

Sex-specific transcriptome similarity networks elucidate comorbidity relationships

Jon Sánchez-Valle^{1*}, María Flores-Rodero^{1,2}, Felipe Xavier Costa^{3,4}, Jose Carbonell-Caballero¹, Iker Núñez-Carpintero^{1,5}, Rafael Tabarés-Seisdedos², Luis Mateus Rocha^{3,4}, Davide Cirillo⁵ and Alfonso Valencia^{1,6*}

¹Computational Biology, Barcelona Supercomputing Center, Barcelona, 08034, Spain

²Department of Medicine, University of Valencia, CIBERSAM, INCLIVA, 46010, Valencia, Spain

³Universidade Católica Portuguesa, Católica Medical School, Católica Biomedical Research Centre, 1649-023 Lisbon, Portugal

⁴School of Systems Science and Industrial Engineering, Binghamton University (State University of New York), Binghamton, NY 13902, USA

⁵Machine Learning for Biomedical Research, Barcelona Supercomputing Center, Barcelona, 08034, Spain

⁶ICREA, Barcelona, 08010 Spain

*Corresponding authors (jon.sanchez@bsc.es & alfonso.valencia@bsc.es)

Abstract

Humans present sex-driven biological differences. Consequently, the prevalence of analyzing specific diseases and comorbidities differs between the sexes, directly impacting patients' management and treatment. Despite its relevance and the growing evidence of said differences across numerous diseases (with 4,370 PubMed results published within the past year), knowledge at the comorbidity level remains limited. In fact, to date, no study has attempted to identify the biological processes altered differently in women and men, promoting differences in comorbidities. To shed light on this problem, we analyze expression data for more than 100 diseases from public repositories, analyzing each sex independently. We calculate similarities between differential expression profiles by disease pairs and find that 13-16% of transcriptomically similar disease pairs are sex-specific. By comparing these results with epidemiological evidence, we recapitulate 53-60% of known comorbidities distinctly described for men and women, finding sex-specific transcriptomic similarities between sex-specific comorbid diseases. The analysis of shared underlying pathways shows that diseases can co-occur in men and women by altering alternative biological processes. Finally, we identify different drugs differentially associated with comorbid diseases depending on patients' sex, highlighting the need to consider this relevant variable in the administration of drugs due to their possible influence on comorbidities.

Introduction

Comorbidity and multimorbidity, defined as the presence of more than one medical condition in individuals^{1,2}, have been studied by conducting cohort, case-control, and nested case-control studies. Such studies have described, for example, that patients with type II diabetes (T2D) have a higher prevalence of dementia and cancer, with men specifically showing a higher risk of Alzheimer's disease^{3,4}. However, the development of a digital medical record system⁵ and the implementation of the Veterans Health Information System and Technology Architecture⁶ paved the way for the collection of medical information in the so-called Electronic Health Records (EHR). Since then, several studies have systematically analyzed comorbidity relations and constructed networks based on epidemiological evidence. For example, Hidalgo *et al.* generated a network exploring Medicare data from more than 30 million patients in the United States⁷. They found insightful differences in comorbidity relationships between genders, such as a higher risk of nephropathies in women or acute myocardial infarction in men with T2D. In addition, Jensen *et al.* generated chains of disease diagnoses - temporal disease trajectories - concatenating statistically significant co-occurrences using EHRs from Denmark, identifying key diseases that contribute to the progression of other diseases and increase mortality⁸. In the same population, Westergaard *et al.* found that women present more comorbidities than men and a higher number of disease co-occurrences with an increased timespan between diagnoses⁹. Therefore, those works suggest that men and women present essential differences in disease comorbidities, highlighting the importance of understanding such differences from the molecular point of view.

Since the publication of the first human disease network in 2007 by Goh *et al.*, where diseases were connected if they shared at least one altered gene (diseasome)¹⁰, dozens of works have studied molecular similarities between diseases, some of them trying to understand comorbidities better. Lee *et al.* described that diseases with coupled metabolic reactions co-occur 3 to 7 times more than those without coupling reactions¹¹. Measuring distances between disease modules in the interactome showed that diseases with overlapping modules tend to co-occur more than expected by chance^{12,13}. Transcriptomic similarities between diseases significantly recovered epidemiologically described co-occurrences^{14,15}. Recently, Dong and colleagues combined EHR and Genome-wide association studies (GWAS) data from the UK Biobank to recapitulate 46% of the multimorbidities¹⁶.

We have already pointed out the lack of specific analysis of disease networks in terms of sex differences, despite its potential importance at the scientific and medical level¹⁷, even more when we know that 37% of all genes exhibit sex-biased expression in at least one tissue¹⁸.

Liu *et al.* found disease-associated sex-specific polymorphisms in GWAS ¹⁹. Constructing sample-specific gene regulatory networks from healthy human tissues, Lopes-Ramos *et al.* found that many transcription factors have different regulatory targets depending on the sex of the sample. Interestingly, those differentially targeted genes are enriched in tissue-related functions and diseases. As an example, Alzheimer's disease genes are targeted by different transcription factors depending on the sex of the sample ²⁰.

Given the knowledge gap regarding sex differences at the comorbidity level, we have generated disease transcriptomic similarity networks separately for men and women. To our knowledge, this represents the first analysis using this approach in the field (see Figure 1). To avoid the constant change of terminology (female/male and women/men), throughout the text, we will always use women and men to refer both to transcriptomic data (female/male), and epidemiological data (women/men). The generated networks recover a representative set of known comorbidities previously described for women and men ⁹. Analyzing pathways altered in the same direction in comorbid diseases allows us to propose hypotheses for differences in disease co-occurrence between women and men. In addition, we find that disease pairs may co-occur more than expected by chance in both sexes through different biological processes. Finally, we translate the results to potentially related drugs, highlighting the relevance of understanding the scientific and medical implications of studying sex-specific differences in diseases and comorbidities.

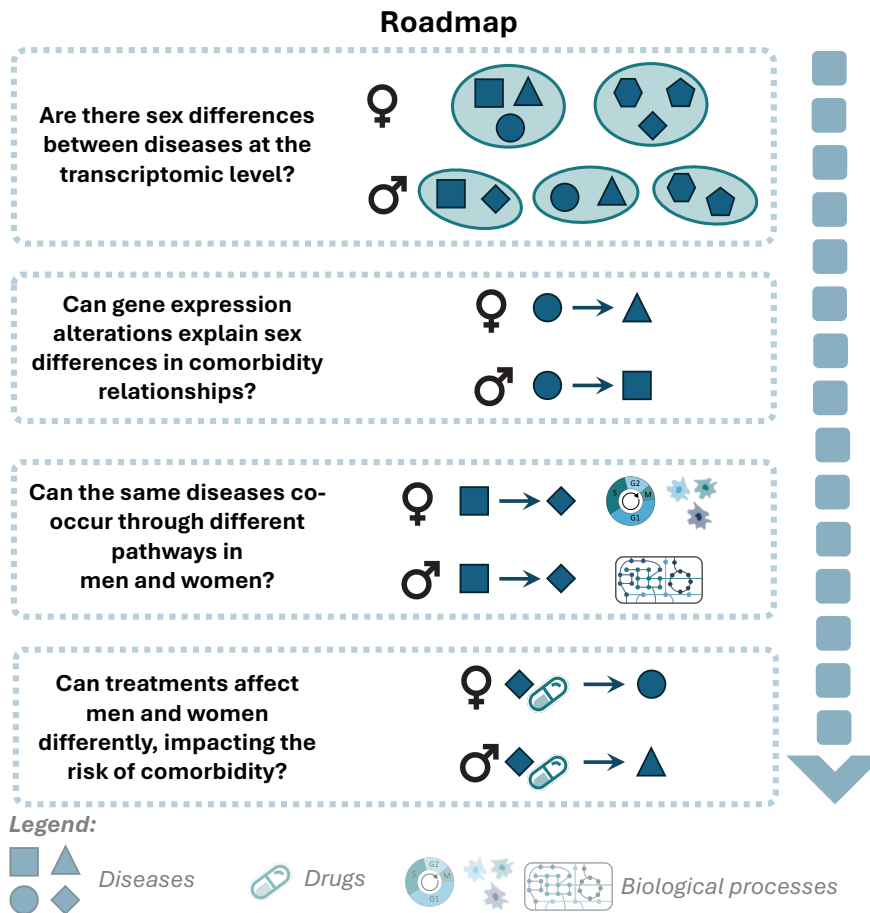


Figure 1. Study roadmap. Schematic representation of the main questions raised in the study. Squares, diamonds, triangles and circles denote diseases. Arrows indicate a higher-than-expected risk of developing a disease when suffering from a previous one.

Methods

Gene expression analysis

Raw expression data was downloaded from the Gene Expression Omnibus (GEO) and ArrayExpress (<https://www.ebi.ac.uk/arrayexpress>). Studies conducted on the *HG U133Plus 2 Affymetrix* microarray platform were chosen based on their cost-effectiveness and reproducibility, given their potential to translate findings into clinical practice. This selection also considered the large sample sizes necessary for analyzing disease-disease associations while minimizing biases that could arise from using different platforms or technologies. After removing low-quality samples (GNUSE median values higher than 1.25²¹), 128 diseases with at least three cases and three controls were kept. As samples' sex information is only given for 52.23% (6,685/12,797) of the samples, the *massiR* package was used to infer samples' sex, properly recovering the sex for 94.24% of the samples with annotated sex²². This method classifies samples as male or female by analyzing the expression of the probes for Y

chromosome genes. To homogenize the terms and to make the transcriptomic similarity networks that were generated in the following steps comparable to previously published epidemiological networks⁹, the disease names were transformed into 3-digit codes of the International Classification of Diseases, version 10 (ICD10), of the World Health Organization, grouping specific diseases into a single 3-digit ICD code. As a result, we identified 76 (2,465 cases and 1,370 controls) and 77 (3,200 cases and 1,871 controls) ICD10 codes in women and men, 59 of which are common to both (see Supplementary Figure 1). 43% (42%) of the cases (controls) are women, with cases accounting for 63-64% of the samples in men and women, respectively. MAS 5.0 algorithm²³ was used to identify and remove lowly expressed genes (detection p -value <0.05). Background correction, normalization, and summarization were done using the frozen Robust Multiarray Analysis (fRMA) preprocessing algorithm (default parameters)²⁴. Differential expression analyses comparing the expression of samples with disease (cases) vs. disease-free samples (controls) were conducted using the linear regression model provided by the LIMMA package²⁵. These comparisons were performed separately and jointly for women and men, adjusting for confounding variables (study of origin and sex). Those genes with a False Discovery Rate (FDR) ≤ 0.05 and a log Fold Change (logFC) lower (higher) than zero were considered as significantly down- (up-) regulated.

Gene set enrichment analysis

We performed functional enrichment analyses using the Gene Set Enrichment Analysis (GSEA) method²⁶, which was applied to the entire list of logFC-ranked genes obtained from the differential expression analysis (see previous section). The resources used in GSEA were Reactome, Gene Ontology, and KEGG. We performed disease clustering using Reactome pathways specifically, whose hierarchical structure allows us to identify 29 lowest-level pathway diagram annotations or categories. Pairwise Euclidean distances were calculated between Reactome pathway enrichment profiles (Normalized Enrichment Score values provided by GSEA). Ward2 method²⁷ was used to generate the clusters, identifying significant disease clusters through bootstrapping using the *pvclust* R package²⁸. The resulting dendrograms obtained for women and men were compared using *tanglegrams* from the *dendextend* R package²⁹, which gives two dendrograms (with the same set of labels), one facing the other, and having their labels connected by lines.

Network construction

Transcriptional similarities were calculated using three distinct sets: (i) the complete list of annotated genes, (ii) the union of annotated genes with significant differential expression

(sDEGs), and (iii) their intersection based on differential expression values (logFC) as previously done ¹⁵. Six similarity metrics were calculated: Pearson's and Spearman's coefficients, cosine similarity, and Euclidean, Canberra, and Manhattan distances. Empirical p-values were calculated through 10,000 permutations for the cosine similarity and the Euclidean, Canberra, and Manhattan distances, correcting for multiple testing by the Bonferroni approach and considering those similarities with an $FDR \leq 0.05$ as significant. In the case of Euclidean, Canberra, and Manhattan distances, the mean of the random distances was compared with the actual distances, obtaining positive (negative) values indicating a greater (lesser) similarity than expected by chance (see [Supplementary Figure 2](#)). The similarity values – obtained from the comparison between real and random distances in the case of Euclidean, Canberra and Manhattan distances, and from the coefficients in the case of Pearson and Spearman correlations and cosine similarity – were binarized, converting those coefficients greater than 0 to +1 and those less than 0 to -1. The results for the different metrics are similar, especially for Pearson and Spearman correlations, cosine similarity, and Euclidean distance (see [Supplementary Figure 3](#)). For simplicity, we will show the results obtained with Euclidean distances, which are the ones that allow us to recover more comorbidity relationships in a meaningful way (see [Supplementary Table 1](#)). Since the number of sDEGs strongly affects the ability to find similarities between diseases (correlations of 0.75 and 0.86 are obtained when working with the sDEGs union and intersection, compared to 0.34 when working with all genes, see [Supplementary Figure 4](#)), we focused on the calculations performed on the complete list of genes.

Overlap with epidemiology

We used a previously published epidemiological network ⁹ ([Supplementary Note 1](#)) to identify the comorbidity relationships recovered by the disease transcriptomic similarity networks (DTSN) generated by comparing similarities between differential expression profiles and their ability to explain comorbidity relationships.

The overlap between networks was performed on the shared set of diseases (present in both the DTSNs and epidemiological networks). Specifically, the overlap of positive and negative transcriptomic similarities with the epidemiological networks was analyzed separately. Overlaps were measured by sex (women vs. women, men vs. men). The significance of the overlap was assessed by Fisher's tests and the use of a set of specifically defined randomisations (generating 10,000 random networks shuffling the edges of the DTSNs while maintaining the degree distribution).

Disease-drug associations

To study the potential sex-specific role of drugs in comorbidities, we retrieved drug targets from the DrugBank³⁰. Since the number of targets per drug is relatively small for enrichment analyses, we used the protein-protein interaction network extracted from IID³¹ - selecting only those protein-protein interactions in humans that have been experimentally verified - to expand the number of targets associated with a given drug by mapping the targets on the network and selecting the first neighbours of the targets for each drug. We then conducted a GSEA²⁶ to associate drugs targeting the products of up or down-regulated genes with the corresponding disease, separately by sex. Disease-drug associations were extracted from the SIDER database³². Disease names were transformed into ICD10 codes using the Unified Medical Language System³³, and DrugBank IDs were mapped into drug names.

Results and discussion

Sex-associated differences in gene expression

To examine the insights gained by studying the similarities between diseases at the transcriptional level according to the patient's sex, we first downloaded microarray raw data from Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo>) and ArrayExpress (<https://www.ebi.ac.uk/arrayexpress>). In order to study the insights gained by calculating similarities between diseases separately for women and men - rather than together - differential expression and enrichment analyses were performed comparing expression in cases vs. controls by analyzing both sexes together and individually (adjusting for sex differences; [see methods](#)).

To get a first overview of the differences in diseases between women and men, we performed hierarchical clustering on the Normalized Enrichment Score of their enriched pathways ([see methods](#)). On average, clusters were larger and more heterogeneous - comprising diseases from different categories - in women than men (66%/25% of the clusters in women/men were composed of diseases from different categories, [see Figure 2](#)). Indeed, the overall agreement between men and women clusters was quite limited. Parkinson's and Alzheimer's diseases (G20 and G30), previously linked at the molecular level³⁴, clustered significantly in men, sharing 112 pathways in which differentially expressed genes were enriched, mainly involved in signal transduction, immune system, and neuronal system. However, Alzheimer's disease clustered with schizophrenia (F20) in women, sharing 116 pathways, including metabolic (glucose metabolism and respiratory electron transport)^{35,36}, metabolism of proteins (protein ubiquitination)^{37,38} and neuronal system pathways (serotonin neurotransmitter release cycle and GABA synthesis release reuptake and degradation)^{39,40}, which have been previously

associated with both diseases. Interestingly, a significant co-occurrence of dementia (including Alzheimer's disease) has been described in patients with schizophrenia, with the risks being higher in women ⁴¹. In contrast, schizophrenia in men clustered with hepatitis C virus (B17) infection, two diseases that co-occur more than expected by chance in the population ⁴², with that risk being higher in men.

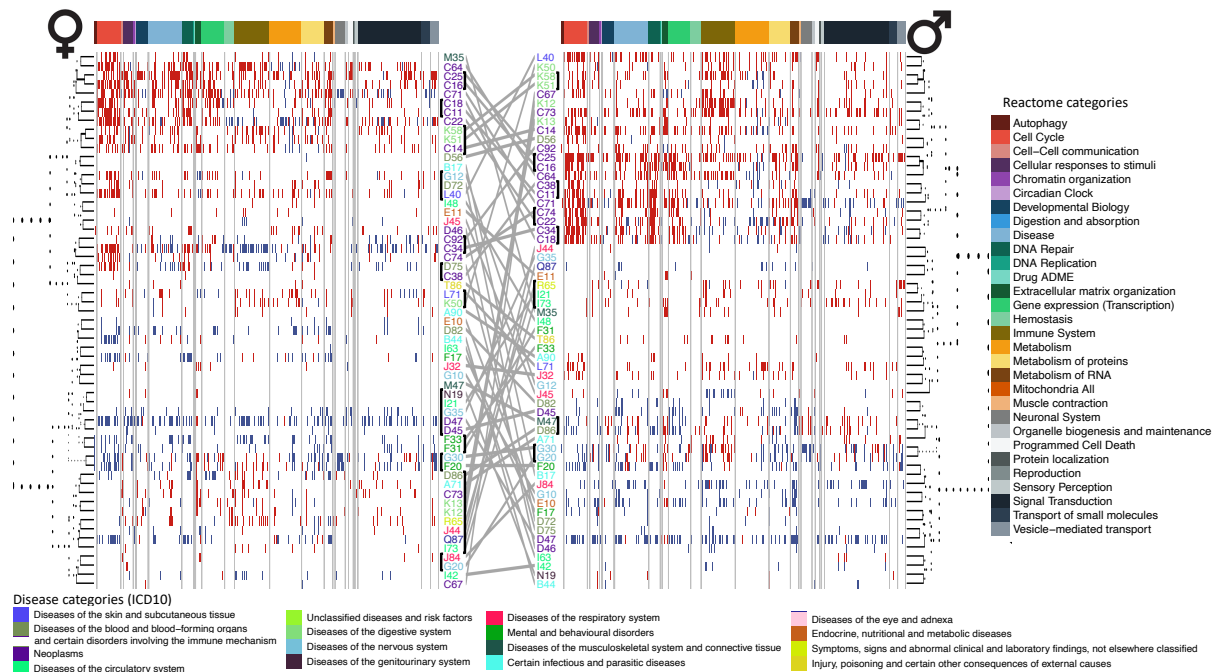


Figure 2. Reactome pathways significantly altered in disease in women and men. The diseases are represented in rows, and the pathways are in columns. The colors of the diseases refer to the category of the ICD10 to which they belong. The color bar in the columns represents the Reactome parents to which 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 on the Normalized Enrichment Scores. The gray lines connect diseases in women and men. Black lines near ICD codes indicate significant clusters after bootstrapping.

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 both sexes, cell cycle processes dominated the similarity in men and the immune system in women. Of particular note is the over-activation of interleukin signaling pathways, which has been proposed as a promising candidate for developing immunomodulatory therapeutic approaches ⁴³. Indeed, considerable divergence in the tumor immune system has been described in tumors in women and men, highlighting its possible implications for differences in the efficacy of immunotherapy between sexes ⁴⁴. In the case of the two digestive diseases (K58 and K51), they, in turn, clustered with oral cavity cancer (C14) in women, which co-occurs more than

expected by chance, coinciding with the risk being higher in women ⁴⁵. The increasing HPV prevalence amplifies the higher risk of oral cavity cancer in IBS patients and might be increased due to the use of immunosuppressive therapy ⁴⁶, denoting the importance of treatments in comorbidity relationships. A total of 37 pathways were found to be altered in this cluster in women, 28 of them up-regulated (principally signal transduction pathways) and nine down-regulated (primarily metabolic pathways). Among them is the overactivation of NOTCH4 signaling, which has been proposed as a marker of oral cancer ⁴⁷ and whose expression is induced in macrophages after activation by Toll-like receptors (TLR) and interferon- γ (IFN- γ), which have been extensively studied in the context of IBS ⁴⁸. These results highlight large differences at the molecular level between women and men in diseases belonging to different categories, opening the question of whether such differences may lead to the development of different comorbidities.

Disease Transcriptomic Similarity Networks

Given the difference in how diseases cluster in women and men based on pathways, we built Disease Transcriptomic Similarity Networks (DTSN) by calculating similarities between pairs of diseases, adjusting for sex differences, and separately for women (wDTSN) and men (mDTSN) to study differences at the pairwise level. Similarities were calculated between diseases on the logFC ([see methods](#)).

Between 45% and 48% of all the potential edges in the wDTSN and mDTSN were positive, meaning the distance was smaller than expected at random. In contrast, 25-27% of the edges were negative, indicating larger-than-expected distances. As described in epidemiology ⁹, significantly more positive interactions were observed between diseases of the same category than between diseases of different categories in both women and men (Odds Ratio (OR)=2.81 in women vs. OR=1.83 in men, [Supplementary Table 2](#)), pointing out that transcriptional similarity may help to understand why diseases affecting the same system are more likely to co-occur versus those affecting different systems. These ORs were even larger when we considered only the connections that preserve the most similar comorbidity trajectories, i.e. the network backbone ⁴⁹, (4.4 in women and 3.2 in men), denoting a greater relevance of intra-category connections in the flow of information, which is also observed in epidemiology ([see Supplementary Note 2](#)). Analyzing specific categories on mDTSN and wDTSN, diseases of the digestive and musculoskeletal system in men and skin diseases in women had a clustering coefficient of 1 (were all connected), followed by mental illnesses which, in women, had a density of 0.83 compared to 0.4 in men, correlated with the higher co-occurrence of mental diseases in women than in men ⁵⁰ ([see Supplementary Figure 5](#)). Focusing on the set of diseases common to men and women, we found significantly fewer positive interactions

in women than in men between the digestive system and nervous system diseases (OR= 0.21) and digestive system and blood diseases (OR= 0.27, [see Supplementary Table 3](#)). These results correlate with the higher risk of co-occurrence of IBS and multiple sclerosis ⁵¹, Parkinson's disease ⁵², or dementia ⁵³. In fact, no positive interactions were found between IBS and multiple sclerosis in men. In contrast, women had significantly more negative interactions between digestive and nervous system diseases (OR = 10.81, [see Supplementary Table 4](#)). When comparing the networks generated considering both sexes together and separately, 16.22% and 13.49% of the positive interactions were detected only in women and men, respectively, values that hover around 9% and 4% in the epidemiology ⁹ ([see Supplementary Figure 6A-B](#)). These results demonstrate that, as in epidemiology, relationships relevant to only one of the sexes can be lost by analysing women and men together.

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

Previous studies have shown that transcriptional similarity can significantly recover comorbidity relationships ^{14,15}, even when analyzing very different populations, highlighting the relevance of the approach ([see Supplementary Note 3](#)). As reflected in the introduction, to date, this phenomenon has not been evaluated separately in women and men ¹⁷. Our results showed that 53.12% and 60.68% of the disease co-occurrences described in women and men by Westergaard et al. ⁹ were also connected in the wDTSN and mDTSN ([see Figure 3, Supplementary Figure 6C and Supplementary Table 1](#)). Metastases and myelodysplastic syndromes were the ICD10 codes with the highest number of comorbidities retrieved by the wDTSN and mDTSN respectively (92.3% and 87.5% respectively, [see Figure 3 and Supplementary Figure 7](#)), where the former co-occurred only with other neoplasms, while the latter also co-occurred with blood disorders (autosomal dominant monocytopenia and primary myelofibrosis). As expected, pathways such as positive regulation of chromosome segregation, DNA unwinding involved in DNA replication, deposition of new CENP-A containing nucleosomes at the centromere, and DNA strand elongation were found to be up-regulated in eleven of the twelve cancers co-occurring with metastases (including women-specific cancers such as ovary and breast cancer). The role of these pathways in cancer has been extensively described ⁵⁴, and they have been shown to correlate with tumor stage and be enriched in relapsed and metastatic tumor patients ⁵⁵ and to promote epithelial-mesenchymal transition (precursor of metastasis) ⁵⁶. In the case of myelodysplastic syndromes, the same biological processes were found to be altered in the same direction in up to five of the seven recovered comorbidities. This is the case of down-regulation of T cell receptor signaling, altered in juvenile myelomonocytic leukemia (C93), primary myelofibrosis

(D75), autosomal dominant monocytopenia (D72), neoplasms of uncertain behavior of lymphoid, hematopoietic and related tissue (D47) and polycythemia vera (D45). Immune deregulation has been shown to play a prominent role in the pathogenesis and progression of the mentioned disorders^{57,58}, opening the door to a possible role also in their co-occurrence. Indeed, there is a syndrome called MonoMAC, characterized by severe monocytopenia and a high risk of progression to myelodysplasia⁵⁹. Other additional processes related to the immune system, such as antigen receptor-mediated signaling, B-cell receptor signaling, and primary immunodeficiency, were also jointly altered in the last four. Genes such as JAK1 were also found to be changed in expression in most diseases that co-occurred with myelodysplastic syndromes (four out of seven). The role of JAKs has been widely described in immunity, immunodeficiency, and cancer⁶⁰. Likewise, it is known that interleukin-7 is crucial for lymphocyte development and for the survival, proliferation, differentiation, and activity of mature T- and B-cells⁶¹. Four of the seven diseases comorbid with myelodysplastic syndromes had altered IL7R levels.

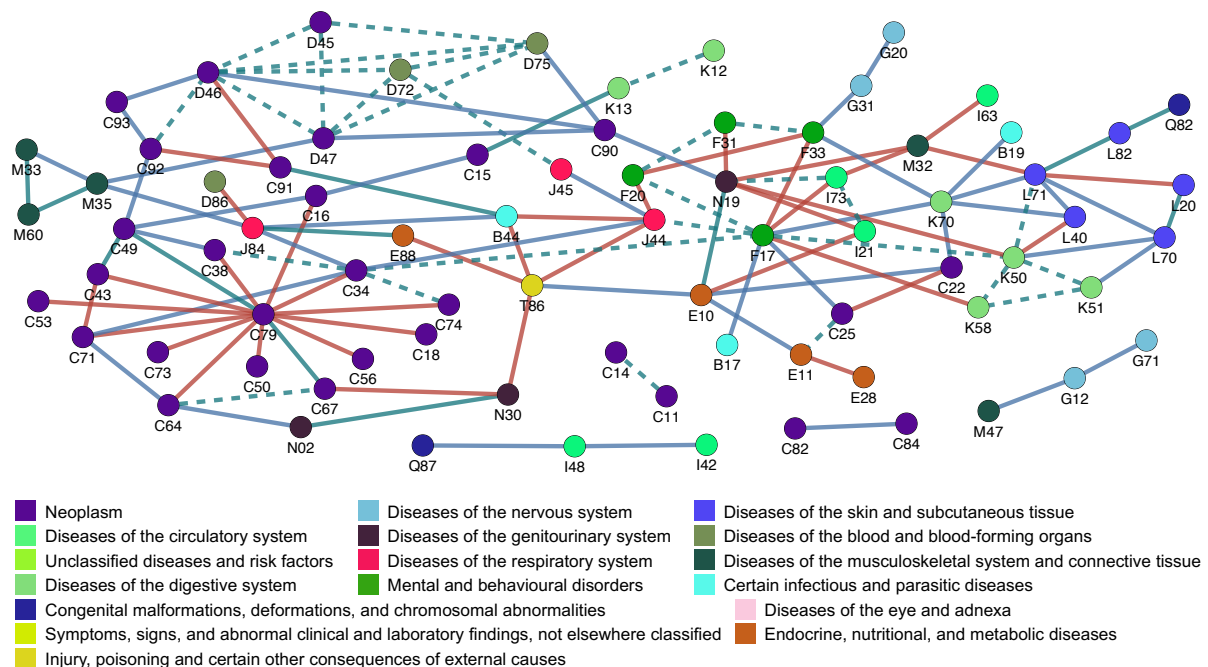


Figure 3. Comorbidities explained by the Disease Transcriptomic Similarity Network. Network of comorbidities recovered based on transcriptomic similarities. Each node represents a disease, colored based on the disease category they belong to (ICD10). Red, blue, and green edges represent comorbidities (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.

In addition to co-occurrences of diseases of the same category, sex-specific DTSNs also identified altered biological processes in comorbid diseases of different categories. For

example, interstitial cystitis (N30) has been shown to increase the risk of bladder cancer (C67)⁶², being the risk only significant for women in the Danish population (RR=1.138, 97.5% CI: 1.032, 1.252). Nonsense-mediated decay, seleno-amino acid metabolism, and EIF2AK4 (GCN2) to amino acid deficiency pathways were down-regulated in both diseases in the wDTSN. Several studies have shown that inflammation plays a prominent role in interstitial cystitis⁶³. Interestingly, it has been described how upregulation of nonsense-mediated decay during chronic inflammation can help restore immune homeostasis⁶⁴, and in turn mutations that can elicit nonsense-mediated decay have been detected in half of patients with bladder cancer⁶⁵. In the same direction, oxidative stress is implicated in the pathogenesis of interstitial cystitis⁶⁶, and seleno-amino acids and their metabolic reprogramming regulate antioxidant defense, enzyme activity, and tumorigenesis⁶⁷, highlighting the widely known phenomenon that chronic inflammation and the resulting oxidative stress can increase the risk of cancer development⁶⁸.

Similarly, psoriasis was found to increase the risk of alcoholic hepatitis⁶⁹, two diseases positively connected in our mDTSN. Psoriasis is a clinically heterogeneous lifelong skin disease mediated by chronic inflammation⁷⁰, and alcoholic hepatitis is characterized by the degeneration of hepatocytes, neutrophil infiltration, and fibrosis⁷¹. Previous studies have described that their co-occurrence might be driven by shared risk factors such as smoking and alcohol consumption, shared pathophysiological pathways including insulin resistance and oxidative stress, as well as psoriasis drugs potentially inducing liver damage^{72,73}. The mDTSN showed that most altered biological processes in both diseases implicated the cell cycle, signal transduction, and the immune system. Notably, several of these pathways were involved in the activation and signaling of nuclear regulatory factor κ B (NF κ B), which is a crucial mediator of the pathogenesis of psoriasis⁷⁴ and at the same time is activated by chronic alcohol intake, promoting the hepatic macrophage expression of pro-inflammatory mediators⁷⁵ (including the tumor necrosis factor involved in hepatocyte injury⁷⁶ and psoriasis⁷⁷). Supporting the existence of common pathways that could explain the co-occurrence of both diseases, it has been described that anti-TNF drugs can induce the onset of psoriasis⁷⁸ and, in turn, significantly increase mortality in patients with moderate to severe alcoholic hepatitis⁷⁹. These diseases also co-occur more than expected in the other sex. Unfortunately, the lack of sufficient samples ([see methods](#)) has not allowed us to analyze these relationships.

Focusing now on the common set of diseases for which we have sufficient samples in both women and men, it has been described that patients with schizophrenia (F20) are at a higher risk of developing COPD (J44)⁸⁰, even when compared to smoking controls⁸¹. At the core of such co-occurrence of diseases, it has been shown that patients with schizophrenia have impaired lung function⁸². Interestingly, lung function SNPs acting as eQTLs have been linked to neuropsychiatric diseases, including schizophrenia⁸³. Clinical observations have shown

that there are sex differences in prevalence, symptoms, and response to treatment in patients with schizophrenia⁸⁴, as well as in comorbidity relationships, where, for example, the risk of developing COPD would be higher in women⁶⁹ – indeed, in the Danish population the risk is only significant for women⁹. The wDTSN recovered this comorbidity relationship between both diseases in women, finding pathways associated with mitochondria (mitochondrial translation, mitochondrial RNA metabolic process, and mitochondrial gene expression) and the immune system (interleukin 10 signaling, neutrophil degranulation, macrophage cytokine production, and antimicrobial peptides) altered in the same direction. Interestingly, previous studies have described an enrichment in mitochondrial processes in women vs men in schizophrenia⁸⁵ and COPD⁸⁶, as well as differences in the immune system^{87,88}, supporting the hypothesis that differences between women and men at the disease level may explain the development of different comorbidities.

Similarly, a higher risk of developing IBS (K58) in smokers (F17) has been observed, being higher in women⁸⁹. Interestingly, the impact of smoking is greater in women than in men⁹⁰ and the prevalence of IBS is higher in women⁹¹, potentially driven by hormonal differences between both sexes⁹². As in other comorbidities, most pathways in common detected in the wDTSN involved mitochondria (mitochondrial respiratory chain complex assembly, respiratory electron transport, complex I biogenesis, detoxification of reactive oxygen species) and immune system (neutrophil activation involved in immune response and immunoregulatory interactions between a lymphoid and non-lymphoid cell). Previous studies support the identified differences in each of the diseases separately^{93–95}, pointing to these processes as possible future targets for reducing the risk of disease co-occurrence. Other diseases that co-occur more than expected by chance in women but not men and were uniquely connected in the wDTSN include type I diabetes (T1D, E10) - myocardial infarction (I21)⁹⁶, bipolar disorder (F31) – uremia (N19), and COPD (J44) - chronic lung allograft dysfunction (B44, [see Figure 3](#)).

On the other hand, a higher-than-expected risk of developing liver cancer (C22) in men with T1D (E10) has been described⁹⁷. Interestingly, the prevalence of T1D is larger in prepubertal women and afterwards occurs twice as much in men⁹⁸. Most of the biological processes altered in the same direction in men in both diseases included mainly pathways 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 the immune system (humoral immune response, positive regulation of immune effector process, and regulation of t helper 1 type immune response). Finally, by calculating similarities between diseases separately for each Reactome parent category ([see Supplementary Note 4](#)) we observed that gene expression and the immune system were the categories that recovered the most comorbidities (41-47%), and drug ADME

the least (4-5%, [see Supplementary Table 5](#)). Despite the diversity in the amount of comorbidities identified for the different categories, all overlaps with the epidemiology were significant ([see Supplementary Figure 8A-B and Supplementary Table 5](#)). All the results presented in this section provide evidence of differences in the biological processes altered in diseases according to the sex of the patients and that these differences may help to understand better why disease co-occurrences are different between women and men. Among the most relevant biological processes, we find the immune system, metabolism, and mitochondria-associated processes as those that best explain the differences in comorbidity relationships between women and men, previously described to drive differences between sexes^{99,100}.

Comorbidities occur through different mechanisms in women and men

Once verified that differences in transcriptional similarities between diseases in women and men can explain differences in comorbidity relationships, the question is, is it possible that the mechanisms by which some diseases co-occur are different for each sex? This has been previously described for mechanical pain hypersensitivity, which is mediated through different immune cells in male (microglia) and female (T-cell) mice¹⁰¹. As shown in [Figure 4](#), 29 disease pairs co-occur in women and men more than expected by chance (12 belong to different categories), and such comorbidities are recovered in both wDTSN and mDTSN. However, there are sex-specific biological alterations.

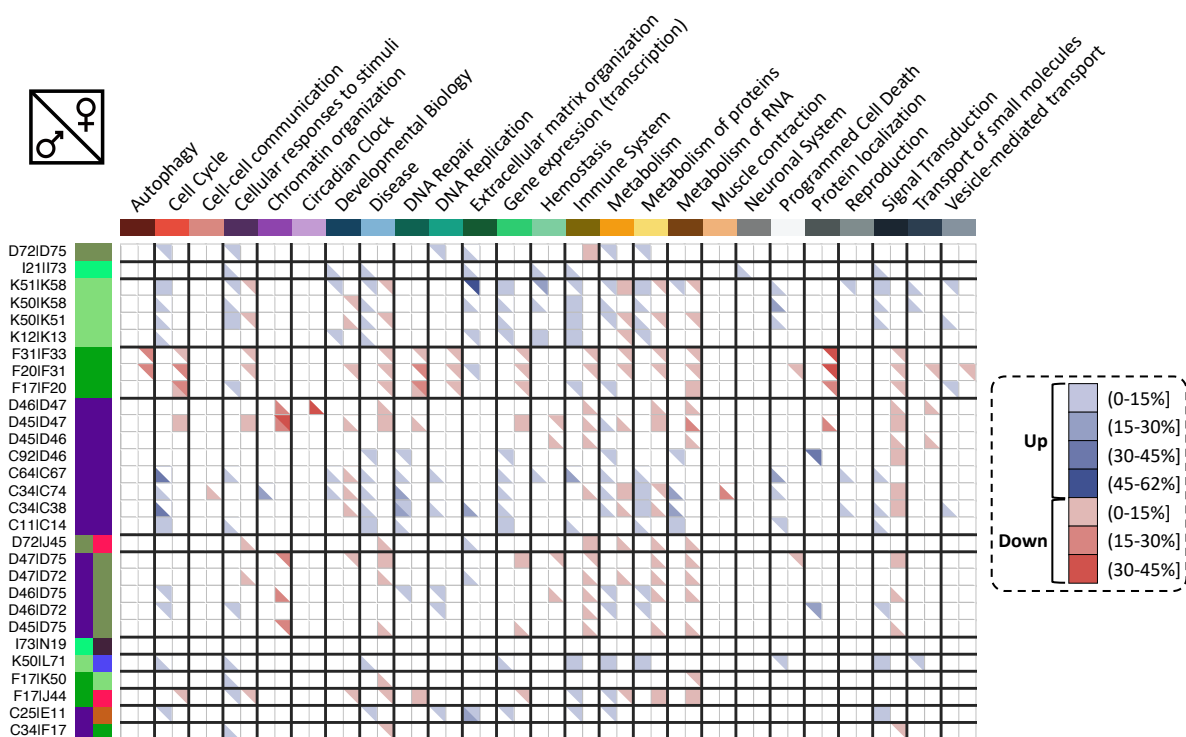


Figure 4. Pathways shared between comorbid diseases in women or men only. Heatmap with disease pairs in rows and Reactome categories in columns representing the percentages of pathways in each category that are up- (blue) or down-regulated (red) in both diseases. Each square is divided into two triangles, the one on the top (bottom) refers to women (men).

Smoking (F17) is a widely described risk factor for the development of COPD (J44), affecting both women and men ¹⁰². However, the mechanisms by which smoking increases the risk of COPD may differ by sex. As shown in [Figure 4](#), in women, alterations common to both diseases were identified in the cell cycle, response to stimuli, metabolism, immune system, and developmental biology that did not occur in men; in men, alterations were observed in DNA repair and protein and RNA metabolism that did not occur in women. It has been previously proposed that cigarette metabolism may vary due to sex differences in the expression and activity of cytochrome P450 enzymes ¹⁰³, a process that was found to be overexpressed in women in smokers and COPD but not in men (cytochrome p450 arranged by substrate type). As previously described, mitochondrial functional pathways were altered in both diseases in women but not men, which might be potential drivers of sex differences in lung diseases ⁸⁶. In contrast, the processes of sumoylation of transcription cofactors, processing of capper intro containing pre-mRNA, and base excision repair were downregulated in men in both diseases but not in women.

Another example is the significant co-occurrence of pancreatic cancer (C25) and T2D (E11) ¹⁰⁴. It is interesting to note that men and women with pancreatic cancer share a high number of altered pathways in the same direction (Jaccard indexes for up and down-regulated pathways of 0.78 and 0.58, respectively), while the overlaps were low in the case of T2D (0.03 and 0, see [Supplementary Table 6](#)). These results point to greater sex differences in T2D than in pancreatic cancer. Therefore, the development of pancreatic cancer may be different according to sex in patients with T2D due to differences in its development ¹⁰⁵. In men, T2D and pancreatic cancer shared mainly extracellular matrix organization pathways (collagen biosynthesis and modifying enzymes, laminin interactions, ECM proteoglycans...) and signal transduction pathways altered in the same direction (not shared in women). On the other hand, women shared immune system (neutrophil degranulation, antiviral mechanism by IFN stimulated genes...) and metabolism pathways (sphingolipid de novo biosynthesis and regulation of cholesterol biosynthesis by SREBP (SREBF)). The results highlight the usefulness of the proposed approach in identifying biological processes that are differentially altered in women and men and may be fundamental in the development of diseases and comorbidities differently in both sexes. Unfortunately, there is still a considerable gap in biological knowledge regarding the differences between women and men in the development

of diseases, which is fundamental to understanding the differences in their comorbidities better.

Sex differences in drug effects

In previous studies, we have identified drugs that could play a role in disease co-occurrence, both increasing and decreasing the risk, proposing this approach as a possible avenue for drug repositioning¹⁰⁶. Going a step further, we have identified subgroups of patients with a disease for which a specific drug association might be related to the observed increased risk of developing secondary diseases¹⁴. In this instance, we investigate how different drugs may be associated with diseases differently in women and men and thus explain their specific comorbidity relationships. Due to the lack of studies analyzing differences in expression changes after drug intake as a function of sex, we extracted drug targets from DrugBank³⁰, expanded the association using a protein-protein interaction network³¹, and performed enrichment analyses²⁶ to identify drugs whose expanded targets tend to be over- or under-expressed in different diseases as a function of sex. 3,878 DrugBank IDs were significantly enriched in at least one disease. Of them, 616 had associations (3,997) to 568 ICD10 codes in the SIDER database (a database of drugs and adverse drug reactions which contains drug indications, extracted from the package inserts using Natural Language Processing)³². Focusing on the diseases for which there were sufficient samples in women and men, 43 of the 59 had at least one enriched drug associated with any diseases under analysis in SIDER. Studying the top 10 drugs associated with more diseases in women and men, only three were found in common (metformin, clofarabine and irinotecan), drugs known to treat diabetes, acute lymphoblastic leukemia, and metastasis. Arginine, bortezomib, and carfilzomib in women, and dexrazoxane, etoposide, and idarubicin in men were other drugs associated with more than 25% of the diseases (used to treat high blood pressure, cardiomyopathies, and cancer). Seven diseases did not share a single drug associated with both men and women: Aspergillus colonization of lung allograft, T1D, amyotrophic lateral sclerosis, interstitial lung disease, rosacea, connective tissue disorders, and axial spondyloarthritis (see [Supplementary Figure 9](#)), while 8 presented a Jaccard index above 0.5: pancreatic, colon, and kidney cancer, neoplasms of uncertain behavior of lymphoid, hematopoietic and related tissue, oral dysplasia, Job's syndrome, thalassemia and IBS, highlighting that there was great heterogeneity in the number of drugs shared between sexes depending on the disease. In total, women had more drugs associated with them in 19 diseases versus 20 diseases in which more drugs were associated with men. These differences between sexes were not due to differences in the number of samples of each sex (p-value of the correlation

= 0.64). In fact, in diseases such as schizophrenia, 3.13 more drugs were identified in women than in men, while there were 2.7 more men than women. When analyzing the drugs enriched for diseases in women and men (and the diseases for which they are indicated), we found interesting differences according to sex (see Figure 5 and Supplementary Figure 10).

Drug-mediated disease associations may reflect that both diseases share symptoms, patients suffering from one disease also have the other disease, or the treatment of one disease plays a role in the development of the other. In the first scenario, IBS (K58) and major depression (F33), two comorbid diseases with shared pathogenesis¹⁰⁷ were connected by lubiprostone, used to treat constipation, a consequence of both diseases¹⁰⁸ (see Figure 5). The drug was indicated for treating IBS and enriched in the differential expression profile of major depression in women (but not men). Interestingly, constipation is more likely in women with depression than in men¹⁰⁹. In the second one, the association between T2D (E11) and schizophrenia (F20), where drugs used to treat diabetes (metformin and gliclazide) were enriched in women with schizophrenia, could indicate that such patients were indeed taking the aforementioned drugs or suffering from T2D. This hypothesis could be supported by the high representation of elderly patients with schizophrenia in the study dataset, since it has been reported that the prevalence of diabetes in late-life schizophrenia patients is indeed significantly higher in women than in men (35% vs. 21.53%)¹¹⁰. In the third one, it has been described that patients with essential thrombocythemia have an increased risk of transformation to myelofibrosis and that this risk is higher in patients taking anagrelide¹¹¹ (especially in women), a drug indicated to treat essential thrombocythemia (D47) and enriched in women with myelofibrosis (D75). However, the origin of the data and the lack of metadata preclude further investigation of this hypothesis.

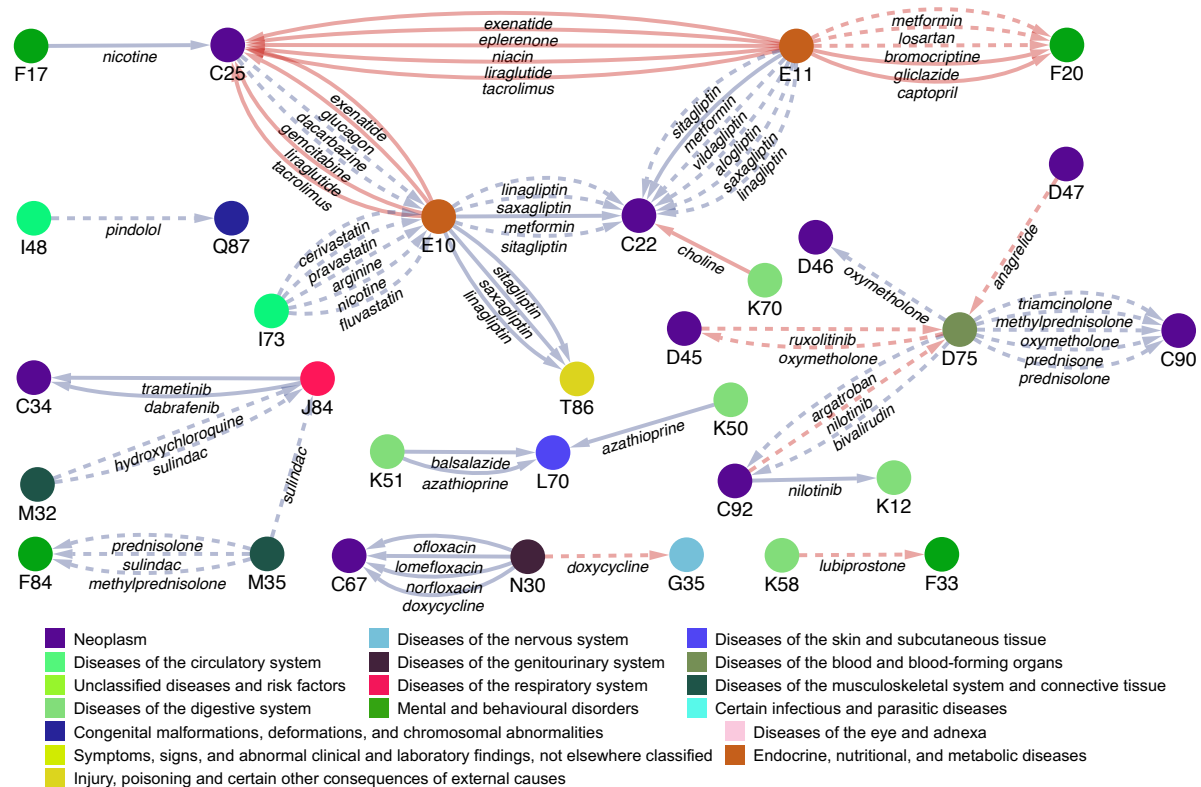


Figure 5. Sex-specific disease-drug associations. The nodes represent diseases, colored according to the disease category (ICD10) to which they belong. An edge connects two diseases if the drug indicated to treat one disease (source node) is enriched in the differential expression profile of another disease (sink node). Only associations between comorbid diseases⁹ from different ICD10 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 is positive (negative), for all other associations see [Supplementary Figure 10](#).

Other comorbidity relationships with pharmacological associations not captured by Westergaard *et al.*⁹ included the association between T1D and asthma¹¹², where the risk is higher in boys compared to girls¹¹³. We found 15 T1D-enriched drugs indicated to treat asthma in men, including salbutamol and salmeterol, not found in women. Interestingly, an increase in blood glucose levels associated with salbutamol use has been described¹¹⁴, highlighting the need to pay special attention to the use of certain treatments depending on the sex of the patients. Similarly, we found two drugs, risperidone and citalopram, significantly enriched in women with schizophrenia, which have also been associated with Alzheimer's disease. The first is an atypical antipsychotic used to treat schizophrenia¹¹⁵ and behavioral symptoms in Alzheimer's disease patients¹¹⁶. At the same time, the second is an antidepressant used to treat depression and negative symptoms in schizophrenia¹¹⁷ and agitation in patients with Alzheimer's disease¹¹⁸. Interestingly, as mentioned, a significant risk

of developing Alzheimer's disease has been described in schizophrenia patients, being higher in women⁴¹, and the response to citalopram has also shown sex differences¹¹⁹. These results support the hypothesis that drug-disease associations are different depending on the sex of the patients, and that this may also play a role in comorbidity relationships. However, further studies are needed that directly analyze molecular-level differences in the effect of drugs as a function of the sex of the patient.

Conclusions

For the first time, disease networks have been generated using separate molecular data for women and men, observing how, based on their differential expression profiles, diseases cluster differently in women and men, being the clusters more heterogeneous in terms of disease categories in women. Calculating transcriptomic similarities between pairs of diseases, we have identified relevant sex-specific differences that align well with those observed at the level of disease co-occurrence. This is the case for T1D and liver cancer in men, or T1D and myocardial infarction and schizophrenia and COPD in women. Interestingly, in the set of comorbidity relationships common to both sexes, we have found differences related to the role of the immune system. Finally, the above-mentioned sex differences have potential consequences for the efficacy of the treatments and the side effects in women and men. The results have shown that it is essential to study comorbidity patterns in women and men separately, investigating the underlying molecular mechanisms and aiming to improve treatments.

References

1. 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).
2. Feinstein, A. R. The pre-therapeutic classification of co-morbidity in chronic disease. *J. Chronic Dis.* **23**, 455–468 (1970).
3. 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).
4. 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).
5. McDonald, C. J. The barriers to electronic medical record systems and how to overcome them. *J. Am. Med. Inform. Assoc. JAMIA* **4**, 213–221 (1997).
6. *Computerizing Large Integrated Health Networks.* (Springer, New York, NY, 1997). doi:10.1007/978-1-4612-0655-2.
7. 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).
8. Jensen, A. B. *et al.* Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. *Nat. Commun.* **5**, 4022 (2014).
9. 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).
10. Goh, K.-I. *et al.* The human disease network. *Proc. Natl. Acad. Sci.* **104**, 8685–8690 (2007).
11. Lee, D. S. *et al.* The implications of human metabolic network topology for disease comorbidity. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 9880–9885 (2008).
12. Menche, J. *et al.* Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science* **347**, 841 (2015).

13. Morselli Gysi, D. & Barabási, A.-L. Noncoding RNAs improve the predictive power of network medicine. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2301342120 (2023).
14. Sánchez-Valle, J. *et al.* Interpreting molecular similarity between patients as a determinant of disease comorbidity relationships. *Nat. Commun.* **11**, (2020).
15. Urda-García, B., Sánchez-Valle, J., Lepore, R. & Valencia, A. Patient stratification reveals the molecular basis of disease comorbidities. *medRxiv* 2021.07.22.21260979 (2021) doi:10.1101/2021.07.22.21260979.
16. 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**, (2021).
17. Sánchez-Valle, J. & Valencia, A. Molecular bases of comorbidities: present and future perspectives. *Trends Genet. TIG* **39**, 773–786 (2023).
18. Aguet, F. *et al.* The impact of sex on gene expression across human tissues. *Science* **369**, (2020).
19. 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).
20. Lopes-Ramos, C. M. *et al.* Sex Differences in Gene Expression and Regulatory Networks across 29 Human Tissues. *Cell Rep.* **31**, (2020).
21. McCall, M. N., Murakami, P. N., Lukk, M., Huber, W. & Irizarry, R. A. Assessing affymetrix GeneChip microarray quality. *BMC Bioinformatics* **12**, 137 (2011).
22. 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. *Bioinforma. Oxf. Engl.* **30**, 2084–2085 (2014).
23. Hubbell, E., Liu, W.-M. & Mei, R. Robust estimators for expression analysis. *Bioinforma. Oxf. Engl.* **18**, 1585–1592 (2002).

24. McCall, M. N., Jaffee, H. A. & Irizarry, R. A. fRMA ST: frozen robust multiarray analysis for Affymetrix Exon and Gene ST arrays. *Bioinforma. Oxf. Engl.* **28**, 3153–4 (2012).
25. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
26. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–50 (2005).
27. Murtagh, F. & Legendre, P. Ward’s Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward’s Criterion? *J. Classif.* **31**, 274–295 (2014).
28. Suzuki, R. & Shimodaira, H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinforma. Oxf. Engl.* **22**, 1540–1542 (2006).
29. Galili, T. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinforma. Oxf. Engl.* **31**, 3718–3720 (2015).
30. Wishart, D. S. *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074–D1082 (2018).
31. 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).
32. Kuhn, M., Letunic, I., Jensen, L. J. & Bork, P. The SIDER database of drugs and side effects. *Nucleic Acids Res.* **44**, D1075–1079 (2016).
33. Bodenreider, O. The Unified Medical Language System (UMLS): Integrating biomedical terminology. *Nucleic Acids Res.* **32**, (2004).
34. Kelly, J., Moyeed, R., Carroll, C., Luo, S. & Li, X. Genetic networks in Parkinson’s and Alzheimer’s disease. *Aging* **12**, 5221–5243 (2020).
35. Cunnane, S. *et al.* Brain fuel metabolism, aging, and Alzheimer’s disease. *Nutr. Burbank Los Angel. Cty. Calif* **27**, 3–20 (2011).

36. 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).
37. Harris, L. D., Jasem, S. & Licchesi, J. D. F. The Ubiquitin System in Alzheimer's Disease. *Adv. Exp. Med. Biol.* **1233**, 195–221 (2020).
38. Bousman, C. A. *et al.* Elevated ubiquitinated proteins in brain and blood of individuals with schizophrenia. *Sci. Rep.* **9**, 2307 (2019).
39. Eggers, A. E. A serotonin hypothesis of schizophrenia. *Med. Hypotheses* **80**, 791–794 (2013).
40. 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).
41. Cai, L. & Huang, J. Schizophrenia and risk of dementia: a meta-analysis study. *Neuropsychiatr. Dis. Treat.* **14**, 2047–2055 (2018).
42. Hughes, E., Bassi, S., Gilbody, S., Bland, M. & Martin, F. Prevalence of HIV, hepatitis B, and hepatitis C in people with severe mental illness: a systematic review and meta-analysis. *Lancet Psychiatry* **3**, 40–48 (2016).
43. Mirlekar, B. & Pylayeva-Gupta, Y. IL-12 Family Cytokines in Cancer and Immunotherapy. *Cancers* **13**, 167 (2021).
44. Ye, Y. *et al.* Sex-associated molecular differences for cancer immunotherapy. *Nat. Commun.* **11**, 1779 (2020).
45. 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. Off. Clin. Pract. J. Am. Gastroenterol. Assoc.* **14**, 413–420 (2016).

46. 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).
47. 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).
48. Lu, Y., Li, X., Liu, S., Zhang, Y. & Zhang, D. Toll-like Receptors and Inflammatory Bowel Disease. *Front. Immunol.* **9**, 72 (2018).
49. Simas, T., Correia, R. B. & Rocha, L. M. The distance backbone of complex networks. *J. Complex Netw.* **9**, 1–35 (2021).
50. Astbury, J. Gender disparities in mental health. in 73–92 (Who, Geneva, Switzerland, 2001).
51. 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).
52. 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).
53. Liu, N. *et al.* Inflammatory bowel disease and risk of dementia: An updated meta-analysis. *Front. Aging Neurosci.* **14**, 962681 (2022).
54. Dhar, S., Datta, A. & Brosh, R. M. DNA helicases and their roles in cancer. *DNA Repair* **96**, 102994 (2020).
55. Bakhoun, S. F. & Cantley, L. C. The Multifaceted Role of Chromosomal Instability in Cancer and Its Microenvironment. *Cell* **174**, 1347–1360 (2018).
56. Jeffery, D. *et al.* CENP-A overexpression promotes distinct fates in human cells, depending on p53 status. *Commun. Biol.* **4**, 417 (2021).

57. Rodriguez-Sevilla, J. J. & Colla, S. T-cell dysfunctions in myelodysplastic syndromes. *Blood* **143**, 1329–1343 (2024).
58. Cardoso, E. M. *et al.* Distinctive CD8+ T cell and MHC class I signatures in polycythemia vera patients. *Ann. Hematol.* **97**, 1563–1575 (2018).
59. Calvo, K. R. *et al.* Myelodysplasia in autosomal dominant and sporadic monocytopenia immunodeficiency syndrome: diagnostic features and clinical implications. *Haematologica* **96**, 1221–1225 (2011).
60. O’Shea, J. J., Holland, S. M. & Staudt, L. M. JAKs and STATs in immunity, immunodeficiency, and cancer. *N. Engl. J. Med.* **368**, 161–170 (2013).
61. Mazzucchelli, R. & Durum, S. K. Interleukin-7 receptor expression: intelligent design. *Nat. Rev. Immunol.* **7**, 144–154 (2007).
62. Vermeulen, S. H. *et al.* Recurrent urinary tract infection and risk of bladder cancer in the Nijmegen bladder cancer study. *Br. J. Cancer* **112**, 594–600 (2015).
63. Moldwin, R. M. *et al.* Immune cell profiles of patients with interstitial cystitis/bladder pain syndrome. *J. Transl. Med.* **20**, 97 (2022).
64. Hug, N., Longman, D. & Cáceres, J. F. Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Res.* **44**, 1483–1495 (2016).
65. Hu, Z., Yau, C. & Ahmed, A. A. A pan-cancer genome-wide analysis reveals tumour dependencies by induction of nonsense-mediated decay. *Nat. Commun.* **8**, 15943 (2017).
66. Mohammad, A., Laboulaye, M. A., Shenhar, C. & Dobberfuhl, A. D. Mechanisms of oxidative stress in interstitial cystitis/bladder pain syndrome. *Nat. Rev. Urol.* **21**, 433–449 (2024).
67. Liang, R. *et al.* Seleno-amino Acid Metabolism Reshapes the Tumor Microenvironment: from Cytotoxicity to Immunotherapy. *Int. J. Biol. Sci.* **20**, 2779–2789 (2024).
68. Zhao, H. *et al.* Inflammation and tumor progression: signaling pathways and targeted intervention. *Signal Transduct. Target. Ther.* **6**, 263 (2021).

69. 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).
70. Raharja, A., Mahil, S. K. & Barker, J. N. Psoriasis: a brief overview. *Clin. Med. Lond. Engl.* **21**, 170–173 (2021).
71. Turner, A. W., Wong, D., Dreisbach, C. N. & Miller, C. L. GWAS Reveal Targets in Vessel Wall Pathways to Treat Coronary Artery Disease. *Front. Cardiovasc. Med.* **5**, 72 (2018).
72. Cassano, N., Vestita, M., Apruzzi, D. & Vena, G. A. Alcohol, psoriasis, liver disease, and anti-psoriasis drugs. *Int. J. Dermatol.* **50**, 1323–1331 (2011).
73. Balak, D. M. W., Piaserico, S. & Kasujee, I. Non-Alcoholic Fatty Liver Disease (NAFLD) in Patients with Psoriasis: A Review of the Hepatic Effects of Systemic Therapies. *Psoriasis Auckl. NZ* **11**, 151–168 (2021).
74. Goldminz, A. M., Au, S. C., Kim, N., Gottlieb, A. B. & Lizzul, P. F. NF- κ B: an essential transcription factor in psoriasis. *J. Dermatol. Sci.* **69**, 89–94 (2013).
75. Mandrekar, P. & Szabo, G. Signalling pathways in alcohol-induced liver inflammation. *J. Hepatol.* **50**, 1258–1266 (2009).
76. Kawaratani, H. *et al.* The effect of inflammatory cytokines in alcoholic liver disease. *Mediators Inflamm.* **2013**, 495156 (2013).
77. Mylonas, A. & Conrad, C. Psoriasis: Classical vs. Paradoxical. The Yin-Yang of TNF and Type I Interferon. *Front. Immunol.* **9**, 2746 (2018).
78. Joulfayan, H., Makunts, T. & Abagyan, R. Anti-TNF- α therapy induced psoriasis in rheumatoid arthritis patients according to FDA postmarketing surveillance data. *Sci. Rep.* **13**, 10448 (2023).
79. Boetticher, N. C. *et al.* A randomized, double-blinded, placebo-controlled multicenter trial of etanercept in the treatment of alcoholic hepatitis. *Gastroenterology* **135**, 1953–1960 (2008).

80. 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).
81. 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).
82. Ruiz-Rull, C. *et al.* Low lung function in Bipolar Disorder and Schizophrenia: a hidden risk. *Front. Physiol.* **15**, 1335798 (2024).
83. Obeidat, M. *et al.* Molecular mechanisms underlying variations in lung function: a systems genetics analysis. *Lancet Respir. Med.* **3**, 782–795 (2015).
84. Li, R., Ma, X., Wang, G., Yang, J. & Wang, C. Why sex differences in schizophrenia? *J. Transl. Neurosci.* **1**, 37–42 (2016).
85. Yu, Z. *et al.* Sex-Specific Differences in the Transcriptome of the Human Dorsolateral Prefrontal Cortex in Schizophrenia. *Mol. Neurobiol.* **60**, 1083–1098 (2023).
86. Glass, K. *et al.* Sexually-dimorphic targeting of functionally-related genes in COPD. *BMC Syst. Biol.* **8**, 118 (2014).
87. 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).
88. Troianova, N. *et al.* Impact of Sex on Circulating Leukocytes Composition in COPD Patients. *Int. J. Chron. Obstruct. Pulmon. Dis.* **16**, 3539–3550 (2021).
89. 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).
90. Langhammer, A., Johnsen, R., Gulsvik, A., Holmen, T. L. & Bjermer, L. Sex differences in lung vulnerability to tobacco smoking. *Eur. Respir. J.* **21**, 1017–1023 (2003).

91. Kim, Y. S. & Kim, N. Sex-Gender Differences in Irritable Bowel Syndrome. *J. Neurogastroenterol. Motil.* **24**, 544–558 (2018).
92. Meleine, M. & Matricon, J. Gender-related differences in irritable bowel syndrome: potential mechanisms of sex hormones. *World J. Gastroenterol.* **20**, 6725–6743 (2014).
93. Yang, C. X. *et al.* Widespread Sexual Dimorphism in the Transcriptome of Human Airway Epithelium in Response to Smoking. *Sci. Rep.* **9**, 17600 (2019).
94. Koo, H.-K. *et al.* Sex-specific associations with DNA methylation in lung tissue demonstrate smoking interactions. *Epigenetics* **16**, 692–703 (2021).
95. Hoang, N., Brooks, K. & Edwards, K. Sex-specific colonic mitochondrial dysfunction in the indomethacin-induced inflammatory bowel disease model in rats. *Res. Sq.* rs.3.rs-2626257 (2023) doi:10.21203/rs.3.rs-2626257/v1.
96. 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).
97. Carstensen, B. *et al.* Cancer incidence in persons with type 1 diabetes: a five-country study of 9,000 cancers in type 1 diabetic individuals. *Diabetologia* **59**, 980–988 (2016).
98. Williams, J. M., Poudel, B. & Shields, C. A. Chapter 15 - Sex Differences in Autoimmune Type-1 Diabetes. in *Sex Differences in Cardiovascular Physiology and Pathophysiology* (eds. LaMarca, B. & Alexander, B. T.) 239–249 (Academic Press, 2019). doi:10.1016/B978-0-12-813197-8.00015-4.
99. Klein, S. L. & Flanagan, K. L. Sex differences in immune responses. *Nat. Rev. Immunol.* **16**, 626–638 (2016).
100. Mauvais-Jarvis, F. Sex differences in energy metabolism: natural selection, mechanisms and consequences. *Nat. Rev. Nephrol.* **20**, 56–69 (2024).
101. Sorge, R. E. *et al.* Different immune cells mediate mechanical pain hypersensitivity in male and female mice. *Nat. Neurosci.* **18**, 1081–1083 (2015).

102. Adeloje, 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).
103. Barnes, P. J. Sex Differences in Chronic Obstructive Pulmonary Disease Mechanisms. *Am. J. Respir. Crit. Care Med.* **193**, 813–814 (2016).
104. 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).
105. Kautzky-Willer, A., Leutner, M. & Harreiter, J. Sex differences in type 2 diabetes. *Diabetologia* **66**, 986–1002 (2023).
106. Sánchez-Valle, J. *et al.* A molecular hypothesis to explain direct and inverse comorbidities between Alzheimer’s Disease, Glioblastoma and Lung cancer. *Sci. Rep.* **7**, 4474 (2017).
107. 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).
108. Lacy, B. E. & Levy, L. C. Lubiprostone: a novel treatment for chronic constipation. *Clin. Interv. Aging* **3**, 357–364 (2008).
109. 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).
110. Huo, L. *et al.* Diabetes in late-life schizophrenia: Prevalence, factors, and association with clinical symptoms. *J. Psychiatr. Res.* **132**, 44–49 (2021).
111. 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).

112. Smew, A. I., Lundholm, C., Sävendahl, L., Lichtenstein, P. & Almqvist, C. Familial Coaggregation of Asthma and Type 1 Diabetes in Children. *JAMA Netw. Open* **3**, e200834 (2020).
113. Hsiao, Y.-T. *et al.* Type 1 Diabetes and Increased Risk of Subsequent Asthma: A Nationwide Population-Based Cohort Study. *Medicine (Baltimore)* **94**, e1466 (2015).
114. 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**, (2019).
115. Marder, S. R. & Meibach, R. C. Risperidone in the treatment of schizophrenia. *Am. J. Psychiatry* **151**, 825–835 (1994).
116. 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).
117. Goff, D. C. *et al.* Citalopram in first episode schizophrenia: The DECIFER trial. *Schizophr. Res.* **208**, 331–337 (2019).
118. Porsteinsson, A. P. *et al.* Effect of citalopram on agitation in Alzheimer disease: the CitAD randomized clinical trial. *JAMA* **311**, 682–691 (2014).
119. Young, E. A. *et al.* Sex differences in response to citalopram: a STAR*D report. *J. Psychiatr. Res.* **43**, 503–511 (2009).

Acknowledgements

J.S.-V. was funded by the Spanish Ministry of Economics and Competitiveness (PID2022-141809OB-I00). J.S.-V. acknowledge funding from the Horizon Europe project COMMUTE under the grant agreement No. 101136957. Funded by the European Union. 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, and by the National Science Foundation Research Traineeship “Interdisciplinary Training in Complex Networks and Systems” (grant 1735095). 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](https://www.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 (C005/24-ED CV1), 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. The authors declare that they have no competing interests. We thank Beatriz Urda (Barcelona Supercomputing Center) for critical reading of the manuscript.

Data availability

All data needed to understand and assess the conclusions of this research are available in the main text, supplementary materials, and <https://github.com/jonsv89/SHDC/blob/main/Datasets.txt>. The raw datasets are publicly available and can be downloaded from GEO (<https://www.ncbi.nlm.nih.gov/geo>) and ArrayExpress (<https://www.ebi.ac.uk/arrayexpress>).

Code availability

The code is available at <https://github.com/jonsv89/SHDC>.

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. and AV. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Competing interests

The authors declare no competing interests.