bioRxiv preprint doi: https://doi.org/10.1101/2022.05.15.492007; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

MultiCens: Multilayer network centrality measures to uncover molecular mediators of tissue-tissue communication

Tarun Kumar[1,3,4], Ramanathan Sethuraman[2], Sanga Mitra[1], Balaraman Ravindran[1,3,4], and Manikandan Narayanan[1,3,4,5,a]

[1] Department of Computer Science and Engineering, Indian Institute of Technology (IIT) Madras, Chennai, India
[2] Intel Corporation, Bangalore India
[3] Robert Bosch Center for Data Science and Artificial Intelligence (RBCDSAI), IIT Madras, Chennai, India
[4] The Centre for Integrative Biology and Systems medicinE (IBSE), IIT Madras, Chennai, India
[5] Multiscale Digital Neuroanatomy (MDN), IIT Madras, Chennai, India

Abstract. With the evolution of multicellularity, communication among cells in different organs/tissues became pivotal to life. Molecular basis of such communication has long been studied, but genome-wide screens for biomolecules/genes mediating tissue-tissue signaling are lacking. To systematically identify inter-tissue mediators, we present a novel computational approach MultiCens (Multilayer/Multi-tissue network Centrality measures). Unlike single-layer network methods, MultiCens can distinguish within- vs. across-layer connectivity to quantify the "influence" of any gene in a tissue on a query set of genes of interest in another tissue. MultiCens enjoys theoretical guarantees on convergence and decomposability, and excels on synthetic benchmarks. On human multi-tissue datasets, MultiCens predicts known and novel genes linked to hormones. MultiCens further reveals shifts in gene network architecture among four brain regions in Alzheimer's disease. MultiCens-prioritized hypotheses from these two diverse applications, and potential future ones like "Multi-tissue-expanded Gene Ontology" analysis, can enable whole-body yet molecular-level investigations in humans.

Main

For any multicellular organism with specialized tissue or organ structures, communication among the different tissues/organs is essential for coherent integrated functioning of the whole body. The molecular mechanisms of such inter-organ communication, be it canonical communication routes such as the nervous system and hormonal system (or) non-canonical recently-discovered routes such as ones mediated by fat-derived extracellular vesicles [1] and microbiota-derived metabolites in the gut-brain axis, can be represented as a network of interactions among the biomolecules residing in different tissues/organs (and called the inter-organ communication network or ICN) [2]. Rapidly gaining interest in the mapping of ICN [3] and detailed mechanistic characterization of specific interactions in the ICN [4] have revealed a large ICN network among secreted proteins in model organisms like Drosophila, and the key roles played by certain ICN molecules or interactions in healthy and disease conditions. But these experimental techniques for ICN mapping or ICN analysis are predominantly *in vivo* and hence of limited use in non-model organisms including humans, and also quite time-consuming even in model organisms (due to the potentially huge experimental space to cover the quadratic number of all pairwise interactions among thousands of biomolecules in tens of tissues of interest). As a result, the ICN is vastly under-explored in both model as well as non-model organisms, and there is an immediate need to accelerate mapping and analysis of ICNs in health and disease.

In this study, we propose a computational approach to rapidly map and analyze a multi-tissue network, comprising both inter- and intra-tissue gene-gene interactions. Our work is enabled by the recently accumulating multi-tissue genomic datasets (e.g., [5],[6]), which can be used to infer inter/intra-tissue networks using the concept of gene-gene correlation or coexpression. Coexpression network mapping and analysis have been done before, for instance using the popular WGCNA method [7], and gene prioritization using network based measures have also successfully guided downstream experiments before [8,9,10]; but these existing studies have mostly focused on a single tissue of interest in a healthy condition or the single most affected tissue in a disease. Our proposed centrality framework, termed MultiCens, works in a multi-tissue setting and offers a systematic data/computation-driven prioritization of genes to be key regulators of inter-tissue signaling.

[a] *email:* nmanik@cse.iitm.ac.in

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.15.492007; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Specifically, a main contribution of our work is the design and application of a gene centrality measure that quantifies the extent to which each gene in a tissue influences a query set of genes of interest in another tissue via direct and multi-hop inter-/intra-tissue interactions. We extend traditional centrality measures like PageRank [11] that work for a single-layer system to design new measures for a multilayer network model, wherein each layer corresponds to a tissue and nodes (genes) can have within-layer and across-layer connections (gene-gene interactions). We demonstrate the effectiveness of MultiCens in capturing multi-hop effects using both synthetic multilayer networks as well as real-world multi-tissue datasets. On a real-world human multi-tissue gene expression dataset for instance, we can uncover genes responsible for inter-tissue communication via mediating hormones, specifically genes involved either in the production/processing/release of hormones in a source tissue or those that respond to hormones in the target tissues. Even with well-studied hormones such as insulin, our study identifies not only known but also novel regulators of insulin signaling, including lncRNAs (long non-coding RNAs) as well. MultiCens can also be applied to multi-brain-region gene expression datasets obtained from postmortem brain samples of Alzheimer's disease vs. control individuals to highlight the large-scale changes in the centrality of specific genes and pathways in Alzheimer's disease. The diverse applications of MultiCens to find the molecular mediators of inter-tissue hormonal signaling in healthy tissue or inter-brain-region dysregulation in disease is promising for its broader applicability and robustness to dissect communication amongst other functional structures within the body of humans and other species.

Results

Overview of our proposed centrality measures: MultiCens

We introduce a set of centrality measures, termed MultiCens, to quantify the influence or effect a gene has at different levels of granularity, such as the effect a gene has (i) "locally" within a tissue due to its connections to other genes in the tissue, or (ii) "globally" across all tissues due to within- as well as across-tissue connections, or specifically (iii) to a particular tissue, or (iv) to a query set of genes in a particular tissue. MultiCens measures account for the multilayer, multi-hop network connectivity of the underlying system in a hierarchical fashion, by decomposing the overall centrality (*versatility* pioneered by Domenico et al. [12]) of a gene into *local* vs. *global* centrality, and further into *layer-specific* centrality specific to a tissue (referred to interchangeably as layer) or *query-set* centrality specific to a gene set in a tissue (see hierarchical organization in Fig. 1 and Methods). We prove theoretical guarantees on the convergence and decomposability of MultiCens measures (see Theorems in Methods section), and demonstrate empirical applications of MultiCens to simulated networks as well as real-world healthy and disease multi-tissue datasets below. Our overall pipeline starts with a multilayer network model (constructed for instance from transcriptomic data of a multi-tissue system), represents it as a supra-adjacency matrix comprising two matrices (one for capturing within-layer connections alone, and another for across-layer connections), and then uses these two matrices to define different centrality measures (see Fig. 1 and Methods). Ranking nodes/genes by their centrality scores can readily help predict key genes involved in inter-layer communication, amongst other applications.

Capturing multi-hop effects in synthetic multilayer networks

We first evaluate MultiCens on synthetic networks that simulate a real-world application scenario of identifying genes involved in tissue-tissue hormonal signaling. In this scenario, we test if MultiCens assigns top ranks to hormone-producing genes in a hormone's source layer, when hormone-responsive genes in its target layer are provided as the query set. Since "ground-truth" hormone-producing genes could be linked to the "query" hormone-responsive genes via a mixture of direct connections (edges) or indirect one/more-hop connections (paths), we model our synthetic networks accordingly as a two-layered network with a fixed query set in layer 2, and two communities *source-set 1* and *source-set 2* in layer 1 that are strongly connected to layer 2 by direct and multi-hop connections respectively (Fig. 2a-b). We start with a ground truth set of nodes that has all *source-set 1* nodes alone, and then replace a fraction of these nodes with nodes from *source-set 2* (Fig. 2c).

A *recall-at-100* analysis shows that two existing methods, as well as MultiCens, can recover the ground truth nodes when they are directly connected to the *query-set* (Fig. 2c, 0% curves). However, as we increase the fraction of nodes from *source-set 2* in the ground truth, our MultiCens *query-set* centrality performs increasingly better than other methods (Fig. 2c). These benchmarks show MultiCens *query-set* can rank nodes with direct as well as indirect (multi-hop) connections to a cross-layer *query-set* towards the top, due to its ability to distinguish intra- vs. inter-layer edges (unlike the existing versatility method [12], which cannot make this distinction) and handle multi-hop connectivity (unlike the existing inter-layer degree based method S_{sec}, proposed in a pioneering work on data-driven discovery of endocrine hormone interactions [13], which handles only direct interactions). For completeness, we compared these methods to local and global centrality-based rankings as well (see Suppl Fig. S1), and observed that MultiCens query-set centrality performs better – this encourages the use of a query set of genes, whenever this information is available, in our MultiCens applications.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.15.492007; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

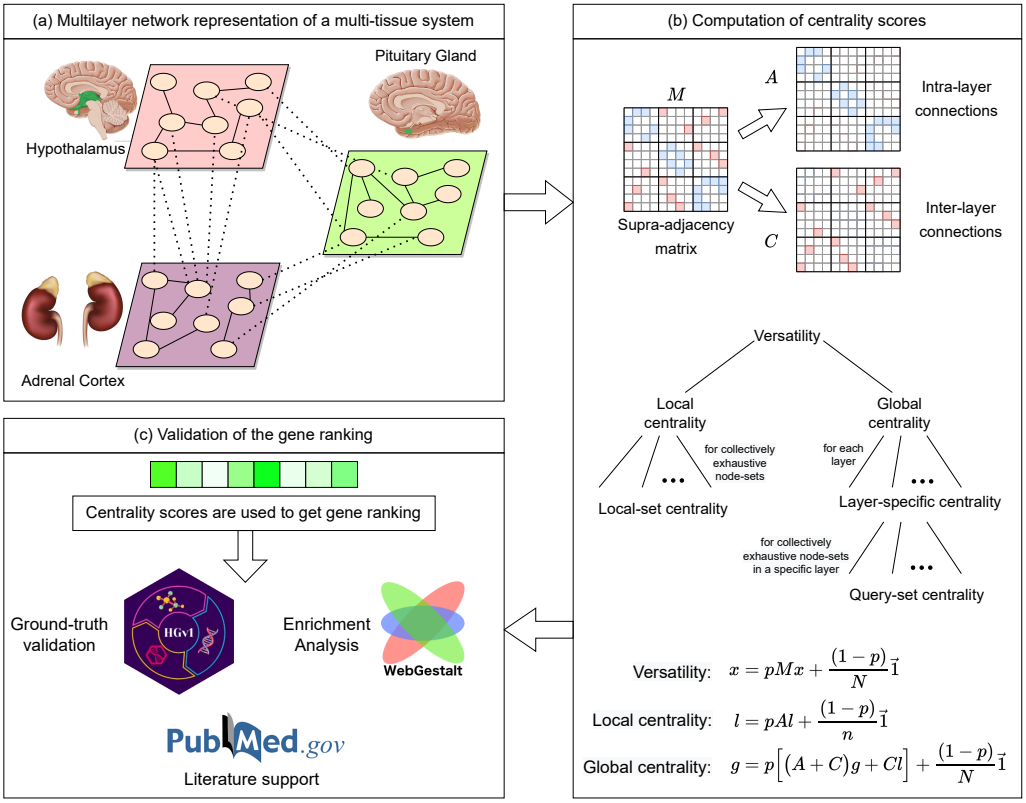


Fig. 1: **Workflow of our MultiCens measures.** (a) Each layer in the network represents a tissue, and connections represent gene-gene interactions (e.g., inferred from transcriptomic data). (b) Supra-adjacency matrix (M) contains within-tissue connections on the diagonal blocks (intra-layer matrix A), and across-tissue connections on the off-diagonal blocks (inter-layer matrix C). The A, C matrices are used to compute different hierarchically-organized centralities as shown. The centrality vectors (x, l, and g) have an entry for each gene in every tissue. (c) The centrality scores are used to obtain gene rankings which are further validated using different methods, and interpreted to predict novel mediators of inter-tissue signaling.

MultiCens ranks inter-tissue signaling genes at the top

After verifying MultiCens on synthetic multilayer networks, we now apply it to human multilayer networks, comprising gene-gene coexpression relations inferred from a multi-tissue dataset GTEx (Genotype-Tissue Expression [14]) and tissue-specific protein-protein interactions from a repository SNAP/BioSNAP (Stanford Biomedical Network Dataset Collection [15]) (see Methods). To validate the MultiCens-based gene rankings obtained from any human multilayer network of interest, we use a Gene Ontology (GO) based database of hormone-related genes HGv1 (Hormone-Gene version 1 [16]) as the ground truth. Our task is to predict hormone-producing genes when only a query-set of hormone-responding genes is given as input, and vice versa. To capture the communication paths between a hormone's producing and responding set of genes in the multilayer network, both sets should be sufficiently large. Hence, we restrict our evaluation to hormones with at least 10 hormone-producing and at least 10 responding genes. Four hormones pass this threshold, and are referred to as the *primary* hormones. For all but one of these primary hormones, viz., for Insulin, Somatotropin, and Progesterone, our MultiCens *query-set centrality* ranks the ground truth hormone-related genes towards the top (see *recall-at-k* plots in Fig. 3a). The complete gene ranking for these hormones is provided in Suppl Dataset SD1.

We then expanded our application to all hormones with at least 10 genes in the hormone-producing set *or* the responding set or both sets, and report such hormone's Area Under *recall-at-k* Curve or AUC in Fig. 3b (see also Suppl Fig. S2 and Suppl Table S1 for results on all tested hormones). For a majority of these hormones (all but 5 of the corresponding 16 prediction tasks in Fig. 3b), our MultiCens gene rankings yield AUCs better than that of random rankings. When we remove SNAP-based protein interactions and keep only coexpression edges in the human multilayer networks (Fig. 3b; lighter dots), performance drops slightly, but otherwise the trend of AUCs remain similar. Taken together, these results affirm the robustness of MultiCens in ranking genes associated to hormonal inter-tissue signaling at the top.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.15.492007; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

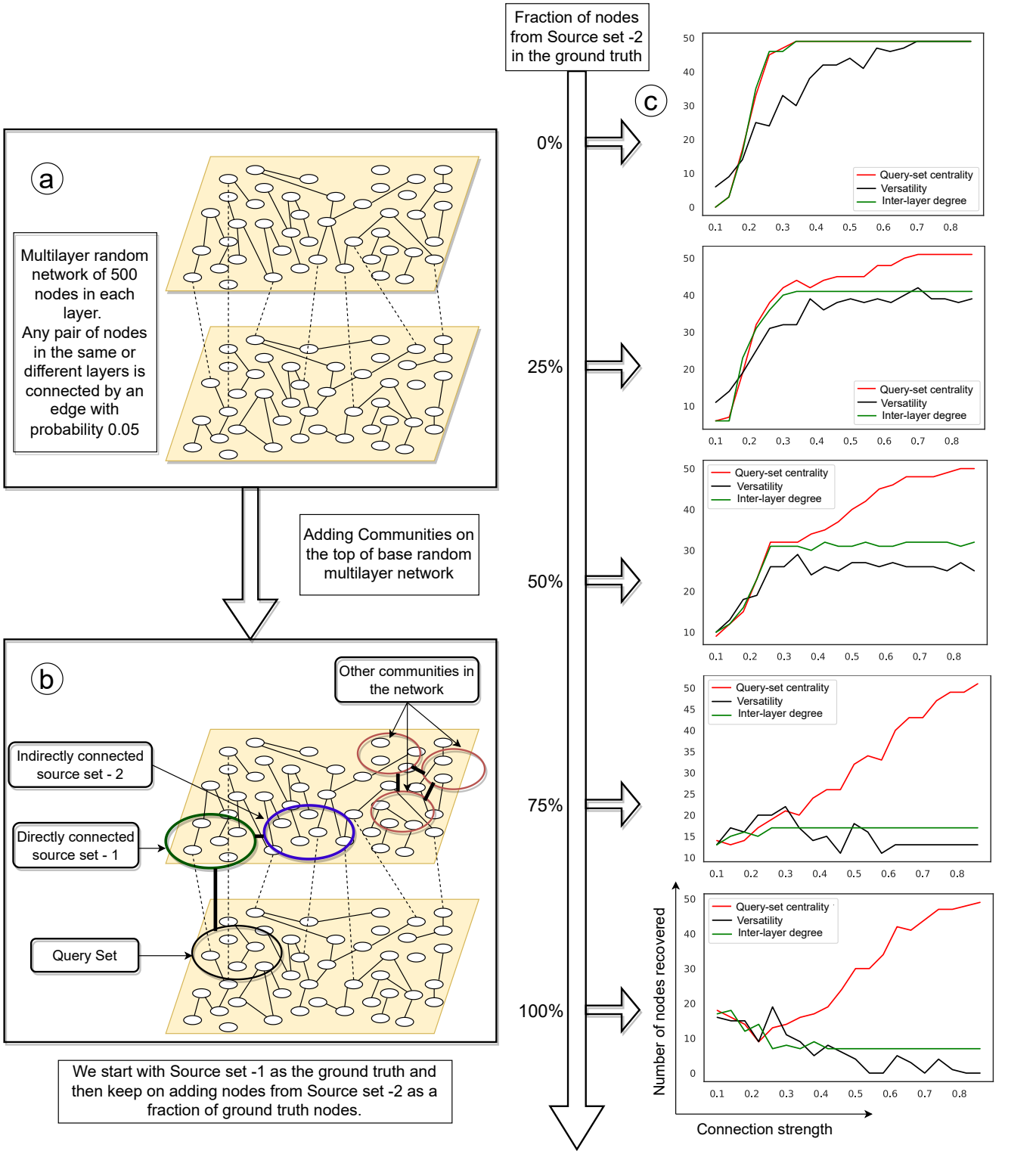


Fig. 2: **Synthetic multilayer network construction and evaluation.** (a) Synthetic network construction starts with a base random multilayer network with edge probability 0.05; (b) More edges are then added, according to the connection strength desired, both within the selected communities (indicated by circles) and between certain pairs of communities (indicated by thick dark edges connecting the pair; e.g. between *source-set 1* and *source-set 2*). (c) As more nodes from *source-set-2* become part of the ground truth (shown as increasing percentages), our MultiCens query-set centrality outperforms the existing methods to a larger extent. Each plot shows the connection strength (x-axis) against the number of ground truth nodes in the top 100 ranked nodes (y-axis). See also Suppl Fig. S1.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.15.492007; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

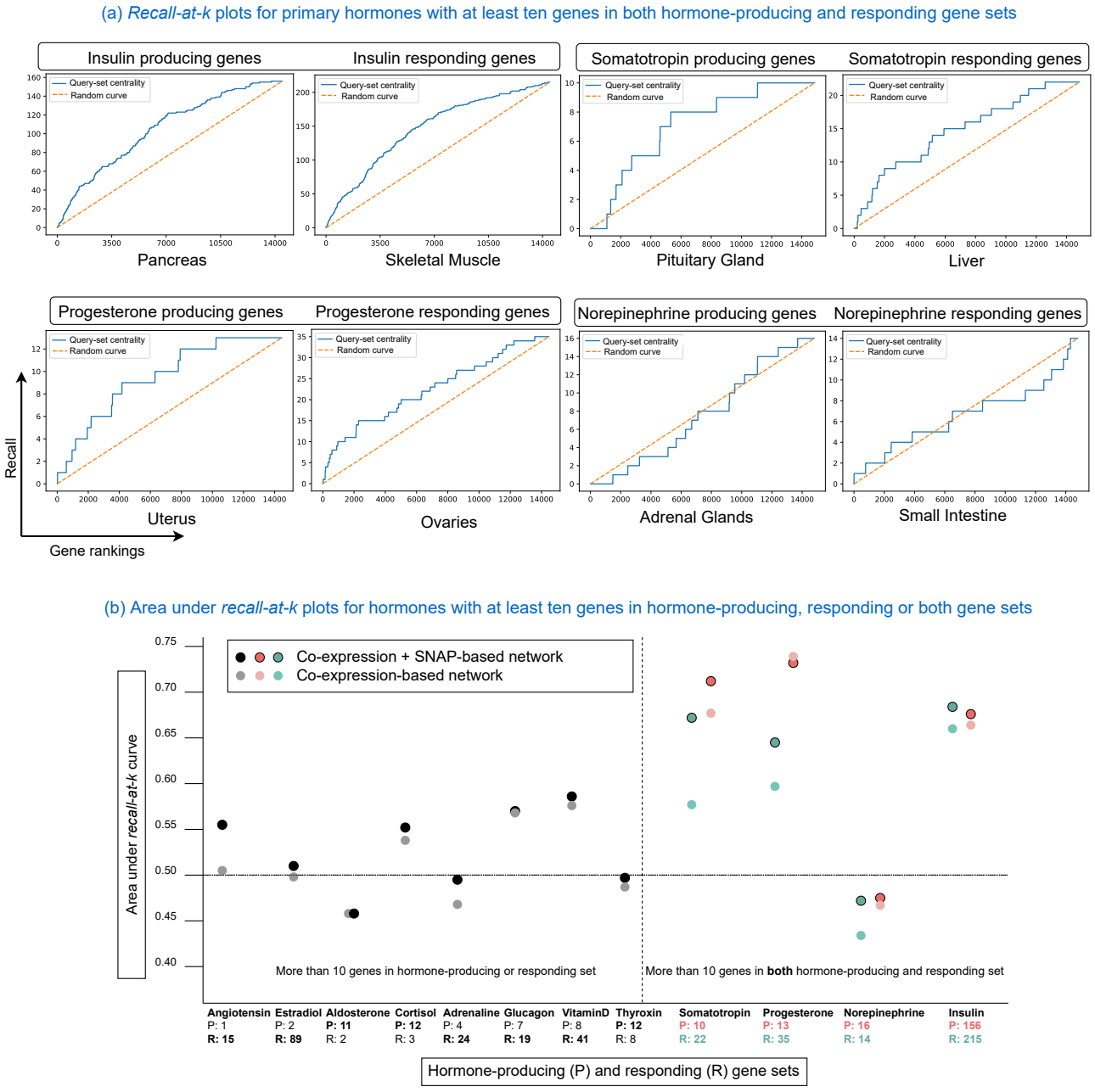


Fig. 3: **MultiCens on human multilayer networks: ground-truth validation.** (a) Recall (# of ground truth genes recovered; y-axis) in the top k ranked genes (x-axis) are plotted using MultiCens query-set centrality based ranking *vis-à-vis* a random ranking (random curve). Only primary hormones shown here; see Suppl Fig. S2 for plots for the other tested hormones. (b) For hormones with 10 or more genes in either producing or responding set, the smaller set is used as the query set, and the plot reports AUC score for predicting the bigger set (marked in bold-face font in x-axis). For the four primary hormones having at least 10 genes on both producing and responding sets, plot reports AUC for predicting both sets. See also Suppl Table S1.

MultiCens gene rankings are enriched for hormone-related diseases

The promising validation of MultiCens-based gene rankings using the ground truth HGv1 database encouraged us to test if our top-ranking genes are enriched for the corresponding hormone-related disorders/diseases (as in our earlier literature mining study [16]). Among all enriched disease terms at FDR 5% (Fig. 4a), many of them are well-supported in the literature such as enrichment of Type-2 Diabetes for Insulin [17], breast cancer for progesterone [18], and colorectal cancer for somatotropin [19]. Moreover, insulin resistance leads to chronic hyperinsulinemia, which is

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.15.492007; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

further associated with various types of cancer including breast, colorectal, prostate cancer among others [20, 21], as reflected in our enrichment results. Insulin resistance in skeletal muscle leads to a condition less studied called diabetic myopathy, where the strength and mass of skeletal muscle is reduced [22]. In case of somatotropin, a growth hormone secreted by the pituitary gland, our enrichment result confirms its association with increased colon polyps and cancer [23]. Finally, mood-related disorders typically associated with Norepinephrine were not enriched in our analysis, in line with the poor validation of this hormone against HGv1 ground truth; however, this hormone is an etiological factor for different cancer types [24], including the ones found in our enrichment analysis (Suppl Fig. S3). Summarizing, for three of our four primary hormones with sufficient gene associations, our MultiCens ranking reveals meaningful disease enrichments.

PubMed literature analysis of MultiCens predictions reveals known and novel hormone-gene links

Ground-truth databases including our HGv1 could be incomplete and miss certain genuine hormone-gene relations. So we turn to the PubMed literature corpus to search for known vs. novel hormone-related genes amongst the top-ranked genes returned by our MultiCens on the hormone-specific human multilayer networks. We employ two PubMed-derived scores to quantify the evidence for a potential link between a hormone and a gene: (i) co-occurrence or co-mention of a hormone-gene pair in published articles in PubMed (see Methods), and (ii) contextual similarity between a hormone and a gene in the corpus, which can also identify hormone-gene pairs not co-mentioned in any publication. Text-based deep learning methods can successfully capture the contextual similarity between two words via cosine similarity of their corresponding word embedding vectors [25], and this is what we adopt too (see Methods).

In this literature-based analysis, we focus on peptide hormones insulin and somatotropin, so that we can apply a filter to test predictions that are only among genes involved in peptide secretion (see Methods). Fig. 4b shows the top 10 secretory genes in the MultiCens ranking for each hormone (when MultiCens centrality is obtained by taking the hormone-responsive genes as the query set) along with their co-occurrence and contextual similarity scores with the hormone-related terms. While a few genes (yellow background) from our predictions are already present in our ground truth HGv1, there are other genes (green background) not present in HGv1 but whose associations are confirmed by the high PubMed-based similarity scores with at least one of the hormone-related terms. For insulin for example, we obtain two such out-of-ground-truth genes: *LRRC8*, which has been found to enhance insulin secretion in pancreatic β-cells in a recent study [26], with later studies confirming its role in insulin resistance and glucose resistance [27]; similarly, *EGFR* gene is known to mediate diabetes-induced microvascular dysfunction [28].

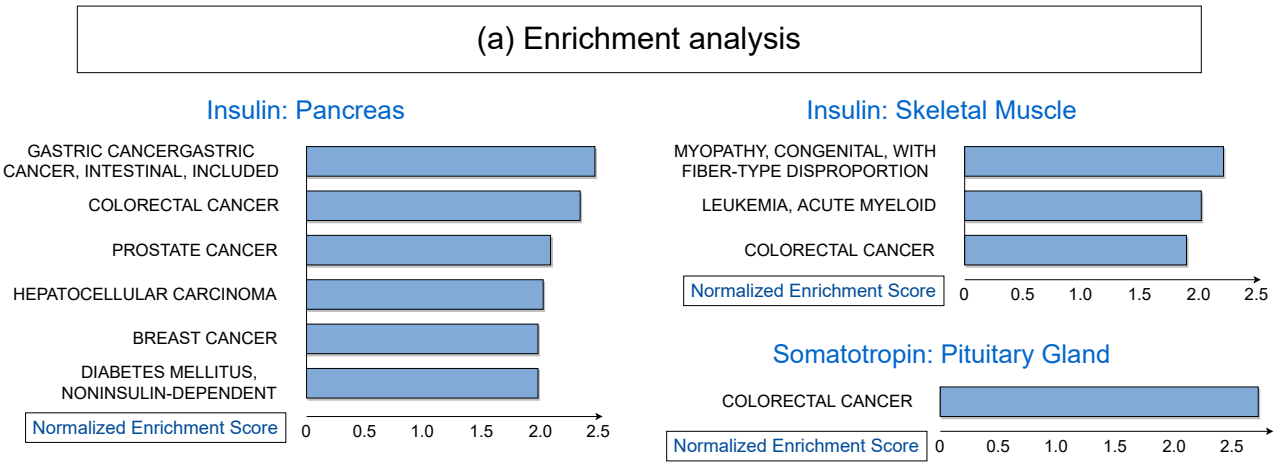
For both hormones, we find certain novel gene predictions that are both absent in our ground truth and have poor PubMed literature support scores (white-background genes in Fig. 4b). One such novel prediction is *CD74* for insulin - this gene's role in insulin secretion and related diseases was not well-established until the recent discovery of its participation in insulin resistance [29]. Another example of a novel prediction is *RFX3* for somatotropin – this gene has no direct co-occurrence with hormone-related terms, but is known to play a role in hydrocephalus disease [30], which is associated with deficiency in this growth hormone [31]. Based on the top centrality ranks and the above-discussed recent or indirect pieces of literature evidence, the role of genes like *CD74* and *RFX3* respectively in insulin and somatotropin signaling warrant further exploration and can be prioritized in future experiments. For further details, please see Suppl Results.

MultiCens identifies lncRNAs as integral part of hormone signaling networks

The role of protein-coding genes in hormonal signaling is well established, but that of long non-coding RNAs (lncRNAs) in the endocrine system is only evolving. Uncovering lncRNA's association to the hormones may provide a ground for innovative treatment strategies for related diseases, and MultiCens provides a systematic data-driven discovery of these associations. Table 1 shows the top 5 lncRNA genes among the top 1000 MultiCens-predicted genes in terms of tissue-specific gene rankings for each primary hormone. Suppl Table S3 provides supporting references for each predicted lncRNA (hence we do not cite all references explicitly in the following text).

For the insulin hormone, MultiCens detected *PRKCQ-AS1*, a natural antisense lncRNA for the diabetes drug-target and insulin signaling regulator *PRKCQ* (Protein kinase C theta). Gene *PRKCQ* has higher activity in muscle from obese diabetic patients and *PRKCQ-AS1* is required to maintain a relatively constant level of *PRKCQ*. Recent evidence indicates that lncRNAs, through β-cell mass modulation, affect insulin synthesis, secretion and signaling, thereby enhancing the progression of type-2 diabetes mellitus (T2DM) [32]. MultiCens-predicted lncRNA *MIR22HG* is reported for instance as a hub node in a competitive endogenous RNA (ceRNA) network related to T2DM, along with other cancer signaling pathways. Further, PWAR6 (Prader Willi/Angelman region RNA 6) is reported to play a major role in the Prader–Willi syndrome (PWS) phenotype, and PWS patients are often diagnosed with T2DM. It will be interesting to find if there is any direct link between PWAR6 and T2DM.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.15.492007; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



(b) Literature support

Insulin: Pancreas

Top Genes	Term: Insulin		Term: Diabetes	
	Co-occurrence score	Cosine similarity of embeddings	Co-occurrence score	Cosine similarity of embeddings
SYBU	3.65	0.22	0.00	0.26
LRP1	4.79	0.33	1.87	0.25
CNR1	3.48	0.27	2.03	0.29
PTPRN2	21.76	0.21	15.04	0.23
SERP1	0.61	0.34	1.59	0.26
CD74	0.80	0.25	0.79	0.22
LRRC8A	5.11	0.19	2.44	0.08
PICK1	1.73	0.32	0.99	0.23
EGFR	1.55	0.36	1.92	0.29
INS	17.37	0.59	6.34	0.42

Somatotropin: Pituitary Gland

Top Genes	Term: Somatotropin		Term: Acromegaly	
	Co-occurrence score	Cosine similarity of embeddings	Co-occurrence score	Cosine similarity of embeddings
HDAC1	1.85	0.20	0.00	0.10
EGFR	5.05	0.24	0.18	0.23
FKBP1B	6.05	0.25	0.00	0.16
FFAR4	1.00	0.31	0.00	0.20
S100A8	0.73	0.24	0.00	0.13
PER2	1.64	0.31	0.00	0.17
RFX3	0.00	0.22	0.00	0.17
PRKCE	1.48	0.28	0.00	0.11
SERP1	3.88	0.32	0.00	0.12
ITSN1	0.00	0.20	0.00	0.08

Fig. 4: **MultiCens on human multilayer networks: prior support and novel predictions.** (a) Shown are all disease gene sets based on OMIM (Online Mendelian Inheritance in Man) that are enriched for top MultiCens centrality scores at FDR 5%, as reported by WebGestalt (see Methods; when predicting somatotropin-responding genes in liver, no disease enrichments pass this FDR cutoff; see also Suppl Fig. S3 for the other two primary hormones' disease enrichments). (b) Literature support for our top 10 predicted genes (ranked only among genes involved in peptide secretion) for the two peptide hormones, along with their co-occurrence scores and similarity in embedding space with hormone-related terms. Genes with a yellow background are present in the ground truth (HGv1 database); from the remaining genes, the green background represents genes supported by scores (co-occurrence score ≥ 1) for either or both hormone-related terms, and white background represents the other genes not supported by scores for both hormone-related terms. See also Suppl Table S2 for gene names corresponding to the gene symbols shown.

Somatotropin, a growth hormone secreted in the anterior pituitary gland, stimulates body growth, and also stimulates liver and other tissues to produce Insulin-like growth factor I (IGF-I), which in turn results in cartilage cell proliferation and bone growth [33, 34]. Reassuringly, lncRNAs predicted for association to somatotropin in liver are involved in many liver diseases and cancer. *NEAT1* (nuclear paraspeckle assembly transcript 1) is significantly increased in non-alcoholic fatty liver disease (NAFLD) and its' high expression is correlated with worse survival in cancer patients. Expression of *MIR210HG* increases in hepatocellular carcinoma (HCC) cells relative to paired adjacent normal liver tissue samples and relative to normal liver cell line. Similarly, *LINC01278* mediates HCC metastasis by regulating miR-1258 expression.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.15.492007; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 1: Top five ranked lncRNAs by MultiCens in source and target tissues of the four considered hormones.

	Insulin	
	Pancreas	**Skeletal Muscle**
1	*LINC00672*	*ZEB1-AS1*
2	*HOXA-AS2*	*TNK2-AS1*
3	*PRR34-AS1*	*PWAR6*
4	*MIR22HG*	*PRRT3-AS1*
5	*LINC00294*	*PRKCQ-AS1*

	Somatotropin	
	Pituitary Gland	**Liver**
1	*LINC01588*	*NEAT1*
2	*PTPRD-AS1*	*ZNF528-AS1*
3	*LINC01132*	*MIR210HG*
4	*UCA1*	*ALMS1-IT1*
5	*LINC01473*	*LINC01278*

	Progesterone	
	Ovaries	**Uterus**
1	*CCDC18-AS1*	*HAGLR*
2	*LINC00641*	*TAF1A-AS1*
3	*MIR210HG*	*LINC00602*
4	*LINC01016*	*PCAT19*
5	*BEAN1-AS1*	*HHIP-AS1*

	Norepinephrine	
	Adrenal Glands	**Small Intestine**
1	*PGM5P4-AS1*	*RNF139-AS1*
2	*CCDC18-AS1*	*CARMN*
3	*MAGI2-AS3*	*SPATA41*
4	*LINC01291*	*GHET1*
5	*TOLLIP-AS1*	*ATP1B3-AS1*

Although lncRNAs are correlated with multiple cancers in general, their molecular mechanisms in the context of hormone signaling remain inadequately understood. Our predictions linking a hormone and its predicted lncRNA to the same cancer type can thus accelerate and prioritize experimental investigations of these mechanisms. For instance, breast, ovary and uterine endometrium are known targets of progesterone, and the lncRNAs with high progesterone-related query set centrality are seen to be involved in cancer of these three regions (see Suppl Results). Suppl Results also discusses how somatotropin's involvement in proliferation is reinforced by MultiCens-detected lncRNAs, most of which are linked to cancer cell growth.

Finally, MultiCens yields interesting lncRNA predictions for norepinephrine, a neurotransmitter which promotes vasoconstriction and controls heart rate and also effects intestinal absorption and secretion by regulating the tone of smooth muscle. *CARMN*, a smooth muscle cell-specific lncRNA, detected by MultiCens, is reported to regulate cardiac cell differentiation and homeostasis. Further, lncRNA *GHET1* has effects in development of pre-eclampsia, a difficult pregnancy indicated by high blood pressure. Based on the role of these lncRNAs, they seem to be influenced by norepinephrine, but exact mechanism of regulation requires further study. MultiCens therefore predicted lncRNAs, a few of which are already present in our ground-truth database, as well as other novel ones with interesting links to hormonal signaling and disorders.

MultiCens detects changes in brain networks between Alzheimer disease and control populations

After recognizing the potential of MultiCens in identifying genes (both protein coding and lncRNAs) in hormone signaling pathways in health, we employ it to understand the change in the gene-gene network structures in disease, specifically Alzheimer's disease (AD) relative to a control (CTL) population. We retrieved data of 264 AD and 372 control human postmortem RNAseq samples from Mount Sinai Brain Bank dataset [35] for four brain regions: frontal pole (FP), superior temporal gyrus (STG), parahippocampal gyrus (PHG), and inferior frontal gyrus (IFG). We construct one multilayer network for the AD group of individuals and another for the CTL group, with four layers in the network representing the four brain regions, and network nodes and edges representing respectively the genes in these brain regions and gene-gene coexpression relations (after adjusting for covariates; see Methods). We use the genes involved in synaptic signaling (SSG) in the PHG region as the query set of genes (134 genes), and identify the disease-driven change in the centrality-based ranking of genes in the remaining three regions. We observed considerable shift in the ordering of these three brain regions in the AD vs. CTL multilayer networks according to their median gene centrality scores (see Fig. 5a, STG region's ordering for instance). In terms of individual genes, *ANKFN1*, *OR10AD1* and *PLCD3* gain the highest positive shift in AD-based ranking in the FP, STG and IFG regions respectively. *ANKFN1* is found to be upregulated in hippocampus tissues of AD patients [36]. Though *OR10AD1* (olfactory receptor family 10 subfamily AD member 1) is not yet connected to AD, olfactory impairments is recently reported to be one of the early phase' pathophysiological changes in AD [37]. *PLCD3* is known to be upregulated in the AD population along with other regulators of lipid metabolism [38]. We provide the complete gene rankings of all three regions for AD vs. CTL networks in Supplementary Dataset SD2.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.15.492007; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

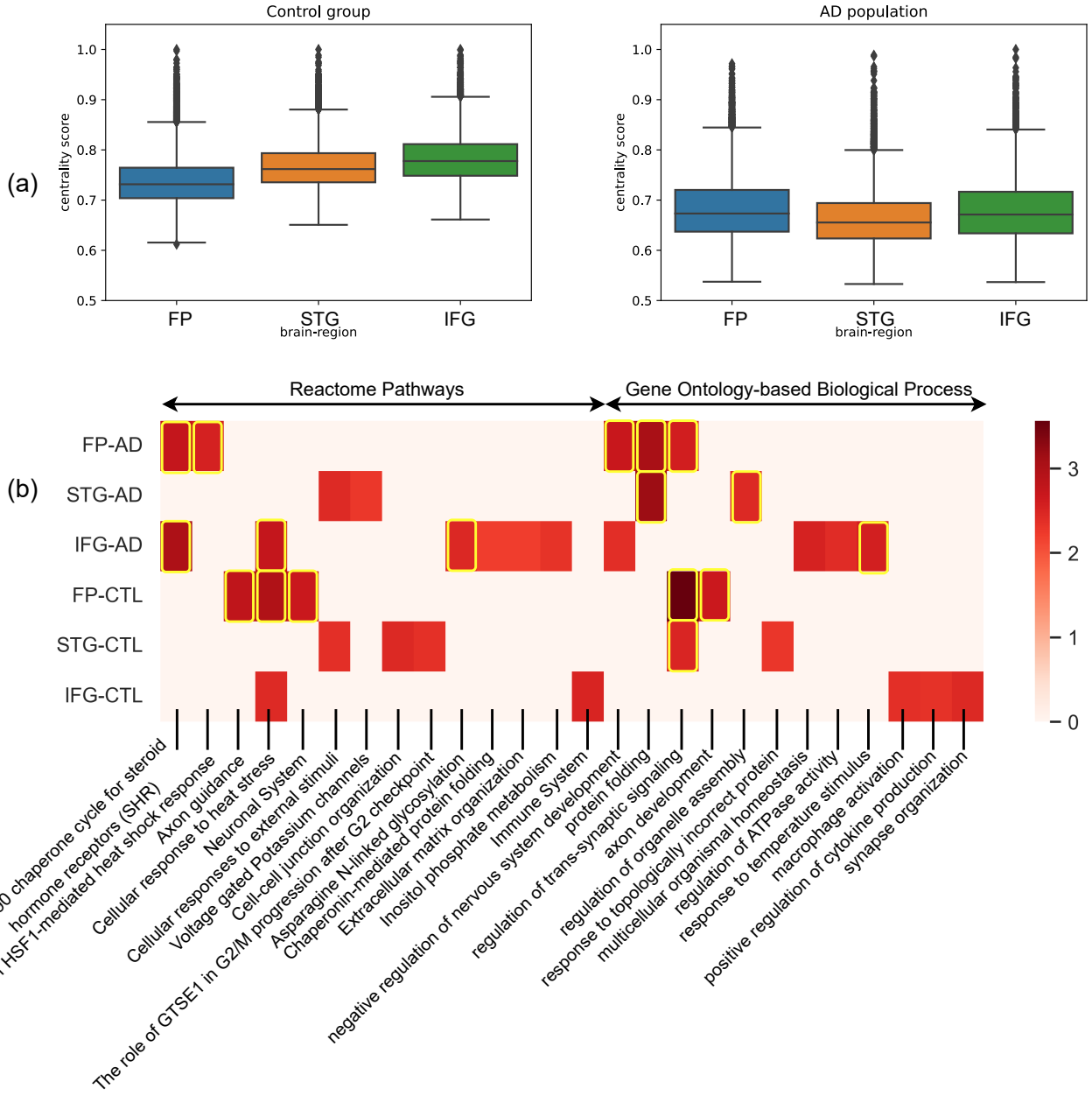


Fig. 5: **MultiCens on multi-brain-region networks in disease:** Study of changes in MultiCens gene rankings of four-layer networks of control and Alzheimer affected population. We rank genes of frontal pole (FP), superior temporal gyrus (STG) and inferior frontal gyrus (IFG) using MultiCens centralities calculated using a query-set of synaptic signaling genes in parahippocampal gyrus (PHG). (a) Bar-plot showing region-wise shift of centrality scores of the three regions. (b) Reactome pathways and Gene Ontology-based process (GO-BP) enrichment analysis of each region in control and AD state. Color map represents the normalized enrichment score from WebGestalt. The highlighted boxes pass the 0.01 FDR cut-off. If centrality-based gene rankings of a region do not pass the 0.05 FDR cut off for an enrichment, we set the corresponding normalized enrichment score to 0.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.15.492007; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

MultiCens also offers an across-region view of gene importance in the AD or CTL multilayer networks. In the AD network, irrespective of brain regions, genes *JMJD6*, *SLC5A3*, *CIRBP*, *TARBP1* and *AHSA1* are among the top ten central genes correlated with the SSG set, of which *AHSA1* (activator of HSP90 ATPase activity 1) is already known to correlated with AD progression by promoting tau fibril formation [39]. On the other hand, *CIRBP* (cold inducible RNA binding protein) shields neurons from amyloid toxicity mediated by antioxidative and antiapoptotic pathways, making it a favourable molecule contending for AD prevention or therapy [40]. It may be worth studying the other three genes experimentally to test their connections to AD pathology. Similar to these individual genes, certain biological pathways were also enriched for top ranks, irrespective of the brain region, in the AD network (see Fig. 5b) – examples include HSP90 chaperone cycle for steroid hormone receptors (R-HSA-3371497) pathway and negative regulation of nervous system development (GO:0051961). Heat shock protein 90 (Hsp90), "a molecular chaperone", is known to induce microglial activation leading to amyloid-beta (Aβ) clearance [41]. The across-region consistency of top-ranking genes/pathways in the AD network is not observed in the CTL multilayer network. For example, gene *CDK5R2* (Cyclin Dependent Kinase 5 Regulatory Subunit 2) is ranked 3rd in FP, rank 224 in STG, and 2076 in IFG. Pathway enrichments are also more region-specific in the CTL network (relative to AD network; see Fig. 5b), such as Axon guidance in FP, Cell-cell junction organization in STG, and immune system in IFG. The intricate links between immune system and neuronal signaling is well-appreciated. Other enrichments that serve as a positive control to increase confidence in our MultiCens rankings are those of biological processes like 'regulation of trans-synaptic signaling' in FP and STG, and 'synapse organization' in IFG.

Finally, to find out whether changes in AD-network is specific to the query pathway or similar across pathways, we further use plaque-induced genes (PIGs, total 57 genes), prominent in the later phase of AD, as query set in PHG instead of the SSG set and repeat the same analysis with MultiCens. We found predominant similarities as well as certain interesting differences in centrality ranks between the two query gene sets. While pathways related to heat stress was common for both query sets, synaptic signalling related process like "cell-cell junction organization" was prominent for SSG set and interleukin signaling was exclusively noted for PIG set (see Suppl Figs. S4,S5 and Suppl Results for a detailed discussion). In aggregate, these results on alterations of brain networks in Alzheimer's disease using different query sets show how MultiCens can provide a new network-centric perpsective and related hypotheses for prioritizing experimental investigations of disease mechanisms.

Discussion

We propose a computational framework for modeling a multi-tissue system as a multilayer network and then introduce a set of centrality measures MultiCens to capture the influence of a gene at the tissue and across-tissue levels. MultiCens specifically harnesses the multilayer network structure to decompose the overall centrality of a gene into its local/within-layer vs. global influences, and further into the gene's influence on a particular tissue or a query set of genes in that tissue. Our extensive set of experiments demonstrates the effectiveness of MultiCens on both synthetic and real-world multilayer networks. For instance, with real-world networks learnt from multi-tissue genomic data, MultiCens revealed gene mediators of endocrine hormonal signaling between human tissues, which were then validated via overlap with known hormone-gene relations in HGv1 ground-truth database or in PubMed literature corpus, and via hormonal disease enrichment analysis. Further, out-of-ground-truth gene predictions supported by PubMed literature corpus can in turn be used to prioritize annotation and curation efforts of ground-truth databases. Specifically, these MultiCens predictions can be used to update the current HGv1 database and underlying GO terms with new hormone-producing or responsive genes. In addition to predicting hormone-gene relations, when applied to a multi-brain-region dataset, MultiCens can point to specific genes and pathways whose centrality scores change between AD vs. CTL groups. The novel predictions/hypotheses generated and ranked by MultiCens in both these applications can guide downstream experiments, and thereby foster the emerging field of studying the whole body at the molecular (gene) yet holistic (multi-organ/tissue) levels.

MultiCens performance in predicting hormone-gene relations depends on the quality of the underlying network and that of the query set. Hence, our method would've difficulty with networks inferred from multi-tissue datasets of small sample sizes, and with poorly-studied hormones with very few known gene regulators that could be used as the query set. We get around the sample size issue by applying MultiCens to data from two tissues at a time (the source and target tissue of a hormone profiled in GTEx; see Methods), rather than all tissues at once, which suffers from small sample sizes. To work around the query set issue, we restrict MultiCens predictions to only hormones with sufficient query genes (i.e., at least 10 hormone-producing or responding genes in the ground-truth database). These workarounds have enabled MultiCens to systematically identify known as well as novel gene regulators of hormone-mediated inter-tissue communication. In addition to identifying the involvement of protein-coding genes in inter-tissue communication, our method recognizes potential lncRNAs that may play a crucial role in hormonal signaling pathways [42]. The participation of lncRNA genes in tissue-tissue communication was not known until very recently. Based on our study, experimental studies can be designed to investigate the top-ranked genes to identify their roles in driving cross-tissue communication.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.15.492007; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The concept of brain gene network structure and its shift in neurodegenerative disease such as AD is emerging rapidly. MultiCens helps to understand this shift from a new perspective – we specifically observe how the influence of a given set of genes in a particular brain region on the genes of other brain regions changes in the AD population relative to the control group. We observe the predominance of heat shock protein related pathway (HSP90 particularly) in AD gene-gene network both under the influence of synaptic signaling and PIG related gene set. This may be AD specific change irrespective of region, or may be the result of influence by PHG on AD pathology. Pathways and biological process specific to network in CTL group are also revealed. Major repositioning of genes is seen between AD and CTL group, expect for a few genes, particularly *RBM3* (RNA Binding Motif Protein 3), which is top ranked gene with high centrality score (> 0.9) in both conditions, in all three brain regions and in case of both the query sets. *RBM3* is known to maintain neural stem cell self-renewal and neurogenesis [43]. Does it act as a hub gene for networks linked to PHG, or is an universal hub gene for most of the brain subnetworks? It will be interesting to find the role *RBM3* in brain gene-gene network. Results from this study will help to design specific experiments and give us much better understanding about the brain network structures that are conserved across regions and disease/healthy states, as well as those that are specific to disease states.

The encouraging results from applying MultiCens to understand hormone-gene signaling network and brain network rewiring in AD holds promise for future applications. For instance, MultiCens can be used for "Multi-tissue(-network)-expanded Gene Ontology" analysis of a given set of genes of interest – i.e., computing MultiCens on this query gene set using the underlying multilayer network and coupling it with enrichment analysis can reveal not only pathways directly enriched in this query set as is usually done, but also pathways enriched in the (within-/across-tissue) neighborhood of this query set. MultiCens applications has been human-centric in this study – our preliminary exploration of applying MultiCens to data from a different species like mouse showed that species-specific tuning of our framework may be required, and would be in the scope of future work. Further, MultiCens can also be extended to interpret ligand-receptor interactions. Thus, applicability of MultiCens to study biological systems is manifold.

Beyond the field of biological networks, our measures represent an advance in the overall field of network centrality as well, since existing measures are primarily based on either direct inter-layer interactions [13], or handle multi-hop connectivity but fail to distinguish between within- vs. across-layer interactions [12]. MultiCens accounts for the multilayer multi-hop network connectivity structure of the underlying system. For these reasons and the diverse applications we've demonstrated above, we believe our work on multilayer centrality opens up several future application areas in the area of multi-organ systems-level modeling.

Methods

Our MultiCens framework: context and rationale

Recently, multilayer network modeling has been used to predict functional categories of proteins in multi-tissue systems [44], corroborate experiment findings with open access databases [45], determine potential avenues in the advancement in biomedicine [46], etc. The existing methods of finding multilayer centrality either utilize only the inter-layer degree of the nodes [13] or do not distinguish between within-layer and across-layer connections [12]. While predicting genes involved in inter-tissue communication, we need to emphasize the inter-tissue connections, specifically, connections to hormone-producing or responding tissues/gene-sets. Also, we rely on the hypothesis that hormonal signaling is not simply caused by merely direct connections between hormone-producing and responding genes; other intermediatory genes within the same tissue or in other tissues play the part of mediators in carrying these signals. To accommodate our requirements, we propose a set of centrality measures that can capture the effect of genes at different levels, such as within the same tissue, across tissues, to a specific tissue, or a query set of genes in a particular tissue. This section introduces our proposed methods, starting with degree-based centrality scores and a version of well-known PageRank centrality for multilayer networks, called *versatility*.

Background and Preliminaries

Multilayer network representation

A multilayer network is represented by $G = (V, \mathbb{L}, E)$, where V represent the set of n nodes which is the same across all layers, $\mathbb{L}$ is the set of L number of layers and E represents the set of inter and intra layer edges. The set of nodes in layer α is represented by $V = \{v_1^\alpha, v_2^\alpha, \ldots, v_n^\alpha\}$. The total number of nodes in the multilayer network is $N = n \times L$. Following the convention used in ([47, 48]), we represent the multilayer network by a supra-adjacency matrix M of dimension $N \times N$ as,

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.15.492007; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

$$\mathbf{M}(i_\alpha, j_\beta) = \begin{cases} w(i_\alpha, j_\beta) & \text{if } (v_i^\alpha, v_j^\beta) \in E \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

The supra-adjacency matrix can further be decomposed to represent the network with only intra-tissue edges by A and the network with only inter-tissue edges by C such that,

$$\tilde{A} = A + C$$

$$\begin{bmatrix} A^{[1]} & C^{[1,2]} & C^{[1,3]} & \dots \\ C^{[2,1]} & A^{[2]} & C^{[2,3]} & \ddots \\ C^{[3,1]} & C^{[3,2]} & A^{[3]} & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{bmatrix} = \begin{bmatrix} A^{[1]} & 0 & 0 & \dots \\ 0 & A^{[2]} & 0 & \ddots \\ 0 & 0 & A^{[3]} & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{bmatrix} + \begin{bmatrix} 0 & C^{[1,2]} & C^{[1,3]} & \dots \\ C^{[2,1]} & 0 & C^{[2,3]} & \ddots \\ C^{[3,1]} & C^{[3,2]} & 0 & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{bmatrix}$$

Degree-based methods

Definition 1 *Intra-layer centrality vector of a multilayer network can be computed by the following equation.*

$$deg_{intra} = A\vec{1} \tag{2}$$

Inter-layer degree is a count of the edges that cross the layers. These edges make the backbone of layer-layer communication. The inter-layer degree can be computed using the C matrix as follows.

Definition 2 *Inter-layer centrality vector of a multilayer network can be computed by the following equation.*

$$deg_{inter} = C\vec{1} \tag{3}$$

A more robust way of computing degree-based centrality uses p-values associated with edges instead of weights. A recent method known as S_{sec} makes use of this method to find the prominent set of hormones between a pair of tissues ([13]). S_{sec} is based only on the inter-layer degree of nodes and is used to find hormonal genes that are strongly connected in a pair of tissues.

Definition 3 *For a given p-value matrix P, the S_{sec} score can be computed as follows*

$$S_{sec} = -ln(P)\vec{1} \tag{4}$$

Recently, degree and connectivity patterns such as shortest paths in multilayer networks are being deployed to complete private data with the help of open datasets [45]. Apart from degree-based centrality, there are methods such as PageRank centrality that can capture multihop effects in a network. We will now discuss an existing framework that extends PageRank centrality to a multilayer network.

Versatility

Domenico et al., in their seminal paper ([12]), described a mathematical framework for centrality computation in multiplex networks. The proposed approach assigns a ranking to the nodes based on their interconnectedness. By setting proper weights of the layers (based on the number of nodes/edges), such a ranking method can reveal versatile nodes in the network. For a user-defined constant $p \in [0, 1)$, and N dimensional vector $\vec{1}$, the *versatility* vector can be defined as follows:

Definition 4 *Multilayer network PageRank centrality (also known as versatility ([12])) x of a supra-adjacency network can be defined by the following equation.*

$$x = pMx + \frac{(1-p)}{N}\vec{1} \tag{5}$$

$$x = \Big(I - pM\Big)^{-1}\Big(\frac{(1-p)}{N}\vec{1}\Big)$$

The method itself does not distinguish between the within-layer and cross-layer edges, thus making it unavailing to distinguish the local vs. global effect of nodes. However, the mathematical formulation of a multilayer network described in this work can be extended to define the desired centrality measures, as we will discuss in the upcoming parts. There exist another line of work that focuses on centrality methods for multilayer networks with either no inter layer connections or restricted between identical nodes [49,50,51,52]. We model our multi-tissue datasets by multilayer networks with inter layer connections between any pair of nodes.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.15.492007; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Our proposed methods - MultiCens measures

In the previous section, we discussed methods based on inter-layer degrees and PageRank. Both these methods have shown their usefulness in revealing information about the underlying system. However, the multilayer structure of the network allows us to capture the effect of nodes at multiple levels such as within layer, across the layer, to a target layer, or a query set of genes in a target layer. The existing methods do not capture these effects and thus are limited in their usability. Capturing such effects can have immediate applications in several areas, such as systems biology, where we can identify genes that regulate hormonal communication between two tissues via multiple hops. In order to capture such effects, we propose the following set of centrality measures, as shown in the Fig. 1.

Local centrality

A node in a layer can affect other nodes in the same layer as well as different layers. In order to capture the within layer effect of a node, we define the local centrality as follows

Definition 5 *Local centrality vector is defined as the following iterative equation*

$$l = pAl + \frac{(1-p)}{n}\vec{1} \tag{6}$$

It can be noticed that the local centrality of a node is defined by using only within-layer connections; thus, it does not capture any effects beyond the layer where the node is located. Since *local centrality* considers the effect of only within-layer connections, the remaining effect is captured by *global centrality*.

Global centrality

The global centrality of a node is a measure of its influence on all nodes irrespective of their layers. While computing this centrality score, we use both - within and across tissue connections in the following manner.

Definition 6 *For a given local centrality l, global centrality vector in a multi-layer network can be defined by the following iterative equation*

$$g = p\Big[(A+C)g + Cl\Big] + \frac{(1-p)}{N}\vec{1} \tag{7}$$

The *global centrality* of a node can be thought of as seeing an infinite length random walker on that node where at each step, the random walker can do one of the following.

1. With probability p,
   (a) Jump to a neighboring node $v_{n'}$ in the same layer with probability proportional to the weight of the connection.
   (b) Jump to a neighboring node $v_{n'}$ in a different layer with probability proportional to the weight of the connection and the local centrality of $v_{n'}$.
2. Restart the walk from any node in the network with probability $(1-p)$.

Convergence and Decomposability We now prove the convergence of the proposed centrality measures. The *local centrality* measure is similar to the Pagerank centrality and its convergence follows from the Pagerank centrality convergence itself. Whereas, *global centrality* has additional terms in the equation and we provide a proof for its convergence.

Theorem 1 *For $0 \leq p < 1$, global centrality, as defined by Equation 7 always converges.*

Proof From equation 7,

$$\begin{aligned}
g &= p\Big[(A+C)g + Cl\Big] + \frac{(1-p)}{N}\vec{1} \\
&= p\left[(A+C)\left(p\big[(A+C)g + Cl\big] + \frac{(1-p)}{N}\vec{1}\right) + Cl\right] + \frac{(1-p)}{N}\vec{1} \\
&= p\left[p(A+C)^2 g + p(A+C)Cl + (A+C)\frac{(1-p)}{N}\vec{1} + Cl\right] + \frac{(1-p)}{N} \\
&\vdots \\
&= p^k(A+C)^k g + p\sum_k p^k(A+C)^k Cl + \sum_k p^k(A+C)^k\frac{(1-p)}{N}\vec{1} + \frac{(1-p)}{N}\vec{1}
\end{aligned}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.15.492007; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The first term on the right side converges as k grows larger. The second and third terms give rise to two geometric series generated by $p(A+C)$. We know that $(A+C)$ is a row stochastic matrix and the product $(p(A+C))$ can have maximum eigenvalue, $|\lambda'| < 1$. A geometric series generated by a matrix with eigenvalues less than 1 always converges. This completes the proof.

Kindly refer to the supplementary text for the convergence proofs of the remaining centrality measures defined in this section.

We define *global centrality* and *local centrality* in a way that they add up to the *versatility* in the multilayer network, which the following proof can verify.

Theorem 2 *Versatility of a multilayer network can be decomposed into local centrality and global centrality with a scaling factor.*

$$l + g = x$$

Proof

From equation 6

$$l = pAl + \frac{(1-p)}{n}\vec{1}$$

from equation 7

$$\begin{aligned}
g &= p\Big[(A+C)g + Cl\Big] + \frac{(1-p)}{N}\vec{1} \\
(l+g) &= p\Big[(A+C)g + (A+C)l\Big] + (1-p)(\frac{1}{n} + \frac{1}{N})\vec{1} \\
(l+g) &= p\Big[(A+C)(l+g)\Big] + \frac{(L+1)(1-p)}{N}(\vec{1}) \\
(l+g) &= \Big(I - p(A+C)\Big)^{-1}\Big(\frac{(L+1)(1-p)}{N}\vec{1}\Big) \\
(l+g) &= (L+1)\Big(I - p(\tilde{A})\Big)^{-1}\Big(\frac{(1-p)}{N}\vec{1}\Big) \\
(l+g) &= (L+1)x
\end{aligned}$$

Where L is the total number of layers. Since l, g, and x are centrality vectors, they are scale agnostic, so the constant factor $(L+1)$ on the right side of the equation can be ignored. This completes the proof.

In the remaining section, we decompose *global centrality* into *layer-specific* centrality and further into *query set* centrality. The theoretical proofs for the same are given in the supplement section.

Layer-specific centrality

We are interested in finding the effect of node(s) on a specific layer (target layer) in the multilayer network. In doing so, we define the *layer-specific* centrality as follows.

Definition 7 *For a given l, layer-specific centrality vector in a multi-layer network can be defined by the following iterative equation*

$$g_{layer(i)} = p\Big[(A+C)g_{layer(i)} + C^{[i]}l\Big] + \frac{(1-p)}{N}\vec{1} \tag{8}$$

where $C^{[i]}$ represents the matrix C with all but ith column-block entries set to 0.

Theorem 3 *For $0 \leq p < 1$, $g_{layer(i)}$ defined by Equation 8 always converges.*

Proof Kindly refer to Theorem 3 in the supplementary section.

Theorem 4 *Global centrality, as defined in the main text, can be decomposed for each layer being a specific target layer.*

$$\sum_{i=1}^{L} g_{layer(i)} = g \tag{9}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.15.492007; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Proof Kindly refer to Theorem 4 in the supplementary section.

Our proposed centrality framework is highly generic, and the definition of centrality can further be customized to capture the effect of a node on a set of nodes on a specific target layer. We propose another refinement in the *layer-specific* centrality by decomposing it into multiple query-node sets in the specific target layer.

Query set centrality

We introduce query-set centrality that can capture the effect of a node on a query set of nodes present in any specific layer in the multilayer network. We begin by defining $local - set\ centrality$, a variation of *local centrality* which assigns scores to the nodes in a specific target layer.

Definition 8 *For a given set of query nodes $set(k)$, the local-set centrality in a multilayer network can be defined by the following equation.*

$$l^{set(k)} = pAl^{set(k)} + \frac{(1-p)}{n}\vec{1^k} \tag{10}$$

Where $\vec{1^k}$ represents the vector of $1's$ at indices corresponding to the node-set k and 0 otherwise.

Theorem 5 *For $0 \leq p < 1$, $l^{set(k)}$ defined by Equation 10 always converges.*

Proof Kindly refer to Theorem 5 in the supplementary section.

Theorem 6 *Local-set centrality defined by equation 10 can be added for each set k to obtain the local centrality l.*

$$\sum_{k=1}^{K} l^{set(k)} = l \tag{11}$$

Proof Kindly refer to Theorem 6 in the supplementary section.

We use this *local-set* centrality to define *query-set* centrality as follows.

Definition 9 *For a given set of query genes $set(k)$ in a layer i, the query-set centrality in a multilayer network can be defined by the following equation.*

$$g_{layer(i)}^{set(k)} = p\Big[(A+C)g_{layer(i)}^{set(k)} + C^{[i]}l^{set(k)}\Big] + \frac{(1-p)}{N}\vec{1^k} \tag{12}$$

The *query-set centrality* is defined in order to capture the effect of nodes on a query set of genes in a specific target layer. As shown in Fig. 1, our centrality equations are based on the principle of decomposability.

Theorem 7 *For $0 \leq p < 1$, $g_{layer(i)}^{set(k)}$ defined by Equation 12 always converges.*

Proof Kindly refer to Theorem 7 in the supplementary section.

Theorem 8 *Layer-specific centrality defined by equation 8 can be decomposed into query-set centrality defined over collectively exhaustive subsets of nodes.*

$$\sum_{k} g_{layer(i)}^{set(k)} = g_{layer(i)} \tag{13}$$

Proof Kindly refer to Theorem 8 in the supplementary section.

We restrict our experiments to multilayer networks of only two tissues at once. Having more tissues leaves us with a tiny number of common samples, resulting in a dubious network structure. Our centrality method is designed to handle multiple tissues at once, as we will discuss these experiments in the later section.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.15.492007; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Synthetic multilayer networks

To understand the working of our MultiCens measures, we generate an extensive set of synthetic multilayer networks. As shown in Fig. 2, we begin with a two-layered multilayer network where each layer has 500 nodes. Following the popular ER-random graph generation model [53], we consider all possible pairs of nodes (within and across layer) and put an edge with probability $p = 0.05$. This multilayer network is called the *base* network, and we mark 50 nodes in layer two as the query set. On top of the base network, we add additional edges among the nodes in the query set by another ER-based process of adding random edges. To add these additional edges, we vary this additional edge probability p (called *connection strength*) from $p = 0.05$ to $p = 1$ at steps of 0.05, and obtain a network structure at each step. If a node pair, say (i, j), gets connected in the base network and gets another edge while adding additional edges, we assign weight of two units to the original edge. Similarly, in the first layer, we mark a community of 50 nodes directly connected to the query-set, and call it *source set 1*. Another community of 50 nodes, *source set 2*, is connected to *source set 1*. The connection strength within these two communities and between *source set 1* and *source set 2*, and between *source set 1* and *query set* is varied from 0.05 to 1. In our hormonal signaling example, *query-set* can be thought of as a set of genes that respond to a hormone, say insulin in skeletal muscle tissue. *Source set 1* and *source set 2* can be considered as genes in the pancreas tissue that interact with the *query set* either by direct or two-hop long dense connections. Since the tissues will have multiple other clusters of genes that are not in the proximity of insulin-related genes, we mark three such communities of 50 nodes each. Connection strength within these three communities and across them is also varied.

In this synthetic multilayer network structure, our goal is to understand whether genes from *source set 1* (direct connections) and *source set 2* (two-hop connections) get top centrality-based ranks for a given *query set*, across different values of connection strength.

Real-world Application I: Hormone-related multilayer data, networks, and gene ranking evaluations

Hormone-related multi-tissue data

We work with human multi-tissue datasets and use the following resources.

1. GTEx.v8 Single-Tissue cis-QTL Data [5][1]: This data is a result of the Genotype-Tissue Expression (GTEx) project. The dataset contains gene expression profiles of hundreds of individuals from over 30 tissues. The dataset is pre-processed to account for some known as well as derived covariates[2].
2. Stanford Biomedical Network Dataset Collection [15][3]: This dataset provides a tissue-specific protein-protein edge list for humans. The data is derived from a global protein-protein network. In the global interactions, if a pair of proteins is tissue-specific or if one protein is tissue-specific and the other protein is ubiquitous, then the tissue information is associated with the interaction, and hence the tissue-specific networks are obtained. Physical protein-protein interactions experimentally support the edges in the networks.

We retrieve the hormone-producing and responding gene sets from HGv1 database [16][4]. In HGv1, the source and target genes of hormones are first retrieved in a tissue-agnostic manner, and then through biomedical literature mining source and target tissues of a given hormone is designated. We treat these hormone producing and responding gene sets as the ground truth genes for hormonal signaling.

Hormone-related network construction

Gene coexpression networks are known to capture the patterns of underlying gene expression data that can reveal important biological biomarkers, functional associations between different genes, etc. In human experiments, we make use of the *GTEx.v8 Single-Tissue cis-QTL* data and compute Spearman correlation to find the correlation coefficients between all gene pairs (within and across tissue) and use it as an edge weight (absolute value) to signify the strength of interactions. In order to avoid the blowup in the size of the multilayer network, we only use the top $10k$ varying genes in each tissue and take the union of these genes while constructing the multilayer network.

[1] File "GTEx_Analysis_v8_eQTL_expression_matrices.tar" accessed from "https://gtexportal.org/home/datasets" on Sep 25, 2020.

[2] List of covariates in file "GTEx_Analysis_v8_eQTL_covariates.tar.gz" accessed from "https://gtexportal.org/home/datasets" on Sep 25, 2020.

[3] File "PPT-Ohmnet_tissues-combined.edgelist" accessed from "https://snap.stanford.edu/biodata/datasets/10013/10013-PPT-Ohmnet.html" on Sep 25, 2020.

[4] Files accessed from "https://cross-tissue-signaling.herokuapp.com/" on Jan 10, 2021.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.15.492007; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

We also use the protein-protein interaction data as described earlier, in addition to using a gene coexpression network. For every gene-gene pair, if it is present in the protein interaction data, we increase its weight by 1 unit (adding edge weights) and work with the resultant network. In this paper and its supplementary text, we report results on this resultant network unless mentioned otherwise.

In GTEx dataset, combining multiple tissues in a network leads to fewer common samples and, hence, a less robust network; we restrict these experiments to multilayer networks only with two tissues (the predominant source and target tissue for a hormone; so these multilayer networks we construct and analyze are hormone-specific). However, our network generation mechanism as well as the MultiCens framework to compute centrality can be readily used for any number of tissues, as we illustrate in the Alzheimer's brain network application with four brain regions/tissues.

Evaluation of hormone-gene predictions

In one of MultiCens' applications, we use hormone-producing set as the *query-set* of genes and rank all genes in the target tissue to predict the hormone-responsive set; this process is repeated vice versa to predict hormone-producing genes from an input query set of hormone-responsive genes. We use the HGv1 database [16] as ground truth and validate our gene rankings against it. We also perform disease enrichment analysis to find that whether our centrality-based gene rankings are enriched for hormone-related diseases using WebGestalt[5]. To obtain the enriched set of diseases for human gene rankings, we use the WebGestalt portal and select "Homo sapiens" as the Organism of Interest. Method of interest and Functional Database are set to Gene Set Enrichment Analysis (GSEA) and disease, respectively. We select OMIM functional database and set the significance level to 0.05 FDR. We give the gene symbols, and their corresponding centrality scores as input, and the portal returns the set of diseases enriched at the given FDR cut-off. The gene symbols and their corresponding centrality scores are shared in the supplementary file SD1.

From the gene rankings obtained using our centrality measure, we find the support for top protein-coding genes based on co-occurrence with hormone-related terms in the PubMed corpus[6]. More information about these evaluation approaches are given below.

1. Recall-at-k plot: This plot can be used to validate the results visually. Both in synthetic as well as real-world datasets, we have a set of ground truth genes that we expect to come at the top as per their centrality scores. This can be verified by visualizing *recall-at-k* plots where the x-axis reports the top k predictions and the y-axis marks the number of hits from the ground truth at any given k.
2. Area under *recall-at-k* curve: Higher recall-at-k curve implies the better performance of a method. One way to quantify it is by calculating the area under it. We normalize the maximum possible area under *recall-at-k* curve to be 1 and report the area obtained by curves corresponding to the proposed method.
3. Support from literature: The evaluation metrics discussed above require the ground truth for evaluation. Many times, especially in biology, it is tough to have access to the complete ground set of hormone-producing/responding genes. Continuous research like this study pushes our knowledge boundaries, and we get access to more reliable and more complete ground truth datasets. In order to validate the novel findings, we rely on support from literature and use the following two metrics.
   (a) Co-occurrence in the PubMed database: We use articles present in the PubMed data and find the support for predicted genes. The support is calculated as an overlap between the gene name and the hormone/disease name. The support is calculated using the following formula.

$$Support = \frac{H \cap G}{\frac{H}{\text{number of articles on PubMed}} \times G}$$

   Where H and G denote the number of articles that mention the hormone name and gene name, respectively, and $H \cap G$ denote the number of articles that contain both the hormone name and gene name. While finding support for the gene-disease association, we use articles that contain the disease name instead of hormone name. We use 27 million as the number of articles present in the PubMed database.
   (b) Cosine similarity in the embedding space: We find cosine similarity between the embedding vector of a gene symbol and that of a hormone or disease name. Since cosine similarity can range between -1 and 1, a positive number indicates that the gene-hormone or gene-disease association is supported in the embedded space. Our embeddings (also called as word embeddings or embedding vectors) are from BioWordVec[7], a deep learning model pretrained on the PubMed corpus [25].

[5] Tool http://webgestalt.org/ accessed on Aug 5, 2021.
[6] Data accessed from "https://pubmed.ncbi.nlm.nih.gov/" on Aug 1, 2021.
[7] BioWordVec model/embeddings are downloaded from https://github.com/ncbi-nlp/BioSentVec.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.15.492007; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Both these metrics use articles present in the PubMed database, but they differ because the co-occurrence is based solely on the presence of two terms in an article, whereas the second metric also captures the contextual dependencies in the embedding space.

Our PubMed literature analysis focuses only on the peptide hormones insulin and somatotropin (out of all the four primary hormones considered), since we wanted to apply an informative filter to inspect predictions that are only among genes involved in peptide secretion[8]. This filter was inspired by a similar filter applied in an earlier study on endocrine interactions [13].

Real-world Application II: Alzheimer's vs. Control multilayer data, networks, and rankings

Multi-brain-region data - preprocessing and correction

The covariate-adjusted transcriptomic (RNA-sequencing) data with the following synapse ids - syn16795931 – Brodmann Area (BM10) – frontal pole (FP), syn16795934 - BM22 - superior temporal gyrus (STG), syn16795937 - BM36 - parahippocampal gyrus (PHG), syn16795940 – BM44 - inferior frontal gyrus (IFG), were downloaded from AD Knowledge Portal – The Mount Sinai/JJ Peters VA Medical Center Brain Bank cohort (MSBB) study [35] (10.7303/syn3159438). The preprocessed data is corrected for library size differences using the trimmed mean of M-values normalization (TMM method – edge R package) and linearly corrected for sex, race, age, RIN (RNA Integration Number), PMI (Post-Mortem Interval), sequencing batch, exonic rate and rRNA (ribosomal RNA) rate. The normalization procedure was performed on the concatenated data from all four brain regions to avoid any artificial regional difference as before [35].

The clinical (MSBB_clinical.csv) and experimental metadata (MSBB_RNAseq_covariates_November2018Update.csv) files available on the portal are used to classify the samples into control (CTL) and Alzheimer's disease (AD) based on CERAD score (Consortium to Establish a Registry for AD). CERAD score 1 was used to define CTL samples, and 2 ('Definite AD') was used for defining AD samples [35]. Probable AD (CERAD = 3) and Possible AD (CERAD = 4) samples were not considered for this study.

To mitigate the confounding effect of cellular composition on gene-gene coexpression relations, we corrected (linearly adjusted) the RNAseq gene expression data for cell type frequencies of four major brain cell types: astrocytes, microglia, neuron, and oligodendrocytes. We estimated these cell type frequencies in each brain region/tissue separately from the bulk tissue expression of the marker genes of these cell types using a cellular deconvolution method called CellCODE (Cell-type Computational Differential Estimation) [54]. Specifically, we used the getAllSPVs function from the CellCODE, and provided its input arguments to select robust marker genes that do not change between AD vs. CTL groups (specified via the mix.par argument set at 0.3) from a starting set of 80 marker genes (top 20 per cell type, obtained from the BRETIGEA (BRain cEll Type specIfic Gene Expression Analysis) meta-analysis study [55].

Network construction and enrichment analysis of gene rankings

AD and CTL networks are separately constructed as before by computing the Spearman correlation between all pairs of genes in the four brain regions and taking absolute value of these correlations as the edge weights. To make the analysis computationally tractable, we restrict our focus to a subset of genes as follows - identify the 9000 most varying genes in each region for both AD and CTL populations, and then consider the union of all these gene sets as the final set of nodes in each layer of the multilayer network.

MultiCens scores are then calculated for all the nodes in the AD or CTL multilayer networks to obtain gene rankings, which were then subjected to enrichment analysis with WebGestalt as described before. Additionally, we applied redundancy reduction methods (affinity propagation and weighted set cover) and selected the significantly enriched terms, which passed both the methods. We use the centrality score of each of the three brain regions other than the query brain region to find the significantly enriched terms considering both Reactome pathways and Gene Ontology based Biological Process (GO-BP).

Code and Data Availability

The code that implements both network construction and MultiCens measures is available here: https://github.com/BIRDSgroup/MultiCens.

[8] List of genes involved in peptide secretion accessed from this URL- www.ebi.ac.uk/QuickGO/GTerm?id=GO:0002790 on Dec 1, 2020

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.15.492007; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Acknowledgments

We thank members of our BIRDS (Bioinformatics and Integrative Data Science) research group for their valuable inputs during presentations of this work. The research presented in this work was supported by Wellcome Trust/DBT grant IA/I/17/2/503323 awarded to MN, Intel research grant RB/18-19/CSE/002/INTI/BRAV to BR, and Intel PhD Fellowship awarded to TK.

Author Contributions

TK, MN, and BR conceived this study, and formulated it with additional inputs from RS; TK developed MultiCens measures, proved theoretical guarantees, and tested them on synthetic networks with inputs from BR and MN; TK and MN conceived the two biological applications of MultiCens with inputs from SM, RS, and BR; SM contributed to biological interpretation of the results (gene rankings) in the two applications, with inputs from TK, MN, and RS; TK, MN, and SM wrote the manuscript with inputs from BR and RS; MN guided and supervised the computational as well as biological aspects of the study.

References

1. Huang, Z. & Xu, A. Adipose extracellular vesicles in intercellular and inter-organ crosstalk in metabolic health and diseases. *Frontiers in Immunology* **12**, 463 (2021).
2. Droujinine, I. & Perrimon, N. Defining the interorgan communication network: systemic coordination of organismal cellular processes under homeostasis and localized stress. *Frontiers in cellular and infection microbiology* **3**, 82 (2013).
3. Droujinine, I. A. *et al.* Proteomics of protein trafficking by in vivo tissue-specific labeling. *Nature communications* **12**, 1–22 (2021).
4. Bodine, S. C. *et al.* An american physiological society cross-journal call for papers on "inter-organ communication in homeostasis and disease" (2021).
5. Lonsdale, J. *et al.* The genotype-tissue expression (gtex) project. *Nature genetics* **45**, 580–585 (2013).
6. Wang, M. *et al.* The mount sinai cohort of large-scale genomic, transcriptomic and proteomic data in alzheimer's disease. *Scientific data* **5**, 1–16 (2018).
7. Langfelder, P. & Horvath, S. Wgcna: an r package for weighted correlation network analysis. *BMC bioinformatics* **9**, 1–13 (2008).
8. Moreau, Y. & Tranchevent, L.-C. Computational tools for prioritizing candidate genes: boosting disease gene discovery. *Nature Reviews Genetics* **13**, 523–536 (2012).
9. López-Cortés, A. *et al.* Gene prioritization, communality analysis, networking and metabolic integrated pathway to better understand breast cancer pathogenesis. *Scientific reports* **8**, 1–15 (2018).
10. Kolosov, N., Daly, M. J. & Artomov, M. Prioritization of disease genes from gwas using ensemble-based positive-unlabeled learning. *European Journal of Human Genetics* 1–9 (2021).
11. Page, L., Brin, S., Motwani, R. & Winograd, T. The pagerank citation ranking: Bringing order to the web. Tech. Rep., Stanford InfoLab (1999).
12. De Domenico, M., Solé-Ribalta, A., Omodei, E., Gómez, S. & Arenas, A. Ranking in interconnected multilayer networks reveals versatile nodes. *Nature communications* **6**, 6868 (2015).
13. Seldin, M. M. *et al.* A strategy for discovery of endocrine interactions with application to whole-body metabolism. *Cell metabolism* **27**, 1138–1155 (2018).
14. Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
15. Zitnik, M., Rok Sosic, S. & Leskovec, J. Biosnap datasets: Stanford biomedical network dataset collection. *Note: http://snap. stanford. edu/biodata Cited by* **5** (2018).
16. Jadhav, A., Kumar, T., Raghavendra, M., Loganathan, T. & Narayanan, M. Predicting cross-tissue hormone-gene relations using balanced word embeddings. *bioRxiv* (2021).
17. Reaven, G. M. Insulin-independent diabetes mellitus: metabolic characteristics. *Metabolism* **29**, 445–454 (1980).
18. Trabert, B., Sherman, M. E., Kannan, N. & Stanczyk, F. Z. Progesterone and breast cancer. *Endocrine reviews* **41**, 320–344 (2020).
19. Yang, X. *et al.* Growth hormone receptor expression in human colorectal cancer. *Digestive diseases and sciences* **49**, 1493–1498 (2004).
20. Gallagher, E. J. & LeRoith, D. Hyperinsulinaemia in cancer. *Nat Rev Cancer* **20**, 629–644 (2020).
21. Orgel, E. & Mittelman, S. D. The links between insulin resistance, diabetes, and cancer. *Curr Diab Rep* **13**, 213–222 (2013).
22. D'Souza, D. M., Al-Sajee, D. & Hawke, T. J. Diabetic myopathy: impact of diabetes mellitus on skeletal muscle progenitor cells. *Front Physiol* **4**, 379 (2013).
23. Chesnokova, V. *et al.* Growth hormone is permissive for neoplastic colon growth. *Proc Natl Acad Sci U S A* **113**, E3250–3259 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.15.492007; this version posted May 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

24. Fitzgerald, P. J. Is norepinephrine an etiological factor in some types of cancer? *International journal of cancer* **124**, 257–263 (2009).
25. Zhang, Y., Chen, Q., Yang, Z., Lin, H. & Lu, Z. BioWordVec, improving biomedical word embeddings with subword information and MeSH. *Scientific Data* **6** (2019).
26. Stuhlmann, T., Planells-Cases, R. & Jentsch, T. J. Lrrc8/vrac anion channels enhance β-cell glucose sensing and insulin secretion. *Nature communications* **9**, 1–12 (2018).
27. Kumar, A. *et al.* Swell1 regulates skeletal muscle cell size, intracellular signaling, adiposity and glucose metabolism. *Elife* **9**, e58941 (2020).
28. Shraim, B. A., Moursi, M. O., Benter, I. F., Habib, A. M. & Akhtar, S. The Role of Epidermal Growth Factor Receptor Family of Receptor Tyrosine Kinases in Mediating Diabetes-Induced Cardiovascular Complications. *Front Pharmacol* **12**, 701390 (2021).
29. Chan, P.-C. *et al.* Targeted inhibition of cd74 attenuates adipose cox-2-mif-mediated m1 macrophage polarization and retards obesity-related adipose tissue inflammation and insulin resistance. *Clinical Science* **132**, 1581–1596 (2018).
30. Baas, D. *et al.* A deficiency in rfx3 causes hydrocephalus associated with abnormal differentiation of ependymal cells. *European Journal of Neuroscience* **24**, 1020–1030 (2006).
31. Wen, M.-H., Hsiao, H.-P., Chao, M.-C. & Tsai, F.-J. Growth hormone deficiency in a case of crouzon syndrome with hydrocephalus. *International journal of pediatric endocrinology* **2010**, 1–4 (2010).
32. López-Noriega, L. & Rutter, G. A. Long Non-Coding RNAs as Key Modulators of Pancreatic β-Cell Mass and Function. *Front Endocrinol (Lausanne)* **11**, 610213 (2020).
33. Devesa, J., Almengló, C. & Devesa, P. Multiple Effects of Growth Hormone in the Body: Is it Really the Hormone for Growth? *Clin Med Insights Endocrinol Diabetes* **9**, 47–71 (2016).
34. Giustina, A., Mazziotti, G. & Canalis, E. Growth hormone, insulin-like growth factors, and the skeleton. *Endocr Rev* **29**, 535–559 (2008).
35. Wang, M. *et al.* The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease. *Sci Data* **5**, 180185 (2018).
36. Yan, T., Ding, F. & Zhao, Y. Integrated identification of key genes and pathways in Alzheimer's disease via comprehensive bioinformatical analyses. *Hereditas* **156**, 25 (2019).
37. Alves, J., Petrosyan, A. & Magalhães, R. Olfactory dysfunction in dementia. *World J Clin Cases* **2**, 661–667 (2014).
38. Zhang, Q. *et al.* Integrated proteomics and network analysis identifies protein hubs and network alterations in Alzheimer's disease. *Acta Neuropathol Commun* **6**, 19 (2018).
39. Shelton, L. B. *et al.* Hsp90 activator Aha1 drives production of pathological tau aggregates. *Proc Natl Acad Sci U S A* **114**, 9707–9712 (2017).
40. Su, F. *et al.* CIRBP Ameliorates Neuronal Amyloid Toxicity via Antioxidative and Antiapoptotic Pathways in Primary Cortical Neurons. *Oxid Med Cell Longev* **2020**, 2786139 (2020).
41. Ou, J. R., Tan, M. S., Xie, A. M., Yu, J. T. & Tan, L. Heat shock protein 90 in Alzheimer's disease. *Biomed Res Int* **2014**, 796869 (2014).
42. Sun, M. & Kraus, W. L. From discovery to function: the expanding roles of long noncoding rnas in physiology and disease. *Endocrine reviews* **36**, 25–64 (2015).
43. Yan, J. *et al.* The RNA-Binding Protein RBM3 Promotes Neural Stem Cell (NSC) Proliferation Under Hypoxia. *Front Cell Dev Biol* **7**, 288 (2019).
44. Zitnik, M. & Leskovec, J. Predicting multicellular function through multi-layer tissue networks. *Bioinformatics* **33**, i190–i198 (2017).
45. Malek, M., Zorzan, S. & Ghoniem, M. A methodology for multilayer networks analysis in the context of open and private data: biological application. *Applied Network Science* **5**, 1–28 (2020).
46. Hammoud, Z. & Kramer, F. Multilayer networks: aspects, implementations, and application in biomedicine. *Big Data Analytics* **5**, 1–18 (2020).
47. Gomez, S. *et al.* Diffusion dynamics on multiplex networks. *Physical review letters* **110**, 028701 (2013).
48. Kumar, T., Narayanan, M. & Ravindran, B. Effect of inter-layer coupling on multilayer network centrality measures. *Journal of the Indian Institute of Science* **99**, 237–246 (2019).
49. Óskarsdóttir, M. & Bravo, C. Multilayer network analysis for improved credit risk prediction. *Omega* **105**, 102520 (2021).
50. Lv, L. *et al.* Hits centrality based on inter-layer similarity for multilayer temporal networks. *Neurocomputing* **423**, 220–235 (2021).
51. Frost, H. R. Eigenvector centrality for multilayer networks with dependent node importance. *arXiv preprint arXiv:2205.01478* (2022).
52. Wang, D., Wang, H. & Zou, X. Identifying key nodes in multilayer networks based on tensor decomposition. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **27**, 063108 (2017).
53. Erdos, P., Rényi, A. *et al.* On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci* **5**, 17–60 (1960).
54. Chikina, M., Zaslavsky, E. & Sealfon, S. C. CellCODE: a robust latent variable approach to differential expression analysis for heterogeneous cell populations. *Bioinformatics* **31**, 1584–1591 (2015).
55. McKenzie, A. T. *et al.* Brain Cell Type Specific Gene Expression and Co-expression Network Architectures. *Sci Rep* **8**, 8868 (2018).