Michael C O'Donovan* and Michael J Owen

The implications of the shared genetics of psychiatric disorders

MRC Centre for Neuropsychiatric Genetics and Genomics,
Division of Psychiatry and Clinical Neurosciences
Cardiff University School of Medicine
Cardiff, UK
CF24 4HQ
*corresponding author; email: odonovanmc@cardiff.ac.uk

Abstract

Recent genomic studies have revealed the highly polygenic nature of psychiatric disorders including schizophrenia, bipolar disorder, and major depressive disorder. Many of the individual genetic associations are shared across multiple disorders in a way that points to extensive biological pleiotropy, and further challenges the biological validity of existing diagnostic approaches. Here, we argue it is unlikely that risk alleles exist that are specific to a single diagnostic category. We also highlight some of the important clinical repercussions of pleiotropy.

Introduction

Psychiatric disorders represent 21st Century Medicine's greatest global challenge. They are the major cause, worldwide, of non-fatal burden of disease[1]. They account for around 30% of all years lived with disability, a contribution that is rising, especially in developing countries as the burden moves from communicable to non-communicable diseases[1]. With a life time prevalence greater than 10%, major depressive disorder accounts for a higher burden of disability than any other disorder, while schizophrenia, which contributes less to global burden due to its lower prevalence (around 1%), is the most severely disabling of all medical conditions[2]. It is stating the obvious that we need to develop and deliver more effective psychiatric prevention and treatment, but despite years of effort there have been few

significant advances. There are a number of reasons for this, but prominent among them has been our lack of understanding of aetiology and pathogenesis, compounded by our reliance on observational and syndromic systems of diagnosis and classification.

Concerns about psychiatric diagnosis and classification have been thrown into sharp relief by recent genomic studies that appear to show that risk alleles tend not to be specific to any particular disorder. In this perspective, we discuss the nature and extent of the evidence for shared risk alleles across psychiatric disorders and interpret that evidence within the context of how psychiatric diagnoses are made. We consider whether it is likely that risk variants for specific disorders exist, and how future studies might usefully illuminate alternative genotype-phenotype relationships. We also consider some of the clinical implications emerging from pleiotropy. The focus of our discussion is the major psychiatric disorders such as schizophrenia, major mood disorders, autism spectrum disorders (ASD) and attention deficit hyperactivity disorder (ADHD), disorders for which research has faced conceptual and practical challenges that place them apart from dementias and from acute disturbances of mental function that are secondary to trauma, toxicity, or medical conditions.

Psychiatric Diagnosis

The Making of Psychiatric Diagnosis

Psychiatric diagnoses are made on the basis of patient descriptions of their subjective experiences (e.g. energy, mood, perception, beliefs, appetite), and from observations of behaviour (e.g. bizarre activity, attention, self-care, social interaction) made by clinicians or reported by informants (e.g. family or carers, neighbours, teachers). Other factors are taken into consideration including functional impairment, developmental trajectory, and outcome. Ultimately a diagnosis is assigned to individuals who exhibit a minimum number of symptoms, behaviours, or outcomes, usually for a minimum period of time, with the proviso they do not meet criteria that exclude that diagnosis. The exclusion criteria are often subjective, requiring clinicians to judge that the clinical picture is not ‘better accounted for’ by another diagnosis, or that the picture is not ‘clearly caused by’ the effects of a psychoactive agent for example. In clinical practice, experience and intuition play a role, although semi-standardized data acquisition tools and operationalized diagnostic criteria have been developed to minimize the impact of these subjective factors. These are used primarily in research, but they are sometimes employed to aid diagnosis in the clinic (e.g. the Autism Diagnostic Observation Schedule).

The Validity of Psychiatric Diagnosis

As research data have accumulated it has become clear that the boundaries between diagnostic groups and between illness and wellness are not clear-cut, there is considerable heterogeneity within diagnostic categories, patients often have the clinical features of more than one disorder[3], and the preponderance of those features in a particular individual can change markedly over time and development. Even with the most fastidious application of diagnostic criteria, there is no avoiding the fact that none of the clinical features are pathognomonic. For example, the occurrence of psychotic symptoms such as hallucinations and delusions, mood changes, and alterations in speech, activity level, behaviour and sleep, can indicate either a diagnosis of schizophrenia or bipolar disorder (BD). The frequent co-occurrence of symptoms that could imply either major diagnostic label has led to a third category, schizoaffective disorder. Archetypal versions of each diagnosis exist, but for a large number of people, the distinction is based on relatively subjective judgments about the duration, quality, and severity of component signs and symptoms[4]. Outcome within diagnostic groups also varies widely, for example some people with a diagnosis of schizophrenia remain chronically symptomatic and impaired while others make a complete recovery[5]. Finally, as better-powered epidemiological studies have been carried out, it has become clear that the relatives of an individual with one psychiatric diagnosis are at increased risk for other diagnoses, undermining the genetic validity of current diagnostic approaches[6].

Biomarkers

There have been extensive efforts to identify biomarkers indexing pathogenic mechanisms, including studies of blood markers (e.g. metabolites, cytokines, cortisol suppression), behavioural and cognitive measures, and various neuroimaging modalities[7]. However, this work has failed to deliver markers that can reliably distinguish between diagnoses, or to identify disease subgroups and, currently, there are no biomarkers in routine clinical use. For example, despite the extensive use of ever more sophisticated neuroimaging approaches, no measures have emerged that can separate people with a particular diagnosis from healthy individuals, much less distinguish between those with different diagnoses[8].

Molecular Genetic Findings in Psychiatry

The robust identification of risk factors for psychiatric disease as indicated by DNA variation has been eagerly awaited for the insights this might provide into the basic biological architecture of, and relationships between, psychiatric phenotypes, as well as for its

contributions to understanding disease mechanisms. In the last few years, genomic studies have begun to identify risk alleles in large numbers; although success has largely been confined to ASD, schizophrenia, and, to a lesser extent, BD and Major Depressive Disorder (MDD). Other psychiatric phenotypes have yet to be subjected to large-scale genome-wide studies.

In ASD, the evidence implicating specific risk genes comes primarily from mutations that occur *de novo* in the form of large insertion-deletion mutations called copy number variants (CNVs) or rare coding variants (RCVs) that change the DNA sequence at a single, or a few, nucleotides. A recent synthesis of the ASD data (5,563 cases for *de novo* RCVs, 4687 cases for CNVs) reported high confidence associations to 65 genes and an additional 6 CNV loci[9]. All loci identified thus far confer large effects on risk, but with population frequencies less than one in a thousand, however this might simply reflect low power to detect smaller effect sizes. It should also be noted that there is emerging evidence that common genetic variation makes a substantial contribution to the variance in liability to ASD[10] although, individual common alleles have not yet been robustly implicated.

In schizophrenia, identified risk alleles span the full spectrum of frequencies. The largest analysis of genome-wide association (GWAS) data (up to 36,989 cases and 113,075 controls) identified a total of 108 loci containing common alleles while that of rare CNVs (12,029-21,269 cases; 24,815- 81,821 controls) identified 11 strongly supported loci[12]. The latter was largely based on a meta-analysis of candidate CNVs, and a systematic genome wide CNV meta-analysis is awaited. Exome sequencing studies in schizophrenia have been smaller than those in ASD, and the evidence for RCVs is largely restricted to enrichments in pathways rather than specific genes[13,14], although recently, a meta-analysis (4,264 schizophrenia cases, 9,343 controls, 1,077 parent-proband trios) obtained genome-wide significant association between schizophrenia and Loss-of-Function (LoF) RCVs in a gene which encodes the histone methytransferase SETD1A[15]. That study also reported a specific mutation in *SETD1A* that occurred in people with the disorder as a *de novo* mutation at a frequency far in excess of that expected by chance, providing confidence for pathogenicity of that specific mutation.

In BD, GWAS and rare variant datasets are smaller than those of schizophrenia. The largest GWAS study (9,747 patients and 14,278 controls) identified 5 risk loci while, at the rare variant end of the spectrum, the only finding that meets a statistical threshold equivalent to genome wide significance is a duplication CNV at 16p.11.2[16]. Finally, a recent GWAS[17]

based upon MDD, as self-reported by customers of a consumer genetics company, identified 15 loci for the disorder. It is particularly notable that only 15 loci were identified in a study including up to 130,620 cases and 347,620 controls. This underscores the fact that, while sample size may be critical for discovery genetics, it is not the only factor. Differences in other properties of disorders (e.g. disease prevalence, heterogeneity, phenotype definition, variance in risk contributed by individual alleles) can have a major impact.

Pleiotropy

The Nature of Pleiotropy

The meaning of pleiotropy (Figure 1) depends on context[18,19]. We refer to genic pleiotropy when altered function of a gene influences multiple traits (note the term trait includes phenotypes that are not necessarily abnormal or disorders). Allelic pleiotropy, a subtype of genic pleiotropy, occurs when the same gene variant influences multiple traits. This is exemplified by phenylketonuria (PKU) in which causative mutations are pleiotropic for intellectual disability, lack of pigmentation, as well as various metabolic changes that can be measured in the blood. These two forms of biological pleiotropy, genic and allelic, suggest shared biology between disorders, but this is not the only explanation.

Mediated pleiotropy occurs when an allele influences two traits, but its effects on one are secondary to more direct effects on the other. For example genetic variation at the *fat mass and obesity associated* (*FTO*) locus is pleiotropic for body mass index (BMI) and type 2 diabetes (T2D), but the effects on T2D are secondary to those on BMI. In the case of PKU, the effects on intellectual function and pigmentation are mediated by the effects on the metabolic traits. As in these examples, mediated pleiotropy can be informative for understanding causal pathways to disease and, as we shall see, is often implicitly assumed in endophenotype studies, but the mediating relationship between the two traits can be complex and it does not necessarily imply that the two phenotypes share biological mechanisms.

There are also numerous sources of false or pseudo pleiotropy. Pseudo pleiotropy can arise as a result of imprecision in gene mapping where two phenotypes are influenced by different genes in close proximity (Figure 1) but it can also arise from poor study design, associations that are due to chance (type II errors), and publication biases favouring reports of overlaps.

Pleiotropy in Psychiatry and Developmental Disorders

Evidence for cross disorder effects of genetic variation has come from studies showing that CNVs that influence risk for schizophrenia also often do so for ASD, intellectual disability (ID) developmental delay (Figure 2), and ADHD[20]. The majority of these apparently pleiotropic CNVs are multigenic, and therefore we cannot exclude pseudo pleiotropy in which distinct genes within the CNV cause each associated phenotype (Figure 1). However, the observation that every CNV known to increase risk of schizophrenia also does so for ID[21] makes co-localization alone an unlikely explanation. Moreover, the only 'single gene' CNV that is unequivocally associated with schizophrenia, deletion of the gene *NRXN1* encoding the pre-synaptic protein neurexin 1 is also associated with ASD and with ID[22]. Sequencing studies have shown that as a group, genes impacted by LoF *de novo* mutations in schizophrenia are enriched for those affected by this same class of mutation in people with ASD and ID[13]. Moreover, several genes have been definitively implicated by *de novo* LoF mutations in each of developmental delay and ASD[9,23] , while at an even finer level of resolution, the same LoF mutation in *SETD1A* that contributes high risk to schizophrenia also does so for severe ID and developmental delay[15].

The hypothesis of true pleiotropy in psychiatric and developmental disorders is also supported by common variants identified by GWAS. The International Schizophrenia Consortium (ISC) showed that hundreds, and perhaps thousands, of common alleles that increase risk for schizophrenia also do so for BD[24] and it is now clear they also do so for MDD, and to a lesser extent, ASD, ADHD, Anorexia Nervosa, Obsessive Compulsive Disorder (Figure 3), as well as personality traits such as neuroticism[25-27]. A problem with inferring biological pleiotropy from GWAS is that the functional alleles (i.e. the alleles that changes function or expression of the gene and directly cause the association) responsible for the vast majority of the GWAS associations have not been identified. It is therefore possible that for any single cross-disorder association, different functional variants within the same or different genes might be responsible. However, the substantial genetic correlations between pairs of psychiatric phenotypes (Figure 3) are less readily explained by pseudo pleiotropy as this would require different functional alleles to be systematically and consistently tagged by the same GWAS allele across large numbers of loci.

Taking the genomic data as a whole, true pleiotropy is by far the most parsimonious explanation for the majority of published cross disorder effects, and most of the findings support extensive allelic pleiotropy. A proviso here is that we must exclude mediated pleiotropy as an explanation. By definition, for one trait to be secondary to (or mediated by) another, the mediating trait must occur first. It follows that childhood onset disorders (e.g. ADHD) cannot be mediated by disorders with typically later ages of onset (e.g.

schizophrenia, MDD). However, it is theoretically possible that the converse is true, and that where alleles are pleiotropic for ID, schizophrenia, and ASD, ID is the primary phenotype influenced by those alleles, and that having ID causally increases risk of ASD and schizophrenia. There is certainly evidence that CNVs and *de novo* LoF mutations occur more frequently in people with psychiatric disorders who additionally have cognitive impairment[9,13,28], an observation that has sometimes been interpreted as indicative of pleiotropy mediated through ID. However, this pattern of co-morbidity is not sufficient to establish mediated pleiotropy, indeed it is to be expected in cases where mutations have direct effects on two phenotypes. There are also powerful arguments against mediated pleiotropy as the sole explanation for this. First, in ASD, LoF *de novo* mutations tend to occur in the same sets of genes in probands with and without intellectual disability[9]. Second, at *SETD1A,* although LoF mutations are associated with both ID and schizophrenia, ID is not a prerequisite for schizophrenia in mutation carriers[15]. Third, ID is not universally seen in people with schizophrenia who carry *de novo* CNVs that are pleiotropic for both disorders[28]. Fourth, in the only study we are aware of that has explicitly undertaken a formal mediation analyses based on a rare variant, the 22q11 deletion CNV was found to have independent effects on cognitive and psychiatric traits (e.g. ADHD and ASD)[30]. The rare variant data are therefore inconsistent with the hypothesis that cross disorder findings are explained by mediated rather than allelic pleiotropy. The common variant findings are more complex, and will be considered further below.

Pleiotropy in the context of complex disorders

Pleiotropy is a challenging phenomenon in the context of highly polygenic disorders. Consider CNVs associated with at least two clinical outcomes, schizophrenia and intellectual disability, as well as being present in apparently unaffected carriers with no clinical phenotype. It has recently been shown[31] that clinically unaffected CNV carriers perform worse on a range of measures of cognitive performance than do non-carrier controls, but better than people with either of the clinical diagnoses associated with the CNVs. Cognitive phenotyping therefore empirically demonstrates that CNVs impact on liability to quantitative traits that are overlooked when the only definition of 'affected' is that of a clinical diagnosis. What determines the final manifestations of increased liability in CNV carriers is not well understood, but an individual's burden of common schizophrenia risk alleles is one important factor[32]. What might then be perceived as pleiotropic manifestations of a particular mutation (e.g. a CNV) may in fact more generally represent the net effects of an individual's polygenic and environmental background on multiple traits representing various domains of brain function.

Specific genes for psychiatric diagnoses

Whether it is possible to link genotype to psychiatric phenotype is generally couched in terms of linear relationships between a gene and a single categorical diagnosis. In our opinion, the evidence summarized above suggests the outlook for relating genotype and phenotype in this way is not promising, although we recognize that there is a bias towards observing pleiotropy since studies are better powered to identify genetic similarities rather than differences.

We do not suggest that risk alleles impact on psychiatric outcomes indiscriminately. For example duplication at 22q11 increases risk of ID and ASD, but is neutral for bipolar disorder, and protective for schizophrenia[33]. Damaging rare mutations play a greater role in ID than in schizophrenia, in schizophrenia compared with mood disorder, and in psychiatric disorders with comorbid cognitive impairment[16]. With regard to common alleles, although many psychiatric disorders are genetically correlated, the degree of correlation between diagnostic classes is usually less than the degree of within disorder correlation[26,34]. These observations suggest that current diagnostic schemes do to some extent capture groups whose members have more in common with each other than they do with members of a general class 'psychiatric disorder'. However, until we can directly measure liability, it is impossible to distinguish the phenotypic heterogeneity arising from true pleiotropic effects of a specific allele (even an allele of large effect) from that resulting from a person's unique blend of risk factors. Directly measuring liability remains a distant goal; for now, identifying alternative approaches to patient stratification that index liability better than current diagnostic categories, and therefore might link more specifically to particular genotypes, is a more realistic aim. Some approaches to doing so are outlined in Box 1.

Implications of Pleiotropy.

The current system of psychiatric classification is not optimal, and alternative approaches are urgently required for clinical and fundamental research. The genetic findings do not, however, imply a similar urgency for fundamental changes in clinical practice as they do not provide the basis for a system with clear clinical value. Given the complexity of the relationships between disorders, and the likelihood that people with psychiatric illnesses differ quantitatively on multiple dimensions of function rather than categorically, seeking hard categorical boundaries that validly reflect aetiology seems a fool's errand. Ultimately, we

suspect the advances in genomic research will allow mapping between pathophysiological processes and domains of brain function (perhaps those outlined in RDoC, perhaps not) and between domains of brain function and the clinical picture and in doing so, will allow clinical measurements (for example types of cognitive test, brain imaging) that highlight perturbations that are pertinent to, and suggest interventions for, particular groups of patients. But what measures are likely to best achieve this, much less how to implement them in a clinical setting, is far from clear. Nevertheless, even now, the pervasive nature of shared risk factors, pleiotropy, and arbitrary diagnostic boundaries between disorders has clinical implications.

As clinicians, we recognize the utility of diagnostic boundaries for therapeutic decision making, communication, and predicting (in a general way) certain outcomes and we do not suggest that clinicians abandon diagnosis using existing categories. However, rigid adherence to categories makes it easy to either overlook co-morbidity or, where it is detected, to inappropriately ascribe it to a diagnosis that has greater weight in the current diagnostic hierarchy. As a result, co-morbid syndromes may not be optimally treated. Given that pleiotropy implies that a person with one syndrome is at enhanced risk for a second syndrome, far from implying lax assessment, pleiotropy emphasizes the need for detailed on-going clinical monitoring, and assessments that go beyond the bare requirements of arriving at the best fitting diagnostic category. Moreover, by appreciating the increasing empirical basis for pleiotropy, clinicians can engage better in discussion with patients who are often bewildered by the range of diagnoses they may receive across their lifespans. Clinicians in other medical disciplines would not assign to a single clinical entity all the physical ailments associated with a pleotropic risk factor such as smoking, and there is no reason why psychiatrists should either.

Children with congenital malformations, developmental delay, and ASD are already being referred for molecular diagnostics, particularly for known pathogenic CNVs, but as the data continue to accumulate, more types of genetic findings will be incorporated. It has been argued that CNV testing should be offered to people with other forms of psychiatric disorders; for now, the case is strongest for schizophrenia[40] but we predict ADHD is likely to follow suit. The range of arguments for and against this are beyond the scope of this article[41]; here, we note that identifying carriers of high penetrance mutations is currently of limited value in psychiatry for precision medicine, but should testing be offered for counseling or predictive purposes, it is important to consider the pleiotropic effects of mutations. CNVs detected in children referred for testing may have important adult psychiatric implications, and conversely if adults are tested, pleiotropy has implications for

their children and other relatives. The counseling challenges are substantial given the wide range of possible outcomes, and much of the data that are required to do this with precision, even for well-documented pathogenic CNVs, is lacking.

The extensive pleiotropy reveled by psychiatric genetics also has important implications for interpreting mechanistic studies, whether in humans, using endophenotypes (Box 1), or in animal and cellular models. Even for high penetrance alleles, the possibility of pleiotropy implies the need for caution in ascribing a causal role in disease for particular brain imaging correlates of that mutation, or in a rodent or stem cell model, neurobiological outcomes. This issue has been discussed conceptually in the case of human endophenotypes and some of the statistical approaches to identifying mediation outlined[42,43]. The challenges to interpreting results from model systems are more testing and will require researchers to cast the net wider than is often the case in seeking the consequences of genetic risk factors and to relate their findings to comparable findings from clinical neuroscience. This will require the use of translatable measures and direct comparisons of the effects of genetic risk across levels of complexity[3].

Finally, on a positive note, pleiotropy may offer unsuspected therapeutic opportunities if it turns out that this is reflected in shared pathophysiology. It is not uncommon for psychiatrists to offer (off-label) treatments to patients with a particular diagnosis that are known to be effective in a different psychiatric disorder. In a very general sense, pleiotropy can be seen as offering some *post-hoc* justification for this, although we stress currently not at the level of any specific treatment. As new treatments are developed to target one disorder, it is likely that treatment will have a broader therapeutic role, and that wider patient populations may benefit from advances in research into a particular disorder.

References

1 Whiteford, H. A. *et al.* Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *Lancet* **382**, 1575 – 1586, doi: 10.1016/S0140-6736(13)61611-6 (2013)

2 Salomon, J. A. *et al.* Common values in assessing health outcomes from disease and injury: disability weights measurement study for the Global Burden of Disease Study 2010. *Lancet* **380**, 2129-2143, doi: 10.1016/S0140-6736(12)61680-8 (2013)

3 Owen, M. J. New approaches to psychiatric diagnostic classification. *Neuron* **84**, 564-571, doi:10.1016/j.neuron.2014.10.028 (2014).

4 Malaspina, D. *et al.* Schizoaffective Disorder in the DSM-5. *Schizophr Res* **150**, 21-25, doi:10.1016/j.schres.2013.04.026 (2013).

5 van Os, J. & Kapur, S. Schizophrenia. *Lancet* **374**, 635-645, doi:10.1016/S0140-6736(09)60995-8 (2009).

6 Lichtenstein, P. *et al.* Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet* **373**, 234-239, doi:10.1016/S0140-6736(09)60072-6 (2009).

7 Kapur, S., Phillips, A. G. & Insel, T. R. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol Psychiatr* **17**, 1174-1179, doi:10.1038/mp.2012.105 (2012).

8 Linden, D. E. The challenges and promise of neuroimaging in psychiatry. *Neuron* **73**, 8-22, doi:10.1016/j.neuron.2011.12.014 (2012).

9 Sanders, S. J. *et al.* Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* **87**, 1215-1233, doi:10.1016/j.neuron.2015.09.016 (2015).

10 Gaugler, T. *et al.* Most genetic risk for autism resides with common variation. *Nat Genet* **46**, 881-885, doi:10.1038/ng.3039 (2014).

11 Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427, doi:10.1038/nature13595 (2014).

12 Rees, E. *et al.* Analysis of copy number variations at 15 schizophrenia-associated loci. *Brit J Psychiat* **204**, 108-114, doi:10.1192/bjp.bp.113.131052 (2014).

13 Fromer, M. *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* **506**, 179-184, doi:10.1038/nature12929 (2014).

14 Purcell, S. M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**, 185-190 (2014).

15 Singh, T. *et al.* Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nat Neurosci* **19**, 571-577, doi:10.1038/nn.4267 (2016).

16 Green, E. K. *et al.* Copy number variation in bipolar disorder. *Mol Psychiatr* **21**, 89-93, doi:10.1038/mp.2014.174 (2016).

17 Hyde, C. L. *et al.* Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat Genet* **48**, 1031 – 1036, doi: 10.1038/ng.3623

Formatted: German (Germany)

Formatted: Spanish (Spain)

Formatted: French (France)

18 Paaby, A. B. & Rockman, M. V. The many faces of pleiotropy. *Trends Genet* **29**, 66-73, doi:10.1016/j.tig.2012.10.010 (2013).
19 Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M. & Smoller, J. W. Pleiotropy in complex traits: challenges and strategies. *Nat Rev Genet* **14**, 483-495, doi:10.1038/nrg3461 (2013).
20 Kirov, G. *et al.* The penetrance of copy number variations for schizophrenia and developmental delay. *Biol Psychiatr* **75**, 378-385, doi:10.1016/j.biopsych.2013.07.022 (2014).
21 Rees, E. Analysis of intellectual disability copy number variants implicates novel schizophrenia risk loci. *JAMA Psychiatry* (2016; in press).
22 Rujescu, D. *et al.* Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum Mol Genet genetics* **18**, 988-996, doi:10.1093/hmg/ddn351 (2009).
23 Deciphering Developmental Disorders, S. Large-scale discovery of novel genetic causes of developmental disorders. *Nature* **519**, 223-228, doi:10.1038/nature14135 (2015).
24 International Schizophrenia, C. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748-752, doi:10.1038/nature08185 (2009).
25 Cross-Disorder Group of the Psychiatric Genomics, C. *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* **45**, 984-994, doi:10.1038/ng.2711 (2013).
26 Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-1241, doi:10.1038/ng.3406 (2015).
27 Hamshere, M. L. *et al.* Shared polygenic contribution between childhood attention-deficit hyperactivity disorder and adult schizophrenia. *Brit J Psychiat* **203**, 107-111, doi:10.1192/bjp.bp.112.117432 (2013).
28 Samocha, K. E. *et al.* A framework for the interpretation of de novo mutation in human disease. *Nat Genet* **46**, 944-950, doi:10.1038/ng.3050 (2014).
29 Kirov, G. *et al.* De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol Psychiatr* **17**, 142-153, doi:10.1038/mp.2011.154 (2012).
30 Niarchou, M. *et al.* Psychopathology and cognition in children with 22q11.2 deletion syndrome. *Brit J Psych* **204**, 46-54, doi:10.1192/bjp.bp.113.132324 (2014).
31 Stefansson, H. *et al.* CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature* **505**, 361-366, doi:10.1038/nature12818 (2014).
32 Tansey, K. E. *et al.* Common alleles contribute to schizophrenia in CNV carriers. *Mol Psychiatr*, doi:10.1038/mp.2015.143 (2015).

Formatted: German (Germany)

Formatted: French (France)

Formatted: French (France)

Formatted: French (France)

33 Rees, E. *et al.* Evidence that duplications of 22q11.2 protect against schizophrenia. *Mol Psychiatr* **19**, 37-40, doi:10.1038/mp.2013.156 (2014).
34 Lee, S. *et al.* Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet* **91**, 224-237, doi:10.1016/j.ajhg.2012.06.007 (2012).
35 Ruderfer, D. M. *et al.* A family-based study of common polygenic variation and risk of schizophrenia. *Mol Psychiatr* **16**, 887-888, doi:10.1038/mp.2011.34 (2011).
36 Hamshere, M. L. *et al.* Genome-wide significant associations in schizophrenia to ITIH3/4, CACNA1C and SDCCAG8, and extensive replication of associations reported by the Schizophrenia PGC. *Mol Psychiatr* **18**, 708-712, doi:10.1038/mp.2012.67 (2013).
37 Flint, J., Timpson, N. & Munafo, M. Assessing the utility of intermediate phenotypes for genetic mapping of psychiatric disease. *Trends Neurosci* **37**, 733-741, doi:10.1016/j.tins.2014.08.007 (2014).
38 Franke, B. *et al.* Genetic influences on schizophrenia and subcortical brain volumes: large-scale proof of concept. *Nat Neurosci*, doi:10.1038/nn.4228 (2016).
39 Cuthbert, B. N. The RDoC framework: facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. *World Psychiatry* **13**, 28-35, doi:10.1002/wps.20087 (2014).
40 Costain, G. *et al.* Pathogenic rare copy number variants in community-based schizophrenia suggest a potential role for clinical mircoarrays. *Hum Mol Genet* **22**, 4485-4501 doi: 10.1093/hmg/ddt297 (2013).
41 Gershon, E. S. & Alliey-Rodriguez, N. New ethical issues for genetic counseling in common mental disorders. *Am J Psychiat* **170**, 968-976, doi:10.1176/appi.ajp.2013.12121558 (2013).
42 Walters, J. T. R. & Owen, M. J. Endophenotypes in psychiatric genetics. *MolPsychiatr* **12**, 886–890, doi:10.1038/sj.mp.4002068 (2007).
43 Kendler K. S. & Neale, M. C. Endophenotype: a conceptual analysis. *MolPsychiatr* **15**, 789-97, doi: 10.1038/mp.2010.8 (2010).
44. Antilla, V. *et al.* Analysis of shared heritability in common disorders of the brain. Preprint at *bioRxiv* http://dx.doi.org/10.1101/048991 (2016)**.**

Formatted: German (Germany)

Formatted: German (Germany)

Figure Legends

Figure 1. Types of pleiotropy**.**

Adjacent genes containing functional variants (FV; yellow circles) that together directly influence four distinct phenotypes (blue shapes). The three phenotypes directly influenced by FV_1 and FV_2 are examples of genic pleiotropy. The circle and pentagon phenotypes influenced by FV_1 are examples of allelic pleiotropy in which the same variant rather than just the same gene influences multiple phenotypes. FV_1 influences the triangle phenotype but this is indirect and only occurs through the direct effects of FV_1 on the circle phenotype. This is mediated pleiotropy. Alleles at FV_2 and FV_3 are correlated (through linkage disequilibrium; LD) with the same single nucleotide polymorphism (SNP) Accordingly, that SNP will be associated with both phenotypes that are caused by those functional variants. The SNP is depicted midpoint between the genes but be positioned anywhere within the region of LD, including within one of the genes. This is pseudo pleiotropy due to co-localization. This region is also prone to a deletion CNV which results in complete loss of function of both genes by virtue of which it is associated with all five phenotypes. In a literal sense, all of the blue phenotypes in this instance are now examples of allelic pleiotropy (being directly caused by the same CNV allele at a single locus).

Figure 2. Relative CNV Frequencies.

Relative frequencies for schizophrenia associated CNVs. Frequency is expressed as fold increase in each disorder relative to the estimated population frequency. Data are taken from[21] based on loci reported as schizophrenia associated[12,21]. CNVs are described by cytogenetic position or the named syndrome most strongly affiliated with the locus. The approximate lifetime population risk for SZ is approximately 1% and for ID/ASD combined 4%[20]. Abbreviations: SZ schizophrenia; ASD/ID autism spectrum disorder and ID intellectual disability combined. WBS Williams-Beuren Syndrome. PWS/AS. Prader-Willi Syndrome/Angelman Syndrome; VCFS Velo-cardio-facial syndrome; CNV copy number variant; del deletion; dup duplication.

Figure 3. Genetic correlation between schizophrenia and selected psychiatric disorders.

Psychiatric disorders showing significant evidence ($P \leq 0.001$) for overlaps between common variant contributions to schizophrenia and other psychiatric disorders. Overlaps are

513 expressed as correlation in heritability (r_g) captured by SNPs. Data are from[44]. Abbreviations:
514 ADHD Attention Deficit Hyperactivity Disorder; OCD Obsessive Compulsive Disorder.
515
516 **Box 1: Patient Stratification**
517 There is general agreement that we need new approaches to patient stratification in
518 research if we are to better understand gene-phenotype relationships, accelerate
519 understanding of aetiology and pathogenesis, and inform mechanistic studies and treatment
520 trials. Generally speaking three ways in which we can move beyond the constraints of
521 current diagnostic approaches have been proposed. Rather than being mutually exclusive,
522 these can be thought of as targeting psychiatric disorders at different levels of conceptual,
523 and aetiological complexity, from the molecular at one end to the function and behaviour of
524 the whole human at the other. Models that attempt to capture this hierarchical complexity of
525 have been proposed and discussed in detail elsewhere[3].
526
527 First, we can use clinical classifications that cut across or divide current diagnostic groups.
528 These might be based upon the presence of absence of particular symptoms (e.g.
529 hallucinations), syndromes (e.g. psychosis, depression) or other features such as course or
530 outcome. This may aid the identification of risk factors and pathogenic mechanisms
531 providing the strata map more closely onto these than do current diagnostic groupings. This
532 approach also has the potential to help our understanding of the basis of heterogeneity.
533 There is some evidence to support this type of approach, for example stratifying people with
534 BD for the presence of psychotic symptoms predicts a higher burden of schizophrenia risk
535 alleles, and, conversely, stratifying people with schizophrenia for presence of manic type
536 symptoms predicts a higher burden of bipolar risk alleles[35,36]. These preliminary findings
537 suggest that, across disorders, sets of syndromes have some shared biological basis, and
538 support a model where disorders, as manifest in individuals, may be viewed as the
539 confluence of partly orthogonal symptom dimensions.
540
541 Second, stratification can be based on the presence of a particular aetiological factor (e.g. a
542 rare high penetrance mutation, a particular environmental exposure) rather than clinical
543 features. The assumption is that constraining the risk architecture will increase biological
544 homogeneity, and allow researchers to focus on specific risk mechanisms and understand
545 what factors lead to different outcomes, including resilience as well as risk. This type of
546 approach also lends itself to complementary studies in cells and animals as well as humans.
547 In psychiatry, this has yet to yield unqualified success, and, given evolutionary multi
548 purposing of proteins, which may have different functions in different cells or cell
549 compartments, even a single genetic variant might map onto different pathogenic

mechanisms in carriers. While this is a theoretical concern, the fact that regardless of the specific psychiatric diagnoses (ID, ASD, schizophrenia), rare *de novo* and LoF mutations tend to impact upon similar broadly similar processes (e.g. glutamatergic pathways regulating synaptic plasticity, chromatin modifiers, and targets of fragile X mental retardation protein) suggests that individual mutations are likely to influence the same pathogenic mechanisms across disorders[13] .

Third, in attempting to relate risk factors and clinical phenotypes to underlying pathophysiology and mechanisms, stratification can be performed at the level of endophenotypes (intermediate phenotypes). One problem with this approach is the large number of potential endophenotypes including measures of cognition, brain structure, electrophysiology, and biochemistry. Moreover, initial claims that endophenotypes are likely to be less complex genetically than clinical disorders have not in general been supported[37] and perhaps this explains why failures to link endophenotypes to genetic risk[38] are for now more notable than any reproducible successes. Nevertheless, this approach offers a means by which genetic risk can be linked to disturbances of brain function, and a framework for doing so has been implemented in the Research Domain Criteria (RDoC) project of the National Institutes of Mental Health[39]. The pleiotropic effects of many risk alleles are clear reminders that there are pitfalls associated with using this approach to chart the pathways mediating the effects of genetic risk on clinical phenotypes (see main text).