9.

IDENTIFYING UNANTICIPATED GENES AND MECHANISMS IN SERIOUS MENTAL ILLNESS

GWAS-BASED IMAGING GENETICS STRATEGIES

Steven G. Potkin, Theo G. M. van Erp, Shichun Ling, Fabio Macciardi, and Xiaohui Xie

INTRODUCTION

Genome-wide association studies (GWAS) investigate the relationship or association between genetic markers, usually single-nucleotide polymorphisms (SNPs), along the entire genome with categorical (e.g., diagnoses) or quantitative traits (e.g., imaging phenotypes). In contrast, candidate gene association studies investigate the relationship between genetic markers at a particular locus (e.g., a gene hypothesized to have biological relevance for a particular disorder) and either categorical or quantitative traits. The GWAS approach has an advantage compared to the candidate gene approach because it does not require a priori knowledge of a trait's etiological connection, or its function singly or as a part of a network (Potkin et al. 2009a; Stranger, Stahl, and Raj 2011); thus this approach can lead to unanticipated discoveries of genes associated with mental illness. While the advantages of this approach apply to mental illness in general, we will confine our examples to schizophrenia and bipolar subjects, as several SNPs are risk factors for the development of both disorders and for some of their shared clinical characteristics. We would prefer to include only relevant studies with both bipolar and schizophrenia subjects; however, given that there are very few such studies, we will include studies with either disorder.

Why combine genetic and imaging data? Mental disorders are brain disorders that can be studied with functional brain imaging measures, which are known to be highly heritable. Studying brain imaging measures in neuropsychiatry without considering genetics neglects part of the risk for developing mental disorders such as schizophrenia and bipolar disorder. Similarly, studying genetics in psychiatric disorders without determining their brain effects fails to understand (at least part of) the consequences of these genetic influences. The rationale for considering both schizophrenia and bipolar disorder is due to their overlap in clinical symptoms, medication response, and risk genes (Tesli et al. 2014), although we are aware that the two disorders also have significant differences in these domains. The approach to jointly study multiple psychiatric syndromes is in accord with the National Institute of Mental Health's Research Domain Criteria (RDoC) approach of examining possible common mechanisms across disorders. The Psychiatric Genomics Consortium (PGC) cross-diagnosis working group has identified six candidate regions that are shared risk factors for both schizophrenia and bipolar disorders (Ruderfer et al. 2014). However, these data to date lack a functional phenotype that is likely to be critical in understanding the candidate regions' role in these disorders.

In this chapter we selectively review studies that combine both genome-wide genetic and functional imaging data in schizophrenia and/or bipolar disorder. An advantage of the GWAS approach is that it surveys the entire genome for associations between genetic markers and phenotypes (e.g., diagnosis, imaging, and/or clinical characteristics). Adequate SNP coverage is important, and in current studies at least 750,000 SNPs proportionately distributed across the genome are needed for detailed association studies. With imputation of SNPs absent in early SNP chips used for genotyping, this has become a realistic goal. Rich multimodal quantitative imaging phenotype data offer the potential of linking genetic risk with functional brain consequences and have several additional advantages over case-control studies that will be discussed

later in this chapter. We did not include linkage studies, which survey the whole genome with less than 10,000 SNP markers, because of the large number of genes contained in a typical linkage peak.

We searched the PubMed database with the following criteria: the "All Fields" field contained "imaging," "fMRI," "functional MRI," "resting state," "resting state fMRI," "GWAS," "genome-wide association study," "genetics," "schizophrenia" and/or "bipolar"; or "MeSH Major Topic" contained "MRI," "fMRI," "functional MRI," "imaging," "GWAS," "genome-wide association study," "schizophrenia" and/or "bipolar"; and various combinations of these terms.

We integrated the search results and limited them to papers that contained both the "GWAS" and "imaging" stems. We manually reviewed all the abstracts prior to September 1, 2014, and identified 12 relevant functional imaging genetics publications related to schizophrenia and/or bipolar through this extensive search. We have summarized these publications in Table 9.1 (in the following section).

BACKGROUND

The integration of imaging and genetic data has the potential to not only avoid the limitations of studying genetics or brain imaging in isolation but can leverage their synergism. Brain imaging has productively been used to clarify the brain effects of candidate genes. These candidate genes can be a priori chosen or identified in GWAS. One example is the study by Walton et al. (2013), who examined the structural and functional imaging associations with neurogranin SNPs that were identified in a GWAS of 12,945 schizophrenia and 34,591 healthy controls (Stefansson et al. 2009). Walton et al. (2013) observed increased cortical inefficiency and increased gray matter thinning in schizophrenia patients with the neurogranin risk alleles. Prefrontal cortical inefficiency and gray matter thinning are both well-documented findings in schizophrenia. Walton et al. (2013) linked these brain findings with specific neurogranin genetic risk, thereby moving toward potentially clarifying an underlying etiological mechanism.

Brain imaging can also be used as a quantitative trait to discover unexpected risk genes for schizophrenia and bipolar disorder (see Table 9.1 and Figure 9.1). For example, in a sample of less than 200 subjects, the quantitative trait (QT) dorsolateral prefrontal cortex (DLPFC) BOLD activation measured during a working memory task identified several unanticipated risk SNPs for schizophrenia. These statistically significant SNPs included TNIK (TRAF2 and NCK-interacting kinase) (Potkin et al. 2009b). Interestingly, the previously identified candidate gene *DISC1* modulates TNIK kinase activity. *DISC1* and TNIK are co-localized in dendritic spines of the hippocampus. Together they co-regulate synaptic GLUR1 and AMPA activity (Wang et al. 2011). Dysregulation in both receptors has been hypothesized in schizophrenia (Harrison and Weinberger 2005) and bipolar disorder (Wang et al. 2011).

In addition to these biological mechanistic implications, a QT approach has distinct advantages in power over categorical diagnoses (i.e., healthy control vs. schizophrenia and/or bipolar disorder). QT approaches have approximately 4–10 times more statistical power (Purcell, Cherny, and Sham 2003; Potkin et al. 2009a; Shen et al. 2014;) by making use of the entire distribution of trait values. This approach also avoids the often arbitrary or error-prone cutoff distinctions, for example, thresholds or differing criteria for the diagnosis of schizophrenia are different in *ICD-10, DSM-IV*, and *DSM-5*. In addition, a GWAS-QT analysis, compared to a GWAS-CT (categorical trait) analysis, offers an alternative strategy to discover unanticipated genes associated with schizophrenia and/or bipolar risk as described for TNIK. In identifying TNIK, DLPFC activation during a working memory task was the starting place, as aberrant activation is well-described in schizophrenia. The GWAS SNPs that were statistically significantly associated with the degree of DLPFC activation were then identified. This approach began with a brain imaging characteristic of schizophrenia and/or bipolar disorder and then identified genes (or SNPs, or other types of genetic variation) associated with that imaging phenotype. Such intermediate phenotypes may have greater sensitivity in clarifying the functional links related to the schizophrenia and/or bipolar risk genes than diagnoses, in part because they are quantitative and possibly also because they are closer to the biological mechanisms underlying the illness than symptom patterns.

Symptoms or cognitive performance may be affected by many factors in addition to genetic risk, such as medication and social factors, as well as modifying genes. One option to avoid the potential confounds of medication and disease is to study subjects at genetic or clinical high risk for developing these disorders, usually non-ill relatives or at-risk adolescents (see Table 9.1). Studies of genetic risk influences in high-risk subjects have the advantages of avoiding medication and illness effects that can affect brain activation

and interpretation of the results. On the other hand, clarifying mechanisms of disease are the primary motivation for imaging genetics studies. Non-ill high-risk subjects who are past the age of illness onset have a genetic vulnerability, but the vulnerability has been insufficient to produce the disease, possibly because of the lack of additional genetic or environmental influences or the presence of compensatory mechanisms. Studies of adolescents at clinical high risk (CHR) or offspring from parents with the illness have an advantage because they can be followed longitudinally such that both risk- and disease-related effects (in those who become ill) can be studied. However, these studies take many years to complete and therefore require considerable time and financial investment. Others have chosen to investigate the function of GWAS-identified candidates by conducting brain imaging studies in healthy controls. These studies also avoid confounds of medication and disease, but may ignore the perturbations of genetic networks that are responsible for the risk liability for disease.

Previous GWAS have provided candidates that have subsequently been studied with brain imaging, producing useful data regarding the brain effects of these candidate genes. We do not discuss these studies, as their designs are straightforward. We focus on publications that combine GWAS and brain imaging in a single study. These studies offer, in our view, the most robust strategy for identifying unanticipated genetic influences in serious mental illness (SMI) and for understanding the underlying mechanisms, while presenting the most challenging methodological and statistical issues. This chapter focuses on addressing those challenges and highlighting the promise of this method.

STRUCTURAL AND FUNCTIONAL NEUROIMAGING

Structural neuroimaging has been the most studied of brain phenotypes in psychiatry. Brain-genome associations can be identified at several levels. From a genomic perspective, we can examine candidate genes (identified by surrogate SNPs), part of the genome involved in biological pathways or networks, the entire genome (GWAS), or the DNA sequence (e.g., exome, or whole genome seqeuncing (WGS)). Current multi-center collaborative research efforts such as ENIGMA (Enhancing Neuro Imaging Genetics through Meta Analysis) can leverage imaging and genetic data on thousands of subjects and are beginning to reveal the genomic underpinning of brain structural differences across disorders (Stein et al. 2012).

From an imaging perspective, individual region of interest (ROI), brain circuits including connectivity between multiple ROIs, or the whole brain (e.g., voxel-wise or cortical surface) activations can represent the entry point (phenotype/dependent measure[s]) of an investigation. Figure 9.2 shows examples of studies in each category. In this chapter we review fMRI (functional MRI) and rsfMRI (resting state fMRI) but not EEG/EVP/ERP, MEG, PET, SPECT, or fNIR because we were unable to find GWAS combined with these imaging methods in schizophrenia and/or bipolar subjects, although the discussion and points raised also apply to these methods. Structural imaging, which is discussed in detail in other chapters of this book, is a more indirect measure of brain function than physiological activations, though the relationship between structural and functional brain measures remains insufficiently defined.

In this chapter, we focus on functional brain imaging, as it likely reflects a more direct measure of the functional consequences of genetic risk than those reflected by structural imaging, though some genetic variation may contribute to both or be associated with functional and not structural brain changes, and vice versa. Moreover, abnormal brain structure may result in abnormal brain function, and abnormal brain function may result in changes in brain structure, shape, and/or volume. However, we believe that to understand underlying mechanisms, functional changes may be more sensitive and have greater specificity than structural alterations. We review an unbiased whole genome approach using SNP associations with functional phenotypes. As genes do not function in isolation but rather function within pathways or networks, we also consider genetic association of functional networks with targeted functional phenotypes. All of these analyses require corrections for multiple testing to minimize the risk of false positives (type 1 errors) when searching millions of variants in our genetic code; therefore statistical threshold corrections are discussed first.

IMAGING GENETICS STATISTICS

Genetic studies have begun to implement a high statistical threshold (e.g., p-value < 5×10^{-8} in case-control studies) to implicate a genetic variant with a trait or a disorder. Currently there is considerable debate of what an appropriate statistical threshold is for imaging genetic findings to avoid type 1 errors without over-corrections. The widely accepted threshold (Risch and Merikangas 1996; Barsh et al. 2012) for a conventional GWAS is nominal

TABLE 9.1 IMAGING GENETICS PUBLICATIONS IN SCHIZOPHRENIA AND BIPOLAR DISORDER

Imaging	Author	Year	Cohort (n)	Imaging Cohort	fMRI Phenotype	Positive SNPs/Total	Top SNPs	Genes	Analysis
FA	Sprooten et al.	2013	70 high risk for BD; 80 HC	70 high risk for BD; 80 HC	Fractional anisotropy	730K	10 SNPs	LPP, HEPACAM, HEPN1, ROBO4, CAMK2D	Association
FA	Whalley et al.	2013	70 BD/ MDD family members; 62 HC	70 BD/ MDD family members; 62 HC	Fractional anisotropy	Polygenic risk score using 730K GWAS	PRS with FA in R superior longitudinal fasciculus in MDD but not BD	N/A	Association w/multi-dimensional scaling
Resting state	Martin et al.	2014	78 SZ	33 SZ	Resting state cognitive control and default network	900K	Total deletion burden		Association
Resting state	Meda et al.	2014	296 SZ; 300 BD; 324 HC; 385 relatives	190 SZ; 189 BD; 170 HC	Resting state default mode network (DMN)				Parallel imaging and genetic ICA
Motor	Chen et al.	2012	92 SZ; 116 HC	92 SZ; 116 HC	Motor task in pre/post central gyri activation	253 of 5157 selected from 778K	253 SNPs comprising an ICA component		Parallel imaging and genetic ICA
Auditory	Liu et al.	2012	48 SZ; 40 HC	48 SZ; 40 HC	Auditory oddball task in thalamus anterior and posterior cingulate gyri	300K	26		Parallel imaging and genetic ICA; association, permutation
Cognitive Control	Rietschel et al.	2012	1169 SZ; 3714 HC for GWAS	122 HC	Flanker cognitive control task in anterior cingulate	475K	top GWAS SNP rs11154491; rs7112229, rs11819869, rs7130141 and rs12575668	ARGHAP18; AMBRA1, DGKZ, and CHRM4	ANOVA
Faces	Ousdal et al.	2012	51 SZ; 64 BD; 94 HC		Face processing in amygdala	546K	rs10014254, rs11722038, rs17529323	PHOX2B	Association

Main Effect *p*-Value	*p*-Value of Interaction	Percent of Variance Explained	Results	PubMed URL
10^-5 - 10^-6	10^-5 - 10^-7; top 10 SNPs associated with FA within each group, but in opposite directions of effect.		KEGG: axon guidance, ErbB-signaling neurotrophin signaling, phosphatidylinositol signaling, and cell adhesion	http://www.ncbi.nlm.nih.gov/pubmed/23218918
N/A	Sig negative association between MDD PRS and FA values in the right SLF, ILF, and IFOF in the parietal region (pFWE < 0.05) but not BD		Negative correlation between PRS* and FA in MDD but not BD	http://www.ncbi.nlm.nih.gov/pubmed/23453289
$p < .001$	N/A		DLPFC and putamen cognitive control connectivity decreaed as deletion burden increased. Dysregulated DMN positively associated with deletion burden.	http://www.ncbi.nlm.nih.gov/pubmed/25036426
3 significant DMNs (p = 4.8E-4; 8.7E-5;5.3E-7)	NS	<3.5%	Pathways: NMDA-related long-term potentiation, PKA, immune response signaling, axon guidance, and synaptogenesis	http://www.ncbi.nlm.nih.gov/pubmed/24778245
r = .29, p<10^-5, permutation corrected $p < .03$	$r = .2$, p = 10^-3		Pathways: GABA receptor signaling, dopamine receptor signaling, neuregulin signaling, and glutamate receptor signaling	http://www.ncbi.nlm.nih.gov/pubmed/22440650
$r = 0.39$ ($p < 0.0001$)	$p < 0.0001$	15%	Thalamus and cingulate activation with chr. 7q21 region (rs10953026, rs2286696, rs758706, rs3824039, rs2188240, rs1110457, rs7669821, rs2381387, and rs1419005) mainly involving gene CDK14, NXT1, and UGDH; and chr 5q35 region (rs4105175, rs2173096, rs2067011, rs1445846, rs1445845, rs1445844, and rs11750568) mainly involving gene GRM6 and ZNF354C.	http://www.ncbi.nlm.nih.gov/pubmed/22371699
$p < 0.05$ FWE corrected	N/A		AMBRA1, DGKZ, CHRM4 (muscarinic receptor), MDK, and ARGHAP18 (imaging)	http://www.ncbi.nlm.nih.gov/pubmed/21747397
rs10014254 was 10^-8	NS		Survival of adrenergic neurons; monoamine synthesis	http://www.ncbi.nlm.nih.gov/pubmed/22856363

TABLE 9.1 CONTINUED

Imaging	Author	Year	Cohort (n)	Imaging Cohort	fMRI Phenotype	Positive SNPs/Total	Top SNPs	Genes	Analysis
Faces	Liu et al.	2010	39 BD; 29 HC	39 BD; 29 HC	Face processing in amygdala	104K	rs2023454; rs6354	DOK5; serotonin 5HTT	Association
Sentence Completion	Whalley et al.	2012	87 BDHR; 71 HC	73 BDHR; 52 HC	Sentence completion in anterior cingulate and amygdala	730K GWAS used to generate PRS	N/A	N/A	Association
Working memory	Potkin et al.	2009b	64 SZ; 74 HC	64 SZ; 74 HC	Sternberg working memory in DLPFC	300K	rs7610746; rs9836484; rs2088885; rs7627954; rs245178; rs245201; rs1574192; rs9491640; rs10133111	ROBO1-ROBO2, TNIK, and CTXN3-SLC12A2; POU3F2, TRAF, and GPC1	Association with permutation
Working memory	Potkin et al.	2009c	28 SZ (discovery); 82 SZ, 91 HC	28 SZ	Sternberg working memory task in DLPFC and associated circuitry	99K	rs1915935, rs1624094, rs1729997; rs12664247, rs 4509146	RSRC1 and ARHGAP18	Association

*The polygenic risk score (PRS) term includes risk profile score (RPS), PPRS (pathway polygenic risk score), and GRS (genetic risk score)

p-value < 5×10^{-8}. This threshold was based on a FWER (Familywise Error rate, like Bonferroni's, for example) correction for a 1-tailed test, assuming 100,000 genes with an average of five SNPs per gene, and there is debate regarding the appropriateness of this threshold (Dudbridge and Gusnanto 2008; Xu et al. 2014). In brief, an appropriate threshold for inferring significance can be proposed a priori but also requires an a priori knowledge of the specific effect size (or odds ratio) of the risk factors being investigated. While such information is available in epidemiological studies, we generally do not know the actual effect size or odds ratio attributed to a genetic DNA variant; hence it is often difficult to justify the pre-specified fixed significance threshold used in GWAS.

A major issue in setting the nominal p-value threshold is the multiple hypotheses testing problem. Since many genetic markers (~1 million in typical GWAS) are tested for association with the phenotypes, the nominal significance levels derived from single tests must be corrected by the total number of independent tests carried out within the entire study. The simplest correction approach is through the Bonferroni correction, which assumes that all the individual tests applied to each marker are independent of each other. Such an assumption is unfounded, since many markers are in linkage disequilibrium with each other. As a result, the Bonferroni correction is overly conservative.

An alternative approach to multiple testing correction without assuming marker independence is through permutation testing (Dudbridge, Gusnanto, and Koeleman 2006) (see, e.g., http://pngu.mgh.harvard.edu/~purcell/plink/perm.shtml). In permutation testing, the phenotype labels are repeatedly randomly and are iteratively reassigned, and the statistical tests are run thousands or hundreds of thousands of times to produce an empirical distribution and, hence, an empirically derived p-value threshold for each SNP. This approach addresses several statistical issues by letting the collected data set determine the correct threshold rather than relying on generalized corrections and avoiding dependency on various assumptions, such as normal distribution of data and the degree of linkage disequilibrium across SNPs, and so on. Through permutation testing, Dudbridge and Gusnanto (2008) found that a nominal p-value < 7.2×10^{-8} would correspond to an experiment-wise p-value of < 0.05. However appealing, a permutation test procedure is computationally heavy when using millions of DNA variants.

Main Effect *p*-Value	*p*-Value of Interaction	Percent of Variance Explained	Results	PubMed URL
10–7 with FDR correction; $p < .005$	NS	33% for BD; 12% for HC; 11% for serotonin transporter	Neurotrophin signaling pathways, neuronal development, and differentiation; and serotonin regulation	http://www.ncbi.nlm.nih.gov/pubmed/20215924
PRS and anterior cingulate ($r = .2$ for BDHR; $r = .5$ for HC; $p < .01$) and amygdala ($r = .18$ for BDHR; $r = .45$ for HC; $p = .02$)	NS		PRS associated with increased anterior cingulate cortex and amygdala activation across both groups.	http://www.ncbi.nlm.nih.gov/pubmed/22760554
N/A	10^-6		Neurodevelopment and response to stress	http://www.ncbi.nlm.nih.gov/pubmed/19023125
10^-7	10^-4 - 10^-7		DLPFC and anatomically connected neocortical circuitry associated with RSRC1 (D1 function) and ARHGAP18 (cell migration and differentiation)	http://www.ncbi.nlm.nih.gov/pubmed/19065146

Instead of correcting nominal *p*-values, another approach to multiple testing is to monitor the False Discovery Rate (FDR) (Benjamini and Hochberg 1995), or the rate that the rejected or significant features are truly null, essentially to control for the expected proportion of incorrectly rejected null hypotheses (i.e., false discoveries). One can then choose a nominal *p*-value threshold according to an acceptable level of FDR. FDR can be calculated through the standard Benjamini-Hochberg method (1995) or through permutation studies. Several other methods have been recently developed to control the portion of false positives, such as the positive false discovery rate (pFDR) and the proportion of false positives (PFP) (see review by Pan 2013).

In order to generate greater power to offset the multiple test correction penalty when testing for the statistical significance of multiple SNPs, several methods have been developed to calculate a composite risk score or polygenic risk score (PRS). Having a single total risk metric is attractive for disease prediction purposes; however, it negates much of the value of GWAS in its ability to discover the functional role of unanticipated genes. Examples of PRS include risk profile scores (RPS) (Purcell et al. 2009; Ripke et al. 2013), genetic risk scores (GRS) (Walton et al. 2013), and pathway polygenic risk scores (PPRS) (Nicodemus et al. 2104), each of which combine GWAS findings to generate a total risk score without regard to the function of individual genes, though the latter takes into account known pathways.

RISK PROFILE SCORE

Risk score profiling was first proposed in the initial schizophrenia case-control study involving 3,322 European individuals with schizophrenia and 3,587 controls carried out by the International Schizophrenia Consortium (Purcell et al. 2009). Only a few SNPs reached genome-wide significance in this study. The authors speculated that this might be due to the polygenic nature of the schizophrenia—many SNPs (possibly in order of thousands) with very small individual effects collectively account for a large fraction of variation in risk. To test this hypothesis, the authors introduced a method to summarize variation across nominally associated loci into a single quantitative score, called a risk profile score (RPS), for each patient.

The risk score profiling analysis is typically carried out using two sets of samples—one called the discovery

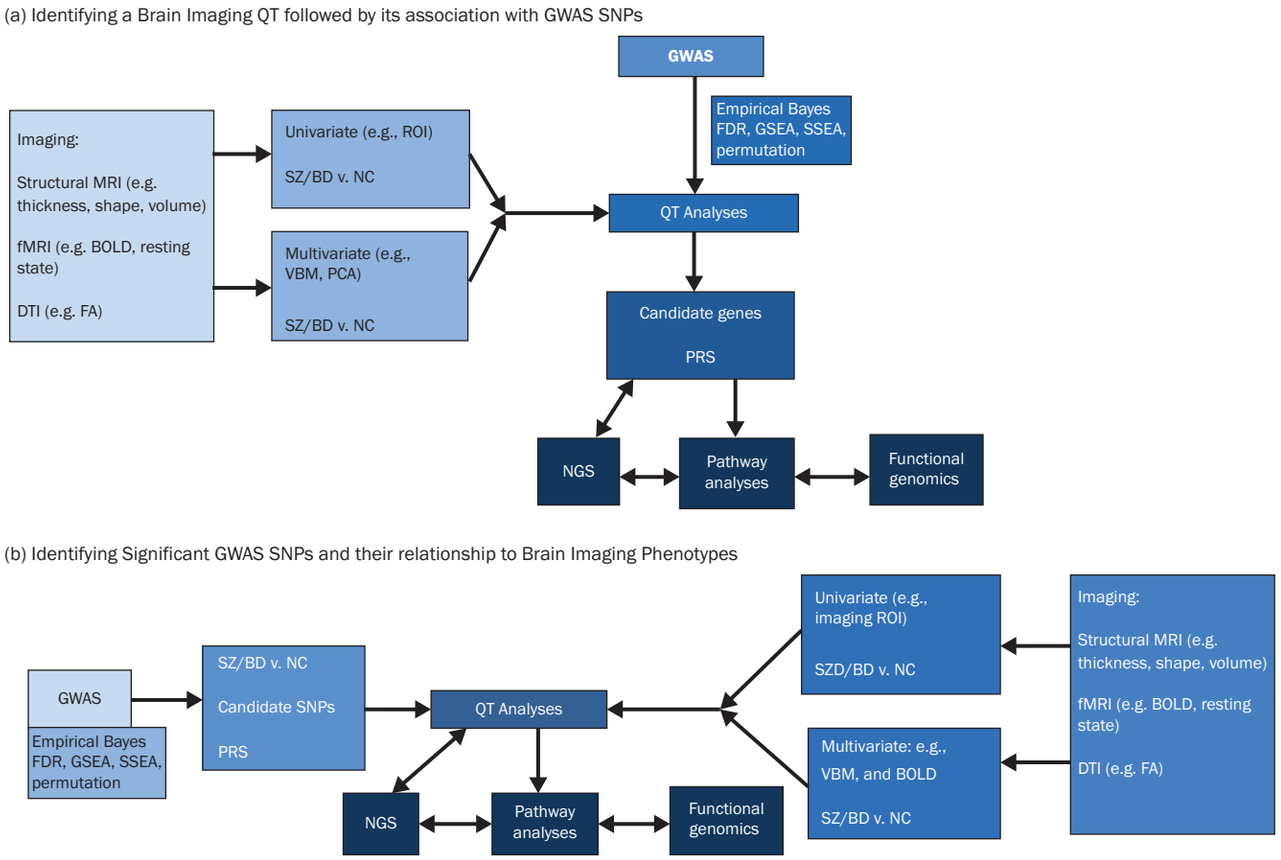

Figure 9.1 FLOW DIAGRAM OF TWO IMAGING GENETICS DATA-DRIVING ANALYSIS STRATEGIES. The upper figure (a) begins with brain imaging to define imaging phenotypes, which are then integrated with GWAS data in a QT analysis to understand which genetic variants influence these imaging phenotypes. The lower figure (b) begins with GWAS data to identify genetic candidates that differentiate diagnoses or groups. The candidates are combined with imaging data in a QT analysis to determine the brain effects of those identified genetic variants. In both approaches, the results of the QT analyses that represent integrated imaging genomic data are more fully explored with deep sequencing, pathway analysis, and functional genomics to achieve a deeper biological understanding. The color shading from light to dark represents the chronological order of analysis steps. MRI = magnetic resonance imaging; fMRI = functional MRI; BOLD = blood oxygen level dependent; DTI = diffusion tensor imaging; FA = fractional anisotropia; ROI = region of interest; SZ = schizophrenia; BD = bipolar disorder; NC = normal healthy controls; VBM = voxel-based morphometry; PCA = principal component analysis; GWAS = genome-wide association study; FDR = false discovery rate; GSEA = gene-set enrichment analysis; SSEA = SNP-set enrichment analysis; QT = quantitative trait; PRS = polygenic risk score; NGS = next generation sequencing

sample, and the other called the target sample. The analysis begins by selecting high-quality SNPs by filtering on minor allele frequency, genotyping rate, and linkage disequilibrium, but without regard to association with phenotypes. Different subsets of SNPs are then selected from the list of high-quality SNPs by increasing *p*-value thresholds on the discovery sample (e.g., from 0.001 to 0.5). These *p*-values thresholds are quite liberal so that a large number of the SNPs will be selected. For a given *p*-value threshold, the subset of SNPs with association *p*-values below the threshold in the discovery sample are called *score alleles*. The RPS for each individual in the target sample is then calculated to be the number of score alleles weighted by the log odds ratio, or the logistic regression coefficient from the discovery sample. Finally, logistic regression is carried out to test whether the cases in the target sample have significant burden of score alleles when compared to the controls. The entire risk score profiling analysis can be done using the "–score" function in PLINK.

Applying this method to two schizophrenia samples from the International Schizophrenia Consortium, in which the discovery sample consisted of selected males and the target sample consisted of selected females, Purcell et al. (2009) were able to show that the score alleles from the discovery sample were significantly enriched in the target cases, and the effect was larger for increasingly liberal *p*-value thresholds. Applying the same method to the Swedish National Sample and the Psychiatric Genomic Consortium (PGC) schizophrenia sample, Ripke et al. (2013) were able to show that the PGC score alleles were highly correlated with the case-control status in the independent Swedish samples.

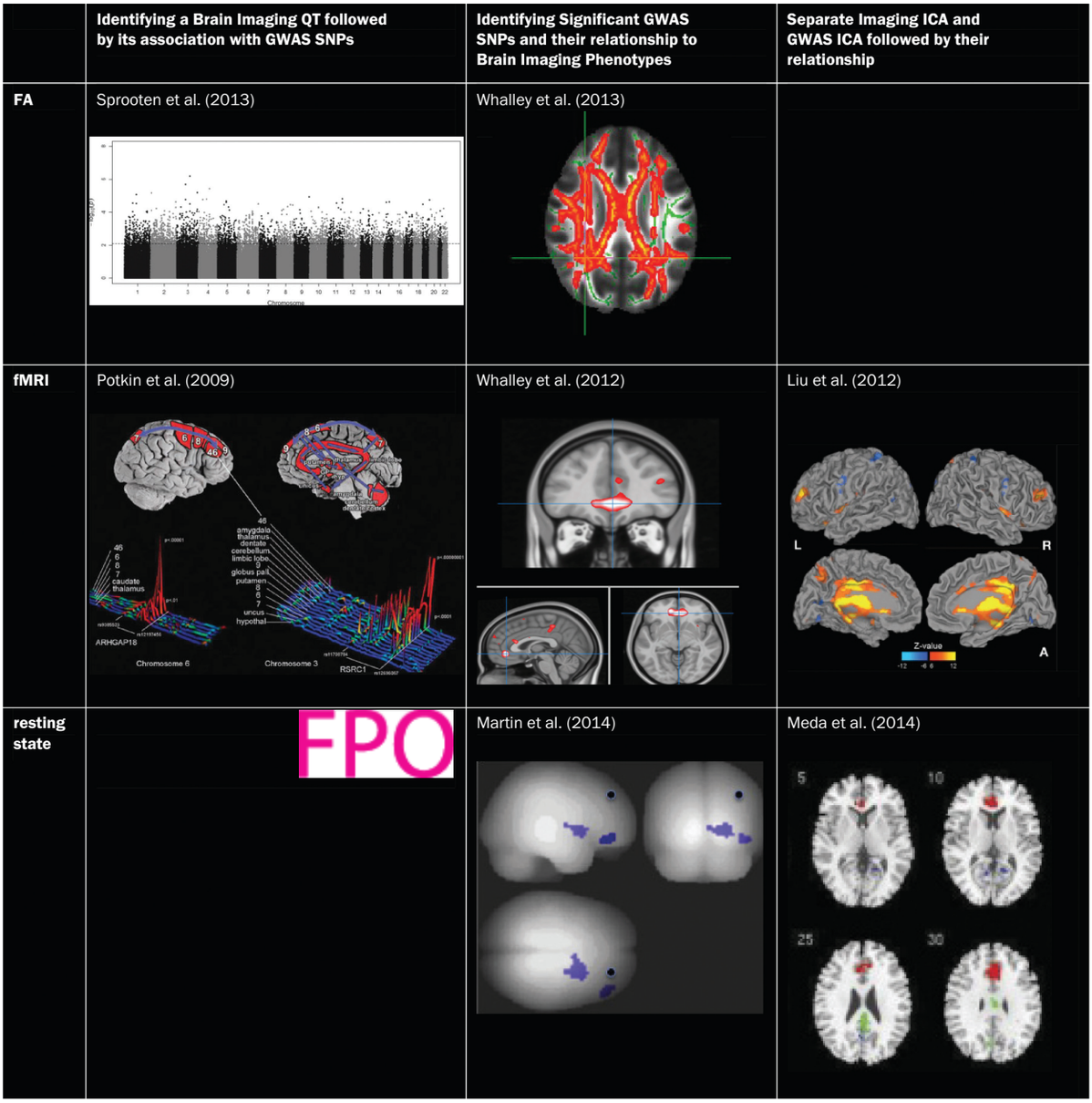

Figure 9.2 IMAGING GENETICS ANALYSIS EXAMPLES. The figure illustrates a range of selected functional imaging genetic results beginning with brain imaging QTs (left column), GWAS-generated genetic risk (middle column), or simultaneous parallel brain imaging and genetic QTs (right column). Sprooten et al. (2013): Using FA as a QT measure of white matter integrity, SNPs related to cell adhesion, white matter development, and neuronal plasticity were associated with FA and differentiated subjects at high risk for bipolar from healthy controls. Potkin et al. (2009): Using DLPFC activation in a working memory task as a QT, SNPs related to dopamine function (RSRC1) and cell migration and differentiation (ARHGAP18) were associated with BOLD activation in anatomical areas forming memory circuitry. Whalley et al. (2013): Using a GWAS-generated polygene risk score as a genetic QT, associations were found with decreased FA in high-risk subjects for developing MDD but not BD. Whalley et al. (2012): Using a GWAS-generated polygene score as a genetic QT, positive associations were found between increased activation in the subgenual anterior cingulate during a sentence completion task. Martin et al. (2014): Using a GWAS-generated total deletion burden as a genetic QT, associations were found with decreased connectivity in cognitive control and default mode network (DMN). Liu et al. (2012): The fMRI auditory oddball component is positively correlated (red to yellow) with the SNP component. Meda et al. (2014): Topology of ICA-derived DMNs demonstrating between-group differences that are subsequently correlated with an independent GWAS ICA-derived top SNPs. These images have been reproduced with permission from Elsevier, Nature, and PNAS.

Adapted with permission from Appendix A: Supplementary Materials in Sprooten et al. (2013) *White matter integrity as an intermediate phenotype: Exploratory genome-wide association analysis in individuals at high risk of bipolar disorder.* Psychiatry Research.

Adapted with permission from Fig. 1 in Whalley et al. (2013) *Polygenic Risk and White Matter Integrity in Individuals at High Risk of Mood Disorder.* Biological Psychiatry.

Reproduced with permission from Fig. 3 in Potkin et al. (2009c) *Gene discovery through imaging genetics: identification of two novel genes associated with schizophrenia.* Molecular Psychiatry.

Adapted with permission from Fig. 1 in Whalley et al. (2012) *The influence of polygenic risk for bipolar disorder on neural activation assessed using fMRI.* Translational Psychiatry.

Reproduced with permission from Fig 3 in Liu et al. (2012) *An ICA with reference approach in identification of genetic variation and associated brain networks.* Frontiers in Human Neuroscience.

Adapted with permission from Fig. 4 in Martin et al. (2014) *Copy number deletion burden is associated with cognitive, structural, and resting-state network differences in patients with schizophrenia.* Behavioural Brain Research.

Adapted with permission from Fig. 2 in Meda et al. (2014) *Multivariate analysis reveals genetic associations of the resting default mode network in psychotic bipolar disorder and schizophrenia.* Proceedings of the National Academy of Sciences USA.

Although the previous studies have convincingly shown that the RPS has a highly significant capacity in predicting case-control status, supporting the polygenic hypothesis of schizophrenia, the risk score profiling analysis has a number of limitations. First, since the RPS is a composite score comprising the contributions of many risk alleles, it offers little insight as to which genes or pathways are involved in the disease, thereby negating the purpose of GWAS. Second, the risk variances explained by RPSs in the published schizophrenia studies are typically small (e.g., Nagelkerke pseudo $R^2 = 0.06$ in the study by Ripke et al. 2013), and the scores are insufficient for diagnostic purposes (the area under ROC curves < 0.65 in the study by Ripke et al. 2013). Similarly, in a smaller GWAS of 1,420 postpartum depression cases and 9,473 controls, no significant SNPs were found, although an RPS analysis produced statistically significant results ($p < 0.004$), but the Nagelkerke pseudo R^2 was less than 1% (Byrne et al. 2014).

GENETIC RISK SCORE

The genetic risk score (GRS) is another composite measure of risk genes (as represented by SNPs), representing cumulative genetic risk as opposed to single genes. Like the RPS, the GRS has the potential to explain more variance with greater statistical power than individual SNPs. However, the GRS, in contrast to the RPS, only uses genome-wide-significant SNPs to create a unique predictor representing cumulative shared schizophrenia and bipolar genetic risk. Additional statistical power can be achieved when a GRS is combined with quantitative imaging or cognitive phenotypes as contrasted with a categorical diagnosis (e.g., case versus control) analysis strategy (Potkin et al. 2009a).

There are several methods to calculate a GRS. We summarize the method described by Walton et al. (2013) as an example in which SNPs are selected based on the continuously updated meta-analysis of genetic studies on schizophrenia available at www.schizophreniaresearchforum.org, using the latest update at the time of analysis. Additional data, such as those from the latest PGC results, can also be included. For SNPs in linkage disequilibrium (LD; $r^2 > .8$), SNPs with the smallest *p*-value are chosen. An additive GRS for each subject is calculated (Walton et al. 2013):

$$\mathbf{GRS} = \sum_{i \in \{1,\ldots,41\}} wiXi,$$

where w is the log-transformed odds ratio (OR) for each SNP with $w = \ln(\mathrm{OR}_{\mathrm{SNP}})$ and X is the number of risk alleles (0, 1, or 2). Stratification is assessed with PLINK and any needed adjustments and/or any censored analyses are performed. In large heterogeneous samples, it is essential that admixture be assessed and addressed. Walton et al. (2013) demonstrated that the GRS has "good" power for association with brain activation in a sample of schizophrenia ($n = 79$) and control subjects ($n = 99$).

PATHWAY POLYGENIC RISK SCORE

A pathway polygenic risk score (PPRS) is another way of calculating a composite risk score for each individual. Instead of using all SNPs passing a specific quality control, the PPRS focuses on SNPs involved in a specific pathway (Nicodemus et al. 2014). Similar to the RPS, the PPRS analysis is also carried out using two sets of samples—a discovery sample for computing association *p*-values and selecting risk alleles, and a target sample for testing the association of the PPRS with the phenotype. Given a *p*-value threshold and a specific pathway, the PPRS for each individual in the target sample is calculated to be the number of risk alleles that pass the *p*-value threshold in the discovery sample and are located in the genes belonging to the given pathway, weighted by the log-odd ratio in the discovery sample.

Nicodemus et al. (2014) used the Psychiatric Genomic Consortium to calculate a PPRS associated with the ZNF804A pathway, and correlated it with cognitive measures, a QT trait. Individual ZNF804A SNPs have been previously associated with cognition. Higher ZNF804A pathway polygenic scores were associated with poorer cognitive performance, explaining 1%–3% of the variation. The percentage of variation explained increased to 4.8% when two-way SNP-by-SNP interactions were added in the model. To reduce multiple testing constraints, the 2x2 epistatic interactions were limited to SNPs significantly associated with cognitive performance. Such epistatic interactions can be evaluated. The dopamine transporter (DAT) and glutamate-related G72 genes (a.k.a. *DAOA*) both have been independently identified in schizophrenia and bipolar disorder, and are hypothesized to play an interactive role in schizophrenia. DAT genotype (DAT 10/10 VNTR in the 3′UTR) increased fMRI activation in the postcentral gyrus, while the G72 genotype (GG; G72 SNP rs746187) increased widespread activation during the "hard" version of the verbal fluency task. The G72 GG and DAT 10/10-repeat subjects had the greatest activation in the hippocampus and basal ganglia, while the G72 AA and DAT 9-repeat carriers had the lowest activations, with the heterozygotes being intermediate. In general, significant

gene-gene interaction effects result in larger effect sizes than single SNPs, thereby achieving significance with a smaller number of subjects (Pauli et al. 2013).

The composite risk scores described here provide ways of summarizing genetic effects among an ensemble of markers that individually might give rise to very small significance in genome-wide association studies. There are pros and cons associated with each of these methods. The RPS uses all SNPs passing quality control and *p*-value thresholds, and thus offers an unbiased and systematic approach for testing the cumulative effects of common variants in associating with phenotypes. However, because the score is composed of a large number of genetic markers, it offers little insight on genes and biological mechanisms underlying the association. The PPRS directly tests the association between markers within a pathway and phenotypes. As such, it is most useful in revealing genes and pathways involved in a significant association. However, application of the method requires a priori knowledge of pathways or genes involved in the disease. The GRS uses previously discovered genome-wide significant SNPs in calculating the composite risk score. Although it is easy to use, it may have limited power in explain genetic risks if most of the genetic markers have yet to be discovered, which is often the case for complex psychiatric diseases.

By combining the Swedish National Sample, the Psychiatric Genomic Consortium (PGC), and the replicated schizophrenia samples, Ripke et al. (2013) estimated that 8,300 independent SNPs contributed to the risk for schizophrenia and collectively accounted for 32% of the variance in risk liability. These results may begin to address the missing heritability in schizophrenia, but leave us with the daunting task of finding biological mechanisms of action for such a large number of implicated genes. Current large GWAS have identified many SNPs associated with schizophrenia. Given the relatively large confidence interval for each of these SNPs, most of the entire genome lies within these confidence intervals (Consortium SWGotPG 2014). If this is accurate and not a reflection of measurement imprecision, it challenges our biological interpretation of the findings. Brain imaging may represent an approach that can direct and refine our interpretation of such findings. Methods like brain imaging genetics, compared to case-control studies, may more readily lend themselves to biologically sound interpretation and follow-up studies for identification of true culprit genes and their functional consequences. The functional consequences of SNPs associated with regional brain imaging phenotypes can be examined at multiple levels of investigation, including regional expression of these associated genes in postmortem tissue, and investigation of their functional roles in cellular and animal models.

GENE NETWORK AND PATHWAY ANALYSIS

Gene regulatory networks can be inferred from expression profiles, the location of regulatory motifs, and microRNAs (miR) and their targets (Subramanian et al. 2005). We adapted the GSEA (gene-set enrichment analysis) originally developed for gene expression analyses to identify gene sets or pathways that contribute to schizophrenia based on SNP data (Potkin et al. 2010). Essentially, the GSEA determines whether a group of genes is over-enriched with SNPs associated with a disease or a disease trait (e.g., cognition or brain activation). Gene sets are based on prior biological knowledge and can include canonical pathways, genetic perturbations, microRNA targets, and transcription factors (Potkin et al. 2010). Our GSEA application with brain imaging as a QT identified several microRNAs associated with schizophrenia for the first time and in particular miR-137 (Potkin et al. 2010), which has been subsequently confirmed in very large case-control studies and has been described as an "etiological mechanism for schizophrenia" (Consortium SPG-WASG 2011; Ripke et al. 2013).

A challenge is how to combine SNP evidence of association over multiple SNPs within a gene and multiple genes within a pathway. Most methods ignore the joint action of multiple SNPs within a gene. The SSEA (SNP-set enrichment analysis) method we developed considers the joint action of multiple SNPs within a gene, controlling for gene length, and we have applied this method to schizophrenia case-control GWAS data (Weng et al. 2011). It uses an adaptive truncated product statistic to address the potentially confounding feature that some genes are represented by many more SNPs than others. We applied this method to European and African American case-control GWAS data, finding eight important pathways common to both. An advantage to this method is that many canonical databases can be applied, though currently there is no metric for choosing among them. Our method also avoids considering individual SNPs by themselves, as both SNPs and genes operate within biological pathways. These pathways can interact with one another as part of biological networks.

Table 9.1's imaging genetics analyses implicate several pathways with schizophrenia and/or bipolar disorder,

although the results are not entirely consistent and should be reviewed as preliminary. In summary, the two fractional anisotropy (FA) studies, which measure white matter integrity, found decreased FA in family members of bipolar disorder (Sprooten et al. 2013) or unipolar depression (Whalley et al. 2013) subjects. The biological pathways implicated relate to axonal guidance and brain development. The two rsfMRI studies found that the resting state default mode network was associated with increased deletion burden (Martin et al. 2014) in pathways related to axon guidance, synaptogenesis, NMDA, and immune signaling (Meda et al. 2014). Task-related fMRI findings were generally related to brain areas activated by the specific task. Motor task activation revealed pathways related to GABA, dopamine, and glutamate signaling (Chen et al. 2012). The auditory oddball discrimination task was associated with Wnt signaling and metabotropic glutamate receptor signaling (Liu et al. 2012). Cognitive control was related to hippocampal and cholinergic function as well as cell clearance mechanisms (Rietschel et al. 2012). Amygdala activation to faces was associated with serotonin regulation, MAP kinase, and neuronal development pathways (Liu et al. 2010; Ousdal et al. 2012). DLPFC activation during a working memory task was associated with cell differentiation and neuronal development, stress response, and D1 receptor functioning in schizophrenia (Potkin et al. 2009b; Potkin et al. 2009c). The RPS was associated with increased activation in the anterior cingulate and amygdala during a sentence completion task (Whalley et al. 2012) and decreased FA (Whalley et al. 2013) in bipolar high-risk subjects. However, RPS analyses do not allow identification of particular pathways involved in the disorder.

A recently developed statistical method (Bayes FDR GWAS) for analyzing GWAS data demonstrated that functional genic elements differentially contribute to phenotypic variance (e.g., 5' UTR > exon) (Schork et al. 2013). Using this approach and including pleiotropy in the analysis, Andreassen et al. (2013) greatly increased the discovery rate for loci related to both schizophrenia and bipolar disorder by approximately 15-fold. This method's advantage includes covariate modulated local FDR-corrections to efficiently re-rank SNPs based on functional annotation priors (Zablocki et al. 2014). The weightings from this Bayes FDR approach can be incorporated in the SSEA. An SSEA based on QT GWAS SNPs will also benefit from the additional statistical power of a QT analysis (> 4-fold), as we have demonstrated (Potkin et al. 2009a; Shen et al. 2014).

BRAIN CIRCUITRY AS A QUANTITATIVE TRAIT

An alternative approach to examining the functional consequences of case-control GWAS-identified SNPs/genes begins with brain circuitry as a quantitative phenotype (e.g., correlation between DLPFC activation and hippocampal activation on fMRI) and identifies the SNPs/genes that contribute to a given imaging phenotype (Potkin et al. 2010). Abnormal circuitry may be closer to the underlying disease mechanism than the abnormal structure or function of any single ROI. There may be multiple ways to affect dysfunctional brain circuitry. Circuitry thus may infer greater risk than individual components of the circuitry. Similarly, pathways may infer greater risk than the individual genes that comprise them. Nevertheless, identification of genes, pathways, or networks with these approaches will need follow-up experiments to pinpoint the molecular mechanisms involved, and to identify new therapeutic targets.

REPLICATION

Replication of any of the current results in independent samples remains of critical importance for confirmation. Replication involves complexities that are not readily apparent. In basic science, replication implies testing for the same results when all parameters are held constant, such as sex, animal strain, developmental stage, experimental procedures, and analyses. This is rarely the case in clinical studies in which patient characteristics are typically not standardized. Clinical samples typically differ by geography, comorbidities, diet, and treatment. Additionally, imaging task paradigms are rarely standardized or reported in sufficient detail to allow standardization across studies. Finally, the genetic background of the subjects is typically heterogeneous (Evans and Cardon 2005; Evangelou and Ioannidis 2013). Given these sources of variability, heterogeneity in results is not unexpected.

However, lack of convergent results on the same SNPs across different samples or populations may nonetheless converge on the same network or pathway. One example is the convergence of results from two independent GWA imaging genetics studies that examined DLPFC inefficiency (see Table 9.2). MiR-137, recently identified as a risk factor for schizophrenia (Consortium SPG-WASG 2011), did not appear as a significant finding in either of the two previous imaging GWAS. Nevertheless, when the data

TABLE 9.2 TOP 12 GENE SETS ENRICHED WITH MOST SIGNIFICANT P-VALUE GENES IN TWO INDEPENDENT IMAGING GENETICS GWAS DATA SETS: SCZ DATA SET *N* = 24 (POTKIN ET AL., 2009B) AND SIRP DATA SET *N* = 138 (POTKIN ET AL., 2009C)

Gene set category	Gene Set Name	SIRP Data Set*		SIRP Data Set*		Number of genes
		Mann–Whitney *Z*-score	KS −log10 (*P* value)	Mann–Whitney *Z*-score	KS −log10 (*P* value)	
c3.mir	ATATGCA,MIR-448	8.8657	17.1214	6.9348	9.7789	177
c2.cgp	UVC_TTD_ALL_DN	8.6874	12.9778	7.4691	9.9803	318
c2.cgp	UVC_XPCS_ALL_DN	8.1193	12.5773	7.9788	12.225	419
c2.cgp	UVC_XPCS_8HR_DN	7.6156	10.3729	7.5681	10.7036	358
c2.cgp	UVC_TTD_8HR_DN	7.3774	8.0467	6.4858	6.8742	144
c2.cgp.	UVC_TTD_4HR_DN	7.3072	10.244	6.1085	7.1162	262
c3.tft	V$OCT1_03	6.783	8.168	2.6201	1.6422	147
c2.cgp	UVC_XPCS_4HR_DN	6.6389	9.8161	5.7022	7.1296	208
c2.cgp	BAF57_BT549_UP	6.6375	9.8099	5.2192	4.502	203
c3.mir	AAGCACA,MIR-218	6.5172	7.8408	6.9509	11.0283	332
c3.tft	YNGTTNNNATT_UNKNOWN	5.9645	6.611	7.3309	12.7903	235
c3.mir	AAGCAAT,MIR-137	5.8681	7.2371	6.5203	7.6126	168

Gene set codes: c2 refers to curated gene sets from canonical pathways and chemical and genetic perturbations; c3 refers to motif gene sets in MSigDB. MicroRNA data sets are indicated by .mir.

Adapted with permission from Table 2 in Potkin et al (2010) *Identifying Gene Regulatory Networks in Schizophrenia.* Neuroimage.

sets from the two independent studies were included in a GSEA, the miR-137 target gene set appeared as a statistically significant risk factor for schizophrenia, despite the fact that miR-137 was not identified in either study individually based on direct association (the rs1625579 SNP identified by Ripke et al. [2013] was not present on the SNP chip used by the studies). More recently, we genotyped the rs1625579 SNP in one of our samples and showed that it was associated with DLPFC hyperactivation during working memory performance, a known schizophrenia risk phenotype (van Erp et al. 2014).

RESTING STATE

We have discussed the use of task-related brain imaging activation patterns and their use in imaging genetics studies to discover unanticipated genetic variants related to SMI. Brain activity can be measured in a non-task condition (a.k.a. at rest) and similarly used as an imaging phenotype. Resting state functional connectivity identifies networks of coactivated brain areas. These analyses can be performed using a seed ROI to explore the temporally based correlations with other brain areas (other ROIs or whole brain). The choice of seeds raises the limitations and advantages of using prior knowledge (Ford et al. 2015). Data-driven (e.g., ICA [independent components analysis] and graph-based) methods are also employed to identify resting state networks and to compare functional brain connectivity between groups (Damaraju et al. 2014; Tomasi and Volkow 2014). The area of investigation that attempts to identify genetic variation associated with structural and functional brain connectivity has been coined "connectome" genetics, and several candidate gene imaging genetics association studies have recently been reviewed by Thompson et al. (2013). In addition to the candidate gene studies, a recent whole brain analysis showed significant associations between copy number deletion burden and resting state functional connectivity (Martin et al. 2014). Finally, a first GWAS using parallel ICA with rsfMRI as the quantitative phenotype in a schizophrenia/bipolar sample was recently published, and it identified pathways related to axon guidance, synaptogenesis, NMDA, and immune signaling (Meda et al. 2014).

MULTIVARIATE METHODS

An area of intense research is the development of multivariate methods that jointly consider the variability of phenotypes (a.k.a., the brain imaging phenotypes) and genotypes. These analytical strategies are usually collectively referred to as data reduction methods. Among them, a well-known procedure is the principal component analysis (PCA),

> a mathematical algorithm that reduces the dimensionality of the data while retaining most of the variation in the data set. It accomplishes this reduction by identifying directions, called principal components, along which the variation in the data is maximal. By using a few components, each sample can be represented by relatively few numbers instead of by values for thousands of variables. Samples can then be plotted, making it possible to visually assess similarities and differences between samples and determine whether samples can be grouped. (Ringner 2008)

Despite its appealing properties, however, PCA has not been largely accepted by the genetics community. Its current applications are primarily limited to exploratory analyses and quality control data performed at the very early stages of genomics analyses.

PARALLEL INDEPENDENT COMPONENT ANALYSIS

The ICA (independent component analysis) is a blind source separation technique that can separate multivariate data (signals) into independent additive subcomponents. When applied to fMRI time-series data, it identifies independent spatiotemporal components, such as sets of voxels that show time series BOLD fluctuations associated with a presented task (e.g., checkerboard versus rest), fluctuations during rest, fluctuations associated with artifacts (e.g., motion), and even physiology (e.g., cardiac pulsation). A parallel ICA performs two separate ICAs, one on imaging data and one on genetic data. It then examines the correlation between the two parallel ICAs, creating fMRI-SNP pairs. Dimension reduction of SNPs is necessary to optimize the parallel ICA (Liu et al. 2012). Although not required, the selection of SNPs can be reduced based on LD, thresholded p-values, or published genetic SZ susceptibility genetic studies. Chen et al. (2012), using parallel ICA in 92 SZ and 116 HC, identified pathways involving GABA, dopamine, neuregulin, and glutamate receptor signaling that were correlated with pre-central and post-central gyri activation during a motor task. Interestingly, a number of SNPs that were significant in the analysis were unanticipated, requiring replication but supporting the value of data-driving approaches in generating novel findings.

A last note of caution relates to the meaning of the association with any given genetic variant. While the results are usually presented and discussed with a major emphasis on the given variant found significant, like a SNP where, for example, the nucleotide "A" rather than the original "G" is associated with the trait under investigation, such description of results does not completely address the underlying biological reality. Since humans have a diploid genome, it is of critical importance to consider the effects of both chromosomes, that is, the specific genotypes rather than the alleles in isolation. There is debate as to whether an additive model with three levels of genotype should be used (e.g., AA > AG > GG) or whether two levels are appropriate when a "dominant" model of transmission is invoked for the phenotype or there is a small number of homogeneous individuals (e.g., AA > AG+GG or AA+AG > GG). While presenting genotypes is more complete than presenting "allelic" associations, the interpretation of genetic association is more complex when considering the effects of multiple genes and genotypes. Risk genes do not function in isolation; they interact with one another, and function in multiple networks and pathways.

SUMMARY

There are at least two major approaches for understanding the function of genes at the anatomical or brain circuitry level. One can begin with genetic studies of individuals or families in which risk SNPs whose frequencies are different between schizophrenia and/or bipolar probands versus controls are identified. The functional role of these top SNPs can be clarified by association studies using functional imaging. The PGC (Consortium C-DGotPG 2013; Consortium SWGotPG 2014) and other consortia have usefully identified more than 100 risk genes. An alternative approach is to begin with imaging phenotypes that characterize schizophrenia and/or bipolar circuitry abnormalities and then to determine which genes influence this phenotype. This approach has been used to identify brain circuitry associated genetic risk candidates, many of which were unanticipated (Potkin et al. 2009b; Potkin et al.

2010). Confirmation can be achieved through additional studies that test the association of imaging phenotypes with the candidates (van Erp et al. 2014), or by determining the frequency of these new candidates in large standard case-control genetic studies (Consortium SPG-WASG 2011; Consortium SWGotPG 2014). A third approach is to study the genetics and imaging in disorders that have a known risk for the development of schizophrenia and/or bipolar disorder, such as 22Q deletion syndrome, which has a risk of 25%–30% for schizophrenia and has also been linked to bipolar disorder (Papolos et al. 1996).

To discover the underlying mechanisms of SMI, we need more than one approach. Candidate gene studies are essential; however, they are limited based on our current but incomplete knowledge. Data-driven approaches can identify unanticipated genetic risk variants. Imaging genetic approaches can link genetic risk factors with brain structural and functional effects. Jointly studying imaging and genomic data leverages the power of both brain imaging and genomics, recognizing that risk genes must affect brain function and that brain structure and function have strong genetic components. Finally, identifying risk genes is only the first step, as the effects of these risk genes and the networks in which they participate must be understood at their molecular basis to develop new and more effective therapeutics. The goal of precision medicine will be met by identifying the risk factors present in individuals, as opposed to groups of individuals. Such future studies will allow the development of mechanistically based targeted therapies.

ACKNOWLEDGMENTS

Research reported in this publication was supported by the National Institutes of Health under award numbers R01HG006870 and U01AG024904. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

Andreassen OA, Thompson WK, Schork AJ, et al. (2013, Apr). Improved detection of common variants associated with schizophrenia and bipolar disorder using pleiotropy-informed conditional false discovery rate. *PLoS Genet.* 9(4): e1003455.

Barsh GS, Copenhaver GP, Gibson G, Williams SM. (2012, Jul). Guidelines for genome-wide association studies. *PLoS Genet.* 8(7): e1002812.

Benjamini Y, Hochberg Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Statis Soc B (Methodological).* 57(1): 289–300.

Byrne EM, Carrillo-Roa T, Penninx BW, et al. (2014, Jul). Applying polygenic risk scores to postpartum depression. *Arch Womens Ment Health.*

AQ1: Please provide vol and pp nos.

Chen J, Calhoun VD, Pearlson GD, et al. (2012, Jul). Multifaceted genomic risk for brain function in schizophrenia. *Neuroimage.* 61(4): 866–875.

Consortium C-DGotPG. (2013, Apr). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet.* 381(9875): 1371–1379. See article: http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(12)62129-1/abstract

Consortium SPG-WASG. (2011, Oct). Genome-wide association study identifies five new schizophrenia loci. *Nat Genet.* 43(10): 969–976. See article: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3303194/

Consortium SWGotPG. (2014, Jul). Biological insights from 108 schizophrenia-associated genetic loci. *Nature.* 511(7510): 421–427. See article: http://www.nature.com/nature/journal/v511/n7510/full/nature13595.html

Damaraju E, Allen EA, Belger A, et al. (2014). Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia. *Neuroimage Clin.* 5: 298–308.

Dudbridge F, Gusnanto A. (2008, Apr). Estimation of significance thresholds for genomewide association scans. *Genet Epidemiol.* 32(3): 227–234.

Dudbridge F, Gusnanto A, Koeleman BP. (2006, Mar). Detecting multiple associations in genome-wide studies. *Hum Genomics.* 2(5): 310–317.

Evangelou E, Ioannidis JP. (2013, Jun). Meta-analysis methods for genome-wide association studies and beyond. *Nat Rev Genet.* 14(6): 379–389.

Evans DM, Cardon LR. (2005, Apr). A comparison of linkage disequilibrium patterns and estimated population recombination rates across multiple populations. *Am J Hum Genet.* 76(4): 681–687.

Ford JM, Palzes VA, Roach BJ, et al. (2015, Jan). Visual hallucinations are associated with hyperconnectivity between the amygdala and visual cortex in people with a diagnosis of schizophrenia. *Schizophr Bull.* 41(1): 223–232. doi: 10.1093/schbul/sbu031

Harrison PJ, Weinberger DR. (2005, Jan). Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence. *Mol Psychiatry.* 10(1): 40–68; image 45.

Liu J, Ghassemi MM, Michael AM, et al. (2012). An ICA with reference approach in identification of genetic variation and associated brain networks. *Front Hum Neurosci.* 6: 21.

Liu X, Akula N, Skup M, Brotman MA, Leibenluft E, McMahon FJ. (2010, Jan). A genome-wide association study of amygdala activation in youths with and without bipolar disorder. *J Am Acad Child Adolesc Psychiatry.* 49(1): 33–41.

Martin AK, Robinson G, Reutens D, Mowry B. (2014, Oct). Copy number deletion burden is associated with cognitive, structural, and resting-state network differences in patients with schizophrenia. *Behav Brain Res.* 272: 324–334.

Meda SA, Ruano G, Windemuth A, et al. (2014, May). Multivariate analysis reveals genetic associations of the resting default mode network in psychotic bipolar disorder and schizophrenia. *Proc Natl Acad Sci U S A.* 111(19): E2066–2075.

Nicodemus KK, Hargreaves A, Morris D, et al. (2014, Jul). Variability in working memory performance explained by epistasis vs polygenic scores in the ZNF804A pathway. *JAMA Psychiatry.* 71(7): 778–785.

Ousdal OT, Anand Brown A, Jensen J, et al. (2012, Jun). Associations between variants near a monoaminergic pathways gene (PHOX2B)

and amygdala reactivity: a genome-wide functional imaging study. *Twin Res Hum Genet.* 15(3): 273–285.

Pan Q. (2013). Multiple hypotheses testing procedures in clinical trials and genomic studies. *Front Public Health.* 1:63.

Papolos DF, Faedda GL, Veit S, et al. (1996, Dec). Bipolar spectrum disorders in patients diagnosed with velo-cardio-facial syndrome: does a hemizygous deletion of chromosome 22q11 result in bipolar affective disorder? *Am J Psychiatry.* 153(12): 1541–1547.

Pauli A, Prata DP, Mechelli A, et al. (2013, Sep). Interaction between effects of genes coding for dopamine and glutamate transmission on striatal and parahippocampal function. *Hum Brain Mapp.* 34(9): 2244–2258.

Potkin SG, Macciardi F, Guffanti G, et al. (2010, Nov). Identifying gene regulatory networks in schizophrenia. *Neuroimage.* 53(3): 839–847.

Potkin SG, Turner JA, Guffanti G, et al. (2009a). Genome-wide strategies for discovering genetic influences on cognition and cognitive disorders: methodological considerations. *Cogn Neuropsychiatry.* 14(4–5): 391–418.

Potkin SG, Turner JA, Guffanti G, et al. (2009b, Jan). A genome-wide association study of schizophrenia using brain activation as a quantitative phenotype. *Schizophr Bull.* 35(1):96–108.

Potkin SG, Turner JA, Fallon JA, et al. (2009c, Apr). Gene discovery through imaging genetics: identification of two novel genes associated with schizophrenia. *Mol Psychiatry.* 14(4): 416–428.

Purcell S, Cherny SS, Sham PC. (2003, Jan). Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics.* 19(1): 149–150.

Purcell SM, Wray NR, Stone JL, et al. (2009, Aug). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature.* 460(7256): 748–752.

Rietschel M, Mattheisen M, Degenhardt F, et al. (2012, Sep). Association between genetic variation in a region on chromosome 11 and schizophrenia in large samples from Europe. *Mol Psychiatry.* 17(9):906–917.

Ringner M. (2008, Mar). What is principal component analysis? *Nat Biotechnol.* 26(3):303–304.

Ripke S, O'Dushlaine C, Chambert K, et al. (2013, Oct). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Genet.* 45(10):1150–1159.

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

Ruderfer DM, Fanous AH, Ripke S, et al. (2014, Sep). Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. *Mol Psychiatry.* 19(9): 1017–1024. doi: 10.1038/mp.2013.

Schork AJ, Thompson WK, Pham P, et al. (2013, Apr). All SNPs are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. *PLoS Genet.* 9(4): e1003449.

Shen L, Thompson PM, Potkin SG, et al. (2014, Jun). Genetic analysis of quantitative phenotypes in AD and MCI: imaging, cognition and biomarkers. *Brain Imaging Behav.* 8(2): 183–207.

Sprooten E, Fleming KM, Thomson PA, et al. (2013). White matter integrity as an intermediate phenotype: Exploratory genome-wide association analysis in individuals at high risk of bipolar disorder. *Psychiatry Research.* 206: 223–231.

Stefansson H, Ophoff RA, Steinberg S, et al. (2009, Aug). Common variants conferring risk of schizophrenia. *Nature.* 460(7256): 744–747.

Stein JL, Medland SE, Vasquez AA, et al. (2012, May). Identification of common variants associated with human hippocampal and intracranial volumes. *Nat Genet.* 44(5): 552–561.

Stranger BE, Stahl EA, Raj T. (2011, Feb). Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics.* 187(2): 367–383.

Subramanian A, Tamayo P, Mootha VK, et al. (2005, Oct). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 102(43): 15545–15550.

Tesli M, Espeseth T, Bettella F, et al. (2014, Jun). Polygenic risk score and the psychosis continuum model. *Acta Psychiatr Scand.* 130(4): 311–317. doi: 10.1111/acps.12307.

Thompson PM, Ge T, Glahn DC, Jahanshad N, Nichols TE. (2013, Oct). Genetics of the connectome. *Neuroimage.* 80: 475–488

Tomasi D, Volkow ND. (2014). Mapping small-world properties through development in the human brain: disruption in schizophrenia. *PLoS One.* 9(4): e96176.

van Erp TG, Guella I, Vawter MP, et al. (2014, Mar). Schizophrenia miR-137 locus risk genotype is associated with dorsolateral prefrontal cortex hyperactivation. *Biol Psychiatry.* 75(5): 398–405.

Walton E, Geisler D, Hass J, et al. (2013). The impact of genome-wide supported schizophrenia risk variants in the neurogranin gene on brain structure and function. *PLoS One.* 8(10): e76815.

Walton E, Turner J, Gollub RL, et al. (2013, May). Cumulative genetic risk and prefrontal activity in patients with schizophrenia. *Schizophr Bull.* 39(3): 703–711.

Wang Q, Charych EI, Pulito VL, et al. (2011, Oct). The psychiatric disease risk factors DISC1 and TNIK interact to regulate synapse composition and function. *Mol Psychiatry.* 16(10): 1006–1023.

Weng L, Macciardi F, Subramanian A, et al. (2011). SNP-based pathway enrichment analysis for genome-wide association studies. *BMC Bioinformatics.* 12: 99.

Whalley HC, Papmeyer M, Sprooten E, et al. (2012). The influence of polygenic risk for bipolar disorder on neural activation assessed using fMRI. *Transl Psychiatry.* 2: e130.

Whalley HC, Sprooten E, Hackett S, et al. (2013, Aug). Polygenic risk and white matter integrity in individuals at high risk of mood disorder. *Biol Psychiatry.* 74(4): 280–286.

Xu C, Tachmazidou I, Walter K, Ciampi A, Zeggini E, Greenwood CM. (2014, May). Estimating genome-wide significance for whole-genome sequencing studies. *Genet Epidemiol.* 38(4): 281–290.

Zablocki RW, Schork AJ, Levine RA, Andreassen OA, Dale AM, Thompson WK. (2014, Aug). Covariate-modulated local false discovery rate for genome-wide association studies. *Bioinformatics.* 30(15): 2098–2104.