DISEASE MECHANISMS

Genetic architectures of psychiatric disorders: the emerging picture and its implications

Patrick F. Sullivan[1], Mark J. Daly[2] and Michael O'Donovan[3]

Abstract | Psychiatric disorders are among the most intractable enigmas in medicine. In the past 5 years, there has been unprecedented progress on the genetics of many of these conditions. In this Review, we discuss the genetics of nine cardinal psychiatric disorders (namely, Alzheimer's disease, attention-deficit hyperactivity disorder, alcohol dependence, anorexia nervosa, autism spectrum disorder, bipolar disorder, major depressive disorder, nicotine dependence and schizophrenia). Empirical approaches have yielded new hypotheses about aetiology and now provide data on the often debated genetic architectures of these conditions, which have implications for future research strategies. Further study using a balanced portfolio of methods to assess multiple forms of genetic variation is likely to yield many additional new findings.

A core set of psychiatric conditions — madness, mania and melancholia — has been perplexing for millennia. Although mortality is increased for many psychiatric disorders[1], their major impact is on morbidity: psychiatric disorders account for around one-third of disability worldwide[2] and cause enormous personal and societal burdens[3].

In the past century, considerable efforts towards understanding the nature of psychiatric disorders have been undertaken. There have been successes, and a few diseases (for example, pellagra[4] and neurosyphilis[5]) with prominent psychiatric manifestations that were previously prevalent are now rare in many parts of the world. These few triumphs stand in contrast to decades of frustration and occasional notoriety, when highly publicized and plausible findings failed to be replicated. Indeed, most psychiatric disorders have been intractable to approaches that were fruitful in other areas of medicine. Thus, psychiatric syndromes are generally referred to as 'disorders' (which are illnesses that disrupt normal function), and only a few are referred to as 'diseases' (which are disorders with a known pathophysiology or structural pathology). An obvious goal of psychiatric research is to convert idiopathic disorders into pathophysiologically defined diseases.

Since 2007, numerous robust and replicable genetic findings have been reported for psychiatric disorders. These advances have mostly been through genome-wide association studies (GWASs) and structural variation studies, although studies of uncommon or rare exonic variation are likely to have a prominent role in the next few years. These results meet community standards in human genetics for significance and replication[6]. Although these findings often appear in high-profile journals, sentiments such as 'genetics has failed in psychiatry' or 'there are no genes for psychiatric disorders' are still heard. A review of psychiatric genetics is thus particularly opportune.

Over 300 psychiatric disorders have been described, and nine are covered in this Review. The conditions selected are all psychiatric disorders that have been subjected to intensive genetic study and for which genome-wide results (usually GWASs and structural variation but also genome-wide linkage and resequencing) have been obtained. The disorders and their abbreviations are defined in TABLE 1 and Supplementary information S1 (table), and the heritabilites and lifetime prevalences are depicted in FIG. 1a. Intellectual disability could have been included, but the voluminous literature on this topic has been thoroughly reviewed[7–9]. Studies of other psychiatric disorders are in progress, but the published data are few (for example, in the cases of obsessive-compulsive disorder, Tourette's syndrome and panic disorder).

The genetic dissection of complex traits has been frequently reviewed[6,10–12], so here we provide a brief overview of the approaches and study design considerations in BOX 1. Advances in genetics are often yoked to

Pellagra
A disease caused by niacin (vitamin B_3) deficiency that has prominent neuropsychiatric symptoms.

Neurosyphilis
Infection of the central nervous system by *Treponema pallidum* (which is the spirochete that causes syphilis) and often causes prominent neuropsychiatric symptoms.

[1]*Departments of Genetics and Psychiatry, CB# 7264, 5097 Genomic Medicine, University of North Carolina at Chapel Hill, North Carolina 27599–27264, USA.*
[2]*Department of Genetics, Harvard University, Cambridge, Massachusetts, USA.*
[3]*MRC Centre for Neuropsychiatric Genetics, Cardiff University, Cardiff CF14 4XN, UK.*
Correspondence to P.F.S.
e-mail: pfsulliv@med.unc.edu
doi:10.1038/nrg3240
Published online 10 July 2012

Table 1 | **Defining features of nine psychiatric disorders***

Name	Life prevalence	Heritability	Essential characteristics	Notable feature
Alzheimer's disease	0.132	0.58	Dementia, defining neuropathology	Of the top ten causes of death in the United States, Alzheimer's disease alone has increasing mortality
Attention-deficit hyperactivity disorder (ADHD)	0.053	0.75	Persistent inattention, hyperactivity, impulsivity	Costs estimated at ~$US100 × 10^9 per year
Alcohol dependence (ALC)	0.178	0.57	Persistent ethanol use despite tolerance, withdrawal, dysfunction	Most expensive psychiatric disorder (total costs exceed US$225 × 10^9 per year)
Anorexia nervosa	0.006	0.56	Dangerously low weight from self-starvation	Notably high standardized mortality ratio
Autism spectrum disorder (ASD)	0.001	0.80	Markedly abnormal social interaction and communication beginning before age 3	Huge range of function, from people requiring complete daily care to exceptional occupational achievement
Bipolar disorder (BIP)	0.007	0.75	Manic-depressive illness, episodes of mania, usually with major depressive disorder	As a group, nearly as disabling as schizophrenia
Major depressive disorder (MDD)	0.130	0.37	Unipolar depression, marked and persistent dysphoria with physical and cognitive symptoms	Ranks number one in the burden of disease in the world
Nicotine dependence (NIC)	0.240	0.67	Persistent nicotine use with physical dependence (usually cigarettes)	Major preventable risk factor for many diseases
Schizophrenia (SCZ)	0.004	0.81	Long-standing delusions and hallucinations	Life expectancy decreased by 12–15 years

*Most of these definitions are made more restrictive by requiring persistence over time (for example, the criteria for SCZ require ≥6 months of symptoms), substantial impairment and presence across multiple different contexts. See Supplementary information S1 (table) for more detail. Additional sources are REFS 1,2,181–183).

Genome-wide association studies
(GWASs). Unbiased genome screens of unrelated cases and appropriately matched controls or parent-affected child trios. The dominant technology has been individual genotyping using highly multiplexed SNP arrays.

Structural variation
A genomic alteration that changes the number of copies or the arrangement of the genome. Copy number variants are one type of structural variation.

Genome-wide linkage
A type of unbiased genome screen based on multiplex pedigrees. Genotyping approaches have included restriction fragment length polymorphisms, microsatellites and SNP arrays. After adjustment for multiple comparisons, the signal is the co-segregation of a genotype with a disease phenotype within the pedigrees.

Karyotyping
Determination of the microscopic appearance and gross changes in chromosomal content and structure of a cell.

Meta-analyses
These are methods for summarizing and combining results across multiple studies and are widely used in complex trait genetics. Meta-analyses combine the summary results from each study.

technological advancements. Major approaches that have been informative in psychiatric genetics include assessment of: structural variation through karyotyping, array-based methods and high-throughput sequencing[13–16] (TABLE 2 and Supplementary information S2 (table)); GWASs using highly multiplexed SNP arrays and, potentially, high-throughput sequencing[12,17–19] (TABLE 3); and high-throughput sequencing to uncover rare variants of fairly strong effect (perhaps arising *de novo*)[20,21]. Genome-wide linkage and hypothesis-driven candidate gene association studies have also been conducted but, as in many areas of biomedicine, with uncertain yield[22–26].

In this Review, we first summarize the literature for the nine disorders listed in TABLE 1 with a particular emphasis on the findings that appear to meet community standards for replication in human genetics (that is, robustly significant with consistent effects across samples)[6]. We then highlight new hypotheses that have emerged across the allelic spectrum, including *de novo* and rare exonic mutations, rare structural variation and common variation from GWASs. Crucially, these results provide empirical insights into the genetic architectures of these disorders: data that are essential for guiding future work in this area.

Alzheimer's disease

Rare variation. Before 2007, rare autosomal dominant mutations in amyloid beta precursor protein (*APP*), presenilin 1 (*PSEN1*) and *PSEN2* were known to cause early-onset familial Alzheimer's disease[27]. These loci have atypically large effect sizes, thereby facilitating identification using 'past generation' technologies, such as candidate gene association and genome-wide linkage studies (Supplementary information S2 (table)). Treatments for Alzheimer's disease based on these findings have been developed and are undergoing testing. Rare structural variation duplications containing *APP* have been associated with Alzheimer's disease[28,29]. Small-exome sequencing studies of Alzheimer's disease have been published[30,31], and larger studies are in progress and should provide a more nuanced understanding of the role of rare exonic mutations in the pathogenesis of the disease.

Common variation. In the early 1990s, apolipoprotein E (*APOE*) was identified by candidate gene association as a susceptibility gene for late-onset Alzheimer's disease, in no small part owing to its unusually large effect size[27,32] (TABLE 3). In 2009, GWASs from two large consortia[33,34] implicated three novel loci, and six additional loci were identified in 2011 (REFS 35,36). Full meta-analyses are keenly awaited, but the ten loci that have been identified to date account for ~20% of the total variation in risk or ~33% of the risk attributable to genetic effects, with the major contribution being from *APOE*. Note that the association of one gene identified by GWASs, complement component (3b/4b) receptor 1 (*CR1*), might result from structural variation[37].

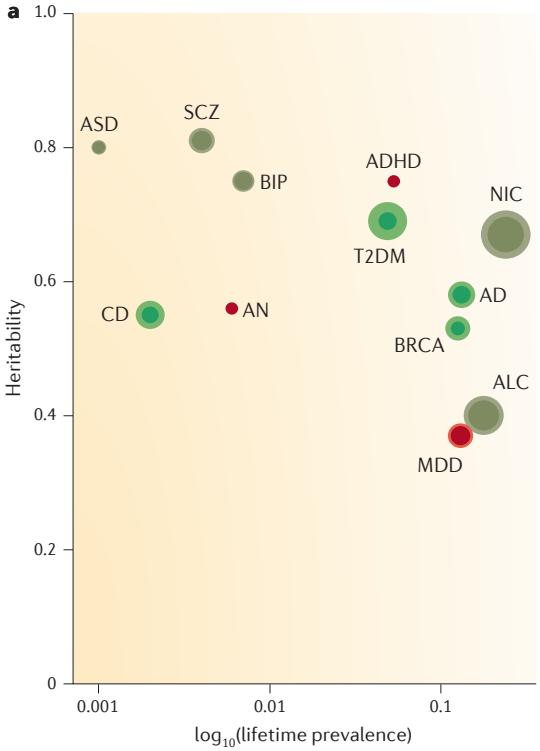
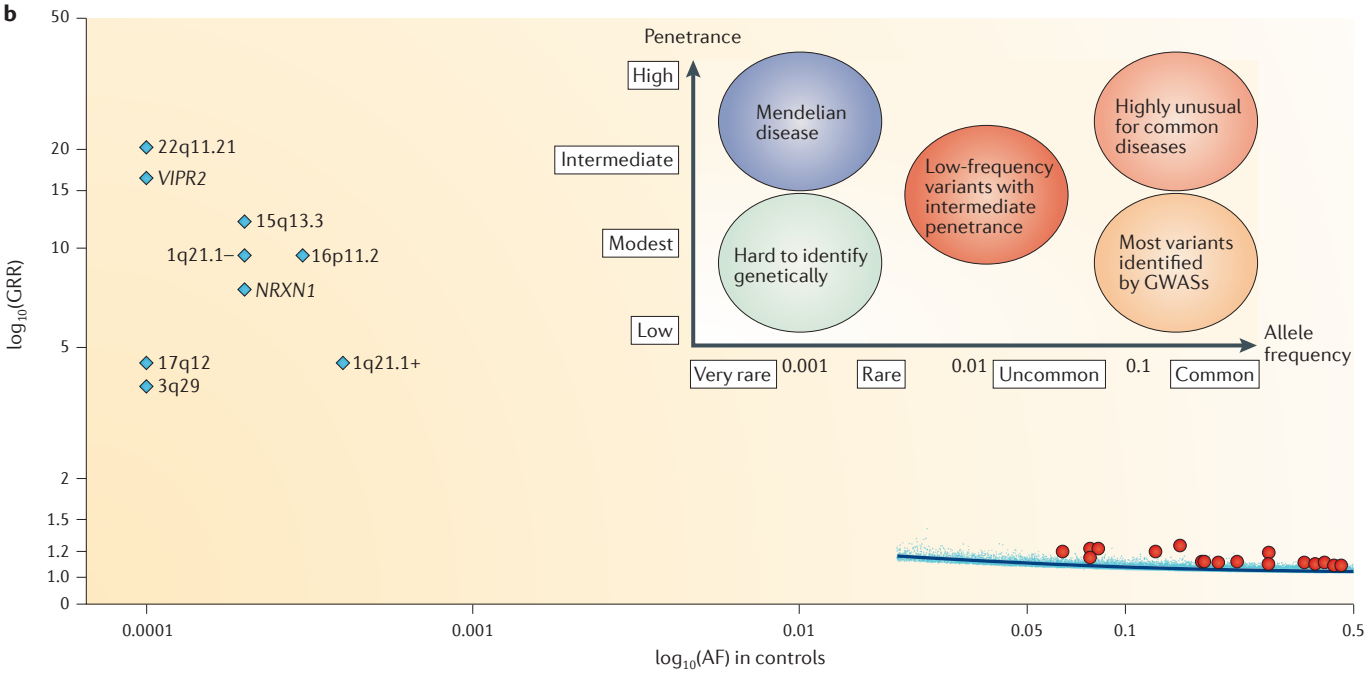


Figure 1 | **Results pertaining to genetic architecture.** **a** | Plot of heritability by $\log_{10}$(lifetime prevalence) for the nine psychiatric disorders considered in this Review plus three complex diseases for which genetic dissection has been particularly successful (TABLE 1; Supplementary information S1 (table)). Each disorder is plotted as heritability by lifetime prevalence. Colour indicates qualitative success in identifying aetiological genetic variation (with bright green meaning notably successful, dark green meaning some successes and red meaning minimal or no clear success to date). The bubble sizes are proportional to the numbers of cases studied in genome-wide association studies (GWASs; the smaller bubble indicates discovery number of cases (N_{case}), and the larger bubble indicates the total N_{case} for discovery plus replication samples). **b** | Allelic spectrum of schizophrenia (SCZ). The inset is a conceptual schematic from a 2008 *Nature Reviews Genetics* article[10]. The lower part of the figure depicts empirical results for SCZ. The *x*-axis is $\log_{10}$(allele frequency (AF)) in controls. The *y*-axis is the point estimate for $\log_{10}$(genotypic relative risk (GRR). For clarity, confidence intervals are not shown. There are no known Mendelian variants for SCZ (AF << 0.0001, GRR >> 50). There are no known common variants (AF > 0.1) with GRR > 1.5, and these can be excluded with >99% statistical power. Nine structural variants associated with SCZ are shown as light blue diamonds (TABLE 2; 1q21.1– is the deletion and 1q21.1+ is the duplication). If AF in controls was 0, AF was set to 0.0001. These structural variants do not have a corresponding region in the inset. Seventeen common variants have been associated with SCZ (red circles; TABLE 3). SNPs contributing to the Psychiatric Genomics Consortium SCZ risk profile score[59] (21,171 autosomal SNPs with $P_T < 0.1$; BOX 3, panel **b**) are shown in light blue dots with a lowess smoother in dark blue. AD, Alzheimer's disease; ADHD, attention-deficit hyperactivity disorder; ALC, alcohol dependence; AN, anorexia nervosa; ASD, autism spectrum disorder; BIP, bipolar disorder; BRCA, breast cancer; CD, Crohn's disease; MDD, major depressive disorder; NIC, nicotine usage (maximum cigarettes per day); SCZ, schizophrenia; T2DM, type 2 diabetes mellitus. The inset in panel **b** is adapted, with permission, from REF. 10 © (2008) Macmillan Publishers Ltd. All rights reserved.



Intriguingly, pathway analyses (BOX 2) of Alzheimer's disease implicate cholesterol metabolism and the innate immune response[38]. Genes attaining genome-wide significance point towards immune and inflammatory processes (clusterin (*CLU*) and *CR1*), lipid processing (*APOE*, *CLU* and *ABCA7*) and endocytosis (phosphatidylinositol-binding clathrin assembly protein (*PICALM*), bridging integrator 1 (*BIN1*), CD2-associated protein (*CD2AP*) and *CD33*). Altered immune function and lipid metabolism had previously been proposed as Alzheimer's disease risk factors, but whether these represented causation or reverse causation was unclear[39]. The genetic findings now strongly point to reverse causation.

It is unclear how the above findings relate to accumulation of β-amyloid in Alzheimer's disease pathogenesis, but some relationship seems likely. For example, phosphatidylinositol-binding clathrin assembly protein (PICALM) and other endocytic molecules can modify

Lowess smoother
Locally weighted scatterplot smoothing, which is one technique to fit a curve to a scatterplot.

Box 1 | **Study design considerations: simplex and multiplex**

Study design is a crucial component of human genetics research. The major designs are case–control studies and pedigree-based studies. The most common design is the case–control study, in which the frequency of a genetic variant in those cases with a disorder is contrasted against the appropriate control group. Case–control designs are used in most genome-wide association studies (GWASs)[156] and next-generation sequencing studies, as they are efficient and conceptually straightforward[157]. Case–control studies are simpler, and most biases can be surmounted by careful study procedures, but they cannot delineate rare inherited variation from *de novo* variation. Family-based designs are more complex but can be used for association testing as well as for linkage evaluation of co-segregation of genotypes and phenotypes within pedigrees. They provide protection against a key form of bias (namely, population stratification artefacts) but are less efficient given that the unit of analysis is a set of relatives; however, it is possible to identify mutations that arise *de novo*.

An additional decision is whether to focus on the presence or absence of other affected family members (multiplex pedigrees and simplex pedigrees, respectively). Human genetic studies have classically focused on multiplex pedigrees under the assumption that these pedigrees are enriched for causal genetic variation with higher penetrance. A focus on multiplex pedigrees has led to the identification of specific mutations underlying hundreds of Mendelian disorders (including autism spectrum disorder (ASD) and Alzheimer's disease). Simplex pedigrees have become popular for ASD and schizophrenia (SCZ). Simplex-based ascertainment is tailored to evaluate *de novo* mutations and is predicated on a model in which disorder with dramatically reduced fecundity and a proven role of *de novo* structural variation might be explicable as a series of Mendelian disorders that can be attributed to recent high-penetrance mutations in any of a large number of genes.

However, this important choice is not simple, and it continues to be moderately controversial. Some investigators believe a focus on simplex pedigrees to be optimal, and other investigators have concerns about the implications of this decision. Some of the issues are listed below.

- Correct classification as simplex or multiplex requires confident knowledge of family history — many people either do not know their family psychiatric histories, true episodes of illness may have been kept private from other relatives, and some affected individuals can over-call illness in their relatives (for example, an individual with alcohol dependence labelling all relatives who drink as the same).
- Fecundity is a major confounder. If there are greater numbers of relatives, there is a greater chance of multiplex classification. In addition, the presence of a psychiatric disorder can reduce fecundity (for example, fecundity is reduced in SCZ, and having a child with ASD can be a powerful inducement not to reproduce further). If fecundity had not been inhibited because of a psychiatric disorder, some apparently simplex families might have been revealed to be multiplex.
- Simplex designs often require both parents. This complicates recruitment, increases genetic assay costs and becomes increasingly less practical for disorders with later ages of onset.
- Both designs have a hidden weakness in the possibility of enriching for environmental causes of illness. Many psychiatric disorders have multiple different but rare environmental risk factors that are sufficient to cause a disorder. These potent exposures are sometimes very difficult to detect or are not routinely evaluated. Examples include mercury poisoning and ASD or viral meningitis and SCZ. Contrary to its intent, simplex cases may be enriched for difficult-to-detect, individual-specific environmental causes. Multiplex ascertainment could enrich for shared environmental causes.

Some recent data pertain to this choice. Unexpectedly, simplex and multiplex ASD pedigrees show fairly similar *de novo* mutation rates for structural variation[81,82] and exonic variation[84–86]. It is possible that larger studies will find differences in *de novo* mutation rates between simplex and multiplex families but the magnitude is likely to be smaller than anticipated. For SCZ, the available data are insufficient to resolve this issue[45,50]. It has also been pointed out that *de novo* events must confer risk in multiplex families, as such mutations increase the chance that an individual is affected and increase risk in that person's offspring. Intriguingly, there are three instances of ASD cases with *de novo* deletions of 16p11.2 who also had an affected sibling without this deletion[158–160], along with similar observations for structural variants in 1q21.1 and 17p12 (REF. 159).

β-amyloid toxicity in yeast and other model systems[40]. Although these genetic findings provide support for novel causal relationships that could be targeted by treatments, the association data point to genomic regions, not genes. Moreover, the proximal steps from genotype to phenotype are unclear, and many of the implicated genes are plausibly involved in multiple relevant functions (for example, *CLU* is involved in altered immune function and lipid processing).

Psychotic disorders

Rare variation. Unfortunately, unlike the case for Alzheimer's disease, no Mendelian forms of bipolar disorder (BIP) and schizophrenia (SCZ) have been identified[41]. However, rare but potent structural variants (with a frequency of <0.5% and a genotypic relative risk (GRR) of 5–20) have a role in a small proportion of cases with SCZ (TABLE 2; Supplementary information S3 (figure)). None is fully penetrant, and nearly all appear to be nonspecific, as risk is often increased for SCZ, autism spectrum disorder (ASD), developmental delay, intellectual disability, epilepsy, somatic dysmorphism and extremes of body mass and head size. Most of these structural variation regions are fairly large (hundreds of kilobases to hundreds of megabases) and generally centre on structural variation hotspots[42]. Two rare structural variants affect single genes (namely, neurexin 1 (*NRXN1*) and vasoactive intestinal peptide receptor 2 (*VIPR2*))[43,44], offering opportunities for downstream functional studies. Pathway analyses of genes that are intersected by rare structural variants suggest enrichment for neuronal processes of plausible aetiological relevance (for example, postsynaptic signalling)[45–47]. The structural variation regions in TABLE 2 probably represent 'low-hanging fruit', and more discoveries are likely when improved technologies for structural variation detection are applied to larger samples[15].

A complementary approach is to evaluate structural variation 'burden' in cases compared to controls (for example, the number of structural variants per person)[48,49]. This approach tests an explicitly multigenic model whereby many rare but different genomic disruptions have an impact on disease risk. Increased structural variation burden in SCZ cases has been reported by multiple groups[47,48,50]. One report found more rare structural variation in SCZ cases (odds ratio = 1.15), particularly for large deletions (odds ratio = 3.6)[48]. *De novo* structural variants are also more common in SCZ cases[45]. For BIP, there are reports of increased[51–53] and similar structural variation burden in cases versus controls[54,55]. *De novo* structural variation may be relevant in BIP (odds ratio = 4.8), particularly in cases with earlier ages of onset (odds ratio = 6.3)[53].

Multiple studies are now evaluating the role of *de novo*, rare and uncommon exonic variation in BIP and SCZ using resequencing or genotyping approaches. Two small-exome-sequencing studies[56,57] reported rates of putatively functional mutations that exceeded null expectations in SCZ cases (although the rate of *de novo* point mutations was not elevated in cases, and specific genes were not identified). A recent study[58] conducted high-throughput sequencing on 166 cases with SCZ and

Table 2 | **Structural variation associated with psychiatric disorders**

Structural variant	Location (Mb)	Genes	Type	Disorder	Frequency in cases	Frequency in controls	Odds ratio	*P* value	Other associations	Refs
1q21.1	chr1: 145.0–148.0	34	Deletion	SCZ	0.0018	0.0002	9.5	8×10^{-6}	Developmental delay, intellectual disability, micro- and macrocephaly, dysmorphia, epilepsy, cataracts, cardiac defects, possibly ASD[185], thrombocytopenia-absent radius syndrome[48,159,184–188]	184
			Duplication	SCZ	0.0013	0.0004	4.5	0.02		184
2p16.3	chr2: 50.1–51.2	*NRXN1* exons	Deletion	ASD				0.004	Developmental delay, intellectual disability, epilepsy, Pitt–Hopkins-like syndrome 2	81
			Deletion	SCZ	0.0018	0.0002	7.5	1×10^{-6}		184
3q29	chr3: 195.7–197.3	19	Deletion	SCZ	0.0010	0.0	3.8	4×10^{-4}	Developmental delay, intellectual disability, possibly ASD	184
7q11.23	chr7: 72.7–74.1	25	Duplication	ASD	0.0011			0.003	Developmental delay, intellectual disability. Deletion: Williams–Beuren syndrome	81
7q36.3	chr7: 158.8–158.9	*VIPR2*	Duplication	SCZ	0.0024	0.0001	16.4	4×10^{-5}		44, 184
15q11.2	chr15: 23.6–28.4	70	Duplication	ASD	0.0018			4×10^{-9}	Developmental delay, intellectual disability, Prader–Willi and Angelman syndromes[188]	81
15q13.3	chr15: 30.9–33.5	12	Duplication	ADHD	0.0125	0.0061	2.1	2×10^{-4}	Developmental delay, intellectual disability, epilepsy[188,189]	120
			Duplication	ASD	0.0013			2×10^{-5}		81
			Deletion	SCZ	0.0019	0.0002	12.1	7×10^{-7}		184
16p13.11	chr16: 15.4–16.3	8	Duplication	ADHD	0.0164	0.0009	13.9	8×10^{-4}	Deletion: developmental delay, epilepsy[188,189]	119
16p11.2	chr16: 29.5–30.2	29	Deletion	ASD	0.0037			5×10^{-29}	Developmental delay, intellectual disability, epilepsy, macrocephaly, obesity[190,191]	81
			Duplication	ASD	0.0013			2×10^{-5}	Developmental delay, intellectual disability, epilepsy, microcephaly, low body mass index[190,191]	81
			Duplication	SCZ	0.0031	0.0003	9.5	3×10^{-8}		184
17q12	chr17: 34.8–36.2	18	Deletion	ASD	0.0017	0.0	6.12	9×10^{-4}		192
			Deletion	SCZ	0.0006	0.0	4.49	3×10^{-4}		
22q11.21	chr22: 18.7–21.8	53	Deletion or duplication	ASD	0.0013			0.002	Developmental delay, intellectual disability, velocard-iofacial-DiGeorge syndrome	81
			Deletion	SCZ	0.0031	0.0	20.3	7×10^{-13}		184

Locations are US National Center for Biotechnology Information (NCBI) Build 37 and University of California, Santa Cruz (UCSC) hg19. The positions of these structural variants are denoted in Supplementary information S3 (figure) with yellow circles. For succinctness, the citations refer to the most comprehensive study rather than to an initial report. 'Genes' refers to the number from the UCSC Known Genes data set. ADHD, attention-deficit hyperactivity disorder; ASD, autism spectrum disorder; SCZ, schizophrenia.

β-amyloid
Neuronal accumulation of this peptide contributes to the aetiology of Alzheimer's disease.

Multiplex pedigrees
Family constellations containing more than one affected individual.

Simplex pedigrees
Family constellations containing one affected individual.

on 307 controls, followed by genotyping of over 5,000 variants in 2,617 independent cases and 1,800 controls (the cases were enriched for treatment resistance or strong family histories). No finding met genome-wide significance. Larger studies are ongoing and will aid understanding this area in 2012–2013.

Common variation. The Psychiatric Genomics Consortium (PGC) recently published mega-analyses for SCZ and BIP[59,60]. In SCZ, 9,394 cases and 12,462 controls were combined in a single analysis, and the top 81 statistically independent loci from that analysis were then tested in over 8,000 cases. The mega-analysis identified seven significant loci (TABLE 3). A sign test for consistency between the mega-analysis and the follow-up stage was highly significant, implying that many of the 81 top loci include true risk loci but that power was insufficient. For BIP, the discovery phase consisted of 7,481 cases and 9,250 controls with a follow-up of 34 statistically independent loci in around 4,500 cases. Two loci exceeded genome-wide significance (TABLE 3). Similarly, a sign test between the discovery and follow-up results was highly significant, again suggesting insufficient power[60].

In BIP, the genome-wide significant association at calcium channel, voltage-dependent, L type, alpha 1C subunit (*CACNA1C*) deserves specific comment, given

Table 3 | **Genome-wide association study findings for psychiatric disorders**

Phenotype	SNP	Location	Discovery GWAS (cases/controls)	Largest meta-analysis (cases/controls)	*P* value	Odds ratio	Nearest gene
Alzheimer's disease	rs3818361	chr1:207784968	2,018/5,324 (REF. 34)	<19,870/39,846 (REF. 35)	3.7×10^{-14}	1.18	*CR1*
	rs744373	chr2:127894615	3,006/14,642 (REF. 193)	<19,870/39,846 (REF. 35)	2.6×10^{-14}	1.17	*BIN1*
	rs9349407	chr6:47453378	8,309/7,366 (REF. 36)	18,762/29,827 (REF. 36)	8.6×10^{-9}	1.11	*CD2AP*
	rs11767557	chr7:143109139	8,309/7,366 (REF. 36)	18,762/35,597 (REF. 36)	6.0×10^{-10}	1.11	*EPHA1*
	rs11136000	chr8:27464519	3,941/7,848 (REF. 33)	8,371/26,965 (REF. 193)	1.6×10^{-16}	1.18	*CLU*
	rs610932	chr11:59939307	6,688/13,251 (REF. 35)	>19,000/38,000 (REF. 35)	1.2×10^{-16}	1.10	*MS4A* cluster
	rs3851179	chr11:85868640	3,941/7,849 (REF. 33)	8,371/26,966 (REF. 193)	3.2×10^{-12}	1.15	*PICALM*
	rs3764650	chr19:1046520	5,509/11,531 (REF. 35)	>17,000/34,000 (REF. 35)	5.0×10^{-21}	1.23	*ABCA7*
	rs2075650	chr19:45395619		8,371/26,966 (REF. 193)	1×10^{-295}	2.53	*APOE*, *TOMM40*
	rs3865444	chr19:51727962	8,309/7,366 (REF. 36)	18,762/29,827 (REF. 36)	1.6×10^{-9}	1.10	*CD33*
Alcohol consumption	rs1229984	chr4:100239319	REF. 102		1.3×10^{-11}		*ADH1B*
	rs6943555	chr7:69806023	REF. 101		4.1×10^{-9}		*AUTS2*
	rs671	chr12:112241766	REF. 100		3×10^{-211}		*ALDH2*
Bipolar disorder	rs12576775	chr11:79077193	7,481/9,251 (REF. 60)	11,974/51,793 (REF. 60)	4.4×10^{-8}	1.14	*ODZ4*
	rs4765913	chr12:2419896	7,481/9,250 (REF. 60)	11,974/51,792 (REF. 60)	1.5×10^{-8}	1.14	*CACNA1C*
	rs1064395	chr19:19361735	682/1300 (REF. 194)	8,441/35,362 (REF. 194)	2.1×10^{-9}	1.17	*NCAN*
Nicotine consumption	rs1329650	chr10:93348120	38,181 (REF. 93)	73,853 (REF. 93)	5.7×10^{-10}		*LOC100188947*
	rs1051730	chr15:78894339	38,181 (REF. 93)	73,853 (REF. 93)	2.8×10^{-73}		*CHRNA3*
	rs3733829	chr19:41310571	38,181 (REF. 93)	73,853 (REF. 93)	1.0×10^{-8}		*EGLN2*, *CYP2A6*
Smoking cessation	rs3025343	chr9:136478355	41,278 (REF. 93)	64,924 (REF. 93)	3.6×10^{-8}	1.13	*DBH*
Smoking initiation	rs6265	chr11:27679916	74,035 (REF. 93)	143,023 (REF. 93)	1.8×10^{-8}	0.94	*BDNF*
Schizophrenia	rs1625579	chr1:98502934	9,394/12,462 (REF. 59)	17,839/33,859 (REF. 59)	1.6×10^{-11}	1.12	*MIR137*
	rs2312147	chr2:58222928		18,206/42,536 (REF. 195)	1.9×10^{-9}	1.09	*VRK2*
	rs1344706	chr2:185778428	479/2,937 (REF. 174)	18,945/38,675 (REF. 196)	2.5×10^{-11}	1.10	*ZNF804A*
	rs17662626	chr2:193984621	9,394/12,463 (REF. 59)	17,839/33,860 (REF. 59)	4.6×10^{-8}	1.20	*PCGEM1*
	rs13211507	chr6:28257377	3,322/3,587 (REF. 70)	18,206/42,536 (REF. 195)	1.4×10^{-13}	1.22	*MHC*
	rs7004635	chr8:3360967	9,394/12,465 (REF. 59)	17,839/33,862 (REF. 59)	2.7×10^{-8}	1.10	*MMP16*
	rs10503253	chr8:4180844	9,394/12,464 (REF. 59)	17,839/33,861 (REF. 59)	4.1×10^{-8}	1.11	*CSMD1*
	rs16887244	chr8:38031345	3,750/6,468 (REF. 68)	8,133/11,007 (REF. 68)	1.3×10^{-10}	1.19	*LSM1*
	rs7914558	chr10:104775908	9,394/12,466 (REF. 59)	17,839/33,863 (REF. 59)	1.8×10^{-9}	1.10	*CNNM2*
	rs11191580	chr10:104906211	9,394/12,467 (REF. 59)	17,839/33,864 (REF. 59)	1.1×10^{-8}	1.15	*NT5C2*
	rs11819869	chr11:46560680	1,169/3,714 (REF. 197)	3,738/7,802 (REF. 197)	3.9×10^{-9}	1.25	*AMBRA1*
	rs12807809	chr11:124606285		18,206/42,536 (REF. 195)	2.8×10^{-9}	1.12	*NRGN*
	rs12966547	chr18:52752017	9,394/12,468 (REF. 59)	17,839/33,865 (REF. 59)	2.6×10^{-10}	1.09	*CCDC68*
	rs9960767	chr18:53155002		18,206/42,537 (REF. 195)	4.2×10^{-9}	1.20	*TCF4*

its mechanistic implications. Indeed, multiple voltage gated calcium channel subunits were among the top 34 loci followed up in the BIP GWAS. Calcium channels are a treatment for BIP and regulate neuronal excitability and multiple brain functions, including long-term potentiation and synaptic plasticity. Combined analysis of the PGC BIP and SCZ samples strengthened the association in the *CACNA1C* region. Further, results from SGENE+[61] implicate neurogranin (*NRGN*), which may act as a calcium sensor[62]. Therefore, a detailed investigation of brain calcium biology is warranted for both BIP and SCZ.

For SCZ, the strongest association is in the extended major histocompatability complex region (MHC region; chr6:27–33 Mb). The evidence for association is compelling, but high gene density and exceptionally high linkage disequilibrium complicate the identification of specific sequence variation. Although it is tempting to propose that the association supports long-standing hypotheses

Genotypic relative risk (GRR). A measure of the effect size of a genetic variant ranging from zero to infinity. A GRR of 1 means no change in risk, GRR < 1 is protective and GRR > 1 is predisposing.

Table 3 (cont.) | **Genome-wide association study findings for psychiatric disorders**

Phenotype	SNP	Location	Discovery GWAS (cases/controls)	Largest meta-analysis (cases/controls)	*P* value	Odds ratio	Nearest gene
Schizophrenia and bipolar disorder	rs1344706	chr2:185778428	479/2,937 (REF. 174)	21,274/38,675 (REF. 196)	4.1×10^{-13}	1.11	*ZNF804A*
	rs2239547	chr3:52855229	9,394/12,471 (REF. 59)	16,374/14,046 (REF. 59)	7.8×10^{-9}	1.12	*ITIH3–ITIH4*
	rs10994359	chr10:62222107	9,394/12,470 (REF. 59)	16,374/14,045 (REF. 59)	2.4×10^{-8}	1.22	*ANK3*
	rs4765905	chr12:2349584	9,394/12,469 (REF. 59)	16,374/14,044	7.0×10^{-9}	1.11	*CACNA1C*

This table focuses on results achieving genome-wide significance in large samples. We use a significance threshold of 5×10^{-8} (REF. 198). Most associations that achieve this level of significance are secure, but some may ultimately prove not to be. Included are SNPs with $P < 5 \times 10^{-8}$ that were evaluated in samples of a minimum of around 10,000 cases and 10,000 controls. Discovery sample sizes reflect the primary samples for which full genome-wide association studies (GWASs) were conducted. In most cases, discovery *P* values were $> 5 \times 10^{-8}$ but met a threshold (typically 1×10^{-5}) for inclusion in replication efforts. In some instances, simultaneous publications based on overlapping samples were considered to be 'discovery' studies. Where this occurred, providing samples of roughly equivalent sizes, the most significant primary GWAS findings are given, otherwise the largest discovery samples are favoured. For many studies, it was not possible to extract the exact sample size used for each locus, so the sample sizes above are approximate. *P* values and odds ratios are from the meta-analysis with the largest sample sizes. If two meta-analyses based on overlapping samples reported similar results, the 'discovery' study is cited. The genes that are nearest to each locus are provided, but aetiological variants are generally unknown, and it remains likely that some of the associations do not alter the function of the designated genes (for example, *ITIH3–ITIH4*, in which multiple correlated SNPs span many genes, and *TCF4–CCDC68*, in which statistically independent associations occur in *TCF4* and closer to *CCDC68*). In the *TCF4–CCDC68* example, it may be that both associations point to the same functional element, but it is also possible that independent aetiological variants occur in adjacent genes. For the schizophrenia loci attributed to REF. 195, no discovery sample size is listed because the initial *P* values were modest, and as the authors conducted multiple follow-up analyses, there is no obvious discovery sample. For the major histocompatibility complex (MHC), the International Schizophrenia Consortium[70] is designated a discovery study, as it was the only primary GWAS for which genome-wide significance at the MHC was obtained. REF. 195 is cited for the meta-analysis at the MHC, as it reported the most significant MHC association[195]. The most significant SNP at the MHC across the two studies is not identical, and the one listed is taken from REF. 195. Multiple statistically independent SNPs have been reported at the MHC[59,186]. We note that genome-wide significance had been reported in bipolar disorder for ankyrin 3 (*ANK3*) (REF. 199) but not in a larger mega-analysis that included the same samples[60]. Others have reported genome-wide significance for composite phenotype studies of *ITIH3–ITIH4* (REF. 200) (but see REF. 201) and calcium channel, voltage-dependent, L type, alpha 1C subunit (*CACNA1C*)[202] but in samples smaller than required for inclusion in the above table. For alcohol consumption, the rs671 association was in East Asian samples. The alcohol dehydrogenase 1B (class I), beta polypeptide (*ADH1B*) locus was also associated with alcohol dependence. Nicotine consumption is measured in terms of maximal number of cigarettes smoked per day. Smoking initiation refers to ever versus never began smoking. Smoking cessation is whether regular smokers had quit at the time of interview.

concerning roles in SCZ for intra-uterine infection, autoimmunity or even synaptic pruning (in which MHC genes have a role), this lack of precision renders such propositions speculative.

A novel association for SCZ is in Ensembl gene *RP11-490G2.1*, which encodes the primary transcript for the microRNA miR-137 (*MIR137*)[63]. Supporting the hypothesis that this association implicates *MIR137*, predicted targets of miR-137 were significantly enriched for smaller GWAS *P* values ($P < 0.01$), and four of the genes that achieved genome-wide significance contain verified miR-137-binding sites[64]. miR-137 is a key regulator of neuronal development with roles in neurogenesis and maturation[65,66] and is highly expressed at synapses in the cortex and hippocampus[67]. Future studies of networks regulated by miR-137 offer the possibility of insights into SCZ pathophysiology.

GWASs of BIP and SCZ have been predominantly based on subjects of European ancestry, but there are increasing reports from other world ancestries[68,69]. Although those findings do not yet provide additional pathophysiological insights, it is worth noting that a chromosome 8 locus found in an East Asian sample[68] has support in the PGC data set, suggesting that planned mega-analyses across world populations will be informative.

Some of the most intriguing findings for SCZ and BIP are from large sets of genetic markers[70] (BOX 3). There are now replicated data that vulnerability to SCZ is influenced by common genetic variation in hundreds of different loci, and this vulnerability partially overlaps that for BIP[70]. Indeed, the large-scale impact of large numbers of common variants may be a general feature of human complex traits[71–78]

Odds ratio
Similar to genotypic relative risk: a measure of the change in risk associated with a genetic variant.

Mega-analyses
These are methods (that are less widely used than meta-analyses) for summarizing and combining results across multiple studies. Mega-analysis combines individual-level genotype and phenotype data from all subjects in each study.

Major histocompatibility complex region
(MHC). A region of approximately 3 Mb on human chromosome 6p22.1 that is exceptionally complex and has considerable importance to disease. It contains genes encoding cell surface molecules that are important to immunity and disease susceptibility and many other functions.

Autism spectrum disorder

Rare variation. For ASD, there is a notably strong prima facie case for there being a cardinal role for rare variation. Karyotyping studies suggest that on the order of 5% of ASD cases have one of a large number of rare but fairly gross chromosomal abnormalities[14,79]. In addition, ASD has been noted as a co-morbid feature of >100 single-gene Mendelian medical genetic syndromes[80], although the penetrance and confidence of the clinical associations are variable. Indeed, ASD mutations with a high penetrance are exceptional (that is, Rett's syndrome mutations in methyl-CpG-binding protein 2 (*MECP2*) and cyclin-dependent kinase-like 5 (*CDKL5*)), and Mendelian diseases that are enriched for ASD have far less than complete penetrance (for example, fragile X syndrome and tuberous sclerosis)[79].

Analysis of structural variation has been a major focus in ASD research (TABLE 2; Supplementary information S3 (figure)). Implicated loci to date are generally rare and potent risk factors but are incompletely penetrant and are not specific to ASD. As these large events have an impact on the dosages of many genes, biological insight has been slow to emerge; however, pathway analyses of genes within structural variants do implicate neuronal processes of aetiological relevance[45–47]. Large structural variants are present in 5–10% of ASD cases, and the number of ASD structural variants could total 130–234 (REF. 81). There is also consistent evidence for increased structural variation burden in ASD[49,81–83]. For example, 5.8% of ASD probands had ≥1 rare *de novo* structural variant versus 1.7% of their unaffected siblings (odds ratio = 3.5), and this difference was more pronounced for structural variants that intersected genes[81]. The 16p11.2 structural variant that is associated with ASD

Box 2 | **Pathway analysis**

Pathway analysis is based on the assumption that risk variants for a disease will converge on sets of genes with functions that are more closely related to each other than to random sets of genes. For example, dominant forms of Alzheimer's disease are caused by mutations in amyloid beta precursor protein (*APP*), presenilin 1 (*PSEN1*) and *PSEN2*; the PSEN genes encode protein components of γ-secretase, a protease that cleaves APP. The availability of genome-wide association study (GWAS) and structural variation results for many psychiatric disorders, along with increasing amounts of sequence data, have generated interest in using analytic methods for exploiting nonrandom functional relationships between genes containing risk variants. Many approaches have been developed (for example, ALIGATOR[161], INRICH[162], DAPPLE[163] and GRAIL[164]) and reviewed in detail elsewhere[147,165,166]. Although the algorithms differ, the principle behind these methods is to evaluate whether a given set of genomic regions (that is, a broadly inclusive definition of 'pathway') is enriched for genetic variants that show some relationship with disease compared to a null expectation.

There are important subsidiary considerations. The first is the definition of a 'pathway'. Standard pathways consist of sets of genes found in the Gene Ontology[167], the Kyoto Encyclopedia of Genes and Genomes[168] or PANTHER databases[169]. Other pathway gene sets are manually curated by experts in a particular area (for example, genes that are known to make proteins that function at the synapse)[170,171]. In addition, a 'pathway' can consist of genomic regions that are selected for a particular property, such as a high degree of conservation[172] or expression quantitative trait locus (eQTL) associations[173]. Finally, other pathways consist of genes that are known to be connected by experimental data (for example, by protein–protein interaction screens, microRNA target sites or gene expression modules).

It is advantageous that these pathway data sets are defined independently of genetic studies of psychiatric disorders, but they do have limitations and can contain errors of omission and commission. Standard gene sets can have highly overlapping pathways that complicate some analyses, and the pathway content can have variable quality. Expert-curated pathways can be more specific but can be vulnerable to post hoc bias (that is, including genes in a pathway on the basis of results from genetic studies). Pathways based on empirical approaches depend on the quality and completeness of the primary data (for example, existing protein–protein interaction databases cover the interaction space partially).

A second question concerns what is required before a member of a pathway is accepted as having some relationship with disease. For common variation, the analysis might be restricted to genes within recombination regions containing SNPs that have genome-wide significance, an approach that was used successfully to implicate broad biological pathways that are relevant to height[76]. However, much of the interest in pathway analysis involves exploiting much weaker associations under the assumption that these associations more reflect true associations in the context of limited power (signal) rather than chance (noise). If so, those weakly associated SNPs may also be nonrandom with respect to gene sets (BOX 3). The threshold at which SNPs or genes are chosen is arbitrary, and the signal-to-noise ratio for a given arbitrary threshold can vary substantially with sample size and genetic architecture. For rare variants in complex diseases, in light of recent empirical results, pathway analysis will by necessity be based on sets of genes for which the involvement in disease is unclear (for example, genes with a single observed *de novo* exonic deleterious mutation)[84,85].

A third consideration concerns the null expectation to which the observed pathways are compared. Early structural variant pathway analysis did not fully account for important biases, such as that large genes are more likely to be intersected by copy number variants by chance and that some functional pathways — often related to brain development — are enriched for large genes[164]. Early GWAS pathway analysis sometimes did not fully allow for the variable numbers of SNPs per gene and their degree of linkage disequilibrium, both of which have an impact on the probability of a high-ranking association[161]. Thus, it is necessary to be cautious about the use of pathway-based approaches. In psychiatric disorders, some results give cause for optimism[38,45,82].

Finally, in pathway analyses, the unit of inference is the pathway. Tempting although it may be, it is generally inappropriate to make strong inferences about specific variants or genes on the basis of their membership of pathways that attain some level of significance. It may be possible to do so if the variants or genes are subsequently evaluated in data sets independently of those from which the importance of the pathways are derived, using a statistical framework that adequately deals with multiple testing.

and SCZ has been termed a 'mirror image' structural variant, as the deletion and duplication are associated with increased and reduced head and body size. However, it is difficult to understand the clinical features of ASD and SCZ as mirror images, and more importantly ASD is associated with both 16p11.2 deletions and duplications.

ASD is the first psychiatric disorder for which exome sequencing using substantial numbers of samples has been published. Three recent papers describe the results from exome sequencing of ~600 trios and identify roles for *de novo* exonic mutations in sodium channel protein type 2 subunit alpha (*SCN2A*), katanin p60 subunit A-like 2 (*KATNAL2*) and chromodomain helicase DNA-binding protein 8 (*CHD8*) in the pathogenesis of ASD[84–86]. Intriguingly, all three studies noted an increased rate of *de novo* exonic mutations in older parents (with the mutations generally being of paternal origin)[84–86], and pathway analyses reported in two of the studies found that genes containing *de novo* exonic variation were more closely connected in reference to protein–protein interaction databases[84,85]. Additional sequencing studies are in progress.

However, a central finding from these papers was that only a minority of cases had a *de novo* putatively functional variant, suggesting that this class of genetic variation is unlikely fully to explain the clinical entity of ASD. Indeed, estimates from *de novo* exonic mutations (which are similar to those from structural variation data) suggest that ASD is highly polygenic (estimates ranged from 400–1,000 genes)[85,86]. Importantly, a hypothetical model of ASD that it is caused by rare but fully penetrant mutations in 100 different genes could be confidently rejected[84].

Common variation. Evaluation of rare structural variation and exonic variation in ASD is particularly advanced. By contrast, evaluation of common variation is far more limited (FIG. 1a), and the published GWASs for ASD are small by current standards[87–90]. It is currently not possible to discern or dismiss a role for common genetic variation in risk for ASD. In our opinion, GWASs with larger samples are needed for ASD, given that detailed studies of rare variation currently explain a fraction of risk and that common variation plays a clear part in other psychiatric disorders. Indeed, there were few confident findings for GWASs of SCZ when the sample sizes were similar to those that are now available for ASD. Additional support for our recommendation for more GWASs is provided by Voineagu *et al.*[91], who identified a gene expression module that had attenuated expression in post-mortem brain samples of individuals with ASD and that also had enrichment for GWAS signals[91].

Alcohol and nicotine dependence

Alcohol dependence (ALC) and nicotine dependence (NIC) are complex conditions to study, given the requirement for ingestion of a psychoactive substance and cohort effects due to temporal and geographic variation in the availability of ethanol and nicotine. Many investigators focus on ALC and NIC, which are clinically salient but multi-component syndromes[92]. As a part of the Tobacco and Genetics Consortium[93], we determined

Box 3 | **Common variant risk profile**

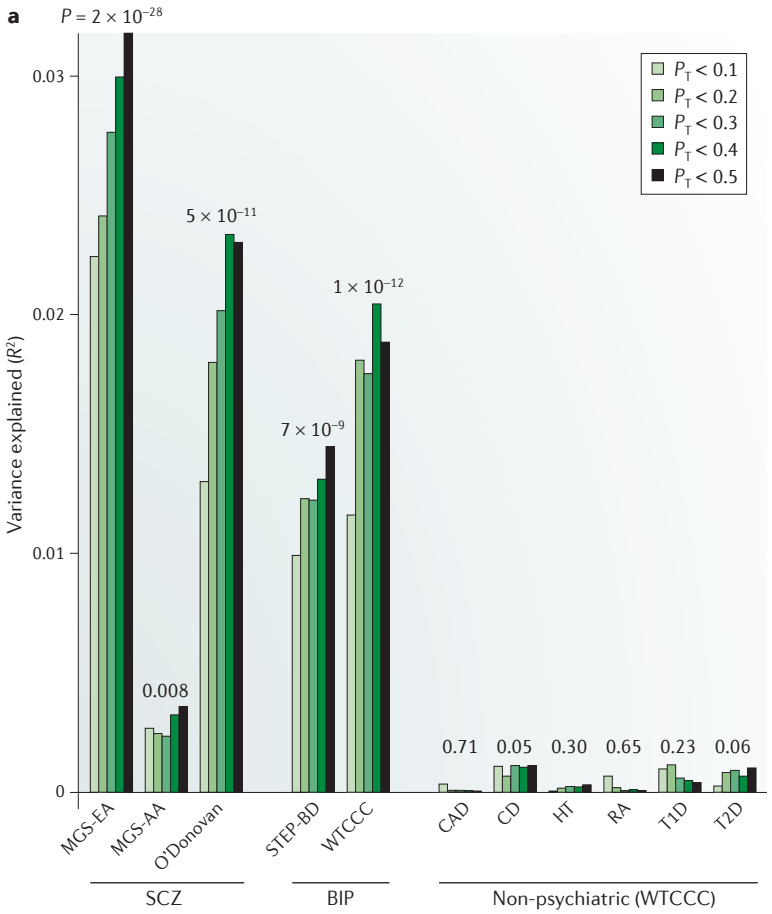
For schizophrenia (SCZ) and bipolar disorder (BIP), sign tests comparing the consistency of association tests from discovery genome-wide association studies (GWASs) and replication samples for sets of top signals are usually highly significant even if most loci do not meet genome-wide significance[59,60,174]. This implies that sample sizes are insufficient and that additional loci can be discovered in larger samples. Tests of the existence of large numbers of true but weakly associated variants have been conducted for SCZ, BIP and many other biomedical disorders.

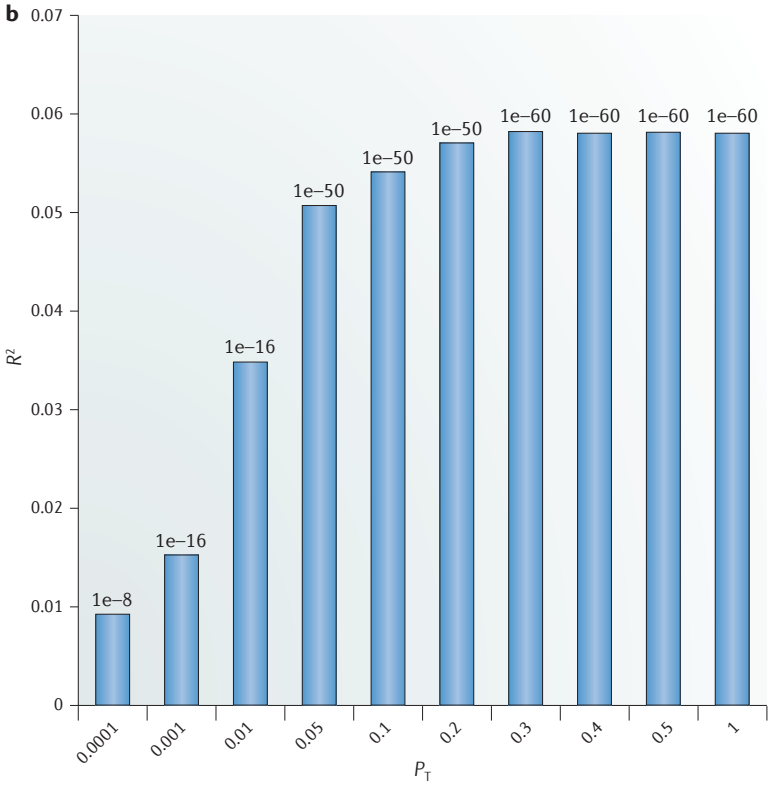
In light of theoretical work by Visscher and colleagues[71], one study used GWAS results as a discovery set (after removing correlated SNPs), and subjects in 11 independent test GWAS data sets were assigned risk profile scores (that is, the number of SCZ risk alleles weighted by their effect sizes in the discovery set). The mean risk profile scores for cases were compared to the mean scores for controls in these independent data sets. Panel **a** of the figure shows risk profile scores for extremely relaxed *P* value thresholds ($P_T < 0.1$, 0.2, 0.3, 0.4 and 0.5, light to dark green bars). Risk profile scores were derived (after linkage-disequilibrium-based SNP pruning) from a discovery SCZ sample and then applied to three independent SCZ samples[174,175], two BIP samples[176,177] and six non-psychiatric diseases[177]. In three independent GWASs, SCZ cases had significantly higher risk profile scores than controls. Remarkably, the same set of markers also discriminated BIP cases from controls, indicating substantial genetic contributions between SCZ and BIP. As an important test of specificity, the SCZ risk profile was not predictive of case status for any of six non-psychiatric diseases[177]. A recent paper evaluated risk profile scores in a trio sample and could exclude population stratification as an explanation[178].

The proportion of variance explained by the risk profile score increased with relaxation of the significance thresholds. This suggests that the discovery sample was insufficiently large to identify many true risk loci at even nominal levels of significance: adding more SNPs contributed more genetic signal than noise. This is partly a feature of sample size. Panel **b** of the figure shows a similar analysis based on a larger discovery sample, in which the proportion of variance explained is approximately double[59], and instead of increasing with P_T the proportion of variance reaches a plateau. If the sample size were truly adequate, the first P_T bin would explain the greatest amount of variance, and relaxing P_T would decrease the proportion of variation explained.

Finally, estimates from two different methods indicated that the risk profile component for SCZ contributes between one-quarter and one-third of the overall variance in liability to SCZ[70,131], a substantial fraction of the 65–81% heritability of SCZ[179,180]. These estimates suggest that 'missing heritability' is merely hidden and imperfectly assayed by current genotyping technologies. CAD, coronary artery disease; CD, Crohn's disease; HT, hypertension; MGS-AA, Molecular Genetics of Schizophrenia — African American; MGS-EA, Molecular Genetics of Schizophrenia — European American; RA, rheumatoid arthritis; STEP-BD, Systematic Treatment Enhancement Program for Bipolar Disorder; T1D, type 1 diabetes; T2D, type 2 diabetes; WTCCC, Wellcome Trust Case Control Consortium.

Data in panel **b** are taken from REF. 59.





that the components of the Fagerstrom test for nicotine dependence (which is a measure of NIC) had heritabilities ranging from fairly high to near zero with important common environmental effects. Other investigators evaluated self-reported lifetime maximum use of ethanol (measured in grams per day) or nicotine (measured in number of cigarettes per day), and such continuous measures of consumption are often available for secondary analysis of samples studied for other diseases.

For ALC, the published GWASs are small, and no large-scale meta-analysis has been conducted[94–98]. In our opinion, there are clear needs for a high-quality meta-analysis and for increasing the number of samples with GWAS data — particularly given that risk profile analysis (BOX 3) suggested that larger samples would yield more associations[98]. For alcohol consumption, GWASs in East Asian samples confirmed the role of aldehyde dehydrogenase 2 (*ALDH2*)[99,100], and autism susceptibility candidate 2 (*AUTS2*) was implicated in alcohol consumption in European subjects[101]. Using a candidate gene approach, the association of alcohol dehydrogenase 1B (*ADH1B*) with ALC and alcohol consumption was extended to European ancestry subjects[102].

For NIC, a field-wide meta-analysis is also needed. For smoking behaviour, large meta-analyses have been conducted[93,103,104]. The strongest finding is an association of smoking quantity with a cluster of nicotinic receptor genes (*CHRNA5–CHRNA3–CHRNB4*) with an effect size that corresponds to one cigarette per day, and there may be several independent associations[105]. Associations to this region have also been reported for lung cancer[106,107]. A recent study showed that *Chrna5*-null mice had higher nicotine intake owing to loss of an inhibitory effect on brain reward systems[108].

Major depressive disorder

The PGC GWAS mega-analysis of 9,240 major depressive disorder (MDD) cases and 9,519 controls (replication in 6,783 MDD cases) revealed no findings of genome-wide significance[109]. These null results are intriguing, as almost all other published GWASs with $N > 11{,}000$ for any disease have found at least one genome-wide significant finding. The most likely reasons for these results are particularly high heterogeneity of MDD and insufficient power arising from its lower heritability[109]. There are few published data on structural variation, although one study found increased structural variation burden in MDD cases versus controls (odds ratio = 1.31)[110].

A provocative finding from 2003 was that risk for MDD might be influenced by a gene–environment interaction with genetic variation near the serotonin transporter[111]. Meta-analyses have supported[112,113] and not supported[114,115] this finding. This association did not replicate in an independent but similar study from the same geographic region, casting particular doubt on the reported association[116].

Other disorders

The published GWASs for attention-deficit hyperactivity disorder (ADHD)[117] and anorexia nervosa[118] are small, but studies using larger sample sizes are currently in progress (for example, by the Wellcome Trust Case-Control Consortium for anorexia nervosa). Given low power, no conclusions about common variation can be made. In ADHD, increased structural variation burden has been reported (odds ratio = 2.1)[119,120], an effect that is higher in ADHD cases with intellectual disability (odds ratio = 5.7)[119]. Pathway analysis in ADHD found association signals enriched in the same Gene Ontology categories also overrepresented for large structural variants[121]. The weak signals in ADHD GWASs are not randomly distributed but index the same pathophysiological pathways as rare structural variants. Thus, it appears that the reason no common variants are yet to be confidently implicated in ADHD by GWASs is a lack of power and not a lack of variants to be found.

What is the emerging picture?

Knowledge of psychiatric genetics is vastly greater than it was 5 years ago. Specifically, there are now multiple high-confidence structural variants (TABLE 2), rare exonic variants (currently only for ASD, Alzheimer's disease and ALC) and an increasing number of robustly significant and replicated common variants (TABLE 3). The data support multiple novel biological hypotheses — for example, cholesterol metabolism and the innate immune response in Alzheimer's disease, a network involving miR-137 for SCZ, calcium signalling for BIP and SCZ, and chromatin remodelling for ASD — and reinforce previous hypotheses, such as synaptic biology for SCZ and ASD.

Genetic architecture. These results also provide insights into genetic architecture that are crucial for planning more complete attempts at the genetic dissection of these major public health conditions. We can now make informed predictions about the types of future studies that can increase understanding in order to generate well-grounded biological hypotheses.

For several disorders, there are now data to replace the interminable debate about the fundamental nature of these illnesses[122]. These occasionally vociferous debates[71] have generally been of an 'either/or' nature: psychiatric disorders as collections of Mendelian-like, single-gene disorders (multiple rare variant models) 'versus' psychiatric disorders as caused by many common variants of small effects (common disease–common variant models)[15,123]. Although we were initially agnostic[124], we now believe that the data support both positions.

For disorders with sufficient data (Alzheimer's disease, BIP and SCZ), the results are consistent with an allelic spectrum and an aetiological role for both rare and common variation. As an example, FIG. 1b synthesizes current knowledge of SCZ as an empirical allelic spectrum map compared with a conceptual schematic from a 2008 Review in this journal[10]. There are no known Mendelian variants, and power analyses can exclude common variants of modest effect (genotypic relative risk >1.5 for allele frequencies >0.1). There are multiple structural variants that are rare, strong but nonspecific risk factors (TABLE 2), and there are 17 common variant associations of subtle effects (TABLE 3). There is an important component arising from common variation in hundreds of

Expression quantitative trait locus
(eQTL). DNA variation that is strongly associated with the expression of a particular mRNA transcript.

Polygenic
Meaning 'many genes'. As a description of genetic architecture, polygenic gives no indications about the frequencies, modes of action or effect sizes of any relevant genetic variation.

different loci (BOX 3), and larger sample sizes are likely to convert many of these to genome-wide significance. The frequency region between 0.001 and 0.05 is under investigation by studies that evaluate exon variation, and more should be known in 2012–2013. Early data for ASD and SCZ suggest that this will not necessarily yield a treasure trove of new findings. At the very least, the 'many Mendelian model' seems to be extremely unlikely. This allelic spectrum map might well be replicated for other psychiatric disorders should larger studies of both rare, uncommon and common variation be achieved.

Hypothesized genetic architectures that consist entirely of rare variants are inconsistent with the data for Alzheimer's disease, ASD, BIP and SCZ (as well as for multiple other complex biomedical diseases)[71–78]. The Procrustean theory that common variant signals inevitably reflect 'synthetic associations'[125] to rare, high-penetrance mutations is not credible[78,126–128].

Psychiatric disorders are polygenic. The evidence is strong that many genes are involved in the aetiology of Alzheimer's disease (currently evidence of rare exonic, rare structural variation and common variation), ALC (currently evidence of common variation), ASD (currently evidence of *de novo* exonic variation and structural variation), BIP (currently evidence of common variation), NIC (currently evidence of common variation) and SCZ (currently evidence of rare structural variation and common variation). Projections for ASD and SCZ suggest that variation at hundreds of different genes will ultimately be shown to be involved[59,81,86]. There are statistical hints that ADHD and MDD might also be polygenic.

Polygenicity may be a general feature of complex biomedical diseases[129,130]. Common variant SNP effects have been estimated to explain large proportions of the phenotypic heritability for a wide range of diseases: BIP and SCZ[70,131]; type 1 diabetes mellitus (T1DM), T2DM, Crohn's disease, rheumatoid arthritis, coeliac disease and coronary artery disease[71,78]; and continuous traits (namely, height, intelligence and body mass)[71,132,133]. These results are consistent with suggestions that the 'missing heritability'[134] is merely hidden[133].

As discussed further below, currently we do not now possess a comprehensive enumeration of loci that are associated with any psychiatric disorder (that is, the 'parts list'), regardless of where genetic variation might lie in the allelic spectrum-effect size space.

Implications and future directions

Why these successes matter. As other commentators have written[71,129,135,136], and as we argued in early 2009 (REF. 124), the proximal purpose of genetic studies is to gain insight into biology. This goal is crucial for psychiatric disorders as so little is known about pathophysiology, and as highly publicized but ultimately false leads have occurred. For this primary goal, there have been unequivocal successes for many psychiatric disorders. This crucial point is sometimes overlooked: the knowledge base in psychiatric genetics is vastly greater than it was 5 years ago, and the rate of change is unprecedented in the history of the field.

What about clinical utility? So-called personalized medicine has been touted as the crucial yardstick against which to measure the success of genetic studies. We believe clinical utility is the ultimate goal but is an inappropriate proximal goal. Still, there are a number of findings for which clinical importance should be evaluated. For example, structural variant testing is often a part of the clinical evaluation of ASD, and careful evaluation of its use in psychosis is warranted. As another example, Perlis and colleagues are undertaking clinical trials on the 'repurposing' of isradipine (an approved antihypertensive that interacts with the protein product of *CACNA1C*) for the treatment of BIP. It is possible that risk profile scores, structural variant burden or rare variant burden could have clinical utility. If these assess latent liability, they might be useful in selected clinical scenarios (for example, for predicting which patients require aggressive treatment in the psychosis prodrome)[137].

The polygenicity of psychiatric disorders poses intriguing difficulties: how can these many genes be coherently tied together? A parsimonious hypothesis is that the polygenic basis of a psychiatric disorder is manifested in the regulation or function of one or more known or novel pathways. Genetic variation at many different loci could introduce numerous slight alterations that result in a pathway that is insufficiently robust in response to an environmental insult or that leads to an inappropriate developmental programme[138]. Risk for a complex psychiatric disorder could be conferred by the emergent properties of the pathway itself rather than any single component. For SCZ, this conceptualization is supported by the risk profile findings for SCZ and by the miR-137 results that hint at an underlying regulatory network. For ASD, typical patterns of cortical gene expression in frontal and temporal cortex have been found to be attenuated in ASD cases compared with controls, and an empirically derived gene expression module that is under-expressed in ASD was found to be enriched for known ASD susceptibility genes and genetic association signals[91].

Alternative modes of investigation, such as network medicine, are needed to further our understanding of the roles of pathways in complex biological traits[139]. If polygenicity is indeed fundamental to complex psychiatric disorders and if some psychiatric disorders eventually prove to be pathway diseases[138], then we need to confront this directly and to develop innovative methods. Developing such methods is more constructive and more likely to advance our understanding of these devastating diseases than raging against nature for not delivering common diseases in simpler Mendelian units.

Indeed, if one or more psychiatric disorders eventually prove to be pathway diseases, there could be clinical benefit. We conjecture that it might be considerably easier to coax an existing but dysfunctional biological pathway into the normal range than to replace components that have been broken by Mendelian mutations. Moreover, in an era in which many drug companies have moved away from drug development for psychiatric disorders[140], the ability to measure such a hypothetical pathway in an appropriate cellular system could enable chemical biology screens of existing and novel compounds, as well as the evaluation of the rational use of multiple compounds simultaneously.

Prodrome
Premonitory symptoms. In this case, a collection of psychotic symptoms that sometimes evolves into schizophrenia or a mood disorder or that resolves without further sequella.

Implications for strategy. A comprehensive portrait of genetic architecture does not now exist for any psychiatric disorder. Gaining a more complete knowledge of the 'parts list' for each disorder — the specific loci that are aetiologically involved plus the identity, frequency and impact of genetic variation at each locus — would be of exceptional importance. Such an enumeration would catalyse an array of specific, targeted and nuanced scientific studies. For example, such studies might lead to the elucidation of biological mechanisms between the genotype and psychiatric phenotype, enablement of cell-based chemical biology and pharmacological screening, evaluation of gene action over developmental time, addressing the important roles of gene–gene and gene–environment interactions, understanding the part played by epigenetic modifications, evaluation of disease prediction models, and so forth.

This is an attainable goal. The genomic search space is large but finite and so, in theory, elucidating the parts list for a psychiatric disorder could be achieved. On the basis of the evidence acquired to date, thorough and well-powered genomic evaluations across the allelic spectrum are needed. We believe that a balanced portfolio of genomic assessments is required, as there are clear roles for common variation, structural variation, rare variation and *de novo* variation for most disorders. Most discoveries in psychiatric genetics to date are from GWASs and structural variant evaluation (both of which are often based on the use of GWAS chips), and larger and more comprehensive GWAS and structural variation studies are highly likely to increase knowledge[136]. It is possible to provide realistic estimates of power and to predict the number of new associations for each increment in sample size[129,130,141] (for example, predictions have been made for GWASs and structural variation studies in 50,000 SCZ cases and 50,000 controls[141]). In light of the recent ASD studies, sequencing that is directed at rare and *de novo* variation will have a role in a balanced portfolio of approaches[84–86]. Indeed, with improvements in accuracy, coverage and pricing, it is possible that sequencing could evolve into the technology of choice for genotyping all major classes of genetic variation.

Such a series of studies would be costly, so a crucial challenge is funding. These costs deserve to be placed in the context of the public health implications of these disorders, and historically psychiatric research has been underfunded in comparison to public health impact (with the possible exception of Alzheimer's disease)[142–145]. For example, the lifetime cost per person with SCZ is on the order of US$1.4 million[146]: if this program of research were eventually able to prevent only several dozen cases, it would probably prove to be cost-effective.

Continued cooperation. The successful study of any type of genetic variation in complex biomedical diseases requires large sample sizes as a means to cut the Gordian knot posed by genetic architecture, aetiological complexity and phenotypic uncertainty. To achieve this end, there have been multiple meta-analysis[147,148] consortia in psychiatric genetics, of which the PGC is the largest and most encompassing[124,149,150]. Indeed, a GWAS co-authorship network graph shows the high connectedness of researchers in the field (Supplementary information S4 (figure)).

An initial concern regarding GWAS meta-analysis was that increased signal from combining multiple samples would be negated by 'noise' owing to inter-site differences. This theoretical concern has not been borne out in practice, as shown above with the examples in SCZ[59], BIP[60], smoking behaviour[93], alcohol consumption[101] and Alzheimer's disease[35,36]. These meta-analyses are designed to identify risk or protective loci that have fairly similar effects across populations and that are not particularly sensitive to sample-specific factors. For example, T2DM and breast cancer loci identified in European samples tend to be replicated in samples of East Asian ancestry[151,152]. It is possible that some genetic variants associated with phenotype risk are only found in certain population groups and are missed in meta-analysis; however, conclusive identification of such loci is likely to be challenging unless the effect sizes are fairly large.

Statistical rigour. In our opinion, a key ingredient of progress in psychiatric genetics has been uncompromising statistical rigour. Genomic technologies routinely posit 10^5–10^8 hypotheses, and false positives are a serious concern. For some investigators, suggestive statistical evidence combined with intriguing biology is sufficient. However, we fear that any benefit from relaxing statistical standards will be outweighed by the negative consequences of false-positive claims.

This issue is particularly salient for exonic variation. Humans carry a huge pool of phenotypically neutral background variation that adds noise to genetic analyses (for example, each person has ~100 loss-of-function variants, most of which are rare in a population)[153,154], and the presence of such variation complicates identification of disease-relevant variants. Thus, owing to chance, a researcher would expect to find a functional exonic mutation — possibly in a gene with intriguing biology — in one case sample and in none of their control samples. More quantitatively, if 1% of cases are caused by fully penetrant mutations in a single gene with no background confounding variation, then observing 10 deleterious mutations in 1,000 cases and 0 deleterious mutations in 1,000 controls would not stand out in test statistics from 20,000 genes. More realistic scenarios (including locus heterogeneity, incomplete penetrance and background variation) will substantially erode the signal. The published results for ASD[84–86] and unpublished data on SCZ suggest that these issues will be important and will underscore the need for sequencing studies to have the same emphasis on statistical rigour and large sample sizes that has enabled GWASs to realize success for multiple psychiatric disorders[155].

Psychiatric genetics is not 'post-genomic'. In psychiatric genetics, we are at the end of the beginning, not the beginning of the end. Remarkably, in a field that is characterized by a checkered history and few confident aetiological clues, the genetics knowledge base has advanced

considerably during the past 5 years, and results to date contain clear indications that further study will yield greater insight. Elucidation of the genetic architectures of psychiatric disorders is an attainable goal with existing technologies (albeit that are both costly and cost-effective). Few predictions are perfectly safe, but we would argue that genetics is a particularly good bet for psychiatry.

1. Eaton, W. W. *et al.* The burden of mental disorders. *Epidemiol. Rev.* **30**, 1–14 (2008).
2. World Health Organization. *The Global Burden of Disease: 2004 Update* (WHO Press, 2008).
3. Collins, P. Y. *et al.* Grand challenges in global mental health. *Nature* **475**, 27–30 (2011).
4. Park, Y. K., Sempos, C. T., Barton, C. N., Vanderveen, J. E. & Yetley, E. A. Effectiveness of food fortification in the United States: the case of pellagra. *Am. J. Public Health* **90**, 727–738 (2000).
5. Centers for Disease Control and Prevention. *Sexually Transmitted Disease Surveillance 2009* (US Department of Health and Human Services, 2010).
6. Chanock, S. J. *et al.* Replicating genotype–phenotype associations. *Nature* **447**, 655–660 (2007).
7. McKusick, V. A. Mendelian Inheritance in Man and its online version, OMIM. *Am. J. Hum. Genet.* **80**, 588–604 (2007).
8. Chiurazzi, P., Schwartz, C. E., Gecz, J. & Neri, G. XLMR genes: update 2007. *Eur. J. Hum. Genet.* **16**, 422–434 (2008).
9. Inlow, J. K. & Restifo, L. L. Molecular and comparative genetics of mental retardation. *Genetics* **166**, 835–881 (2004).
10. McCarthy, M. I. *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Rev. Genet.* **9**, 356–369 (2008).
11. Altshuler, D. & Daly, M. Guilt beyond a reasonable doubt. *Nature Genet.* **39**, 813–815 (2007).
12. Corvin, A., Craddock, N. & Sullivan, P. F. Genome-wide association studies: a primer. *Psychol. Med.* **40**, 1063–1077 (2010).
13. Bassett, A. S., Chow, E. W. & Weksberg, R. Chromosomal abnormalities and schizophrenia. *Am. J. Med. Genet.* **97**, 45–51 (2000).
14. Vorstman, J. A. *et al.* Identification of novel autism candidate regions through analysis of reported cytogenetic abnormalities associated with autism. *Mol. Psychiatry* **11**, 18–28 (2006).
15. Malhotra, D. & Sebat, J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell* **148**, 1223–1241 (2012).
16. Alkan, C., Coe, B. P. & Eichler, E. E. Genome structural variation discovery and genotyping. *Nature Rev. Genet.* **12**, 363–376 (2011).
17. Attia, J. *et al.* How to use an article about genetic association. A: background concepts. *JAMA* **301**, 74–81 (2009).
18. Hindorff, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA* **106**, 9362–9367 (2009).
19. Pasaniuc, B. *et al.* Extremely low-coverage sequencing and imputation increases power for genome-wide association studies. *Nature Genet.* 20 May 2012 (doi:10.1038/ng.2283).
20. Ng, S. B., Nickerson, D. A., Bamshad, M. J. & Shendure, J. Massively parallel sequencing and rare disease. *Hum. Mol. Genet.* **19**, R119–R124 (2010).
21. Cirulli, E. T. & Goldstein, D. B. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature Rev. Genet.* **11**, 415–425 (2010).
22. Collins, A. L. *et al.* Hypothesis-driven candidate genes for schizophrenia compared to genome-wide association results. *Psychol. Med.* **42**, 607–616 (2012).
23. Ng, M. Y. *et al.* Meta-analysis of 32 genome-wide linkage studies of schizophrenia. *Mol. Psychiatry* **14**, 774–785 (2009).
24. McQueen, M. B. *et al.* Combined analysis from eleven linkage studies of bipolar disorder provides strong evidence of susceptibility loci on chromosomes 6q and 8q. *Am. J. Hum. Genet.* **77**, 582–595 (2005).
25. Trikalinos, T. A. *et al.* A heterogeneity-based genome search meta-analysis for autism-spectrum disorders. *Mol. Psychiatry* **11**, 29–36 (2006).
26. Zhou, K. *et al.* Meta-analysis of genome-wide linkage scans of attention deficit hyperactivity disorder. *Am. J. Med. Genet. B* **147**, 1392–1398 (2008).
27. Bertram, L. & Tanzi, R. E. Thirty years of Alzheimer's disease genetics: the implications of systematic meta-analyses. *Nature Rev. Neurosci.* **9**, 768–778 (2008).
28. McNaughton, D. *et al.* Duplication of amyloid precursor protein (*APP*), but not prion protein (*PRNP*) gene is a significant cause of early onset dementia in a large UK series. *Neurobiol. Aging* **33**, 426.e13–426.e21 (2012).
29. Rovelet-Lecrux, A. *et al.* APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nature Genet.* **38**, 24–26 (2006).
30. Guerreiro, R. J. *et al.* Exome sequencing reveals an unexpected genetic cause of disease: *NOTCH3* mutation in a Turkish family with Alzheimer's disease. *Neurobiol. Aging* **33**, 1008.e17–1008.e23 (2012).
31. Pottier, C. *et al.* High frequency of potentially pathogenic *SORL1* mutations in autosomal dominant early-onset Alzheimer disease. *Mol. Psychiatry* 3 Apr 2012 (doi:10.1038/mp.2012.15).
32. Strittmatter, W. J. *et al.* Apolipoprotein E: high-avidity binding to β-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proc. Natl Acad. Sci. USA* **90**, 1977–1981 (1993).
33. Harold, D. *et al.* Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nature Genet.* **41**, 1088–1093 (2009).
34. Lambert, J. C. *et al.* Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nature Genet.* **41**, 1094–1099 (2009).
35. Hollingworth, P. *et al.* Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nature Genet.* **43**, 429–435 (2011).
36. Naj, A. C. *et al.* Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nature Genet.* **43**, 436–441 (2011).
37. Brouwers, N. *et al.* Alzheimer risk associated with a copy number variation in the complement receptor 1 increasing C3b/C4b binding sites. *Mol. Psychiatry* **17**, 223–233 (2012).
38. Jones, L. *et al.* Genetic evidence implicates the immune system and cholesterol metabolism in the aetiology of Alzheimer's disease. *PLoS ONE* **5**, e13950 (2010).
39. Guerreiro, R. J. & Hardy, J. Alzheimer's disease genetics: lessons to improve disease modelling. *Biochem. Soc. Trans.* **39**, 910–916 (2011).
40. Treusch, S. *et al.* Functional links between Aβ toxicity, endocytic trafficking, and Alzheimer's disease risk factors in yeast. *Science* **334**, 1241–1245 (2011).
41. Owen, M. J., Craddock, N. & O'Donovan, M. C. Suggestion of roles for both common and rare risk variants in genome-wide studies of schizophrenia. *Arch. Gen. Psychiatry* **67**, 667–673 (2010).
42. Itsara, A. *et al.* Population analysis of large copy number variants and hotspots of human genetic disease. *Am. J. Hum. Genet.* **84**, 148–161 (2009).
43. Rujescu, D. *et al.* Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum. Mol. Genet.* **18**, 988–996 (2009).
44. Vacic, V. *et al.* Duplications of the neuropeptide receptor gene *VIPR2* confer significant risk for schizophrenia. *Nature* **471**, 499–503 (2011).
45. Kirov, G. *et al. De novo* CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol. Psychiatry* **17**, 142–153 (2011).
46. Raychaudhuri, S. *et al.* Accurately assessing the risk of schizophrenia conferred by rare copy-number variation affecting genes with brain function. *PLoS Genet.* **6**, e1001097 (2010).
47. Walsh, T. *et al.* Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* **320**, 539–543 (2008).
48. International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455**, 237–241 (2008).
49. Sebat, J. *et al.* Strong association of *de novo* copy number mutations with autism. *Science* **316**, 445–449 (2007).
50. Xu, B. *et al.* Strong association of *de novo* copy number mutations with sporadic schizophrenia. *Nature Genet.* **40**, 880–885 (2008).
51. Zhang, D. *et al.* Singleton deletions throughout the genome increase risk of bipolar disorder. *Mol. Psychiatry* **14**, 376–380 (2009).
52. Priebe, L. *et al.* Genome-wide survey implicates the influence of copy number variants (CNVs) in the development of early-onset bipolar disorder. *Mol. Psychiatry* **17**, 421–432 (2011).
53. Malhotra, D. *et al.* High frequencies of *de novo* CNVs in bipolar disorder and schizophrenia. *Neuron* **72**, 951–963 (2011).
54. Grozeva, D. *et al.* Rare copy number variants: a point of rarity in genetic risk for bipolar disorder and schizophrenia. *Arch. Gen. Psychiatry* **67**, 318–327 (2010).
55. McQuillin, A. *et al.* Analysis of genetic deletions and duplications in the University College London bipolar disorder case control sample. *Eur. J. Hum. Genet.* **19**, 588–592 (2011).
56. Girard, S. L. *et al.* Increased exonic *de novo* mutation rate in individuals with schizophrenia. *Nature Genet.* **43**, 860–863 (2011).
57. Xu, B. *et al.* Exome sequencing supports a *de novo* mutational paradigm for schizophrenia. *Nature Genet.* **43**, 864–868 (2011).
58. Need, A. C. *et al.* Exome sequencing followed by large-scale genotyping suggests a limited role for moderately rare risk factors of strong effect in schizophrenia. *Am. J. Hum. Genet.* (in the press).
In the largest study yet reported, the authors report exome resequencing of 166 cases with schizophrenia enriched for strong family histories of illness and treatment non-response. No definitive associations were identified: this intriguing finding makes the 'many Mendelians' model of this disorder far less tenable.
59. Schizophrenia Psychiatric Genome-Wide Association Study Consortium. Genome-wide association study identifies five new schizophrenia loci. *Nature Genet.* **43**, 969–976 (2011).
60. Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature Genet.* **43**, 977–983 (2011).
These two papers are the largest and most comprehensive GWAS mega-analyses conducted for psychotic disorders.
61. Stefansson, H. *et al.* Common variants conferring risk of schizophrenia. *Nature* **460**, 744–747 (2009).
62. Zhong, L., Cherry, T., Bies, C. E., Florence, M. A. & Gerges, N. Z. Neurogranin enhances synaptic strength through its interaction with calmodulin. *EMBO J.* **28**, 3027–3039 (2009).
63. Bemis, L. T. *et al.* MicroRNA-137 targets microphthalmia-associated transcription factor in melanoma cell lines. *Cancer Res.* **68**, 1362–1368 (2008).
64. Kwon, E., Wang, W. & Tsai, L. H. Validation of schizophrenia-associated genes CSMD1, C10orf26, CACNA1C and TCF4 as miR-137 targets. *Mol. Psychiatry* 20 Dec 2011 (doi:10.1038/mp.2011.170).
65. Szulwach, K. E. *et al.* Cross talk between microRNA and epigenetic regulation in adult neurogenesis. *J. Cell Biol.* **189**, 127–141 (2010).
66. Smrt, R. D. *et al.* MicroRNA miR-137 regulates neuronal maturation by targeting ubiquitin ligase mind bomb-1. *Stem Cells* **28**, 1060–1070 (2010).
67. Willemsen, M. H. *et al.* Chromosome 1p21.3 microdeletions comprising DPYD and MIR137 are associated with intellectual disability. *J. Med. Genet.* **48**, 810–818 (2011).
68. Shi, Y. *et al.* Common variants on 8p12 and 1q24.2 confer risk of schizophrenia. *Nature Genet.* **43**, 1224–1227 (2011).
69. Yue, W. H. *et al.* Genome-wide association study identifies a susceptibility locus for schizophrenia in Han Chinese at 11p11.2. *Nature Genet.* **43**, 1228–1231 (2011).

70. International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
71. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90**, 7–24 (2012).
72. Voight, B. F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nature Genet.* **42**, 579–589 (2010).
73. Franke, A. *et al.* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nature Genet.* **42**, 1118–1125 (2010).
74. Houlston, R. S. *et al.* Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nature Genet.* **42**, 973–977 (2010).
75. Turnbull, C. *et al.* Genome-wide association study identifies five new breast cancer susceptibility loci. *Nature Genet.* **42**, 504–507 (2010).
76. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).
77. Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genet.* **42**, 937–948 (2010).
78. Stahl, E. A. *et al.* Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nature Genet.* **44**, 483–489 (2012).
79. Folstein, S. E. & Rosen-Sheidley, B. Genetics of autism: complex aetiology for a heterogeneous disorder. *Nature Rev. Genet.* **2**, 943–955 (2001).
80. Betancur, C. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Res.* **1380**, 42–77 (2011).
81. Sanders, S. J. *et al.* Multiple recurrent *de novo* CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* **70**, 863–885 (2011).
There have been many excellent studies of CNVs in autism. This is one of the largest and is particularly technically well done. The authors report compelling findings regarding specific CNVs and a greater burden of rare CNVs in cases with ASD.
82. Pinto, D. *et al.* Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466**, 368–372 (2010).
83. Marshall, C. R. *et al.* Structural variation of chromosomes in autism spectrum disorder. *Am. J. Hum. Genet.* **82**, 477–488 (2008).
84. Neale, B. M. *et al.* Patterns and rates of exonic *de novo* mutations in autism spectrum disorders. *Nature* **485**, 242–245 (2012).
85. O'Roak, B. J. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. *Nature* **485**, 246–250 (2012).
86. Sanders, S. J. *et al. De novo* mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012).
These three papers report exome sequencing of trios with ASD. Support was found for 2–4 *de novo* mutations that are plausibly causal. However, as for schizophrenia (reference 58), an abundance of Mendelian variants was not identified.
87. Wang, K. *et al.* Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* **459**, 528–533 (2009).
88. Weiss, L. A., Arking, D. E., Daly, M. J. & Chakravarti, A. A genome-wide linkage and association scan reveals novel loci for autism. *Nature* **461**, 802–808 (2009).
89. Anney, R. *et al.* A genome-wide scan for common alleles affecting risk for autism. *Hum. Mol. Genet.* **19**, 4072–4082 (2010).
90. Ma, D. *et al.* A genome-wide association study of autism reveals a common novel risk locus at 5p14.1. *Ann. Hum. Genet.* **73**, 263–273 (2009).
91. Voineagu, I. *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380–384 (2011).
92. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 1994).
93. Tobacco and Genetics Consortium. Meta-analyses of genome-wide association studies implicate multiple loci for smoking behavior. *Nature Genet.* **42**, 441–447 (2010).
94. Edenberg, H. J. *et al.* Genome-wide association study of alcohol dependence implicates a region on chromosome 11. *Alcohol. Clin. Exp. Res.* **34**, 840–852 (2010).
95. Bierut, L. J. *et al.* A genome-wide association study of alcohol dependence. *Proc. Natl Acad. Sci. USA* **107**, 5082–5087 (2010).
96. Kendler, K. S. *et al.* Genomewide association analysis of symptoms of alcohol dependence in the molecular genetics of schizophrenia (MGS2) control sample. *Alcohol. Clin. Exp. Res.* **35**, 963–975 (2011).
97. Heath, A. C. *et al.* A quantitative-trait genome-wide association study of alcoholism risk in the community: findings and implications. *Biol. Psychiatry* **70**, 513–518 (2011).
98. Frank, J. *et al.* Genome-wide significant association between alcohol dependence and a variant in the ADH gene cluster. *Addict. Biol.* **17**, 171–180 (2012).
99. Baik, I., Cho, N. H., Kim, S. H., Han, B. G. & Shin, C. Genome-wide association studies identify genetic loci related to alcohol consumption in Korean men. *Am. J. Clin. Nutr.* **93**, 809–816 (2011).
100. Takeuchi, F. *et al.* Confirmation of ALDH2 as a major locus of drinking behavior and of its variants regulating multiple metabolic phenotypes in a Japanese population. *Circ. J.* **75**, 911–918 (2011).
101. Schumann, G. *et al.* Genome-wide association and genetic functional studies identify autism susceptibility candidate 2 gene (AUTS2) in the regulation of alcohol consumption. *Proc. Natl Acad. Sci. USA* **108**, 7119–7124 (2011).
102. Bierut, L. J. *et al.* ADH1B is associated with alcohol dependence and alcohol consumption in populations of European and African ancestry. *Mol. Psychiatry* **7**, 445–450 (2011).
103. Thorgeirsson, T. E. *et al.* Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. *Nature Genet.* **42**, 448–453 (2010).
104. Liu, J. Z. *et al.* Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nature Genet.* **42**, 436–440 (2010).
105. Saccone, N. L. *et al.* Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a meta-analysis and comparison with lung cancer and COPD. *PLoS Genet.* **6**, e1001053 (2010).
106. Thorgeirsson, T. E. *et al.* A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* **452**, 638–642 (2008).
107. Truong, T. *et al.* Replication of lung cancer susceptibility loci at chromosomes 15q25, 5p15, and 6p21: a pooled analysis from the International Lung Cancer Consortium. *J. Natl Cancer Inst.* **102**, 959–971 (2010).
108. Fowler, C. D., Lu, Q., Johnson, P. M., Marks, M. J. & Kenny, P. J. Habenular α5 nicotinic receptor subunit signalling controls nicotine intake. *Nature* **471**, 597–601 (2011).
109. Major Depressive Disorder Working Group of the PGC. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol. Psychiatry* 3 Apr 2012 (doi:10.1038/mp.2012.21).
Despite a heritability and lifetime prevalence that is not dissimilar to T2DM and in contrast to similarly sized studies of BIP and SCZ, this study of MDD is largely negative. Higher heterogeneity is a prime candidate for the divergent results.
110. Rucker, J. J. *et al.* Genome-wide association analysis of copy number variation in recurrent depressive disorder. *Mol. Psychiatry* 1 Nov 2011 (doi:10.1038/mp.2011.144).
111. Caspi, A. *et al.* Influence of life stress on depression: moderation by a polymorphism in the *5-HTT* gene. *Science* **301**, 386–389 (2003).
112. Karg, K., Burmeister, M., Shedden, K. & Sen, S. The serotonin transporter promoter variant (*5-HTTLPR*), stress, and depression meta-analysis revisited: evidence of genetic moderation. *Arch. Gen. Psychiatry* **68**, 444–454 (2011).
113. Uher, R. & McGuffin, P. The moderation by the serotonin transporter gene of environmental adversity in the etiology of depression: 2009 update. *Mol. Psychiatry* **15**, 18–22 (2010).
114. Munafo, M. R., Durrant, C., Lewis, G. & Flint, J. Gene X environment interactions at the serotonin transporter locus. *Biol. Psychiatry* **65**, 211–219 (2009).
115. Risch, N. *et al.* Interaction between the serotonin transporter gene (*5-HTTLPR*), stressful life events, and risk of depression: a meta-analysis. *JAMA* **301**, 2462–2471 (2009).
116. Fergusson, D. M., Horwood, L. J., Miller, A. L. & Kennedy, M. A. Life stress, *5-HTTLPR* and mental disorder: findings from a 30-year longitudinal study. *Br. J. Psychiatry* **198**, 129–135 (2011).
117. Neale, B. M. *et al.* Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *J. Am. Acad. Child Adolesc. Psychiatry* **49**, 884–897 (2010).
118. Wang, K. *et al.* A genome-wide association study on common SNPs and rare CNVs in anorexia nervosa. *Mol. Psychiatry* **16**, 949–959 (2011).
119. Williams, N. M. *et al.* Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. *Lancet* **376**, 1401–1408 (2010).
120. Williams, N. M. *et al.* Genome-wide analysis of copy number variants in attention deficit hyperactivity disorder: the role of rare variants and duplications at 15q13.3. *Am. J. Psychiatry* **169**, **195**–204 (2012).
121. Stergiakouli, E. *et al.* Investigating the contribution of common genetic variants to the risk and pathogenesis of ADHD. *Am. J. Psychiatry* **169**, 186–194 (2012).
122. Wray, N. R. & Visscher, P. M. Narrowing the boundaries of the genetic architecture of schizophrenia. *Schizophr. Bull.* **36**, 14–23 (2010).
123. McClellan, J. & King, M. C. Genomic analysis of mental illness: a changing landscape. *JAMA* **303**, 2523–2524 (2010).
124. Psychiatric GWAS Consortium. A framework for interpreting genomewide association studies of psychiatric disorders. *Mol. Psychiatry* **14**, 10–17 (2009).
125. Dickson, S. P., Wang, K., Krantz, I., Hakonarson, H. & Goldstein, D. B. Rare variants create synthetic genome-wide associations. *PLoS Biol.* **8**, e1000294 (2010).
126. Wray, N. R., Purcell, S. M. & Visscher, P. M. Synthetic associations created by rare variants do not explain most GWAS results. *PLoS Biol.* **9**, e1000579 (2011).
127. Anderson, C. A., Soranzo, N., Zeggini, E. & Barrett, J. C. Synthetic associations are unlikely to account for many common disease genome-wide association signals. *PLoS Biol.* **9**, e1000580 (2011).
128. Orozco, G., Barrett, J. C. & Zeggini, E. Synthetic associations in the context of genome-wide association scan signals. *Hum. Mol. Genet.* **19**, R137–R144 (2010).
129. Lander, E. S. Initial impact of the sequencing of the human genome. *Nature* **470**, 187–197 (2011).
130. Kim, Y., Zerwas, S., Trace, S. E. & Sullivan, P. F. Schizophrenia genetics: where next? *Schizophr. Bull.* **37**, 456–463 (2011).
131. Lee, S. *et al.* Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nature Genet.* **44**, 247–250 (2012).
This is a provocative paper showing that liability to SCZ is substantial and distributed among genetic variants across the common variant spectrum and across the genome. The authors suggest that the so-called missing heritability problem is hidden and that the architecture of SCZ is polygenic.
132. Davies, G. *et al.* Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Mol. Psychiatry* **16**, 996–1005 (2011).
133. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nature Genet.* **42**, 565–569 (2010).
134. Eichler, E. E. *et al.* Missing heritability and strategies for finding the underlying causes of complex disease. *Nature Rev. Genet.* **11**, 446–450 (2010).
135. Hirschhorn, J. N. Genomewide association studies—illuminating biologic pathways. *N. Engl. J. Med.* **360**, 1699–1701 (2009).
136. Sullivan, P. F. Don't give up on GWAS. *Mol. Psychiatry* **17**, 2–3 (2012).
137. Addington, J. & Heinssen, R. Prediction and prevention of psychosis in youth at clinical high risk. *Annu. Rev. Clin. Psychol.* **8**, 269–289 (2012).
138. Sullivan, P. F. Schizophrenia as a pathway disease. *Nature Med.* **18**, 210–211 (2012).
139. Barabasi, A. L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nature Rev. Genet.* **12**, 56–68 (2011).
140. Miller, G. Is pharma running out of brainy ideas? *Science* **329**, 502–504 (2010).

141. Yang, J., Wray, N. R. & Visscher, P. M. Comparing apples and oranges: equating the power of case–control and quantitative trait association studies. *Genet. Epidemiol.* **34**, 254–257 (2010).
142. Gross, C. P., Anderson, G. F. & Powe, N. R. The relation between funding by the National Institutes of Health and the burden of disease. *New Engl. J. Med.* **340**, 1881–1887 (1999).
143. Aoun, S., Pennebaker, D. & Pascal, R. To what extent is health and medical research funding associated with the burden of disease in Australia? *Aust. N. Z. J. Public Health* **28**, 80–86 (2004).
144. Gillum, L. A. *et al.* NIH disease funding levels and burden of disease. *PLoS ONE* **6**, e16837 (2011).
145. Lamarre-Cliche, M., Castilloux, A. M. & LeLorier, J. Association between the burden of disease and research funding by the Medical Research Council of Canada and the National Institutes of Health. A cross-sectional study. *Clin. Invest. Med.* **24**, 83–89 (2001).
146. Wu, E. Q. *et al.* The economic burden of schizophrenia in the United States in 2002. *J. Clin. Psychiatry* **66**, 1122–1129 (2005).
147. Cantor, R. M., Lange, K. & Sinsheimer, J. S. Prioritizing GWAS results: a review of statistical methods and recommendations for their application. *Am. J. Hum. Genet.* **86**, 6–22 (2010).
148. de Bakker, P. I., Neale, B. M. & Daly, M. J. Meta-analysis of genome-wide association studies. *Cold Spring Harb. Protoc.* **2010**, pdb.top81 (2010).
149. Psychiatric GWAS Consortium. Genome-wide association studies: history, rationale, and prospects for psychiatric disorders. *Am. J. Psychiatry* **166**, 540–546 (2009).
150. Sullivan, P. F. The Psychiatric GWAS Consortium: big science comes to psychiatry. *Neuron* **68**, 182–186 (2010).
151. Waters, K. M. *et al.* Consistent association of type 2 diabetes risk variants found in Europeans in diverse racial and ethnic groups. *PLoS Genet.* **6**, e1001078 (2010).
152. Easton, D. F. *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087–1093 (2007).
153. MacArthur, D. G. *et al.* A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**, 823–828 (2012).
154. Durbin, R. M. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
155. Kiezun, A. *et al.* Exome sequencing and the genetic basis of complex traits. *Nature Genet.* **44**, 623–630 (2012).
156. de Bakker, P. I. *et al.* Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum. Mol. Genet.* **17**, R122–R128 (2008).
157. Ott, J. *Analysis of Human Genetic Linkage* (Johns Hopkins Univ. Press, 1999).
158. Weiss, L. A. *et al.* Association between microdeletion and microduplication at 16p11.2 and autism. *N. Engl. J. Med.* **358**, 667–675 (2008).
159. Szatmari, P. *et al.* Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nature Genet.* **39**, 319–328 (2007).
160. Kumar, R. A. *et al.* Recurrent 16p11.2 microdeletions in autism. *Hum. Mol. Genet.* **17**, 628–638 (2008).
161. Holmans, P. *et al.* Gene ontology analysis of GWA study data sets provides insights into the biology of bipolar disorder. *Am. J. Hum. Genet.* **85**, 13–24 (2009).
162. Lee, P., O'Dushlaine, C., Thomas, B. & Purcell, S. INRICH: interval-based enrichment analysis for genome-wide association studies. *Bioinformatics* 17 Apr 2012 (doi:10.1093/bioinformatics/bts191).
163. Rossin, E. J. *et al.* Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet.* **7**, e1001273 (2011).
164. Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.* **5**, e1000534 (2009).
165. Khatri, P., Sirota, M. & Butte, A. J. Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput. Biol.* **8**, e1002375 (2012).
166. Yaspan, B. L. & Veatch, O. J. Strategies for pathway analysis from GWAS data. *Curr. Protoc. Hum. Genet.* **71**, 1.20.1–1.20.15 (2011).
167. GO Project. The Gene Ontology: enhancements for 2011. *Nucleic Acids Res.* **40**, D559–D564 (2012).
168. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40**, D109–D114 (2012).
169. Mi, H. *et al.* PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Res.* **38**, D204–D210 (2010).
170. Croning, M. D., Marshall, M. C., McLaren, P., Armstrong, J. D. & Grant, S. G. G2Cdb: the Genes to Cognition database. *Nucleic Acids Res.* **37**, D846–D851 (2009).
171. Ruano, D. *et al.* Functional gene group analysis reveals a role of synaptic heterotrimeric G proteins in cognitive ability. *Am. J. Hum. Genet.* **86**, 113–125 (2010).
172. Lindblad-Toh, K. *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476–482 (2011).
173. Nicolae, D. L. *et al.* Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* **6**, e1000888 (2010).
174. O'Donovan, M. *et al.* Identification of novel schizophrenia loci by genome-wide association and follow-up. *Nature Genet.* **40**, 1053–1055 (2008).
175. Shi, J. *et al.* Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* **460**, 753–757 (2009).
176. Sklar, P. *et al.* Whole-genome association study of bipolar disorder. *Mol. Psychiatry* **13**, 558–569 (2008).
177. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
178. Ruderfer, D. M. *et al.* A family-based study of common polygenic variation and risk of schizophrenia. *Mol. Psychiatry* **16**, 887–888 (2011).
179. Sullivan, P. F., Kendler, K. S. & Neale, M. C. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch. Gen. Psychiatry* **60**, 1187–1192 (2003).
180. Lichtenstein, P. *et al.* Common genetic influences for schizophrenia and bipolar disorder: a population-based study of 2 million nuclear families. *Lancet* **373**, 234–239 (2009).
181. Ropers, H. H. Genetics of early onset cognitive impairment. *Annu. Rev. Genom. Hum. Genet.* **11**, 161–187 (2010).
182. Arcelus, J., Mitchell, A. J., Wales, J. & Nielsen, S. Mortality rates in patients with anorexia nervosa and other eating disorders. A meta-analysis of 36 studies. *Arch. Gen. Psychiatry* **68**, 724–731 (2011).
183. Biederman, J. & Faraone, S. V. The effects of attention-deficit/hyperactivity disorder on employment and household income. *MedGenMed.* **8**, 12 (2006).
184. Levinson, D. F. *et al.* Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. *Am. J. Psychiatry* **168**, 302–316 (2011).
This is an excellent large study and meta-analysis of copy number variation in SCZ.
185. Mefford, H. C. *et al.* Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *New Engl. J. Med.* **359**, 1685–1699 (2008).
186. Stefansson, H. *et al.* Large recurrent microdeletions associated with schizophrenia. *Nature* **455**, 232–236 (2008).
187. Brunetti-Pierri, N. *et al.* Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nature Genet.* **40**, 1466–1471 (2008).
188. Cooper, G. M. *et al.* A copy number variation morbidity map of developmental delay. *Nature Genet.* **43**, 838–846 (2011).
189. Mefford, H. C. *et al.* Genome-wide copy number variation in epilepsy: novel susceptibility loci in idiopathic generalized and focal epilepsies. *PLoS Genet.* **6**, e1000962 (2010).
190. Jacquemont, S. *et al.* Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. *Nature* **478**, 97–102 (2011).
191. Shinawi, M. *et al.* Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size. *J. Med. Genet.* **47**, 332–341 (2010).
192. Moreno-De-Luca, D. *et al.* Deletion 17q12 is a recurrent copy number variant that confers high risk of autism and schizophrenia. *Am. J. Hum. Genet.* **87**, 618–630 (2010).
193. Seshadri, S. *et al.* Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA* **303**, 1832–1840 (2010).
194. Cichon, S. *et al.* Genome-wide association study identifies genetic variation in neurocan as a susceptibility factor for bipolar disorder. *Am. J. Hum. Genet.* **88**, 372–381 (2011).
195. Steinberg, S. *et al.* Common variants at *VRK2* and *TCF4* conferring risk of schizophrenia. *Hum. Mol. Genet.* **20**, 4076–4081 (2011).
196. Williams, H. J. *et al.* Fine mapping of ZNF804A and genome-wide significant evidence for its involvement in schizophrenia and bipolar disorder. *Mol. Psychiatry* **16**, 429–441 (2010).
197. Rietschel, M. *et al.* Association between genetic variation in a region on chromosome 11 and schizophrenia in large samples from Europe. *Mol. Psychiatry* 12 July 2011 (doi:10.1038/mp.2011.80).
198. Pe'er, I., Yelensky, R., Altshuler, D. & Daly, M. J. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet. Epidemiol.* **32**, 381–385 (2008).
199. Ferreira, M. *et al.* Collaborative genome-wide association analysis of 10,596 individuals supports a role for Ankyrin-G (*ANK3*) and the ⊠-1C subunit of the L-type voltage-gated calcium channel (*CACNA1C*) in bipolar disorder. *Nature Genet.* **40**, 1056–1058 (2008).
200. McMahon, F. J. *et al.* Meta-analysis of genome-wide association data identifies a risk locus for major mood disorders on 3p21.1. *Nature Genet.* **42**, 128–131 (2010).
201. Breen, G. *et al.* Replication of association of 3p21.1 with susceptibility to bipolar disorder but not major depression. *Nature Genet.* **43**, 3–5 (2011).
202. Green, E. K. *et al.* The bipolar disorder risk allele at CACNA1C also confers risk of recurrent major depression and of schizophrenia. *Mol. Psychiatry* **15**, 1016–1022 (2010).

Acknowledgements

We remain indebted to the tens of thousand of individuals with psychiatric disorders for their participation in genetic studies of these disorders. We thank many colleagues for helpful critiques, plus a wealth of conversations with colleagues in the Psychiatric Genomics Consortium. We thank the US National Institutes of Health (NIH) for funding (MH077139 and MH085520), T. Lehner of the US National Institute of Mental Health (NIMH) for his continued support, B. Voight and J. Hirschhorn for data relevant to power estimation, S. Ripke for help with figures, J. Szatkiewicz for assistance in reviewing the structural variation literature, and C. Bulik and J. Crowley for helpful comments. M.O. is supported for schizophrenia research by grants from the UK Medical Research Council (MRC) (G0800509, G0801418), the European Community's Seventh Framework Programme (HEALTH-F2-2010-241909 (Project EU-GEI)) and by NIMH (5P50MH066392-09). This study makes use of data generated by DECIPHER. A full list of centres that contributed to the generation of the data is available from http://decipher.sanger.ac.uk. Funding for DECIPHER was provided by the Wellcome Trust. The authors jointly conceived and wrote this paper and take responsibility for its content.

Competing interests statement

The authors declare no competing financial interests.

FURTHER INFORMATION

Patrick F. Sullivan's homepage: http://sullivanlab.unc.edu
Mark J. Daly's homepage: http://atgu.mgh.harvard.edu
Michael O'Donovan's homepage (at University of Cardiff): http://medicine.cf.ac.uk/person/prof-michael-conlon-odonovan
Michael O'Donovan's profile at the National Centre for Mental Health: http://ncmh.info/about-us/our-team/principal-investigators/michael-odonovan
DECIPHER: http://decipher.sanger.ac.uk
***Nature Reviews Genetics* Series on Disease mechanisms:** http://www.nature.com/nrg/series/disease
Psychiatric Genomics Consortium: http://pgc.unc.edu
Ricopili (a tool for visualizing regions of interest in select GWAS data sets): http://www.broadinstitute.org/mpg/ricopili

SUPPLEMENTARY INFORMATION

See online article: S1 (table) | S2 (table) | S3 (figure) | S4 (figure)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF