).
56. Krieger, I., Tzur Bitan, D., Comaneshter, D., Cohen, A. & Feingold, D. Increased risk of smoking-related illnesses in schizophrenia patients: a nationwide cohort study. *Schizophr. Res.* **212**, 121–125 (2019).
57. Ruiz-Rull, C. et al. Low lung function in bipolar disorder and schizophrenia: a hidden risk. *Front. Physiol.* **15**, 1335798 (2024).
58. Obeidat, M. et al. Molecular mechanisms underlying variations in lung function: a systems genetics analysis. *Lancet Respir. Med.* **3**, 782–795 (2015).
59. Li, R., Ma, X., Wang, G., Yang, J. & Wang, C. Why sex differences in schizophrenia? *J. Transl. Neurosci.* **1**, 37–42 (2016).
60. Ogdie, A. et al. Risk of incident liver disease in patients with psoriasis, psoriatic arthritis, and rheumatoid arthritis: a population-based study. *J. Invest. Dermatol.* **138**, 760–767 (2018).
61. Yu, Z. et al. Sex-specific differences in the transcriptome of the human dorsolateral prefrontal cortex in schizophrenia. *Mol. Neurobiol.* **60**, 1083–1098 (2023).
62. Glass, K. et al. Sexually-dimorphic targeting of functionally-related genes in COPD. *BMC Syst. Biol.* **8**, 118 (2014).
63. Benjamin, K. J. M. et al. Sex affects transcriptional associations with schizophrenia across the dorsolateral prefrontal cortex, hippocampus, and caudate nucleus. *Nat. Commun.* **15**, 3980 (2024).
64. Troianova, N. et al. Impact of sex on circulating leukocytes composition in COPD patients. *Int. J. Chron. Obstruct. Pulmon. Dis.* **16**, 3539–3550 (2021).
65. Lohasz, C. et al. A microphysiological cell-culturing system for pharmacokinetic drug exposure and high-resolution imaging of arrays of 3D microtissues. *Front. Pharmacol.* **12**, 785851 (2021).
66. Langhammer, A., Johnsen, R., Gulsvik, A., Holmen, T. L. & Bjerner, L. Sex differences in lung vulnerability to tobacco smoking. *Eur. Respir. J.* **21**, 1017–1023 (2003).
67. Kim, Y. S. & Kim, N. Sex-gender differences in irritable bowel syndrome. *J. Neurogastroenterol. Motil.* **24**, 544–558 (2018).
68. Meleine, M. & Matricon, J. Gender-related differences in irritable bowel syndrome: potential mechanisms of sex hormones. *World J. Gastroenterol.* **20**, 6725–6743 (2014).
69. Yang, C. X. et al. Widespread sexual dimorphism in the transcriptome of human airway epithelium in response to smoking. *Sci. Rep.* **9**, 17600 (2019).
70. Koo, H.-K. et al. Sex-specific associations with DNA methylation in lung tissue demonstrate smoking interactions. *Epigenetics* **16**, 692–703 (2021).
71. Hoang, N., Brooks, K. & Edwards, K. Sex-specific colonic mitochondrial dysfunction in the indomethacin-induced rat model of inflammatory bowel disease. *Front. Physiol.* **15**, 1341742 (2024).
72. Millett, E. R. C., Peters, S. A. E. & Woodward, M. Sex differences in risk factors for myocardial infarction: cohort study of UK Biobank participants. *BMJ* **363**, k4247 (2018).
73. Carstensen, B. et al. Cancer incidence in persons with type 1 diabetes: a five-country study of 9000 cancers in type 1 diabetic individuals. *Diabetologia* **59**, 980–988 (2016).
74. Williams, J. M., Poudel, B. & Shields, C. A. in *Sex Differences in Cardiovascular Physiology and Pathophysiology* (eds LaMarca, B. & Alexander, B. T.) Ch. 15 (Academic Press, 2019).
75. Klein, S. L. & Flanagan, K. L. Sex differences in immune responses. *Nat. Rev. Immunol.* **16**, 626–638 (2016).
76. Mauvais-Jarvis, F. Sex differences in energy metabolism: natural selection, mechanisms and consequences. *Nat. Rev. Nephrol.* **20**, 56–69 (2024).
77. Sorge, R. E. et al. Different immune cells mediate mechanical pain hypersensitivity in male and female mice. *Nat. Neurosci.* **18**, 1081–1083 (2015).
78. Adeloye, D. et al. Global, regional, and national prevalence of, and risk factors for, chronic obstructive pulmonary disease (COPD) in 2019: a systematic review and modelling analysis. *Lancet Respir. Med.* **10**, 447–458 (2022).
79. Barnes, P. J. Sex differences in chronic obstructive pulmonary disease mechanisms. *Am. J. Respir. Crit. Care Med.* **193**, 813–814 (2016).
80. Shen, B. et al. Association between age at diabetes onset or diabetes duration and subsequent risk of pancreatic cancer: results

- from a longitudinal cohort and mendelian randomization study. *Lancet Reg. Health West. Pac.* **30**, 100596 (2023).
81. Kautzky-Willer, A., Leutner, M. & Harreiter, J. Sex differences in type 2 diabetes. *Diabetologia* **66**, 986–1002 (2023).
82. Sánchez-Valle, J. et al. A molecular hypothesis to explain direct and inverse co-morbidities between Alzheimer's disease, glioblastoma and lung cancer. *Sci. Rep.* **7**, 4474 (2017).
83. Mudyanadzo, T. A., Hauzaree, C., Yerokhina, O., Architha, N. N. & Ashqar, H. M. Irritable bowel syndrome and depression: a shared pathogenesis. *Cureus* **10**, e3178 (2018).
84. Lacy, B. E. & Levy, L. C. Lubiprostone: a novel treatment for chronic constipation. *Clin. Interv. Aging* **3**, 357–364 (2008).
85. Wang, P., Shen, X., Wang, Y. & Jia, X. Association between constipation and major depression in adult Americans: evidence from NHANES 2005–2010. *Front. Psychiatry* **14**, 1152435 (2023).
86. Huo, L. et al. Diabetes in late-life schizophrenia: prevalence, factors, and association with clinical symptoms. *J. Psychiatr. Res.* **132**, 44–49 (2021).
87. Birgegård, G. et al. Treatment of essential thrombocythemia in Europe: a prospective long-term observational study of 3649 high-risk patients in the evaluation of anagrelide efficacy and long-term safety study. *Haematologica* **103**, 51–60 (2018).
88. Wang, Y. et al. Type 2 diabetes and gender differences in liver cancer by considering different confounding factors: a meta-analysis of cohort studies. *Ann. Epidemiol.* **26**, 764–772 (2016).
89. Manieri, E. et al. Adiponectin accounts for gender differences in hepatocellular carcinoma incidence. *J. Exp. Med.* **216**, 1108–1119 (2019).
90. Chen, H.-P. et al. Metformin decreases hepatocellular carcinoma risk in a dose-dependent manner: population-based and in vitro studies. *Gut* **62**, 606–615 (2013).
91. Cunha, V., Cotrim, H. P., Rocha, R., Carvalho, K. & Lins-Kusterer, L. Metformin in the prevention of hepatocellular carcinoma in diabetic patients: a systematic review. *Ann. Hepatol.* **19**, 232–237 (2020).
92. Lee, M.-S. et al. Type 2 diabetes increases and metformin reduces total, colorectal, liver and pancreatic cancer incidences in Taiwanese: a representative population prospective cohort study of 800,000 individuals. *BMC Cancer* **11**, 20 (2011).
93. Smew, A. I., Lundholm, C., Säwendahl, L., Lichtenstein, P. & Almqvist, C. Familial coaggregation of asthma and type 1 diabetes in children. *JAMA Netw. Open* **3**, e200834 (2020).
94. Hsiao, Y.-T. et al. Type 1 diabetes and increased risk of subsequent asthma: a nationwide population-based cohort study. *Medicine* **94**, e1466 (2015).
95. Mortimer, C. A preliminary study to determine the effects of nebulised salbutamol on blood glucose levels during an acute asthma exacerbation. *Eur. Respir. J.* **54**, PA4243 (2019).
96. Marder, S. R. & Meibach, R. C. Risperidone in the treatment of schizophrenia. *Am. J. Psychiatry* **151**, 825–835 (1994).
97. Onor, M. L., Saina, M., Trevisiol, M., Cristante, T. & Aguglia, E. Clinical experience with risperidone in the treatment of behavioral and psychological symptoms of dementia. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **31**, 205–209 (2007).
98. Goff, D. C. et al. Citalopram in first episode schizophrenia: the DECIFER trial. *Schizophr. Res.* **208**, 331–337 (2019).
99. Porsteinsson, A. P. et al. Effect of citalopram on agitation in Alzheimer disease: the CitAD randomized clinical trial. *JAMA* **311**, 682–691 (2014).
100. Young, E. A. et al. Sex differences in response to citalopram: a STAR\*D report. *J. Psychiatr. Res.* **43**, 503–511 (2009).
101. Park, J., Lee, D.-S., Christakis, N. A. & Barabási, A.-L. The impact of cellular networks on disease comorbidity. *Mol. Syst. Biol.* **5**, 262 (2009).
102. Sirota, M. et al. Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Sci. Transl. Med.* **3**, 96ra77 (2011).
103. He, H. et al. Computational drug repurposing by exploiting large-scale gene expression data: strategy, methods and applications. *Comput. Biol. Med.* **155**, 106671 (2023).
104. Ruiz-Serra, V. et al. Analyzing sex imbalance in EGA and dbGaP biological databases: recommendations for better practices. *iScience* **27**, 110831 (2024).
105. Sriram, V., Woerner, J., Ahn, Y.-Y. & Kim, D. The interplay of sex and genotype in disease associations: a comprehensive network analysis in the UK Biobank. *Hum. Genomics* **19**, 4 (2025).
106. Gualtierotti, R. Bridging the gap: time to integrate sex and gender differences into research and clinical practice for improved health outcomes. *Eur. J. Intern. Med.* **134**, 9–16 (2025).
107. Hanlon, P. et al. Representation of people with comorbidity and multimorbidity in clinical trials of novel drug therapies: an individual-level participant data analysis. *BMC Med.* **17**, 201 (2019).
108. Gracia Gutiérrez, A., Poblador-Plou, B., Prados-Torres, A., Ruiz Laiglesia, F. J. & Gimeno-Miguel, A. Sex differences in comorbidity, therapy, and health services' use of heart failure in Spain: evidence from real-world data. *Int. J. Environ. Res. Public Health* **17**, 2136 (2020).
109. Kuan, V. et al. Identifying and visualising multimorbidity and comorbidity patterns in patients in the English National Health Service: a population-based study. *Lancet Digit. Health* **5**, e16–e27 (2023).
110. Barnett, K. et al. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet* **380**, 37–43 (2012).
111. Jon & Fernández, J. M. bsc-life/SHDC: v1.0.0. *Zenodo* <https://doi.org/10.5281/ZENODO.17724395> (2025).
112. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

## Acknowledgements

We thank Beatriz Urda (Barcelona Supercomputing Center) for the critical reading of the manuscript. This publication has been funded by the project PID2022-141809OB-I00, “Health trajectories and the role of sex, gender, ageing, and treatments in comorbidities”, funded by MICIU/AEI/10.13039/501100011033/10.13039/501100011033 and by FEDER, EU. J.S.-V. was funded by the European Community's Horizon Europe Program under grant agreement No. 101136957 (COMMUTE). Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency (HADEA). Neither the European Union nor the granting authority can be held responsible for them. L.M.R. was partially funded by the National Institutes of Health (NIH), National Library of Medicine (grants R01-LM011945 and R01-LM012832), by a Fulbright Commission fellowship, by the National Science Foundation Research Traineeship “Interdisciplinary Training in Complex Networks and Systems” (grant 1735095), and by the European Union under Grant Agreement No. 101186558. J.S.-V., F.X.C., and L.M.R. were funded by the Fundação para a Ciência e a Tecnologia Grant No. 2022.09122.PTDC <https://doi.org/10.54499/2022.09122.PTDC>. J.C.-C. was hired under the Generation D initiative, promoted by Red.es, an entity attached to the Ministry for Digital Transformation and the Civil Service, to attract and retain talent through scholarships and training contracts, financed by the Recovery, Transformation and Resilience Plan through Next Generation funds. M.F.-R. and R.T.-S. were funded by the Ministry of Education of the Valencian Regional Government (PROMETEO/CIPROM/2022/58). R.T.-S. was funded by the Spanish Ministry of Science, Innovation and Universities (PID2021-129099OB-I00). The funders had no role in study

design, data collection and analysis, publication decisions, or manuscript preparation.

### Author contributions

A.V. and J.S.-V. designed all experiments. J.S.-V., M.F.-R., F.X.C., J.C.-C., and I.N.-C. performed the experiments. D.C., L.M.R., J.C.-C., and R.T.-S. provided technical advice. J.S.-V., D.C., and A.V. wrote the manuscript. All authors discussed the results and commented on the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s43856-025-01329-0>.

**Correspondence** and requests for materials should be addressed to Jon Sánchez-Valle or Alfonso Valencia.

**Peer review information** *Communications Medicine* thanks Lei Guo and Woo Ri Chae for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025