**Supplementary Material**

**Review**

**Genomic and Neuroimaging Approaches to Bipolar Disorder**

**Table S1. Glossary of key terms**

|  |  |
| --- | --- |
| **Term** | **Definition** |
| Common genetic variant | Alternative forms of a gene that are present with a minor allele frequency (MAF) of more than 1% in the population (1) |
| Rare genetic variant | Alternative forms of a gene that are present with a minor allele frequency (MAF) of less than 1% in the population (1) |
| Epigenetics | The study of molecular processes that influence the flow of information between a constant DNA sequence and variable gene expression patterns. This includes investigation of nuclear organization, DNA methylation, histone modification and RNA transcription (1) |
| Complex disorder | A disease that is caused by the interaction of multiple genes and environmental factors. It is also called multifactorial disorder (2) |
| Genetic association study | A study that aims to test whether a given sequence, such as a region of a chromosome, a haplotype, a gene, or an allele, has involvement in controlling the phenotype of a specific trait, metabolic pathway, or disease. This usually involves comparing the frequencies between two groups of individuals (often diseased subjects versus healthy controls) (1) |
| Genetic linkage study | A study that is a family-based method used to map a trait to a genomic location by demonstrating co-segregation of the disease with genetic markers of known chromosomal location; locations identified are more likely to contain a causal genetic variant. This technique is particularly useful for the identification of disorders that are inherited in a Mendelian fashion (1) |
| **Candidate gene** | A gene whose chromosomal location is associated with a particular disease or other phenotype. Because of its location, the gene is suspected of causing the disease or other phenotype. Therefore, candidate gene approach is in contrast to GWAS, which scan the entire genome for common genetic variation in a hypothesis-free mode (2) |
| **Genome-wide association studies (GWASs)** | Unbiased genome screens of unrelated individuals and appropriately matched controls to establish whether any genetic variant is associated with a trait. These studies typically focus on associations between single-nucleotide polymorphisms (SNPs) and major diseases (1) |
| **Genotype-by-environment (G×E)** **interaction** | An influence on the expression of a trait that results from the interplay between genes and the environment. Some traits are strongly influenced by genes, while other traits are strongly influenced by the environment. Most traits, however, are influenced by one or more genes interacting in complex ways with the environment (2) |
| **Genetic loci** | A locus (plural: loci) is the specific physical location of a gene or other DNA sequence on a chromosome, like a genetic street address (2) |
| **Haplotype** | A set of DNA variations, or polymorphisms, that tend to be inherited together. A haplotype can refer to a combination of alleles or to a set of single nucleotide polymorphisms (SNPs) found on the same chromosome. Information about haplotypes is being collected by the International HapMap Project and is used to investigate the influence of genes on disease (2) |
| **HapMap** | HapMap (short for "haplotype map") is the nickname of the International HapMap Project, an international project that seeks to relate variations in human DNA sequences with genes associated with health. The HapMap describes common patterns of genetic variation among people (2) |
| **Single nucleotide polymorphism (SNP)** | A type of polymorphism involving variation of a single base pair. Scientists are studying how single nucleotide polymorphisms, or SNPs (pronounced "snips"), in the human genome correlate with disease, drug response, and other phenotypes (2) |
| **Noncoding regions of the genome** | Non-coding DNA sequences do not code for amino acids. Most non-coding DNA lies between genes on the chromosome and has no known function. Other non-coding DNA, called introns, is found within genes. Some non-coding DNA plays a role in the regulation of gene expression (2) |
| **Genomic control (GC) approach** | Due to population substructure or cryptic relatedness, false associations might occur in case-control studies, increasing the variance of the trend test. The genomic control (GC) approach is used to calculate the variance inflation factor and adjust the test statistic (3) |
| **Linkage disequilibrium (LD)** | A non-random association of alleles at two or more loci in a given population (4) |
| **LOD score** | "Logarithm of the odds"; is a statistical estimate of whether two genes are likely to be located near each other on a chromosome and are therefore likely to be inherited. A LOD score of 3 or higher is generally understood to mean that two genes are located close to each other on the chromosome(2) |
| **Genome-wide correction** | In GWASs, where a large numbers of SNPs is analyzed, a large number of tests are performed. In order to correct these results for multiple testing (type-I error), it has become a standard in the field to use a P-value threshold of 5x10-8. SNPs that survive a genome-wide correction in a GWAS are considered to be very strong candidates |
| Whole exome sequencing | A laboratory process that is used to determine the nucleotide sequence primarily of the exonic (or protein-coding) regions of an individual’s genome and related sequences, representing approximately 1% of the complete DNA sequence (5) |
| Whole genome sequencing | A laboratory process that is used to determine nearly all of the approximately 3 billion nucleotides of an individual’s complete DNA sequence, including non-coding sequence (5) |
| **Exon** | The portion of a gene that codes for amino acids. The parts of the gene sequence that are expressed in the protein are called exons, because they are expressed, while the parts of the gene sequence that are not expressed in the protein are called introns, because they come in between--or interfere with--the exons (2) |
| **Polygenic risk scores (PRS)** | Also known as polygenic scores and genetic risk scores, represent the individual genetic burden for a particular disease/trait. They are calculated adding up the effects of hundreds-to-thousands of genetic variants |
| **Magnetic resonance imaging (MRI)** | A non-invasive biomedical imaging technique that uses a strong oscillating magnetic field to induce endogenous atoms such as hydrogen, or exogenously added contrast agents, to emit radiowaves that are detected and used to generate two and three-dimensional images of a living subject in the MRI scanner (1) |
| **Endophenotypes** | Measurable components unseen by the unaided eye along the pathway between disease and distal genotype. They may be neurophysiological, biochemical, endocrinological, neuroanatomical, cognitive, or neuropsychological in nature (6) |
| **Pleiotropy** | The phenomenon of one gene or one mutation affecting multiple traits (1) |
| **Bipolar disorder type I** | At least one episode of full-blown mania or mixed episode (manic and depressive symptoms), usually has at least one depressive episode |
| **Bipolar disorder type II** | Several protracted depressive episodes and at least one hypomanic episode, but no full-blown manic episodes |
| **Identity-by-descent (IBD)** | The shared inheritance of an identical portion of the genome between two individuals (7) |

**Table S2. Genes Abbreviation** (8-10)

|  |  |
| --- | --- |
| ***Gene Abbreviation*** | **Spelled out Name** |
| ***ADCY2*** | adenylate cyclase 2 |
| ***ADD3*** | adducin 3 |
| ***ADGB*** | androglobin |
| ***AHNAK*** | AHNAK nucleoprotein |
| ***ANK3*** | ankyrin 3 or G |
| ***ANKS4B*** | ankyrin repeat and sterile alpha motif domain containing 4B |
| ***AOAH*** | acyloxyacyl hydrolase |
| ***APOC3*** | apolipoprotein C3 |
| ***ARHGAP9*** | Rho GTPase activating protein 9 |
| ***ASTN2*** | astrotactin 2 |
| ***ATR*** | ATR serine/threonine kinase |
| ***BCL2L10*** | B-cell lymphoma 2 like 10 |
| ***BIG-2*** | contactin 4 |
| ***BOD1*** | biorientation of chromosomes in cell division 1 |
| ***CACNA1C*** | calcium voltage-gated channel subunit alpha 1C |
| ***CACNA1D*** | calcium voltage-gated channel subunit alpha1D |
| ***CAND2*** | Cullin associated and neddylation dissociated 2 (Putative) |
| ***CAP2*** | cyclase‐associated actin cytoskeleton regulatory protein 2 |
| ***CAPN2*** | calpain 2 |
| ***CDH23*** | cadherin related 23 |
| ***CDHR1*** | cadherin related family member 1 |
| ***CNNM4*** | cyclin and CBS domain divalent metal cation transport mediator 4 |
| ***CNTN6*** | contactin 6 |
| ***COL3A1*** | collagen type III alpha 1 chain |
| ***CPG2*** | candidate plasticity gene 2 |
| ***CRHR2*** | corticotropin-releasing hormone receptor 2 |
| ***CSNK1G3*** | casein kinase 1 gamma 3 |
| ***DCAF5*** | DDB1 and CUL4 associated factor 5 |
| ***DCTN5*** | dynactin subunit 5 |
| ***DDN*** | dendrin |
| ***DGKH*** | diacylglycerol kinase eta |
| ***DHH*** | desert hedgehog homolog |
| ***DIDO1*** | death inducer-obliterator 1 |
| ***DISC1*** | disrupted in schizophrenia 1 |
| ***DNAH7*** | dynein axonemal heavy chain 7 |
| ***DOCK5*** | dedicator of cytokinesis 5 |
| ***DRD5*** | D(1B) dopamine receptor |
| ***DUSP6*** | dual specificity phosphatase 6 |
| ***DYDC2*** | [DPY30 domain containing 2](https://www.ncbi.nlm.nih.gov/gene/84332) |
| ***EHD1*** | EH domain containing 1 |
| ***ELAVL2*** | ELAV-like RNA binding protein 2 |
| ***ERBB2*** | Erb-B2 receptor tyrosine kinase 2 |
| ***EVC*** | EvC ciliary complex subunit 1 |
| ***FADS1*** | fatty acid desaturase 1 |
| ***FADS2*** | fatty acid desaturase 2 |
| ***FADS3*** | fatty acid desaturase 3 |
| ***FAM71B*** | family with sequence similarity 71 member B |
| ***FER1L5*** | Fer-1 like family member 5 |
| ***GABRR1*** | gamma-aminobutyric acid type A receptor subunit Rho1 |
| ***GABRR2*** | gamma-aminobutyric acid type A receptor subunit Rho2 |
| ***GALNT13*** | polypeptide n-acetyl galactosaminyl transferase 13 |
| ***GGA3*** | golgi-associated gamma adaptin ear containing ARF binding protein 3 |
| ***GHITM*** | growth hormone inducible transmembrane protein |
| ***GRID1*** | glutamate ionotropic receptor delta type subunit 1 |
| ***GRIN2A*** | glutamate ionotropic receptor NMDA type subunit 2A |
| ***GRM1*** | metabotropic glutamate receptor 1 |
| ***GRM1*** | glutamate metabotropic receptor 1 |
| ***GRM4*** | metabotropic glutamate receptor 4 |
| ***HELLS*** | helicase, lymphoid specific |
| ***HIP1R*** | Huntingtin-interacting protein 1-related |
| ***HTR1B*** | 5-hydroxytryptamine receptor 1B |
| ***IRS4*** | insulin receptor substrate 4 |
| ***ITIH1*** | inter-alpha-trypsin inhibitor heavy chain 1 |
| ***ITIH3*** | inter-alpha-trypsin inhibitor heavy chain 3 |
| ***KANK4*** | KN motif and ankyrin repeat domains 4 |
| ***KLF4*** | kruppel like factor 4 |
| ***KMT2C*** | lysine methyltransferase 2C |
| ***LAMB1*** | laminin subunit beta 1 |
| ***LBA1*** | tetratricopeptide repeat and ankyrin repeat containing 1 |
| ***LMAN2L*** | lectin, mannose-binding 2-like |
| ***LOC100505933*** | ADD3 antisense RNA 1 |
| ***MACF1*** | microtubule actin crosslinking factor 1 |
| ***MAD1L1*** | mitotic arrest deficient 1 Like 1 |
| ***MDN1*** | midasin AAA ATPase 1 |
| ***MHC*** | major histocompatibility complex |
| ***MINPP1*** | multiple inositol-polyphosphate phosphatase 1 |
| ***MIR2113*** | microRNA 2113 |
| ***MKL2*** | myocardin related transcription factor B |
| ***MYH7*** | myosin heavy chain 7 |
| ***MYO10*** | myosin X |
| ***MYO5B*** | myosin VB |
| ***NCAN*** | neurocan |
| ***NCKAP5*** | NCK associated protein 5 |
| ***NCL*** | nucleolin |
| ***NDUFAB1*** | NADH:ubiquinone oxidoreductase subunit AB1 |
| ***NEK3*** | NIMA related kinase 3 |
| ***NF1*** | neurofibromin type 1 |
| ***NFIX*** | nuclear family I/X |
| ***NIPBL*** | NIPBL cohesin loading factor |
| ***NMUR2*** | neuromedin U receptor 2 |
| ***NOS1*** | nitric oxide synthase 1 |
| ***NRBF2*** | nuclear receptor binding factor 2 |
| ***NTRK1*** | neurotrophic receptor tyrosine kinase 1 |
| ***ODZ2*** | teneurin transmembrane protein 2 |
| ***ODZ4*** | teneurin transmembrane protein 4 |
| ***OLFML2B*** | olfactomedin like 2B |
| ***PALB2*** | partner and localizer of BRCA2 |
| ***PCDH15*** | protocadherin-related 15 |
| ***PCDHA*** | protocadherin alpha |
| ***PDE10A*** | phosphodiesterase 10A |
| ***PIK3C2A*** | phosphatidylinositol-4-phosphate 3-kinase catalytic subunit type 2 alpha |
| ***PITPNM2*** | phosphatidylinositol transfer protein membrane associated 2 |
| ***PKHD1L1*** | PKHD1 Like 1 |
| ***PLCXD3*** | phosphatidylinositol specific phospholipase C X domain containing 3 |
| ***PLET1*** | placenta expressed transcript 1 |
| ***PNLDC1*** | poly(A)-specific ribonuclease (PARN)-like domain containing 1 |
| ***POLN*** | DNA polymerase Nu |
| ***POU3F2*** | POU class 3 homeobox 2 |
| ***PRODH*** | proline dehydrogenase 1 |
| ***PTGFR*** | prostaglandin F receptor |
| ***PTS*** | 6-pyruvoyl tetrahydropterin synthase |
| ***RBM12*** | RNA-binding-motif protein 12 |
| ***RCCD1*** | RCC1 domain containing 1 |
| ***RGS12*** | regulator of G protein signaling 12 |
| ***RHEBL1*** | ras homolog enriched in brain like 1 |
| ***RIMS1*** | regulating synaptic membrane exocytosis 1 |
| ***SCN2A*** | sodium voltage-gated channel alpha subunit 2 |
| ***SELENOO*** | selenoprotein O |
| ***SERPING1*** | serpin family G member 1 |
| ***SLC22A23*** | solute carrier family 22 member 23 |
| ***SLC26A9*** | solute carrier family 26 member 9 |
| ***SLC4A1*** | solute carrier family 4 member 1 |
| ***SMARCC2*** | SWI/SNF Related, matrix associated, actin dependent regulator of chromatin subfamily C member 2 |
| ***SNTG2*** | syntrophin gamma 2 |
| ***ST6GALNAC4*** | ST6 n-acetylgalactosaminide alpha-2,6-sialyltransferase 4 |
| ***SYNE1*** | spectrin repeat containing nuclear envelope protein 1 |
| ***TCF7L1*** | transcription factor 7 like 1 |
| ***THSD7A*** | thrombospondin Type 1 domain containing 7A |
| ***THYN1*** | thymocyte nuclear protein 1 |
| ***TMEM220*** | transmembrane protein 220 |
| ***TRANK1*** | tetratricopeptide repeat and ankyrin repeat containing 1 |
| ***TRPC4AP*** | transient receptor potential cation channel subfamily C member 4 associated protein |
| ***TTC15*** | trafficking protein particle complex subunit 12 |
| ***TUBB1*** | tubulin beta 1 class VI |
| ***UNC13B*** | Unc-13 homolog B |
| ***UPF2*** | UPF2 regulator of nonsense mediated mRNA decay |
| ***WDFY4*** | WDFY family member 4 |
| ***WDR37*** | WD repeat domain 37 |
| ***XPNPEP1*** | X-Prolyl aminopeptidase 1 |
| ***XPO4*** | exportin 4 |
| ***ZAN*** | zonadhesin |
| ***ZNF433*** | zinc finger protein 433 |
| ***ZNF732*** | zinc finger protein 732 |

**References**

1. https://nature.com/subjects
2. <https://www.genome.gov/genetics-glossary>
3. Zheng G, Freidlin B, Li Z, Gastwirth JL. Genomic control for association studies under various genetic models. Biometrics. 2005 Mar;61(1):186–92.
4. Slatkin M. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. Nat Rev Genet. 2008 Jun;9(6):477-85. doi: 10.1038/nrg2361. PMID: 18427557; PMCID: PMC5124487.
5. <https://www.cancer.gov/publications/dictionaries/genetics-dictionary>
6. Gottesman II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. Am J Psychiatry. 2003 Apr;160(4):636-45. doi: 10.1176/appi.ajp.160.4.636. PMID: 12668349.
7. Sticca EL, Belbin GM, Gignoux CR. Current Developments in Detection of Identity-by-Descent Methods and Applications. Front Genet. 2021 Sep 10;12:722602. doi: 10.3389/fgene.2021.722602. PMID: 34567074; PMCID: PMC8461052.
8. [https://www.genecards.org](https://www.genecards.org/)
9. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. Curr Protoc Bioinformatics. 2016 Jun 20;54:1.30.1-1.30.33.
10. https://www.ncbi.nlm.nih.gov