

Patterns, Volume 4

Supplemental information

**Self-supervised graph representation learning
integrates multiple molecular networks and
decodes gene-disease relationships**

Yi Wang, Zijun Sun, Qiushun He, Jiwei Li, Ming Ni, and Meng Yang

Supplementary Figures

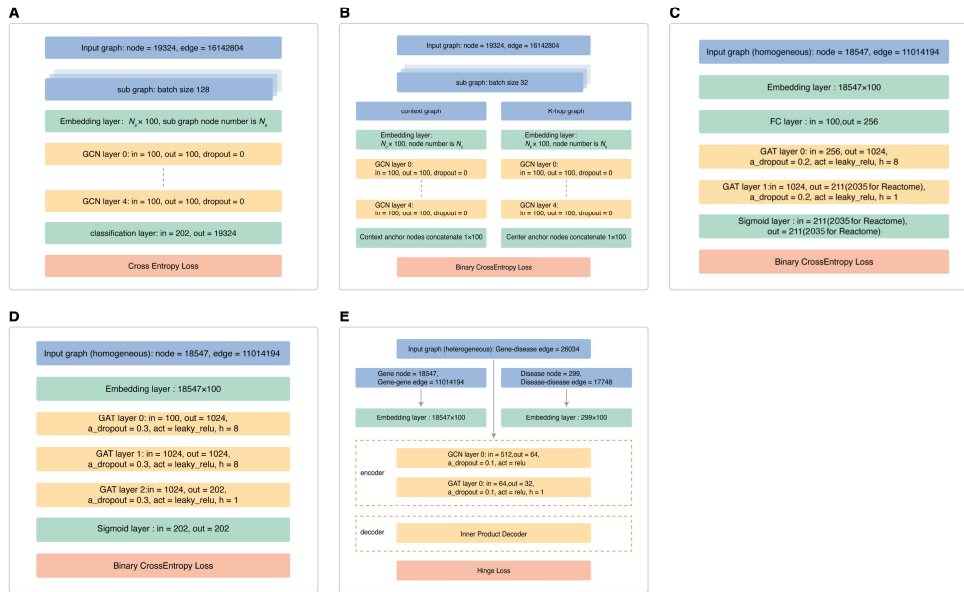


Figure S1. Schematic diagram of Graphene model architecture. (A) Pre-training for masked node recovery (B) Pre-training for context prediction (C) Re-training for gene set member identification task (D) Re-training for disease gene reprioritization task (E) Bipartite architecture for comorbidity prediction task.

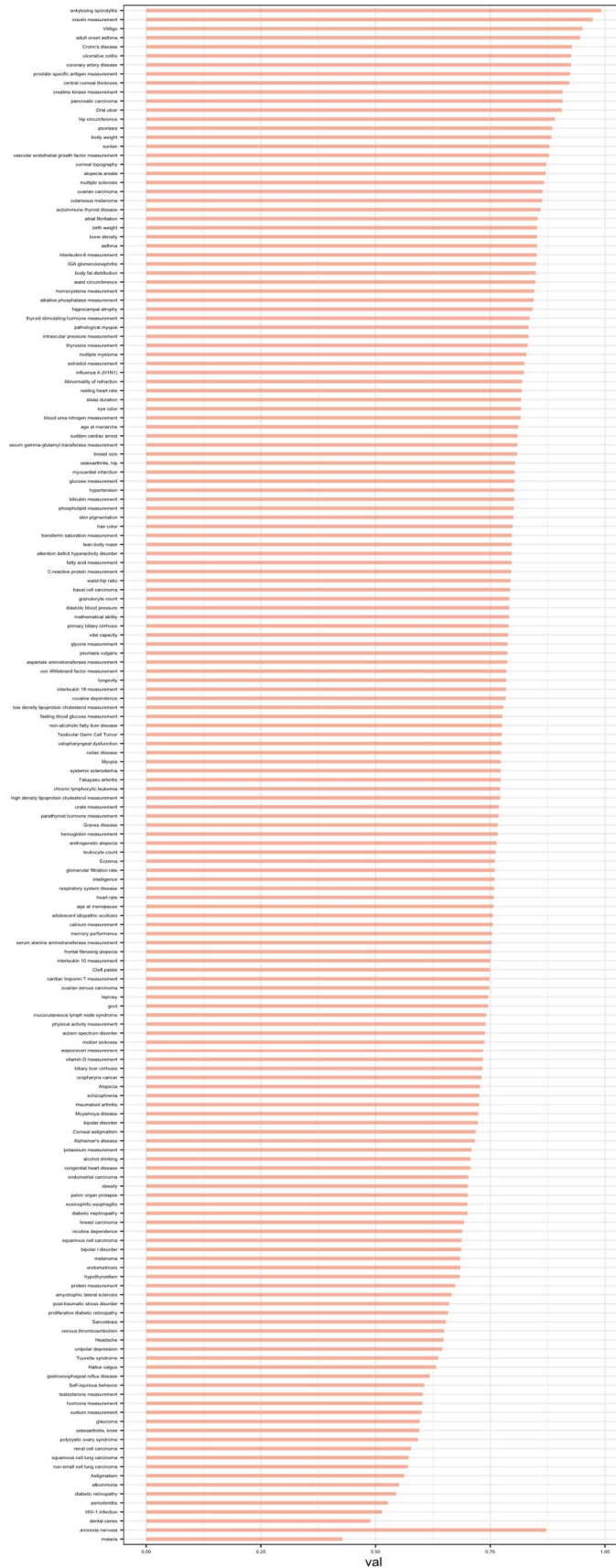


Figure S2. AUROC results for all 171 DisGeNET diseases. Diseases are represented by UMLS CUI ids.

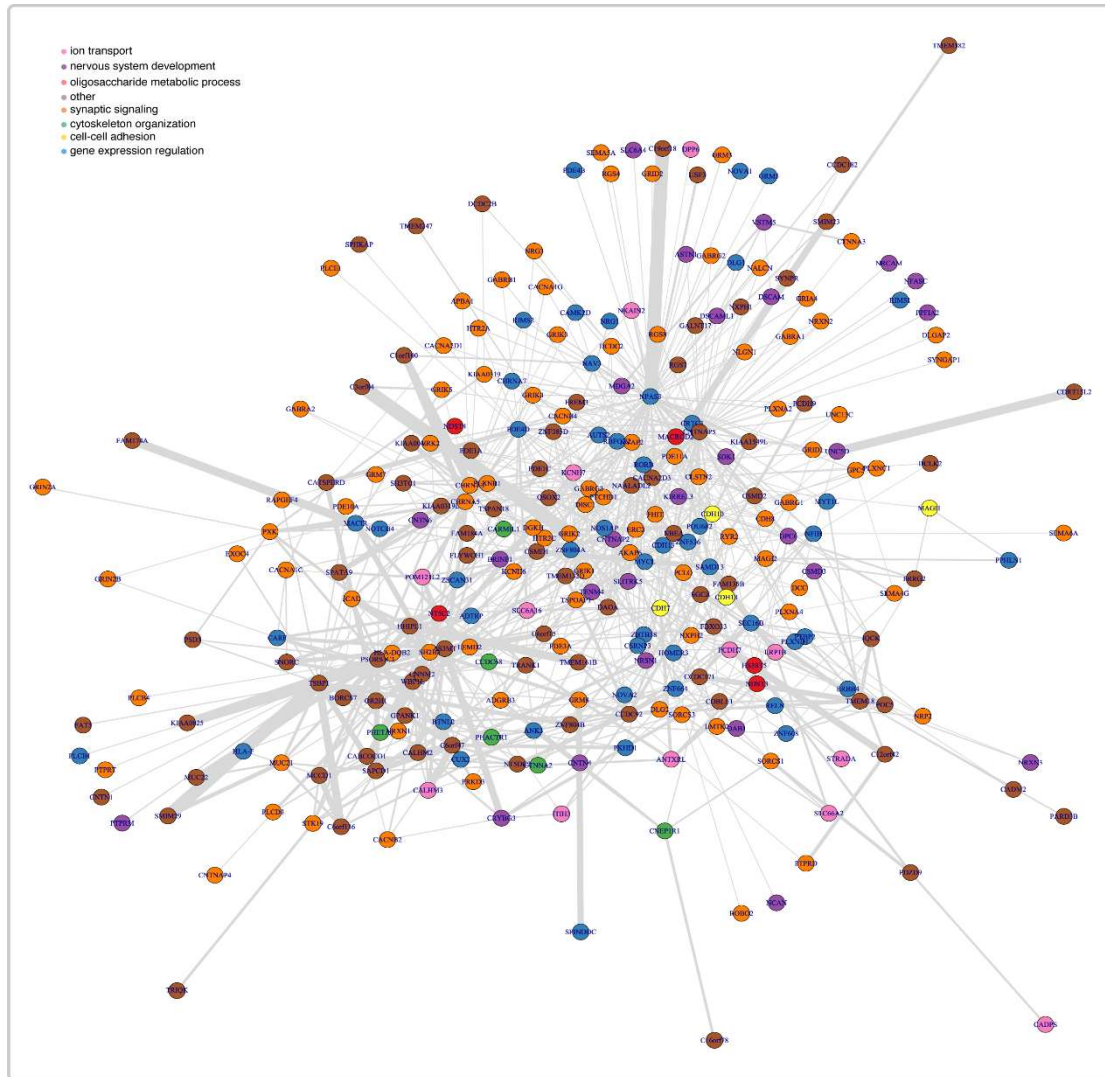


Figure S4. Attention weights showcase interaction patterns of 300 Graphene TPGs for SCZ.

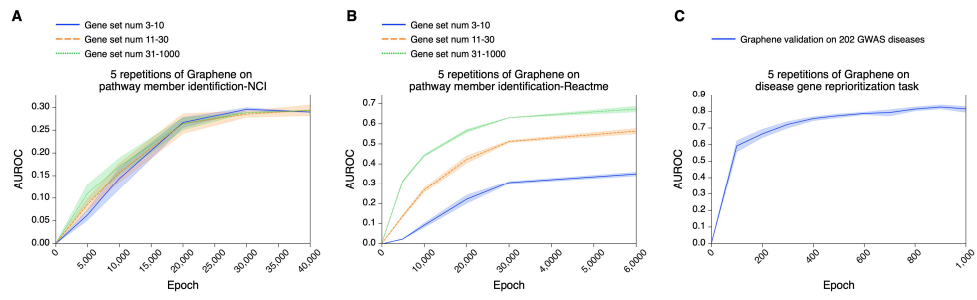


Figure S5. Evaluation with 5 repetitions of Graphene on pathway member identification task and disease gene re-prioritization task.

Supplementary Tables

Table S1. Datasets used in this study

Dataset	Download link
Humanbase	https://hb.flatironinstitute.org/download
HuRI	http://www.interactome-atlas.org/download
STRING9606	https://string-db.org
PCNet	http://www.ndexbio.org
NCI(PID)	http://www.ndexbio.org
Reactome	https://reactome.org/download-data
GWAS Catalog	https://www.ebi.ac.uk/gwas/docs/file-downloads
DisGeNET	https://www.disgenet.org/downloads
UK BioBank	https://pan.ukbb.broadinstitute.org/downloads

Table S2. Runtime required to obtain the results of Figure 2C, 2D in downstream re-training of Graphene and competing methods, evaluated on DisGeNET cross-dataset test (171 diseases) and UK Biobank cross-dataset test (81 diseases) by a single Quadro RTX 6000 GPU.

	GenePanda	Graphene	GWAS P-value	N2V	NAGA	Set2Gaussian
UK Biobank cross-dataset test	2,840 seconds	295 seconds	243 seconds	>2 days	2,358 seconds	1,615 seconds
DisGeNET cross-dataset test	5,960 seconds	306 seconds	513 seconds	>4 days	4,980 seconds	1,620 seconds

Table S3. Methods comparison for pathway gene set recovery

Dataset	Method	AUPRC		
		3-10	11-30	31-1000
NCI	Mean	0.30	0.22	0.19
	Set2Gaussian	0.31	0.27	0.26
	Graphene (random input embedding)	0.24	0.27	0.29
	Graphene (pre-trained input embedding)	0.29	0.31	0.29
Reactome	Mean	0.40	0.47	0.49
	Set2Gaussian	0.43	0.56	0.63
	Graphene (random input embedding)	0.29	0.52	0.67
	Graphene (pre-trained input embedding)	0.42	0.58	0.69

Table S4. Normalized Jaccard Index of GWAS hits for eight diseases

Disease1	Disease2	value
unipolar depression	post-traumatic stress disorder	0.0792
unipolar depression	attention deficit hyperactivity disorder	0.368
unipolar depression	bipolar disorder	0.324
unipolar depression	schizophrenia	0.863
unipolar depression	autism spectrum disorder	0.154
unipolar depression	Tourette syndrome	0.102
unipolar depression	anorexia nervosa	0.121
post-traumatic stress disorder	attention deficit hyperactivity disorder	0.122
post-traumatic stress disorder	bipolar disorder	0.0298
post-traumatic stress disorder	schizophrenia	0.0847
post-traumatic stress disorder	autism spectrum disorder	0
post-traumatic stress disorder	Tourette syndrome	0
post-traumatic stress disorder	anorexia nervosa	0.133
attention deficit hyperactivity disorder	bipolar disorder	0.141
attention deficit hyperactivity disorder	schizophrenia	0.267
attention deficit hyperactivity disorder	autism spectrum disorder	0.388
attention deficit hyperactivity disorder	Tourette syndrome	0.0428
attention deficit hyperactivity disorder	anorexia nervosa	0.126
bipolar disorder	schizophrenia	1
bipolar disorder	autism spectrum disorder	0.114
bipolar disorder	Tourette syndrome	0.0934
bipolar disorder	anorexia nervosa	0.185
schizophrenia	autism spectrum disorder	0.179
schizophrenia	Tourette syndrome	0.0490
schizophrenia	anorexia nervosa	0.0486
autism spectrum disorder	Tourette syndrome	0
autism spectrum disorder	anorexia nervosa	0.106
Tourette syndrome	anorexia nervosa	0.495

Table S5. Normalized Jaccard Index of DisGeNET genes for eight diseases

Disease1	Disease2	value
unipolar depression	post-traumatic stress disorder	0.464
unipolar depression	attention deficit hyperactivity disorder	0.115
unipolar depression	bipolar disorder	0.775
unipolar depression	schizophrenia	0.560
unipolar depression	autism spectrum disorder	0.0562
unipolar depression	Tourette syndrome	0.341
unipolar depression	anorexia nervosa	0.343
post-traumatic stress disorder	attention deficit hyperactivity disorder	0.0562
post-traumatic stress disorder	bipolar disorder	0.507
post-traumatic stress disorder	schizophrenia	0.446
post-traumatic stress disorder	autism spectrum disorder	0.0137
post-traumatic stress disorder	Tourette syndrome	0.297
post-traumatic stress disorder	anorexia nervosa	0.487
attention deficit hyperactivity disorder	bipolar disorder	0.0542
attention deficit hyperactivity disorder	schizophrenia	0.0258
attention deficit hyperactivity disorder	autism spectrum disorder	0
attention deficit hyperactivity disorder	Tourette syndrome	0.157
attention deficit hyperactivity disorder	anorexia nervosa	0.111
bipolar disorder	schizophrenia	1
bipolar disorder	autism spectrum disorder	0.0613
bipolar disorder	Tourette syndrome	0.307
bipolar disorder	anorexia nervosa	0.293
schizophrenia	autism spectrum disorder	0.116
schizophrenia	Tourette syndrome	0.258
schizophrenia	anorexia nervosa	0.200
autism spectrum disorder	Tourette syndrome	0.108
autism spectrum disorder	anorexia nervosa	0.0413
Tourette syndrome	anorexia nervosa	0.329

Table S6. Normalized Jaccard Index of Graphene TPGs for eight diseases

Disease1	Disease2	value
unipolar depression	post-traumatic stress disorder	0.455
unipolar depression	attention deficit hyperactivity disorder	0.479
unipolar depression	bipolar disorder	0.938
unipolar depression	schizophrenia	1
unipolar depression	autism spectrum disorder	0.536
unipolar depression	Tourette syndrome	0.0203
unipolar depression	anorexia nervosa	0.346
post-traumatic stress disorder	attention deficit hyperactivity disorder	0.407
post-traumatic stress disorder	bipolar disorder	0.596
post-traumatic stress disorder	schizophrenia	0.452
post-traumatic stress disorder	autism spectrum disorder	0.937
post-traumatic stress disorder	Tourette syndrome	0.363
post-traumatic stress disorder	anorexia nervosa	0.315
attention deficit hyperactivity disorder	bipolar disorder	0.455
attention deficit hyperactivity disorder	schizophrenia	0.458
attention deficit hyperactivity disorder	autism spectrum disorder	0.406
attention deficit hyperactivity disorder	Tourette syndrome	0.0662
attention deficit hyperactivity disorder	anorexia nervosa	0.504
bipolar disorder	schizophrenia	0.955
bipolar disorder	autism spectrum disorder	0.405
bipolar disorder	Tourette syndrome	0
bipolar disorder	anorexia nervosa	0.372
schizophrenia	autism spectrum disorder	0.343
schizophrenia	Tourette syndrome	0.0271
schizophrenia	anorexia nervosa	0.309
autism spectrum disorder	Tourette syndrome	0.404
autism spectrum disorder	anorexia nervosa	0.439
Tourette syndrome	anorexia nervosa	0.0324

Table S7. Normalized Jaccard Index of NAGA rankings for eight diseases

Disease1	Disease2	value
unipolar depression	post-traumatic stress disorder	0.09
unipolar depression	attention deficit hyperactivity disorder	0.10
unipolar depression	bipolar disorder	0.36
unipolar depression	schizophrenia	1.00
unipolar depression	autism spectrum disorder	0.26
unipolar depression	Tourette syndrome	0.00
unipolar depression	anorexia nervosa	0.15
post-traumatic stress disorder	attention deficit hyperactivity disorder	0.01
post-traumatic stress disorder	bipolar disorder	0.05
post-traumatic stress disorder	schizophrenia	0.05
post-traumatic stress disorder	autism spectrum disorder	0.09
post-traumatic stress disorder	Tourette syndrome	0.07
post-traumatic stress disorder	anorexia nervosa	0.07
attention deficit hyperactivity disorder	bipolar disorder	0.06
attention deficit hyperactivity disorder	schizophrenia	0.06
attention deficit hyperactivity disorder	autism spectrum disorder	0.14
attention deficit hyperactivity disorder	Tourette syndrome	0.04
attention deficit hyperactivity disorder	anorexia nervosa	0.01
bipolar disorder	schizophrenia	0.95
bipolar disorder	autism spectrum disorder	0.03
bipolar disorder	Tourette syndrome	0.03
bipolar disorder	anorexia nervosa	0.05
schizophrenia	autism spectrum disorder	0.11
schizophrenia	Tourette syndrome	0.04
schizophrenia	anorexia nervosa	0.05
autism spectrum disorder	Tourette syndrome	0.03
autism spectrum disorder	anorexia nervosa	0.52
Tourette syndrome	anorexia nervosa	0.17

Table S8. cross-trait LD score for eight diseases

Disease1	Disease2	value
unipolar depression	attention deficit hyperactivity disorder	0.629
unipolar depression	anorexia nervosa	0.400
unipolar depression	autism spectrum disorder	0.486
unipolar depression	Tourette syndrome	0.329
post-traumatic stress disorder	unipolar depression	0
post-traumatic stress disorder	attention deficit hyperactivity disorder	0
post-traumatic stress disorder	bipolar disorder	0
post-traumatic stress disorder	schizophrenia	0
post-traumatic stress disorder	autism spectrum disorder	0
post-traumatic stress disorder	Tourette syndrome	0
post-traumatic stress disorder	anorexia nervosa	0
attention deficit hyperactivity disorder	anorexia nervosa	0.0143
attention deficit hyperactivity disorder	autism spectrum disorder	0.529
attention deficit hyperactivity disorder	Tourette syndrome	0.271
bipolar disorder	unipolar depression	0.471
bipolar disorder	attention deficit hyperactivity disorder	0.257
bipolar disorder	anorexia nervosa	0.143
bipolar disorder	autism spectrum disorder	0.243
bipolar disorder	Tourette syndrome	0.114
schizophrenia	bipolar disorder	1
schizophrenia	unipolar depression	0.443
schizophrenia	attention deficit hyperactivity disorder	0.229
schizophrenia	anorexia nervosa	0.371
schizophrenia	autism spectrum disorder	0.357
schizophrenia	Tourette syndrome	0.157
autism spectrum disorder	Tourette syndrome	0.229
Tourette syndrome	anorexia nervosa	0.114
anorexia nervosa	autism spectrum disorder	0.157

Table S9. Spearman correlation of the four methods (Graphene, DisGeNet, NAGA, GWAS) against ct-LDSC.

Methods	Graphene	DisGeNet	NAGA	GWAS
Spearman correlation	0.767	0.596	0.770	0.731

Table S10. Overlapping TPGs identified by Graphene for 4 disease pairs.

Disease Pair	Overlapping genes
BIP vs MDD	PDE1A,GRID1,GRIA4,GRID2,KCNC2,PLCH1,MDGA2,GABRB2,FAM104A,CARMIL1,GRM3,RAPGEF4,NRXN1,GRM5,DGKH,NCAN,ANTXR1,CTNNA2,TMPRSS7,GABRG3,GABRB3,CNTNAP2,CNTN6,SCN8A,RIMS1,PDE1C,GRM1,NYAP2,MAGI2,BTNL2,ZNF804A,PDE11A,FHIT,CBLN4,RBFOX1,DAOA,DISC1,GRIK2,GRIK4,CACNA1E,GRM8,GRIA1,KCNA4,SGCZ,NXPH1,NBEA,SAMD13,NBPF3,CSRNP3,TMEM247,KCNB2,MTUS2,MACROD2,CNTNAP5,CLSTN2,TMEM108,CSMD1,DAB1,KCNQ3,CPNE4,FAM135B,DLG2,RAPGEF5,GABRG2,GRM7,CDH8,RGS7,STPG3,AKAP6,KCND2,CACNB4,ERC2,GABRA2,CADM2,CACNA1C,CACNA2D3,CSMD3,TMEM161B,KCNV1,KCNH7,RIMS2,LRP1B,GRIK1,RESF1,AUTS2,SDK1,CACNA2D1,NKAIN2,PTPRT,CNTNAP4,OPCML,PCLO,CNTN1,NAV3,TRANK1,SPINDOC,PTPRM,THSD7B,SMIM22,CFAP54,NRG3,KCNH6,NALCN,PTPRD,NXPH2,SCN1A,SCN2A,KCND3,KCNC1,ZNF664,PCDH7,GABRB1,NPAS3,CDH13,CSMD2,VSTM5,GRIN2A,NAALADL2,KCNB1,RGS8,NRXN3,NELL1
SCZ vs MDD	PDE1A,CADPS,GRID1,GRIA2,MUC21,GRIA4,GRID2,FREM3,PLCH1,SMIM29,HLA-G,NOVA2,MDGA2,GABRB2,SNAP91,CARMIL1,GRM3,NRXN1,MAGI1,RAPGEF4,GRM5,DGKH,NCAN,ANTXR1,CTNNA2,SORCS3,FAM174A,FAT3,GABRB3,GABRG3,CNTNAP2,PTPRK,SCN8A,NETO1,CNTN6,TSPAN18,CTNND2,POU6F2,RIMS1,PDE1C,GPC6,GRM1,NYAP2,MAGI2,UNC13C,BTNL2,CDH9,ROBO2,ZNF804A,ASIC2,ZNF536,RAB3C,PDE11A,FHIT,RBFOX1,GRIN2B,DAOA,DISC1,GRIK2,GRIK4,GRIA1,CACNA1E,GRM8,CNTN4,ANKS1B,SGCZ,PPFIA2,CTNNA3,NXPH1,NBEA,SAMD13,UNC5D,TMEM247,MTUS2,CSRNP3,MACROD2,NRG1,CNTNAP5,JCAD,RELN,CLSTN2,SAPCD1,TRIM26,SYNPR,TMEM108,LMTK2,PCDH9,CSMD1,DAB1,C6orf47,CPNE4,FAM135B,DLG2,RAPGEF5,GABRG2,GRM7,CDH8,ASTN1,ZNF385D,STUM,PLXNC1,RGS7,DSCAML1,SEMA6D,DLG1,STPG3,RBFOX2,AKAP6,NPIP5,NDST4,CACNB4,ERC2,GABRA2,CADM2,SNORC,TMEM161B,CSMD3,CACNA1C,PCDH17,CACNA2D3,ROBO1,KCNH7,NOVA1,RIMS2,LRP1B,GRIK1,CNTN5,RESF1,SDK1,C3orf84,AUTS2,CACNA2D1,CNEPIR1,NKAIN2,PTPRT,OPCML,CNTNAP4,PCLO,DSCAM,CNTN1,GPC5,NAV3,PRRG2,TRANK1,CDRT15L2,DPP6,NRCAM,SPINDOC,PTPRM,NRXN2,SMIM22,SEMA5A,GALNT17,KCNQ5,NRG3,CBLN2,KCNH6,NALCN,PTPRD,NXPH2,CDH7,SCN1A,SCN2A,DCC,QSOX2,CDH18,HOMER3,ZNF664,PCDH7,GABRB1,NLGN1,NFASC,NRXN3,NPAS3,CDH13,CSMD2,GRIK3,VSTM5,GRIN2A,NAALADL2,RGS8,CDH10,C6orf136,UNC5C,NELL1,NCAM2
MDD vs ADHD	FHIT,CADPS,RBFOX1,DAOA,GRM7,SPINDOC,CDH8,PTPRM,ZNF385D,GRID2,THSD7B,SMIM22,CFAP54,SEMA5A,NRG3,SGCZ,SEMA6D,PCDHGA1,AKAP6,NRXN1,NXPH1,CTNNA3,NPIP5,SAMD13,ERC2,UNC5D,NALCN,PTPRD,CADM2,CTNNA2,MACROD2,CNTNAP5,RELN,DCC,CLSTN2,CACNA2D3,GABRG3,CNTNAP2,PCDH7,NLGN1,ROBO1,CTNND2,NPAS3,GPC6,LRP1B,CDH13,PCDH9,CSMD1,GRIK1,CSMD2,NYAP2,SDK1,C3orf84,CPNE4,ROBO2,CDH9,CACNA2D1,NAALADL2,FAM135B,NRXN3,UNC5C,GPC5,NELL1,CDH10,NAV3
PTSD vs ASD	CBLN4,CSMD1,NCAN,FAM135B,CNTNAP2,SGCZ,SAMD13,TSBP1,NRXN2,LRP1B,SMIM22,GPC5,CNTNAP5,NRXN3,GRID1,MACROD2,GALNT17,HS3ST3A1,EXTL2,GPC6,HS3ST1,DAOA,HS2ST1,HS3ST3B1,NRXN1,HS3ST5,NDST4,RBFOX1,NDST2,CADPS,NKAIN2,PCDH9,GLCE,NDST3,NLGN1,HS3ST2,CNTNAP4,BRINP1,ARSK,GRID2,ERC2,FAM20B,UST,CADM2,GPC2,CACNA2D3,MACIR

Table S11. Non-overlapping CC-GWAS genes identified by Graphene for 9 disease pairs.

Disease Pair	Graphene prediction of CC-GWAS Gene(symbol)
ANO vs TS	POU3F2
BIP vs ADHD	TRANK1, SCN2A
SCZ vs ADHD	TRANK1
SCZ vs MDD	ZNF536, GRM3, PTBP2, DGKI, TRANK1, AS3MT, SATB2, ZNF804A, SDCCAG8
SCZ vs ANO	SPHKAP, IMMP2L, SEMA6D(not in training set), GRIA1, DLG2, CHRNA3, CNNM2
SCZ vs ASD	CCDC68, NEGR1, TRANK1, AS3MT, SATB2, DRD2
SCZ vs BIP	MAF (not in training set)
SCZ vs MDD	KCNV1(not in training set), IMMP2L, HS3ST5(not in training set), DGKI, CACNB2, PRKD1, CNNM2, SATB2, SNX19, PSD3, PCDHA7(not in training set), CHRNA3, ERBB4
SCZ vs TS	NFIB (not in training set), ZNF536, POU3F2(not in training set)