Genetic pleiotropy between mood disorders, metabolic, and endocrine traits in a multigenerational pedigree

Rachel L. Kember[1], Liping Hou[2], Xiao Ji[3], Lars H. Andersen[4], Arpita Ghorai[1], Lisa N. Estrella[4], Laura Almasy[1,5], Francis J. McMahon[2], Christopher Brown[1], Maja Bucan[1,6]

1. Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104
2. Human Genetics Branch, National Institute of Mental Health Intramural Research Program, National Institutes of Health, Bethesda, MD 20892
3. Genomics and Computational Biology Program, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104
4. Lancaster General Health/Penn Medicine, University of Pennsylvania Health System, Lancaster, PA 17602
5. Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA 19104
6. Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104

Corresponding author: Maja Bucan, bucan@upenn.edu

Abstract

Bipolar disorder (BD) is a mental disorder characterized by alternating periods of depression and mania. Individuals with BD have higher levels of early mortality than the general population, and a substantial proportion of this may be due to increased risk for co-morbid diseases. Recent evidence suggests that genetic pleiotropy, either in the form of a single risk-allele or the combination of multiple loci genome-wide, may underlie medical co-morbidity between traits and diseases. In order to identify the molecular events that underlie BD and related medical co-morbidities, we generated imputed whole genome sequence (WGS) data, using a population specific reference panel, for an extended multigenerational Old Order Amish pedigree (400 family members) segregating BD and related disorders. We identified all disease-causing variants at known Mendelian loci present in this pedigree. In parallel, we performed genomic profiling using polygenic risk scores to establish each family members risk for several complex diseases. To explore the contribution of disease genes to BD we performed gene-based and variant-based association tests for BD, and found that Mendelian disease genes are enriched in the top results from both tests (OR=20.3, $p<0.001$; OR=2.2, $p<0.01$). We next identified a set of Mendelian diseases and variants that co-occur in individuals with BD more frequently than their unaffected family members, including the R3527Q mutation in *APOB* associated with hypercholesterolemia. Using polygenic risk scores, we demonstrated that BD in this pedigree is driven by the same common risk-alleles as in the general population ($\beta=0.416$, $p<0.0006$). Furthermore, in the extended Amish family we find evidence for a common genetic etiology between BD and clinical autoimmune thyroid disease ($p<0.001$), diabetes ($p<0.0004$) and lipid traits ($p<0.0003$). We identify local risk regions for these traits that drive the differences between BD individuals and unaffected family members. Our findings provide evidence for the extensive genetic pleiotropy that can drive epidemiological findings of co-morbidities between diseases and traits. We propose that identifying such patterns of phenotypic and genetically-defined pleiotropy may enable subtyping of complex diseases and facilitate their genetic dissection.

Introduction

Psychiatric disorders frequently co-occur with other medical illnesses, markedly reducing patients' quality of life. Individuals with mood disorders have higher levels of early mortality than the general population, in part due to co-morbid medical disease (Parks et al., 2006). In addition, individuals with a higher burden of medical illness also display an increased severity of psychiatric symptoms (Beyer et al., 2005). Bipolar disorder (BD) is a highly heritable mood disorder characterized by recurrent periods of depression and mania. Individuals with BD have increased rates of asthma, diabetes, hyperlipidemia, epilepsy and thyroid disease, among other diseases (Forty et al., 2014).

Historically, increased rates of medical illness in patients with psychiatric disorders had been attributed to the side effects of antipsychotic medication or to a reduced ability to maintain a healthy lifestyle. However, recent evidence suggests that shared genetic risk loci or common biological pathways may underlie the pervasive pleiotropy between psychiatric and non-psychiatric disorders (Bulik-Sullivan et al., 2015; Pickrell et al., 2016; Prieto et al., 2016; Sivakumaran et al., 2011). Genetic pleiotropy can be identified at the level of individual alleles, or genetic correlations between disorders can be calculated genome-wide to quantify the proportion of shared associated loci between traits (Bulik-Sullivan et al., 2015).

Co-morbidity arising from pleiotropic loci has been noted in Mendelian disorders, with a significant number of Mendelian disease-causing variants leading to complex phenotypes (Zhu et al., 2014). In individual-level data gathered from medical records of over 110 million patients, Mendelian variants were found to contribute non-additively to risk for a subset of complex diseases (Blair et al., 2013). Furthermore, common variants associated with complex disease are enriched in Mendelian disease genes (Blair et al., 2013). Shared genetic influences between common complex traits have also been identified. Using data from genome-wide association studies (GWAS), several groups have identified loci underlying multiple traits (Pickrell et al., 2016; Sivakumaran et al., 2011). Such cross-phenotype associations (Solovieff et al., 2013) have been found even between distinct traits; for example, a nonsynonymous variant in *SLC39A8* is associated with both schizophrenia and height, among others (Pickrell et al., 2016).

Genetic correlations have demonstrated the shared genetic influences between multiple clusters of diseases, including significant correlations between psychiatric disorders (Gale et al., 2016; Hammerschlag et al., 2017; Lee et al., 2013; Lo et al., 2017). Shared genetic etiology has been identified between BD and both schizophrenia and major depressive disorder (Lee et al., 2013), suggesting extensive biological pleiotropy between these psychiatric conditions (O'Donovan and Owen, 2016). Given the polygenic nature of these disorders (Wray et al., 2014), this result is unsurprising, as polygenicity is consistent with comorbidity and pleiotropy (Gratten et al., 2014). Gratten et al. (2014) also notes that genetic correlations between BD data sets are more variable, possibly suggesting greater genetic heterogeneity within BD compared to other psychiatric phenotypes.

Population isolates are frequently utilized in genetic studies of disease in order to reduce the genetic and phenotypic heterogeneity found in outbred populations. Disease-gene identification in population isolates usually attempts to identify both common and low-frequency variants, observed within and across families. Moreover, founder effects can lead to an increase in allele frequencies for many deleterious alleles and clusters of deleterious variants on shared haplotypes. Genetic studies of the Old Order Amish led to the identification of over 200 Mendelian disease loci (Puffenberger et al., 2012; Strauss and Puffenberger, 2009; Strauss et al., 2012), and the same population has been extensively utilized in studies of complex metabolic and psychiatric diseases (Georgi et al., 2014; Kember et al., 2015; Rampersaud et al., 2007; Steinle et al., 2005; Strauss et al., 2014; Xu et al., 2017).

The Amish Study of Major Affective Disorder (ASMAD) is a large extended pedigree collected over 30 years ago with an initial focus on bipolar disorder. In addition to a significant enrichment of mood disorders in this family (~150 family members, corresponding to one third of the pedigree), the pedigree includes individuals with autoimmune thyroid disorder (Jaume et al., 1999) and Ellis-van Creveld Syndrome, an autosomal recessively inherited chondrodysplastic dwarfism (Ginns et al., 2015; McKusick et al., 1964). Our previous work on this pedigree revealed a complex and polygenic inheritance of bipolar disorder with multiple linkage regions and clusters of BD risk-alleles on different haplotypes, supporting a high degree of locus and allelic heterogeneity (Georgi et al., 2014; Kember et al., 2015). Further dissection of the genetic architecture of mental illness in this pedigree has been greatly enhanced by the recently established whole-genome sequence-based imputation reference panel for the Anabaptist population (Amish and Mennonite; Anabaptist Genome Reference Panel, AGRP; (Hou et al., 2017)).

In this study, we used the AGRP in combination with genotypes for the ASMAD pedigree to permit the identification of all known disease-causing and loss-of-function (LoF) variants at known Mendelian loci in subjects with mood disorders. In parallel, we performed genomic profiling using polygenic risk scores to establish individuals' risk for several complex traits and diseases. Long-range phased haplotypes estimated using the extended pedigree permitted the exploration of co-segregation between bipolar risk factors and medical disease loci. We find that a set of Mendelian diseases co-occur in BD individuals more frequently than in their unaffected family-members, and that risk scores for metabolic traits are higher in BD individuals, indicating a common genetic etiology for these traits in the extended Amish family.

Results

Identifying the spectrum of mutations in disease genes

By combining SNP genotype data from 394 ASMAD family members with the Anabaptist Genome Reference Panel (Hou et al., 2017) we established high quality phased whole genome sequence for the entire ASMAD extended family. Although available phenotype records for this collection are limited to information about bipolar and related neuropsychiatric disorders (Supplemental table 1), the availability of imputed genotypes allows a "sequence-first" approach to identify ASMAD family members who are carriers of known Mendelian disease variants; and to ascertain the individual-level burden of common and rare variants in genes associated with a wide range of medical conditions.

Disease genes were identified using the Human Genome Mutation Database (HGMD, (Stenson et al., 2014)) by selecting genes with known disease causing mutations (HGMD-DM). Out of 3,456 HGMD-DM genes, 239 contain either known disease causing variants or LoF variants in the ASMAD imputed whole genome sequence. The set of HGMD-DM genes were enriched among all genes with predicted damaging variants that are rare in 1000 Genomes and ExAC (<2%) and common in ASMAD (>5%) (57 HGMD-DM out of 239 genes, OR=1.58, p-val=0.003747, two-sided Fisher's exact test, Supplemental Table 2). Also, HGMD-DM genes are enriched among both sets of genes (Gene-based test for BD association – MONSTER) and sets of variants (variant-based test for association – EMMAX) that are nominally associated with BD in this pedigree. Out of the top 6 genes that have p-values<0.001, 5 of these are known disease genes (OR=20.3, p-value=0.001, two-sided Fisher's exact test, Supplemental Table 3). Out of the top 44 exonic variants that have p-values <0.001, 17 of these are in known disease genes (OR=2.2, p-value=0.01, two-sided Fisher's exact test, Supplemental Table 4).

In light of these initial results, and the previously noted trend towards a higher burden of CNVs in HGMD-DM genes in subjects with bipolar disorder (Kember et al., 2015), we closely examined variants in disease genes to screen for the presence of Mendelian diseases in this pedigree, and to query their contribution to the clinical heterogeneity of BD. Following extensive curation of variants (see Methods), 62 variants were identified as high confidence disease-causing, of which 17 are known to be more common in the Amish (>5% difference) (Supplemental Figure 1, Supplemental Table 5). Based on the allelic state of these variants in individuals, we predict that 22 of these are likely to cause disease due to being present in a homozygous (for recessive diseases) or heterozygous (for dominant diseases) state (Supplemental Table 6). Burden of genomically-predicted medical conditions varies between nuclear families, with some families (and individuals within these families) carrying variants for up to 5 Mendelian diseases (Supplemental Figures 2-19). Furthermore, each individual has on average 31 LoF variants in disease genes with 5 disease genes predicted to be completely inactivated.

BD individuals show co-morbidity with a specific set of Mendelian diseases

Overall burden of disease causing variants and LoF variants in disease genes is not associated with BD status in the pedigree. However, the percentage of individuals carrying specific Mendelian disease-causing variants differs between BD individuals and unaffected individuals (Figure 1). To identify if a specific set of Mendelian diseases show co-morbidity with BD, we applied association rule discovery to determine diseases that co-occur more frequently than expected with affected or unaffected status of mood disorders (Figure 1, A and B; rule confidence: 0.015-0.123, rule lift: 1.012-1.694). In the ASMAD pedigree, BD individuals are more likely to carry variants causing familial hypercholesterolemia, small fiber neuropathy, and Xerocytosis hereditary (Figure 1, C). Variants causing Adrenocortical hyperplasia and Apolipoprotein C-III deficiency are found more frequently in unaffected individuals.

Half of the Mendelian diseases (6 out of 12) that were identified by association rule discovery as either enriched or depleted in BD individuals are associated with cardiovascular disease or endocrine traits. We therefore explored additional damaging variants associated with lipid phenotypes (Table 1), and found that they are present at a higher frequency in individuals with BD compared to unaffected individuals, including the G574R mutation in *ABCG8* associated with hyperabsorption and sisterolemia, and the R3527Q mutation in *APOB*. In addition to the heterozygotes detected, three homozygotes for the R3527Q mutation were found within the same nuclear family. Strikingly, protective lipid variants are found at a lower frequency in BD individuals compared to unaffected individuals, including a variant in *LPA*, associated with a reduction in thermogenic lipoprotein. Variants associated with heart rhythm (in *KCNH2, SCN5A* and *SNTA1*) show no allele frequency difference between affected and unaffected individuals. We also found that a higher percentage of BD individuals than expected have multiple lipid variants, whereas a lower percentage of unaffected individuals carry multiple lipid variants (Supplemental figure 20). Overall, over one third (34.0%) of affected individuals (broad extended phenotype) carry at least one variant associated with damaging lipid or iron overload effects, compared to just under one quarter of unaffected individuals (24.7%).

Common polygenic risk for bipolar disorder

To compare the genetic architecture of BD in the Amish with a non-founder population, and to further explore the enrichment of variants associated with lipid traits in BD subjects, we sought to measure the aggregate effect of common variants underlying both psychiatric and endocrine/metabolic traits in family members with and without mood disorders. To achieve this, we evaluated different phenotypic models from the most defined (BPI individuals only) to the most general (all individuals with any psychiatric phenotype, including minor psychiatric symptoms). Heritability estimates generated for this pedigree previously have shown significant heritability for mood disorder phenotypes (Georgi et al., 2014). Using a refined pedigree structure, updated from accurate genealogical records, we estimated narrow sense heritability for the following phenotype models: narrow (BPI and BPII only), broad (BPI, BPII, and MDDR), broad extended (presence of any psychiatric phenotype), depression (MDDR, MDD, and Minor

Depression), and well. We found significant heritability for all phenotype models (Supplemental Table 7), with broad extended being the most heritable phenotype (h^2=0.81, p=9.76E-10). We generated polygenic risk scores based on GWAS summary data (from Europeans) for multiple traits and tested for an association between the risk score and presence of mood disorder (Figure 2, Table 2, Supplemental Table 8). Polygenic risk scores (PRS) incorporate multiple SNPs, including those that do not reach the threshold for genome-wide significance, to predict an individual's risk for disorder. First, we used summary statistics from a GWAS for bipolar disorder by the Psychiatric Genomics Consortium (PGC BD2, (Stahl et al., 2017)), to generate risk scores for bipolar disorder for all individuals in the ASMAD pedigree. Linear mixed model analyses implemented using the pedigreemm R package (binomial test, (Vazquez et al., 2010)) permitted an association test for risk scores while accounting for relatedness within the pedigree. We tested the difference in risk scores between unaffected and affected individuals for the above phenotype models.

Bipolar disorder risk score was significantly associated with affected phenotype for broad extended (β=0.416, SE=0.121, p=0.000588), broad (β=0.523, SE=0.129, p=0.000048), and narrow phenotypes (β=0.499, SE=0.138, p=0.000298), but not for depression phenotype (β=0.273, SE=0.155, p=0.078473). To verify this result, we used a modified polygenic transmission disequilibrium test (see methods) and also found a statistically significant deviation in risk scores, with affected individuals having a higher polygenic risk than their unaffected siblings (mean deviation 0.27, p=0.03). This increase is mostly driven by higher risk scores of individuals with BPI, BPII, BP:NOS, and MDDR (Supplemental Figure 21). The percentage of affected individuals, on average, increases with increasing deciles of PRS (Figure 3A). Furthermore, the percentage of individuals with BPI, BPII and BP:NOS is highest in the 10th PRS decile, whereas the percentage of individuals with minor depression or other non-specified disorder is lowest in this decile (Figure 3B). Offspring with an affected parent (either one or both parents affected) have higher risk scores than offspring from unaffected parents (Supplemental Figure 22). Furthermore, descendants that can trace their lineage back to the two pioneer members of this pedigree ("in family") have significantly higher risk scores for bipolar disorder than Amish individuals who are "married in" (β=1.653, SE=0.370, p=0.000008, Supplemental Figure 23), suggesting that the pioneer individuals carried high risk for bipolar disorder. Risk scores vary across the pedigree, although they are more similar within nuclear families (Supplemental Figure 24).

Common risk variants for disease suggest pleiotropy between common complex traits and bipolar disorder

Risk score analyses can also be used to query genetic associations between complex disorders that have shown co-morbidity in individuals but were otherwise thought to be causally unrelated (Wray et al., 2014). Therefore we compiled published GWAS summary statistics data (Supplemental table 9) and generated risk scores for 22 additional traits, including common psychiatric disorders, personality measures, educational attainment, Alzheimer's, Type 2 diabetes, autoimmune thyroid disease, lipid traits, and cardiovascular disease (Figure 2, Table 2). Risk scores for clinical autoimmune thyroid

disease, lipid traits (Total cholesterol, Triglycerides, LDL, HDL) and type 2 diabetes were found to be significantly higher in individuals with mood disorder, suggesting that in this pedigree there are shared loci between bipolar disorder and both lipid traits and diabetes. Thyroid hormone levels, positive thyroid antibodies, HDL, LDL, triglyceride, total cholesterol, and diabetes risk scores were significantly associated with affected phenotype (bonferroni corrected p-value=0.0023, broad extended thyroid hormone p=0.000997, thyroid antibodies p=0.001915, HDL p=0.001453, LDL p=0.000620, TG p=0.000269, TC p=0.001866, diabetes p=0.001027; broad thyroid hormone p=0.003233, thyroid antibodies p=0.005292, HDL p=0.000530, LDL p=0.000716, TG p=0.000166, TC p=0.000900, diabetes p=0.000339; narrow thyroid hormone p=0.003294, thyroid antibodies p=0.001068, HDL p=0.005891, LDL p=0.004083, TG p=0.003742, TC p=0.008997, diabetes p=0.003486), but only thyroid hormone, LDL and triglyceride risk scores were nominally associated with depression phenotype (thyroid hormone p=0.003417, LDL p=0.038238, TG p=0.006943) (for full results, see Supplemental table 8).

In the extended Amish pedigree, risk scores for bipolar disorder are positively correlated with risk scores for lipid traits (TG r=0.55, p<2.2e-16, TC r=0.55, p<2.2e-16, LDL r=0.52, p<2.2e-16, HDL r=0.57, p<2.2e-16, Supplemental Figure 25), diabetes (r=0.54, p<2.2e-16), thyroid levels (thyroid hormone levels r=0.31, p=1.692e-10), major depression (r=0.39, p=6.376e-16), schizophrenia (r=0.55, p<2.2e-16), and Alzheimer's disease (r=0.40, p=2.922e-16), and negatively correlated with years of education (r=-0.21, p=1.961e-05), suggesting an overlap between loci that contribute to risk for these traits. Furthermore, we explored the effect of inbreeding on risk score values, as increased homozygosity in offspring of closely related parents is known to increase the incidence of recessive Mendelian disorders, but the effect on complex disorders is less well quantified. We found that higher inbreeding values (measured as a higher number of observed homozygous genotypes than expected), but not increased average length of homozygous regions, are correlated with an increased risk score for bipolar disorder (r=0.69, $p=2.2x10^{-15}$, Supplemental Figure 26), suggesting a role for increased homozygosity in common disease risk.

Identifying pleiotropic regions of the genome

To establish which genomic regions contribute to the genome-wide risk score differences between affected and unaffected individuals identified above, we calculated risk scores across the genome using a sliding window of 5Mb (see methods). This allowed us to partition the genome and analyze risk for each 5Mb region to identify local polygenic risk for bipolar disorder, total cholesterol, LDL, HDL, triglycerides, diabetes and thyroid disorder associated with BD. Using linear mixed model analyses (as above) we examined 1,112 5Mb regions (556 non-overlapping regions) per risk score. We identified 6 genomic regions that had significantly different (p<0.01; uncorrected for multiple testing, since many non-overlapping regions are still highly correlated due to LD structure) risk scores between affected (broad extended phenotype) and unaffected individuals for bipolar disorder, 22 for total cholesterol, 19 for LDL, 14 for HDL, 11 for triglycerides, 11 for diabetes, and 8 for thyroid hormone levels (Figure 4, Supplemental Table 10). Of the

91 regions identified, 15 regions overlapped between traits. Total cholesterol, HDL and diabetes local polygenic risk associated with BD overlapped in two neighboring regions on chromosome 3 (3p14.1-p12.3), a region previously found to be associated with hypertension (Koivukoski et al., 2004) that harbors multiple genes known to be involved in HDL and cholesterol traits. This region has also been associated with bipolar disorder, depression and schizophrenia in multiple studies (Chuang and Kuo, 2017; Goes et al., 2015; Rudd et al., 2015; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). Total cholesterol and LDL local polygenic risk associated with BD overlapped in one region on chromosome 20 (20q11.23-q12), a region previously associated with hypertension and diabetes, although not found to be consistently associated with bipolar disorder. Bipolar disorder and LDL local polygenic risk associated with BD overlapped in one region on chromosome 16 (16q21), which contains a single gene called cadherin 8 type 2 (*CDH8*). CDH8 is widely expressed in the brain and in the aorta, and is thought to be involved in synaptic adhesion and early cortical development (Pagnamenta et al., 2011). The other regions were not found to overlap between traits, suggesting that the association of bipolar disorder with risk for other traits in this pedigree is due to multiple pleiotropic regions spread across the genome.

Discussion

Co-morbidity of medical disease in individuals with psychiatric disorder is a major contributor to early mortality and severity of phenotype. Although previously attributed to poor lifestyle choices, recent genetic evidence suggests that pleiotropic loci that predispose to both mental illness and non-psychiatric disease may underlie at least some co-morbidities. Here we present a large multigenerational pedigree significantly enriched for Bipolar and related mood disorders, that we further genetically diagnose with segregating Mendelian disease variation. In addition to identifying a unique pattern of co-segregating low-frequency Mendelian disease variants associated with mood disorder status, polygenic risk profiling, using several sets of common variants associated with a wide range of complex diseases and traits, provides strong support for a shared genetic architecture between mood disorders and specific metabolic and endocrine traits. Although our detected associations may be unique to the Amish extended pedigree, we propose that such patterns of phenotypic and genetically-defined pleiotropy may enable subtyping of complex diseases and facilitate their genetic dissection.

The expectation in studies of complex disease in founder populations is that low-frequency disease-causing variants with strong effect sizes, that are rare in the general population but enriched or specific to the isolate, can be more easily identified compared to studies in outbred populations. However, previously published findings from the ASMAD population (Georgi et al., 2014; Ginns et al., 2015; Kember et al., 2015) reveal a complex polygenic mode of inheritance, in line with findings from general population studies. Here we corroborate this conclusion by reporting that bipolar disorder polygenic risk profiling unequivocally supports a shared genetic architecture (i.e. shared common risk variants) between bipolar disorder in the Amish and in an outbred population of European ancestry. Furthermore, as expected from studies on the shared genetic etiology of psychiatric disorders (Lee et al., 2013), polygenic risk scores for major depressive disorder are also associated with bipolar disorder in this pedigree. Interestingly, despite the high genetic correlation between bipolar disorder and schizophrenia in the general population, the schizophrenia polygenic risk score was limited in its predictive ability in the ASMAD sample (R^2=0.019). This may reflect an ascertainment bias in recruitment or a unique aspect of the phenotype in this extended pedigree. For example, out of 195 individuals with any reported psychiatric condition in the family, only a single individual presented with schizoaffective disorder and none with schizophrenia. Similarly, a study of major psychiatric disorders in mid-Western Amish in Ohio and Indiana reported lower rates of psychosis compared to other populations (Hou et al., 2013).

Our previous analysis of regions of homozygosity did not reveal recessive risk variants (SNV or CNVs) for BD in this pedigree (Georgi et al., 2014; Kember et al., 2015). Conversely, here we report that polygenic risk scores for bipolar disorder and several metabolic traits are positively correlated with the inbreeding co-efficient. Interestingly, in an effort to investigate genetic diversity within Anabaptist groups represented in the Anabaptist Genome Reference Panel (AGRP) (Hou et al., 2017), we showed that Amish living in Lancaster County, Pennsylvania, where the ASMAD pedigree originates, have

the longest homozygous-by-descent-regions. Therefore, we suggest that reduced genetic diversity may underlie the accumulation of both common and low-frequency, homozygous and compound heterozygous, risk alleles for a range of diseases and traits.

Our identification of individuals who should present with a Mendelian disease, based on their carrier status of highly penetrant Mendelian variants, established a set of diseases that differ in frequency between those with and without mood disorder. Moreover, individuals carrying these variants in a manner insufficient for presentation of disease also demonstrated differences in mood disorder status. For example, a dominant familial hypercholesterolemia-associated APOB R3527Q variant, which has been previously identified as being carried by around 12% of Old Order Amish individuals (Shen et al., 2010), was found in 19.7% of individuals with BPI, BPII or BP:NOS. This finding suggests that rather than one trait having a causal effect on the other (e.g. hypercholesterolemia "causing" bipolar disorder) it is more likely that a gene variant, or the haplotype containing this Mendelian variant harbors other co-segregating variants, that convey risk for both disorders.

To further explore the biological implications of our finding that a specific set of Mendelian variants are enriched in individuals with mood disorders, we generated risk scores for multiple traits using common disease associated variants. We identified a number of complex metabolic and endocrinological diseases (lipid traits, diabetes and clinical thyroid disease) that were significantly associated with mood disorder in this pedigree. The relationship between thyroid disorders and mood disorders has been acknowledged for many years (Whybrow et al., 1969), with thyroid function associated with depressive symptoms, anxiety, and mania (Chakrabarti, 2011). While metabolic syndrome in general (Fiedorowicz et al., 2008; Vancampfort et al., 2013), and diabetes (Golden et al., 2008) and lipid levels (Enger et al., 2013; Patel et al., 2007) specifically, have been shown to be elevated in individuals with mood disorder, cross-trait analysis has not provided evidence for a significant genetic correlation between bipolar disorder and any metabolic trait (Bulik-Sullivan et al., 2015). Our study on individual level data in a large extended family reveals evidence for a genome-wide genetic correlation between mood disorder and specific metabolic traits. As this finding has not yet been replicated in an outbred population, we hypothesize that BD in the ASMAD family could represent a subtype of BD with high levels of metabolic and endocrinological disorders. This has been termed subgroup heterogeneity (Han et al., 2016), where a genetically distinct subset of individuals within a patient cohort is also genetically similar to individuals with another disease. There is already emerging evidence that such a subgroup exists for major depressive disorder (Howard et al., 2017), and the identification of a similar subgroup in BD cohorts may help explain the epidemiological findings of associations between these disorders.

In an effort to quantify the regions which contribute to differences in risk scores, we take advantage of the long haplotypes shared between multiple related individuals to analyze risk scores for 5Mb regions of each chromosome and identify windows for which risk scores differs between individuals with mood disorder and their unaffected relatives. We identified a relatively small number of regions for each trait, covering 0.92-2.08% of the

genome, that are significantly associated with differences in risk scores. The majority of these regions are unique to each trait; out of the five regions identified as driving bipolar risk differences between affected and unaffected individuals, only one of these overlaps with another trait (HDL). This suggests that regions that drive the difference in polygenic risk for bipolar disorder are not the same regions that drive polygenic risk for diabetes and lipid traits. However, we assume that these regions are pleiotropic in that they carry risk for bipolar disorder in this pedigree (given their difference between BD and unaffected individuals) and also carry risk for diabetes/lipid traits based on the summary statistics from an independent data set.

We acknowledge several limitations within our current work that highlight avenues for future study. There are many cases of disease genes with different alleles leading to different phenotypes or a specific allele leading to a range of diverse phenotypes (Lupski et al., 2011). Although there are clear examples of homozygous variants causing a recessive Mendelian disease and a complex condition in carriers (Lupski et al., 2011), our study does not provide the resolution required to distinguish between a single gene vs. haplotype effect. Specifically, we can't determine whether variation in the Mendelian disease gene alone, or variation at the level of the haplotype, is responsible for the shared genetic etiology of disease in this population. We expect that the long shared haplotypes observed in a genetic isolate are more likely to harbor multiple disease-contributing variants, and therefore produce the observed genetic pleiotropy. Furthermore, multiple mechanisms, such as zygosity, epistasis and other interacting genes (genetic background) or environmental effects may modify expressivity or penetrance of a specific allele (Zhu et al., 2014). While every attempt was made to limit our analyses to Mendelian variants predicted to cause disease, the presentation of Mendelian disease in members of the ASMAD pedigree could not be confirmed due to restrictions on re-contacting individuals in this legacy collection. As pleiotropic alleles continue to be identified, future studies would benefit from broadly phenotyping cases to fully capture the combination of traits and diseases present in each individual.

In conclusion, we demonstrate a case of genetic pleiotropy between a complex psychiatric disease with both Mendelian and complex metabolic and endocrine traits. We suggest this indicates a common genetic etiology for these traits in the extended Amish family. Taken together, our results denote that medical co-morbidity between complex diseases and Mendelian disorders arises as a combination of chromosomal proximity of disease causing variants and pleiotropy of disease genes.

Methods

Sample

The genetic-epidemiologic study of bipolar disorder among the Old Order Amish in Pennsylvania (The Amish Study of Major Affective Disorder) has been well documented (Egeland et al., 1990; Hostetter et al., 1983) and the sample collection methods previously described in detail (Georgi et al., 2014; Kember et al., 2015). Briefly, it consists of a large, extended bipolar disorder pedigree of 700 individuals with a small number of founders. Diagnoses of individuals were made following structured interviews (SADS-L) and a review of medical records by a psychiatric board using strict Research Diagnostic Criteria (RDC) and the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV) for uniform clinical criteria (Egeland et al., 1990). The majority of affected individuals in the current pedigree are diagnosed as either BPI, BPII, or Major Depressive Disorder. Collection of blood samples followed diagnostic consensus, and lymphoblastoid cell lines were established by the Coriell Institute of Medical Research (CIMR). Signed informed consents were obtained, using language appropriate for Old Order Amish, to a) access medical records for the Amish Study clinicians exclusively to do diagnostic evaluations and clinical studies, and b) to perform collection of blood/tissue samples. In addition, all work contained within this study was approved by the IRB of the Perelman School of Medicine at the University of Pennsylvania.

Phasing and Imputation

Genotyping was performed on 394 samples from the extended Amish pedigree using Illumina Omni 2.5 M SNP arrays at the Center for Applied Genomics (Children's Hospital of Pennsylvania, Philadelphia, PA). Quality control of the raw genotype calls and imputed genotype calls was conducted using PLINK (Purcell et al., 2007). Individuals were excluded if they (1) had a call rate < 97% or (2) exhibited elevated levels of Mendelian inconsistencies. Variants were excluded from analysis if they: (1) had a call rate $< 97\%$ or (2) had minor allele frequencies $< 0.002\%$ (i.e., were singletons). Genotypes were phased and imputed with SHAPEIT2 (Delaneau et al., 2011) and IMPUTE2 (Howie et al., 2009), respectively, using the Anabaptist reference panel (Hou et al., 2017). Phasing was performed with SHAPEIT's duoHMM option to account for known familial relationships, using the known ASMAD pedigree, with the genetic map from the HapMap phase II (Frazer et al., 2007), in 5Mb windows. Imputation was performed with IMPUTE2 in 5Mb windows with the following options: -use_prephased_g -known_haps_g $SAMPLE.haps -phase -buffer 500'. Following imputation, genotype dosages were converted to hard calls if above/below a threshold of 1.9/0.1, or otherwise set to missing. In total, 2,379,855 variants were imputed in 394 individuals. Quality control of the imputed calls included removing individuals who had a call rate < 97% (0 individuals), setting Mendelian errors to missing (124,668 variants), and removing variants with a call rate of <99% (120,611 variants) or had minor allele frequencies $< 0.002\%$ (869,901 variants). 1,372,783 variants and 394 individuals passed

QC. Imputation accuracy was assessed by comparing to whole genome sequence data available for a subset of family members (99.3% concordance).

Whole genome sequencing

Whole genome sequencing (WGS) for 80 Old Order Amish family members (including 30 parent child trios) was performed by Complete Genomics Inc. (CGI; Mountain View, CA) using a sequence-by-ligation method (Drmanac et al., 2010). Paired-end reads of length 70 bp (35 bp at each end) were mapped to the National Center for Biotechnology Information (NCBI) human reference genome (build 37.2) using a Bayesian mapping pipeline (Carnevali et al., 2012). Variant calls were performed by CGI using version 2.0.3.1 of their pipeline. False discovery rate estimates for SNP calls of the CGI platform are 0.2–0.6% (Drmanac et al., 2010). Gene annotations were based on the NCBI build 37.2 seq_gene file contained in a NCBI annotation build. The variant calls within the WGS were processed using the cgatools software (version 1.5.0, build 31) made available by CGI. The listvar tool was used to generate a master list of the 11.1 M variants present in the 80 Amish samples. The testvar tool was used to determine presence and absence of each variant within the 80 Amish WGS. Only variants with high variant call scores ("VQHIGH" tag in the data files) were included.

Human disease catalog

The Human Genome Mutation Database (HGMD) catalogs known disease associated variants (http://www.hgmd.org/; (Stenson et al., 2014)). Most of the clinical phenotypes in the database are monogenic diseases. In the June 2013 release it contained 141,000 different variants in ~5,700 genes ("HGMD disease genes"). We examined all variants present in ASMAD in 3456 HGMD disease genes ('DM' tag in HGMD).

Curation of HGMD disease causing variants

251 variants present in ASMAD were annotated as being a 'disease mutation' in HGMD. In order to further refine this list to identify true disease causing mutations, we first removed all variants found to be present at >1% frequency in 1000 Genomes (Auton et al., 2015) or ExAC (Lek et al., 2016), assuming that disease causing alleles will be rare in a population (total remaining variants=154). Next, we merged this list with annotation from ClinVar (Landrum et al., 2016) and selected all variants annotated as pathogenic (n=62). We then identified whether each variant caused disease in a recessive or dominant model, and used this information to identify individuals in the ASMAD family predicted to display the disease phenotype based on their allelic status (n=25 variants). For these 25 variants, we expanded upon the HGMD and ClinVar annotation by applying criteria recommended by the American College of Medical Genetics and Genomics (ACMG; (Richards et al., 2015)) for the interpretation of sequence variants, and classified the variants into the categories "pathogenic", "likely pathogenic", "uncertain significance", "likely benign" and "benign". Out of the 25, ACMG criteria classified 3 variants as "benign", 6 variants as "uncertain significance", 10 variants as "likely

pathogenic" and 6 variants as "pathogenic". For analyses we removed the 3 variants classified as benign and retained the others (n=22 variants).

EMMAX and MONSTER

Association analysis for all variants was performed using EMMAX (Version from February 2012, (Kang et al., 2010)), a statistical test for association analysis using mixed models that accounts for the population structure within the sample. Gene-based association tests were performed using MONSTER (Jiang and McPeek, 2014), a statistical test that generalizes the SKAT-O method and uses a mixed effects model to account for population structure.

Loss of function variants

Variants were annotated using VEP LOFTEE (McLaren et al., 2016). We detect 1177 putative protein truncating (frameshift, splice donor, splice acceptor, and stop-gained) variants. Using filters supplied by LOFTEE, we removed 176 variants (see tables below), resulting in 1001 high confidence loss of function variants (HC-LoF).

Filter	Description	Number of variants
END_TRUNC	Variant falls in the last 5% of the transcript	28
EXON_INTRON_UNDEF	Exon or intron boundaries undefined for this transcript	1
NON_CAN_SPLICE	Variant falls in non-canonical splice site	3
NON_CAN_SPLICE_SURR	Variant falls in exon with non-canonical splice site	7

Flag	Description	Number of variants
NAGNAG_SITE	Splice acceptor has in-frame AG acceptor site one codon away	24
SINGLE_EXON	Variant falls in a single exon transcript	113

We identified 167 HC-LoF variants in disease genes, of which 71 are in a homozygous state in one or more individuals. There are 83 rare (<1% in ExAC and 1000G) HC-LoF in disease genes, 17 of which are found in a homozygous state in one or more individuals.

Association rule discovery

Association rules for the co-morbidity of mood disorder and Mendelian diseases in individuals within the ASMAD pedigree were determined using the apriori algorithm from the arules package in R (Hahsler et al., 2005). An itemset was created for each individual, consisting of the affected status for mood disorder ("unaffected" or "broad extended" phenotype) and any co-morbid Mendelian disease as determined by the allelic

status of the disease causing variant in that individual. Association rules for the frequent itemsets were generated using the apriori command. Rules were then limited to those with either "broad extended" or "unaffected" on the left hand side of the rule (antecedents) and Mendelian diseases on the right hand side of the rule (consequents).

Polygenic risk scores

A polygenic risk score is generated for each individual as the sum of all variants they carry, weighted by the effect that variant has on phenotype. Polygenic risk scores were generated using the PRSice package (Euesden et al., 2015), with multiple GWAS summary statistics as the base dataset (see Supplemental table 9), and imputed whole genome sequence data in ASMAD as the target dataset. As recommended in the software, we performed p-value informed clumping on the genotype data with an r2 = 0.1 and a distance threshold of 250kb, following exclusion of the MHC region. The optimal p-value threshold for PRS was defined as that which explained the most phenotypic variation for mood disorder, out of a set of pre-determined thresholds ($p \leq 0.0001$, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5). For traits identified as significantly associated with mood disorder in this pedigree, we show barplots indicating the model fit and p-value for association at each p-value threshold tested (Supplemental figure 27).

Statistical analysis

Polygenic risk scores for the optimal p-value threshold for each trait were standardized mean=0 and standard deviation=1. Linear mixed model analyses were selected to model outcomes (Narrow phenotype, Broad phenotype, Broad extended phenotype, Depression phenotype) while accounting for relatedness. The analysis was performed using the pedigreemm package in R (Vazquez et al., 2010), with PRS as the independent variable and phenotype as the dependent variable. An empirical kinship matrix constructed from a vcf file by the Balding-Nicols method using rvtests (Zhan et al., 2016) was fitted as a random effect to account for relatedness between individuals. Risk scores were evaluated for 22 traits, and so a bonferroni corrected p-value<0.0023 (0.05/22) was selected as the level for statistical significance.

Polygenic transmission disequilibrium

A test for polygenic transmission disequilibrium has recently been described by Weiner et al. (Weiner et al., 2017). We modified this test to allow the comparison of multiple affected and unaffected siblings. We selected all nuclear families with at least one child with Broad phenotype (BPI, BPII, BP:NOS, MDDR, number of families=46). For each family, the polygenic transmission disequilibrium deviation was calculated as follows:

$$pTDTdeviation = \frac{\text{Average PRS affected} - \text{Average PRS unaffected}}{SD(PRS\ all\ siblings)}$$

The test statistic for pTDT deviation was then calculated based on all families as follows:

$$tpTDT = \frac{\text{Mean(pTDT deviation)}}{SD(pTDT\ deviation)/\sqrt{n}}$$

Where n is the number of families.

Risk score chromosome windows

Drawing on work from Shi et al. (Shi et al., 2016), on local genetic correlation, we developed a method for establishing local polygenic risk. Files containing summary statistics for GWAS of bipolar disorder, thyroid levels, lipid traits and diabetes were split into chromosomal regions using a sliding window (5 Mb window, sliding 2.5Mb). Risk scores were then generated for each individual in the pedigree using just the summary statistics for each 5Mb genetic region. As above (see Methods: statistical analysis), risk scores were standardized and tested using linear mixed model analyses to account for relatedness. Around 1,112 regions were tested per score (556 non-overlapping regions). P-values reported for risk score regions are uncorrected for multiple testing, since many non-overlapping regions are still highly correlated due to LD structure. A conservative Bonferroni threshold accounting for 556 independent tests at $P<0.05$ is $p<0.00009$. We chose a less stringent cut-off of $P<0.01$ for reporting these results.

Acknowledgments

The authors are indebted to the members of the Old Order Amish settlements who participated in the Amish Study of Major Affective Disorder and Dr. Janice Egeland who designed and directed the initial recruitments and clinical evaluation. This study was supported by the NIH grant R01MH093415, and supported in part by the NIMH Intramural Research Program. The authors are thankful to Drs. Wade Berrettini, Elliot Gershon and Daniel Rader for informative discussions and their feedback on the manuscript. We are grateful to Marco Medici and Anne Cappola, and the many other studies and consortia that made their summary statistics readily available.

References

Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature *526*, 68-74. 4750478

Beyer, J., Kuchibhatla, M., Gersing, K., and Krishnan, K.R. (2005). Medical comorbidity in a bipolar outpatient clinical population. Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology *30*, 401-404.

Blair, D.R., Lyttle, C.S., Mortensen, J.M., Bearden, C.F., Jensen, A.B., Khiabanian, H., Melamed, R., Rabadan, R., Bernstam, E.V., Brunak, S., Jensen, L.J., Nicolae, D., Shah, N.H., Grossman, R.L., Cox, N.J., White, K.P., and Rzhetsky, A. (2013). A nondegenerate code of deleterious variants in Mendelian loci contributes to complex disease risk. Cell *155*, 70-80. 3844554

Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.R., Duncan, L., Perry, J.R., Patterson, N., Robinson, E.B., Daly, M.J., Price, A.L., and Neale, B.M. (2015). An atlas of genetic correlations across human diseases and traits. Nature genetics *47*, 1236-1241. 4797329

Carnevali, P., Baccash, J., Halpern, A.L., Nazarenko, I., Nilsen, G.B., Pant, K.P., Ebert, J.C., Brownley, A., Morenzoni, M., Karpinchyk, V., Martin, B., Ballinger, D.G., and Drmanac, R. (2012). Computational techniques for human genome resequencing using mated gapped reads. Journal of computational biology : a journal of computational molecular cell biology *19*, 279-292.

Chakrabarti, S. (2011). Thyroid functions and bipolar affective disorder. Journal of thyroid research *2011*, 306367. 3144691

Chuang, L.C., and Kuo, P.H. (2017). Building a genetic risk model for bipolar disorder from genome-wide association data with random forest algorithm. Scientific reports *7*, 39943. 5206749

Delaneau, O., Marchini, J., and Zagury, J.F. (2011). A linear complexity phasing method for thousands of genomes. Nature methods *9*, 179-181.

Drmanac, R., Sparks, A.B., Callow, M.J., Halpern, A.L., Burns, N.L., Kermani, B.G., Carnevali, P., Nazarenko, I., Nilsen, G.B., Yeung, G., Dahl, F., Fernandez, A., Staker, B., Pant, K.P., Baccash, J., Borcherding, A.P., Brownley, A., Cedeno, R., Chen, L., Chernikoff, D., Cheung, A., Chirita, R., Curson, B., Ebert, J.C., Hacker, C.R., Hartlage, R., Hauser, B., Huang, S., Jiang, Y., Karpinchyk, V., Koenig, M., Kong, C., Landers, T., Le, C., Liu, J., McBride, C.E., Morenzoni, M., Morey, R.E., Mutch, K., Perazich, H., Perry, K., Peters, B.A., Peterson, J., Pethiyagoda, C.L., Pothuraju, K., Richter, C., Rosenbaum, A.M., Roy, S., Shafto, J., Sharanhovich, U., Shannon, K.W., Sheppy, C.G., Sun, M., Thakuria, J.V., Tran, A., Vu, D., Zaranek, A.W., Wu, X., Drmanac, S.,

Oliphant, A.R., Banyai, W.C., Martin, B., Ballinger, D.G., Church, G.M., and Reid, C.A. (2010). Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science *327*, 78-81.

Egeland, J.A., Sussex, J.N., Endicott, J., Hostetter, A.M., Offord, D.R., Schwab, J.J., Allen, C.R., and Pauls, D.L. (1990). The Impact of Diagnoses on Genetic Linkage Study for Bipolar Affective Disorders among the Amish. Psychiatric Genetics *1*, 5&hyhen;18.

Enger, C., Jones, M.E., Kryzhanovskaya, L., Doherty, M., and McAfee, A.T. (2013). Risk of developing diabetes and dyslipidemia among adolescents with bipolar disorder or schizophrenia. International journal of adolescent medicine and health *25*, 3-11.

Euesden, J., Lewis, C.M., and O'Reilly, P.F. (2015). PRSice: Polygenic Risk Score software. Bioinformatics *31*, 1466-1468. 4410663

Fiedorowicz, J.G., Palagummi, N.M., Forman-Hoffman, V.L., Miller, D.D., and Haynes, W.G. (2008). Elevated prevalence of obesity, metabolic syndrome, and cardiovascular risk factors in bipolar disorder. Annals of clinical psychiatry : official journal of the American Academy of Clinical Psychiatrists *20*, 131-137. 2776768

Forty, L., Ulanova, A., Jones, L., Jones, I., Gordon-Smith, K., Fraser, C., Farmer, A., McGuffin, P., Lewis, C.M., Hosang, G.M., Rivera, M., and Craddock, N. (2014). Comorbid medical illness in bipolar disorder. The British journal of psychiatry : the journal of mental science *205*, 465-472. 4248234

Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., Pasternak, S., Wheeler, D.A., Willis, T.D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhao, H., Zhou, J., Gabriel, S.B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R.C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Shen, Y., Sun, W., Wang, H., Wang, Y., Wang, Y., Xiong, X., Xu, L., Waye, M.M., Tsui, S.K., Xue, H., Wong, J.T., Galver, L.M., Fan, J.B., Gunderson, K., Murray, S.S., Oliphant, A.R., Chee, M.S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J.F., Phillips, M.S., Roumy, S., Sallee, C., Verner, A., Hudson, T.J., Kwok, P.Y., Cai, D., Koboldt, D.C., Miller, R.D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L.C., Mak, W., Song, Y.Q., Tam, P.K., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A., Tanaka, T., Tsunoda, T., Deloukas, P., Bird, C.P., Delgado, M., Dermitzakis, E.T., Gwilliam, R., Hunt, S., Morrison, J., Powell, D., Stranger, B.E., Whittaker, P., Bentley, D.R., Daly, M.J., de Bakker, P.I., Barrett, J., Chretien, Y.R., Maller, J., McCarroll, S., Patterson, N., Pe'er, I., Price, A., Purcell, S., Richter, D.J., Sabeti, P., Saxena, R., Schaffner, S.F., Sham, P.C., Varilly, P., Altshuler, D., Stein, L.D., Krishnan, L., Smith, A.V., Tello-Ruiz, M.K., Thorisson, G.A., Chakravarti, A., Chen, P.E., Cutler, D.J., Kashuk, C.S., Lin, S., Abecasis, G.R., Guan, W., Li, Y., Munro, H.M., Qin, Z.S., Thomas, D.J., McVean, G., Auton, A., Bottolo, L., Cardin, N., Eyheramendy, S.,

Freeman, C., Marchini, J., Myers, S., Spencer, C., Stephens, M., Donnelly, P., Cardon, L.R., Clarke, G., Evans, D.M., Morris, A.P., Weir, B.S., Tsunoda, T., Mullikin, J.C., Sherry, S.T., Feolo, M., Skol, A., Zhang, H., Zeng, C., Zhao, H., Matsuda, I., Fukushima, Y., Macer, D.R., Suda, E., Rotimi, C.N., Adebamowo, C.A., Ajayi, I., Aniagwu, T., Marshall, P.A., Nkwodimmah, C., Royal, C.D., Leppert, M.F., Dixon, M., Peiffer, A., Qiu, R., Kent, A., Kato, K., Niikawa, N., Adewole, I.F., Knoppers, B.M., Foster, M.W., Clayton, E.W., Watkin, J., Gibbs, R.A., Belmont, J.W., Muzny, D., Nazareth, L., Sodergren, E., Weinstock, G.M., Wheeler, D.A., Yakub, I., Gabriel, S.B., Onofrio, R.C., Richter, D.J., Ziaugra, L., Birren, B.W., Daly, M.J., Altshuler, D., Wilson, R.K., Fulton, L.L., Rogers, J., Burton, J., Carter, N.P., Clee, C.M., Griffiths, M., Jones, M.C., McLay, K., Plumb, R.W., Ross, M.T., Sims, S.K., Willey, D.L., Chen, Z., Han, H., Kang, L., Godbout, M., Wallenburg, J.C., L'Archeveque, P., Bellemare, G., Saeki, K., Wang, H., An, D., Fu, H., Li, Q., Wang, Z., Wang, R., Holden, A.L., Brooks, L.D., McEwen, J.E., Guyer, M.S., Wang, V.O., Peterson, J.L., Shi, M., Spiegel, J., Sung, L.M., Zacharia, L.F., Collins, F.S., Kennedy, K., Jamieson, R., and Stewart, J. (2007). A second generation human haplotype map of over 3.1 million SNPs. Nature *449*, 851-861. 2689609

Gale, C.R., Hagenaars, S.P., Davies, G., Hill, W.D., Liewald, D.C., Cullen, B., Penninx, B.W., Boomsma, D.I., Pell, J., McIntosh, A.M., Smith, D.J., Deary, I.J., and Harris, S.E. (2016). Pleiotropy between neuroticism and physical and mental health: findings from 108 038 men and women in UK Biobank. Translational psychiatry *6*, e791. 4872414

Georgi, B., Craig, D., Kember, R.L., Liu, W., Lindquist, I., Nasser, S., Brown, C., Egeland, J.A., Paul, S.M., and Bucan, M. (2014). Genomic view of bipolar disorder revealed by whole genome sequencing in a genetic isolate. PLoS genetics *10*, e1004229. 3953017

Ginns, E.I., Galdzicka, M., Elston, R.C., Song, Y.E., Paul, S.M., and Egeland, J.A. (2015). Disruption of sonic hedgehog signaling in Ellis-van Creveld dwarfism confers protection against bipolar affective disorder. Molecular psychiatry *20*, 1212-1218.

Goes, F.S., McGrath, J., Avramopoulos, D., Wolyniec, P., Pirooznia, M., Ruczinski, I., Nestadt, G., Kenny, E.E., Vacic, V., Peters, I., Lencz, T., Darvasi, A., Mulle, J.G., Warren, S.T., and Pulver, A.E. (2015). Genome-wide association study of schizophrenia in Ashkenazi Jews. American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics *168*, 649-659.

Golden, S.H., Lazo, M., Carnethon, M., Bertoni, A.G., Schreiner, P.J., Diez Roux, A.V., Lee, H.B., and Lyketsos, C. (2008). Examining a bidirectional association between depressive symptoms and diabetes. Jama *299*, 2751-2759. 2648841

Gratten, J., Wray, N.R., Keller, M.C., and Visscher, P.M. (2014). Large-scale genomics unveils the genetic architecture of psychiatric disorders. Nature neuroscience *17*, 782-790. 4112149

Hahsler, M., Grün, B., and Hornik, K. (2005). arules - A Computational Environment for Mining Association Rules and Frequent Item Sets. 2005 *14*, 25.

Hammerschlag, A.R., Stringer, S., de Leeuw, C.A., Sniekers, S., Taskesen, E., Watanabe, K., Blanken, T.F., Dekker, K., Te Lindert, B.H.W., Wassing, R., Jonsdottir, I., Thorleifsson, G., Stefansson, H., Gislason, T., Berger, K., Schormair, B., Wellmann, J., Winkelmann, J., Stefansson, K., Oexle, K., Van Someren, E.J.W., and Posthuma, D. (2017). Genome-wide association analysis of insomnia complaints identifies risk genes and genetic overlap with psychiatric and metabolic traits. Nature genetics.

Han, B., Pouget, J.G., Slowikowski, K., Stahl, E., Lee, C.H., Diogo, D., Hu, X., Park, Y.R., Kim, E., Gregersen, P.K., Dahlqvist, S.R., Worthington, J., Martin, J., Eyre, S., Klareskog, L., Huizinga, T., Chen, W.M., Onengut-Gumuscu, S., Rich, S.S., Wray, N.R., and Raychaudhuri, S. (2016). A method to decipher pleiotropy by detecting underlying heterogeneity driven by hidden subgroups applied to autoimmune and neuropsychiatric diseases. Nature genetics *48*, 803-810. 4925284

Hostetter, A.M., Egeland, J.A., and Endicott, J. (1983). Amish Study, II: Consensus diagnoses and reliability results. The American journal of psychiatry *140*, 62-66.

Hou, L., Faraci, G., Chen, D.T., Kassem, L., Schulze, T.G., Shugart, Y.Y., and McMahon, F.J. (2013). Amish revisited: next-generation sequencing studies of psychiatric disorders among the Plain people. Trends in genetics : TIG *29*, 412-418. 3941079

Hou, L., Kember, R.L., Roach, J.C., O'Connell, J.R., Craig, D.W., Bucan, M., Scott, W.K., Pericak-Vance, M., Haines, J.L., Crawford, M.H., Shuldiner, A.R., and McMahon, F.J. (2017). A population-specific reference panel empowers genetic studies of Anabaptist populations. Scientific reports *7*, 6079. 5519631

Howard, D.M., Clarke, T.-K., Adams, M.J., Hafferty, J.D., Wigmore, E.M., Zeng, Y., Hall, L.S., Gibson, J., Boutin, T.S., Hayward, C., Thomson, P.A., Porteous, D.J., Smith, B.H., Murray, A.D., Haley, C.S., Deary, I.J., Whalley, H.C., and McIntosh, A.M. (2017). The Stratification Of Major Depressive Disorder Into Genetic Subgroups. bioRxiv.

Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS genetics *5*, e1000529. 2689936

Jaume, J.C., Guo, J., Pauls, D.L., Zakarija, M., McKenzie, J.M., Egeland, J.A., Burek, C.L., Rose, N.R., Hoffman, W.H., Rapoport, B., and McLachlan, S.M. (1999). Evidence for genetic transmission of thyroid peroxidase autoantibody epitopic "fingerprints". The Journal of clinical endocrinology and metabolism *84*, 1424-1431.

Jiang, D., and McPeek, M.S. (2014). Robust rare variant association testing for quantitative traits in samples with related individuals. Genetic epidemiology *38*, 10-20. 4510991

Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S.Y., Freimer, N.B., Sabatti, C., and Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. Nature genetics *42*, 348-354. 3092069

Kember, R.L., Georgi, B., Bailey-Wilson, J.E., Stambolian, D., Paul, S.M., and Bucan, M. (2015). Copy number variants encompassing Mendelian disease genes in a large multigenerational family segregating bipolar disorder. BMC genetics *16*, 27. 4382929

Koivukoski, L., Fisher, S.A., Kanninen, T., Lewis, C.M., von Wowern, F., Hunt, S., Kardia, S.L., Levy, D., Perola, M., Rankinen, T., Rao, D.C., Rice, T., Thiel, B.A., and Melander, O. (2004). Meta-analysis of genome-wide scans for hypertension and blood pressure in Caucasians shows evidence of susceptibility regions on chromosomes 2 and 3. Human molecular genetics *13*, 2325-2332.

Landrum, M.J., Lee, J.M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Hoover, J., Jang, W., Katz, K., Ovetsky, M., Riley, G., Sethi, A., Tully, R., Villamarin-Salomon, R., Rubinstein, W., and Maglott, D.R. (2016). ClinVar: public archive of interpretations of clinically relevant variants. Nucleic acids research *44*, D862-868. 4702865

Lee, S.H., Ripke, S., Neale, B.M., Faraone, S.V., Purcell, S.M., Perlis, R.H., Mowry, B.J., Thapar, A., Goddard, M.E., Witte, J.S., Absher, D., Agartz, I., Akil, H., Amin, F., Andreassen, O.A., Anjorin, A., Anney, R., Anttila, V., Arking, D.E., Asherson, P., Azevedo, M.H., Backlund, L., Badner, J.A., Bailey, A.J., Banaschewski, T., Barchas, J.D., Barnes, M.R., Barrett, T.B., Bass, N., Battaglia, A., Bauer, M., Bayes, M., Bellivier, F., Bergen, S.E., Berrettini, W., Betancur, C., Bettecken, T., Biederman, J., Binder, E.B., Black, D.W., Blackwood, D.H., Bloss, C.S., Boehnke, M., Boomsma, D.I., Breen, G., Breuer, R., Bruggeman, R., Cormican, P., Buccola, N.G., Buitelaar, J.K., Bunney, W.E., Buxbaum, J.D., Byerley, W.F., Byrne, E.M., Caesar, S., Cahn, W., Cantor, R.M., Casas, M., Chakravarti, A., Chambert, K., Choudhury, K., Cichon, S., Cloninger, C.R., Collier, D.A., Cook, E.H., Coon, H., Cormand, B., Corvin, A., Coryell, W.H., Craig, D.W., Craig, I.W., Crosbie, J., Cuccaro, M.L., Curtis, D., Czamara, D., Datta, S., Dawson, G., Day, R., De Geus, E.J., Degenhardt, F., Djurovic, S., Donohoe, G.J., Doyle, A.E., Duan, J., Dudbridge, F., Duketis, E., Ebstein, R.P., Edenberg, H.J., Elia, J., Ennis, S., Etain, B., Fanous, A., Farmer, A.E., Ferrier, I.N., Flickinger, M., Fombonne, E., Foroud, T., Frank, J., Franke, B., Fraser, C., Freedman, R., Freimer, N.B., Freitag, C.M., Friedl, M., Frisen, L., Gallagher, L., Gejman, P.V., Georgieva, L., Gershon, E.S., Geschwind, D.H., Giegling, I., Gill, M., Gordon, S.D., Gordon-Smith, K., Green, E.K., Greenwood, T.A., Grice, D.E., Gross, M., Grozeva, D., Guan, W., Gurling, H., De Haan, L., Haines, J.L., Hakonarson, H., Hallmayer, J., Hamilton, S.P., Hamshere, M.L., Hansen, T.F., Hartmann, A.M., Hautzinger, M., Heath, A.C., Henders, A.K., Herms, S., Hickie, I.B., Hipolito, M., Hoefels, S., Holmans, P.A., Holsboer, F., Hoogendijk, W.J., Hottenga, J.J., Hultman, C.M., Hus, V., Ingason, A., Ising, M., Jamain, S., Jones, E.G., Jones, I., Jones, L., Tzeng, J.Y., Kahler, A.K., Kahn, R.S., Kandaswamy, R., Keller, M.C., Kennedy, J.L., Kenny, E., Kent, L., Kim, Y., Kirov, G.K., Klauck, S.M., Klei, L., Knowles, J.A., Kohli, M.A., Koller, D.L., Konte, B., Korszun, A., Krabbendam, L., Krasucki, R., Kuntsi, J., Kwan, P., Landen, M., Langstrom, N., Lathrop, M., Lawrence, J., Lawson, W.B.,

Leboyer, M., Ledbetter, D.H., Lee, P.H., Lencz, T., Lesch, K.P., Levinson, D.F., Lewis, C.M., Li, J., Lichtenstein, P., Lieberman, J.A., Lin, D.Y., Linszen, D.H., Liu, C., Lohoff, F.W., Loo, S.K., Lord, C., Lowe, J.K., Lucae, S., MacIntyre, D.J., Madden, P.A., Maestrini, E., Magnusson, P.K., Mahon, P.B., Maier, W., Malhotra, A.K., Mane, S.M., Martin, C.L., Martin, N.G., Mattheisen, M., Matthews, K., Mattingsdal, M., McCarroll, S.A., McGhee, K.A., McGough, J.J., McGrath, P.J., McGuffin, P., McInnis, M.G., McIntosh, A., McKinney, R., McLean, A.W., McMahon, F.J., McMahon, W.M., McQuillin, A., Medeiros, H., Medland, S.E., Meier, S., Melle, I., Meng, F., Meyer, J., Middeldorp, C.M., Middleton, L., Milanova, V., Miranda, A., Monaco, A.P., Montgomery, G.W., Moran, J.L., Moreno-De-Luca, D., Morken, G., Morris, D.W., Morrow, E.M., Moskvina, V., Muglia, P., Muhleisen, T.W., Muir, W.J., Muller-Myhsok, B., Murtha, M., Myers, R.M., Myin-Germeys, I., Neale, M.C., Nelson, S.F., Nievergelt, C.M., Nikolov, I., Nimgaonkar, V., Nolen, W.A., Nothen, M.M., Nurnberger, J.I., Nwulia, E.A., Nyholt, D.R., O'Dushlaine, C., Oades, R.D., Olincy, A., Oliveira, G., Olsen, L., Ophoff, R.A., Osby, U., Owen, M.J., Palotie, A., Parr, J.R., Paterson, A.D., Pato, C.N., Pato, M.T., Penninx, B.W., Pergadia, M.L., Pericak-Vance, M.A., Pickard, B.S., Pimm, J., Piven, J., Posthuma, D., Potash, J.B., Poustka, F., Propping, P., Puri, V., Quested, D.J., Quinn, E.M., Ramos-Quiroga, J.A., Rasmussen, H.B., Raychaudhuri, S., Rehnstrom, K., Reif, A., Ribases, M., Rice, J.P., Rietschel, M., Roeder, K., Roeyers, H., Rossin, L., Rothenberger, A., Rouleau, G., Ruderfer, D., Rujescu, D., Sanders, A.R., Sanders, S.J., Santangelo, S.L., Sergeant, J.A., Schachar, R., Schalling, M., Schatzberg, A.F., Scheftner, W.A., Schellenberg, G.D., Scherer, S.W., Schork, N.J., Schulze, T.G., Schumacher, J., Schwarz, M., Scolnick, E., Scott, L.J., Shi, J., Shilling, P.D., Shyn, S.I., Silverman, J.M., Slager, S.L., Smalley, S.L., Smit, J.H., Smith, E.N., Sonuga-Barke, E.J., St Clair, D., State, M., Steffens, M., Steinhausen, H.C., Strauss, J.S., Strohmaier, J., Stroup, T.S., Sutcliffe, J.S., Szatmari, P., Szelinger, S., Thirumalai, S., Thompson, R.C., Todorov, A.A., Tozzi, F., Treutlein, J., Uhr, M., van den Oord, E.J., Van Grootheest, G., Van Os, J., Vicente, A.M., Vieland, V.J., Vincent, J.B., Visscher, P.M., Walsh, C.A., Wassink, T.H., Watson, S.J., Weissman, M.M., Werge, T., Wienker, T.F., Wijsman, E.M., Willemsen, G., Williams, N., Willsey, A.J., Witt, S.H., Xu, W., Young, A.H., Yu, T.W., Zammit, S., Zandi, P.P., Zhang, P., Zitman, F.G., Zollner, S., Devlin, B., Kelsoe, J.R., Sklar, P., Daly, M.J., O'Donovan, M.C., Craddock, N., Sullivan, P.F., Smoller, J.W., Kendler, K.S., and Wray, N.R. (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nature genetics *45*, 984-994. 3800159

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., Tukiainen, T., Birnbaum, D.P., Kosmicki, J.A., Duncan, L.E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., Cooper, D.N., Deflaux, N., DePristo, M., Do, R., Flannick, J., Fromer, M., Gauthier, L., Goldstein, J., Gupta, N., Howrigan, D., Kiezun, A., Kurki, M.I., Moonshine, A.L., Natarajan, P., Orozco, L., Peloso, G.M., Poplin, R., Rivas, M.A., Ruano-Rubio, V., Rose, S.A., Ruderfer, D.M., Shakir, K., Stenson, P.D., Stevens, C., Thomas, B.P., Tiao, G., Tusie-Luna, M.T., Weisburd, B., Won, H.H., Yu, D., Altshuler, D.M., Ardissino, D., Boehnke, M., Danesh, J., Donnelly, S., Elosua, R., Florez, J.C., Gabriel, S.B., Getz, G., Glatt, S.J., Hultman, C.M., Kathiresan, S., Laakso, M., McCarroll, S., McCarthy, M.I., McGovern, D., McPherson, R., Neale, B.M., Palotie, A., Purcell, S.M., Saleheen, D., Scharf, J.M., Sklar, P., Sullivan, P.F., Tuomilehto, J., Tsuang, M.T., Watkins, H.C.,

Wilson, J.G., Daly, M.J., and MacArthur, D.G. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285-291. 5018207

Lo, M.T., Hinds, D.A., Tung, J.Y., Franz, C., Fan, C.C., Wang, Y., Smeland, O.B., Schork, A., Holland, D., Kauppi, K., Sanyal, N., Escott-Price, V., Smith, D.J., O'Donovan, M., Stefansson, H., Bjornsdottir, G., Thorgeirsson, T.E., Stefansson, K., McEvoy, L.K., Dale, A.M., Andreassen, O.A., and Chen, C.H. (2017). Genome-wide analyses for personality traits identify six genomic loci and show correlations with psychiatric disorders. Nature genetics *49*, 152-156. 5278898

Lupski, J.R., Belmont, J.W., Boerwinkle, E., and Gibbs, R.A. (2011). Clan genomics and the complex architecture of human disease. Cell *147*, 32-43. 3656718

McKusick, V.A., Egeland, J.A., Eldridge, R., and Krusen, D.E. (1964). Dwarfism in the Amish I. The Ellis-Van Creveld Syndrome. Bulletin of the Johns Hopkins Hospital *115*, 306-336.

McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome biology *17*, 122. 4893825

O'Donovan, M.C., and Owen, M.J. (2016). The implications of the shared genetics of psychiatric disorders. Nature medicine *22*, 1214-1219.

Pagnamenta, A.T., Khan, H., Walker, S., Gerrelli, D., Wing, K., Bonaglia, M.C., Giorda, R., Berney, T., Mani, E., Molteni, M., Pinto, D., Le Couteur, A., Hallmayer, J., Sutcliffe, J.S., Szatmari, P., Paterson, A.D., Scherer, S.W., Vieland, V.J., and Monaco, A.P. (2011). Rare familial 16q21 microdeletions under a linkage peak implicate cadherin 8 (CDH8) in susceptibility to autism and learning disability. Journal of medical genetics *48*, 48-54. 3003876

Parks, J., Svendsen, D., Singer, P., Foti, M., and Mauer, B. (2006). Morbidity and Mortality in People with Serious Mental Illness. In National Association of State Mental Health Program Directors.

Patel, N.C., Hariparsad, M., Matias-Akthar, M., Sorter, M.T., Barzman, D.H., Morrison, J.A., Stanford, K.E., Strakowski, S.M., and DelBello, M.P. (2007). Body mass indexes and lipid profiles in hospitalized children and adolescents exposed to atypical antipsychotics. Journal of child and adolescent psychopharmacology *17*, 303-311.

Pickrell, J.K., Berisa, T., Liu, J.Z., Segurel, L., Tung, J.Y., and Hinds, D.A. (2016). Detection and interpretation of shared genetic influences on 42 human traits. Nature genetics *48*, 709-717. 5207801

Prieto, M.L., Ryu, E., Jenkins, G.D., Batzler, A., Nassan, M.M., Cuellar-Barboza, A.B., Pathak, J., McElroy, S.L., Frye, M.A., and Biernacka, J.M. (2016). Leveraging electronic health records to study pleiotropic effects on bipolar disorder and medical comorbidities. Translational psychiatry *6*, e870. 5022084

Puffenberger, E.G., Jinks, R.N., Sougnez, C., Cibulskis, K., Willert, R.A., Achilly, N.P., Cassidy, R.P., Fiorentini, C.J., Heiken, K.F., Lawrence, J.J., Mahoney, M.H., Miller, C.J., Nair, D.T., Politi, K.A., Worcester, K.N., Setton, R.A., Dipiazza, R., Sherman, E.A., Eastman, J.T., Francklyn, C., Robey-Bond, S., Rider, N.L., Gabriel, S., Morton, D.H., and Strauss, K.A. (2012). Genetic mapping and exome sequencing identify variants associated with five novel diseases. PloS one *7*, e28936. 3260153

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics *81*, 559-575. 1950838

Rampersaud, E., Damcott, C.M., Fu, M., Shen, H., McArdle, P., Shi, X., Shelton, J., Yin, J., Chang, Y.P., Ott, S.H., Zhang, L., Zhao, Y., Mitchell, B.D., O'Connell, J., and Shuldiner, A.R. (2007). Identification of novel candidate genes for type 2 diabetes from a genome-wide association scan in the Old Order Amish: evidence for replication from diabetes-related quantitative traits and from independent populations. Diabetes *56*, 3053-3062.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., and Rehm, H.L. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genetics in medicine : official journal of the American College of Medical Genetics *17*, 405-424. 4544753

Rudd, D., Axelsen, M., Epping, E.A., Andreasen, N., and Wassink, T. (2015). Childhood-onset schizophrenia case with 2.2 Mb deletion at chromosome 3p12.2–p12.1 and two large chromosomal abnormalities at 16q22.3–q24.3 and Xq23–q28. Clinical Case Reports *3*, 201-207.

Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421-427. 4112379

Shen, H., Damcott, C.M., Rampersaud, E., Pollin, T.I., Horenstein, R.B., McArdle, P.F., Peyser, P.A., Bielak, L.F., Post, W.S., Chang, Y.P., Ryan, K.A., Miller, M., Rumberger, J.A., Sheedy, P.F., 2nd, Shelton, J., O'Connell, J.R., Shuldiner, A.R., and Mitchell, B.D. (2010). Familial defective apolipoprotein B-100 and increased low-density lipoprotein cholesterol and coronary artery calcification in the old order amish. Archives of internal medicine *170*, 1850-1855. 3587042

Shi, H., Mancuso, N., Spendlove, S., and Pasaniuc, B. (2016). Local genetic correlation gives insights into the shared genetic architecture of complex traits. bioRxiv.

Sivakumaran, S., Agakov, F., Theodoratou, E., Prendergast, J.G., Zgaga, L., Manolio, T., Rudan, I., McKeigue, P., Wilson, J.F., and Campbell, H. (2011). Abundant pleiotropy in

human complex diseases and traits. American journal of human genetics *89*, 607-618. 3213397

Solovieff, N., Cotsapas, C., Lee, P.H., Purcell, S.M., and Smoller, J.W. (2013). Pleiotropy in complex traits: challenges and strategies. Nature reviews Genetics *14*, 483-495. 4104202

Stahl, E., Forstner, A., McQuillin, A., Ripke, S., Ophoff, R., Scott, L., Cichon, S., Andreassen, O.A., Sklar, P., Kelsoe, J., and Breen, G. (2017). Genomewide association study identifies 30 loci associated with bipolar disorder. bioRxiv.

Steinle, N.I., Pollin, T.I., O'Connell, J.R., Mitchell, B.D., and Shuldiner, A.R. (2005). Variants in the ghrelin gene are associated with metabolic syndrome in the Old Order Amish. The Journal of clinical endocrinology and metabolism *90*, 6672-6677.

Stenson, P.D., Mort, M., Ball, E.V., Shaw, K., Phillips, A., and Cooper, D.N. (2014). The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. Human genetics *133*, 1-9. 3898141

Strauss, K.A., Markx, S., Georgi, B., Paul, S.M., Jinks, R.N., Hoshi, T., McDonald, A., First, M.B., Liu, W., Benkert, A.R., Heaps, A.D., Tian, Y., Chakravarti, A., Bucan, M., and Puffenberger, E.G. (2014). A population-based study of KCNH7 p.Arg394His and bipolar spectrum disorder. Human molecular genetics *23*, 6395-6406. 4222358

Strauss, K.A., and Puffenberger, E.G. (2009). Genetics, medicine, and the Plain people. Annual review of genomics and human genetics *10*, 513-536.

Strauss, K.A., Puffenberger, E.G., and Morton, D.H. (2012). One community's effort to control genetic disease. American journal of public health *102*, 1300-1306. 3477994

Vancampfort, D., Vansteelandt, K., Correll, C.U., Mitchell, A.J., De Herdt, A., Sienaert, P., Probst, M., and De Hert, M. (2013). Metabolic syndrome and metabolic abnormalities in bipolar disorder: a meta-analysis of prevalence rates and moderators. The American journal of psychiatry *170*, 265-274.

Vazquez, A.I., Bates, D.M., Rosa, G.J., Gianola, D., and Weigel, K.A. (2010). Technical note: an R package for fitting generalized linear mixed models in animal breeding. Journal of animal science *88*, 497-504.

Weiner, D.J., Wigdor, E.M., Ripke, S., Walters, R.K., Kosmicki, J.A., Grove, J., Samocha, K.E., Goldstein, J.I., Okbay, A., Bybjerg-Grauholm, J., Werge, T., Hougaard, D.M., Taylor, J., Skuse, D., Devlin, B., Anney, R., Sanders, S.J., Bishop, S., Mortensen, P.B., Borglum, A.D., Smith, G.D., Daly, M.J., and Robinson, E.B. (2017). Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. Nature genetics *49*, 978-985. 5552240

Whybrow, P.C., Prange, A.J., Jr., and Treadway, C.R. (1969). Mental changes accompanying thyroid gland dysfunction. A reappraisal using objective psychological measurement. Archives of general psychiatry *20*, 48-63.

Wray, N.R., Lee, S.H., Mehta, D., Vinkhuyzen, A.A., Dudbridge, F., and Middeldorp, C.M. (2014). Research review: Polygenic methods and their application to psychiatric traits. Journal of child psychology and psychiatry, and allied disciplines *55*, 1068-1087.

Xu, H., Ryan, K.A., Jaworek, T.J., Southam, L., Reid, J.G., Overton, J.D., Baras, A., Puurunen, M.K., Zeggini, E., Taylor, S.I., Shuldiner, A.R., and Mitchell, B.D. (2017). Familial Hypercholesterolemia and Type 2 Diabetes in the Old Order Amish. Diabetes *66*, 2054-2058. 5482078

Zhan, X., Hu, Y., Li, B., Abecasis, G.R., and Liu, D.J. (2016). RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. Bioinformatics *32*, 1423-1426. 4848408

Zhu, X., Need, A.C., Petrovski, S., and Goldstein, D.B. (2014). One gene, many neuropsychiatric disorders: lessons from Mendelian diseases. Nature neuroscience *17*, 773-781.

Tables and Figures

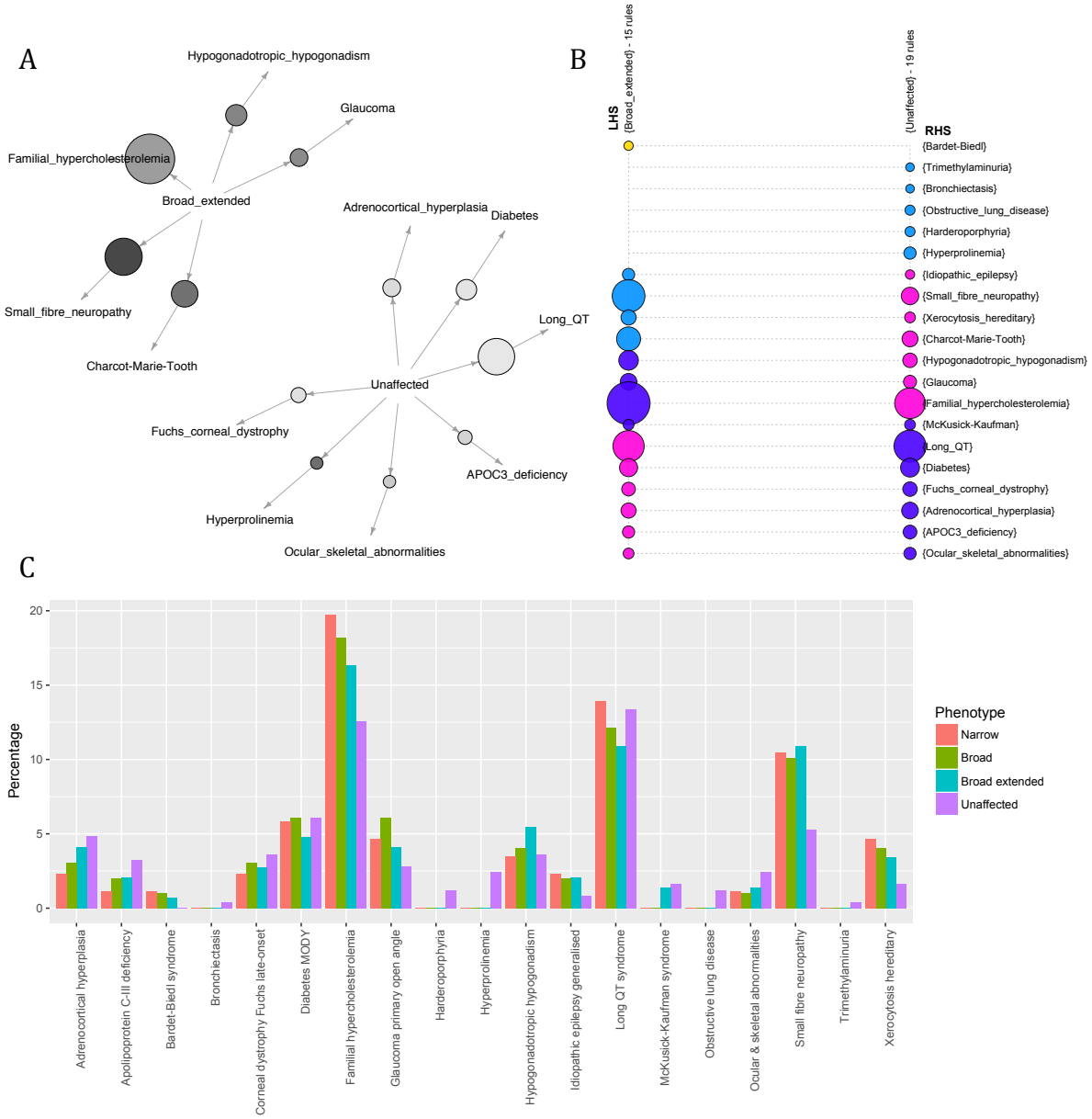

Figure 1: Profile of co-occurring Mendelian diseases varies between affected and unaffected individuals. A and B: Individuals genetically diagnosed with Mendelian diseases. Association rule discovery looks for combinations of variables that occur together more frequently than expected by chance. Diseases that co-occur more than expected differ between individuals affected with mood disorder and unaffected individuals. A: Graph for 12 rules and B: Grouped matrix for 34 rules with mood disorder affected status (either “broad extended” or “unaffected”) as the antecedent and Mendelian disease as the consequent. Size of the circle demonstrates confidence in that rule. Color indicates lift (A: darker=higher lift, B: blue=higher lift, pink=lower lift). C: Individuals carrying disease causing variants. Percentage of Narrow, Broad and Broad extended individuals (see “common polygenic risk for bipolar disorder” in results section for

explanation of phenotype) carrying disease causing variants compared to percentage of unaffected individuals with disease causing variants. Some variants are found in a higher percentage of affected individuals.

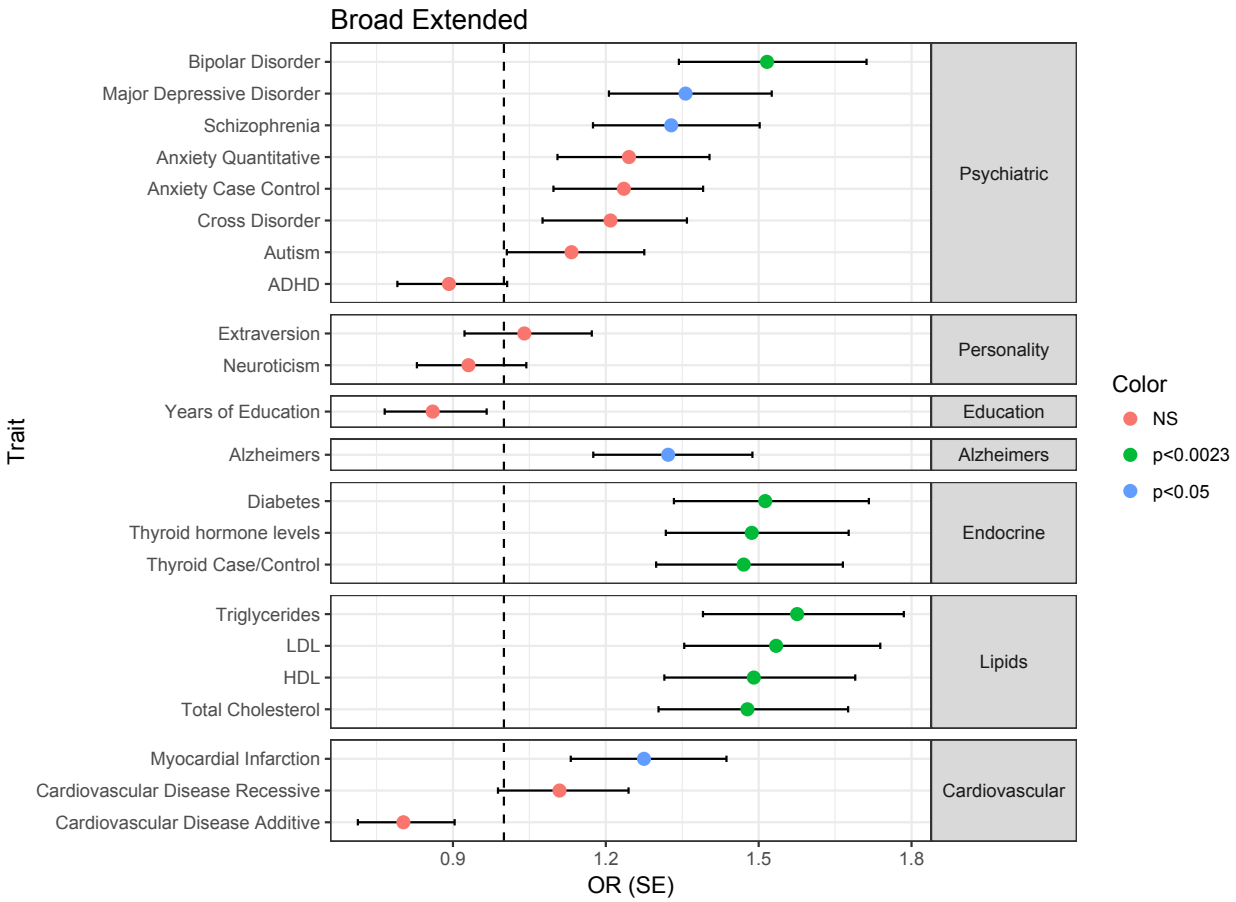

Figure 2: Association of polygenic risk scores for multiple traits with bipolar disorder (broad extended phenotype) in ASMAD. Risk scores for all traits tested are displayed, split by trait type. Circles indicate the association of the risk scores with mood disorder in the ASMAD pedigree, color of circles shows the level of significance, and error bars show +/- standard error.

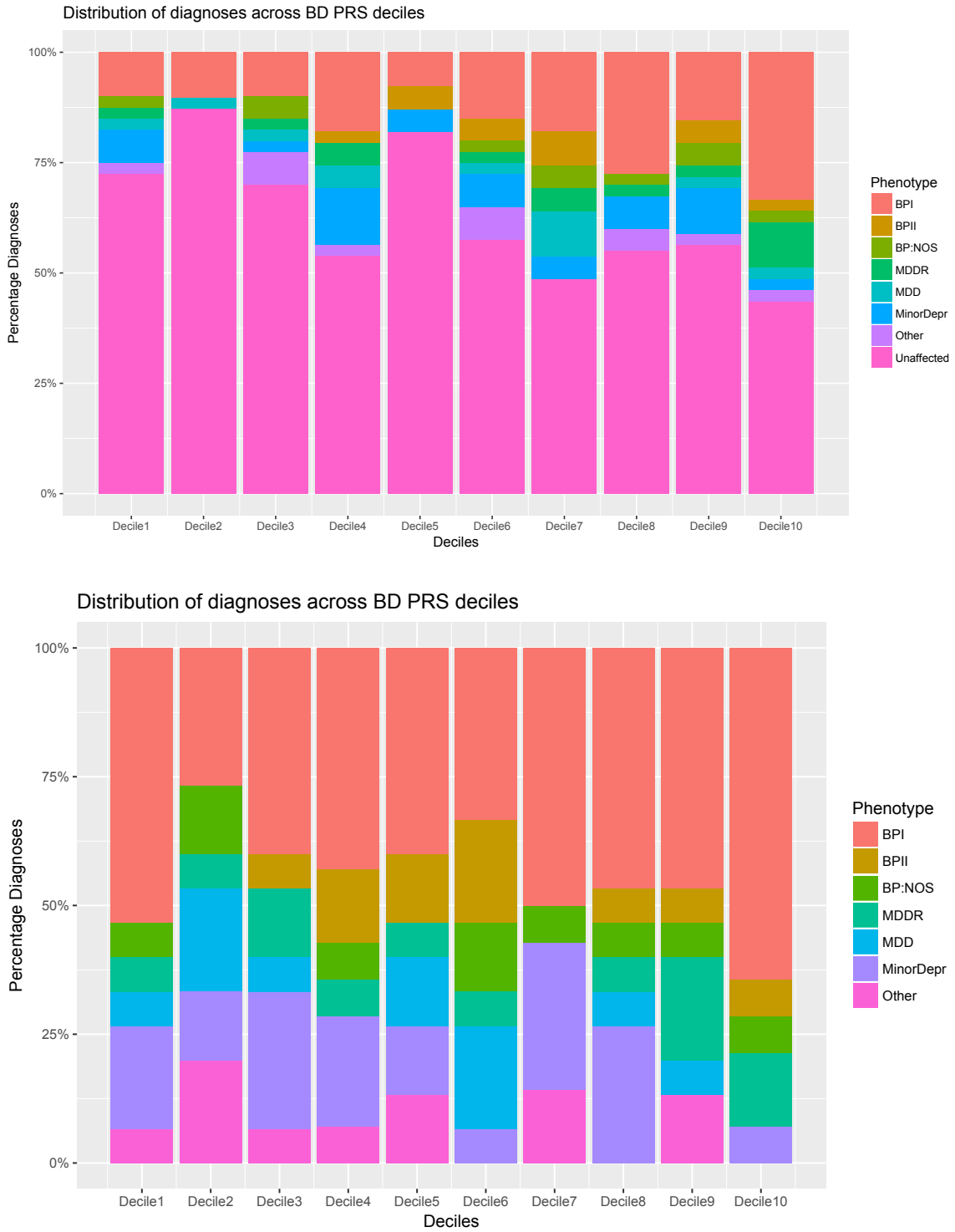

Figure 3: Percentage of individuals with affected status for each decile of polygenic risk score. A: All individuals, including unaffected individuals, are shown. B: Affected individuals only are shown.

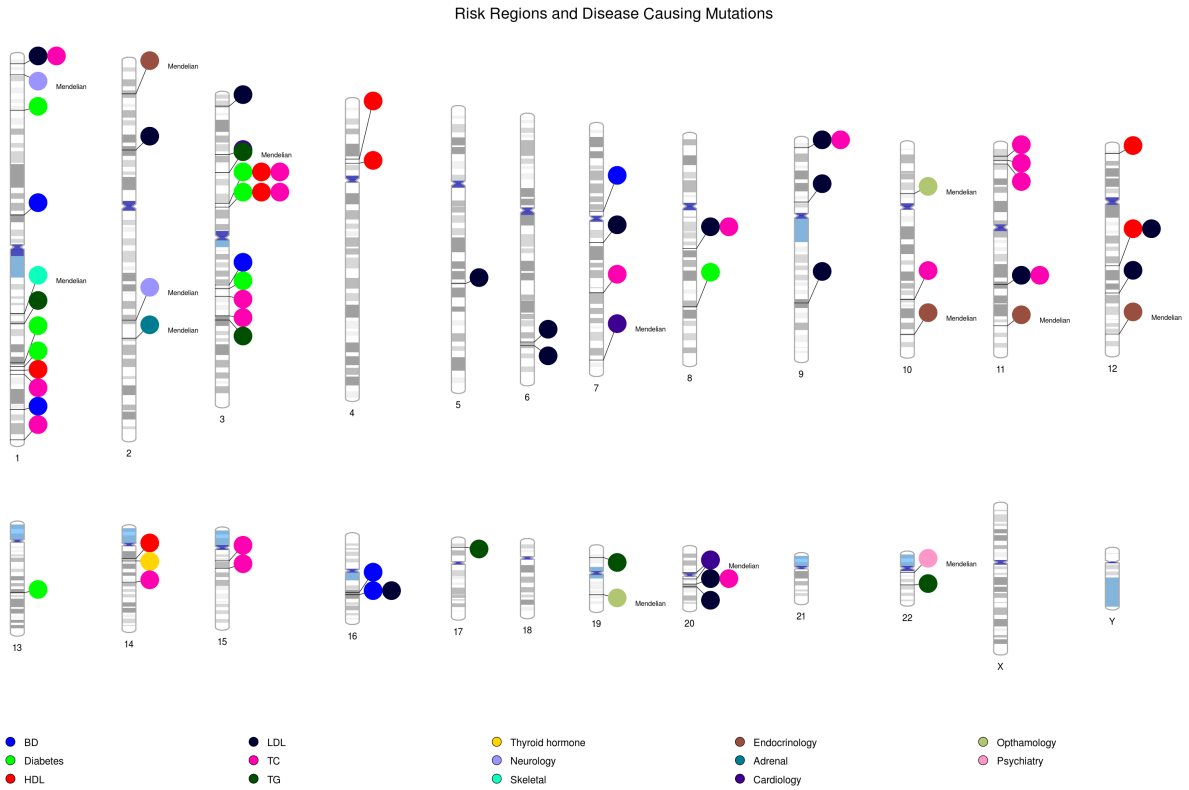

Figure 4: Locations of genomic regions that contribute to the genome-wide risk score differences between affected and unaffected individuals for bipolar disorder (blue), diabetes (green), HDL (red), LDL (black), Total cholesterol (pink), triglycerides (dark green) and thyroid hormone (yellow). The location of Mendelian variants which are associated with either affected or unaffected status in this pedigree are also included (denoted by the notation: Mendelian).

Table 1: Variants associated with lipid and cardiovascular phenotypes

Gene	Variant	rsID	Effect	% Unaffected	% BPI/BPII	% Other psychiatric phenotype
APOB	Arg3527Gln	rs5742904	Damaging lipid variant	12.6	21.3	10.3
ABCG8	Gly574Arg	rs137852988		5.7	11.3	12.3
HFE	Cys282Tyr	rs1800562	Iron overload	8.9	15.0	8.8
CYBRD1	Arg226His	rs62181680		1.2	3.8	4.4
LPA	c.4289+1G>A	rs41272114	Cardioprotective	5.3	3.8	1.5
APOC3	Arg19Ter	rs76353203		3.3	2.5	1.5
KCNH2	Gly803Arg	rs199473669	Damaging variant affecting heart rhythm	11.0	11.3	5.9
SCN5A	Glu462Lys	rs199473572		2.8	1.3	0
SNTA1	Ala257Gly	rs56157422		1.2	1.3	0