Human Behavioral Informatics in Genetic Studies of Neuropsychiatric Disease: Multivariate Profile-Based Analysis

Cinnamon S. Bloss[1], **Kelly M. Schiabor**[1], and **Nicholas J. Schork**[1,2]
[1]Scripps Genomic Medicine, Scripps Translational Science Institute, Scripps Health

[2]Department of Molecular and Experimental Medicine, The Scripps Research Institute

1. Introduction

The current research environment is virtually unprecedented in regards to both the number and sophistication of methods available to investigate the genetic determinants of disease. In particular, high-throughput DNA microarrays now enable scientists to conduct genome-wide association (GWA) studies; i.e., to essentially scan the entire human genome for single nucleotide polymorphisms (SNPs) that underlie disease risk. While GWA studies have yielded many important findings with respect to several human conditions (e.g., see Table 1) including obesity [1], diabetes [2,3], and glaucoma [4], among others [5–12], identification of the genetic determinants of neuropsychiatric disease via GWA studies has been more challenging [13,14]. For example, although GWA studies of both schizophrenia [15–17] and bipolar disorder [13,18–21] have revealed some notable results (Table 2), it has been difficult to detect signals associated with single genetic variations that meet criteria for genome-wide statistical significance. Furthermore, replication of findings has been difficult.

There are several reasons for the lack of consistently replicated, compelling findings in genetic studies of neuropsychiatric disorders. Some of these reasons include inadequate power to detect allelic effects of modest sizes, population-specific locus effects, epistatic interactions involving multiple genes each with modest independent effects, gene-environment interactions, effects of copy number variants or other forms of genetic variation not well captured by the panels of common SNPs that have been used, and the influence of multiple rare variations [18,22]. In addition to these reasons, however, an important and plausible explanation for the lack of consistency and compelling findings in genetic studies of neuropsychiatric disease relates to phenotypic heterogeneity and the notion that there is simply inherent imprecision associated with diagnostic categories in neuropsychiatric disease [23,24]. Indeed, traditional categorical diagnostic criteria for neuropsychiatric disorders as delineated by the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR) [25] essentially serve to categorize affected individuals based on highly complex clinical, behavioral, and neurocognitive profiles (see Figure 1 for DSM-IV-TR criteria for schizophrenia). Thus, when these simplified diagnostic categories are used in genetic association studies of neuropsychiatric disease, critical phenotypic information is unnecessarily lost. This is particularly problematic given that genes likely do not encode for

Address for Correspondence: Nicholas J. Schork, Ph.D., Scripps Genomic Medicine, 3344 North Torrey Pines Court, Suite 300, La Jolla, CA 92037, nschork@scripps.edu, (858) 554-5704 (phone), (858) 546-9284 (fax).

categorical 'diagnosis', but rather, as has been discussed widely in the neuropsychiatric literature, genes are likely associated with clusters of symptoms defined as characteristic or representative of a neuropsychiatric disorder (e.g., types of delusions and hallucinations and extent of cognitive impairment in schizophrenia, severity of depression and mania and level of functional impairment in bipolar disorder). Hence, the search for genes that mediate heterogeneous and imprecise clinical diagnostic categorizations may be ill-conceived and should be replaced – or at least extended – by genetic studies that leverage behavioral profile-based phenotypes (i.e., data obtained from a battery of clinical assessment measures that reflect the phenotypic complexity and heterogeneity observed in neuropsychiatric disease).

In this review we begin by outlining the GWA study paradigm, and discuss, in greater detail, possible explanations for why this approach has been less fruitful than anticipated with respect to neuropsychiatric diseases, including schizophrenia and bipolar disorder. We then propose a general approach to addressing the problem of phenotypic heterogeneity in genetic association studies of neuropsychiatric disease that draws on behavioral informatics concepts and involves the use of multivariate behavioral profile-based phenotypes as opposed to single-trait/case-control/diagnostic status-based approaches. Although there is a rich history in the use of multivariate methods in behavioral and behavior genetics research, including a journal devoted to 'Multivariate Behavior Research' (www.tandf.co.uk/journals/titles/00273171.asp), many of the multivariate methods that have been proposed are designed to test specific hypotheses or are based on data reduction methods such as cluster analysis or factor analysis and hence may not exploit all the variation present in a given data set. In addition to discussing the advantages of multivariate profiling, we discuss common behavioral assessment measures that are used in neuropsychiatry, as well as more novel telemetric monitoring devices and strategies that could also be used to assess behavior and construct profile-based phenotypes. We present four data analysis methods that can be used to implement this strategy and thus integrate multivariate behavioral profile data and high-dimensional genomic data. We further outline the advantages and disadvantages of each approach. We conclude that behavioral profile-based phenotypes may provide a meaningful alternative to the use of single measures, such as diagnostic category, in genetic association studies of neuropsychiatric disease.

2. Developing the GWA Study Paradigm

GWA studies allow researchers to statistically determine if variation in a phenotype can be explained by any of a large number of common genetic variations or differences within a population. Common genetic variations assessed in GWA studies are typically single nucleotide polymorphisms (or 'SNPs') and reflect a single base pair change in a DNA sequence that occurs at a frequency of at least 1% in a population [14]. The presence of a SNP at a particular site in the genome creates sequence variation as an individual can possess either two copies of the original base (known as 'homozygous wild type'), two copies of the SNP variant (known as 'homozygous derived allele'), or one copy of each ('heterozygous'). SNPs occur frequently throughout the human genome, and no fewer than 10 million SNPs are believed to populate a single human genome [26].

With such substantial SNP variation, unbiased genome-wide searches for variations associated with a disease phenotype could potentially entail a comparison of the frequency of all the SNPs present at over 10 million individual sites in the genome between diseased and non-diseased cohorts [14]. Testing >10 million variations, however, would be highly problematic due to enormous multiple comparisons issues. Fortunately, GWA studies involving fewer variations are feasible due to a structural characteristic of the genome known as linkage disequilibrium (LD) [27]. LD reflects the fact that SNPs are not

individually inherited; instead, blocks of the genome, often containing multiple SNPs, are inherited together as stretches of maternal or paternal chromosome comprising a gamete resulting from the recombination process that occurs during meiosis [14]. The more proximal variations are on a chromosome, the less likely they are to be separated via recombination over a large number of generations. By studying patterns of LD in a population, 'tag SNPs' can be selected which essentially function as proxies for other SNPs. A collection of adjacent variations that appear to be inherited together as a single unit are known as a 'haplotype.' Genome-wide SNP variation can thus be indirectly assessed through the direct genotyping of selected tag SNPs that capture all the variations in defined haplotypes [27].

The International HapMap Consortium (www.hapmap.org) was initiated to lead an effort to determine LD relationships in the genome, identify appropriate haplotypes, and facilitate the identification of tag SNPs for different populations [26]. The success of this consortium in identifying patterns of LD that might facilitate GWA studies also encouraged technological advances in genotyping that would essentially make GWA studies financially feasible. Thus, high-throughput microarray-based genotyping chips have been developed to allow GWA studies [27]. With both an understanding of the genetic structure of common SNP variation and economically viable technological platforms to assess or assay this genetic variation in large number of individuals, GWA studies have become feasible and have resulted in an unprecedented number of studies designed to interrogate the genome for the genetic determinants of common diseases, including neuropsychiatric disease.

2.1. GWA Studies in Practice

As researchers learn more about the etiology of disease through genetic investigations, it has become apparent that the underlying genetic architecture of many highly heritable diseases is more complex than originally thought. The high penetrance, single gene associations found to explain diseases such as Huntington's disease and, to an extent, cystic fibrosis, have proven to be the exception instead of the norm as explanations for the genetic basis of disease [14]. Indeed, as research progresses, the genetic research community is beginning to appreciate that it is more than likely that several genomic variations, environmental factors, gene-environment interactions, and epigenetic alterations all contribute to a disease phenotype. Furthermore, phenotypic heterogeneity can thwart efforts to identify genetic susceptibility variants via association studies. Ultimately, it appears that some common diseases are more amenable to successful study via GWA methods than others, but knowing, *a priori*, which might be amenable to such methods and which might not is difficult at best.

Obesity is an example of a complex disease that has been successfully studied via GWA studies. Twin and family studies in industrialized countries have consistently shown that 60% to 70% of the variation in obesity phenotypes is due to heritable genetic factors [28]. GWA studies investigating obesity have been successful in identifying contributing genetic variations. Two of the most robust genetic associations identified for the obesity phenotype, as measured through body mass index (BMI), are common SNP variants located near the FTO and MC4R genes [1,28,29]. Although, each of these associations makes a modest contribution to the overall phenotypic variation in BMI in the population at large, together these findings are a testament to the potential biological contribution significant associations can have on the overall understanding of the etiology of a disease. Of note, unlike neuropsychiatric diagnostic categories, BMI represents a continuous subclinical biological phenotype, which is likely very 'close' to (i.e., results directly from) the genetic variations that mediate susceptibility to obesity.

While GWA study of the obesity phenotype represents a successful effort at defining some genetic determinants of a common disease, not all GWA studies have produced similarly

definitive results. In fact, as noted previously, lack of replication and failure to detect genome-wide significant associations in GWA studies seem to disproportionately occur with the analysis of neuropsychiatric diseases (see Table 2) [13].

2.2. GWA Studies in Neuropsychiatric Disease

The first genetic variant to reach genome-wide significance in a GWA study of schizophrenia was the SNP rs41299148, located near the gene CSF2RA. CSF2RA has been classified as a cytokine-related gene and could potentially contribute to the cytokine-mediated neuronal inflammatory response. As such, variation in CSF2RA is seen as a plausible biologically-relevant candidate mechanism for the increased inflammatory response to environmental stressors that has been proposed to underlie the schizophrenia phenotype [15]. Unfortunately, this robust finding has not been replicated in subsequent schizophrenia GWA studies, and subsequent studies have, in general, failed to produce associations that exceed the genome-wide threshold of significance [17,30].

To cope with the difficulty in reaching genome-wide significance thresholds in associations of neuropsychiatric disease, SNPs that fall just below threshold levels of significance are often assessed in additional cohorts to determine the reliability and validity of the association. This replication-based trend seems to be employed often in GWA studies of neuropsychiatric diseases like schizophrenia. For example, in a study conducted by Shifman and colleagues, a common genetic variant in the RELN gene reached significance in women, but only when data from an initial cohort and four additional replication cohorts were jointly analyzed [16]. In another recent GWAS, the SNP rs13344706 in the vicinity of ZNF804A gene was found to be associated with the schizophrenia phenotype, but below the accepted level of genome-wide significance. To test the reliability and reproducibility of the association, the variation was examined in two independent cohorts. Most interestingly, the association reached genome-wide significance when the bipolar cohort from the Wellcome Trust Case Control Consortium (WTCCC) study was added to the schizophrenia cases. This increase in significance of the association could indicate that a common biological mechanism contributes to both the schizophrenia and bipolar phenotypes. This is further supported by the clinical observation that bipolar disorder and schizophrenia can share common psychotic features [24,31].

Similarly, despite producing independently significant results, there has been little convergence across the associations identified by the three GWA studies of bipolar disorder completed to date [20]. For example, Sklar and colleagues found strong associations between SNP variants in MYO5B, TSPAN8, and EGFR genes and the bipolar phenotype [18]. The significant associations from the primary analysis, however, were not replicated when analyzed in an independent cohort of family based trios. Further, the MY05B, TSPAN8, and EGFR localizations do not coincide with the association detected in the genome-wide association analysis of bipolar disorder conducted by the WTCCC [13]. The most notable result from the WTCCC bipolar GWA study was SNP rs420259 located at the locus 16p12. Several genes at this locus, including DCTN5, have compelling biological functions in the context of bipolar disorder. DCTN5 encodes a protein involved in intracellular transport known to interact with DISC1, a gene that has been previously linked with both bipolar disorder and schizophrenia [13]. To further complicate the interpretation of significant associations, a combined GWA study of the bipolar and control cohorts from the WTCCC study, the study by Sklar et al., and a new study, produced a new set of significant association signals. This collaborative effort indentified SNPs within both the ANK3 and CACNA1C genes as significantly associated with the bipolar phenotype [20], though neither association had reached genome-wide significance in the independent cohorts. Independent replication of these most recent genetic localizations has been

observed to some extent [21], though additional studies in large samples will be needed in order to confirm the reliability of the associations [32].

Overall, GWA studies in bipolar disorder lack strong findings and have shown minimal replication with respect to one another. Knowledge from GWA studies of schizophrenia is similarly ambiguous. Given these observations, we are left with determining the reasons for such a discrepancy in the number of robust genetic associations seen for diseases like obesity relative to neuropsychiatric diseases.

3. Genetic Associations in Neuropsychiatric Disease

As previously mentioned, there are likely several reasons for the lack of consistently replicated, compelling findings in genetic association studies of neuropsychiatric disorders. For instance, despite fairly large sample sizes, more statistical power may be required to obtain statistically significant results. It is possible that the effect sizes and minor allele frequencies of the genetic variants associated with neuropsychiatric disease are simply lower than those for other common, complex diseases. For most contemporary, case/control-based GWA studies, the effect sizes of individual SNPs are measured as odds ratios, which reflect the ratio of the probability of the occurrence of the disease phenotype among subjects with a certain genetic variant compared to the probability of the occurrence of the disease phenotype across subjects with the alternate allele [14]. If unidentified genetic variants associated with neuropsychiatric disease have low odds ratios, larger cohorts will be needed to identify and substantiate the small phenotypic effects associated with these variants. A similar expansion in sample size could be necessary if the minor allele frequency of a particular associated SNP variant is unusually low. This explanation is consistent with the research reports cited above in which combining or expanding the size of the case and control cohorts led to increasingly significant associations.

It is also possible that the majority of genetic susceptibility variants that underlie neuropsychiatric disease are not well-covered by current known SNP haplotypes. Other important sources of variation in the human genome that are not well-captured with traditional SNP assays include epistatic (gene-gene) interactions, gene-environment interactions, epigenetic controls, copy number variations, SNPs in regions of the genome with low linkage disequilibrium, and rare disease alleles [33]. The potential advantages of assessing alternate forms of genomic variation have been recently explored with respect to copy number variation and schizophrenia [34]. Developing new methods to assess different forms of genomic variation, including DNA sequencing strategies to identify novel, and possibly rare genetic variations, will enable a more comprehensive analysis of the genetic basis of neuropsychiatric disease.

3.1. Phenotypic Heterogeneity in Neuropsychiatric Disease

In addition to issues relating to statistical power and the nature of the genetic effects underlying neuropsychiatric disease, another important and plausible explanation for the lack of consistent findings in this area involves phenotypic heterogeneity and the notion that there is simply inherent imprecision associated with neuropsychiatric diagnostic categories [23,24]. As previously noted, traditional categorical diagnostic criteria for neuropsychiatric disorders as delineated by the DSM [25] essentially serve to place affected individuals into discrete, diagnostic categories despite the fact that individual members of a category typically possess highly complex and heterogeneous clinical, behavioral, and neurocognitive profiles. Thus, because most GWA studies of neuropsychiatric disease have defined the phenotype according to these simplified diagnostic categories in service of case/control study designs, critical phenotypic information is not leveraged, which likely results in decreased statistical power to detect an effect.

Ignoring phenotypic heterogeneity in genetic association studies of neuropsychiatric disease is particularly problematic given that genes are not likely to encode for categorical 'diagnosis', but rather are likely to be associated with specific clusters of neuropsychiatric symptoms (e.g., paranoid delusions), which can be present across multiple diagnostic categories (e.g., schizophrenia and bipolar disorder) [24]. In their review of genes implicated in schizophrenia and bipolar disorder, Craddock and colleagues note that there is increasing evidence from genetic association studies for an overlap in genetic susceptibility across the traditional classification categories. This suggests the possibility of relatively specific relationships between genotype and certain forms of psychopathology (i.e., rather than DSM-defined diagnostic categories). As an example, they noted evidence that variants in DISC1 and NRG1 may confer susceptibility to a form of neuropsychiatric illness with mixed features of schizophrenia and mania [24]. Thus, findings such as these suggest the need for alternative approaches to phenotype definition in genetic studies of neuropsychiatric disease rather than continued use of the traditional, and limited, DSM-based categories [35,36]. One such alternative may be the use of behavioral profile-based phenotypes (i.e., data obtained from a battery of clinical assessment measures or telemetric behavioral monitoring devices, which may reflect or capture the phenotypic complexity and heterogeneity observed within and across neuropsychiatric diseases). We discuss the use of such profile-based phenotypes below.

4. Behavioral Profile-Based Phenotypes in Genetic Studies of Neuropsychiatric Disease

One strategy to address the issue of phenotypic heterogeneity in genetic association studies of neuropsychiatric disease is to develop and exploit data analysis strategies that leverage all, or most, of the available high-dimensional clinical, diagnostic, and behavioral information for individuals under study. Such strategies draw on behavioral informatics concepts and could involve the construction and use of multivariate behavioral profile-based phenotypes. With respect to psychotic disorders such as schizophrenia and bipolar disorder, these profiles could encompass scores on a battery of clinical assessment measures, which could include psychiatric rating scales, personality measures, and/or neurocognitive tests. Further, such an approach could also leverage data from emerging telemetric monitoring devices (e.g., data collected via handheld devices, such as mobile phones and personal digital assistants), which can record cognitive and affective states [37]. There are a number of potential advantages to such an approach. For instance, the use of behavioral profiles eliminates the need for categorizing individuals with different 'subtypes' of a specific disease into one group. Further, behavioral profile-based phenotypes potentially provide a way to investigate genetic susceptibility among neuropsychiatric disorders that share similar clinical characteristics, like schizophrenia and bipolar disorder. Such an approach has been used successfully in the identification of the genetic determinants of complex behavioral phenotypes in model organisms such as mice; see, for example, the study of the genetic determinants of emotionality in mice, which is defined by a number of behavioral constructs any one of which would not necessarily capture the phenotype in isolation [38]. Finally, behavioral profiles are a direct, quantitative representation of the psychological and behavioral functioning of the individuals being studied, and as such, the use of these profiles may provide increased statistical power to detect genetic associations [39–41]. Notably, profiles across neurocognitive measures may be particularly salient in this regard given that neurocognition is considered an 'endophenotypic' domain underlying many neuropsychiatric conditions. Endophenotypes are quantitative, heritable, 'sub-clinical' phenotypes that are seen as reflecting underlying pathophysiologies associated with disease. Endophenotypes are also seen as being 'closer,' in terms of direct physiologic links, to

genetic variation than are the more remote clinical symptoms or diagnostic categories they are thought to influence [42].

5. Behavioral Assessment Measures for Profile-Based Phenotypes

A neuropsychiatric assessment, either for research or clinical purposes, typically involves evaluation of an individual's functioning across a wide range of domains, including emotional/psychiatric, personality, and neurocognitive domains, among others. Prior to discussion of analysis methods that can leverage the range of high-dimensional information produced via this type of assessment, as well as via more novel assessment methods such as telemetric monitoring [43], we will briefly discuss the common assessment measures and techniques used and the types of data generated.

5.1. Emotional Functioning and Level of Psychopathology

Although several psychometric instruments have been developed for the assessment of emotional functioning and psychopathology [44] we discuss a small subset of the most common measures. One widely used measure is the Symptom Checklist 90-R (SCL-90-R) [45] and its short form, the Brief Symptom Inventory (BSI) [46]. The SCL-90-R assesses a variety of symptoms as experienced over a one-week interval and consists of a series of 90 descriptions of symptoms that a respondent rates in terms of severity. The items, or symptoms, are scored around nine different dimensions, including somatization, obsessive-compulsive, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, and psychotocism as well as three global indices (i.e., the Global Severity Index, Positive Symptom Distress Index, and Positive Symptom Total). T-scores with a mean of 50 and a standard deviation of 10 are generated for each of the nine symptom dimensions, as well as the global indices, and can be compared with data available for four different normative groups (i.e., psychiatric outpatients, nonpatients, psychiatric inpatients, and nonpatient adolescents). Higher t-scores on the individual dimensions generally suggest elevated levels of psychological distress. In short, the available normative data, coupled with the wide diversity of validity studies on this measure, suggests that the SCL-90-R is appropriate for use with a variety of individuals and psychiatric issues.

In addition to general symptom assessment, it is common in neuropsychiatric assessment to specifically oversample both depressive and anxious symptomatology. With respect to depression, likely the most widely used assessment instruments are the Beck Depression Inventories [47]. The current version, the Beck Depression Inventory-II (BDI-II) [48] is highly congruent with DSM-IV-TR [25] diagnostic criteria for depressive disorders, and the items on the BDI were originally derived from observing the typical symptoms presented by depressed psychiatric patients [44]. On the BDI-II, a total of 21 items that relate to the following areas of difficulty are included: sadness, pessimism, past failure, loss of pleasure, guilty feelings, punishment feelings, self-dislike, self-criticalness, suicidal thoughts or wishes, crying, agitation, loss of interest, indecisiveness, worthlessness, loss of energy, changes in sleep pattern, irritability, changes in appetite, concentration difficulty, tiredness or fatigue, and loss of interest in sex. When completing the inventory, respondents are asked to rate the intensity of these symptoms on a scale from 0 to 3. The possible range of total scores extends from 0 to 63; scores from 0 to 13 indicate no or minimal depression, scores from 14 to 19 indicate mild depression, scores from 20 to 28 indicate moderate depression, and scores greater than 29 indicate severe depression. Similar to the BDI for the assessment of depression, the State-Trait Anxiety Inventory (STAI) [49] is one of the most widely used measures for the assessment of anxiety. The STAI is a 40-item, self-report inventory that has been shown to be sensitive to transitory episodes of anxiety (i.e., states), as well as more stable personality features that predispose an individual to experiencing more chronic levels of anxiety (i.e., traits). The state and trait dimensions of the inventory each have their own

individual items. Trait items were selected based on their having the highest correlations with other trait anxiety measures, as well as being the most stable over time; state items were selected based on their being sensitive to high versus low stress conditions and having the highest internal consistency [44]. Normative data are available for a range of healthy populations, as well as neuropsychiatric and general medical/surgical populations.

5.2. Personality Assessment

The two most commonly used personality assessment measures are the Minnesota Multiphasic Personality Inventory (MMPI) and the Millon Clinical Multiaxial Inventory (MCMI). The MMPI is a standardized questionnaire that elicits a wide range of self-descriptions that are scored to give a quantitative assessment and indicator of an individual's personality. Since its development in 1940, the MMPI has become the most widely used clinical personality inventory, with more than 10,000 published references [44]. The current version of the MMPI, the MMPI-2 [50], consists of 567 items, which are affirmative statements that can be answered 'True' or 'False.' The measure can be either hand or computer scored, and results are then summarized and compared with different normative samples. There are 10 clinical/personality scales, including scales thought to reflect levels of the following: hypochondriasis, depression, hysteria, psychopathic deviance, masculinity-femininity, paranoia, psychasthenia, schizophrenia, hypomania, and social introversion. There are also 7 validity scales that assess a person's attitude toward test-taking, as well as additional options (e.g., the 'content scales') for refining the meaning of the clinical scales and generally obtaining additional information. The content for the majority of individual MMPI questions is relatively obvious, and items deal largely with psychiatric, psychological, neurological, or physical symptoms. After a test profile is scored, interpretation involves taking into account the overall configuration, or profile, across the different scales, as well as relevant demographic characteristics of the respondent. In general, it is notable that the scales are believed to represent measures of personality traits rather than simply diagnostic categories. For example, elevated scores on the Depression scale may suggest characteristics such as mental apathy, self-deprecation, and tendency to worry, rather than simply the presence of an acute depressive disorder.

The Millon Clinical Multiaxial Inventory (MCMI) is a standardized, self-report questionnaire that also assesses a wide range of information related to an individual's personality, emotional adjustment, and test-taking attitude [44]. The current version, the MCMI-III [51,52], is composed of 175 items that represent 28 different scales, which are divided into the following categories: modifying indices, clinical personality patterns, severe personality pathology, clinical syndromes, and severe syndromes. The MCMI utilizes cutoffs related to Base Rate (BR) scores to designate the presence or absence of a particular characteristic. BR scores are derived from the percentage of the population that has been deemed to have a certain characteristic or syndrome. Millon arbitrarily set a BR score of 85 to indicate that the characteristic in question was definitely present; empirical evidence suggests the BR approach increases diagnostic accuracy when compared with the more frequently used t-score approach [53]. The scales, as well as the items that comprise the scales, are closely aligned to Millon's theory of personality and to the DSM-IV-TR [25]. Furthermore, many of the scales have both theoretical and item overlap, and as such, a respondent's profile, or scores across all of the scales, should be interpreted together. Like the MMPI, the MCMI is also not designed to provide a diagnosis; rather, it provides considerable information relevant to diagnosis, which must be taken together and integrated with other clinical information in order to put potential neuropsychiatric pathology a patient might exhibit into context.

5.3. Neurocognitive Functioning

An important aspect of any neuropsychiatric assessment is to get some sense of the individual's brain and cognitive functioning. This can be particularly important for psychotic disorders such as schizophrenia and bipolar disorder given the known cognitive impairments (e.g., in domains such as frontal/executive functioning) that are often observed. Furthermore, neurocognitive variables in genetic studies of neuropsychiatric disease are thought to be particularly useful given that neurocognition is considered an endophenotype for psychotic disorders [42]. Domains of functioning considered important for neurocognitive assessment include current general cognitive functioning, 'premorbid' functioning, intelligence, language/verbal skills, attention/processing speed, learning and memory, visuospatial skills, higher level 'executive' skills, and motor functioning. Because in-depth discussion of tests used to assess each of these domains is beyond the scope of this article, a subset of measures that are thought to tap the domains of higher level executive skills are reviewed given that deficits in this domain are the most common neurocognitive deficits observed in both schizophrenia and bipolar disorder [54].

The domain of executive functioning is thought to represent several components, including abstraction, reasoning, concept formation, mental flexibility, planning, set-shifting, sequencing, and working memory. Although lesions in frontal regions of the brain are often implicated when deficits in executive functioning are observed, these skills are also sensitive to lesions in other anatomic regions as well. Furthermore, over the past several decades, a number of neuropsychological tests have been developed to assess this varied domain. One such test that is widely used for the assessment of concept formation and set-shifting, the Wisconsin Card Sorting Test (WCST) [55,56], as well as a set of measures that was recently developed to assess the range of executive functions, the Delis-Kaplan Executive Function System (D-KEFS) [57], will both be briefly reviewed here.

The WCST has undergone a number of revisions since it was first developed in 1948. In the current and most widely used version [58], the examinee is presented with four 'key' cards each with a different set of shapes on them in different colors and numbers (i.e., one red triangle, two green stars, three yellow crosses, and four blue circles). The respondent is then given a pack of either 64 or 128 cards that they are to 'match,' one by one in a certain order, to one of the four key cards. The examiner does not provide any information about how to match the cards, but rather, tells the respondent each time whether the match was right or wrong. From the pattern of feedback the respondent is getting from the examiner, the respondent must deduce the appropriate placement of the cards. After a run of 10 correct placements in a row, the examiner shifts the sorting principle, indicating the shift only via the changed pattern of 'right' and 'wrong' feedback to the respondent. The WCST scoring scheme provides a score that represents the number of correct categories achieved and the number of perseverative errors. Categories achieved refers to the number of correct runs of 10 sorts, which can range from 0, if the respondent never learns the task, up to 6, at which point the test is normally discontinued. Perseverative errors occur when the subject continues to sort according to a previously successful principle, which is a useful measure of impaired concept formation, inability to incorporate feedback into one's problem solving strategies, and conceptual inflexibility. Performance profiles on the WCST are known to be highly useful for detecting brain damage due to epilepsy [59], stroke [60], traumatic brain injury [61], multiple sclerosis [62], alcoholism [63], Parkinson's disease [64], depression [65], and schizophrenia [66], among other etiologies, although different patient groups show different patterns of performance.

The D-KEFS is a set of nine tests selected to be sensitive to the many types of executive impairment seen in patients with brain disease, and both verbal and nonverbal tests are included [57]. Each of the nine tests, including the Trailmaking Test, Verbal Fluency,

Design Fluency, Color-Word Interference, Sorting Test (i.e., with similarities to the WCST, described above), Twenty Questions Test, Tower Test, Proverb Test, and Word Context Test are intended to stand alone. Although for the most part, these tests are variations on several existing, commonly used tests for assessment of executive functions, the major advantage of the D-KEFS collection is that all nine measures are co-normed on 1,750 participants ranging in age from 8 to 89. Furthermore, many of the standard tests have been lengthened in the D-KEFS collection, with additional easy and difficult items to avoid ceiling and floor effects. Additional subtests have also been added that break down performances into the fundamental components required for success on these complex tasks. As such, additional scores for these subtests, as well as additional scores that represent more process oriented measures, are available and easily calculated to obtain a more detailed profile of both an individual's level of functioning as well as the profile of performance. The fact that the D-KEFS does not offer a composite score that represents a person's performance across all nine tests is yet another striking example of a highly useful assessment instrument that is best interpreted based on the respondent's profile of scores across all of the measures.

5.4. Telemetric Monitoring

In addition to traditional measures used in the context of a standard neuropsychiatric assessment (many of which are described above), more novel assessment methods are also being developed, including an emerging class of remote data collection and information technologies called 'telemetrics'. Telemetric devices, such as handheld instruments, wireless monitors, actigraphs, etc. can be used to gather data on individuals repeatedly over extended periods of time and in a wide range of settings [43].

In their review of this novel and expanding research area, Goodwin and colleagues note that telemetrics generally include what have been dubbed 'passive' and 'active' wireless technologies, which collect and transmit data remotely. Passive telemetrics include wearable and ubiquitous computers that record behavioral, physiological, and environmental data from sensors worn on the body or embedded in the environment. Active telemetrics include handheld devices such as mobile phones and personal digital assistants (PDAs), which allow a user to record and receive cognitive and affective information. Aspects of these technologies that have been and are currently being piloted include studies using telemetrics to measure stress in autism, physiological responses in anxiety disorders, and treatment adherence and compliance monitoring in obstructive sleep apnea [67]. Further, Goodwin and colleagues note many potential applications of these technologies to neuropsychiatric and neurological disease. As examples, they discuss the possibility of using these devices to characterize and identify epileptic seizures through accelerometry to develop an ambulatory monitor with a real-time seizure classifier; to monitor the movement states of Parkinson's patients to establish a timeline of symptom severity and motor complications; and finally to develop physiological and behavioral measures to classify emotional states associated with preclinical symptoms of psychosis, mood, anxiety, and personality disorders [68].

While telemetrics will surely provide more statistical power to address a range of questions relating to behavior, the data produced will be a new class of multivariate data, which will include many more 'phenotypic' data points than are typically leveraged and accommodated in genetic studies of neuropsychiatric disease. With respect to these emerging approaches to data collection and behavioral assessment, data analytic and statistical methods that will accommodate and leverage the high-dimensional profile-based data that they produce will be essential.

Our brief review of both traditional and novel human behavioral assessment strategies for neuropsychiatric disease suggests that there is a wide, possibly bewildering array of phenotyping methodologies that go well beyond DSM-based categorical diagnoses. Our

conclusions in this regard are consistent with other previous discussions in the literature centered on the need for convergent approaches to assessing and accounting for heterogeneity across the many composite domains that make up psychiatric diagnoses [69,70]. Unfortunately, genetic association studies of neuropsychiatric disease have rarely leveraged the range of clinical, diagnostic, and behavioral information gleaned from either traditional or novel assessment approaches. Instead, association studies have typically relied on the simple 'case-control' paradigm whereby a phenotype of interest is defined as the presence or absence of disease based on DSM criteria [25]. Notably, this current state of the field is not surprising given that geneticists rarely possess detailed knowledge of neuropsychiatric phenotypes, and clinical researchers do not often have expertise in the conduct of high-dimensional genetic association studies. It is not, however, an optimal design for genetic association studies of neuropsychiatric disease, which we propose would glean many benefits from the use of behavioral profile-based phenotypic definitions (i.e., data obtained from a battery of clinical assessment measures or novel assessment technologies, which reflect the phenotypic complexity and heterogeneity that is actually observed).

5. Analysis of Combined Behavioral Profile and Genetic Data

Neurobehavioral assessments potentially provide an enormous amount of data for each individual in a study, as noted above. Use of the full range of these data in genetic association studies, however, raises some issues about the best way to integrate high-dimensional behavioral data with high-dimensional genetic data. Prominent among these issues are the risk of false positives due to multiple comparisons if multiple univariate analyses are applied to each data point (e.g., each item or subscale score of a particular assessment instrument, actigraph/global positioning system (GPS) data collected over some period of time). Another issue is the potential loss of information that may result if standard multivariate strategies that rely on data reduction approaches (e.g., factor and cluster analysis) are used. Thus, analytical approaches that leverage behavioral profile data and that account for issues related to multiple comparisons and data reduction are needed.

While there are a number of analytical approaches one could pursue in relating a large number of genetic variations to high-dimensional behavioral data, there are some very logical distinctions between many of these approaches. For example, basic approaches can be group into four basic classes: 1. multiple univariate analysis; 2. data reduction approaches; 3. regression and regression-like approaches; 4. and multivariate profile similarity analysis. Of these approaches, relatively few are designed to interrogate the simultaneous relationship between a large number of 'independent' variables (e.g., genetic variations) and clinically meaningful 'dependent' or outcome measures (e.g., neurobehavioral measures), although some good reviews of the more widely-used approaches exist [71,72]. Table 3 lists data analysis techniques that assess the relationship between two sets of variables, each possessing certain advantages and disadvantages. We discuss the four basic approaches to data analysis below, with special reference to those that are truly multivariate in orientation and as such, consider the relationships between sets of variables.

5.1. Multiple Univariate Approaches

Currently the most widely used, and certainly the most simplistic approach for conducting association of high-dimensional genetic data with any type of phenotype, involves relating each genetic variation to each phenotype in focused, independent analyses. Indeed, all GWA studies performed to date have used this approach [14]. For instance, if one is seeking to test the association of 500,000 SNPs with a quantitative phenotype such as intelligence or 'IQ', then one could perform 500,000 analysis-of-variance (ANOVA) tests, which would contrast

IQ scores across the three genotypic categories for each SNP. One could then pursue this analysis for each item, subscale, or assessment measure within each behavioral or neurocognitive domain assessed. The advantages of this approach are that the results of each test are intuitive and every element of the data is statistically assessed, allowing an investigator to pinpoint the exact nature of the relationship between the genetic variant and behavioral phenotypes. Clearly, however, this approach is problematic due to multiple comparisons. To address this issue, one could leverage other statistical strategies, such as false discovery rate (FDR) analysis, to determine how many analyses might have produced results worth considering further in light of the number of tests that were performed [73,74]. Another disadvantage of the multiple univariate analysis approach is that it does not, by itself, provide more 'holistic' insights into how the individual data points may cohere, although methodologies for grouping the results of multiple independent statistical analyses have been devised [75–77].

5.2. Data Reduction Approaches

Another approach to the analysis of high-dimensional data sets involves reducing the data points into a smaller set of homogenous data points that essentially capture the information or variation in the data set as a whole. These reduced data points can then be examined for patterns and associations with other data points. Well-known methods that fall into this class include factor analysis and principal components analysis (PCA) [78,79]. The advantages of data reduction approaches are that they can provide ways of empirically organizing high-dimensional data sets, and they can illuminate patterns that may not have been discernable from the results of multiple univariate analyses. A disadvantage of these approaches for combined behavioral profile and genetic data is that one would have to apply the data reduction technique to the behavioral profile data and genetic data independently and then assess the extent to which the reduced data points from each are associated with one another. In addition, the probabilistic and biological meaningfulness of the number and 'content' of the reduced data points is often in doubt. Thus, the results from such an analysis could be difficult to interpret from a clinical or biological standpoint, especially in the context of genetic variation data. An alternative, and perhaps better strategy, might be to reduce the behavioral profile data and then test for association between the reduced data points (e.g., factors, principal components) and each of the genetic variations or groups of variations in the form of haplotypes. Interpretability of factors derived from high-dimensional behavioral data may still be difficult, however, and an additional problem with many of these approaches is their computational demand.

5.3. Regression-Based Approaches

An additional approach to the analysis of high-dimensional data are methods rooted in multivariate regression and regression-like methodologies in which, e.g., the genetic variations provided are treated as independent predictors of the behavioral profile data, which are treated as a set of dependent variables. Partial least squares and canonical correlation analyses are approaches similar to multivariate regression approaches [41,80–82]. Advantages of these approaches are that they allow one to assess the effect of each data element simultaneously with the others, which could lead to insights about which variables have effects that are essentially independent of the others. Also, these approaches allow one to consider interactions between independent variables. Disadvantages include potential problems with interpretability, computational efficiency, and difficulty identifying redundancies among the variables used. Ways to address this might include leveraging a moving window approach whereby sets of adjacent genetic variations are tested for association with sets of behavioral phenotypes.

5.4. Multivariate Profile Similarity Analysis

The last approach described here for the analysis of high-dimensional data has applicability to combined behavioral profile and genetic data. Importantly, variants of this phenotypic profiling approach have been implemented for genetic analysis of psychotic disorders [36]. One variant of this approach, termed 'multivariate profile similarity analysis,' involves treating behavioral data collected on an individual as a multivariate profile for that individual; profiles for each individual can then be assessed to determine the degree of similarity of each individual to all of the other individuals in the study. Factors that explain or influence the greater or lesser similarity observed between profiles of the individuals in the study can then be tested. Regression-like models relating various factors (e.g., genetic variations) to profile similarity can be fashioned in what has been termed 'Multivariate Distance Matrix Regression (MDMR)' [83–85].

A detailed illustration of the MDMR method was presented by Zapala and Schork (2006) and involved analysis of gene expression patterns in human frontal cortex among individuals who died at various ages [85]. Initially, these authors performed correlational analyses to identify 463 genes that correlated with age, and then calculated a Pearson correlation-based heat map matrix that covered all pair-wise comparisons of individuals. They then analyzed the distance matrix based on the pair-wise correlations using the proposed MDMR-based method. Within the context of this method, the values in the distance matrix served as the dependent variables and age and sex served as independent variables. Although sex was not a significant predictor of the gene expression patterns in frontal cortex, as would be expected, age was significantly associated with gene expression profile similarity. Specifically, among individuals in the study, age appeared to explain approximately 35% of the variance in gene expression profile similarity of the genes (previously selected for their age-relatedness). Thus, through the use of this methodology, similarity/dissimilarity in patterns of variables of interest were assessed for association with a set of predictor variables. Advantages of analysis approaches rooted in multivariate profile similarity are that they can be used with large amounts of data collected on relatively few subjects; they also provide a method for examining the data as a whole. Disadvantages of the approach are that by treating all the data points equally, the 'signal' provided by a few data points may be confounded by the 'noise' given off from the others. In behavioral profile contexts, however, this may be overcome by confining the construction of the profiles to behavioral phenotypes thought to be of most relevance for the diseases being studied.

5.5. Power Advantages of Multivariate Analysis Methods

Although multivariate analysis methods leverage all the data (or large amounts of the data) at hand in a single analysis, they do not always provide power advantages to detect, e.g., genetic associations. The reasons for this have to do with the fact that some variables may be highly redundant in terms of the information they provide for assessing an association. Consider a situation in which an investigator plans to leverage two variables for an association study investigating genetic variations that influence height: height measured in inches and height measured in centimeters. Obviously, these two variables are highly correlated and essentially redundant. Therefore, there will not be added statistical power from studying them in a multivariate analysis as opposed to a simple univariate analysis of either in isolation. The manner in which variables provide new or complementary information in a way that increases power in multivariate analyses has been studied [40,41,86]. To simplify, consider two variables being tested for association with variation at a locus with two alleles, A and T, and hence three genotypes, AA, AT, and TT. The two variables will exhibit some degree of 'overall' correlation which is independent of genotype category. In addition, the variables will exhibit correlations within each genotype category (termed the 'residual' correlation). Multivariate techniques achieve considerable increases in

power to detect associations if the overall correlation between the variables is opposite in sign to the residual correlations. Figure 2 provides a graphical display of this phenomenon, in which the left panel depicts two variables that have an overall positive correlation (indicated by the long black line) and positive residual correlations (indicated by the shorter lines for individuals with the same genotype). The right panel of Figure 2 depicts a situation in which the overall correlation between two variables is positive and the residual correlation, within genotype, is negative. Variables that have overall correlations that are opposite to residual correlations provide a setting in which the variables provide unique and non-redundant information about their relationship to genotype categories, and hence, an increase in statistical power.

6. Conclusions and Future Directions

There is currently an unprecedented array of sophisticated technologies one can leverage to investigate the genetic determinants of neuropsychiatric disease. High-throughput DNA microarrays for genotyping are prominent among these and have paved the way for a number of GWA studies of complex diseases. However, GWA studies of neuropsychiatric disease, relative to other diseases, have often failed to produce consistent, replicated, and compelling findings. As previously mentioned, there are likely several reasons for this, including low statistical power and the possibility of smaller effect sizes associated with neuropsychiatric disease susceptibility variants, as well as the possibility that variants underlying neuropsychiatric disease susceptibility are not well-covered by current known SNP haplotypes. Another reason, however, and one that has been widely discussed in the literature, is that there is simply greater heterogeneity in neuropsychiatric disease that is not well accounted for in traditional case-control analyses. This case-control analysis approach, which is standard in genetic association studies of neuropsychiatric disease, essentially involves placing affected individuals into discrete diagnostic categories despite the fact that individual members of a category typically possess highly complex and heterogeneous clinical, behavioral, and neurocognitive profiles.

In this review, we have argued for an alternative approach to the simplistic case-control phenotypic definition, in which phenotypes are defined based on clinical or behavioral profiles (i.e., data obtained from a battery of clinical assessment measures or more novel assessment approaches such as telemetric monitoring, which reflect the phenotypic complexity and heterogeneity observed in neuropsychiatric disease). Such an approach would potentially eliminate the need for categorizing individuals with different 'subtypes' of a specific disease into one group and would provide a way to investigate genetic susceptibility among neuropsychiatric disorders that share similar clinical characteristics (e.g., schizophrenia and bipolar disorder). Behavioral profiles are a direct, quantitative representation of the psychological functioning of the individuals being studied, and as such, the use of these profiles may provide increased statistical power to detect associations. As we have discussed here, however, taking a behavioral profile approach to phenotypic definition requires appropriate data analysis strategies that can accommodate multiple high-dimensional data types. As such, we have offered some general strategies for this, and have focused on one approach in particular termed 'Multivariate Distance Matrix Regression' (MDMR).

In short, important information may be lost when behavioral profiles are reduced to a single score or diagnostic category, and thus, we encourage the use of clinical or behavioral profiles in defining phenotypes for genetic association studies of neuropsychiatric disease. More broadly, we offer that such profile-based approaches are highly consistent with, and would be suitable for, analysis of data from emerging novel phenotypic assessment technologies, such as wireless telemetric monitoring. Similarly, high-dimensional data from

model systems, including studies emerging in the areas of sociogenomics and polyphenism (i.e., defined as the occurrence of several distinct phenotypes or forms in a given species) [87] would also be amenable to such strategies, whether or not genetic data is involved. Indeed, the existence of journals devoted to behavioral profiling (e.g., *Multivariate Behavioral Research*) as well as technological developments applied to profile-based behavioral assessment (e.g., *Behavior Research Methods*), suggests the importance of using behavioral profile-based phenotypes and the development of appropriate statistical analysis methods for doing so.

Acknowledgments

References

1. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JR, Elliott KS, Lango H, Rayner NW, Shields B, Harries LW, Barrett JC, Ellard S, Groves CJ, Knight B, Patch AM, Ness AR, Ebrahim S, Lawlor DA, Ring SM, Ben-Shlomo Y, Jarvelin MR, Sovio U, Bennett AJ, Melzer D, Ferrucci L, Loos RJ, Barroso I, Wareham NJ, Karpe F, Owen KR, Cardon LR, Walker M, Hitman GA, Palmer CN, Doney AS, Morris AD, Smith GD, Hattersley AT, McCarthy MI. A common variant in the fto gene is associated with body mass index and predisposes to childhood and adult obesity. Science (New York, NY. 2007; 316:889–894.

2. Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V, Bailey R, Nejentsev S, Field SF, Payne F, Lowe CE, Szeszko JS, Hafler JP, Zeitels L, Yang JH, Vella A, Nutland S, Stevens HE, Schuilenburg H, Coleman G, Maisuria M, Meadows W, Smink LJ, Healy B, Burren OS, Lam AA, Ovington NR, Allen J, Adlem E, Leung HT, Wallace C, Howson JM, Guja C, Ionescu-Tirgoviste C, Simmonds MJ, Heward JM, Gough SC, Dunger DB, Wicker LS, Clayton DG. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nature genetics. 2007; 39:857–864. [PubMed: 17554260]

3. Hakonarson H, Grant SF, Bradfield JP, Marchand L, Kim CE, Glessner JT, Grabs R, Casalunovo T, Taback SP, Frackelton EC, Lawson ML, Robinson LJ, Skraban R, Lu Y, Chiavacci RM, Stanley CA, Kirsch SE, Rappaport EF, Orange JS, Monos DS, Devoto M, Qu HQ, Polychronakos C. A genome-wide association study identifies kiaa0350 as a type 1 diabetes gene. Nature. 2007; 448:591–594. [PubMed: 17632545]

4. Thorleifsson G, Magnusson KP, Sulem P, Walters GB, Gudbjartsson DF, Stefansson H, Jonsson T, Jonasdottir A, Jonasdottir A, Stefansdottir G, Masson G, Hardarson GA, Petursson H, Arnarsson A, Motallebipour M, Wallerman O, Wadelius C, Gulcher JR, Thorsteinsdottir U, Kong A, Jonasson F, Stefansson K. Common sequence variants in the loxl1 gene confer susceptibility to exfoliation glaucoma. Science. 2007; 317:1397–1400. [PubMed: 17690259]

5. Eeles RA, Kote-Jarai Z, Giles GG, Olama AA, Guy M, Jugurnauth SK, Mulholland S, Leongamornlert DA, Edwards SM, Morrison J, Field HI, Southey MC, Severi G, Donovan JL, Hamdy FC, Dearnaley DP, Muir KR, Smith C, Bagnato M, Ardern-Jones AT, Hall AL, O'Brien LT, Gehr-Swain BN, Wilkinson RA, Cox A, Lewis S, Brown PM, Jhavar SG, Tymrakiewicz M, Lophatananon A, Bryant SL, Horwich A, Huddart RA, Khoo VS, Parker CC, Woodhouse CJ, Thompson A, Christmas T, Ogden C, Fisher C, Jamieson C, Cooper CS, English DR, Hopper JL, Neal DE, Easton DF. Multiple newly identified loci associated with prostate cancer susceptibility. Nature genetics. 2008; 40:316–321. [PubMed: 18264097]

6. Harley JB, Alarcon-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL, Tsao BP, Vyse TJ, Langefeld CD, Nath SK, Guthridge JM, Cobb BL, Mirel DB, Marion MC, Williams AH, Divers J, Wang W, Frank SG, Namjou B, Gabriel SB, Lee AT, Gregersen PK, Behrens TW, Taylor KE, Fernando M, Zidovetzki R, Gaffney PM, Edberg JC, Rioux JD, Ojwang JO, James JA, Merrill

JT, Gilkeson GS, Seldin MF, Yin H, Baechler EC, Li QZ, Wakeland EK, Bruner GR, Kaufman KM, Kelly JA. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in itgam, pxk, kiaa1542 and other loci. Nature genetics. 2008; 40:204–210. [PubMed: 18204446]

7. Hunt KA, Zhernakova A, Turner G, Heap GA, Franke L, Bruinenberg M, Romanos J, Dinesen LC, Ryan AW, Panesar D, Gwilliam R, Takeuchi F, McLaren WM, Holmes GK, Howdle PD, Walters JR, Sanders DS, Playford RJ, Trynka G, Mulder CJ, Mearin ML, Verbeek WH, Trimble V, Stevens FM, O'Morain C, Kennedy NP, Kelleher D, Pennington DJ, Strachan DP, McArdle WL, Mein CA, Wapenaar MC, Deloukas P, McGinnis R, McManus R, Wijmenga C, van Heel DA. Newly identified genetic risk variants for celiac disease related to the immune response. Nature genetics. 2008; 40:395–402. [PubMed: 18311140]

8. Ridker PM, Pare G, Parker A, Zee RY, Danik JS, Buring JE, Kwiatkowski D, Cook NR, Miletich JP, Chasman DI. Loci related to metabolic-syndrome pathways including lepr, hnf1a, il6r, and gckr associate with plasma c-reactive protein: The women's genome health study. Am J Hum Genet. 2008; 82:1185–1192. [PubMed: 18439548]

9. Tenesa A, Farrington SM, Prendergast JG, Porteous ME, Walker M, Haq N, Barnetson RA, Theodoratou E, Cetnarskyj R, Cartwright N, Semple C, Clark AJ, Reid FJ, Smith LA, Kavoussanakis K, Koessler T, Pharoah PD, Buch S, Schafmayer C, Tepel J, Schreiber S, Volzke H, Schmidt CO, Hampe J, Chang-Claude J, Hoffmeister M, Brenner H, Wilkening S, Canzian F, Capella G, Moreno V, Deary IJ, Starr JM, Tomlinson IP, Kemp Z, Howarth K, Carvajal-Carmona L, Webb E, Broderick P, Vijayakrishnan J, Houlston RS, Rennert G, Ballinger D, Rozek L, Gruber SB, Matsuda K, Kidokoro T, Nakamura Y, Zanke BW, Greenwood CM, Rangrej J, Kustra R, Montpetit A, Hudson TJ, Gallinger S, Campbell H, Dunlop MG. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. Nat Genet. 2008; 40:631–637. [PubMed: 18372901]

10. Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP, Manolescu A, Thorleifsson G, Stefansson H, Ingason A, Stacey SN, Bergthorsson JT, Thorlacius S, Gudmundsson J, Jonsson T, Jakobsdottir M, Saemundsdottir J, Olafsdottir O, Gudmundsson LJ, Bjornsdottir G, Kristjansson K, Skuladottir H, Isaksson HJ, Gudbjartsson T, Jones GT, Mueller T, Gottsater A, Flex A, Aben KK, de Vegt F, Mulders PF, Isla D, Vidal MJ, Asin L, Saez B, Murillo L, Blondal T, Kolbeinsson H, Stefansson JG, Hansdottir I, Runarsdottir V, Pola R, Lindblad B, van Rij AM, Dieplinger B, Haltmayer M, Mayordomo JI, Kiemeney LA, Matthiasson SE, Oskarsson H, Tyrfingsson T, Gudbjartsson DF, Gulcher JR, Jonsson S, Thorsteinsdottir U, Kong A, Stefansson K. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature. 2008; 452:638–642. [PubMed: 18385739]

11. Uda M, Galanello R, Sanna S, Lettre G, Sankaran VG, Chen W, Usala G, Busonero F, Maschio A, Albai G, Piras MG, Sestu N, Lai S, Dei M, Mulas A, Crisponi L, Naitza S, Asunis I, Deiana M, Nagaraja R, Perseu L, Satta S, Cipollina MD, Sollaino C, Moi P, Hirschhorn JN, Orkin SH, Abecasis GR, Schlessinger D, Cao A. Genome-wide association study shows bcl11a associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105:1620–1625. [PubMed: 18245381]

12. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, de Bakker PI, Abecasis GR, Almgren P, Andersen G, Ardlie K, Bostrom KB, Bergman RN, Bonnycastle LL, Borch-Johnsen K, Burtt NP, Chen H, Chines PS, Daly MJ, Deodhar P, Ding CJ, Doney AS, Duren WL, Elliott KS, Erdos MR, Frayling TM, Freathy RM, Gianniny L, Grallert H, Grarup N, Groves CJ, Guiducci C, Hansen T, Herder C, Hitman GA, Hughes TE, Isomaa B, Jackson AU, Jorgensen T, Kong A, Kubalanza K, Kuruvilla FG, Kuusisto J, Langenberg C, Lango H, Lauritzen T, Li Y, Lindgren CM, Lyssenko V, Marvelle AF, Meisinger C, Midthjell K, Mohlke KL, Morken MA, Morris AD, Narisu N, Nilsson P, Owen KR, Palmer CN, Payne F, Perry JR, Pettersen E, Platou C, Prokopenko I, Qi L, Qin L, Rayner NW, Rees M, Roix JJ, Sandbaek A, Shields B, Sjogren M, Steinthorsdottir V, Stringham HM, Swift AJ, Thorleifsson G, Thorsteinsdottir U, Timpson NJ, Tuomi T, Tuomilehto J, Walker M, Watanabe RM, Weedon MN, Willer CJ, Illig T, Hveem K, Hu FB, Laakso M, Stefansson K, Pedersen O, Wareham NJ, Barroso I, Hattersley AT, Collins FS, Groop L, McCarthy MI, Boehnke M, Altshuler D. Meta-analysis of genome-wide association data

and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet. 2008; 40:638–645. [PubMed: 18372903]

13. WTCCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007; 447:661–678. [PubMed: 17554300]

14. Manolio TA, Brooks LD, Collins FS. A hapmap harvest of insights into the genetics of common disease. The Journal of clinical investigation. 2008; 118:1590–1605. [PubMed: 18451988]

15. Lencz T, Morgan TV, Athanasiou M, Dain B, Reed CR, Kane JM, Kucherlapati R, Malhotra AK. Converging evidence for a pseudoautosomal cytokine receptor gene locus in schizophrenia. Molecular psychiatry. 2007; 12:572–580. [PubMed: 17522711]

16. Shifman S, Johannesson M, Bronstein M, Chen SX, Collier DA, Craddock NJ, Kendler KS, Li T, O'Donovan M, O'Neill FA, Owen MJ, Walsh D, Weinberger DR, Sun C, Flint J, Darvasi A. Genome-wide association identifies a common variant in the reelin gene that increases the risk of schizophrenia only in women. PLoS genetics. 2008; 4:e28. [PubMed: 18282107]

17. Kirov G, Zaharieva I, Georgieva L, Moskvina V, Nikolov I, Cichon S, Hillmer A, Toncheva D, Owen MJ, O'Donovan MC. A genome-wide association study in 574 schizophrenia trios using DNA pooling. Mol Psychiatry. 2008

18. Sklar P, Smoller JW, Fan J, Ferreira MA, Perlis RH, Chambert K, Nimgaonkar VL, McQueen MB, Faraone SV, Kirby A, de Bakker PI, Ogdie MN, Thase ME, Sachs GS, Todd-Brown K, Gabriel SB, Sougnez C, Gates C, Blumenstiel B, Defelice M, Ardlie KG, Franklin J, Muir WJ, McGhee KA, MacIntyre DJ, McLean A, VanBeck M, McQuillin A, Bass NJ, Robinson M, Lawrence J, Anjorin A, Curtis D, Scolnick EM, Daly MJ, Blackwood DH, Gurling HM, Purcell SM. Whole-genome association study of bipolar disorder. Molecular psychiatry. 2008; 13:558–569. [PubMed: 18317468]

19. Baum AE, Akula N, Cabanero M, Cardona I, Corona W, Klemens B, Schulze TG, Cichon S, Rietschel M, Nothen MM, Georgi A, Schumacher J, Schwarz M, Abou Jamra R, Hofels S, Propping P, Satagopan J, Detera-Wadleigh SD, Hardy J, McMahon FJ. A genome-wide association study implicates diacylglycerol kinase eta (dgkh) and several other genes in the etiology of bipolar disorder. Molecular psychiatry. 2008; 13:197–207. [PubMed: 17486107]

20. Ferreira MA, O'Donovan MC, Meng YA, Jones IR, Ruderfer DM, Jones L, Fan J, Kirov G, Perlis RH, Green EK, Smoller JW, Grozeva D, Stone J, Nikolov I, Chambert K, Hamshere ML, Nimgaonkar VL, Moskvina V, Thase ME, Caesar S, Sachs GS, Franklin J, Gordon-Smith K, Ardlie KG, Gabriel SB, Fraser C, Blumenstiel B, Defelice M, Breen G, Gill M, Morris DW, Elkin A, Muir WJ, McGhee KA, Williamson R, Macintyre DJ, Maclean AW, St Clair D, Robinson M, Van Beck M, Pereira AC, Kandaswamy R, McQuillin A, Collier DA, Bass NJ, Young AH, Lawrence J, Nicol Ferrier I, Anjorin A, Farmer A, Curtis D, Scolnick EM, McGuffin P, Daly MJ, Corvin AP, Holmans PA, Blackwood DH, Gurling HM, Owen MJ, Purcell SM, Sklar P, Craddock N. Collaborative genome-wide association analysis supports a role for ank3 and cacna1c in bipolar disorder. Nature genetics. 2008

21. Smith EN, Bloss CS, Badner JA, Barrett T, Belmonte PL, Berrettini W, Byerley W, Coryell W, Craig D, Edenberg HJ, Eskin E, Foroud T, Gershon E, Greenwood TA, Hipolito M, Koller DL, Lawson WB, Liu C, Lohoff F, McInnis MG, McMahon FJ, Mirel DB, Murray SS, Nievergelt C, Nurnberger J, Nwulia EA, Paschall J, Potash JB, Rice J, Schulze TG, Scheftner W, Panganiban C, Zaitlen N, Zandi PP, Zollner S, Schork NJ, Kelsoe JR. Genome-wide association study of bipolar disorder in european american and african american individuals. Molecular psychiatry. 2009; 14:755–763. [PubMed: 19488044]

22. Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. Rare allele hypotheses for complex diseases. Current opinion in genetics & development. 2009; 19:212–219. [PubMed: 19481926]

23. Bearden CE, Freimer NB. Endophenotypes for psychiatric disorders: Ready for primetime? Trends Genet. 2006; 22:306–313. [PubMed: 16697071]

24. Craddock N, O'Donovan MC, Owen MJ. Genes for schizophrenia and bipolar disorder? Implications for psychiatric nosology. Schizophrenia bulletin. 2006; 32:9–16. [PubMed: 16319375]

25. APA. Diagnostic and statistical manual of mental disorders. ed 4-Revision Text. Washington DC: Author; 2000.

26. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, 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 SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, 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 MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallee C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Zeng C, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Onofrio RC, Richter DJ, Ziaugra L, Birren BW, Daly MJ, Altshuler D, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archeveque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J. A second generation human haplotype map of over 3.1 million snps. Nature. 2007; 449:851–861. [PubMed: 17943122]

27. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. Science (New York, NY. 2008; 322:881–888.

28. Scuteri A, Sanna S, Chen WM, Uda M, Albai G, Strait J, Najjar S, Nagaraja R, Orru M, Usala G, Dei M, Lai S, Maschio A, Busonero F, Mulas A, Ehret GB, Fink AA, Weder AB, Cooper RS, Galan P, Chakravarti A, Schlessinger D, Cao A, Lakatta E, Abecasis GR. Genome-wide association scan shows genetic variants in the fto gene are associated with obesity-related traits. PLoS genetics. 2007; 3:e115. [PubMed: 17658951]

29. Loos RJ, Lindgren CM, Li S, Wheeler E, Zhao JH, Prokopenko I, Inouye M, Freathy RM, Attwood AP, Beckmann JS, Berndt SI, Jacobs KB, Chanock SJ, Hayes RB, Bergmann S, Bennett AJ, Bingham SA, Bochud M, Brown M, Cauchi S, Connell JM, Cooper C, Smith GD, Day I, Dina C, De S, Dermitzakis ET, Doney AS, Elliott KS, Elliott P, Evans DM, Sadaf Farooqi I, Froguel P, Ghori J, Groves CJ, Gwilliam R, Hadley D, Hall AS, Hattersley AT, Hebebrand J, Heid IM, Lamina C, Gieger C, Illig T, Meitinger T, Wichmann HE, Herrera B, Hinney A, Hunt SE, Jarvelin MR, Johnson T, Jolley JD, Karpe F, Keniry A, Khaw KT, Luben RN, Mangino M, Marchini J, McArdle WL, McGinnis R, Meyre D, Munroe PB, Morris AD, Ness AR, Neville MJ, Nica AC, Ong KK, O'Rahilly S, Owen KR, Palmer CN, Papadakis K, Potter S, Pouta A, Qi L, Randall JC, Rayner NW, Ring SM, Sandhu MS, Scherag A, Sims MA, Song K, Soranzo N, Speliotes EK, Syddall HE, Teichmann SA, Timpson NJ, Tobias JH, Uda M, Vogel CI, Wallace C, Waterworth DM, Weedon MN, Willer CJ, Wraight Yuan X, Zeggini E, Hirschhorn JN, Strachan DP, Ouwehand WH, Caulfield MJ, Samani NJ, Frayling TM, Vollenweider P, Waeber G, Mooser V, Deloukas P, McCarthy MI, Wareham NJ, Barroso I, Jacobs KB, Chanock SJ, Hayes RB, Lamina C, Gieger C, Illig T, Meitinger T, Wichmann HE, Kraft P, Hankinson SE, Hunter DJ, Hu FB, Lyon HN, Voight BF, Ridderstrale M, Groop L, Scheet P, Sanna S, Abecasis GR, Albai G, Nagaraja R, Schlessinger D, Jackson AU, Tuomilehto J, Collins FS, Boehnke M, Mohlke KL. Common variants near mc4r are associated with fat mass, weight and risk of obesity. Nature genetics. 2008; 40:768–775. [PubMed: 18454148]

30. Sullivan PF, Lin D, Tzeng JY, van Oord den E, Perkins D, Stroup TS, Wagner M, Lee S, Wright FA, Zou F, Liu W, Downing AM, Lieberman J, Close SL. Genomewide association for schizophrenia in the catie study: Results of stage 1. Molecular psychiatry. 2008; 13:570–584. [PubMed: 18347602]

31. O'Donovan MC, Craddock N, Norton N, Williams H, Peirce T, Moskvina V, Nikolov I, Hamshere M, Carroll L, Georgieva L, Dwyer S, Holmans P, Marchini JL, Spencer CC, Howie B, Leung HT, Hartmann AM, Moller HJ, Morris DW, Shi Y, Feng G, Hoffmann P, Propping P, Vasilescu C, Maier W, Rietschel M, Zammit S, Schumacher J, Quinn EM, Schulze TG, Williams NM, Giegling I, Iwata N, Ikeda M, Darvasi A, Shifman S, He L, Duan J, Sanders AR, Levinson DF, Gejman PV, Gejman PV, Sanders AR, Duan J, Levinson DF, Buccola NG, Mowry BJ, Freedman R, Amin F, Black DW, Silverman JM, Byerley WF, Cloninger CR, Cichon S, Nothen MM, Gill M, Corvin A, Rujescu D, Kirov G, Owen MJ. Identification of loci associated with schizophrenia by genome-wide association and follow-up. Nature genetics. 2008

32. Cichon S, Craddock N, Daly M, Faraone SV, Gejman PV, Kelsoe J, Lehner T, Levinson DF, Moran A, Sklar P, Sullivan PF. Genomewide association studies: History, rationale, and prospects for psychiatric disorders. The American journal of psychiatry. 2009; 166:540–556. [PubMed: 19339359]

33. Sullivan PF. The dice are rolling for schizophrenia genetics. Psychological medicine. 2008; 38:1693–1696. [PubMed: 18533057]

34. Stefansson H, Rujescu D, Cichon S, Pietilainen OP, Ingason A, Steinberg S, Fossdal R, Sigurdsson E, Sigmundsson T, Buizer-Voskamp JE, Hansen T, Jakobsen KD, Muglia P, Francks C, Matthews PM, Gylfason A, Halldorsson BV, Gudbjartsson D, Thorgeirsson TE, Sigurdsson A, Jonasdottir A, Jonasdottir A, Bjornsson A, Mattiasdottir S, Blondal T, Haraldsson M, Magnusdottir BB, Giegling I, Moller HJ, Hartmann A, Shianna KV, Ge D, Need AC, Crombie C, Fraser G, Walker N, Lonnqvist J, Suvisaari J, Tuulio-Henriksson A, Paunio T, Toulopoulou T, Bramon E, Di Forti M, Murray R, Ruggeri M, Vassos E, Tosato S, Walshe M, Li T, Vasilescu C, Muhleisen TW, Wang AG, Ullum H, Djurovic S, Melle I, Olesen J, Kiemeney LA, Franke B, Sabatti C, Freimer NB, Gulcher JR, Thorsteinsdottir U, Kong A, Andreassen OA, Ophoff RA, Georgi A, Rietschel M, Werge T, Petursson H, Goldstein DB, Nothen MM, Peltonen L, Collier DA, St Clair D, Stefansson K. Large recurrent microdeletions associated with schizophrenia. Nature. 2008; 455:232–236. [PubMed: 18668039]

35. Schork NJ. Genetics of complex disease: Approaches, problems, and solutions. American journal of respiratory and critical care medicine. 1997; 156:S103–S109. [PubMed: 9351588]

36. Niculescu AB, Lulow LL, Ogden CA, Le-Niculescu H, Salomon DR, Schork NJ, Caligiuri MP, Lohr JB. Phenochipping of psychotic disorders: A novel approach for deconstructing and quantitating psychiatric phenotypes. Am J Med Genet B Neuropsychiatr Genet. 2006; 141B:653–662. [PubMed: 16838358]

37. Sobolewski R, O'Mullane B, Knapp RB, Reilly RB. A portable neurological monitor for use in cognitive function studies. Conf Proc IEEE Eng Med Biol Soc. 2007; 2007:2940–2943. [PubMed: 18002611]

38. Flint J, Corley R, DeFries JC, Fulker DW, Gray JA, Miller S, Collins AC. A simple genetic basis for a complex psychological trait in laboratory mice. Science (New York, NY. 1995; 269:1432–1435.

39. Zhang L, Pei YF, Li J, Papasian CJ, Deng HW. Univariate/multivariate genome-wide association scans using data from families and unrelated samples. PloS one. 2009; 4:e6502. [PubMed: 19652719]

40. Allison DB, Thiel B, St Jean P, Elston RC, Infante MC, Schork NJ. Multiple phenotype modeling in gene-mapping studies of quantitative traits: Power advantages. American journal of human genetics. 1998; 63:1190–1201. [PubMed: 9758596]

41. Ferreira MA, Purcell SM. A multivariate test of association. Bioinformatics (Oxford, England). 2009; 25:132–133.

42. Braff DL, Freedman R, Schork NJ, Gottesman II. Deconstructing schizophrenia: An overview of the use of endophenotypes in order to understand a complex disorder. Schizophrenia bulletin. 2007; 33:21–32. [PubMed: 17088422]

43. Goodwin MS, Velicer WF, Intille SS. Telemetric monitoring in the behavior sciences. Behavior research methods. 2008; 40:328–341. [PubMed: 18411557]

44. Groth-Marnat, G. Handbook of psychological assessment. ed Fourth. Hoboken, NJ: John Wiley & Sons, Inc; 2003.

45. Derogatis, LR. Scl-90-r: Administration, scoring, and procedures manual. Minneapolis, MN: National Computer Systems; 1994.

46. Derogatis, LR. Brief symptom inventory (bsi) administration, scoring, and procedures manual. ed 3rd.. Minneapolis, MN: National Computer Systems; 1993.

47. Beck, AT.; Rush, AJ.; Shaw, BF.; Emery, G. Cognitive therapy of depression. New York, NY: Guilford Press; 1979.

48. Beck, AT.; Steer, RA.; Brown, GK. Bdi-ii manual. San Antonio, TX: Psychological Corporation; 1996.

49. Spielberger, CD.; Gorsuch, RL.; Lushene, R.; Vagg, PR.; Jacobs, GA. Manual for the state-trait anxiety inventory. Palo Alto, CA: Consulting Psychologists Press; 1983.

50. Butcher, JN.; Dahlstrom, WG.; Graham, JR.; Tellegen, A.; Kaemmer, B. Manual for administration and scoring: Mmpi-2. Minneapolis, MN: University of Minnesota Press; 1989.

51. Millon, T. Manual for the mcmi-iii. Minneapolis, MN: National Computer Systems; 1994.

52. Millon, T. Millon clinical multiaxial inventory-iii manual. ed 2nd.. Minneapolis, MN: National Computer Systems; 1997.

53. Duthie B, Vincent KR. Diagnostic hit rates of high point codes for the diagnostic inventory of personality and symptoms using random assignment, base rates, and probability scales. Journal of clinical psychology. 1986; 42:612–614. [PubMed: 3745460]

54. Hill SK, Harris MS, Herbener ES, Pavuluri M, Sweeney JA. Neurocognitive allied phenotypes for schizophrenia and bipolar disorder. Schizophrenia bulletin. 2008; 34:743–759. [PubMed: 18448479]

55. Berg EA. A simple objective technique for measuring flexibility in thinking. The Journal of general psychology. 1948; 39:15–22. [PubMed: 18889466]

56. Grant DA, Berg EA. A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a weigl-type card-sorting problem. Journal of experimental psychology. 1948; 38:404–411. [PubMed: 18874598]

57. Delis, DC.; Kaplan, E.; Kramer, J. Delis-kaplan executive function system. San Antonio, TX: Psychological Corporation; 2001.

58. Heaton, RK. Wisconsin card sorting test (wcst). Odessa, FL: Psychological Assessment Resources; 1981.

59. Horner MD, Flashman LA, Freides D, Epstein CM, Bakay RA. Temporal lobe epilepsy and performance on the wisconsin card sorting test. Journal of clinical and experimental neuropsychology. 1996; 18:310–313. [PubMed: 8780965]

60. Taylor, LB. Psychological assessment of neurosurgical patients. In: Rasmussen, T.; Marino, R., editors. Functional neurosurgery. New York, NY: Raven Press; 1979.

61. Robinson AL, Heaton RK, Lehman RA, Stilson DW. The utility of the wisconsin card sorting test in detecting and localizing frontal lobe lesions. Journal of consulting and clinical psychology. 1980; 48:605–614. [PubMed: 7410659]

62. Rao SM, Hammeke TA, Speech TJ. Wisconsin card sorting test performance in relapsing-remitting and chronic-progressive multiple sclerosis. Journal of consulting and clinical psychology. 1987; 55:263–265. [PubMed: 3571684]

63. Parsons OA. Brain damage in alcoholics: Altered states of unconsciousness. Advances in experimental medicine and biology. 1975; 59:569–584. [PubMed: 1180159]

64. Alevriadou A, Katsarou Z, Bostantjopoulou S, Kiosseoglou G, Mentenopoulos G. Wisconsin card sorting test variables in relation to motor symptoms in parkinson's disease. Perceptual and motor skills. 1999; 89:824–830. [PubMed: 10665014]

65. Channon S. Executive dysfunction in depression: The wisconsin card sorting test. Journal of affective disorders. 1996; 39:107–114. [PubMed: 8827419]

66. Gold JM, Carpenter C, Randolph C, Goldberg TE, Weinberger DR. Auditory working memory and wisconsin card sorting test performance in schizophrenia. Archives of general psychiatry. 1997; 54:159–165. [PubMed: 9040284]

67. Aloia MS, Goodwin MS, Velicer WF, Arnedt JT, Zimmerman M, Skrekas J, Harris S, Millman RP. Time series analysis of treatment adherence patterns in individuals with obstructive sleep apnea. Ann Behav Med. 2008; 36:44–53. [PubMed: 18726659]

68. Sung M, Marci C, Pentland A. Wearable feedback systems for rehabilitation. Journal of neuroengineering and rehabilitation. 2005; 2:17. [PubMed: 15987514]

69. Niculescu AB 3rd, Schork NJ, Salomon DR. Mindscape: A convergent perspective on life, mind, consciousness and happiness. Journal of affective disorders. 2009

70. Niculescu AB 3rd. Polypharmacy in oligopopulations: What psychiatric genetics can teach biological psychiatry. Psychiatric genetics. 2006; 16:241–244. [PubMed: 17106426]

71. Sinha A, Hripcsak G, Markatou M. Large data sets in biomedicine: A discussion of salient analytical issues. J Am Med Inform Assoc. 2009

72. Wang Y, Miller DJ, Clarke R. Approaches to working in high-dimensional data spaces: Gene expression microarrays. British journal of cancer. 2008; 98:1023–1028. [PubMed: 18283324]

73. Tang Y, Ghosal S, Roy A. Nonparametric bayesian estimation of positive false discovery rates. Biometrics. 2007; 63:1126–1134. [PubMed: 17501943]

74. Gadbury GL, Xiang Q, Yang L, Barnes S, Page GP, Allison DB. Evaluating statistical methods using plasmode data sets in the age of massive public databases: An illustration using false discovery rates. PLoS genetics. 2008; 4:e1000098. [PubMed: 18566659]

75. Schmidt S, Shao Y, Hauser ER, Slifer SH, Martin ER, Scott WK, Speer MC, Pericak-Vance MA. Life after the screen: Making sense of many p-values. Genetic epidemiology. 2001; 21 Suppl 1:S546–S551. [PubMed: 11793734]

76. Cucala L. A hypothesis-free multiple scan statistic with variable window. Biometrical journal. 2008; 50:299–310. [PubMed: 18311853]

77. Gordon D, Hoh J, Finch SJ, Levenstien MA, Edington J, Li W, Majewski J, Ott J. Two approaches for consolidating results from genome scans of complex traits: Selection methods and scan statistics. Genetic epidemiology. 2001; 21 Suppl 1:S396–S402. [PubMed: 11793706]

78. Anderson, TW. An introduction to multivariate statistical analysis. New York: Wiley; 1984.

79. Tuncer Y, Tanik MM, Allison DB. An overview of statistical decomposition techniques applied to complex systems. Computational Statistics and Data Analysis. 2008; 52:2292–2310. [PubMed: 19724659]

80. McIntosh AR, Bookstein FL, Haxby JV, Grady CL. Spatial pattern analysis of functional brain images using partial least squares. NeuroImage. 1996; 3:143–157. [PubMed: 9345485]

81. Tura E, Turner JA, Fallon JH, Kennedy JL, Potkin SG. Multivariate analyses suggest genetic impacts on neurocircuitry in schizophrenia. Neuroreport. 2008; 19:603–607. [PubMed: 18382271]

82. Nandy RR, Cordes D. Novel nonparametric approach to canonical correlation analysis with applications to low cnr functional mri data. Magn Reson Med. 2003; 50:354–365. [PubMed: 12876712]

83. Nievergelt CM, Libiger O, Schork NJ. Generalized analysis of molecular variance. PLoS genetics. 2007; 3:e51. [PubMed: 17411342]

84. Wessel J, Schork NJ. Generalized genomic distance-based regression methodology for multilocus association analysis. American journal of human genetics. 2006; 79:792–806. [PubMed: 17033957]

85. Zapala MA, Schork NJ. Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables. Proceedings of the National Academy of Sciences of the United States of America. 2006; 103:19430–19435. [PubMed: 17146048]

86. Jiang C, Zeng ZB. Multiple trait analysis of genetic mapping for quantitative trait loci. Genetics. 1995; 140:1111–1127. [PubMed: 7672582]

87. Robinson GE. Development. Sociogenomics takes flight. Science (New York, NY. 2002; 297:204–205.

88. Barkow S, Bleuler S, Prelic A, Zimmermann P, Zitzler E. Bicat: A biclustering analysis toolbox. Bioinformatics (Oxford, England). 2006; 22:1282–1283.

89. Madeira SC, Oliveira AL. Biclustering algorithms for biological data analysis: A survey. IEEE/ACM transactions on computational biology and bioinformatics / IEEE, ACM. 2004; 1:24–45.

90. Anderson, TW. An introduction to multivariate statistical analysis. New York: Wiley; 2003.

91. Johnson, RA.; Wichern, DW. Applied multivariate statistical analysis. New York: Prentice Hall; 2008.

92. Le Cao KA, Rossouw D, Robert-Granie C, Besse P. A sparse pls for variable selection when integrating omics data. Statistical applications in genetics and molecular biology. 2008; 7 Article 35.

93. Mielke, PW.; Berry, KJ. Permutation methods: A distance function approach. New York: Springer; 2001.

94. Fraley C, Raftery AE. Model-based clustering, discriminant analysis, and density estimation. Journal of the American Statistical Association. 2002; 97:611–631.

95. Lubke GH, Muthen B. Investigating population heterogeneity with factor mixture models. Psychological methods. 2005; 10:21–39. [PubMed: 15810867]

DSM-IV-TR Criteria for Schizophrenia (APA, 2000)

A. Characteristic symptoms: Two or more of the following, each present for a significant portion of time during a 1-month period:

1. Delusions
2. Hallucinations
3. Disorganized speech
4. Grossly disorganized or catatonic behavior
5. Negative symptoms, i.e., affective flattening, alogia, or avolition

* Only one criterion A symptom is required if delusions are bizarre or hallucinations consist of a voice keeping up a running commentary on the person's behavior or thoughts, or two or more voices conversing with each other.

B. Social/occupational dysfunction

C. Duration of at least 6 months

Note: In addition, there are currently three recognized subtypes of Schizophrenia, including Paranoid Type, Disorganized Type, and Catatonic Type, each with additional diagnostic criteria over and above those listed above.

In addition, in the broader category of Psychotic Disorders currently recognized in DSM-IV-TR, there are a total of nine subtypes, including Schizophrenia, Schizophreniform Disorder, Schizoaffective Disorder, Delusional Disorder, Brief Psychotic Disorder, Shared Psychotic Disorder, Psychotic Disorder Due to a General Medical Condition, Substance-Induced Psychotic Disorder, and Psychotic Disorder Not Otherwise Specified, each with separate and different diagnostic criteria and different characteristic disease courses and treatment protocols.

Figure 1.
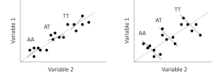DSM-IV-TR Criteria for Schizophrenia (APA, 2000)

Figure 2.
Graphical depiction of the relationship between two variables mediated by genetic locus with two alleles, A and T, that will provide: left panel: small, if any, power increase when studied simultaneously via multivariate methods relative to univariate analysis of each variable; and right panel: large power increase when studied via multivariate methods relative to univariate analysis of each variable.

Table 1

Selected Recent Genome-Wide Association Studies and Discoveries

Disease/Trait	Gene(s)/Loci	Reference
Celiac disease	RGS1, IL1RL1, IL18R1, IL18RAP, SLC9A4, IL12A, SCHIP1	[7]
Colorectal cancer	DQ515897, SMAD7	[9]
C-reactive protein	LEPR, HNF1A, GCKR, IL6R	[8]
Fetal hemoglobin levels	HBB, BCL11A	[11]
Glaucoma	LOXL1	[4]
Nicotine dependence, lung cancer, and peripheral arterial disease	CHRNA3, CHRNA5, CHRNB4	[10]
Obesity	FTO	[1]
Prostate cancer	MSMB, KLK3, SLC22A3, LMTK2	[5]
Systemic lupus erythematosus	ITGAM, KIAA1542, PXK	[6]
Type 2 diabetes	JAZF1, CDC123, CAMK1D, TSPAN8, LGR5, THADA, ADAMTS9	[12]

Table 2

Recent Genome-Wide Association Studies in Schizophrenia and Bipolar Disorder

Disease/Trait	Gene(s)/Loci	Reference
Schizophrenia	CCDC60	[17]
Schizophrenia	RELN (in women)	[16]
Schizophrenia	CSF2RA	[15]
Bipolar Disorder	ANK3, CACNA1C	[20]
Bipolar Disorder	MYO5B, TSPAN8, CACNA1C	[18]
Bipolar Disorder	DGKH	[19]
Bipolar Disorder	PALB2, NDUFAB1, DCTN5	[13]

Table 3

Data Analysis Methods that Interrogate the Relationship Between Two Sets of Variables

Technique	References	Properties	Description
Biclustering; QR-mode factor analysis	[88,89]	Description	Find groups of individuals that share features on subsets of variables
		Advantages	Combines clustering and factor analysis
		Disadvantages	Hard to assess the cluster significance probabilistically
Standard Multivariate Regression	[90,91]	Description	Assess the linear relationships between two sets of variables
		Advantages	Long history and experience; ease of interpretation and significance
		Disadvantages	Does not work with highly correlated, numerous independent variables
Canonical Correlation; PLS	[20,41,92]	Description	Determine relationships between combinations of variables in two sets
		Advantages	Find patterns in relationships between two large data sources
		Disadvantages	Does not necessarily allow intuitive interpretations
Distance Analysis	[85,93]	Description	Predict the distance between individuals based on one set of dependent variables from multiple independent variables
		Advantages	Can accommodate a very large number of dependent variables
		Disadvantages	Does not find optimal subset of 'important' dependent variables
Mixture Models	[94,95]	Description	Identify subgroups of individuals leveraging multiple variables
		Advantages	Parametric formulations have easily interpretable parameters
		Disadvantages	Cannot accommodate a large number of independent of co-variables