Gene modules associated with human diseases revealed by network analysis

Shisong Ma[1,2,*], Jiazhen Gong[1,§], Wanzhu Zuo[1,§], Haiying Geng[1], Yu Zhang[1], Meng Wang[1], Ershang Han[1], Jing Peng[1], Yuzhou Wang[1], Yifan Wang[1], Yanyan Chen[1]

1. Hefei National Laboratory for Physical Sciences at the Microscale, School of Life Sciences, University of Science and Technology of China, Hefei, Anhui 230027, China
2. School of Data Science, University of Science and Technology of China, Hefei, Anhui 230027, China

* Corresponding author: Shisong Ma (sma@ustc.edu.cn)

§ These authors contribute equally.

Abstract

Within biological networks, genes associated with human diseases likely map to modules whose identification facilitates etiology studies but remains challenging. We describe a systematic approach to identify such disease-associated gene modules. A gene co-expression network was constructed using GTEx dataset and assembled into 652 gene modules. Screening these modules identified those with disease genes enrichment for obesity, cardiomyopathy, hypertension, and autism, which revealed the pathways involved in their pathogenesis. Using mammalian phenotypes derived from mouse models, potential disease candidate genes were identified from these modules. Also analyzed were epilepsy, schizophrenia, bipolar disorder, and depressive disorder, revealing shared and distinct disease modules among brain disorders. Thus disease genes converge on modules within our network, which provides a general framework to dissect genetic basis of human diseases.

Main Text

Human diseases often have genetic basis and their corresponding gene-disease associations are widely documented, with many of them possessing multiple disease genes (*1-3*). Within biological networks, disease genes usually do not distribute randomly but map to modules that include subsets of genes functioning together in the same or similar pathways (*4, 5*). It is of great interest to identify such disease gene modules due to their values in elucidating disease etiologies (*6-8*). Yet for most diseases, such modules have not been detected. We describe a systematic approach to identify gene modules associated with human diseases (Fig. 1A). Our analysis revealed modules for obesity, cardiomyopathy, hypertension, autism, and other diseases with major impacts on public health worldwide.

A human gene co-expression network based on the graphical Gaussian model (GGM) was constructed using publicly available transcriptome data from the Genotype-Tissue Expression (GTEx) project (*9*). The project has created a resource of gene expression data from 'normal', non-diseased tissues, which supported a range of studies, including co-expression analysis (*10, 11*). Our work used GTEx V7 release that contains transcriptome data for 11688 samples spanning 53 tissues from 714 postmortem donors

(*9*). The analysis followed a procedure we published previously (*12, 13*), which employed partial correlation coefficient (pcor) for co-expression measurement (*14*). The resulted network, HsGGM2019, contains 166980 co-expressed gene pairs among 18425 protein-coding genes (table S1). Via the MCL algorithm (*15*), 652 gene co-expression modules with 9 or more genes were identified from the network. Considering some genes might participate in multiple pathways, the modules were expanded by including outside genes connected with >= 3 genes within the original modules. Accordingly, 14391 genes assembled into 652 modules containing 9 to 285 genes each, with 2668 genes belonging to multiple modules (Fig. 1B and table S2). The rest 4034 genes organized into smaller modules that were not considered hereafter.

The modules contain co-expressed genes that, according to guilt-by-association, might function in the same or similar pathways. Gene Ontology (GO) analysis identified 341 of them with enriched GO terms (Benjamini-Hochberg adjusted pValue [*P*]<=1E-2) (table S3). Their biological relevance are exemplified by three modules involved in lipid biosynthesis, transport, and storage. Module #71, expressed broadly in various tissues (fig. S1), is enriched with 23 cholesterol biosynthesis genes (*P*=1.92E-42) (Fig. 1C), including genes for enzymes catalyzing 23 of all 24 reaction steps for synthesizing cholesterol from acetyl-CoA (fig. S2). It also contains *LDLR*, *PCSK9*, *SREBF2*, and *INSIG1*, encoding components of the PCSK9-LDLR and SREBP-SCAP-Insig complexes that regulate cholesterol homeostasis (*16*). Module #30, containing genes with restricted expression towards liver (fig. S1), is enriched with high-density lipoprotein (HDL) particle genes (*P*=5.29E-14), such as *APOA1*, *APOA2*, and *APOA5*. This module could function in HDL-mediated reverse cholesterol transport. Module #18, expressed biasedly in fat (fig. S1), is enriched with lipid storage genes (*P*=3.37E-12). It contains *PPARG*, encoding a key transcription factor regulator of adipogenesis, and its target genes involved in adipocyte differentiation and metabolism, like *PLIN1*, *FABP4*, *LEP*, and *ADIPOQ* (*17*). Modules were also revealed for other processes, such as those functioning in specific organelles or tissues (#1, #27, #50), metabolism pathways (#54, #126, #288), immunity pathways (#6, #8, #15), or general cellular pathways (#3, #28, #29) (see Fig. 1D and table S3 for these modules' enriched GO, same as below).

The modules were then screened for association with diseases. A module is considered as disease-associated if it has disease genes enrichment (permutation-based False Discovery Rate [FDR]<=0.05). Curated gene-disease associations for obesity, autism, and other diseases, retrieved from obesity gene map, SFARI, and DisGeNET (*1-3*), were queried against the modules to identify those associated with diseases (Fig. 2A and table S4). As an example, Module #18 for lipid storage is associated with obesity, a disease involving excessive body fat that affects 12% adults worldwide (*18*). Besides environment, genetic factors influence obesity susceptibility. According to obesity gene map, 21 out of the 113 genes within Module #18 are obesity genes (Fig. 2B), including *ADIPOQ* and *LEP*, whose genetic polymorphisms are risk factors for obesity (*19*). Comparing to the whole network, obesity genes are 10.2 fold enriched (Odd Ratio, OR) within the module (FDR<0.0001). We examined phenotypes of transgenic or null mouse models developed for the genes within the module, using Mammalian Phenotype (MP) Ontology assignments from MGI (*20*), to investigate if other genes also relate to obesity. The MP *abnormal body weight* is associated with 33 genes and significantly enriched within the module (*P*=3.02E-07) (Fig. 2B, and see table S5 for MP enrichment analysis results for all modules, and refer to Materials and Methods for analysis details). Among them are 14 obesity genes, and the rest 19 might represent additional candidate genes. Note that these candidate genes derived from mouse models need to be further verified by, i.e., human population genetics studies. Additionally, Module #16 contains 7 genes for insulin resistance (OR=17.7, FDR<0.0001), a syndrome that could result from obesity (*21*). It also has 13 genes with the MP *insulin resistance* (*P*=1.66E-10), again expanding the disease's candidate gene list. Also discovered are 8 additional obesity modules (Fig. 2A and table S4), 7 of which function in lipid transport (#30), extracellular matrix organization (#101), neuropeptide signaling pathway (#277), feeding behavior regulation (#546), dopamine metabolism (#288), circulatory system development (#34), and apoptotic process (#60), respectively. Among them, dopamine relates to obesity via modulating appetite, while within Module #277 are also genes regulating feeding behavior, such as *HCRT* and *PMCH* (*22, 23*).

Besides obesity, modules were identified for diseases affecting specific tissues or organs. Cardiomyopathy is a heart muscle disease, where the heart muscle becomes enlarged, thick or stiff (*24*). Dilated cardiomyopathy (DCM) and hypertrophic cardiomyopathy (HCM) are among the major types of cardiomyopathy. Five gene modules are associated with DCM and/or HCM (Fig. 2A and table S4). All top 3 DCM modules, #85, #92, #39 (OR>=19.2, FDR<0.0001), have enriched GO *muscle contraction* (*P*<=5.41E-20), indicating their muscle-related functions. While Module #85 and #92 are specifically expressed in heart and skeletal muscles, #39 is also expressed in tissues with smooth muscles like artery, colon, and esophagus (fig. S3). Pathway analysis (*25*) indicated that #85 and #92 function in striated muscle contraction and #39 in smooth muscle contraction. Module #70 for mitochondrial fatty acid beta-oxidation is also associated with DCM (OR=22.4, FDR=0.0001). HCM is associated with Modules #85, #39, #70, and another mitochondrial Module #2 (OR=19.4, FDR<0.0001), consistent with that cardiomyopathy could result from mitochondrial dysfunction (*26*). Contained within Module #85, #92 and #39 are 42 genes with the MP *enlarged heart*, including 27 as novel potential candidate genes for cardiomyopathy (Fig. 2C and fig. S4). Other examples of tissue-specific disease modules include those for ciliary motility disorders (#1 and #67), pulmonary alveolar proteinosis (#154), akinesia (#56), hereditary pancreatitis (#105), congenital hypothyroidism (#51), and night blindness (#114) (table S4).

Also identified were modules for hypertension, a complex and multifactorial disease that affects nearly one billion people worldwide (*27*). Using curated gene-disease associations from DisGeNET, 10 hypertension modules were identified (Fig. 2A and table S4), 7 of which form two large categories. First are 2 modules regulating sodium or ion homeostasis. Module #281, expressed specifically in kidneys (fig. S5), is enriched with 7 *ion transmembrane transporter* genes (*P*=2.90E-04). It contains 5 hypertension genes (OR=20.7, FDR=0.0001) (Fig. 3A), including *UMOD* that encodes uromodulin, the major secreted protein in normal urine. Noncoding risk variants increase *UMOD* expression, which causes abnormal activation of renal sodium cotransporter SLC12A1 and leads to hypertension (*28*). Interestingly, *SLC12A1* is also in Module #281, together with *TMEM72*. Both genes were identified as blood pressure loci in a

GWAS study based on genetic analysis of over 1 million people (*29*), but not in DisGeNET. The module contains additional genes with the MP *abnormal kidney physiology*, among which might include novel candidate genes for hypertension, as kidneys are important for blood pressure control (*30*). Module #305, expressed more broadly than #281 (fig. S5), is enriched with 4 hypertension genes (OR=16.5, FDR=0.0022) like *SCNN1B* and *SCNN1G*, which also modulate renal sodium homeostasis (*31*). Second are 5 modules involved in vascular tone regulation. Module #66, over-represented with *extracellular matrix* genes (*P*=1.03E-54), contains 8 hypertension genes (OR=7.1, FDR=0.0006), such as *ELN* and *FBN1* (Fig. 3B). Its enriched MPs include *abnormal blood vessel morphology* and *abnormal blood circulation*, indicating that, similar to *ELN* and *FBN1* (*32*), genes in the module could determine vascular extracellular matrix composition and regulate arterial compliance. Similarly, Module #101 is also enriched with 35 *extracellular matrix* genes as well as 6 hypertension genes (OR=5.1, FDR=0.0314). Module #39, functioning in smooth muscle contraction, encompasses 13 hypertension genes (OR=7.4, FDR<0.0001), among which are *KCNMB1* and *PRKG1*, encoding key modulators of smooth muscle tone (*33, 34*). Genes within the module, enriched with MPs *abnormal vascular smooth muscle physiology* and *impaired smooth muscle contractility*, could regulate arterial wall elasticity. Module #54, for glucocorticoids metabolism, contains 6 hypertension genes (OR=7.2, FDR=0.0037), and #126 (Fig. 3C), participating in cGMP metabolism, includes 4 hypertension genes (OR=9.4, FDR=0.0231). Glucocorticoids and cGMP modulate blood pressure via regulating peripheral vascular resistance and blood vessel relaxation respectively (*35, 36*). The relevance of these 5 modules are further supported by that, besides the hypertension genes in DisGeNET, Module #39, #54, #66, #101, and #126 contain 8, 3, 6, 6, and 3 additional blood pressure loci respectively (Fig. 3, B and C, and fig. S6), as revealed by the above mentioned GWAS study (*29*). The rest 3 hypertension modules, #95, #30, #374 (Fig. 3D) function in inflammatory response, HDL-mediated lipid transport, and mitochondrial ATP synthesis, respectively.

Our approach also identified modules associated with brain disorders, among which is autism, a development disorder characterized by impaired social interaction and by restricted and repetitive behaviors

(*37*). Using SFARI autism genes (*2*), 14 autism modules were detected (Fig. 2A and table S4), 8 of which form three large categories. First is a module involved in *chromatin modification* (#7), which contains 194 genes, including 42 autism genes (OR=5.9, FDR<0.0001) (Fig. 4A), such as *ASH1L*, *CHD8*, and *KMT2C* (*38*). The module has 35 genes with the MP *abnormal brain morphology*, among which 11 are autism genes. It is also over-represented with MPs *abnormal heart morphology*, *abnormal gastrulation*, and *abnormal axial skeleton morphology*. Thus gene mutations in this module might affect multiple developmental processes, including brain development that could lead to autism. Second are 2 modules involved in *nervous system development* (#77) and *axonogenesis* (#9). Among them, Module #9 contains 16 autism genes (OR=3.3, FDR=0.0017) (Fig. 4B), like *NLGN3*, *NLGN4X*, and *NRXN1* (*39*). The module is enriched with autism-related MPs, such as *abnormal nervous system physiology* and *abnormal motor coordination/balance* (*40*). Among the 49 genes with the MP *abnormal nervous system physiology*, 10 are autism genes and the rest might contain additional candidate genes. Third are 5 modules participating in synaptic signaling. Though all enriched with *synaptic signaling* genes (*P*=2.78E-34 to 6.71E-05), these modules express differently. Module #20 and #35 have biased expression in cerebellum and cortex respectively, while #11, #76, and #124 express broadly in various brain regions (fig. S7). All 5 modules are enriched with autism genes and might contain potential candidate genes. For example, Module #20 has 15 autism genes (OR=3.1, FDR=0.0049) (Fig. 4C), like *MYTL1* and *DLGAP1* (*38, 41*). It also contains 49 genes with MPs *abnormal social/conspecific interaction* and/or *abnormal motor capabilities/coordination/movement*, among which are 10 autism genes and the rest merit further investigation. Among the other autism modules, #186 and #131 function in cell-cell adhesion and extracellular matrix organization respectively, #358 is enriched with *neuron part* genes, while #168, #266 and #325 are without clear biological interpretation presently. Thus, these modules indicate abnormality in chromatin modification, nervous system development, synaptic signaling and other processes could lead to autism.

Modules for epilepsy, schizophrenia, bipolar disorder, and depressive disorder are also identified, revealing shared and distinct disease modules among brain disorders (Fig. 4D and table S4). Module #7 for chromatin modification is associated with autism and epilepsy only, but Module #11 and 76 for synaptic signaling are shared by autism, schizophrenia, bipolar disorder and depressive disorder. Also shared by bipolar disorder and depressive disorder is Module #231 involved in circadian rhythm regulation, highlighting circadian rhythm's role in mood disorders (*42*). On the other hand, epilepsy is distinctly associated with modules functioning in mitochondria (#2, #22, #70), lysosomes (#27), and kidneys (#281), indicating its unique etiology (*43-45*). Schizophrenia is uniquely associated with modules for MHC antigen presentation (#83, #264), myelination (#14), and fatty acid biosynthesis (#520), while depressive disorder is particularly associated with Module #95 for inflammatory response (*46-48*). The rest shared or distinct brain disorder modules also function in synaptic signaling (#93, #146, #187, #229), ion transport (#72, #113), neuropeptide signaling (#277), and dopamine metabolism (#288). These modules delineate the pathways associated with brain disorders and should facilitate their etiology studies.

Thus, GGM network analysis identified unbiased data-driven gene modules with enriched functions in a variety of pathways and tissues. Disease genes converge on these modules, although the network was derived from non-diseased samples. Such convergence is not limited to diseases affecting specific tissues but also applies to complex and multifactorial diseases like hypertension and autism. The identified disease modules, mostly with clear biological interpretation, integrate well with previous disease knowledge. They provide useful information about the etiological pathways of the diseases. Based on MP assignments from mouse models, potential disease candidate genes were identified from within the modules. Therefore, the modules can be used to pinpoint the pathways involving in diseases and reveal potential novel disease genes. Our current work focused on coding genes only, but future analysis can also include non-coding genes, such as long non-coding RNAs, to study their roles on diseases development.

Materials and Methods

Co-expression network construction

Open-access and de-identified GTEx V7 transcriptome data were downloaded from the GTEx Portal (http://www.gtexportal.org). The data were fully processed and available as a matrix that contains TPM gene expression values for 53035 genes in 11688 samples spanning 53 tissues from 714 postmortem donors. After filtering out low expressed genes that have TPM values >=1 in less than 10 samples, the expression data were normalized in a tissue-aware manner via qsmooth with default parameters (*49*). A sub-matrix consisting of 18626 protein-coding genes was extracted from the normalized expression matrix and used for GGM network construction, following a procedure described previously (*12, 13*). Briefly, the procedure consisted of 6000 iterations. In each iteration, 2000 genes were randomly selected and used for partial correlation coefficient (pcor) calculation via the GeneNet v 1.2.13 package in R (*14*). After 6000 iterations, every gene pair was sampled in average 69 times with 69 pcors calculated, and the pcor with lowest absolute value was selected as its final pcor. Also calculated were Pearson's correlation coefficient (r) between gene pairs. Finally, gene pairs with pcor>=0.035 and r>=0.35 were chosen for network construction.

Network clustering and module analysis

The network was clustered via the MCL clustering algorithm with parameters "-I 1.55 –scheme 7" (*15*). The identified modules with >=9 genes were kept, and they were further expanded by including outside genes that connect with >= 3 genes within the original modules. The sub-networks for the modules were visualized in Cytoscape v 3.40 (*50*). GO enrichment analysis were performed via hypergeometric test, with GO annotations retrieved from Ensembl BioMart (https://www.ensembl.org/biomart) on 03/06/2019. Mouse Mammalian Phenotype (MP) term assignments were obtained from the MGI database (http://www.informatics.jax.org/downloads/reports/MGI_GenePheno.rpt) on 03/07/2019. MP term assignments derived from mouse models involving 2 or more genes were excluded. Mouse genes' MP

assignments were passed on to their human orthologues and used for MP enrichment analysis via hypergeometric test.

Identification of modules associated with diseases

Gene-disease associations for obesity were obtained from the human obesity gene map (*1*). Autism genes were obtained from the SFARI database on 01/15/2019, and only the autism genes scoring as category 1 to category 4 were used in the analysis (*2*). For other diseases, the curated gene-disease associations registered in DisGeNET v5.0 were used (*3*). For hypertension, besides the disease genes from DisGeNET, blood pressure loci were also obtained from a recent GWAS study, by combining the previous reported blood pressure loci and the newly confirmed variants, as listed in Supplementary Table 4 and 5 in the article by Evangelou *et al.* (*29*).

For every disease, its disease gene list were queried against every gene module to calculate a pValue for disease gene enrichment using hypergeometric test. Suppose a disease has m disease genes within a gene module with the size of k, and M disease genes among all K genes in the whole network. A pValue for that disease and module combination was calculated as:

$$pValue\,(module, disease) = \sum_{l=m}^{\min(k,M)} \frac{\binom{k}{l}\binom{K-k}{M-l}}{\binom{K}{M}}$$

For every disease, the pValues for all modules were adjusted for multiple testing via the Benjamini-Hochberg procedure (*51*).

A permutation based procedure was also used to estimate FDR. In each permutation, every disease's disease gene list were replaced by the same number of genes randomly selected from the whole network and used for enrichment calculation. After conducting 10000 permutations, the results were tallied and used to calculate the FDRs corresponding to original pValues.

References

1. T. Rankinen *et al.*, The human obesity gene map: the 2005 update. *Obesity (Silver Spring)* **14**, 529-644 (2006).
2. B. S. Abrahams *et al.*, SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol Autism* **4**, 36 (2013).
3. J. Pinero *et al.*, DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res* **45**, D833-D839 (2017).
4. A. J. Willsey *et al.*, The Psychiatric Cell Map Initiative: A Convergent Systems Biological Approach to Illuminating Key Molecular Pathways in Neuropsychiatric Disorders. *Cell* **174**, 505-520 (2018).
5. A. L. Barabasi, N. Gulbahce, J. Loscalzo, Network medicine: a network-based approach to human disease. *Nat Rev Genet* **12**, 56-68 (2011).
6. B. Zhang *et al.*, Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* **153**, 707-720 (2013).
7. J. Menche *et al.*, Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science* **347**, 1257601 (2015).
8. F. Hormozdiari, O. Penn, E. Borenstein, E. E. Eichler, The discovery of integrated gene networks for autism and related disorders. *Genome Res* **25**, 142-154 (2015).
9. G. T. Consortium *et al.*, Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213 (2017).
10. A. Saha *et al.*, Co-expression networks reveal the tissue-specific regulation of transcription and splicing. *Genome Res* **27**, 1843-1858 (2017).
11. E. Pierson *et al.*, Sharing and Specificity of Co-expression Networks across 35 Human Tissues. *PLoS Comput Biol* **11**, e1004220 (2015).
12. S. Ma, Q. Gong, H. J. Bohnert, An Arabidopsis gene network based on the graphical Gaussian model. *Genome Res* **17**, 1614-1625 (2007).
13. S. Ma, M. Snyder, S. P. Dinesh-Kumar, Discovery of Novel Human Gene Regulatory Modules from Gene Co-expression and Promoter Motif Analysis. *Scientific reports* **7**, 5557 (2017).
14. J. Schäfer, K. Strimmer, A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat Appl Genet Mol Biol* **4**, Article32 (2005).
15. A. J. Enright, S. Van Dongen, C. A. Ouzounis, An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* **30**, 1575-1584 (2002).
16. L. Goedeke, C. Fernandez-Hernando, Regulation of cholesterol homeostasis. *Cell Mol Life Sci* **69**, 915-930 (2012).
17. R. J. Perera *et al.*, Identification of novel PPARgamma target genes in primary human adipocytes. *Gene* **369**, 90-99 (2006).
18. G. B. D. O. Collaborators *et al.*, Health Effects of Overweight and Obesity in 195 Countries over 25 Years. *N Engl J Med* **377**, 13-27 (2017).
19. J. E. Enns, C. G. Taylor, P. Zahradka, Variations in Adipokine Genes AdipoQ, Lep, and LepR are Associated with Risk for Obesity-Related Metabolic Disease: The Modulatory Role of Gene-Nutrient Interactions. *J Obes* **2011**, 168659 (2011).
20. C. L. Smith, C. A. Goldsmith, J. T. Eppig, The Mammalian Phenotype Ontology as a tool for annotating, analyzing and comparing phenotypic information. *Genome Biol* **6**, R7 (2005).
21. B. B. Kahn, J. S. Flier, Obesity and insulin resistance. *J Clin Invest* **106**, 473-481 (2000).
22. M. Shimada, N. A. Tritos, B. B. Lowell, J. S. Flier, E. Maratos-Flier, Mice lacking melanin-concentrating hormone are hypophagic and lean. *Nature* **396**, 670-674 (1998).
23. G. J. Wang *et al.*, Brain dopamine and obesity. *Lancet (London, England)* **357**, 354-357 (2001).
24. W. M. Franz, O. J. Muller, H. A. Katus, Cardiomyopathies: from genetics to the prospect of treatment. *Lancet (London, England)* **358**, 1627-1637 (2001).

25. A. Fabregat *et al.*, The Reactome Pathway Knowledgebase. *Nucleic Acids Res* **46**, D649-D655 (2018).
26. D. E. Meyers, H. I. Basha, M. K. Koenig, Mitochondrial Cardiomyopathy Pathophysiology, Diagnosis, and Management. *Tex. Heart Inst. J.* **40**, 385-394 (2013).
27. C. M. Lawes, S. Vander Hoorn, A. Rodgers, H. International Society of, Global burden of blood-pressure-related disease, 2001. *Lancet (London, England)* **371**, 1513-1518 (2008).
28. M. Trudu *et al.*, Common noncoding UMOD gene variants induce salt-sensitive hypertension and kidney damage by increasing uromodulin expression. *Nat Med* **19**, 1655-1660 (2013).
29. E. Evangelou *et al.*, Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat Genet* **50**, 1412-1425 (2018).
30. T. M. Coffman, The inextricable role of the kidney in hypertension. *J Clin Invest* **124**, 2341-2347 (2014).
31. R. P. Lifton, A. G. Gharavi, D. S. Geller, Molecular mechanisms of human hypertension. *Cell* **104**, 545-556 (2001).
32. J. E. Wagenseil, R. P. Mecham, Vascular extracellular matrix and arterial mechanics. *Physiol Rev* **89**, 957-989 (2009).
33. J. M. Fernandez-Fernandez *et al.*, Gain-of-function mutation in the KCNMB1 potassium channel subunit is associated with low prevalence of diastolic hypertension. *J. Clin. Invest.* **113**, 1032-1039 (2004).
34. S. K. Michael *et al.*, High blood pressure arising from a defect in vascular function. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 6702-6707 (2008).
35. M. E. Ullian, The role of corticosteroids in the regulation of vascular tone. *Cardiovascular research* **41**, 55-64 (1999).
36. L. J. Ignarro, G. Cirino, A. Casini, C. Napoli, Nitric oxide as a signaling molecule in the vascular system: An overview. *J. Cardiovasc. Pharmacol.* **34**, 879-886 (1999).
37. S. Ozonoff, B. L. Goodlin-Jones, M. Solomon, Evidence-based assessment of autism spectrum disorders in children and adolescents. *J. Clin. Child Adolesc. Psychol.* **34**, 523-540 (2005).
38. S. De Rubeis *et al.*, Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209-215 (2014).
39. B. S. Abrahams, D. H. Geschwind, Advances in autism genetics: on the threshold of a new neurobiology. *Nat Rev Genet* **9**, 341-355 (2008).
40. J. D. Buxbaum *et al.*, Optimizing the phenotyping of rodent ASD models: enrichment analysis of mouse and human neurobiological phenotypes associated with high-risk autism genes identifies morphological, electrophysiological, neurological, and behavioral features. *Mol Autism* **3**, 1 (2012).
41. J. Li *et al.*, Integrated systems analysis reveals a molecular network underlying autism spectrum disorders. *Mol Syst Biol* **10**, 774 (2014).
42. L. M. Lyall *et al.*, Association of disrupted circadian rhythmicity with mood disorders, subjective wellbeing, and cognitive function: a cross-sectional study of 91 105 participants from the UK Biobank. *The lancet. Psychiatry* **5**, 507-514 (2018).
43. D. J. Burn, D. Bates, Neurology and the kidney. *J Neurol Neurosurg Psychiatry* **65**, 810-821 (1998).
44. G. Zsurka, W. S. Kunz, Mitochondrial dysfunction and seizures: the neuronal energy crisis. *Lancet Neurol* **14**, 956-966 (2015).
45. M. L. Zupanc, B. Legros, Progressive myoclonic epilepsy. *The Cerebellum* **3**, 156 (2004).
46. B. D. Peters *et al.*, Brain white matter development is associated with a human-specific haplotype increasing the synthesis of long chain fatty acids. *J Neurosci* **34**, 6367-6376 (2014).
47. A. K. McAllister, Major histocompatibility complex I in brain development and schizophrenia. *Biol Psychiatry* **75**, 262-268 (2014).
48. A. H. Miller, C. L. Raison, The role of inflammation in depression: from evolutionary imperative to modern treatment target. *Nat Rev Immunol* **16**, 22-34 (2016).

49. J. N. Paulson *et al.*, Tissue-aware RNA-Seq processing and normalization for heterogeneous and sparse data. *BMC Bioinformatics* **18**, 437 (2017).
50. P. Shannon *et al.*, Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504 (2003).
51. Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J R Stat Soc B* **57**, 289-300 (1995).

Acknowledgement

We thank the GTEx donors and their families for their contribution to science and the GTEx consortium for generating the resource. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The GTEx transcriptome data used for the analyses described in this manuscript were obtained from the GTEx Portal (http://www.gtexportal.org) on 12/05/2017. The work in Ma lab is supported in part by a grant from NSFC (31770268). The numerical calculations in this manuscript were conducted on the supercomputing systems in USTC Supercomputing Center and in USTC School of Life Sciences Bioinformatics Center.

Supplementary Materials

Table S1: The 166980 co-expressed gene pairs used for network construction

Table S2: The 652 gene modules identified from the network

Table S3: GO enrichment analysis results

Table S4: Disease genes enrichment analysis results

Table S5: MP enrichment analysis results

Figure S1: Gene expression heatmap for Modules #71, #30, and #18

Figure S2: The pathway for synthesizing cholesterol from acetyl-CoA

Figure S3: Gene expression heatmap for Modules #39, #85, and #92

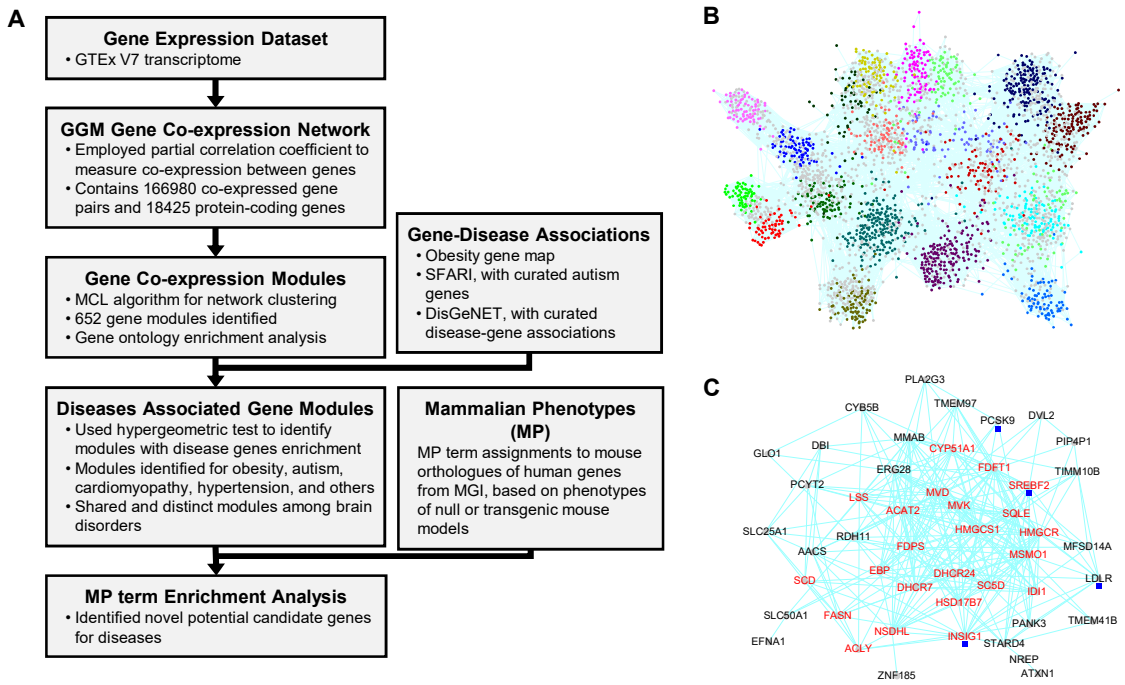
Figure S4: Modules #39 and #92 associated with cardiomyopathy

Figure S5: Gene expression heatmap for Modules #281 and #305

Figure S6: Modules #39, #54, and #101 associated with hypertension

Figure S7: Gene expression heatmap for Modules #11, #20, #35, #76, and #124

Figures



D

M#	Enriched GO	*P*	M#	Enriched GO	*P*	M#	Enriched GO	*P*
1	cilium	3.18E-51	60	apoptotic process	4.54E-06	186	cell-cell adhesion	1.44E-15
2	mitochondrion	8.3E-153	66	extracellular matrix	1.03E-54	187	synaptic signaling	5.26E-05
3	cell cycle	3.9E-135	67	cilium	1.67E-20	197	hyperosmotic salinity response	2.02E-04
6	immune response	4.74E-52	70	fatty acid oxidation	1.00E-27	220	spliceosomal snRNP complex	2.07E-06
7	chromatin organization	1.56E-23	71	cholesterol biosynthetic process	1.92E-42	229	regulation of synaptic plasticity	5.08E-04
8	immune response	5.19E-62	72	anion transmembrane transporter activity	2.11E-12	231	circadian regulation of gene expression	8.27E-07
9	axonogenesis	5.17E-17						
11	synaptic signaling	2.8E-34	76	synaptic signaling	9.39E-22	249	neurotransmitter binding	9.66E-03
14	myelination	1.66E-10	77	nervous system development	4.83E-05	264	MHC class II protein complex	6.70E-31
15	adaptive immune response	5.61E-68	83	MHC class I protein complex	3.16E-15	277	neuropeptide signaling pathway	7.00E-04
18	lipid storage	3.37E-12	85	muscle contraction	5.41E-20	281	ion transmembrane transporter activity	2.90E-04
20	synaptic signaling	1.21E-19	92	muscle contraction	3.69E-21			
22	mitochondrion	1.10E-33	93	synaptic signaling	3.50E-06	288	dopamine metabolic process	3.09E-08
27	lysosome	4.47E-37	95	inflammatory response	1.75E-20	305	sodium ion homeostasis	4.65E-03
28	ribosome biogenesis	1.31E-56	101	extracellular matrix	1.20E-31	358	neuron part	4.31E-03
29	protein folding	9.50E-33	105	serine-type endopeptidase activity	8.09E-10	374	mitochondrial ATP synthesis coupled electron transport	2.19E-20
34	circulatory system development	6.47E-14						
35	synaptic signaling	6.86E-12	113	ion channel activity	2.48E-05	396	regulation of vascular smooth muscle cell differentiation	9.31E-03
39	muscle contraction	6.48E-30	114	visual perception	4.18E-16			
50	brush border	3.03E-11	124	synaptic signaling	6.17E-05	442	inward rectifying potassium channel	1.49E-03
51	thyroid hormone generation	1.51E-08	126	cGMP metabolic process	7.50E-04			
54	glucocorticoid biosynthetic process	4.02E-09	131	extracellular matrix	1.59E-10	466	synapse	1.00E-02
			146	synaptic signaling	8.92E-09	520	fatty acid biosynthetic process	1.12E-08
56	muscle structure development	1.14E-19	154	alveolar lamellar body	8.24E-08	546	feeding behavior	6.44E-04

Fig. 1. Experimental design and overview of the identified modules. (A) The analysis pipeline. **(B)** A sub-network for the largest 20 modules. Nodes represent genes and co-expressed genes are connected by edges. Colors of nodes indicate module identities, except that nodes in grey color are those belong to multiple modules. **(C)** A gene module (#71) for cholesterol biosynthesis. Highlighted in red are cholesterol biosynthesis genes. Labeled with blue square are genes encoding cholesterol homeostasis regulators. **(D)** Enriched GO terms for selected modules discussed in the main text. M#, Module ids; *P*, BH-adjusted pValue for GO enrichment.

A

M#	Disease gene counts	*OR*	*BH Adjusted pValue*	FDR
Obestiy				
18	21	10.2	7.26E-13	<0.0001
288	6	19.8	3.25E-05	0.0001
60	9	7.8	1.54E-04	0.0001
101	9	6.1	7.48E-04	0.0007
34	11	4.9	7.48E-04	0.0007
546	4	21.0	8.44E-04	0.001
30	10	4.6	2.31E-03	0.0029
277	4	14.0	3.81E-03	0.0048
396	3	13.1	2.97E-02	0.0441
Dilated Cardiomyopathy (DCM)				
85	13	60.5	1.06E-17	<0.0001
92	10	42.4	2.68E-12	<0.0001
39	9	19.2	3.98E-08	<0.0001
70	5	22.4	4.55E-05	0.0001
Hypertrophic Cardiomyopathy (HCM)				
2	21	19.4	1.57E-17	<0.0001
85	8	30.9	9.82E-09	<0.0001
70	5	20.0	1.15E-04	0.0002
39	5	9.0	3.88E-03	0.0098
Hypertensive Disease (Hypertension)				
374	7	35.9	4.87E-08	<0.0001
39	13	7.4	3.02E-06	<0.0001
281	5	20.7	1.72E-04	0.0001
95	7	9.7	3.30E-04	0.0002
30	10	5.8	3.98E-04	0.0003
66	8	7.1	5.31E-04	0.0006
305	4	16.5	2.01E-03	0.0022
54	6	7.2	3.93E-03	0.0037
126	4	9.4	1.52E-02	0.0231
101	6	5.1	2.08E-02	0.0314
Autism				
7	42	5.9	1.35E-17	<0.0001
186	17	17.2	2.78E-16	<0.0001
77	13	6.2	1.26E-05	<0.0001
11	21	2.9	1.06E-03	0.0009
168	7	7.9	1.06E-03	0.0009
9	16	3.3	1.76E-03	0.0017
20	15	3.1	5.49E-03	0.0049
35	12	3.5	5.90E-03	0.0049
131	7	5.2	1.07E-02	0.0097
124	7	4.9	1.38E-02	0.0133
358	4	8.0	3.19E-02	0.0325
76	9	3.3	3.66E-02	0.0363
325	4	7.4	3.66E-02	0.0369
266	4	6.5	5.82E-02	0.0492

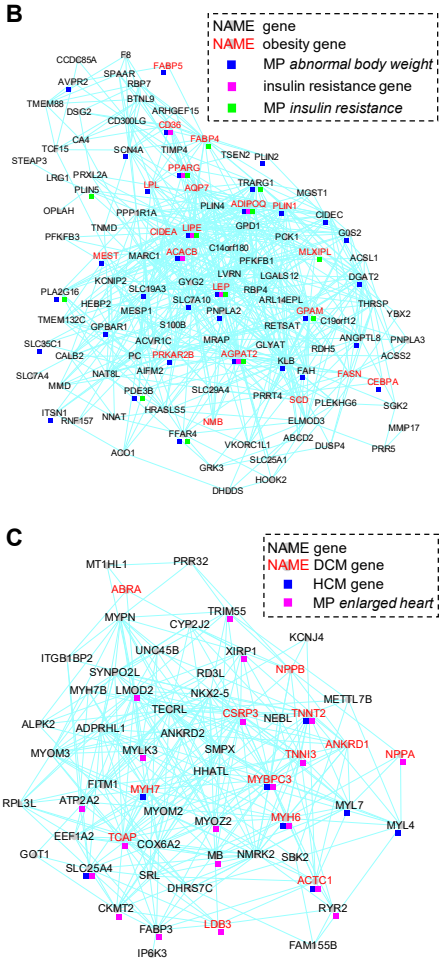


Fig. 2. Gene modules associated with diseases. (A) Modules associated with obesity, cardiomyopathy, hypertension, and autism. M#, module id. OR, odd ratio. **(B)** A module (#18) associated with obesity. Obesity genes, insulin resistance genes, and genes with MPs *abnormal body weight* and *insulin resistance* are indicated. **(C)** A module (#85) associated with dilated cardiomyopathy and hypertrophic cardiomyopathy.

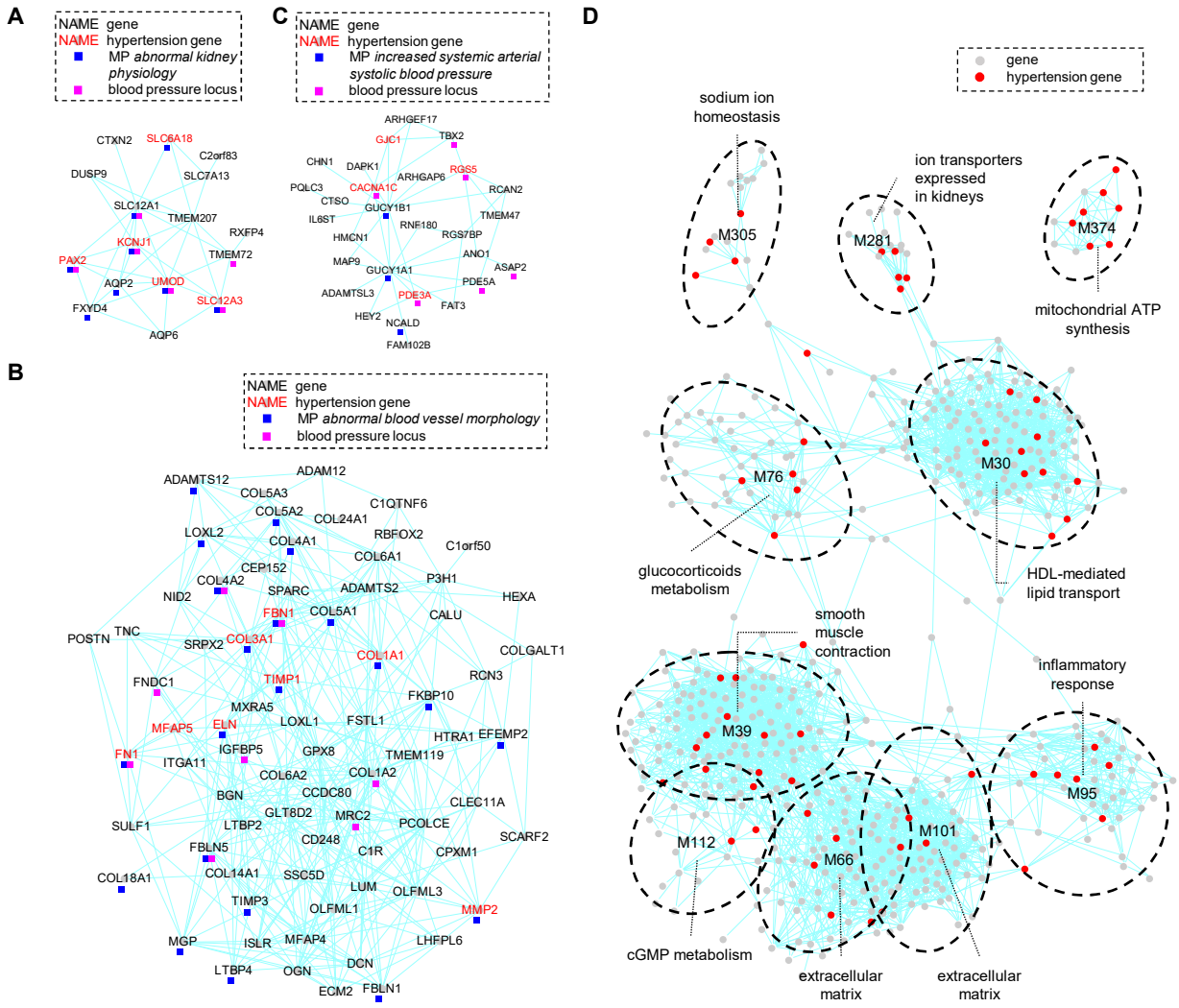
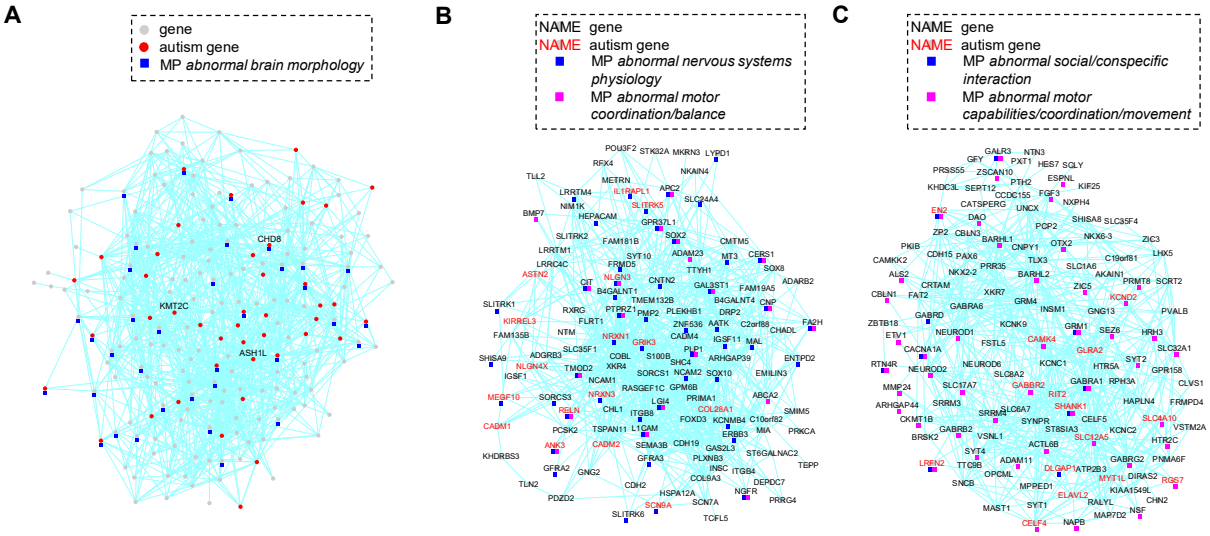


Fig. 3. Gene modules associated with hypertension. (A), **(B)**, and **(C)** Modules #281, #66, and #126 associated with hypertension, respectively. **(D)** A sub-network including all hypertension modules. Gene names are not labeled due to space limitation. Dashed circles outline the approximate positions of the modules.



D

Distinct Modules	Brain Disorder	Shared Modules							
2, 22, 27, 70, 281	epilepsy	7							
131, 168, 186, 266, 325	autism	7	11, 76	9, 20, 35, 124			77, 358		
14, 34, 83, 229, 264, 442, 466, 520, 529, 626	schizophrenia		11, 76	9, 20, 35, 124	93, 288	72, 187			197
113, 146, 249	bipolar disorder		11, 76	9, 20, 35, 124	93, 288	72, 187	77, 358	231	
95, 277, 374	depressive disorder		11, 76		93, 288			231	197

Fig. 4. Gene modules associated with autism and other brain disorders. (A), **(B)**, and **(C)** Modules #7, #9, and #20 associated with autism, respectively. Most gene names are not labeled in (A) due to space limitation. **(D)** Distinct and shared disease modules among epilepsy, autism, schizophrenia, bipolar disorder, and depressive disorder. Numbers indicate the module ids.