Neuroinformatics

1539-2791/06/21–50/$30.00 (Online) 1559-0089
DOI: 10.1385/NI:4:1:21

Reviews

Imaging Phenotypes and Genotypes in Schizophrenia

Jessica A. Turner,[1] Padhraic Smyth,[2] Fabio Macciardi,[3] James H. Fallon,[4] James L. Kennedy,[5] and Steven G. Potkin[*,1]*

[1]Department of Psychiatry and Human Behavior, University of California, Irvine, CA 92697; [2]Department of Computer Science, University of California, Irvine, CA 92697; [3]Policlinico MultiMedica and Department of Science and Biomedical Technologies, University of Milano, Italy; [4]Department of Anatomy and Neurobiology, University of California, Irvine, CA 92697; and [5]Neuroscience Research Department and the Neurogenetics Section, CAMH, University of Toronto, Toronto, Ontario, M5T 1R8, Canada

Abstract

Schizophrenia is associated with subtle structural and functional brain abnormalities. Both recent and classical data suggest that it is a heterogeneous disorder that is clearly heritable. The cause and course of schizophrenia are poorly understood, and classical categories of clinical symptoms have not been particularly useful in identifying its pathophysiology or predicting its treatment. The possible genetic risk factors for schizophrenia are numerous; however, the connection between the genotype and the time-course, or the multifaceted symptoms of the disease, has yet to be established. Brain imaging methods that study the structure or function of the cortical and subcortical regions have also identified distinct patterns that distinguish schizophrenics from controls, and that may identify meaningful subtypes of schizophrenia. The predictive relationship between these imaging phenotypes and disease characteristics such as treatment response is only beginning to be revealed. The emergence of the field of imaging genetics, combining genetic, and neuroimaging data, holds much promise for the deeper understanding and improved treatment of diseases such as schizophrenia. In this article we review some of the key findings in imaging phenotyping and genotyping of schizophrenia, and the initial endeavors at their combination into more meaningful and predictive patterns, or endophenotypes identifying the relationships among clinical symptoms, course, genes, and the underlying pathophysiology.

Index Entries: Schizophrenia; neuroimaging; genetics; endophenotypes; data mining.

(Neuroinformatics DOI: 10.1385/NI:4:1:21)

*Author to whom all correspondence and reprint requests should be addressed.
E-mail: sgpotkin@uci.edu

Introduction

In the past 20 yr, significant strides have been made in imaging the structure and function of the living human brain. Equal steps have been made in understanding the human genome and its role in disease. Particularly unambiguous advancements have occurred in single-gene dominant Mendelian disorders such as cystic fibrosis and Huntington's disease. Most diseases, however, are not single-gene disorders and even single-gene diseases can manifest themselves in a multitude of ways. Complex behavioral diseases with developmental and degenerative components such as schizophrenia appear to involve the combined effects of multiple genes and important interactions with the external and internal environment (Basile et al., 2002; Kennedy et al., 2003). This complexity has created a roadblock in the clinical, genetic, and functional understanding of these diseases. Neuroimaging reveals many important aspects of schizophrenia; however, when considered in isolation, it ignores the strong hereditary aspect. Conversely, although genetic studies are powerful, they are also limited in their explanatory potential, given that monozygotic twin concordance for schizophrenia is approx 40% (for review, *see* Tsuang, 1998; Sivagnansundaram et al., 2003). With its wide range of symptoms, severity, and cognitive dysfunction, schizophrenia is a good model for developing the combined field of imaging genetics. Given the known importance of both genetics and environment in brain function, and the role of neuroimaging in revealing brain dysfunction, the synergism of integrating genetics with brain imaging will fundamentally change our understanding of normal human brain function, and in disease.

Mental illness, as well as normal brain function, has a hereditary component; therefore, it is essential that the genes related to aspects of brain development, and mental function and dysfunction, be considered. It is clear that brain function mediates mental illness; therefore, use of advanced brain imaging techniques can clarify the dysregulation of neural circuits. Although imaging studies have revealed many aspects of dysfunction in schizophrenia (e.g., Weinberger and Berman, 1996; Bunney et al., 1997; Weinberger et al., 2001; Fallon et al., 2003; Potkin et al., 2003) their explanatory power has been limited by ignoring the well-documented heritability of schizophrenia in general and the genetic background of individuals in particular. Clinical symptom patterns in schizophrenia may be more distant from the biological mechanisms underlying the causes of schizophrenia than brain images of neuronal circuit dysfunction (e.g., Basile et al., 2002; Potkin et al., 2002a, 2003b). Discovering new relationships among imaging, clinical, and cognitive phenotypes, considered within an individual's genetic background, may be critical to understanding the heterogeneity within schizophrenia.

The clinical presentation of schizophrenia varies widely, with some clusters of symptoms beginning in childhood and others beginning later in life; the majority of the symptoms, however, begin in early adulthood. Previous classifications based solely on symptoms, i.e., paranoid, simple, disorganized, and undifferentiated, have not been useful to increasing our understanding of the pathophysiology of schizophrenia. Any theory of schizophrenia, however, must deal with the variable pattern of clinical symptoms in schizophrenia: Some patients, but not all, experience hallucinations and delusions (positive symptoms); some exhibit apathy and social withdrawal (negative symptoms); some, difficulties in the timing and coordination of thoughts and memories; some, difficulties in attention and the organization of memory; and some experience mood dysregulation and impulsivity (schizoaffective and suicidality symptoms). The latter suicidal symptoms have increasingly been recognized as a separate symptom domain or subtype (Potkin et al., 2003a), with as many as 40% of schizophrenic persons attempting

suicide and as many as 9% succeeding (Meltzer et al., 2003).

Modern pharmacological and rehabilitation treatments have notably improved the lives of schizophrenic patients, with decreases in both positive and negative symptoms, some cognitive improvement, and perhaps a decrease in depression as well as suicide. Yet despite these advances, few if any patients return to their premorbid level of functioning. Most continue to suffer from some level of hallucination and/or negative symptoms, occupational and functional dysfunction, and continuing cognitive deficits. Consequently, many patients remain only partially responsive to medication and rehabilitation treatments, and as many as 20% are considered refractory to treatment. Just as there are significant individual differences in which clinical symptoms develop, there are also notable individual differences in the efficacy of pharmacological treatment, as well as individual differences in the risk for developing side effects, such as tardive dyskinesia (a movement disorder primarily involving orofacial disfigurement). Thus, there is an obvious need to better understand the pathophysiology of schizophrenia and a corresponding need for improved and possibly individually tailored treatments. It is likely that some differences in outcome, drug response and side-effect risk reflect different endophenotypes, i.e., "measurable components unseen by the unaided eye along the pathway between disease and distal genotype, and thus closer to the biological expression or pathophysiology of the illness than clinical symptoms" (Gottesman and Gould, 2003). These measurable components that distinguish among groups of schizophrenic subjects in our conceptualization include combinations of clinical symptoms, cognitive dysfunction, patterns of brain activation, and genetic influences.

In the following sections we selectively review some of the imaging phenotypes in the research literature on schizophrenia; we overview the basic framework of genetic analysis, and the genetic risk factors and correlations with schizophrenic symptoms. The issues of genetic risk analysis and the complexity of these analyses is explored in their combination in imaging genetics to uncover potential endophenotypes leading to the discovery of validated subgroups within schizophrenia.

Imaging as a Discriminant Among Schizophrenic Subtypes

The functional and structural abnormalities distinguishing schizophrenics from nonpsychiatric subjects have been comprehensively reviewed elsewhere (e.g., Shenton et al., 2001; Fallon et al., 2003). Schizophrenia has long been acknowledged to be a heterogeneous disorder (Winokur, 1975; Jeste et al., 1982), though the codified Diagnostic and Statistical Manual of Mental Disorders [DSM]-IV subtypes have not been particularly useful in defining treatment-related distinctions, with the exception of the paranoid subtype. Classically, paranoids have been defined as having a better prognosis; however, this does not have the support of careful investigations. We will focus in this section on selected literature that uses imaging to distinguish among groups of schizophrenic patients, with the aim of identifying symptom profiles which predict treatment response or clinical progress.

Other symptom-based classifications have shown some physiological underpinnings. Northoff et al. (2004) studied the relatively rare subtype, catatonia. Postacute catatonic, akinetic subjects differed from matched noncatatonic psychiatric subjects in a functional imaging study using emotional stimuli, in that catatonic behavioral symptoms correlated with an orbitofrontal dysfunction (Northoff et al., 2004). The orbitofrontal circuitry is implicated in the control of social function and awareness of social consequences (e.g., Bechara et al., 1994; Stone et al., 1998; Heberlein et al., 2004; Roberts et al., 2004). This dysfunction, in combination with the clinical symptoms, provides a potential

endophenotype for greater understanding of the pathophysiology of catatonia.

The Kraepelinian subtype has been used to characterize a group of patients with severe and unremitting course, and poor response to antipsychotic medication. In an [18F]-Fluoro-2-deoxy-D-glucose (FDG)-positron emission tomography (PET) study of serial verbal learning, Kraepelinian subjects had lower metabolic rates in the frontal lobe and cingulate gyrus, and hypofrontality in comparison with non-Kraepelinian schizophrenic patients (Buchsbaum et al., 2002). Gur et al. (1995) also used FDG-PET to study metabolic differences among schizophrenic subtypes of differing symptom severity, and controls (Gur et al., 1995; Turetsky et al., 1995). Rather than lower metabolic rates in the frontal areas in the more severe patients, greater metabolism in the mid-left temporal area was found generally in schizophrenic patients in comparison with controls. This was particularly pronounced in the negative and Schneiderian (specific types of delusions and hallucinations, such as thought insertion or broadcasting, or voices commenting on subject's behavior) subtypes but not in the paranoid subtypes. Lower metabolism in the left temporal area relative to the right was correlated with better premorbid adjustment and better outcome.

Patients with enduring negative symptoms, also classified as a deficit subtype, differed from nondeficit schizophrenic patients in having a lower frontal but not hippocampal activity during a 0–15 blood flow study of working memory (Heckers et al., 1999). Wolkin et al. (1996) also found that greater negative symptom severity correlated with decreased metabolic response to a conventional antipsychotic treatment (Wolkin et al., 1996). A PET study contrasting patients with predominantly negative symptoms and with predominantly positive symptoms also indicated the physiological distinctions among these subtypes (Potkin et al., 2002b). The ventral, frontal, and temporal systems of negative symptom patients showed decreased activation as well as increased activity of the deep cerebellar nuclei, possibly implicating the cerebellar activity as a mechanism to compensate for the decreased frontal activity via cerebello-thalamic-frontal circuitry. These imaging patterns distinguishing subgroups of schizophrenia hold promise as endophenotypes that could point the way to specifically targeted treatments for these groups.

The time-course of schizophrenic progression is varied, with episodes of increasing and decreasing symptom severity, even under the best medical control. The differences that are found among long-term chronic schizophrenics and controls or among chronic schizophrenics of varying symptom profiles, may be confounded with differing antipsychotic regimens. The study of first-episode patients, either medication-naive, or after first treatment, is critical to understanding the progress of the disease as it controls for chronicity and the effects of treatment. Such subjects show a decrease in left entorhinal cortical volume, which is correlated with the severity of the delusions (Prasad et al., 2004). Deficits in dorsolateral prefrontal cortex (DLPFC) gray matter have also been observed in such patients; these DLPFC volumetric differences have been observed to vary with regulator of G protein signaling subtype-4 (RGS-4) genotype, consistent with microarray studies of showing underexpression of RGS-4 in the DLPFC in schizophrenics (Prasad et al., 2005). Each of these clinical subpopulations of schizophrenia shows an imaging phenotype that reveals more about the pathology of the disease; particularly promising is the repeated finding of hypofrontality correlating with the severity of negative symptoms, a putative endophenotype that may help target treatments more effectively.

Genetic Analyses

Classical studies have involved Mendelian disorders caused by a single gene; most of these

disorders have been mapped and cloned, with resultant increased understanding of the pathophysiology of the disease. Within the Mendelian paradigm, a mutation in a single gene is the only cause of a genetic disease. According to this paradigm, a "mutation" is a rare mistake in the DNA sequence of the gene occurring at a frequency less than 1% in the general population. Because we have two alternate forms of any given gene (alleles) that make up a genotype, one inherited from the father and the other from the mother, a genetic disease can occur (1) if we inherit one mutated allele from the father or from the mother (a condition described as "dominant" transmission of disease) or (2) if we inherit two mutated alleles from both parents (a condition referred to as "recessive"). Typically, Mendelian disorders have a very low prevalence in the population, as mutations are rare events that are always leading to a pathological state.

Alternatively, there are other variations in the DNA sequence of the majority of genes that are much more frequent in the population, but that do not necessarily lead to a pathological condition: this variability generates various different forms of the genes (i.e., they create different alleles at almost any given gene, with minor biochemical/biological consequences within a range of "normality" for each variation) that may be responsible for subtle differences across humans. Considering together all these small differences and knowing that each person differs from one another for approx 3–10 millions of these minor DNA variations, it is easy to understand that such a large allelic variability is also responsible for the observed difference of the humankind.

These minor DNA variations are common in the general population and they can ultimately be responsible for several common diseases, whenever they occur in a particular combination whose overall effect causes a biochemical/biological pathological condition. Under this hypothesis, many common disorders have nonetheless a genetic component presenting with a non-Mendelian mode of transmission that is neither dominant nor recessive, in which many small DNA variations (alleles), each one not pathological *per se*, combine in a "pathological" ensemble. These common differences in genetic sequence can influence brain function, symptom constellation, response to medication, and so forth. Some of the complexities of gene variation and interaction have been summarized in various reviews (Risch and Merikangas, 1996; Risch et al., 2002; Kennedy et al., 2003). Finally, it should be noted that in common diseases, environmental effects may be substantial (e.g., diabetes).

The simplest and most common kind of genetic (DNA sequence) variation is defined as single-nucleotide polymorphism (SNP), i.e., a difference in a nucleotide at a particular site of the DNA sequence. Another kind of variation can be two (or more) nucleotides repeating a variable number of times, defined as micro- or mini-satellites. At a particular chromosomal location, only one variant can be present. When many polymorphic variations, usually SNPs, are very close to each other along a chromosomal region of variable size they can be inherited together and this linear combination of polymorphisms is termed a haplotype. This happens because alleles of neighboring SNPs are not transmitted independently; the measure of their dependency in transmission is called linkage disequilibrium (LD), which ranges from 0 to 1, the last value representing complete identity across two consecutive SNPs. The size of such regions range from three to several hundred kilobases. The haplotypes thus represent another form of normal variation across a given region of a chromosome inherited from parents. Each person has two haplotypes for any such given region, equivalent to having two alleles at multiple sites.

A consortium has been created (HapMap) to build the haplotype map of the entire human

genome: HapMap has developed an initial investigation of the "best" SNPs to generate reliable haplotypes, and further refinements are rapidly progressing (Daly, 2002), making it possible in the future to incorporate these advances into neuroinformatics databases and statistical analyses. To identify the discrete haplotypes across the human genome, the HapMap consortium has already genotyped more than 1.5 million SNPs, with an open end regarding how many SNPs will be needed to build an efficient haplotype map. In fact, the driving hypothesis is that a haplotype map built with a subset (300,000–1 million) of the total number of SNPs in the genome (5–10 million) would enable us to capture the overall information of the genome itself despite the reduction of the number of polymorphisms tested. This hypothesis is supported by our current knowledge of the genome, as we know that each haplotype corresponds to a chromosomal region of low variability (also called a "block") interspersed with short regions of few kilobases characterized by high variability and by the absence of a block-haplotype structure. Ideally, these small regions of high variability correspond to areas of recombination, or cross overs, i.e., regions in which homologous chromosomes break during the meioses and then recombine, allowing for an exchange of DNA material between the "paternal" and the "maternal" chromosomes. Thus, genotyping only those SNPs that univocally detect each haplotype-block (called haplotype-tagging SNPs [htSNP]) in a case–control design, we could identify any gene/variation responsible for any trait that we want to investigate. At present, we still do not know the exact correlation between haplo-blocks and genes: Sometimes one gene corresponds to one block, in many other cases one gene entails several blocks, which may also contribute to explain the evolution of genes and the existence of "families" of genes with a high degree of similarity to each other. Moreover, and probably depending also on the number of recombination events that occurred in a given population, adjacent haplotype-blocks have a variable degree of "LD" with one another, as if each haplotype acts like a single polymorphism despite being made up by some to several SNPs, again supporting the hypothesis that haplo-blocks have a reduced degree of internal variability. Haplotypes interact with each other not only locally but also anywhere across the genome. Understanding the role of these haplotypes in risk of disease, treatment response, and development of side effects is a research priority in for schizophrenia.

To understand the genetic architecture of complex traits and to dissect the genetic component of the trait is a considerable challenge, despite the advances in knowledge on this subject and the sophistication of contemporary technology. A generally accepted view of common complex genetic traits is that the contributing genes are of small or minor effect as opposed to major genes characteristic of simple Mendelian traits (Ott and Hoh, 2000). Opinions differ regarding whether these genes present with common (Risch and Merikangas, 1996; Chakravarty, 1999) or rare variants (Terwilliger and Weiss, 1998; Weiss and Terwilliger, 2000; Pritchard, 2001) and regarding the best way to describe the contribution that each gene exerts on the trait. Consequently, there is considerable debate concerning the appropriate strategy to identify genes for complex disorders.

Linkage approaches have been successfully applied to cloning more than 1200 genes for Mendelian disorders but do not have sufficient power for positional cloning in complex disorders such as schizophrenia, even though they have enabled the identification of linked chromosomal regions. The emphasis has shifted to association methods employing candidate genes and LD mapping with SNPs and haplotypes (Risch and Merikangas, 1996; Tabor et al., 2002; Lohmueller et al., 2003). Within the

association strategy there is a dichotomy between "sequence-based" and "map-based" approaches that may also be termed "functional" vs "positional" (Peltonen and McKusick, 2001; Botstein and Risch, 2003). The functional approach involves selection of candidate genes based on knowledge of disease pathogenesis, functions of the selected genes and where applicable data from an animal model of disease. However, this knowledge is not always available, particularly for psychiatric disorders. A map-based approach employing genome-wide association is today technologically possible, but cost-effectiveness is still a concern.

The current approach proposed for association mapping of a given candidate gene or of a (small) region (e.g., Wang et al., 2003) is to (1) detect informative polymorphisms (=SNPs), (2) determine the LD pattern across the SNPs and construct haplotypes, and (3) genotype the selected htSNPs (e.g., Gabriel et al., 2002). Haplotypes can significantly improve the power of association mapping (Davidson, 2000; Zaykin et al., 2002) especially when we carefully consider the LD block structure that is at the origin of haplotypes (Gabriel et al., 2002). To implement positional cloning by means of LD mapping cost-effectively, the candidate region must be small enough to permit typing SNPs at an average density of 10 kb (Carlson et al., 2003; Reich et al., 2003). If so, a reasonable number of SNPs would be sufficient for the application of positional cloning.

Haplotypes may be directly responsible for the observed variation in the trait of interest, through the combined effects of multiple sequence variants on promoter activity or protein structure and function (e.g., Devlin and Roeder, 1999; Drysdale et al., 2000; Hohe et al., 2000; Joosten et al., 2001). Even when a single, presumably unobserved polymorphism accounts for the trait variation, nearby markers may form haplotypes that are in much higher LD with that functional polymorphism than are the individual markers, because the disequilibrium between a single site and whole haplotypes includes all pairwise as well as higher order disequilibria terms (Bennett, 1954). Thus, even considering that "complex" alleles (i.e., alleles in which there are multiple polymorphic variants jointly present) represent the true risk factors for complex traits (e.g., Kim et al., 2003), a haplotype LD strategy allows them to be detected (Botstein and Risch, 2003).

Genetic Risk Factors and Correlations

Genetic studies of schizophrenia such as those summarized in the following section have classically used the "candidate gene" approach; this utilizes current knowledge of brain chemistry and pharmacology to select genes that are most likely to contribute to specific phenotypes (such as cognitive features, drug response, or distinct side effects). Polymorphisms in or near the coding region of a candidate gene that encode a protein structure important in cognition or with which a drug is thought to interact are assessed using molecular genetic techniques. Subsequently, the frequency of one polymorphic allele is compared with the frequency of the other(s) to substantiate the role of the polymorphism in the expression of a phenotype ("allelic association design"). However, the detection of a positive association between a given genotype and a determined phenotype does not prove a causal relationship. In addition to the possibility of false positives, the marker allele used for genotyping may not be the true function-altering variant, but might be only in close physical proximity at a genomic level.

The examples given below demonstrate the levels and range of complexity of genetic influences on schizophrenia and consequently the need to develop new analytical approaches. These approaches must be multivariate and expandable to reflect the inevitable increasing complexity.

Genetics of Brain Development and Structure

The mechanisms underlying development of the mammalian central nervous system are of fundamental importance for research into schizophrenia and psychiatric disorders in general (*see* Klempan et al., 2005, for a more extensive review). The processes of neurulation, patterning, neuronal specification, and synaptogenesis, as well as the functional dynamics of neurotransmission, are governed by the coordinated actions of products from a wide array of genes. This pattern of gene expression has been related to the etiology of schizophrenia, although not so successfully to clinical subtypes of schizophrenia. Table 1 summarizes many of the genes implicated in schizophrenia.

For example, early in brain development, during the neurulation stage, regionalization within the developing telencephalon is controlled by the ventralizing properties of the sonic hedgehog (Shh) gene expression (Kohtz et al., 1998), and distinct dopaminergic and serotonergic neuronal subpopulations are induced along the anterior–posterior axis at different times by Shh signaling (Hynes and Rosenthal, 1999). A gene downstream from Shh action in the neurulation stage is NOTCH4. The description of a strong association between a promoter base-pair substitution and exon 1 (CTG)*n* repeat of the NOTCH4 gene with schizophrenia (Wei and Hemmings, 2000) in a region (6p21.3) previously associated with schizophrenia (Schwab et al., 1995; Straub et al., 1995) has not been replicated (McGinnis et al., 2001; Sklar et al., 2001; Ujike et al., 2001; Fan et al., 2002; Swift-Scanlan et al., 2002). The first study now appears more likely to be a false-positive association, highlighting the need for better statistical procedures in genetic studies of schizophrenia to control for false-positives without losing power to detect true effects (discussed later in further detail).

Cell adhesion molecules (CAMs) are cell membrane proteins that mediate adhesion between neural cells exerting a key role in cell migration, morphogenesis, and differentiation in the developing brain. Subtle changes in neural CAM (NCAM) gene expression in the schizophrenia brain might be responsible for the observed abnormalities of cell migration in schizophrenia. A number of studies have revealed NCAM alterations in schizophrenia, including decreased polysialylated NCAM in the hippocampus (Barbeau et al., 1995), and increased NCAM in the prefrontal cortex and hippocampus (Honer et al., 1997) suggesting that further genetic investigation is warranted.

Reelin plays a major role in the layering of neurons into their specific brain target area, including the prefrontal cortex and hippocampus, once their migration is completed (Rice and Curran, 2001). In humans, the reelin gene (RELN) maps to 7q22 (DeSilva et al., 1997). RELN is abundantly synthesized and secreted in the extracellular matrix during early stages of brain development. Reelin continues to be synthesized throughout life by a select population of GABA-ergic interneurons including layers I and II of various cortical areas (Rodriguez et al., 2000). The expression pattern of reelin has been investigated as a possible etiological factor underlying the neurodevelopmental anomalies in schizophrenia. Two independent groups observed a 30–60% reduction of RELN expression in the prefrontal cortex and in the hippocampus in schizophrenic patients (Fatemi et al., 2000; Guidotti et al., 2000). Although several polymorphisms have been described for RELN, no detailed sufficiently powered studies investigating the possible role of RELN sequence variants in schizophrenia have been reported.

The primary molecules involved in regulation of neuronal survival and differentiation are the neurotrophins, including neurotrophic factors (NT1–5) and brain-derived neurotrophic

Table 1
Neurodevelopmental Genes and Reports of Expression-Based and Genetic Analysis in Schizophrenia

Gene	Function	Locus of Interest	Genetic and Expression-Based Analysis
Mash1	Neurulation	–	
Notch		NOTCH4 (6p21.3)	Wei and Hemmings (2000)
Delta		DLL1 (6q27)	
Neurogenin		NGN1 (5q23-q31)	
NeuroD		NEUROD (2q32)	
Sonic Hedgehog		SHH (7q36)	
Wnt		WNT1 (12q12-q13)	Cotter et al. (1998); Miyaoka et al. (1999)
Krox20	Patterning	EGR2 (10q21.1-q22.1)	
Hox		HOXB (17q21.3)	Kennedy et al. (1992)
Dlx		DLX1 (2q32)	
Emx		EMX2 (10q26.1)	
Gbx		GBX2 (2q36-q37)	
Nkx		TITF1 (14q13)	
Otx		OTX2 (14q21-q22)	
Pax		PAX6 (11p13)	Stober et al. (1999)
POU		POU3F3 (3p14.2)	
NCAM L1CAM	Cell migration/ neurite extension	NCAM1 (11q23.1)	Doherty et al. (1990); Vicente et al. (1997)
N-Cadherin		NCAD (18q11.2)	
Reelin		RELN (7q22)	Fatemi et al. (2000); Guidotti et al. (2000)
NGF	Neuronal	NGFB (1p13.1)	
BDNF	Survival	BDNF (11p13)	Muglia et al. (2003)
NT-3		NTF3 (12p13)	Nanko et al. (1994)
NT-4/5		NTF5 (19q13.3)	
GDNF		GDNF (5p13.1-p12)	Lee et al. (2001)
CNTF		CNTF (11q12.2)	Thome et al. (1996)
Chondrex	Cell growth/ migration	YKL40 (1q32.1)	Chung et al. (2003)
GSK3		GSK3 (3q13.3)	Nadri et al. (2004)
DISC1		DISC1 (1q42)	Hodgkinson et al. (2004)
SNAP25	Presynaptic/	SNAP25 (20p11.2-p12)	Tachikawa et al. (2001); Wong et al. (2003)
Syntaxin	exocytosis	STX1A (7q11.23)	Wong et al. (2004)
Synaptobrevin		VAMP1 (12p)	
Synapsin		SYN3 (22q12.3)	Ohmori et al. (2000)
Complexin		CPLX2 (5q35.3)	Harrison and Eastwood (1998)
Synaptophysin		SYP (Xp11.23-p11.22)	
Synaptotagmin		SYT1 (12cen-q21)	
NMDA	Postsynaptic	GRIN-1, GRIN-2A,B	Mohn et al. (1999); Martucci et al. (2003)
Myelin Oligodendrocyte Glycoprotein	Myelination	MOG (MAG, MBP)	Malfroy et al. (1995); Zai et al. (2004)

factor (BDNF). The neurotrophins play vital roles in cell survival and differentiation in all regions of the brain. They may be associated with reductions in cortical and whole brain volume (Lawrie and Abukmeil, 1998) and the high-neuronal density (Selemon et al., 1995) seen in morphometric analyses of the schizophrenic brain. Postmortem studies of schizophrenia patients have shown alterations of BDNF protein levels, specifically increases in cortical regions and decreases in the hippocampus (Durany et al., 2001; but not Takahashi et al., 2000, who found increases in both regions).

Genetic investigation of the BDNF gene in schizophrenia was initiated following discovery of a dinucleotide repeat polymorphism located within the intronic region of the gene (Proschel et al., 1992). The initial published report included a Japanese case–control study (Sasaki et al., 1997) and a large Irish study (Hawi et al., 1998), both reporting negative findings for this polymorphism. In contrast, the recent results of transmission disequilibrium test (TDT)-based analysis of a schizophrenia sample of Italian and Canadian families (Muglia et al., 2003) have shown increased transmission of the 170 bp allele ($p = 0.0081$) and reduced transmission of the 174 bp allele ($p = 0.010$) to probands.

The shape and extent of adult brain structures are highly dependent on neuronal activity. An integral part of neuronal activity is the process of neurotransmission. At its most fundamental level, neurotransmission involves exocytosis, which is regulated most centrally by a group of proteins known as soluble *N*-ethylmaleimide-sensitive attachment factor protein receptors (SNAREs). The core SNARE proteins include synaptosomal-associated protein of 25 kDa (SNAP-25), the syntaxins, and vesicle-SNAREs such as the synaptobrevins. Postmortem assays of SNAP-25 immunoreactivity in schizophrenia have revealed changes in the inferior temporal and prefrontal association cortices (Thompson et al., 1998), with decreased expression in certain areas but increased expression in others. A study of hippocampal connectivity found reduced cortical SNAP-25 protein, most notably in the terminal fields of entorhinal cortex projections of schizophrenics (Young et al., 1998), and in the cerebellum of schizophrenics (Mukaetova-Ladinska et al., 2002). The original SNAP25 findings in area 10 (along with reduced synaptophysin expression) was confirmed, yet showed no alteration in SNAP-25 mRNA expression (Karson et al., 1999).

The myelin oligodendrocyte glycoprotein (MOG) gene represents an intriguing candidate gene for schizophrenia and brain structure/function phenotypes based on its role in brain development and the immune system. MOG is directly involved in the development of white matter in the brain (reviewed by Davis and Haroutunian, 2003), and important changes have been found in white matter gene expression in postmortem schizophrenia brains. Hakak et al. (2001) screened homogenized tissue from schizophrenia brains vs matched controls for expression levels of 6500 genes. Of these, only seven were significantly downregulated in schizophrenia. Remarkably, six of those seven genes were myelin related. In a very recent replication study by the same group, the postmortem brains of 13 schizophrenia patients and 13 matched controls were examined for differences in gene expression using the Affymetrix U133 chip. Analyzing the 39,000 transcripts that were interrogated by this chip, Davis et al. (2003, ACNP) found that several genes in the myelin system showed changes in expression. One of the main expression differences was with the MOG gene, which was significantly reduced in the cingulate and in area 44 of the schizophrenia brains (for review, *see* Davis et al., 2003).

The secondary reason for involvement of MOG in schizophrenia is that MOG is known

to mediate the function of the complement cascade in the immune system. Furthermore, the chromosomal location of the MOG gene is on 6p22, close to the region of the loci for the human leukocyte antigens (HLA) on 6p21, which further implicates a secondary immune function of the MOG protein. Both the physical and the genetic distance between MOG and the HLA loci are extremely small, especially considering that the HLA-A Locus is at 0.1 Mb from MOG. The autoimmune hypothesis of schizophrenia has been under consideration for several decades (Cazzullo et al., 1974), and several studies of HLA region markers have found a positive association (Wright et al., 2001). A relatively large number of genome scan studies for schizophrenia have found linkage in this area of 6p22-24 (e.g., Schwab et al., 1995; Straub et al., 1995). Although the gene for dysbindin is also located in the vicinity, and it has been associated with schizophrenia in genetic studies (Straub et al., 2002), the relative risk imparted by this gene is low (1.6) and does not explain the full effect of the linkage findings from the genome scans.

Summarizing the findings outlined in this section on neurodevelopment genes, brain structure, and schizophrenia, there are indeed a large number of candidate genes to consider for phenotypes of brain structure and function and the disease symptoms. Candidate genes can be organized or nested into prioritized subgroups for systematic investigation, such as synaptic, neurotransmitter related, or cognitive. These functionally nested groupings can be used as more complex constraints for combined data analysis.

Epistasis or Gene–Gene Interactions

Given the evidence that all the genes noted above can have significant effects, we must consider how to evaluate the possible presence of epistasis, i.e., the interaction among genes (Rothman and Greenland, 1998). We are interested in whether the observed biological marker is better predicted by a "combination" of genotypes of interest at different loci when they are jointly present in the same subject(s) rather than when they act independently. In our current situation, the definition of the phenotype is qualitative, being either present or absent, and the synergistic effect of the two genes can be measured by estimating the odds ratio (OR) for the combinations of genotypes at different polymorphisms in the genes that we are evaluating (Kleinbaum, 1994). We can examine the joint effect of gene A and gene B by considering the genotypes of interest for polymorphisms at both genes. If the synergistic effect (calculated as the OR of the interaction in a logistic regression model) is significant, the next step is determining the value—and its corresponding meaning—for the OR that considers (gene A) × (gene B). If the OR that refers to the combination of the proposed genotypes has a value greater than the simple sum of the ORs of each single-risk factor (i.e., the implicated genotype for polymorphisms at gene A and the implicated genotype for polymorphisms at gene B), then the gene–gene interaction points to a multiplicative/epistatic, rather than to an additive, effect of the two polymorphisms. The multiplicative nature is evident from the observation that the OR for the interaction is higher than the additive effect of the single-risk genotypes, even at a logarithmic scale as in the logistic regression model. For a qualitative phenotype, the case–control association, or the TDT, must be represented in the form of a logistic regression (Sham, 1998) to allow for the interaction to be evaluated. Coding the genotypes for a given SNP at locus A, with alleles A and a, and another SNP at locus B, with alleles B and b, allows us to test different transmission hypotheses. For each biallelic SNP, we have three genotypes, say AA, Aa, and aa or BB, Bb, and bb and we can test a dominant-by-dominant interaction model,

constraining AA = Aa = 1 and aa = 0 for locus A and BB = Bb = 1 and bb = 0 for locus B. To test for a codominant or a recessive model, we simply adapt the coding procedure as AA = 1, Aa = 0.5, and aa = 0 for a codominant hypothesis and AA = 1 and Aa = aa = 0 for a recessive hypothesis (the same obviously is for locus B). With this simple coding, we can describe up to nine different models of synergy (e.g., dominant-by-dominant, dominant-by-co-dominant, dominant-by recessive, co-dominant-by dominant, codominant-by-co-dominant, etc.) and estimate the presence of a specific form of epistatic interaction (Goodnight, 2000). Considering the relatively simple occurrence that each locus (A and B) can have more than one SNP, whose number depends on the size of each locus, it is clear that the study of a gene–gene interaction rapidly escalates to evaluate tens if not hundreds of complex tests.

Recently a novel gene called G72 (on 13q32-34) has been associated in LD with schizophrenia in two populations (now further confirmed in at least five additional populations), with an OR of ± 2 (Chumakov et al., 2002). By following the G72 pathway to refine the candidate gene strategy, and examining the gene's biochemical partners in yeast, the gene for D-aminio acid oxidase (DAAO) (which acts jointly with G72) was also associated with schizophrenia in a previously unknown linkage region (12q21) with an OR of ± 1.8. Then, when the epistatic (gene–gene) interaction of G72 and DAAO was examined the OR associated to the risk of developing schizophrenia jumped to 5.2.

This example points out the complexity of gene structure, gene–gene interactions, and the need to study extended haplotypes rather than single SNPs or even multiple SNPs/haplotypes restricted to a single gene itself, and the necessity for developing a variety of statistical approaches to such problems. Hierarchical analyses, in which biological relationships have priority, and various nesting procedures based on biology need to be developed.

Endophenotypes and the Combination of Genetics and Imaging

Gottesman and Gould (2003) have discussed the endophenotype concept in neuropsychiatric illness. They defined a phenotype as observable characteristics of an organism that are the product of genetic and environmental influences. Endophenotype was defined as "measurable components unseen by the unaided eye along the pathway between disease and distal genotype." Other terms roughly synonymous include intermediate phenotype, biological marker, subclinical trait, and vulnerability marker. However, in their article they distinguished biological markers as biological characteristics that do not have genetic underpinnings, whereas endophenotypes have heritable characteristics, a convention that we follow, i.e., assuming that the definition of endophenotype includes a heritable component.

By this definition, examples of an endophenotype in schizophrenia include sensory gating abnormalities as measured by attenuated prepulse inhibition (e.g., Anokhin et al., 2003); eye tracking dysfunction as measured by deficiencies in smooth pursuit eye movements (e.g., Katsanis et al., 2000), and neurocognitive dysfunction of various types (e.g., Cannon et al., 2000; Tuulio-Henriksson et al., 2000), all of which have been shown to have heritable components. Gottesman and Gould concluded, "It stands to reason that more optimally reduced measures of neuropsychiatric functioning should be more useful than behavioral 'macros' in studies pursuing the biological and genetic components of psychiatric disorders." A working example of this outside the field of neuropsychiatry can be derived by studies on genetics of hypertension. In fact, despite the fact that hypertension is unequivocally a disease it is a heterogeneous disorder; the clinical phenotype (i.e., the trait) is easy to measure but

there is variability in the response to treatment. Using the amount of Na reabsorption in the distal tubule after a salt load during 48 h as the endophenotype, it is possible to easily identify a more homogeneous form of hypertension linked to a kidney defect, which allowed the clear identification of one gene mostly responsible for this particular subform of the disease (Cusi et al., 1997).

An endophenotype of the Gottesman and Gould type offers the advantage for psychiatric treatment of using a sample that is selected for a particular phenotype, with higher penetrance and more objective diagnosis (Adler et al., 1999). Such endophenotypes or intermediate phenotypes are more closely linked to the underlying genetic variation than clinical symptom clusters that may require a much higher threshold or combination of factors for expression. In order for the concept to be maximally useful, we extend the endophenotype concept in this article to include emergent combinations of imaging, and other neurophysiological measures with genetic information, and may well include specific clinical characteristics, to characterize homogeneous clusters of these measures within the schizophrenic population, leading to greater understanding of the disease characteristics and improve the timeliness and efficacy of treatment. The term carries a particularly applied characteristic, in that endophenotypes are clusters of phenotypic and/or genotypic measurements that allow improved differential diagnosis, improved predictability of disease progression, or improved predictability or understanding of treatment response, side effects, or pathophysiology. Gottesman's definition in contrast to our use would not include any clinical characteristic as part of the cluster constituting an endophenotype.

Some examples of quantifiable endophenotypes using linkage analysis in schizophrenic populations are P50 auditory-evoked potential gating (Freedman et al., 1997) and hippocampal *N*-acetyl asparate levels (Callicott et al., 1998). These are measures show distinctions between schizophrenics and controls; in conjunction with a genetic component they may be potential endophenotypes, leading to more targeted treatments. Chromosome 15q14, and specifically the α-7 nicotinic receptor gene localized to this region, show linkage to P50 auditory sensory gating deficit in schizophrenia (Freedman et al., 1997; Xu et al., 2001). Furthermore, De Luca et al. (2004) have shown the α-7 gene to be associated with the unusually high prevalence of smoking, and heavy smoking, in schizophrenia (for review of the α-7 nicotinic receptor literature regarding schizophrenia, *see* Martin et al., 2004). These and the others noted above are each examples of endophenotypes linked to only a single gene or haplotype; the ability to extend these measures to include epistatic interactions is a current research concern.

A practical example combining clinical outcomes with biological markers is that differential speed and completeness of clinical response to conventional antipsychotic drugs has been related to the interaction between dopamine metabolism and ventricular volume abnormalities in the brain (Garver et al., 2000). Genotypes were not included in this study and thus the heritability of these biological markers was not determined, though the combination of clinical and physiological markers had predictive value for treatment response. Spinks et al. (2004) studied the influence of the exonic polymorphism human opa-containing (HOPA), an X-chromosome gene involved in neuronal growth and differentiation, and the development of schizophrenia. The $HOPA^{12bp}$ allele appears to protect against the development of schizophrenia in these association studies performed in males; for those developing schizophrenia, it is also associated with fewer negative symptoms and better attentional processing than schizophrenics without this genotype, predicting both clinical course and cognitive dysfunction. For the phenotype of

antipsychotic medication response in schizophrenia, Mueller et al. (2003) have evidence that the SNAP-25 gene is involved both in prediction of treatment response and the side-effect of weight gain.

Environmental influences are important and interact with both genes and the displayed phenotype. A noteworthy example is the Caspi et al. (2003) study, in which a group of 400 boys was followed for 20 yr. Stressful events during the past 6 yr were measured and compared with their 5HTT transporter genotypes, studying a specific insertion/deletion polymorphism in the promoter region of the gene. Neither the 5HTT variants (e.g., the long vs the short polymorphism), nor the stress events considered individually predicted depression; however, the combination of the two measures predicted occurrence of depression (Caspi et al., 2003). This analysis highlights the value of combining data from seemingly disparate factors such as environment and genetic variation, and assaying their interaction.

Given the imaging phenotypes reviewed earlier, and the numerous candidate genes implicated in schizophrenia (*see also* Harrison and Weinberger, 2005), the most promising endophenotypes for differential diagnosis should integrate both information sources. Brain imaging usually requires much smaller samples than clinical symptomatology studies because of the greater precision, objectivity, and reliability of imaging measures, that translates into larger effective size. The combination of imaging with a candidate gene approach in case–control studies allows genetic correlations to be determined with dataset sizes on the order of 30 subjects. However, evaluating gene–gene interactions, already an increase in complexity over single-gene association studies, in combination with imaging studies will require considerably larger populations.

One of the most promising current endophenotypes to combine imaging as well as genetics is the combination of catecholamine-*O*-methyl-transferase (COMT) polymorphisms and working memory activations in neuroimaging studies. The catabolic enzyme for dopamine, COMT, the gene for which is on 22q11.2, has been linked to schizophrenia previously (Weinberger et al., 2001; Egan et al., 2002; Egan et al., 2001b). A functional polymorphism (val108met—a nucleotide substitution resulting in a valine-for-methionine change in the COMT protein) results in a fourfold increase in activity resulting in decreased available dopamine, but this increase is limited to brain areas that are relatively devoid of the dopamine transporter, such as the prefrontal cortex. This polymorphism has also been linked to performance on a working memory task (Egan et al., 2001a, b) with poor performance in schizophrenic patients and their siblings as well as comparison volunteers who had the valine variation. This valine variation is the same one found to be transmitted at a higher rate in at least one form of schizophrenia (for review, *see* Weinberger et al., 2001). Interestingly, these valine variation individuals are less efficient in their activation of the DLPFC as measured by functional magnetic resonance imaging (fMRI). Gottesman and Gould (2003), in reviewing these data, concluded that fMRI analysis of subjects undergoing working memory tasks may be a more sensitive endophenotype than purely neuropsychological measures of working memory performance.

This example highlights some of the complexities of combining genetic with imaging data. The brain, unlike many other organs in the body, comprises numerous and variable compartments, each containing multiple cell types and functional subunits which are differentially expressed throughout the various regions of the brain. For example, neurons in the frontal cortex are relatively devoid of the dopamine transporters, unlike neurons elsewhere, even though dopamine has important functions in the frontal cortex. This absence of dopamine transporters has the important functional consequences noted above when

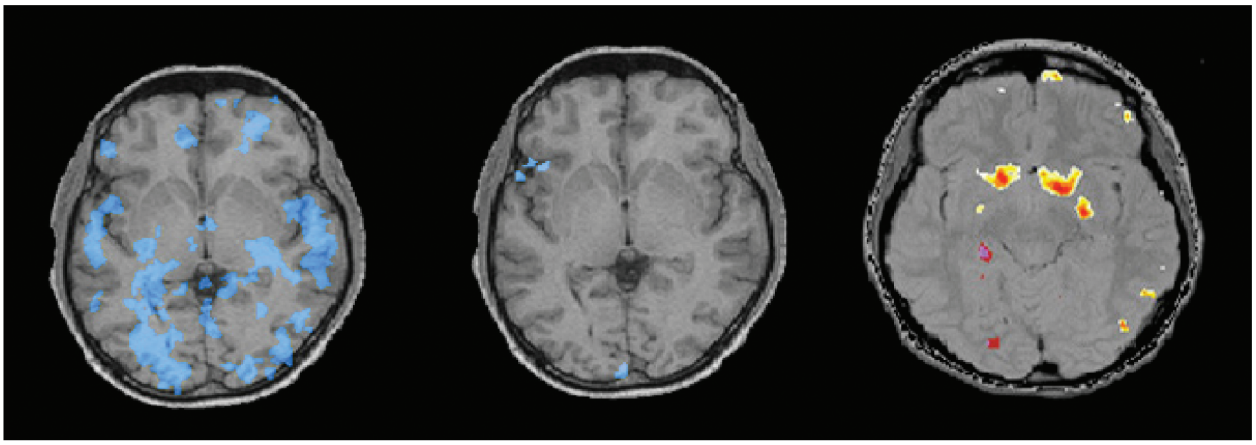


Fig. 1. Combined imaging and genetics data to reveal endophenotypes: **(Left panel)** The regional glucose metabolic changes following clozapine for schizophrenic subjects with the 2,2 alleles for the DRD1 gene. (**Middle panel)** The same changes for the schizophrenic subjects with the 1,2 alleles DRD1. Following clozapine treatment, the 2,2 allele subjects respond clinically and alter their brain metabolic response, whereas 1,2 subjects do not respond clinically or metabolically to clozapine. **(Right panel)** The difference between brain metabolic response in subjects homozygous for the DRD3 glycine allele and those with glycine9serine or serine9serine alleles. The glycine9glycine DRD3 allele subjects who are at increased risk for tardive dyskinesia increase their ventral striatal metabolic response following haloperidol, whereas the other subjects (without the glycine9glycine allele) do not. (Reprinted with permission from Basile et al., 2002; Potkin et al., 2003).

combined with COMT alleles that are relatively unable to metabolize dopamine. In other areas of the brain, the increased dopamine is removed from the synapse by the transporter and thus has little effect on either cognitive function or neuroimaging results. Thus genotypes are differentially expressed throughout the brain and need not be treated as simple categorical variables.

FDG-PET also has the ability to measure regional brain metabolic response whereas the subject is performing a cognitive task. We have shown that dopamine D_1 receptor gene alleles at the DdeI SNP in the first intron are associated with clinical response as well as brain metabolic changes during clozapine treatment (Potkin et al., 2003b). D_3 receptor alleles (gly9gly) have been shown by our group and others to be associated with risk of tardive dyskinesia (Basile et al., 1999; Lerer et al., 2002). Interestingly, the gly9gly subjects dramatically increase their striatal metabolism when treated with the typical antipsychotic haloperidol, whereas the ser9ser or gly9ser subjects do not increase striatal metabolism in response to haloperidol (*see* Fig. 1) (Basile et al., 2002; Potkin et al., 2003). These studies combine genetics and brain imaging to provide new information on how genetic risk influences the production of clinical response to medication as well as the production of side effects, which is critical to developing targeted treatments for schizophrenia.

Statistical Issues in Imaging Genetics

In the above examples, candidate genes are usually chosen *a priori* and used as categorical variables in the analysis of imaging data. Useful information has come out of this approach in both healthy individuals and clinical populations, as demonstrated earlier, as well as in clinical populations such as Alzheimer's disease research with apolipoprotein E (APOE) alleles

(e.g., Bookheimer et al., 2000; Small et al., 2000; Small, 2002). Beyond the well-documented issues with case–control studies in any genetic research (e.g., Cardon and Bell, 2001), there are at least three limitations with this standard approach: The first and critical issue is that candidate genes, by their nature, limit the hypotheses to known genes; they do not facilitate discovery of essential genetic contributions of genes whose function are currently unknown. The second issue is the need for adjustment for the number of candidate genes explored in genetic association studies. The third issue involves the complexity of visualizing both imaging and genetic data for data analysis.

To avoid the risk of inflating false positives in genetic association and imaging studies, we should consider the number of tests performed and correct accordingly. At present, there is no general consensus about how to "correct" for multiple testing in genetic association studies (Krawczak et al., 2001; Nyholt, 2001) or within imaging studies (e.g., Marchini and Presanis, 2004). Despite different opinions, all authors agree that a conservative correction is meaningless in these datasets, and they suggest some kind of "adjustment" in which tests involve data that are independent from each other. The frequently advocated Bonferroni adjustment is not appropriate (Perneger, 1998) because it assumes that all null hypotheses are true simultaneously and independently, which is rarely the case in biological studies (e.g., genetic variation is not necessarily independent, and temporal correlations in dynamic imaging studies can be significant (Purdon and Weisskoff, 1998). The Bonferroni and similar corrections reduce the potential number of false-positive errors, but with the trade-off of dramatically increasing the number of false-negatives; consequently, truly important differences can be easily missed. Strategies that involve split sample or independent sample confirmations mitigate the possibility of both false-positives and false-negatives.

Other methods to correct for multiple testing have been proposed and are reviewed by Brown and Russell (1997), and Cribbie and Kaselman (2003) with a specific focus on genetic associations studies. Similarly in the analysis of brain imaging data, Nichols and Hayasaka (2003) and Marchini and Presanis (2004) review several approaches to the multiple-testing concerns, including nonparametric permutation tests and posterior probability approaches, which are being applied with growing success to imaging studies (e.g., Friston and Penny, 2003; Penny et al., 2003, 2005). The false discovery rate (FDR) procedure, the ratio between the number of false-positives and the number of significant features, is an approach that is currently increasing in popularity. The FDR emphasizes "positive" findings, introducing a correction based on the expected relationship of the sensitivity and sensibility of the test; it has already been proposed as an ideal correction for genome-wide genetic association studies (Benjamini and Hochberg, 1995; Devlin et al., 2003; Storey and Tibshirani, 2003) and has been included in fMRI imaging analysis methods (Genovese et al., 2002). These are applications of FDR to genetic or imaging studies individually; the combination of multiple SNP and brain voxel data will undoubtedly continue to bring new challenges to drive developments in analytic methods.

The third problem in the integration of imaging and genetics involves the combination of such diverse data types in a mathematically valid yet comprehensible way: both genetic and imaging data have a spatial component, for example, but the spatial manifolds, in which each is embedded, are fundamentally different. Both data types are four-dimensional (4D), with internal interactions over time, but they are not represented as such. Imaging data have an inherent three-dimensional (3D) structure, such that voxels near each other in the brain are not independent of each other, although the visualization is of necessity usually a two-dimensional (2D) projection.

Unlike in brain images, the 3D nature of chromosomes is not critical to genetic analyses; genetic data can be represented in 2D space, with the fundamental data thought of as interacting points on many one-dimensional lines. Combining these datasets whereas retaining and utilizing the spatial and temporal information to extract the relevant patterns is a current challenge.

The Role of Data Mining in Analyzing Combined Imaging–Genetics Data

There has been considerable research activity by computer scientists in recent years in the area of data mining (also sometimes referred to as "knowledge discovery in databases," and closely associated with a related research area known as "machine learning"). Broadly speaking, the goal of data mining research is to develop algorithms that can automatically search through data sets to detect patterns that could otherwise go undetected by manual analysis methods (Hand et al., 2001; Smyth et al., 2002). This can be particularly useful in situations in which there are a very large number of measured variables and/or there are a very large number of data points. The interface between data mining and statistics is by necessity somewhat blurred: many data mining techniques have statistical ideas at their core, and there are well-known statistical techniques (in particular in the areas of exploratory data analysis and discriminant analysis) that could be viewed as a type of data mining (Smyth, 2000; Hastie et al., 2001). Many of the successful applications of data mining have occurred in (largely unpublished) commercial applications such as credit-scoring and e-commerce (Berry and Linoff, 2004) in which the very large number of data points available makes it possible to reliably fit relatively complex predictive models for classification and regression with minimal prior assumptions on the functional forms of these models (e.g., Friedman, 2001; Scholkopf and Smola, 2002). High-dimensional problems, with hundreds or thousands of variables, are often addressed by techniques such as tree-based algorithms (Breiman, 2001; Hastie et al., 2001) that search for relatively small subsets of variables (often with nonlinear interactions) that appear empirically useful for the task at hand (whether regression, classification, clustering, etc.). These "dimension-reduction" ideas provide a useful general approach toward combating the problem of "many variables, few samples," and although they are generally useful for high-dimensional problems they do not provide a universal panacea to the fundamental statistical limits governing reliable inference from small samples.

In the area of bioinformatics, data mining and machine learning methods have been successfully applied in succession to problems in protein and DNA/RNA sequence modeling (Durbin et al., 1999; Baldi and Brunak, 2001; Pevzner, 2001), analysis of gene expression microarray data (Kohane et al., 2003; Baldi and Hatfield, 2002; Speed, 2003; McLachlan et al., 2004), and broader problems involving pattern discovery from genomic data (Wang et al., 1999; Scholkopf et al., 2004). However, in the context of the imaging and genetics, the sizes of data sets that are currently available (in terms of number of subjects) are relatively low from a data mining viewpoint. For example, in analyzing web data it is common to have data from millions of subjects to analyze (e.g., Cadez et al., 2003), whereas fMRI studies, for example, are typically only available for order of 100 individuals or fewer (even without a genetic component). The numbers of subjects in these studies will increase as projects involving federated data sets scale up. In the meantime, given that the numbers of imaging (X), genetic (Y), and clinical (Z) variables are often greater than the number of subjects being analyzed, data mining for such data seems best suited to an exploratory (rather than a confirmatory) role, as an aid to direct visualization, suggesting candidate hypotheses for consideration and possible follow-up confirmatory statistical analysis.

Clustering is a particularly popular technique for data mining and exploratory data analysis, in which an algorithm searches for natural groupings of individuals or objects (Jain and Dubes, 1988). Clustering offers the possibility of automated discovery of subgroups that cannot be visualized manually in the high-dimensional combined imaging-genetics space of measurements. Better groupings are those in which the objects within each group are more similar to each other in measurement space, than to objects in other groups. In microarray data analysis, for example automatic clustering of genes can provide a useful starting point for detecting groups of genes whose expression patterns are similar and whose functional roles may be related—a commonly applied approach is to search for a hierarchy of clusters, using Euclidean distance between expression profiles to merge genes into clusters in a bottom-up manner (Eisen et al., 1998; Spellman et al., 1998).

An alternative approach in this context is probabilistic model-based clustering, which hypothesizes the existence of a finite mixture model with K components for modeling the underlying probability distribution (or density) of the observed data, i.e., $p(X,Y) = \Sigma_k p(X,Y \mid C = k)$ where C is a cluster variable. The imaging data $X = [x_1, x_2, \ldots,]$ that would play a role in these clusters could be of various sorts. Neuroimaging methods such as fMRI or PET collect images of the subject's brain either during or following a cognitive task (*see* e.g., Toga and Mazziotta, 2000). These images are usually 3D volumes (made up of 2D images or "slices"), with 3D volumetric elements or voxels being the elements in the image over which measurements are made and statistics are computed. The analysis results are thus not a simple summary statistic, but a 3D brain-shaped arrangement of statistics, one z-score or t-score or other statistic for each 3D voxel in the brain image, for each analysis performed.

At its most computationally demanding, X could be the summary statistic for each voxel x_i in the brain. More realistically, X could be the summary statistic for predefined regions in the brain; for example, in a working memory task the strengths of the cortical activation in a number of relevant frontal and parietal regions x_1, x_2, …., may be represented as the vector X·X could also include the strength of the connection between these areas under the condition of the particular cognitive task, based on a circuitry analysis such as dynamic causal modeling (Mechelli et al., 2003) or structural equation modeling as applied to neuroimaging data (e.g., Kilpatrick and Cahill, 2003).

The genetic data $Y = [y_1, y_2, \ldots,]$ could also represent several different pieces of information. At the lowest and again most computationally demanding level, Y would be the full genomic scan for the person, all base-pairs in order by chromosome. This is more information than the current approaches are using. At its simplest, using SNP data, y_i would simply code a particular SNP type; for example, with the gene APOE, there are three common alleles (coded as 2–4). A person will have two alleles, which can be any unordered combination of the three alleles (2,2; 2,3; 2,4;3,3; 3,4;4,4). If y_1 codes the APOE genotype, it could simply be coded using 1–6 to indicate which of the combinations the subject has. Y_2 would then be a similar representation for another gene such as BDNF or another; y_3 similarly; and so on. A different representation would be a haplotype coding similar to that described in the haplotype blocking section. Thus y_1 would not indicate a particular allele combination but would indicate different haplotypes. This is obviously not the only way to represent these data; the known functional groupings of genes are not explicitly represented, for example, and the best way to incorporate that *a priori* information into the representation is an area for development.

Given these datasets, each of the K component densities, $P(X,Y|C = k)$ in the mixture model corresponds to a particular cluster

(McLachlan and Peel, 2002). Scale and correlation effects can be handled directly, for example, modeled by the covariance matrix for each component in the multivariate Gaussian mixture model. The parameters of this model, namely the parameters for each of the K component densities and the K mixing weights, can be estimated from the data—in particular, a general-purpose estimation method such as the expectation-maximization procedure, is often used in practice. A feature of the model-based approach is that it inherits aspects of both statistical and data mining "thinking": it is grounded in a statistical modeling framework, yet provides a relatively flexible form of model fitting, consistent with the general data mining philosophy of making relatively few assumptions about the data before analysis.

When different data types are present in a data set, such as the combination of imaging and genetics data, it can be difficult to define a suitable distance metric among subjects: how much weight should the genetic data Y get relative to the imaging data X for example? How should replicated data (e.g., multiple images) be handled? Should the measurements be averaged or should the variability in the replications be taken into account? The probabilistic clustering approach provides a starting point to address some of these issues by allowing the data analyst to include different data types and replicated data in a systematic manner. Different data types, for example, can be modeled by hypothesizing parsimonious but realistic conditional independence relations among multiple data modalities and measurements, for example, by assuming conditional independence of imaging and genetic variables given cluster membership, $p(X,Y|C) = p(X|C)\,p(Y|C)$. There is also no need to define *a priori* an explicit distance measure among diverse measurements sets; instead, the statistical learning algorithm in effect can estimate an appropriate implicit distance measure among objects based on their relative likelihoods among different cluster assignments (e.g., Lin et al., 2004).

Other data mining techniques that are potentially relevant to imaging-genetics data include the general class of "dependency modeling." Broadly speaking, the representational language underlying these methods can be thought of as follows: each measurement variable is represented as a node in a graph, and edges in the graph (links among nodes) represent direct dependencies. A data mining algorithm in this context then searches for a sparse set of links that explain the observed dependencies well. For example, given three variables A, B, and C, if B and C are conditionally independent given A, then the ideal directed graph for this probability model consists of a directed edge from A to B and another directed edge from A to C. The ideal model contains no edge between B and C as they do not directly depend on each other but are marginally dependent (if the value of variable A is unknown).

For directed links, there is a well-developed theory for such models under the name of directed graphical models or Bayesian networks (Lauritzen and Spiegelhalter, 1988; Pearl, 1988). There is an accompanying theoretical framework and algorithms for learning the structure of these models from data, in which search algorithms balance goodness-of-fit and model complexity (Jordan, 1999). A recent success of this approach is in automated learning of dependency graphs for gene regulation (Segal et al., 2003), integrating both protein interaction data and microarray measurements using a probabilistic modeling framework. This type of integration of measurements from different sources is somewhat analogous to the problem of analyzing combined genetics, imaging, and clinical data, and probabilistic modeling techniques are likely to play a useful role in this context, allowing the linking together of different information sources using the calculus of probability. The framework of directed graphical models provides a useful starting

point for constructing and computing with complex probabilistic models, with a variety of extensions to include temporal dependencies (such as dynamic Bayesian networks, [Smyth et al., 1998; Murphy 2002]) and spatial information (such as undirected Markov random fields, Winkler, 2003).

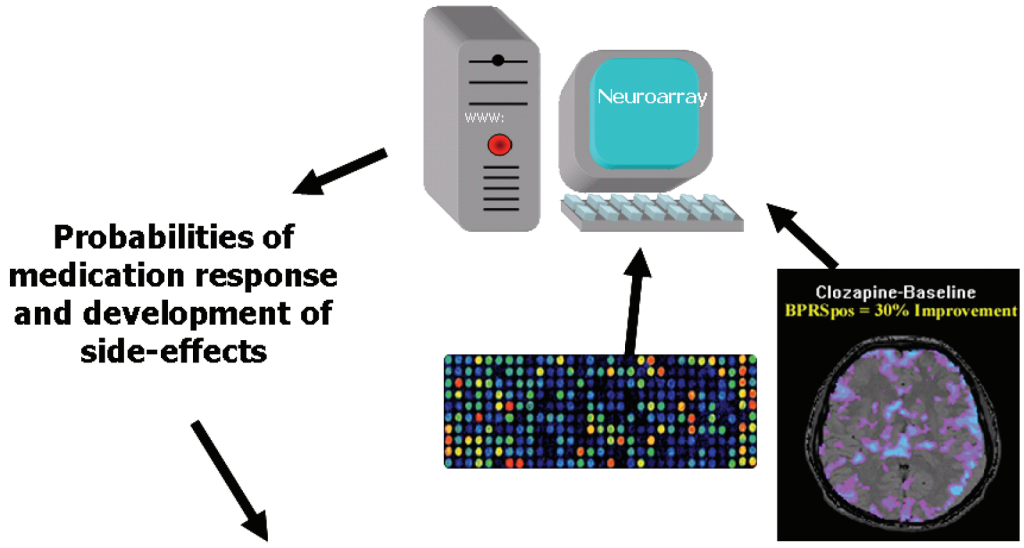
Closer to the theme of this present article is the work of Herskovits and colleagues who have shown how graphical models can be learned that relate structural properties of brain images (estimated lesion volumes, X variables) with clinical function (attention-deficit hyperactivity disorder, Z variables) (Herskovits and Gerring, 2003; Herskovits et al., 2004). In this work it was shown that learning a structured graphical model for $p(Z|X)$, detected nonlinear dependencies between X and Z variables that were undetected by more traditional methods such as t-tests. The utility of being able to automatically learn such dependencies from data is particularly appealing when genetics data is added to the mix, although such studies are still only on the horizon as imaging-genetics datasets scale up to appropriate sizes. Another useful feature of the graphical model approach, as illustrated in Herskovits and Gerring (2003), is the ability to separate variables into different groups that reflect known causal structure and then learn dependencies that are consistent with this structure (e.g., from brain structure X to individual clinical behavior Z, rather than vice versa).

The datasets currently available for imaging-genetic data analysis tend to be somewhat small to support the powerful search-based data mining algorithms that have been developed in the computer science community in recent years. However, as mentioned earlier, as more large-scale studies are performed, and statistical and algorithmic techniques are improved to allow better "pooling" of imaging-genetics data sets collected across different sites and studies, we can expect increased interest and utilization of data mining algorithms, as well the development of new data mining algorithms that leverage the unique aspects of combined imaging and genetic data.

Summary

The understanding of schizophrenia has progressed greatly through the use of neuroimaging studies, genetic studies, and their combination. Numerous physiological distinctions have been identified between schizophrenics and controls, and among schizophrenics of differing clinical profiles. The integration of these approaches into clinically meaningful endophenotypes is the promise of the field of imaging genetics. Currently, although there is a wealth of information regarding different phenotypes and genotypes related to schizophrenia, the choice of treatment for the disease is driven primarily by the patient's subjectively assessed symptoms. The integration of neuroimaging tests and genotyping for each individual can and should be aimed toward identifying the most effective treatment, with the probability of side-effects for that individual also assessed (*see* Fig. 2).

To do this, stable and useful endophenotypes need to be assessed from within large, multimodal datasets. The amount and variety of brain imaging information will continue to expand as new methods such as magnetoencephalography (MEG), diffusion tensor imaging (DTI), and optical imaging become more wide-spread, as the resolution of imaging techniques improve, and the type and complexity of image and pattern analysis and visualization advance. Similarly, the amount of genetic information (SNP, expansions) will soon exceed one million measures of variation per individual. There may be as many as 12 million locations of common variation in the human genome, although they may described by as few as 300,000 haplotypes, again highlighting the promise and complexity of combining these types of data. Quality assurance, quality control, data provenance, and the confidentiality of such amounts of information per subject are challenging. Combining these



Clinical Predictions based on endophenotype

	Efficacy	Negative	Cognitive	DM	Weight	Suicide
Clozapine	90	80	25	50	85	2
Asenapine	90	80	50	10	15	?
Olanzapine	80	70	20	70	90	4
Ziprasidone	85	75	30	20	10	?

Fig. 2. The promise of endophenotypes for improved treatment. Ideally, a patient's genotype and imaging phenotype can be measured as needed, to allow better, individualized predictions of treatment response, both in control of clinical symptoms (e.g., efficacy and negative symptoms, and cognitive function) and side effects such as the development of diabetes mellitus, weight gain, and increased risk of suicide.

large datasets of different data types compounds the problem.

The promise of neuroimaging in enhancing understanding of schizophrenia and in facilitating the development of new treatments has been hampered by the difficulties in comparing and combining imaging data obtained at separate centers, which has constrained investigators from creating the datasets of the size needed for the datamining methods. Different magnet strengths, manufactures, acquisition sequences, and cognitive tasks and paradigms are important aspects of the inherent complexities. National Institutes of Health (NIH) is funding several projects to address these obstacles; the functional imaging research in schizophrenia testbed Biomedical Informatics Research Network (BIRN) project (aka FBIRN) is a transdisciplinary consortium focused on developing the technical infrastructure for combining functional MRI data obtained at multiple sites using diverse imaging equipment, different magnet strengths, and a range of data acquisition methodologies. Anatomical and functional calibration methods have been developed by BIRN to facilitate combining imaging data across sites and over time. FBIRN also provides a clinical federated and distributed database schema, allowing the tracking of demographic, symptomatic, and neurocognitive information. Databases like these and (brain imaging database [BRAID] [Letovsky et al.,

1998; Herskovits, 2000]), in conjunction with those being developed to share medical images for surgical planning (e.g., Cao et al., 2004), form the basis for the broader data sharing and combination that is required for the discovery of endophenotypes.

Tools for integrating these imaging, genetic, and clinical datasets are required, as well as the development of new methods of visualization that take advantage of the human visual system's ability to scale and integrate data. Better methods of pattern recognition are clearly needed. The successful integration of these methods depends on a robust, flexible, and distributed database and IT infrastructure. The transdisciplinary collaborative efforts of neuroscientists, computer scientists, genetics, and clinical investigators are required for the proposed successful integration of these methods in support of new knowledge discovery.

Acknowledgments

The authors gratefully acknowledge the input of the investigators involved in the following consortia and the support of the following grants: the Functional Imaging Research in Schizophrenia Testbed, Biomedical Informatics Research Network (FIRST BIRN; M01RR00827, www.nbirn.net), and the Morphometry Biomedical Informatics Research Network (U24RR021382-01), funded by the National Center for Research Resources at the National Institutes of Health; and through Transdisciplinary Imaging Genetics Center (P20RR020837-01) and the National Alliance for Medical Image Computing (NAMIC; Grant U54 EB005149), funded by the National Institutes of Health through the NIH Roadmap for Medical Research,. Information on the National Centers for Biomedical Computing can be obtained from http://niharoadmap.nih.gov/bioinformatics.

References

Adler, L. E., Freedman, R., Ross, R. G., et al. (1999) Elementary phenotypes in the neurobiological and genetic study of schizophrenia. Biol. Psychiatr. 46(1), 8–18.

Anokhin, A. P., Heath, A. C., Myers, E., et al. (2003) Genetic influences on prepulse inhibition of startle reflex in humans. Neurosci. Lett. 353(1), 45–48.

Baldi, P. and Brunak, S. (2001) Bioinformatics: the Machine Learning Approach. MIT Press: Cambridge, MA, 2nd edition.

Baldi, P. and Hatfield, G. W. (2002) DNA Microarrays and Gene Expression: From Experiments to Data Analysis and Modeling. Cambridge University Press: New York, NY.

Barbeau, D., Liang, J. J., Robitalille, Y., et al. (1995) Decreased expression of the embryonic form of the neural cell adhesion molecule in schizophrenic brains. Proc. Natl. Acad. Sci. USA 92(7), 2785–2789.

Basile, V. S., Masellis, M., Badri, F., et al. (1999) Association of the MscI polymorphism of the dopamine D3 receptor gene with tardive dyskinesia in schizophrenia. Neuropsychopharmacology 21(1), 17–27.

Basile, V. S., Masellis, M., Potkin, S. G., and Kennedy, J. L. (2002) Pharmacogenomics in schizophrenia: the quest for individualized therapy. Hum. Mol. Genet. 11(20), 2517–2730.

Bechara, A., Damasio, A. R., Damasio, H., and Anderson, S. W. (1994) Insensitivity to future consequences following damage to human prefrontal cortex. Cognition 50(1,3), 7–15.

Benjamini and Hochberg (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. B. 57, 289–300.

Berry, M. and Linoff, G. (2004) Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management, Wiley.

Bookheimer, S. Y., Strojwas, M. H., Cohen, M. S., et al. (2000) Patterns of brain activation in people at risk for Alzheimer's disease. N. Engl. J. Med. 343(7), 450–456.

Botstein, D. and Risch, N. (2003) Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat. Genet. 33 Suppl, 228–237.

Breiman, L. (2001) Random forests. Machine Learning 45(1), 5–32.

Brown, B. W. and Russell, K. (1997) Methods of correcting for multiple testing: operating characteristics. Stat. Med. 16(22), 2511–2528.

Buchsbaum, M. S., Shihabuddin, L., Hazlett, E. A., et al. (2002) Kraepelinian and non-Kraepelinian schizophrenia subgroup differences in cerebral metabolic rate. Schizophr. Res. 55(1,2), 25–40.

Bunney, B. G., Potkin, S. G., and Bunney, W. E. (1997) Neuropathological studies of brain tissue in schizophrenia. J. Psychiatr. Res. 31(2), 159–173.

Cadez, I., Heckerman, D., Meek, C., et al. (2003) Visualization of navigation patterns on a Web site using model-based clustering. J. Data Mining Knowledge Discovery 7(4), 399–424.

Callicott, J. H., Egan, M. F., Bertolino, A., et al. (1998) Hippocampal *N*-acetyl aspartate in unaffected siblings of patients with schizophrenia: a possible intermediate neurobiological phenotype. Biol. Psychiatr. 44(10), 941–950.

Cannon, T. D., Huttunen, M. O., Lonnqvist, J., et al. (2000) The inheritance of neuropsychological dysfunction in twins discordant for schizophrenia. Am. J. Hum. Genet. 67(2), 369–382.

Cao, X., Wong, S. T., Hoo, K. S., Jr., et al. (2004) A web-based federated neuroinformatics model for surgical planning and clinical research applications in epilepsy. Neuroinformatics 2(1), 101–118.

Cardon, L. R. and Bell, J. I. (2001) Association study designs for complex diseases. Nat. Rev. Genet. 2, 91–99.

Carlson, C. S., Eberle, M. A., Rieder, M. J., et al. (2003) Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. Nat. Genet. 33, 518–521.

Caspi, A., Sugden, K., Moffitt, T. E., et al. (2003) Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. Science 301(5631), 386–389.

Chakravarty, A. (1999) Population genetics-making sense out of sequence. Nat. Genet. 21, 56–60.

Chumakov, I., Blumenfeld, M., Guerassimenko, O., et al. (2002) Genetic and physiological data implicating the new human gene G72 and the gene for D-amino acid oxidase in schizophrenia. Proc. Natl. Acad. Sci. USA 99(21), 13,675–13,680.

Chung, C., Tallerico, T., Seeman, P. (2003) Schizophrenia hippocampus has elevated expression of chondrex glycoprotein gene. Synapse. Oct;50(1):29–34.

Cotter, D., Kerwin, R., al-Sarraji, S., et al. (1998) Abnormalities of WNT signalling in schizophrenia—evidence for neurodevelopmental abnormality. Neuroreport 9(7), 1379–1383.

Cribbie, R. A. and Keselman, H. J. (2003) Pairwise multiple comparisons: a model comparison approach versus stepwise procedures. Br. J. Math Stat. Psychol. 56(Pt. 1), 167–182.

Cusi, D., Barlassina, C., Azzani, T., et al. (1997) Polymorphisms of alpha-adducin and salt sensitivity in patients with essential hypertension. Lancet 349, 1353–1357.

Daly, M. J. (2002) Estimating the human gene count. Cell 109(3), 283, 284.

Davidson, S. (2000) Research suggests importance of haploytpes over SNPs, Nat. Biotechnol. 18, 1134, 1135.

Davis, K. L. and Haroutunian, V. (2003) Global expression-profiling studies and oligodendrocyte dysfunction in schizophrenia and bipolar disorder. Lancet 362(9386), 758.

Davis, K. L., Stewart, D. G., Friedman, J. I., et al. (2003) White matter changes in schizophrenia: evidence for myelin-related dysfunction. Arch. Gen. Psychiatr. 60(5), 443–456.

De Luca, V., Wong, A. H., Muller, D. J., et al. (2004) Evidence of association between smoking and alpha7 nicotinic receptor subunit gene in schizophrenia patients. Neuropsychopharmacology 29(8), 1522–1526.

DeSilva, U., D'Arcangelo, G., Braden, V. V., et al. (1997) The human reelin gene: isolation, sequencing, and mapping on chromosome 7. Genome Res. 7(2), 157–164.

Devlin, B., Roeder, K. (1999) Genomic control for association studies. Biometrics 55, 997–1004.

Devlin, B., Roeder, K., and Wasserman, L. (2003) Analysis of multilocus models of association. Genet. Epidemiol. 25(1), 36–47.

Doherty, P., Fruns, M., Seaton, P., et al. (1990) A threshold effect of the major isoforms of NCAM on neurite outgrowth. Nature 343(6257), 464–466.

Drysdale, C. M., McGraw, D. W., Stack, C. B., et al. (2000) Complex promoter and coding region b2-adrenergicreceptor haplotypes alter receptor expression and predict in vivo responsiveness. PNAS, 97, 10,483–10,488.

Durany, N., Michel, T., Zochling, R., et al. (2001) Brain-derived neurotrophic factor and neurotrophin 3 in schizophrenic psychoses. Schizophr. Res. 52(1,2), 79–86.

Egan, M., Goldman, D., and Weinberger, D. (2002) The human genome: mutations. Am. J. Psychiatr. 159(1), 12.

Egan, M. F., Goldberg, T. E., Gscheidle, T., et al. (2001a) Relative risk for cognitive impairments in siblings of patients with schizophrenia. Biol. Psychiatr. 50(2), 98–107.

Egan, M. F., Goldberg, T. E., Kolachana, B. S., et al. (2001b) Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for

schizophrenia. Proc. Natl. Acad. Sci. USA 98(12), 6917–6922.
Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA 95(25), 14,863–14,868.
Fallon, J. H., Opole, I. O., and Potkin, S. G. (2003) Neuroanatomy of schizophrenia: circuitry and neurotransmitter systems. Clin. Neurosci. Res. 3(1,2), 77–107.
Fan, J. B., Tang, J. X., Gu, N. F., et al. (2002) A family-based and case-control association study of the NOTCH4 gene and schizophrenia. Mol. Psychiatr. 7(1), 100–103.
Fatemi, S. H., Earle, J. A., and McMenomy, T. (2000) Reduction in Reelin immunoreactivity in hippocampus of subjects with schizophrenia, bipolar disorder and major depression. Mol. Psychiatr. 5(6), 654–663, 571.
Freedman, R., Coon, H., Myles-Worsley, M., et al. (1997) Linkage of a neurophysiological deficit in schizophrenia to a chromosome 15 locus. Proc. Natl. Acad. Sci. USA 94(2), 587–592.
Friedman, J. (2001) Greedy function approximation: a gradient boosting machine. Ann. Stat. 29(5), 1189–1232.
Friston, K. J. and Penny, W. (2003) Posterior probability maps and SPMs. Neuroimage 19(3), 1240–1249.
Gabriel, S. B., Schaffner, S. F., Nguyen, H., et al. (2002) The structure of haplotype blocks in the human genome. Science 296, 2225–2229.
Garver, D. L., Holcomb, J. A., and Christensen, J. D. (2000) Heterogeneity of response to antipsychotics from multiple disorders in the schizophrenia spectrum. J. Clin. Psychiatr. 61(12), 964–972, quiz 973.
Genovese, C. R., Lazar, N. A., and Nichols, T. (2002) Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage 15(4), 870–878.
Goodnight, C. (2000) Modeling gene interaction in structure populations. In: Wolf, J., E. B. III and Wade, M. (eds.) Epistasis and the Evolutionary Process, Oxford University Press, New York, NY, pp. 129–145.
Gottesman, II. and Gould, T. D. (2003) The endophenotype concept in psychiatry: etymology and strategic intentions. Am. J. Psychiatr. 160(4), 636–645.
Guidotti, A., Auta, J., Davis, J. M., et al. (2000) Decrease in reelin and glutamic acid decarboxylase67 (GAD67) expression in schizophrenia and bipolar disorder: a postmortem brain study. Arch. Gen. Psychiatr. 57(11), 1061–1069.
Gur, R. E., Mozley, P. D., Resnick, S. M., et al. (1995) Resting cerebral glucose metabolism in first-episode and previously treated patients with schizophrenia relates to clinical features. Arch. Gen. Psychiatr. 52(8), 657–667.
Hakak, Y., Walker, J. R., Li, C., et al. (2001) Genome-wide expression analysis reveals dysregulation of myelination-related genes in chronic schizophrenia. Proc. Natl. Acad. Sci. USA 98(8), 4746–4751.
Hand, D., Mannila, H., and Smyth, P. (2001) Principles of Data Mining. Boston, MIT Press.
Harrison, P. J. and Eastwood, S. L. (1998) Preferential involvement of excitatory neurons in medial temporal lobe in schizophrenia. Lancet, 352(9141), 1669–1673.
Harrison, P. J. and Weinberger, D. R. (2005) Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence. Mol. Psychiatr. 10(1), 40–68, image 5.
Hastie, T., Tibshirani, R., and Friedman, J. (2001) The elements of statistical learning: data mining, inference, and prediction, Springer-Verlag, Berlin.
Hawi, Z., Straub, R. E., O'Neill, A., et al. (1998) No linkage or linkage disequilibrium between brain-derived neurotrophic factor (BDNF) dinucleotide repeat polymorphism and schizophrenia in Irish families. Psychiatr. Res. 81(2), 111–116.
Heberlein, A. S., Adolphs, R., Tranel, D., and Damasio, H. (2004) Cortical regions for judgments of emotions and personality traits from point-light walkers. J. Cogn. Neurosci. 16(7), 1143–1158.
Heckers, S., Goff, D., Schacter, D. L., et al. (1999) Functional imaging of memory retrieval in deficit vs nondeficit schizophrenia. Arch. Gen. Psychiatr. 56(12), 1117–1123.
Herskovits, E. H. (2000) An architecture for a brain-image database. Methods Inf. Med. 39(4–5), 291–297.
Herskovits, E. H. and Gerring, J. P. (2003) Application of a data-mining method based on Bayesian networks to lesion-deficit analysis. Neuroimage 19(4), 1664–1673.
Herskovits, E. H., Peng, H., and Davatzikos, C. (2004) A Bayesian morphometry algorithm. IEEE Trans. Med. Imaging 23(6), 723–737.
Hodgkinson, C. A., Goldman, D., Jaeger, J., et al. (2004) Disrupted in schizophrenia 1 (DISC1): association with schizophrenia, schizoaffective disorder, and bipolar disorder. Am. J. Hum. Genet. Nov;75(5), 862–872.

Hohe, M. R, Köple, K., Wendel, B., Rohde, K., et al. (2000) Sequence variability and candidate gene analysis in complex disease: association of μ opioid receptor gene variation with substance dependence. Hum. Mol. Genet. 19, 2895–2908.

Honer, W. G., Falkai, P., Young, C., et al. (1997) Cingulate cortex synaptic terminal proteins and neural cell adhesion molecule in schizophrenia. Neuroscience 78(1), 99–110.

Hynes, M. and Rosenthal, A. (1999) Specification of dopaminergic and serotonergic neurons in the vertebrate CNS. Curr. Opin. Neurobiol. 9(1), 26–36.

Jain, A. and Dubes, R. (1988) Algorithms for clustering data. Englewood Cliffs, NJ, Prentice-Hall.

Jeste, D. V., Kleinman, J. E., Potkin, S. G., et al. (1982) Ex uno multi: subtyping the schizophrenic syndrome. Biol. Psychiatr. 17(2), 199–222.

Joosten, P. H. L. J., Toepel, M., Mariman, E. C. M., et al. (2001) Promoter haplotype combinations of the platelet-derived growth factor alpha-receptor gene predispose to human neural tube defects. Nat. Genet. 27, 215–217.

Jordan, M. I. (ed.) (1999) Learning in Graphical Models, Boston, MIT Press.

Karson, C. N., Mrak, R. E., Schluterman, K. O., et al. (1999) Alterations in synaptic proteins and their encoding mRNAs in prefrontal cortex in schizophrenia: a possible neurochemical basis for "hypofrontality". Mol. Psychiatr. 4(1), 39–45.

Katsanis, J., Taylor, J., Iacono, W. G., and Hammer, M. A. (2000) Heritability of different measures of smooth pursuit eye tracking dysfunction: a study of normal twins. Psychophysiology 37(6), 724–730.

Kennedy, J. L., Farrer, L. A., Andreasen, N. C., et al. (2003) The genetics of adult-onset neuropsychiatric disease: complexities and conundra? Science 302(5646), 822–826.

Kennedy, J. L., Honer, W. G., Kaufmann, C. A., Martignetti, J. A., Brosius, J., and Kidd, K. K. (1992) Two RFLPs near HOX2:/NGFR at locus d17s444e. Nucleic Acids Research, 20(5), 1171.

Kilpatrick, L. and Cahill, L. (2003) Amygdala modulation of parahippocampal and frontal regions during emotionally influenced memory storage. Neuroimage 20(4), 2091–2099.

Kim, U. K., Jorgenson, E., Coon, H., et al. (2003) Positional cloning of the human quantitative trait locus underlying taste sensitivity to phenylthiocarbamide. Science 299(5610), 1221–1225.

Kleinbaum, D. (1994) Logistic regression: a self-learning text. Springer-Verlag, New York.

Klempan, T. A., Muglia, P., and Kennedy, J. L. (2005) Genes in Brain Development. In: Keshavan, M., Kennedy, J. L., and Murray, R. (eds.) Neurodevelopment and Schizophrenia. Cambridge, Cambridge University Press.

Kohane, I., Kho, A., and Butte, A. (2003) Microarrays for an Integrative Genomics 2002. MIT Press: Cambridge, MA.

Kohtz, J. D., Baker, D. P., Corte, G., and Fishell, G. (1998) Regionalization within the mammalian telencephalon is mediated by changes in responsiveness to Sonic Hedgehog. Development 125(24), 5079–5089.

Krawczak, M., Cooper, D. N., and Schmidtke, J. (2001) Estimating the efficacy and efficiency of cascade genetic screening. Am. J. Hum. Genet. 69(2), 361–370.

Lauritzen, S. L. and Spiegelhalter, D. J. (1988) Local computations with probabilities on graphical structures and their application to expert systems. J. R. Stat. Soc. B. 50(2), 157–224.

Lawrie, S. M. and Abukmeil, S. S. (1998) Brain abnormality in schizophrenia. A systematic and quantitative review of volumetric magnetic resonance imaging studies. Br. J. Psychiatr. 172, 110–120.

Lee, K., Kunugi, H., and Nanko, S. (2001) Glial cell line-derived neurotrophic factor (GDNF) gene and schizophrenia: Polymorphism screening and association analysis. Psychiatry Res. 104(1), 11–17.

Lerer, B., Segman, R. H., Fangerau, H., et al. (2002) Pharmacogenetics of tardive dyskinesia: combined analysis of 780 patients supports association with dopamine D3 receptor gene Ser9Gly polymorphism. Neuropsychopharmacology 27(1), 105–119.

Letovsky, S. I., Whitehead, S. H., Paik, C. H., et al. (1998) A brain image database for structure/function analysis. AJNR Am. J. Neuroradiol. 19(10), 1869–1877.

Lin, K., Chudova, D., Hatfield, G., et al. (2004) Identification of hair cycle-associated genes from time-course gene expression profile data by using replicate variance. Proc. Natl. Acad. Sci. USA 101(45), 15,955–15,960.

Lohmueller, K. E., Pearce, C. L., Pike, M., et al. (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat. Genet. 33, 177–182.

Malfroy, L., Viratelle, C., Coppin, H., Borot, N., and Roth, M. P. (1995) Polymorphic tri- and tetranucleotide repeats in exons 1 and 8 of the myelin

oligodendrocyte glycoprotein (MOG) gene. Hum. Genet. 96(6), 737, 738.
Marchini, J. and Presanis, A. (2004) Comparing methods of analyzing fMRI statistical parametric maps. Neuroimage 22(3), 1203–1213.
Martin, L. F., Kem, W. R., and Freedman, R. (2004) Alpha-7 nicotinic receptor agonists: potential new candidates for the treatment of schizophrenia. Psychopharmacology (Berl.).
Martucci, L., Wong, A. H., Trakalo, J., et al. (2003) *N*-methyl-D-aspartate receptor nr1 subunit gene (GRIN1) in schizophrenia: Tdt and case–control analyses. Am. J. Med. Genet. 119B(1), 24–27.
McGinnis, R. E., Fox, H., Yates, P., et al. (2001) Failure to confirm NOTCH4 association with schizophrenia in a large population-based sample from Scotland. Nat. Genet. 28(2), 128, 129.
McLachlan, G. J., Do, K. -A., Ambroise, C. (2004) Analysing Microarray Gene Expression Data. Wiley: New York, NY.
McLachlan, G. J. and Peel, D. (2002) Finite Mixture Models, New York, Wiley.
Mechelli, A, Price, C. J., Noppeney, U., and Friston, K. J. (2003) A dynamic causal modeling study on category effects: bottom-up or top-down mediation? J. Cogn. Neurosci. 15(7), 925–934.
Meltzer, H. Y., Alphs, L., Green, A. I., et al. (2003) Clozapine treatment for suicidality in schizophrenia: International Suicide Prevention Trial (InterSePT) Arch. Gen. Psychiatr. 60(1), 82–91.
Miyaoka, T., Seno, H., and Ishino, H. (1999) Increased expression of WNT-1 in schizophrenic brains. Schizophr. Res. 38(1), 1–6.
Mohn, A. R., Gainetdinov, R. R., Caron, M. G., and Koller, B. H. (1999). Mice with reduced NMDA receptor expression display behaviors related to schizophrenia. Cell, 98(4), 427–436.
Muglia, P., Vicente, A. M., Verga, M., et al. (2003) Association between the BDNF gene and schizophrenia. Mol. Psychiatr. 8(2), 146, 147.
Mukaetova-Ladinska, E. B., Hurt, J., Honer, W. G., et al. (2002) Loss of synaptic but not cytoskeletal proteins in the cerebellum of chronic schizophrenics. Neurosci. Lett. 317(3), 161–165.
Murphy, K. P. (2002) Dynamic Bayesian Networks: Representation, Inference, and Learning. PhD Thesis, Computer Science Division, UC Berkeley.
Nadri, C., Dean, B., Scarr, E., and Agam, G. (2004) GSK-3 parameters in postmortem frontal cortex and hippocampus of schizophrenic patients. Schizophr. Res. Dec 1;71(2–3), 377–382.
Nanko, S., Hattori, M., Kuwata, S., et al. (1994) Neurotrophin-3 gene polymorphism associated with schizophrenia. Acta Psychiatr. Scand. 89(6), 390–392.
Nichols, T. and Hayasaka, S. (2003) Controlling the familywise error rate in functional neuroimaging: a comparative review. Stat. Methods Med. Res. 12(5), 419–446.
Northoff, G., Kotter, R., Baumgart, F., et al. (2004) Orbitofrontal cortical dysfunction in akinetic catatonia: a functional magnetic resonance imaging study during negative emotional stimulation. Schizophr. Bull. 30(2), 405–427.
Nyholt, D. R. (2001) Genetic case-control association studies—correcting for multiple testing. Hum. Genet. 109(5), 564–567.
Ohmori, O., Shinkai, T., Hori, H., Kojima, H., and Nakamura, J. (2000). Synapsin III gene polymorphisms and schizophrenia. Neurosci. Lett. 279(2), 125–127.
Ott, J. and Hoh, J. (2000) Statistical approaches to gene mapping. Am. J. Hum. Genet. 67, 289–294.
Pearl, J. (1988) Probabilistic Reasoning in Intelligent Systems. Los Alios, CA, Morgan Kaufmann.
Peltonen, L. and McKusick, V. A. (2001) Genomics and medicine. Dissecting human disease in the postgenomic era. Science 291, 1224–1229.
Penny, W., Kiebel, S., and Friston, K. (2003) Variational Bayesian inference for fMRI time series. Neuroimage 19(3), 727–741.
Penny, W. D., Trujillo-Barreto, N. J., and Friston, K. J. (2005) Bayesian fMRI time series analysis with spatial priors. Neuroimage 24(2), 350–362.
Perneger, T. V. (1998) What's wrong with Bonferroni adjustments. Bt. Med. J. 316(7139), 1236–1238.
Pevzner, P. (2000) Computational Molecular Biology: An Algorithmic Approach. MIT Press: Cambridge, MA.
Potkin, S. G., Basile, V. S., Jin, Y., and Masellis, M. (2002a) D1 receptor alleles predict PET metabolic correlates of clinical response to clozapine. Mol. Psychiatr. 8(1), 109–113.
Potkin, S. G., Alphs, L., Hsu, C., et al. (2003a) Predicting suicidal risk in schizophrenic and schizoaffective patients in a prospective two-year trial. Biol. Psychiatr. 54(4), 444–452.
Potkin, S. G., Alva, G., Fleming, K., et al. (2002b) A PET study of the pathophysiology of negative symptoms in schizophrenia. Positron emission tomography. Am. J. Psychiatr. 159(2), 227–237.
Potkin, S. G., Basile, V. S., Jin, Y., et al. (2003b) D1 receptor alleles predict PET metabolic correlates of clinical response to clozapine. Mol. Psychiatr. 8(1), 109–113.

Prasad, K. M., Chowdari, K. V., Nimgaonkar, V. L., et al. (2005) Genetic polymorphisms of the RGS4 and dorsolateral prefrontal cortex morphometry among first episode schizophrenia patients. Mol. Psychiatr. 10(2), 213–219.

Prasad, K. M., Patel, A. R., Muddasani, S., et al. (2004) The entorhinal cortex in first-episode psychotic disorders: a structural magnetic resonance imaging study. Am. J. Psychiatr. 161(9), 1612–1619.

Pritchard, J. K. (2001) Are rare variants responsible for susceptibility to complex traits? Am. J. Hum. Genet. 46, 222–228.

Proschel, M., Saunders, A., Roses, A. D., and Muller, C. R. (1992) Dinucleotide repeat polymorphism at the human gene for the brain-derived neurotrophic factor (BDNF). Hum. Mol. Genet. 1(5), 353.

Purdon, P. L. and Weisskoff, R. M. (1998) Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI. Hum. Brain Mapp. 6(4), 239–249.

Reich, D. E., Gabriel, S. B., Altshuler, D. (2003) Quality and completeness of SNP databases. Nat. Genet. 33, 457–458.

Rice, D. S. and Curran, T. (2001) Role of the reelin signaling pathway in central nervous system development. Ann. Rev. Neurosci. 24, 1005–1039.

Risch, N. and Merikangas, K. (1996) The future of genetic studies of complex human diseases. Science 273(5281), 1516, 1517.

Risch, N., Burchard, E., Ziv, E., and Tang, H. (2002) Categorization of humans in biomedical research: genes, race and disease. Genome Biol. 3(7), comment 2007.

Roberts, N. A., Beer, J. S., Werner, K. H., et al. (2004) The impact of orbital prefrontal cortex damage on emotional activation to unanticipated and anticipated acoustic startle stimuli. Cogn. Affect Behav. Neurosci. 4(3), 307–316.

Rodriguez, M. A., Pesold, C., Liu, W. S., et al. (2000) Colocalization of integrin receptors and reelin in dendritic spine postsynaptic densities of adult nonhuman primate cortex. Proc. Natl. Acad. Sci. USA 97(7), 3550–3555.

Rothman, K. and Greenland, S. (1998) Concept of interaction. In: Rothman K. and Greenland, S. (eds.) Modern Epidemiology. Philadelphia, Lippincott-Raven, pp. 329–342.

Sasaki, T., Dai, X. Y., Kuwata, S., et al. (1997) Brain-derived neurotrophic factor gene and schizophrenia in Japanese subjects. Am. J. Med. Genet. 74(4), 443, 444.

Scholkopf, B. and Smola, A. (2002) Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, Cambridge, MIT Press.

Scholkopf, B., Tsuda, K., and Vert, J. P. eds. (2004) Kernel Methods in Computational Biology. MIT Press: Cambridge, MA.

Schwab, S. G., Albus, M., Hallmayer, J., et al. (1995) Evaluation of a susceptibility gene for schizophrenia on chromosome 6p by multipoint affected sib-pair linkage analysis. Nat. Genet. 11(3), 325–327.

Segal, M. R., Dahlquist, K. D., and Conklin, B. R. (2003) Regression approaches for microarray data analysis. J. Comput. Biol. 10(6), 961–980.

Selemon, L. D., Rajkowska, G., and Goldman-Rakic, P. S. (1995) Abnormally high neuronal density in the schizophrenic cortex. A morphometric analysis of prefrontal area 9 and occipital area 17. Arch. Gen. Psychiatr. 52(10), 805–818; discussion 819, 820.

Sham, P. C. (1998) Statistical methods in psychiatric genetics. Stat. Methods Med. Res. 7(3), 279–300.

Shenton, M. E., Dickey, C. C., Frumin, M., and McCarley, R. W. (2001) A review of MRI findings in schizophrenia. Schizophr. Res. 49(1,2), 1–52.

Sivagnansundaram, S., Muller, D., Gubanov, A., et al. (2003) Genetics of Schizophrenia: Current Strategies. Clin. Neurosci. Res. 3(1,2), 5–15.

Sklar, P., Schwab, S. G., Williams, N. M., et al. (2001) Association analysis of NOTCH4 loci in schizophrenia using family and population-based controls. Nat. Genet. 28(2), 126–128.

Small, G. W. (2002) Brain-imaging surrogate markers for detection and prevention of age-related memory loss. J. Mol. Neurosci. 19(1,2), 17–21.

Small, G. W., Ercoli, L. M., Silverman, D. H., et al. (2000) Cerebral metabolic and cognitive decline in persons at genetic risk for Alzheimer's disease. Proc. Natl. Acad. Sci. USA 97(11), 6037–6042.

Smyth, P. (2000) Data mining: data analysis on a grand scale? Stat. Methods Med. Res. 9, 309–327.

Smyth, P., Heckerman, D., and Jordan, M. I. (1997) Probabilistic independence networks for hidden Markov models. Neural Computation, 9(2), 227–269.

Smyth, P., Pregibon, D., and Faloutsos, C. (2002) Data-driven evolution of data mining algorithms. Commun. ACM. 45(8), 33–37.

Speed, T. ed. (2003) Statistical Analysis of Gene Expression Microarray Data. Chapman & Hall/CRC: Boca Raton, FL.

Spellman, P. T., Sherlock, G., Zhang, M. Q., et al. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. Mol. Biol. Cell 9(12), 3273–3297.

Spinks, R., Sandhu, H. K., Andreasen, N. C., and Philibert, R. A. (2004) Association of the HOPA12bp allele with a large X-chromosome haplotype and positive symptom schizophrenia. Am. J. Med. Genet. B. Neuropsychiatr Genet. 127(1), 20–27.

Stober, G., Syagailo, Y. V., Okladnova, O., et al. (1999) Functional pax-6 gene-linked polymorphic region: Potential association with paranoid schizophrenia. Biol. Psychiatry. 45(12), 1585–1591.

Stone, V. E., Baron-Cohen, S., and Knight, R. T. (1998) Frontal lobe contributions to theory of mind. J. Cogn. Neurosci. 10(5), 640–656.

Storey, J. D. and Tibshirani, R. (2003) Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. USA 100(16), 9440–9445.

Straub, R. E., Jiang, Y., MacLean, C. J., et al. (2002) Genetic variation in the 6p22.3 gene DTNBP1, the human ortholog of the mouse dysbindin gene, is associated with schizophrenia. Am. J. Hum. Genet. 71(2), 337–348.

Straub, R. E., MacLean, C. J., O'Neill, F. A., et al. (1995) A potential vulnerability locus for schizophrenia on chromosome 6p24-22: evidence for genetic heterogeneity. Nat. Genet. 11(3), 287–293.

Swift-Scanlan, T., Lan, T. H., Fallin, M. D., et al. (2002) Genetic analysis of the (CTG)n NOTCH4 polymorphism in 65 multiplex bipolar pedigrees. Psychiatr. Genet. 12(1), 43–47.

Tabor, H. K., Risch, N. J., and Myers, R. M. (2002) Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat. Rev. Genet. 3, 391–397.

Tachikawa, H., Harada, S., Kawanishi, Y., Okubo, T., and Suzuki, T. (2001) Polymorphism of the 5′-upstream region of the human SNAP-25 gene: An association analysis with schizophrenia. Neuropsychobiology, 43(3), 131–133.

Takahashi, M., Shirakawa, O., Toyooka, K., et al. (2000) Abnormal expression of brain-derived neurotrophic factor and its receptor in the corticolimbic system of schizophrenic patients. Mol. Psychiatr. 5(3), 293–300.

Terwilliger, J. D. and Weiss, K. M. (1998) Linkage disequilibrium mapping of complex disease: fantasy or reality? Cur. Opin.Biotechnol. 9, 578–594.

Thome, J., Durany, N., Harsanyi, A., et al. (1996) A null mutation allele in the CNTF gene and schizophrenic psychoses. Neuroreport, 7(8), 1413–1416.

Thompson, P. M., Sower, A. C., and Perrone-Bizzozero, N. I. (1998) Altered levels of the synaptosomal associated protein SNAP-25 in schizophrenia. Biol. Psychiatr. 43(4), 239–243.

Toga, A. W. and Mazziotta, J. C. (2000) Brain Mapping: The Methods. Academic Press, Amsterdam, Boston.

Tsuang, M. T. (1998) Genetic epidemiology of schizophrenia: review and reassessment. Kaohsiung J. Med. Sci. 14(7), 405–412.

Turetsky, B., Cowell, P. E., Gur, R. C., et al. (1995) Frontal and temporal lobe brain volumes in schizophrenia. Relationship to symptoms and clinical subtype. Arch. Gen. Psychiatr. 52(12), 1061–1070.

Tuulio-Henriksson, A., Ilonen, T., Pirkola, T., and Lonnqvist, J. (2000) Neuropsychology in research and therapy of schizophrenia. Duodecim 116(14), 1453–1458.

Ujike, H., Takehisa, Y., Takaki, M., et al. (2001) NOTCH4 gene polymorphism and susceptibility to schizophrenia and schizoaffective disorder. Neurosci. Lett. 301(1), 41–44.

Vicente, A. M., Macciardi, F., Verga, M., et al. (1997) NCAM and schizophrenia: Genetic studies. Mol. Psychiatr. 2(1), 65–69.

Wang, J., Shapiro, B., Shasha, D. eds. (1999) Pattern Discovery in Biomolecular Data: Tools, Techniques, and Applications. Oxford University Press: New York, NY.

Wang, W. Y. S., Cordell, H. J., and Todd, J. A. (2003) Association mapping of complex diseases in linked regions: estimation of genetic effects and feasibility of testing rare variants. Genet. Epidemiol. 24, 36–43.

Wei, J. and Hemmings, G. P. (2000) The NOTCH4 locus is associated with susceptibility to schizophrenia. Nat. Genet. 25(4), 376, 377.

Weinberger, D. R. and Berman, K. F. (1996) Prefrontal function in schizophrenia: confounds and controversies. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 351(1346), 1495–1503.

Weinberger, D. R., Egan, M. F., Bertolino, A., et al. (2001) Prefrontal neurons and the genetics of schizophrenia. Biol. Psychiatr. 50(11), 825–844.

Weiss, K. M. and Terwilliger, J. D. (2000) How many diseases does it take to map a gene with SNPs? Nat. Genet. 26, 151–157.

Winkler, G. (2003) Image Analysis, Random Fields and Markov Chain Monte Carlo Methods:

A mathematical introduction. Springer: Berlin, 2nd edition.

Winokur, G. (1975) Paranoid vs hebephrenic schizophrenia: clinical and familial (genetic) heterogeneity. Psychopharmacol. Commun. 1(6), 567–577.

Wolkin, A., Sanfilipo, M., Duncan, E., et al. (1996) Blunted change in cerebral glucose utilization after haloperidol treatment in schizophrenic patients with prominent negative symptoms. Am. J. Psychiatry 153(3), 346–354.

Wong, A. H., Macciardi, F., Klempan, T., et al. (2003) Identification of candidate genes for psychosis in rat models, and possible association between schizophrenia and the 14-3-3eta gene. Mol. Psychiatry 8(2), 156–166.

Wong, A. H., Trakalo, J., Likhodi, O., et al. (2004) Association between schizophrenia and the syntaxin 1A gene. Biol. Psychiatry Jul 1;56(1), 24–29.

Wright, P., Nimgaonkar, V. L., Donaldson, P. T., and Murray, R. M. (2001) Schizophrenia and HLA: a review. Schizophr. Res. 47(1), 1–12.

Xu, J., Pato, M. T., Torre, C. D., et al. (2001) Evidence for linkage disequilibrium between the alpha 7-nicotinic receptor gene (CHRNA7) locus and schizophrenia in Azorean families. Am. J. Med. Genet. 105(8), 669–674.

Young, C. E., Arima, K., Xie, J., et al. (1998) SNAP-25 deficit and hippocampal connectivity in schizophrenia. Cereb. Cortex 8(3), 261–268.

Zai, G., King, N., Wigg, K., et al. (2005) Genetic study of the myelin oligodendrocyte glycoprotein (MOG) gene in schizophrenia. Genes Brain Behav. Feb;4(1), 2–9.

Zaykin, D. V., Westfall, P. H., Young, S. S., et al. (2002) Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals. Hum. Heredity 53, 79–91.