OXFORD

Briefings in Bioinformatics, 2025, **26(4)**, bbaf407
https://doi.org/10.1093/bib/bbaf407
Case Study

Differential causal networks highlight sex-based differences in human tissues

Annamaria Defilippo[1,*], Kimberly Glass[2], Federico M. Giorgi[3], Tamer Kahveci[4], Pierangelo Veltri[5], Pietro Hiram Guzzi[1]

[1]Department of Surgical and Medical Sciences, University of Catanzaro, Viale Europa, 88100 Catanzaro, Italy
[2]Channing Division of Network Medicine, Brigham and Women's Hospital & Harvard Medical School,181 Longwood Avenue, Boston, MA 02115, United States
[3]Department of Pharmacy and Biotechnology, University of Bologna,Via Irnerio, 48, 40126 Bologna, Italy
[4]Department of Computer and Information Science and Engineering, University of Florida, 432 Newell Dr, Gainesville, FL 32611, United States
[5]Department of Informatics, Modelling, Electronics and Systems, University of Calabria, Via P. Bucci Edificio 42C, 87036, Rende (CS), Italy

*Corresponding author. Department of Surgical and Medical Sciences, University of Catanzaro, Italy. E-mail: annamaria.defilippo@unicz.it

Abstract

Sex differences appear in healthy and pathological conditions and may influence sex-specific therapeutic responses. Understanding such differences is a key activity for developing precision medicine strategies. This study investigates sex differences in gene expression across 40 human tissues by applying a Differential Causal Network (DCN) analysis using data from the Genotype-Tissue Expression project. We identified sex-based DCNs that highlight distinct molecular mechanisms influencing both health and disease in men and women. For example, in pancreas tissue, genes associated with immune system show significant differences in their regulatory patterns between sexes, demonstrating a possible different response to diseases such as diabetes mellitus and cancer. Our findings provide valuable information on the biological underpinnings of sex differences, offering potential pathways for the development of precision medicine strategies.

Keywords: differential causal networks; gene differentiation; sex differences; human tissues

Introduction

Differential gene expression (GE) analysis is essential for identifying genes whose transcription patterns vary between groups, such as between sexes. Sex-biased genes exhibit differential expression between males and females, characterized as male-biased (higher expression in males) or female-biased (higher expression in females). However, sex-biased expression is not an intrinsic or fixed property of genes; instead, it is observed within specific samples under particular conditions. For instance, sex-biased GE can vary during development, reflecting growth stage specificity [1]. Sex-based differences in GE are observed across numerous human tissues. These differences influence the onset, prevalence, and severity of diverse diseases and are modulated by factors such as hormones, sex chromosomes, behavioral differences, and environmental exposures. Notably, these differences are driven by tissue-specific effects rather than merely differences in cell type abundances [2].

In this context differential network analysis (DiNA) allows for comparing biological networks across different conditions, highlighting unique characteristics. DiNA usually focus solely on comparing the network architectures derived from the same set of genes across conditions without an explicit differential expression step [3]. However, DiNA could integrates differential expression analysis (DEA), identifying genes with significant changes in expression. Lichtblau et al., combined DiNA with DEA to capture complex features associated with changes in biological networks [3].

Despite such advances, causal mechanisms underlying these differences remain poorly understood. Causal networks (CNs) model causal relationships, offering a potential starting point for identifying complex and divergent interactions among genes from experimental data [4, 5]. Consequently, the need to compare different causal mechanisms in two different experimental condition has lead to the introduction of differential causal networks (DCNs) [6].

Other studies have primarily focused on identifying genes that display sex-specific patterns of expression, both at a large scale across multiple tissues and within individual tissues [7, 8]. These efforts have provided important insights into the differential gene regulation associated with sex. In contrast, our work introduces a novel perspective by analyzing the distinct causal mechanisms underlying these sex-biased expression patterns, offering a deeper understanding of the biological processes driving such differences. In addition, other works have investigated sex differences in microRNA expression, which play key regulatory roles in GE, or copy number alteration [9, 10]. In contrast, our analysis is restricted to gene transcripts, allowing us to specifically examine sex-related differences at the level of mRNA expression and their underlying causal mechanisms.

DCNs framework was introduced in [6] to uncover not only whether two CNs differ, but also how they differ structurally,

Received: May 16, 2025. **Revised:** June 26, 2025. **Accepted:** July 07, 2025

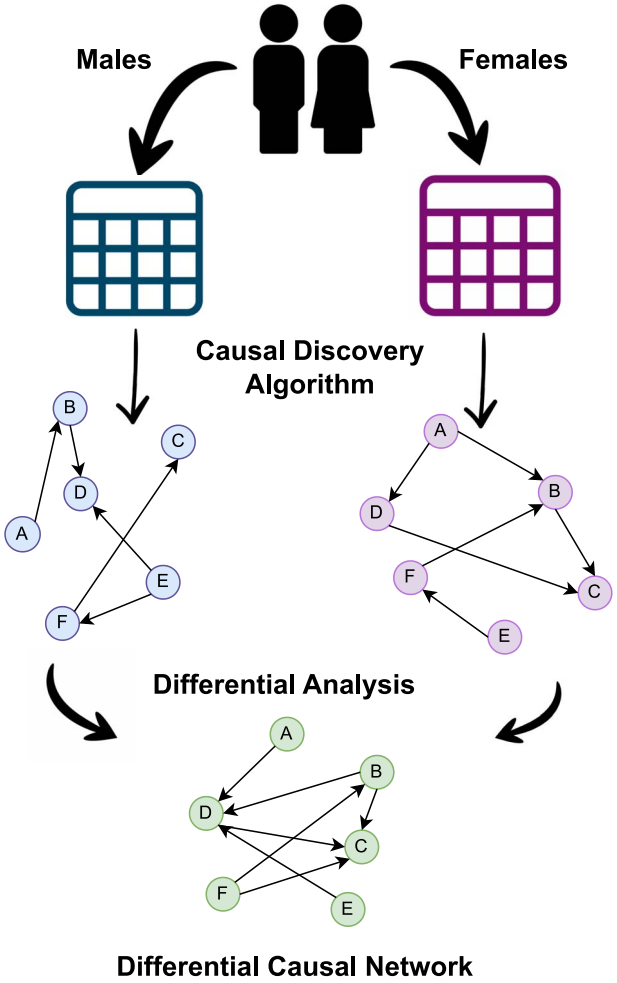

Figure 1. Illustration of the DCN generation process, emphasizing how causal relationships were derived in the study's analysis.

characterizing changes in network topology, such as the presence or absence of specific causal edges.

This study focuses on understanding differences in GE between the sexes using DCNs on most human tissues using publicly available Open Access Data from the Genotype-Tissue Expression (GTEx) portal. Based on differentially expressed genes, four CNs were constructed for each of the 40 selected tissues: one for males and one for females, further stratified by age groups. The age specific networks were aggregated by sex to derive two comprehensive CNs (one for males and one for females). Subsequently, three types of differential analyses [6] were performed to generate the DCNs.

Figure 1 summarizes the workflow followed in this study.

Materials and methods

Data collection

Gene counts and gene TPMs (transcripts per millon) for 59 033 genes across 40 human tissues were extracted using data from the GTEx project (v10 data release) [11]. For this analysis sex-specific tissues and tissues with few samples were excluded. GE data from 981 subjects (654 males and 327 females) were included in the analysis.

Gene counts measure a gene's expression level in a given sample and are often used in DEA [12]. Gene TPMs, on the other hand, represent a normalized measure of GE and are helpful in comparing the relative abundance of different genes [13]. Based on this distinction, gene counts were used for sex gene differentiation, and gene TPMs, filtered by differentiated genes, were

Table 1. Data summary showing the number of selected samples and the corresponding identified differentially expressed genes for each tissue type analyzed, sorted in descending order

Tissue	Samples	Differentially expressed genes
Pancreas	278	249
Adipose subcutaneous	466	165
Adipose visceral omentum	360	88
Artery coronary	200	91
Skin sun exposed (lower leg)	469	80
Small intestine terminal ileum	144	79
Nerve tibial	422	71
Minor salivary gland	104	69
Muscle skeletal	532	68
Artery tibial	440	67
Heart atrial appendage	286	62
Stomach	269	55
Artery aorta	328	55
Brain putamen basal ganglia	122	53
Colon transverse	344	51
Thyroid	450	51
Esophagus muscularis	386	50
Kidney cortex	48	50
Cells ebv-transformed lymphocytes	224	50
Brain hippocampus	130	47
Esophagus mucosa	430	47
Esophagus gastroesophageal junction	274	47
Colon sigmoid	302	46
Skin not sun exposed (suprapubic)	412	45
Liver	156	44
Brain cerebellar hemisphere	150	42
Brain cerebellum	144	42
Brain cortex	150	42
Brain anterior cingulate	112	41
Brain caudate basal ganglia	152	41
Brain hypotalamus	128	40
Brain frontal cortex	136	38
Brain spinal cord cervical	136	34
Brain amygdala	104	33
Whole blood	524	10

Table 2. Pancreas significant pathways

Pathway	N genes	FDR value
Adaptive immune response	28	2.37E-19
Adaptive immunity	21	2.71E-18
Immune system process	49	8.07E-17
Immune response	40	2.31E-16

Table 3. Pancreas gene enrichment

GE description	N genes	*P*-value
Immunoglobulin complex	89	5.77E-139
Adaptive immune response	108	2.59E-94
Antigen binding	71	6.63E-92
Classical antibody-mediated complement activation	49	3.38E-76

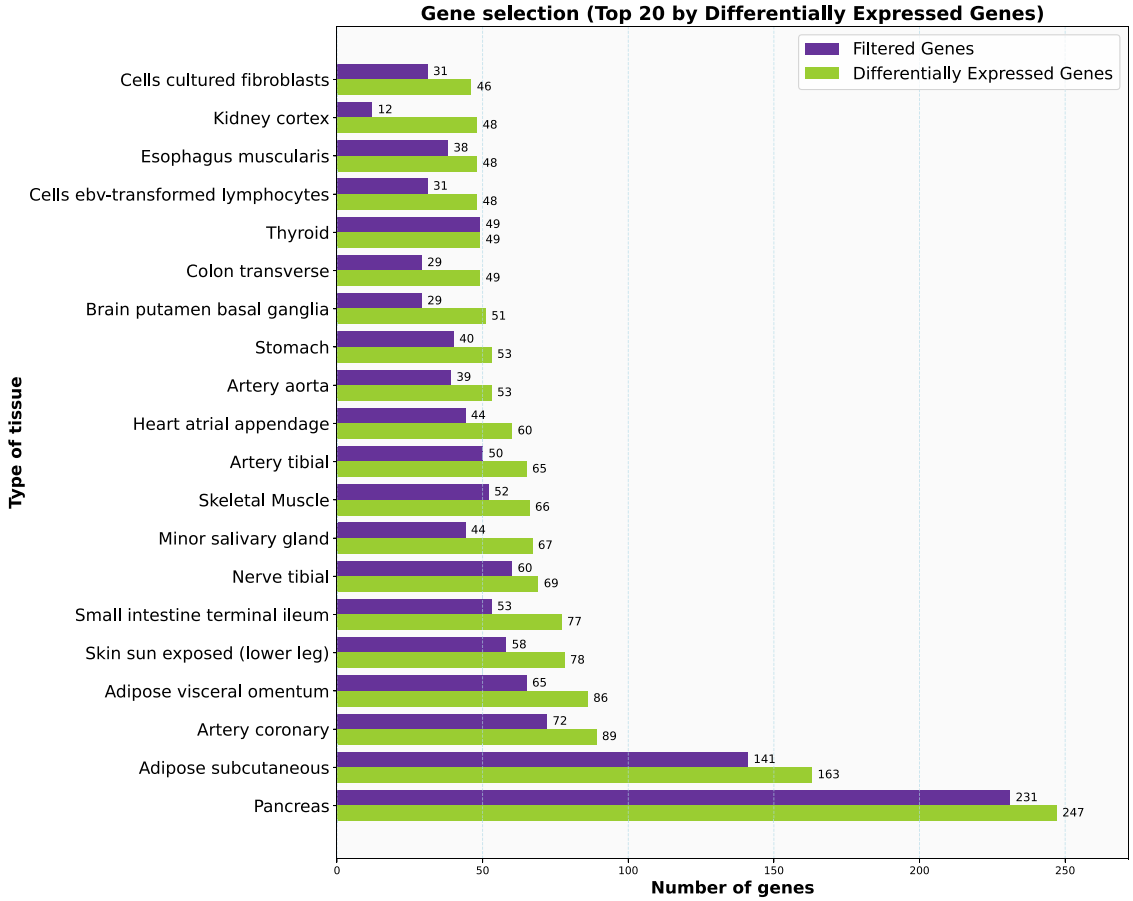

Figure 2. Comparison of differentially expressed and filtered genes between sexes in the top 20 tissues, excluding Y chromosome and low-expression genes.

Table 4. Adipose subcutaneous significant pathways

Pathway	N genes	FDR value
Serum amyloid A proteins	3	0.0011
Acute phase	3	0.0233
HDL	3	0.0233

Table 5. Adipose subcutaneous gene enrichment

GE description	N genes	P-value
Immunoglobulin complex	4	1.89E-3
FCGR3A-mediated IL10 synthesis	5	3.33E-3
Binding and Uptake of Ligands by Scavenger Receptors	5	4.09E-3
Regulation of complement cascade	5	4.99E-3

Table 6. Adipose visceral omentum significant pathways

Pathway	N genes	FDR value
Sex differentiation	7	0.0025
Homeobox protein	4	0.0038
Embryonic limb morphogenesis	5	0.0078
Response to lipid	9	0.0078

subsequently utilized to construct DCNs. Genes expressed on the Y chromosome were also identified and excluded from the analysis. These genes were retrieved from the NCBI Gene Database [14] using the query *Homo sapiens Y chromosome*. Both gene symbols and their aliases were considered during this extraction. The exclusion of Y chromosome-related genes in this analysis is due to the need to prevent the differential analysis from being distorted by sex-specific effects. Y-chromosome genes are intrinsically expressed only in males, and their inclusion would lead to obvious and biologically trivial differences in expression and causal structure.

Data pre-processing

The extracted data's pre-processing phase was divided into two main steps: the identification of differentially expressed genes using gene counts and the filtering and cleaning of TPM data. Subjects with unique values across all fields and complete data in the age and gender fields were included in the study. Duplicate entries and records with incomplete data in the specified fields were excluded. The numbers of males and females were balanced by randomly sampling the larger set to match the size of the smaller set, ensuring that differences in sample sizes between sexes would not affect subsequent analyses. The PyDESeq2 library [15, 16] was used for DEA between sexes.

Genes showing differential expression were defined as those with an adjusted p-value below 0.05 and an absolute fold change greater than 1. Bootstrap resampling involved randomly selecting 90% of the samples with reinsertion during each iteration, ensuring that the proportion of males and females remained constant. Across 100 iterations, the frequency with which each gene was identified as significant was tracked. Only

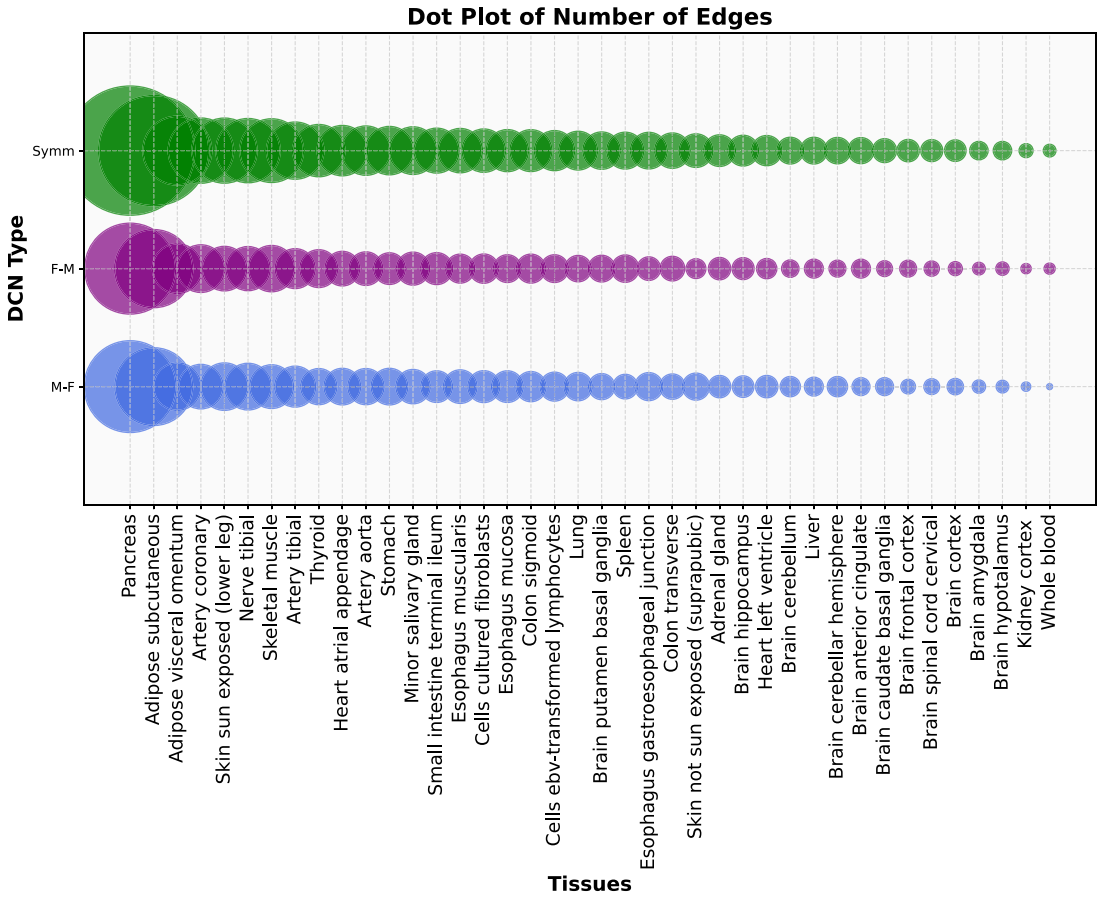

Figure 3. Dot plot illustrating the number of edges across various tissues, categorized by DCN Type: Symmetric (first line), F–M (second line), and M–F (third line), with dot size indicating edge count and tissues sorted by total descending order based on the total number of DCN edges.

Table 7. Lung tissue gene enrichment

GE description	N genes	*P*-value
Factor: HOXA10; motif: NRTCGTAAANN	4	4.03E-2
Inactive sex chromosome	1	4.99E-2
Poly(C) RNA binding	1	4.99E-2

Table 8. Brain tissues gene enrichment

GE description	N genes	*P*-value
HOXA10 Regulatory Pathway	4	2.20E-2
Inactive sex chromosome	1	4.98E-2
Poly(C) RNA binding	1	4.98E-2

Table 9. Brain putamen basal ganglia tissue gene enrichment

GE description	N genes	*P*-value
Transcription regular activity	6	2.08E-2
RNA polymerase II cis-regulatory region sequence-specific DNA binding	5	2.29E-2
cis-regulatory region sequence-specific DNA binding	5	2.5E-2
RNA polymerase II transcription regulatory region sequence-specific DNA binding	5	4.9E-2

genes appearing as significant in at least 75% of iterations were considered robust and included in the final results. As mentioned earlier, a balanced number of samples from males and females was selected prior to the gene differentiation phase to mitigate the potential influence of sample size disparities between sexes. In addition, a possible distribution shift effect could, in theory, influence the causal discovery phase, although in practice, thanks to the rigorous selection and preprocessing procedures of GTEx data, the risk of introducing bias is very low.

The subsequent preprocessing of TPM data was conducted using the differentially expressed genes identified in the previous stage as a filter. The data were subsequently categorized into two distinct groups (males and females) and further refined by excluding genes that had zero counts in more samples than the total number of samples in each dataset minus ten. As a result, CNs for men and women will share the same number and type of genes, but will show different causal edges between them, allowing the DCNs computation.

Differential causal networks

A DCN is produced by subtracting one CN from another or performing a symmetric difference between two CNs with the same nodes but different edges. The PC (Peter–Clark) algorithm [17] was applied to perform the causal discovery phase on filtered TPMs data, obtaining four CNs for each tissue: two for males (one for

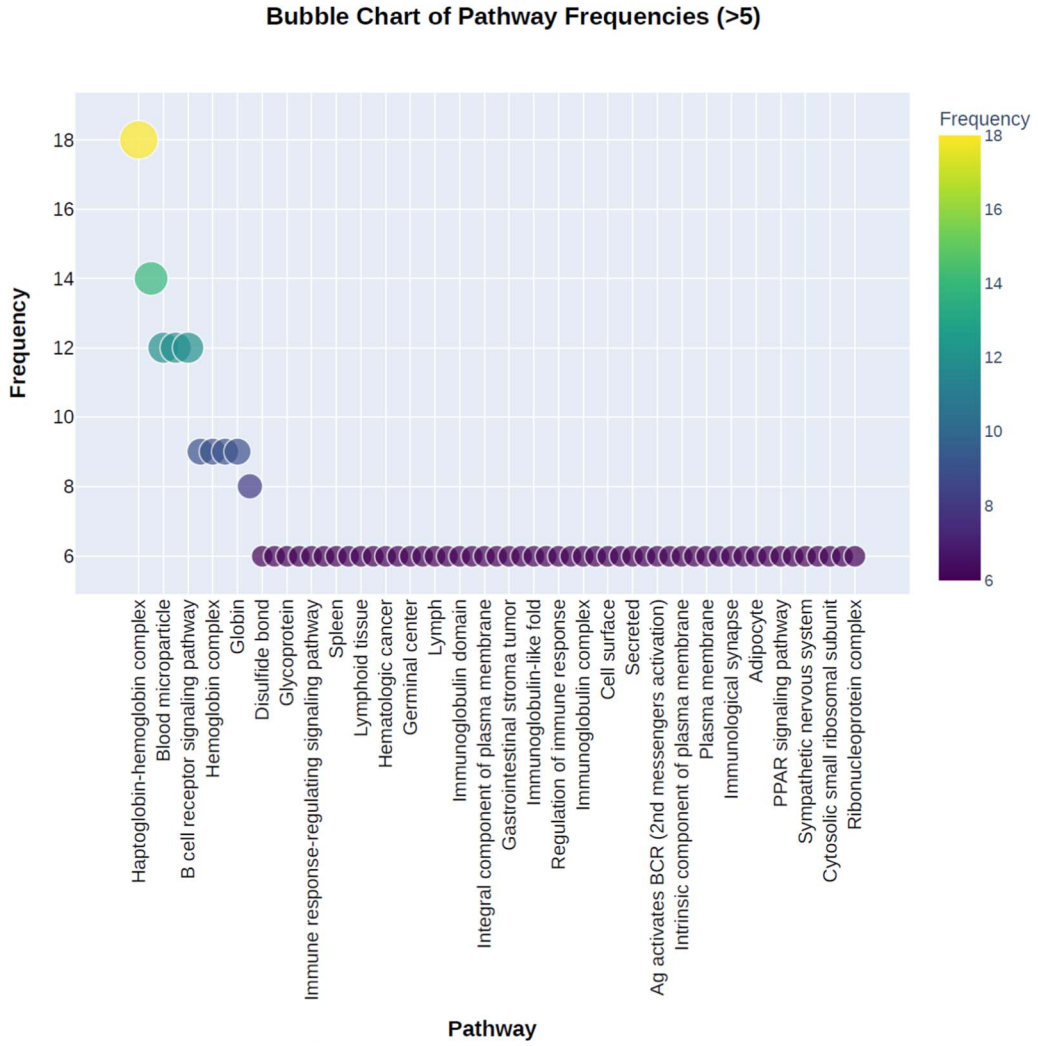

Figure 4. Bubble plot showing pathways occurring more than five times, with bubble size and a purple-to-yellow gradient reflecting frequency.

each age range) and two for females (one for each age range). As a next step, as proposed in [6], three types of differences between male and female CNs for each tissue were computed:

1. Female CN is subtracted from the male CN: the (Asymmetric) DCN contains all the edges of male CN that are not in female CN;
2. Male CN is subtracted from the female CN: the (Asymmetric) DCN contains all the edges of female CN that are not in male CN;
3. Symmetric difference between male and female CN is constructed: the (Symmetric) DCN contains all the edges of the female CN that are not in the male CN and all the edges of the male CN that are not in the female CN. In other words, it is the union of all the edges in the female CN and the male CN described above.

In summary, the investigation of sex-specific mechanisms was conducted using three distinct DCN types constructed based on the presence or absence of causal edges inferred independently in male and female subgroups using the stable PC algorithm. Specifically, edges classified as M–F represent causal interactions detected only in the male subgroup, indicating male-dominant mechanisms; F–M edges, conversely, denote interactions unique to females; while symmetric edges correspond to the union of both sex-specific interactions.

In addition, considering the definitions proposed in [18], the DCNs performed in this study consider the difference in support:

$$E^{\text{diff}} = \left\{(i,j) : A_{ij}^{(1)} \neq A_{ij}^{(2)}\right\} \quad (1)$$

where E^{diff} is the edge difference set, $A^{(1)}$ and $A^{(2)}$ denote the adjacency matrix respectively of the graphs $G^{(1)}$ and $G^{(2)}$ of which the difference is performing [18]. This definition considers structural differences between networks, in line with what has been proposed in [6].

PC algorithm (from the names of its creators, P.S. and C.G.) was used to derive the structure of CNs [17].

Causal directionality was inferred using the PC algorithm, which determines causal structure by testing for conditional independence rather than relying on correlations alone [17]. Assuming causal sufficiency, faithfulness, and the Causal Markov condition, the algorithm identifies v-structures and applies orientation rules to resolve the direction of causal edges. This approach enables the discrimination of genuine causal relationships from spurious associations.

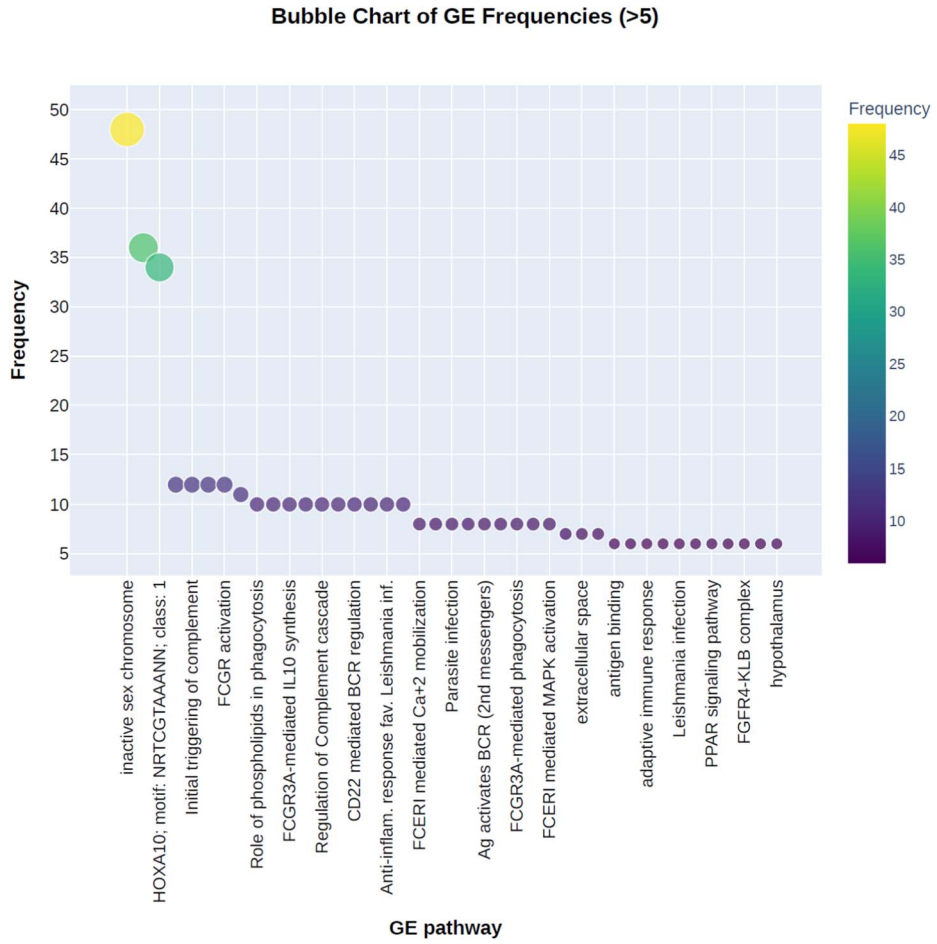

Figure 5. Bubble plot showing GE pathways occurring more than five times, with bubble size and a purple-to-yellow gradient reflecting frequency.

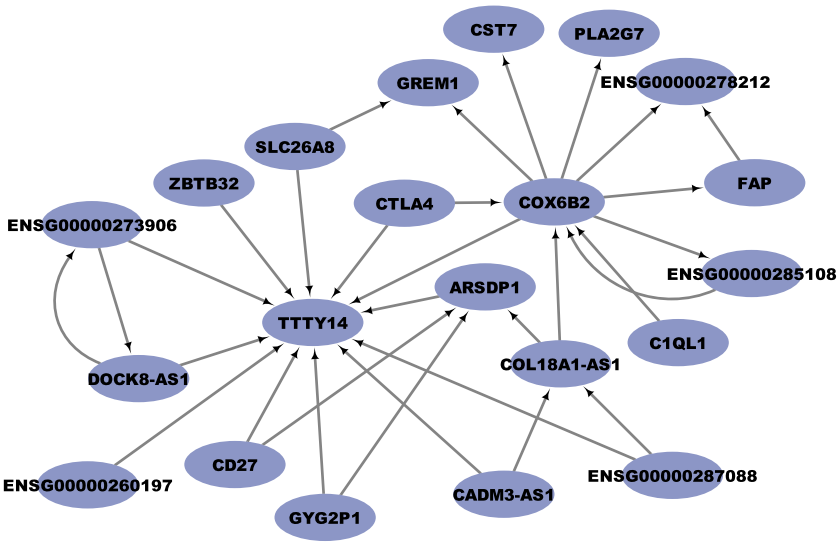

Figure 6. Male-Female DCN for pancreas tissue.

The CNs were obtained using the gCastle Python package [19], which implements different algorithms for PC. In this work, the stable variant of the PC algorithm (variant = "stable") was adopted. It is an efficient method for causal discovery that eliminates order-dependencies in its implementation [20].

The output of the PC algorithm can be order-dependent, i.e. the order in which the variables are presented can affect the results. The authors of [20] proposed the stable version of it with several modifications to the algorithm to remove this dependency in the various stages of the algorithm and ensure reliable results.

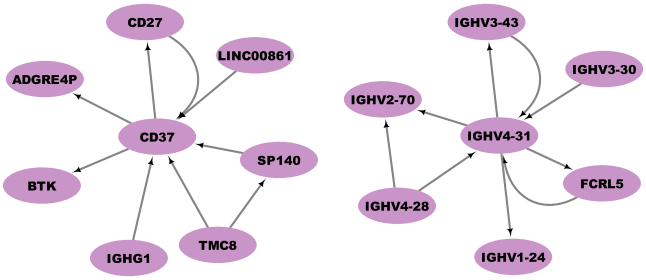

Figure 7. Female-Male DCN for pancreas tissue.

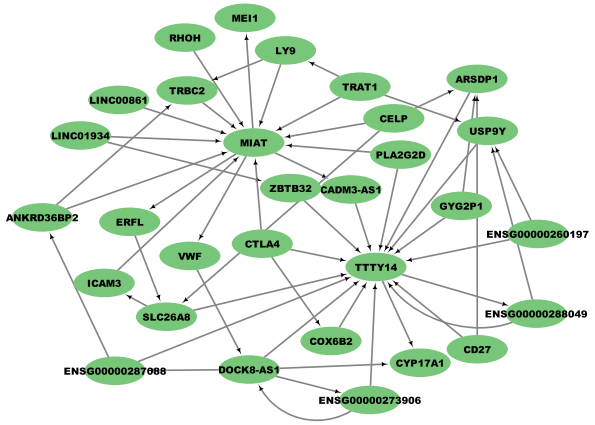

Figure 8. Symmetric DCN for pancreas tissue.

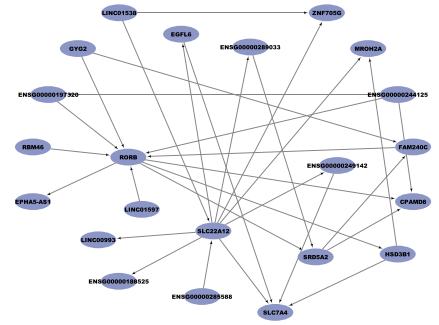

Figure 9. Male-Female DCN for adipose subcutaneous tissue.

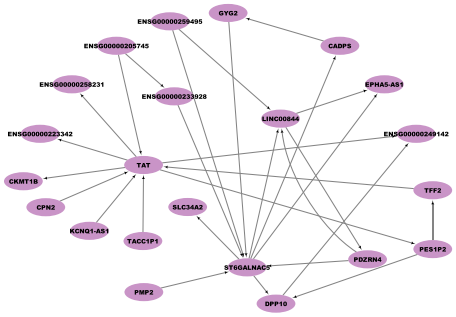

Figure 10. Female-Male DCN for adipose subcutaneous tissue.

The default parameter settings were kept: ci test = "fisherz" (independent test), alpha = 0.05 (significance level), and no background information was added (priori knowledge: None).

Additional evidence supporting this approach is available in the supplementary material.

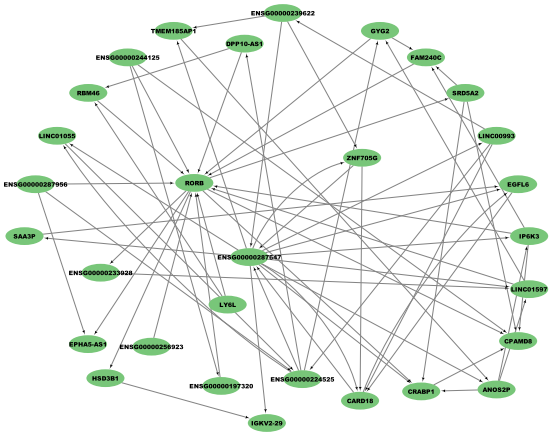

Figure 11. Symmetric DCN for adipose subcutaneous tissue.

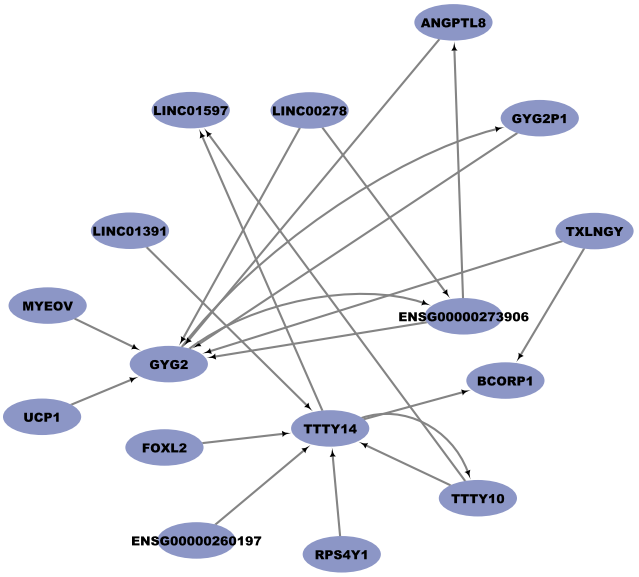

Figure 12. Male-Female DCN for adipose visceral omentum tissue.

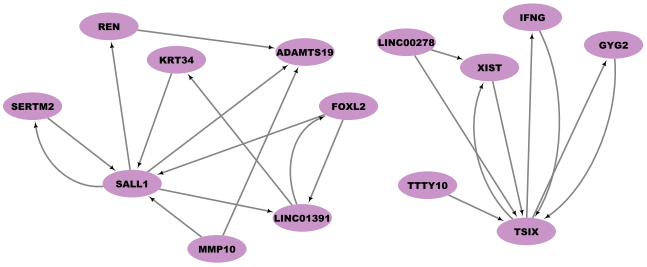

Figure 13. Female - Male DCN for adipose visceral omentum tissue.

Table 10. Heart left ventricle tissue gene enrichment

GE description	N genes	P-value
HOXA10 Regulatory Pathway	4	6.85E-3
Inactive sex chromosome	1	4.98E-2
Poly(C) RNA binding	1	4.98E-2

For each tissue, each of the two groups (men and women) is further divided into two age sets: subjects aged between 20 and 49 years and subjects aged between 50 and 79. Therefore, a CN is generated. To enhance the robustness of the resulting CNs, the

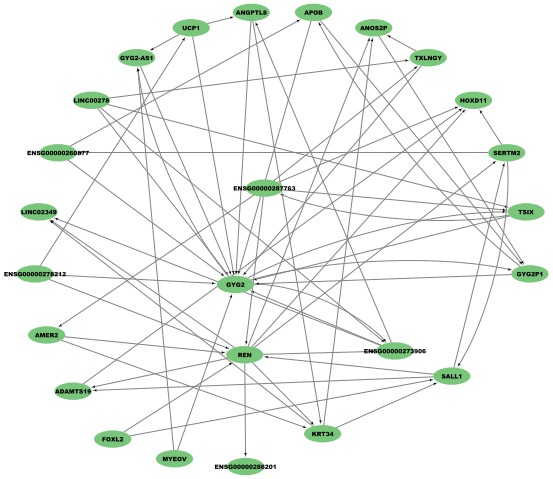

Figure 14. Symmetric DCN for adipose visceral omentum tissue.

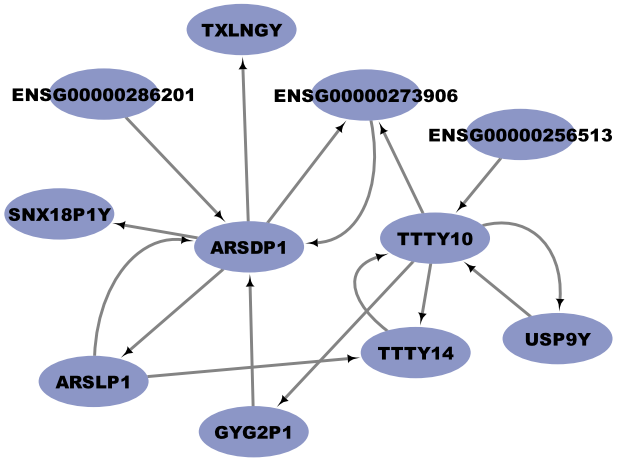

Figure 15. Male-Female DCN for lung tissue.

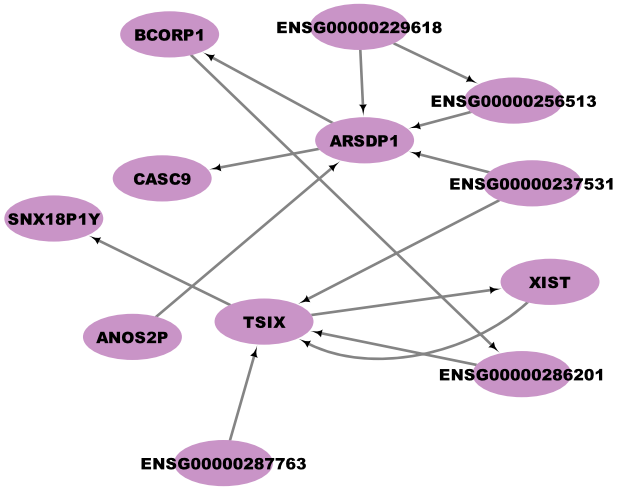

Figure 16. Female-Male DCN for lung tissue.

PC algorithm is applied repeatedly on each group (100 iterations). Only causal edges present in at least 70% of the iterations were considered eligible as causal relationships between genes. The result of this step is four CNs that are subsequently aggregated by combining the relationships found for the two age ranges while maintaining the distinction between sexes. As the last step of this phase, the DCNs were calculated using the NetworkX Python library [21] to identify the differences in causal relationships between two conditions.

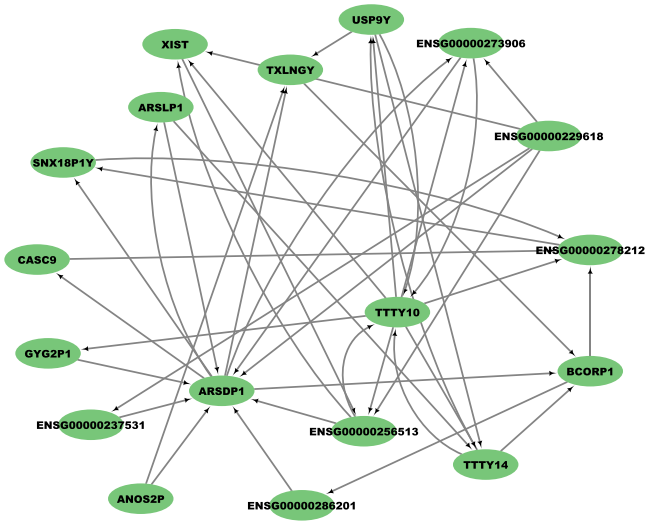

Figure 17. Symmetric DCN for lung tissue.

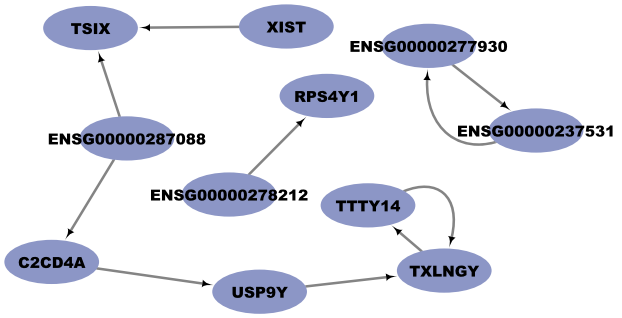

Figure 18. Male-Female DCN for brain amygdala tissue.

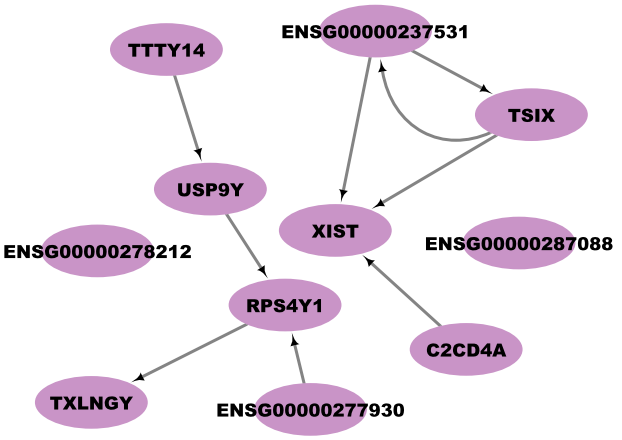

Figure 19. Female-Male DCN for brain amygdala tissue.

Table 11. Heart atrial appendage tissue significant pathways

Pathway	N genes	FDR value
Adipocyte	3	6.80E-3
Preadipocyte	3	6.80E-3
Subcutaneous adipose tissue	2	6.80E-3
Perirenal adipose tissue	2	6.80E-3

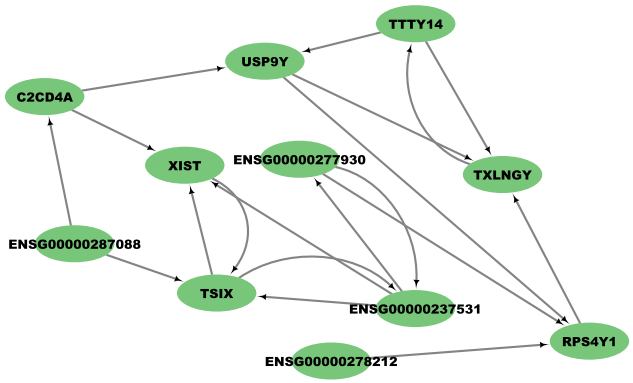

Figure 20. Symmetric DCN for brain amygdala tissue.

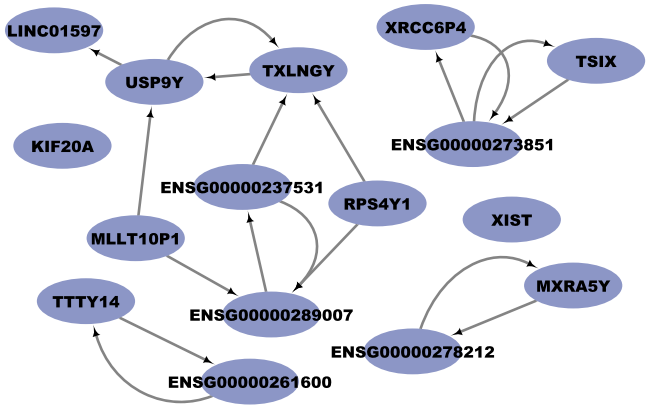

Figure 21. Male-Female DCN for brain anterior cingulate tissue.

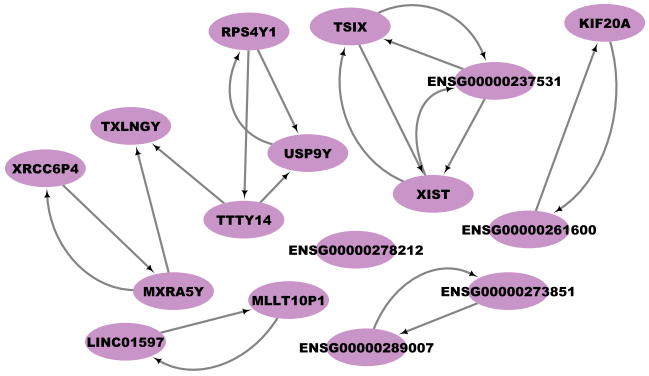

Figure 22. Female-Male DCN for brain anterior cingulate tissue.

Enrichment analysis

Pathway enrichment was performed by utilizing the KEGG pathway database [22] and the Biological Process category of Gene Ontology [23], providing a way to analyze and interpret the functions of genes based on their annotations [24]. Functional enrichment analysis was conducted using the STRING enrichment app of Cytoscape software, while gene enrichment analysis was performed on gProfiler [25].

Results

Figure 2 illustrates the number of differentially expressed genes and the corresponding number of filtered genes, considering only the top 20 tissues. Pancreas tissue exhibited the highest number

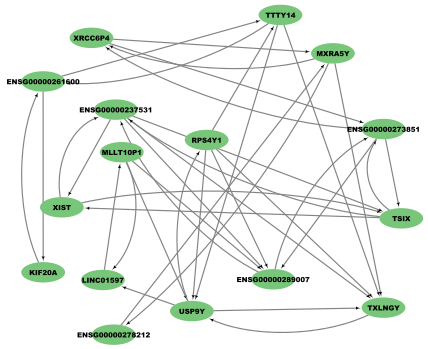

Figure 23. Symmetric DCN for brain anterior cingulate tissue.

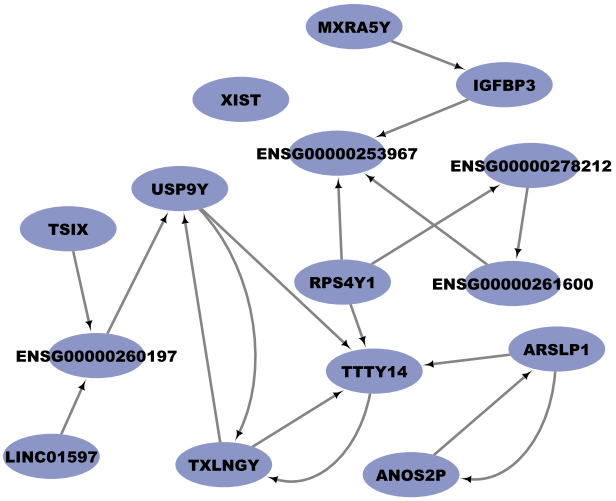

Figure 24. Male-Female DCN for brain caudate basal ganglia tissue.

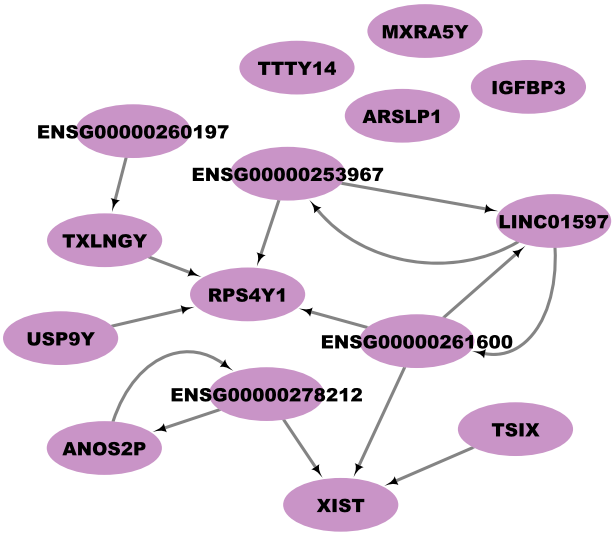

Figure 25. Female-Male DCN for brain caudate basal ganglia tissue.

of sex-specific differentially expressed genes, followed by subcutaneous adipose tissue. In contrast, brain-associated tissues displayed substantially fewer differentially expressed genes, with blood tissue showing the lowest count overall, as detailed in the Table 1.

Focusing on DCN results, DCNs facilitate the exploration of causal edges that differ between two conditions, in this case, sex.

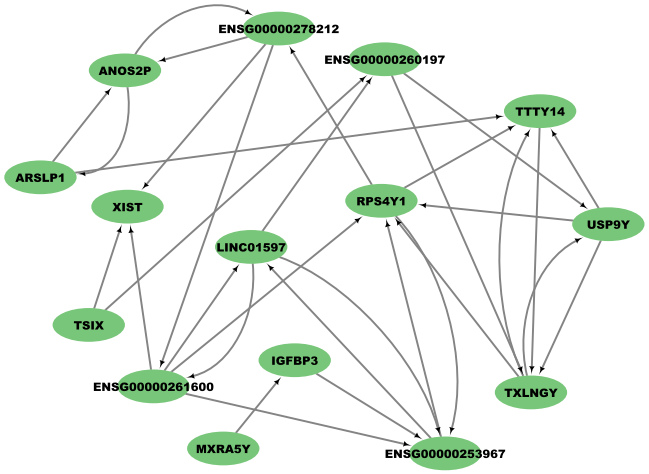

Figure 26. Symmetric DCN for brain caudate basal ganglia tissue.

The numbers of edges in the three types of computed DCNs are summarized in Fig. 3.

Since the symmetric difference represents the union of results achieved by swapping the minuend network and the subtrahend network, the number of edges in all symmetric DCNs equals the sum of the other two DCNs for each tissue, respectively. Furthermore, given the non-zero differences, it can be stated that the causal relationships between the differentially expressed genes differ between men and women, i.e. there are edges that are present in the male sex, but not in the female sex and vice versa.

Additionally, to complete the analysis, functional and gene enrichment analyses were performed. Functional enrichment allows for the contextualization of the evidence found, helping to identify significant pathways in which nodes are involved. On the other hand, gene enrichment analysis may reveal the functional significance of GE changes. The use of both STRING and classical gene enrichment analyses within Cytoscape is beneficial, as they are complementary. Although STRING focuses on protein-protein interactions, hence functional networks and possible therapeutic targets, classical gene enrichment identifies overrepresented pathways and biological functions associated with particular genes, thus contextualizing findings. Together, they enhance the understanding of complex biological mechanisms and support the validation of results, making them valuable tools for comprehensive genomic analyses.

The enrichment analysis shows an overlap between known biological pathways and those highlighted through the structure of DCNs, underscoring the biological relevance of the inferred causal relationships.

The Fig. 4 illustrates the frequency of pathways enriched in more than 5 tissues, taking into account all analyzed tissues in all DCNS.

Haptoglobin-hemoglobin complex, Extracellular region pathways followed by Immunoglobulin, B cell receptor (BCR) signaling and Blood microparticle pathways were the most frequent. Extracellular pathways involve various mechanisms outside the cells, implicating extracellular vesicles, exocytosis and cell signaling, which are crucial for cell communication [26], regulation, and homeostasis. Immunoglobulin pathways include a class of antibodies for infection control being part of the human immune system. Immunoglobulins are vital to the immune system, functioning as both stimulators and inhibitors of inflammation, and their positive effects are noted across multiple diseases [27]. BCR signaling pathways are important for the activation of B cells with subsequent immune response (adaptive immunity). In addition, BCR pathway is central to the onset of B-cell malignancies and consequently, could be a therapeutic target [28]. Blood microparticles, small vesicles released from cells into the bloodstream are involved in pathways related to inflammation, coagulation, cellular communication and apoptosis. High level of blood microparticles are usually related to systemic inflammation and thrombotic disorders [29].

In the Fig. 5, the frequency of the most represented gene enrichment pathways, identified in more than 5 tissues, is reported, considering all the analyzed tissues in all the DCNs.

The most frequent pathways refer to the inactivation of the sex chromosome, the binding of RNA with poly(C) sequences, and the HOXA10 gene, which has been shown to be involved in developmental processes and is regulated by sex steroids [30]. The motif NRTCGTAANN is a specific sequence pattern recognized in these latter pathways. The inactivation of one X chromosome in females (XX) serves to balance the dosage of sex chromosomes [31]. Therefore, the presence of these pathways could serve as evidence of the validity of the proposed methodology, demonstrating its ability to identify different relationships between men and women. The poly(C) binding proteins are involved in mRNA stabilization, translational activation, and translational silencing, for which the pathways that comprise them are quite complex for the role they play [32]. HOXA10 gene with different motifs is associated to pathways for gene regulation, cell migration and cancer progression. In particular, there is an involvement of these gene pathways with both women endometriosis [33] and prostate cancer [34].

Further details on the pathways identified in the DCNs of individual tissues are provided in the following subsections and in Supplementary material.

The top four significant pathways (with the lowest FDR values for pathways identified by String enrichment and the lowest p-values for pathways identified by gene enrichment) highlighted by the applied enrichments are shown. The representations of all tissues DCNs are reduced for better visualization. Only the top 2 genes (nodes) with the highest degree and their 1-hop neighbors for both incoming and outgoing edges are considered. An exception is made for tissues related to the brain, kidney cortex and blood tissues given the small size of the networks.

Pancreas tissue

Pancreatic tissue exhibited the highest number of differentially expressed genes between sexes. As a key component of the endocrine system, the pancreas is implicated in diseases such as diabetes mellitus and cancer, where sex-specific GE may influence disease susceptibility and immune responses [35, 36]. Investigating the underlying causal relationships and associated pathways is therefore of particular interest. Pathway enrichment analysis, summarized in Table 2, revealed that the most significantly affected pathways in the DCNs are immune-related, consistent with the central role of inflammation and immune dysfunction in both diabetes and pancreatic cancer [37, 38].

In addition, gene enrichment analysis indicates which genes belong to interrelated functions, as illustrated in the Table 3. In fact, as already shown in the pathway enrichment, genes related to the immune response turn out to be among the most significant (lower p-value), and, as mentioned, have a non-secondary role in pancreatic-related diseases. Given the size of the DCNs

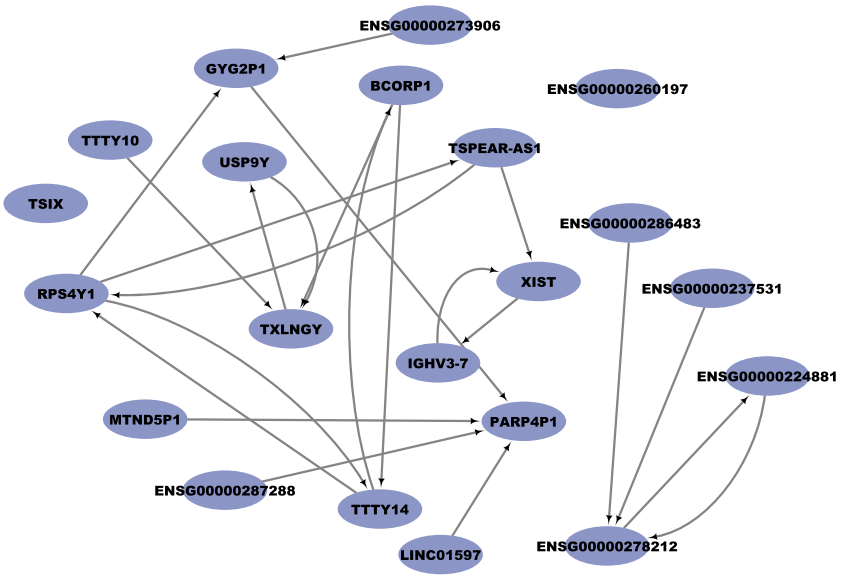

Figure 27. Male-Female DCN for brain cerebellar hemisphere tissue.

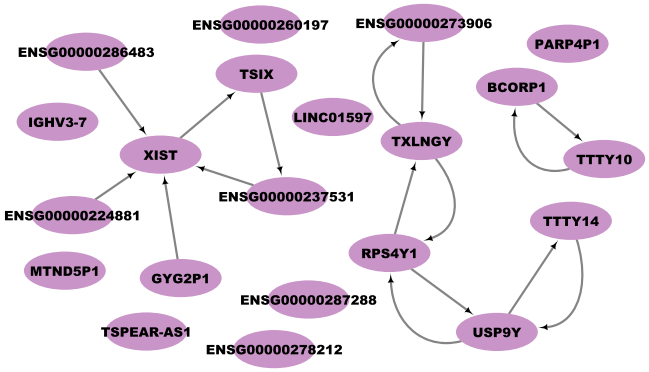

Figure 28. Female-Male DCN for brain cerebellar hemisphere tissue.

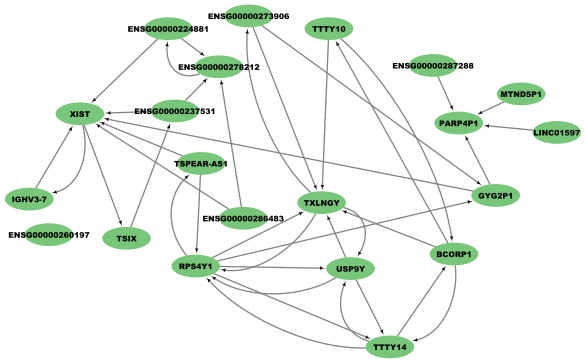

Figure 29. Symmetric DCN for brain cerebellar hemisphere tissue.

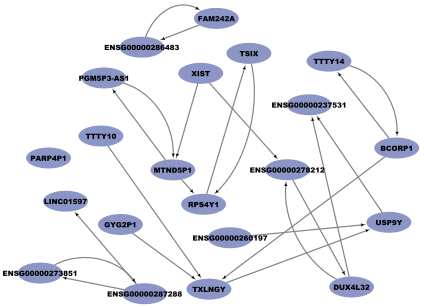

Figure 30. Male-Female DCN for brain cerebellum tissue.

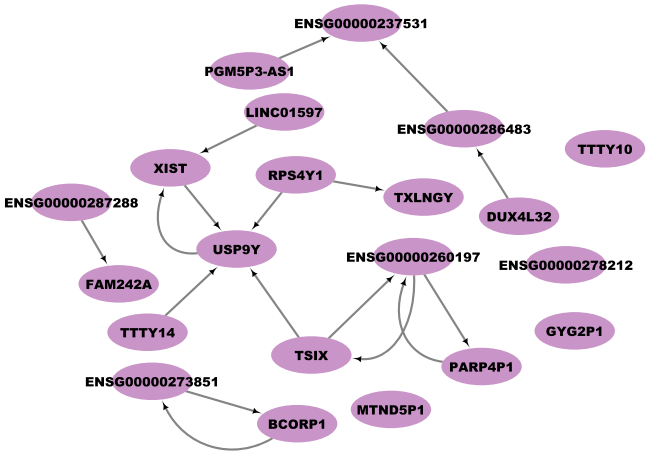

Figure 31. Female-Male DCN for brain cerebellum tissue.

obtained for this tissue, having the largest number of differentially expressed genes, reduced DCNs have been depicted in the Figs 6–8.

Adipose tissues

Adipose tissue, composed predominantly of adipocytes, was analyzed in two forms: subcutaneous and visceral (omentum).

Pathway enrichment analysis of subcutaneous adipose tissue identified three principal pathways linked by their roles in inflammation and lipid metabolism (Table 4). Serum amyloid A proteins, key mediators of the acute-phase response, also regulate cell–cell communication and modulate inflammatory, immunologic, neoplastic, and protective processes [39]. Chronic activation of the acute-phase response by adipose tissue has been implicated in the elevated inflammatory state of diabetes and its cardiovascular sequelae [40].

Additionally, pathways related to high-density lipoprotein (HDL) metabolism were enriched. Given the critical role of HDL

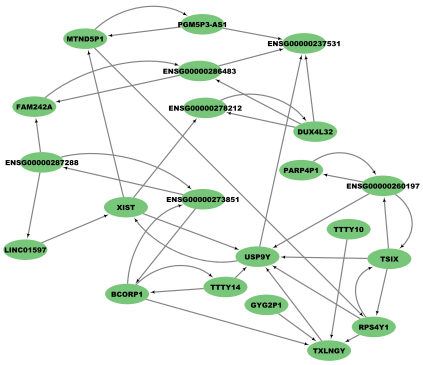

Figure 32. Symmetric DCN for brain cerebellum tissue.

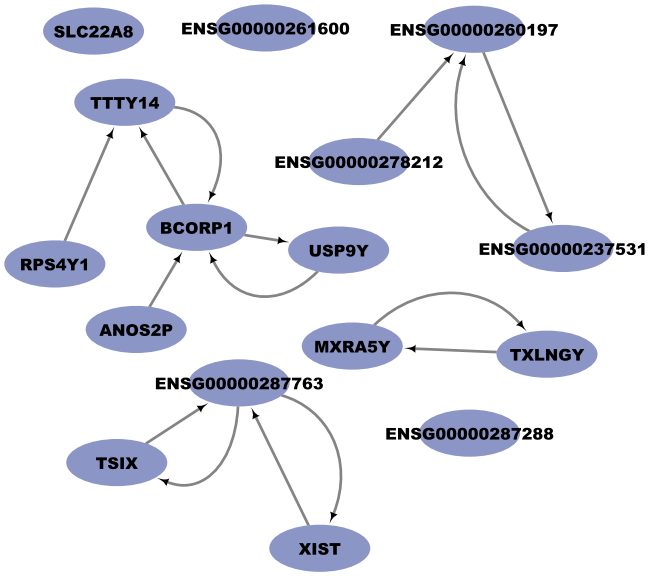

Figure 33. Male-Female DCN for brain cortex tissue.

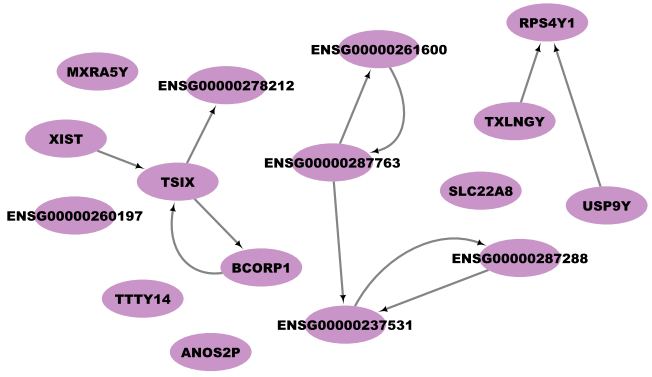

Figure 34. Female-Male DCN for brain cortex tissue.

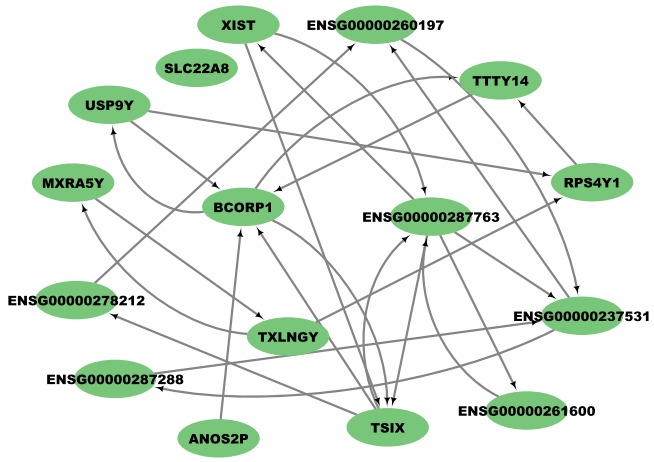

Figure 35. Symmetric DCN for brain cortex tissue.

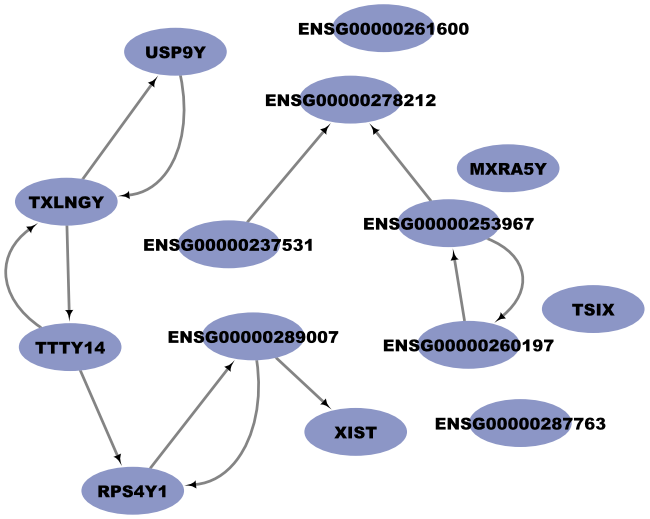

Figure 36. Male-Female DCN for brain frontal cortex tissue.

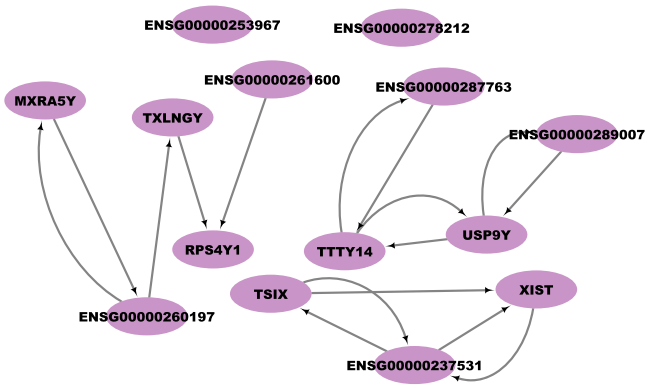

Figure 37. Female-Male DCN for brain frontal cortex tissue.

in cholesterol homeostasis and cardiovascular protection [41], and the importance of adipose tissue in HDL dynamics, these findings offer insights into potential therapeutic targets aimed at enhancing cardiometabolic health.

Gene enrichment for adipose subcutaneous tissue (Table 5) shows the presence of pathways related to immune response and inflammatory processes. Moreover, the pathway related to the binding and the uptake of ligands plays a crucial role in lipid metabolism by mediating the uptake and binding of various ligands, including lipoproteins and fatty acids, in adipose tissue. These receptors are critical for maintaining metabolic homeostasis and facilitating lipid transport [42]. Given the size of the DCNs obtained, reduced DCNs have been depicted in the Figs 9–11.

Table 6 shows significant pathways identified by String analysis applied to adipose visceral omentum tissue. In addition to pathways related to genes involved in sex differentiation, homeobox proteins pathway has also been identified. It involves transcription factors for cell differentiation, development, and

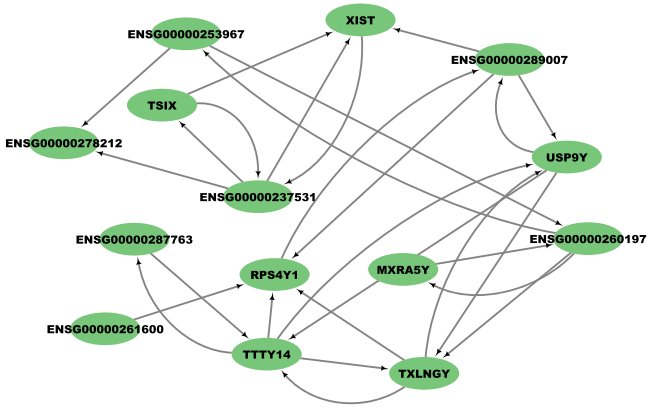

Figure 38. Symmetric DCN for brain frontal cortex tissue.

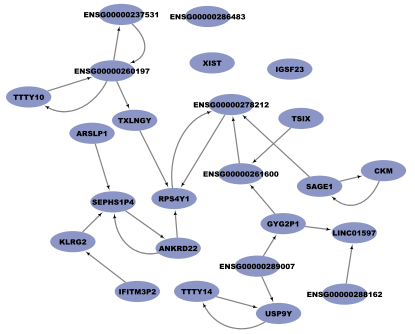

Figure 39. Male-Female DCN for brain hippocampus tissue.

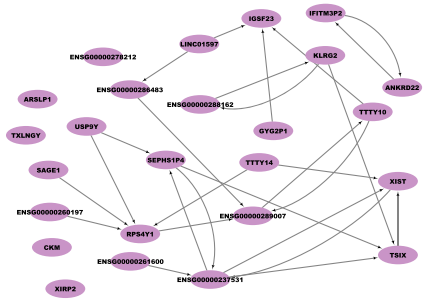

Figure 40. Female-Male DCN for brain hippocampus tissue.

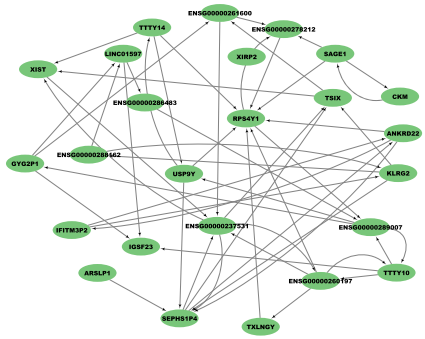

Figure 41. Symmetric DCN for brain hippocampus tissue.

regulation also in the early stages of embryonic development [43]. Moreover, genes involved in the lipid response were detected in the DCNs of the adipose visceral omentum tissue. Reduced DCNs for adipose visceral omentum tissue are illustrated in the Figs 12–14.

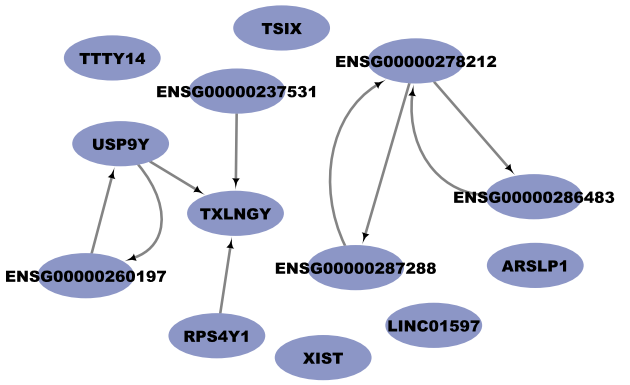

Figure 42. Male-Female DCN for brain hypothalamus tissue.

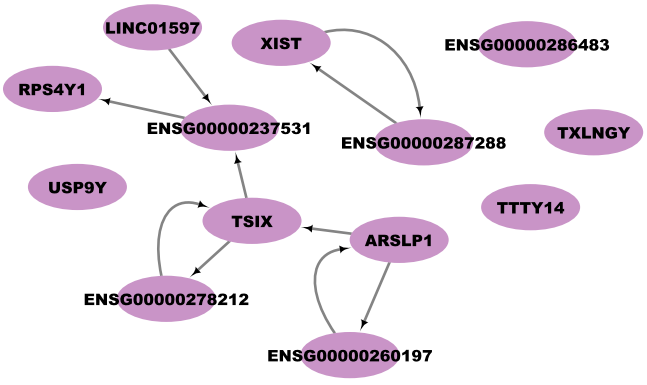

Figure 43. Female-Male DCN for brain hypothalamus tissue.

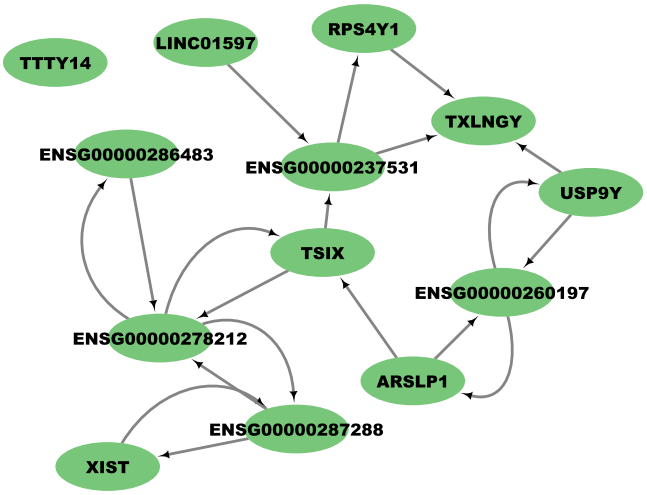

Figure 44. Symmetric DCN for brain hypothalamus tissue.

Lung tissue

The lungs, which are central to the respiratory system, demonstrated a significantly lower number of differentially expressed genes compared to other analyzed tissues. This finding may be partially due to the prolonged ischemic time commonly associated with GTEx lung samples, which can result in increased RNA degradation and, subsequently, reduced RNA quality [44].

Reduced directed CNs (DCNs) for lung tissue are shown in Figs 15–17.

STRING functional enrichment analysis for lung tissue did not identify significant pathways. However, gene enrichment analysis (Table 7) revealed genes implicated in maintaining

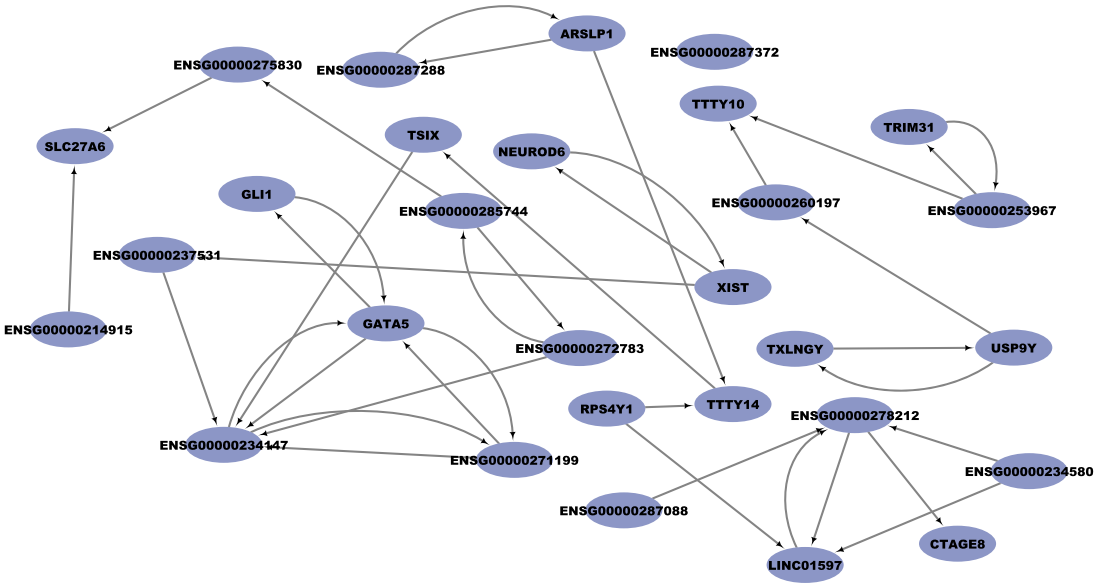

Figure 45. Male-Female DCN for brain putamen basal ganglia tissue.

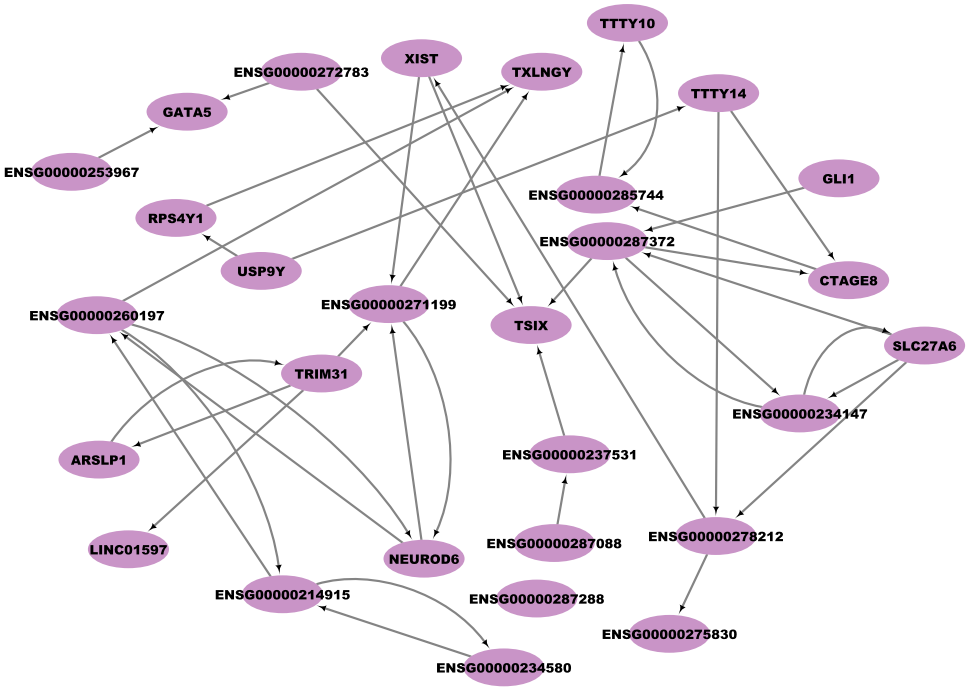

Figure 46. Female-Male DCN for brain putamen basal ganglia tissue.

X-chromosome-related transcriptional suppression (such as HNRNPU, a gene encoding for the nuclear ribonucleoprotein U). The enrichment results suggest the presence of key gene regulatory processes and RNA interactions in lung tissue. Notably, the transcription factor HOXA10, frequently upregulated in lung adenocarcinoma and implicated in cancer progression [45, 46], emerged alongside processes related to sex chromosome inactivation and RNA binding, pointing to potential mechanisms underlying lung-specific molecular regulation.

Lung tissue samples from the GTEx project are known to exhibit reduced RNA integrity due to extended ischemic intervals, potentially leading to RNA degradation and introducing systematic biases in transcriptomic measurements. In this study, additional correction procedures beyond the standard GTEx quality control pipeline were not implemented, as the primary objective was to facilitate exploratory and comparative analyses across tissues under consistent preprocessing conditions. Nevertheless, the notably lower number of differentially expressed genes and the increased sparsity of the inferred CNs in lung tissue align with the anticipated effects of compromised RNA quality. Despite the relatively low number of differentially expressed genes, causal discovery was performed using a robust ensemble strategy, and differential edges were derived from consistent differences between male and female network topologies. This supports the biological relevance of the identified sex-specific regulatory patterns in this tissue.

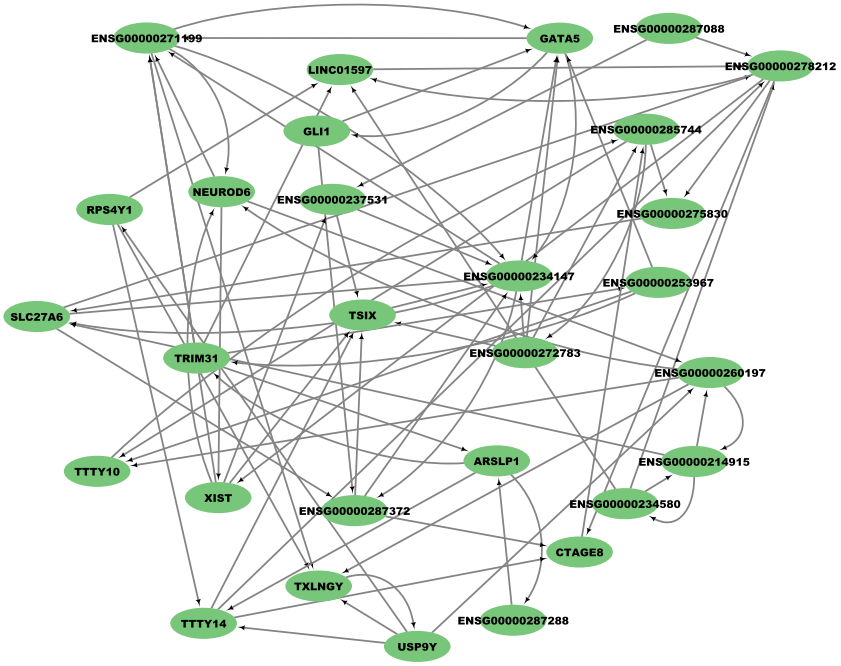

Figure 47. Symmetric DCN for brain putamen basal ganglia tissue.

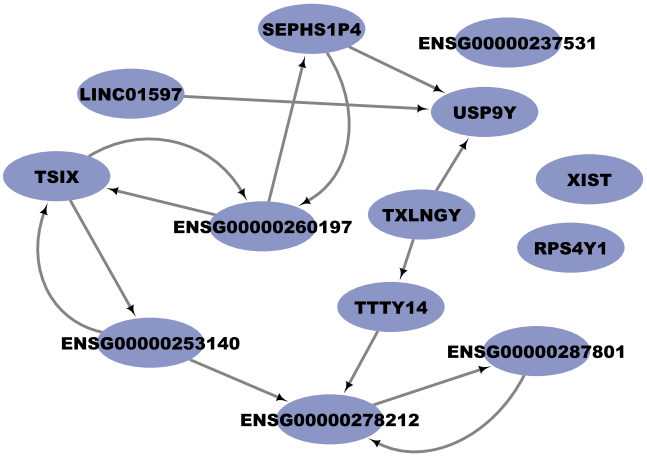

Figure 48. Male-Female DCN for brain spinal cord cervical tissue.

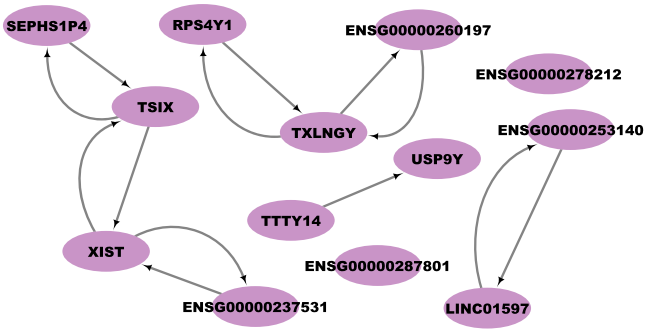

Figure 49. Female-Male DCN for brain spinal cord cervical tissue.

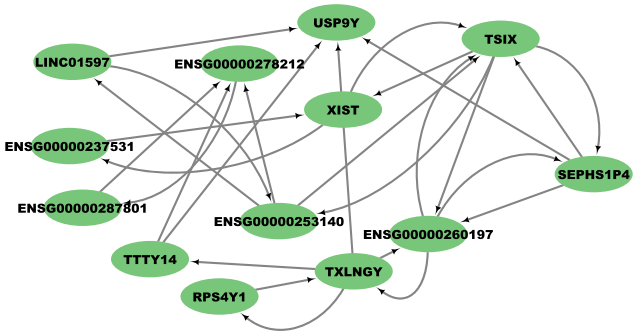

Figure 50. Symmetric DCN for brain spinal cord cervical tissue.

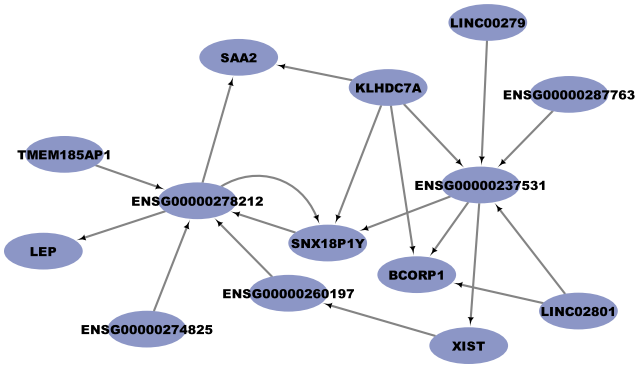

Figure 51. Male-Female DCN for heart atrial appendage tissue.

Brain tissues

GE from 11 brain regions (the amygdala, anterior cingulate cortex, caudate, cerebellar hemisphere, cerebellum, cortex, frontal cortex (BA9), hippocampus, hypothalamus, putamen, and cervical spinal cord -C1-) were analyzed. Pathway enrichment analyses were conducted for each region. Results for the cortex, cerebellum, cerebellar hemisphere, caudate, hypothalamus, frontal cortex (BA9), cervical spinal cord (C1), and amygdala are summarized in Table 8. Reported p-values represent the mean values across these tissues.

The brain hippocampus tissue showed only genes related to the pathways involving inactive sex chromosomes, whereas the brain

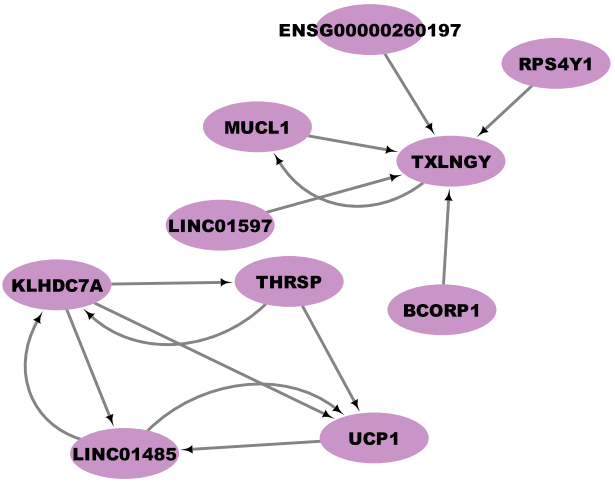

Figure 52. Female-Male DCN for heart atrial appendage tissue.

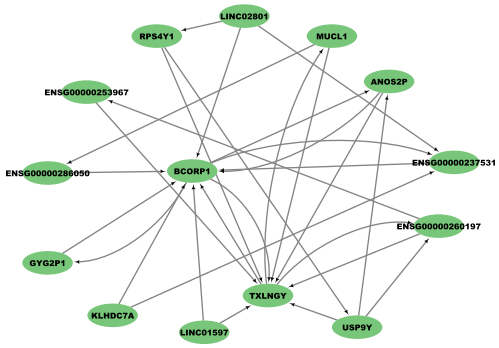

Figure 53. Symmetric DCN for heart atrial appendage tissue.

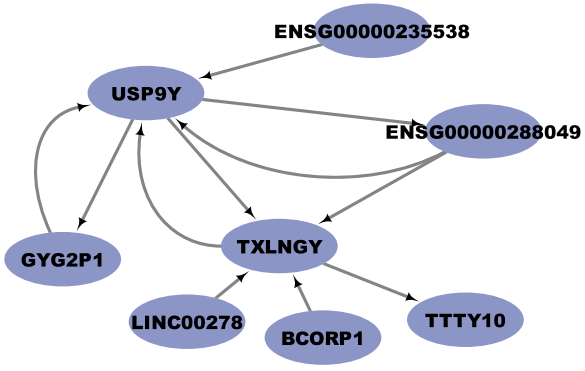

Figure 54. Male-Female DCN for heart left ventricle tissue.

putamen basal ganglia exhibited different pathways, as depicted in Table 9. All the DCNs for the brain tissues are illustrated in Figures 18–50.

Heart tissues

Reduced DCNs for the heart left ventricle and atrial appendage tissues are illustrated in Figs 51–56. In the left ventricle, gene enrichment analysis identified pathways associated with protein synthesis (poly(C) RNA binding), cell proliferation (HOXA10 regulatory pathway), and X chromosome inactivation (Table 10), similarly observed in some brain tissues. For the atrial appendage, pathways related to adipocyte biology were enriched (Table 11),

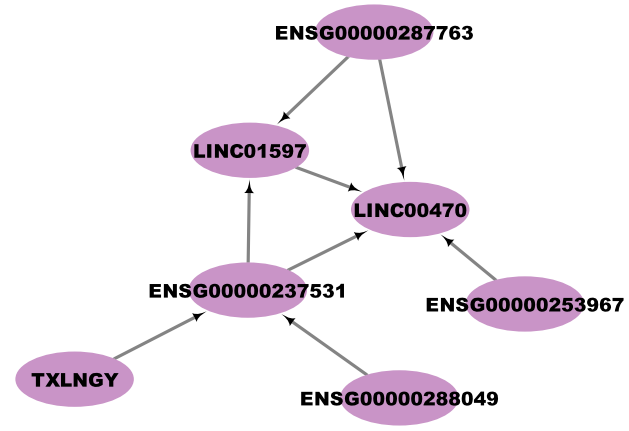

Figure 55. Female-Male DCN for heart left ventricle tissue.

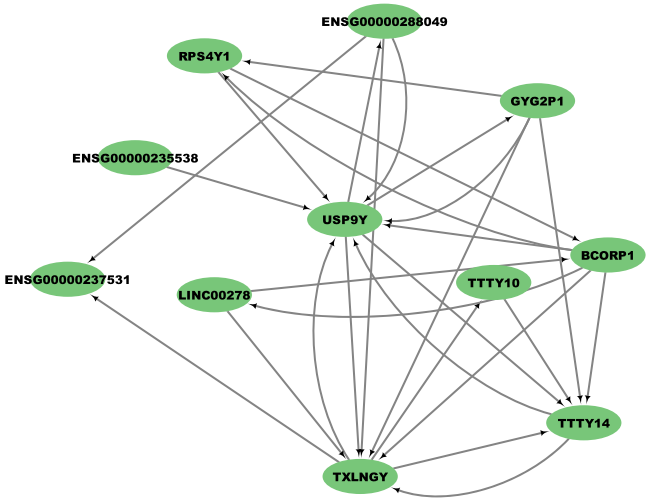

Figure 56. Symmetric DCN for heart left ventricle tissue.

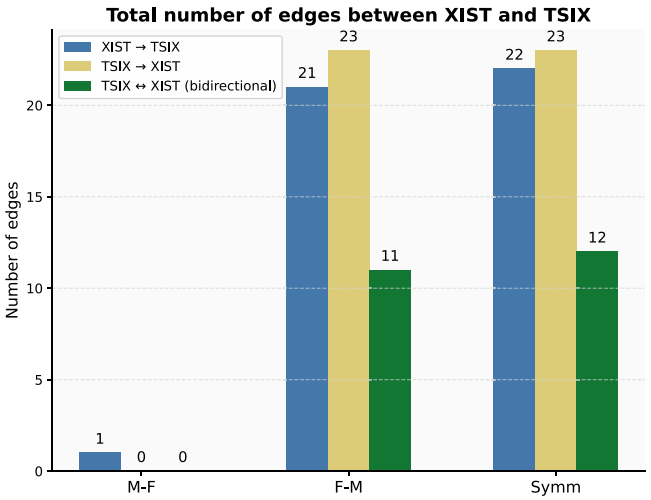

Figure 57. Comprehensive results of DCN connections between XIST and TSIX across all analyzed tissues. *Alt Text:*

consistent with evidence that abnormal adipose tissue accumulation can impact cardiac function and promote thrombus formation [47].

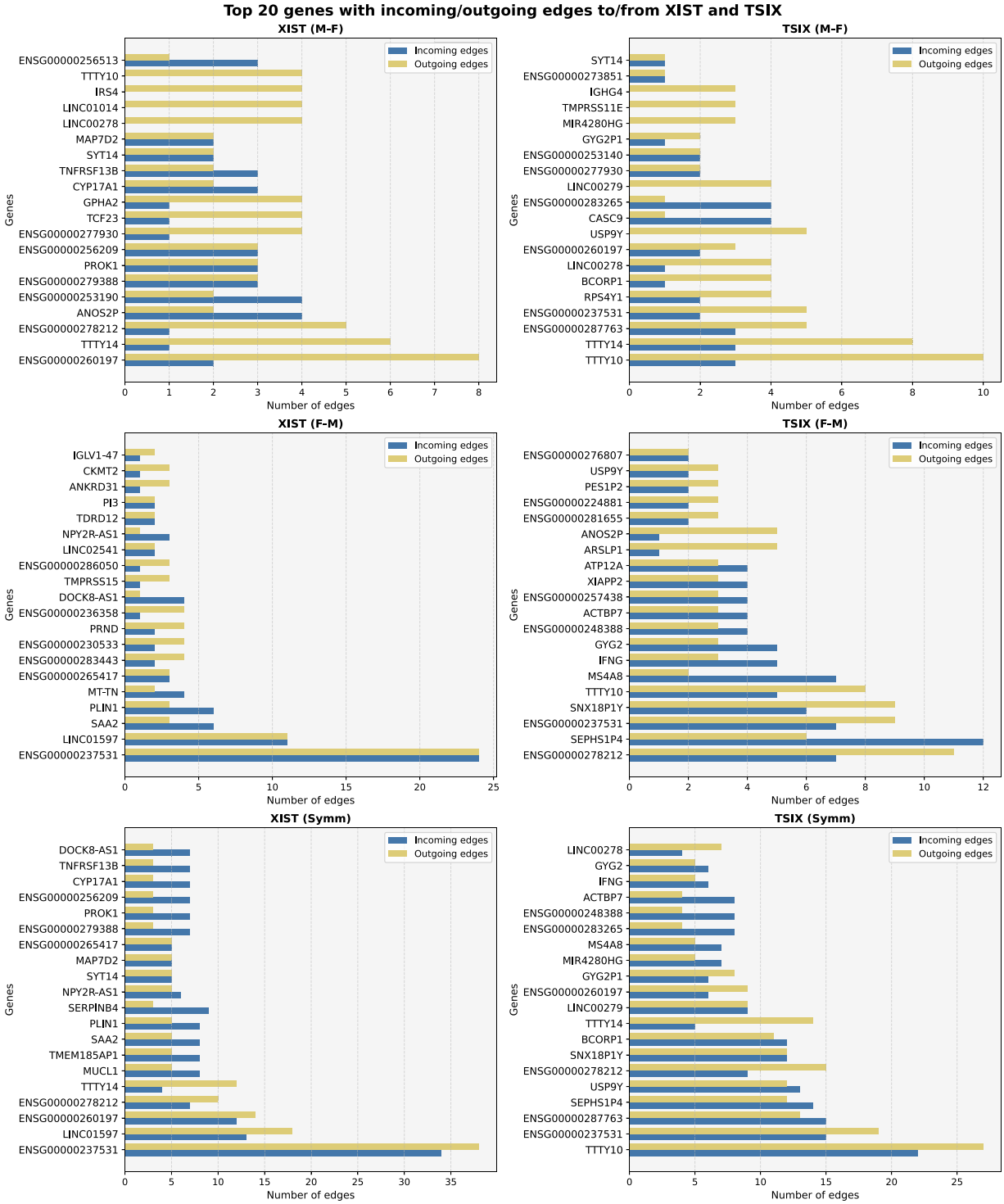

Figure 58. Overview of top 20 genes having incoming/outgoing edges to/from XIST and TSIX in DCNs across all analyzed tissues.

Discussion

This study provides insights into the regulatory mechanisms highlighting the complexity and specificity of genetic interactions in determining both biological and pathological states. The analysis consistently identified strong and reproducible interactions between XIST and TSIX genes that play a central role in the X-chromosome inactivation process, as evidenced by pathway enrichment. As shown in Fig. 57, comparing DCNs between men minus females, only the connection from XIST to TSIX was found in one network. However, females minus males DCNS showed a total of 23 edges linking these two genes, with 22 edges directed from XIST to TSIX and 21 edges directed from TSIX to XIST. Additionally, the presence of bidirectional interactions (20 bidirectional edges reported) suggests the existence of feedback loops or mutual regulation in the inactivation process. As inactivation of the sex chromosome is a specific biological process in women, these findings may highlight the strength of our approach in identifying sex-specific differential causal interactions.

Furthermore, the refined view of the top 20 genes connected to XIST and TSIX, as illustrated in Fig. 58, reinforces the notion that these two genes are pivotal nodes within broader regulatory networks.

In comparison to previous research, our study not only identifies associations but also clarifies the directionality and strength of these relationships, providing a more nuanced view of gene regulation. This methodological advancement is crucial for accurately interpreting genetic data, particularly in the context of complex diseases involving multiple genes and pathways. The CNs identified here serve as fundamental tools for dissecting the genetic basis of such diseases, potentially revolutionizing our approach to diagnosis and treatment.

Despite the innovative techniques and significant insights gained, our approach does have limitations. The construction and analysis of CNs in high-dimensional datasets are challenging, requiring careful consideration to avoid overfitting and ensure the findings' robustness. Nevertheless, these challenges are mitigated by the comprehensive nature and scalability of the developed methodologies, providing a strong foundation for further research and application in the field.

Looking ahead, the implications of this research are substantial. By extending the application of CNs to other biological systems, it could be possible to facilitate the understanding of a wide range of diseases, leading to breakthroughs in genetic research and pharmaceutical development. Moreover, integrating multi-omic data with the presented CNs could enhance the precision and accuracy of the proposed models, facilitating the development of predictive tools for disease progression and treatment responses.

Furthermore, this study lays the groundwork for transforming clinical practices by applying genetic insights to patient care. The causal models developed here could eventually guide clinical decision-making, enabling the practice of truly personalized medicine. By predicting individual responses to treatments based on genetic profiles, the therapeutic outcomes could be optimized and the adverse effects could be minimized, fundamentally changing the healthcare landscape.

This work was designed as an exploratory and systematic application of DCNs across the entire GTEx dataset, with the goal of highlighting both physiological and potentially pathological causal differences between sexes across tissues. A promising direction for future work is to apply this methodology to disease-specific datasets, allowing a more direct link between DCNs and clinical outcomes in precision medicine.

Conclusion

As widely discussed, this study underscores the powerful application of DCNs in analyzing human tissues to uncover sex-specific causal relationships among genes. By leveraging the comprehensive GTEx dataset and selecting 40 tissues, the work highlighted significant GE differences between males and females. These results are reinforced by the use of bootstrap resampling in gene differential analysis and multiple iterations in CN generation. The insights gained emphasize the critical role of sex as a determinant in understanding the underlying mechanisms of both healthy and pathological states, paving the way for the development of targeted therapies and personalized medicine. While sex emerged as the primary factor influencing causal relationships, age-related differences were also considered and accounted for, adding an additional layer of validation to our results. The biological insights provided by enrichment analysis following the DCN analysis offer a solid foundation for future research and clinical trials aimed at validating these findings. In essence, the integration of advanced analytical techniques and complex datasets represents a significant advancement in biomedical research, enhancing the comprehension and the treatment of diseases.

Key Points

- This is the **first comprehensive study** performing **Differential Causal Network (DCN)** analysis across the **entire GTEx dataset**, considering **40 human tissues**.
- Identification of **sex-specific causal relationships** in gene expression through a novel application of DCNs.
- **Integration** of causal discovery techniques with differential expression analysis, ensuring robustness through **bootstrap resampling**.
- Identification of **structural differences** (edges) in gene causal networks between males and females, with implications for **precision medicine**, by providing a causal understanding of sex-based biological differences.

Acknowledgements

Data used for the analyses described in this manuscript were obtained from the GTEx portal. PHG acknowledges the support of the PNRR project FAIR—Future AI Research (PE00000013), Spoke 9—Green-aware AI, under the NRRP MUR program funded by the NextGenerationEU.

Author contributions

ADF, PV, and PHG conceived main ideas of this paper. ADF implemented the software modules and performed the experiments. KG, FMG, and TK contributed to the discussion of the results and on revision of the first draft. All the authors approved the manuscript.

Supplementary data

Supplementary data is available at *Briefings in Bioinformatics* online.

Conflict of interests: None declared.

Funding

K.G. was partially funded by NIH NHGRI grant R01HG011393.

Data availability

The data used in this work are publicly available Open Access Data from the GTEx portal.

Ethics statement

This study is based exclusively on publicly available Genotype-Tissue Expression (GTEx) project data. Therefore, no additional ethical approval or informed consent was required.

References

1. Grath S, Parsch J. Sex-biased gene expression. *Annu Rev Genet* 2016;**50**:29–44. https://doi.org/10.1146/annurev-genet-120215-035429
2. Oliva M, Muñoz-Aguirre M, Kim-Hellmuth S. *et al.* The impact of sex on gene expression across human tissues. *Science* 2020;**369**:eaba3066. https://doi.org/10.1530/ey.18.14.14
3. Lichtblau Y, Zimmermann K, Haldemann B. *et al.* Comparative assessment of differential network analysis methods. *Brief Bioinform* 2017;**18**:837–50. https://doi.org/10.1093/bib/bbw061
4. Yazdani A, Yazdani A, Mendez-Giraldez R. *et al.* From classical mendelian randomization to causal networks for systematic integration of multi-omics. *Front Genet* 2022;**13**:990486. https://doi.org/10.3389/fgene.2022.990486
5. Bang S, Kim J-H, Shin H. Causality modeling for directed disease network. *Bioinformatics* 2016;**32**:i437–44. https://doi.org/10.1093/bioinformatics/btw439
6. Defilippo A, Giorgi FM, Veltri P. *et al.* Understanding complex systems through differential causal networks. *Sci Rep* 2024;**14**:27431.
7. Guo S, Zhou Y, Zeng P. *et al.* Identification and analysis of the human sex-biased genes. *Brief Bioinform* 2018;**19**:188–98. https://doi.org/10.1093/bib/bbw125
8. Cui C, Huang C, Liu K. *et al.* Large-scale in silico identification of drugs exerting sex-specific effects in the heart. *J Transl Med* 2018;**16**:1–7.
9. Cui C, Yang W, Shi J. *et al.* Identification and analysis of human sex-biased microrrnas. *Genom Proteom Bioinform* 2018;**16**:200–11. https://doi.org/10.1016/j.gpb.2018.03.004
10. Zhang C, Yang Y, Cui Q. *et al.* Identification and analysis of sex-biased copy number alterations. *Health Data Sci* 2024;**4**:0121.
11. Lonsdale J, Thomas J, Salvatore M. *et al.* The genotype-tissue expression (GTEx) project. *Nat Genet* 2013;**45**:580–5. https://doi.org/10.1038/ng.2653
12. Anders S, Huber W. Differential expression analysis for sequence count data. *Nat Preced* 2010;1–1. https://doi.org/10.1038/npre.2010.4282.2
13. Zhao S, Ye Z, Stanton R. Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols. *RNA* 2020;**26**:903–9. https://doi.org/10.1261/rna.074922.120
14. Sayers EW, Bolton EE, Rodney Brister J. *et al.* Database resources of the national center for biotechnology information. *Nucleic Acids Res* 2021;**50**:D20–6.
15. Muzellec B, Telenczuk M, Cabeli V. *et al.* PyDESeq2: a python package for bulk RNA-seq differential expression analysis. *Bioinformatics* 2023;**39**. https://doi.org/10.1093/bioinformatics/btad547
16. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;**15**:1–21. https://doi.org/10.1186/s13059-014-0550-8
17. Spirtes P, Glymour C, Scheines R. *Causation, Prediction, and Search*, Vol. 81. In: Dietterich Thomas, Bishop Christopher, Heckerman David, Jordan Michael, Kearns Michael (eds.), *Adaptive Computation and Machine Learning*,Cambridge, MA: MIT Press, 1993.
18. Plaksienko A, Thoresen M, Djordjilović V. Methods for differential network estimation: an empirical comparison arXiv preprint arXiv:2412.17922. 2024.
19. Zhang K, Zhu S, Kalander M. *et al.* gCastle: a python toolbox for causal discovery. 2021.
20. Colombo D, Maathuis MH. *et al.* Order-independent constraint-based causal structure learning. *J Mach Learn Res* 2014;**15**:3741–82.
21. Hagberg A, Swart P, Chult DS. Exploring Network Structure, Dynamics, and Function Using NetworkX. In: Varoquaux G, Vaught T, Millman J, (eds.). *SciPy 2008 Proceedings*. Los Alamos, NM (United States): Los Alamos National Lab (LANL); 2008.
22. Kanehisa M. The KEGG database. In: Bock GR, Goode JA, (eds.). *'In Silico' Simulation of Biological Processes: Novartis Foundation Symposium 247*. Vol. 247. Chichester, UK: John Wiley & Sons; 2002. pp. 91–103.
23. Gene Ontology Consortium. The gene ontology resource: 20 years and still going strong. *Nucleic Acids Res* 2019;**47**:D330–8. https://doi.org/10.1093/nar/gky1055
24. The Gene Ontology Consortium. The gene ontology resource: enriching a GOld mine. *Nucleic Acids Res* 2020;**49**:D325–34.
25. Kolberg L, Raudvere U, Kuzmin I. *et al.* G: Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res* 2023;**51**:W207–12. https://doi.org/10.1093/nar/gkad347
26. Liu Y-J, Wang C. A review of the regulatory mechanisms of extracellular vesicles-mediated intercellular communication. *Cell Communication and Signaling* 2023;**21**:77. https://doi.org/10.1186/s12964-023-01103-6
27. Pasricha C, Bansal N, Kaur R. *et al.* Immunoglobulins: Mechanistic Approaches in Moderation of Various Inflammatory and Anti-Inflammatory Pathways. *Curr Pharm Biotechnol*. 2024. Article in Press. https://doi.org/10.2174/0113892010310906240725072426.
28. Seda V, Mraz M. B-cell receptor signalling and its crosstalk with other pathways in normal and malignant cells. *Eur J Haematol* 2015;**94**:193–205. https://doi.org/10.1111/ejh.12427
29. Morel O, Jesel L, Freyssinet J-M. *et al.* Cellular mechanisms underlying the formation of circulating microparticles. *Arterioscler Thromb Vasc Biol* 2011;**31**:15–26. https://doi.org/10.1161/ATVBAHA.109.200956
30. Taylor HS, Arici A, Olive D. *et al.* HOXA10 is expressed in response to sex steroids at the time of implantation in the human endometrium. *J Clin Invest* 1998;**101**:1379–84.
31. Snell DM, Turner JMA. Sex chromosome effects on male–female differences in mammals. *Curr Biol* 2018;**28**:R1313–24. https://doi.org/10.1016/j.cub.2018.09.018
32. Makeyev A, Liebhaber S. The poly(c)-binding proteins: a multiplicity of functions and a search for mechanisms. *RNA* 2002;**8**:265–78. https://doi.org/10.1017/S1355838202024627
33. Lazim N, Elias MH, Zulazmi Sutaji AK. *et al.* Expression of HOXA10 gene in women with endometriosis: a systematic review. *Int J Mol Sci*. 2023;**24**:12869. https://doi.org/10.3390/ijms241612869
34. Hatanaka Y, De Velasco MD, Oki T. *et al.* HOXA10 expression profiling in prostate cancer. *Prostate* 2019;**79**:554–63. https://doi.org/10.1002/pros.23761
35. Guzzi PH, Cortese F, Mannino GC. *et al.* Differential network analysis between sex of the genes related to comorbidities of type 2 mellitus diabetes. *Appl Netw Sci* 2023;**8**:1–16.
36. Ahmed A, Köhler S, Klotz R. *et al.* Sex differences in the systemic and local immune response of pancreatic cancer patients. *Cancers* 2023;**15**:1815. https://doi.org/10.3390/cancers15061815
37. Padoan A, Plebani M, Basso D. Inflammation and pancreatic cancer: focus on metabolism, cytokines, and immunity. *Int J Mol Sci* 2019;**20**:676. https://doi.org/10.3390/ijms20030676
38. Pickup JC, Crook MA. Is type ii diabetes mellitus a disease of the innate immune system? *Diabetologia* 1998;**41**:1241–8. https://doi.org/10.1007/s001250051058
39. Sack G. Serum amyloid a—A review. *Mol Med* 2018;**24**:46. https://doi.org/10.1186/s10020-018-0047-0

40. Lin Y, Rajala MW, Berger J. *et al.* Hyperglycemia-induced production of acute phase reactants in adipose tissue*. *J Biol Chem* 2001;**276**:42077–83. https://doi.org/10.1074/jbc.M107101200
41. Kwiterovich P. The metabolic pathways of high-density lipoprotein, low-density lipoprotein, and triglycerides: A current review. *Am J Cardiol* 2000;**86**:5L–10L.
42. Reactome Pathway Knowledgebase. Binding and Uptake of Ligands by Scavenger Receptors (R-HSA-2173782). In: Jupe S, Gillespie M, Korninger F, (eds.). Cambridge, UK: European Bioinformatics Institute (EMBL-EBI); 2021.
43. Bürglin TR, Affolter M. Homeodomain proteins: an update. *Chromosoma* 2016;**125**:497–521. https://doi.org/10.1007/s00412-015-0543-8
44. McCall MN, Illei PB, Halushka MK. Complex sources of variation in tissue expression data: Analysis of the GTEx lung transcriptome. *Am J Hum Genet* 2016;**99**:624–35. https://doi.org/10.1016/j.ajhg.2016.07.007
45. Calvo R, West J, Franklin W. *et al.* Altered HOX and WNT7A expression in human lung cancer. *Proc Natl Acad Sci U S A* 2000;**97**:12776–81.
46. Ma T, Yan B, Yanbing H. *et al.* HOXA10 promotion of HDAC1 underpins the development of lung adenocarcinoma through the DNMT1-KLF4 axis. *J Exp Clin Cancer Res.* 2021;**40**. https://doi.org/10.1186/s13046-021-01867-0
47. Karim N, Ho SY, Nicol E. *et al.* The left atrial appendage in humans: structure, physiology, and pathogenesis. *EP Europace*. 2020;**22**:5–18, https://doi.org/10.1093/europace/euz212

Briefings in Bioinformatics, 2025, **26(4)**, bbaf407
https://doi.org/10.1093/bib/bbaf407
Case Study