
This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: http://openaccess.city.ac.uk/15081/

Link to published version: http://dx.doi.org/10.1177/0269881115584470

Copyright and reuse: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

City Research Online: http://openaccess.city.ac.uk/ publications@city.ac.uk
Polygenic risk scores in imaging genetics: usefulness and applications.

Danai Dima\textsuperscript{1,2} and Gerome Breen\textsuperscript{1,3}

\textsuperscript{1} MRC SGDP Centre, Institute of Psychiatry, Psychology and Neuroscience, King’s College London, London, UK

\textsuperscript{2} Clinical Neuroscience Studies (CNS) Center, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, USA

\textsuperscript{3} National Institute of Health Research (NIHR) Biomedical Research Centre for Mental Health, South London and Maudsley NHS Trust and Institute of Psychiatry, London, UK.

\textbf{Corresponding author:} Danai Dima, MRC SGDP Centre, Institute of Psychiatry, Psychology and Neuroscience PO80, King’s College London, 16 De Crespigny Park, London SE5 8AF, UK

E-mail: danai.dima@kcl.ac.uk

Tel: + 44 (0)20 7848 0856
Abstract

Genetic factors account for up to 80% of the liability for schizophrenia and bipolar disorder. Genome-wide association studies (GWAS) have successfully identified several single nucleotide polymorphisms (SNPs) and genes associated with increased risk for both disorders. Single SNP analyses alone do not address the overall genomic or polygenic architecture of psychiatric disorders as the amount of phenotypic variation explained by each GWAS-supported SNP is small whereas the number of SNPs/regions underlying risk for illness is thought to be very large. The polygenic risk score models the aggregate effect of alleles associated with disease status present in each individual and allows us to utilize the power of large GWAS to be applied robustly in small samples. Here we make the case that risk prediction, intervention and personalized medicine can only benefit with the inclusion of polygenic risk scores in imaging genetics research.

Keywords: imaging genetics; polygenic risk score; schizophrenia; bipolar disorder; psychiatry
Psychiatric disorders are highly heritable, especially schizophrenia and bipolar disorder. Heritability is the proportion of trait variation that is accounted for by genetic factors, as opposed to trait variation due to environmental factors. Twin and family studies for schizophrenia have placed heritability around 0.8 (Sullivan et al., 2003; Lichtenstein et al., 2009) and between 0.6 to 0.85 for bipolar disorder (McGuffin et al., 2003; Craddock and Sklar, 2009). After the human genome was first sequenced the genetics field has produced huge technological strides towards estimating heredity (Lander et al., 2001). The most common genetic marker is the single nucleotide polymorphism (SNP), a single subunit change in the DNA sequence. SNPs are located on a DNA chip, which is used to create a genetic profile for each individual. In turn DNA chips are used in genome-wide associations studies (GWAS) that estimate the heritability of a disorder throughout the genome by comparing healthy controls with patients, in a hypothesis-free way. The first disease upon which the GWAS technique was used was age-related macular degeneration in 2005 (Klein et al., 2005), and since then many genes have been implicated in macular degeneration (Fritsche et al., 2013).

After the first GWAS studies in schizophrenia it soon became apparent there were no common variants that had a large influence on risk, but rather that there were thousands of variants of very small effect that together acted to increase or reduce risk. The first signs of success in schizophrenia started to emerge in 2008 with a GWAS
study that implicated the ZNF804A gene that almost reached the significance level $1.61 \times 10^{-7}$ (O’Donovan et al., 2008). In 2009, the SGENE consortium (Stefansson et al., 2009) and the International Schizophrenia Consortium (Shi et al., 2009) published papers on other variants associated with the schizophrenia. These groups then combined their efforts and began publishing under the Psychiatric Genomics Consortium banner with papers in 2011 detailing five variants and 2013 with 13 variants (Ripke et al., 2011; 2013). However, real progress began when the sample size exceeded 15,000 cases or so, with an inflection point being reached where, instead of discovering one variant every couple of thousand cases, variants began to be discovered when, on average, another 250 cases and 250 controls were added to the sample size. Thus, once the size of the sample had increased to >15,000 cases, the power to detect variants reached a point where GWAS could detect the effects of the modest size at which lots of schizophrenia-associated variants are operating.

Then in 2014, Ripke and colleagues (Ripke et al., 2014) published 108 regions associated with the risk of schizophrenia. In this analysis that included over 36,989 patients and 113,075 healthy controls, few if any candidate genes replicated; for example COMT, a gene implicated in the risk of schizophrenia and extensively used in imaging genetics had a $p$-value of 0.56 (Ripke et al., 2014). The same was true for the BDNF gene, with a $p$-value of 0.006. Interestingly the BDNF gene is significantly
associated with smoking and obesity (Hallden et al., 2013), attributes with increased prevalence in schizophrenia (Hennekens et al., 2005). Other genes that had not been identified from previous schizophrenia studies were identified in the PGC2 paper and showed significant association with schizophrenia, such as the DRD2, a target for antipsychotic drugs, and the GRM3 involved in glutamatergic transmission (Ripke et al., 2014).

Although genetic research in bipolar disorder is less published upon than in schizophrenia, susceptibility genes for bipolar disorder have also been identified through GWAS (Alsabban et al., 2011). The most robustly associated SNPs are located in the CACNA1C (Ferreira et al., 2008; Sklar et al., 2008; 2011), the ITIH3 (Sklar et al., 2011), the TRANK1 (Chen et al., 2013), the NCAN (Cichon et al., 2011), the ANK3 (Ferreira et al., 2008; Sklar et al., 2008; Schulze et al., 2009; Scott et al., 2009; Smith et al., 2009; Athanasiu et al., 2010; Tesli et al., 2011), the SYNE1 (Sklar et al., 2011; Green et al., 2013) and the ODZ4 gene (Sklar et al., 2011).

The question that arises is how to utilize the predictive power of GWAS findings. The predictive power of genetic risk prediction is not yet useful for clinical applications but where genetic prediction may have utility is acting as biomarkers for schizophrenia and bipolar disorder traits within individuals in, e.g. imaging studies. A key hypothesised benefit of brain imaging approaches is their ability to decipher if and
how genetic risk factors impact brain structure and function, especially in brain regions associated with disease expression.

In the last decade a substantial amount of literature has focused on the effect of candidate genes on brain structure and function. Most of these studies have used single gene variants and risk haplotypes that were found traditionally from linkage studies and more recently from GWAS studies.

Rasetti and Weinberger (2011) reviewed the literature of candidate genes in schizophrenia in functional magnetic resonance imaging (fMRI) studies, indexed by cognitive task. In the working memory (WM) domain: (i) the GAD1, GRM3, COMT, RGS4, CACNA1C, KCNH2, DTNBPI and MTHFR genes have shown to impact function of the dorso-lateral prefrontal cortex (DLPFC), (ii) the COMT, PRODH and RGS4 have been associated with altered engagement of the ventrolateral prefrontal cortex (VLPFC), (iii) the PRODH with parietal lobe function, (iv) the NRG1, G72 and DISC1 with hippocampal function and last (v) the DRD2 and COMT with supragenua l ACC function. In the cognitive control domain: (i) the DTNBPI, DRD2 and MAOA have been associated with DLPFC function, (ii) the DRD2 with VLPFC function, (iii) the DRD2 and MAOA with parietal lobe function and iv) the COMT, DRD2 and MAOA with ACC function. From all these genes the only ones replicated or discovered in GWAS studies are GRM3, CACNA1C and DRD2 (Ripke et al., 2014). The CACNA1C (Paulus et al., 2014) and
ZNF804A (Rasetti et al., 2011) GWAS schizophrenia risk SNPs were found to be correlated with reduced prefrontal-hippocampal coupling during WM tasks in healthy controls and schizophrenia patients. Tan et al. (2012) using dynamic causal modelling (Friston et al., 2003) on fMRI WM data in healthy participants differentiated the effects of the COMT, DRD2 and AKT1 genes on the prefrontal-parietal WM maintenance and prefrontal-striatal WM manipulation network. Specifically, the prefrontal-parietal circuit is influenced only by the COMT gene, while the prefrontal-striatal circuit populated by the dopamine D2 receptors is affected by all three genes.

Most imaging genetics studies in bipolar disorder have focused on the effect of the CACNA1C risk gene, a bipolar disorder GWAS gene, on structural and functional neuroimaging data, in order to better understand the contribution of genetics to bipolar disorder. Two structural studies have found increased grey matter volume (Kempton et al., 2009) and increased total cortical volume (Wang et al., 2011) associated with the risk CACNA1C variant. Perrier et al. (2011) using a regions of interest approach found increased grey matter density in the right amygdala and hypothalamus in participants with the risk CACNA1C SNP, while another study found that the CACNA1C gene influences the brainstem rather than grey matter volume (Franke et al., 2010).
Studies on the effect of the \textit{CACNA1C} gene on brain function that included bipolar disorder patients, have found increased right amygdala activity during a negative emotional faces task in \textit{CACNA1C} risk carriers (Jogia et al., 2011; Tesli et al., 2013). In another study during an episodic memory task carrying the \textit{CACNA1C} risk variant resulted in a stronger decrease of hippocampal and ACC activation in bipolar disorder relatives, indicating an additive effect of \textit{CACNA1C} variation on familial risk (Erk et al., 2013). Using DCM connectivity analysis, two studies found that during emotional processing, the presence of the \textit{CACNA1C} risk allele was associated with decreased outflow of information from the medial frontal gyrus (Radua et al., 2013) and decreased visual-prefrontal effective connectivity (Dima et al., 2013). In both studies the findings were significantly more marked in patients than in their unaffected relatives and healthy controls. Although all these GWAS candidate genes imaging genetics studies illustrate the potential usefulness of genetic imaging, single SNP analyses alone do not address the overall genomic or polygenic architecture of schizophrenia and bipolar disorder.

\textit{Polygenic scores}

A typical GWAS tests millions of SNPs for association with the disease of choice, but this necessitates the application of a very conservative significance threshold,
usually $5 \times 10^{-8}$ for multiple testing correction. Typically, only a small number of SNPs/regions survive this threshold. Nevertheless, the amount of phenotypic variation explained by each GWAS-supported SNP is very small whereas the number of SNPs underlying the risk for illness is thought to be very large (Purcell et al., 2009). For that reason, candidate genes studies have to be treated with cautiousness since the possibility of a single SNP explaining a large variance of the disease expression is slim, whether this is in e.g. behavioural, cognitive or neuroimaging traits.

By utilizing the vast amount data created by GWAS studies (current sample size in schizophrenia $> 150,000$) we can calculate polygenic risk scores (PRS) for each individual associated with a specific disorder, e.g. schizophrenia or bipolar disorder. A PRS incorporates information from GWAS-SNPs and SNPs that are not genome-wide significant. These non significant genome-wide GWAS SNPs are meeting nominal significance criteria, even though it is accepted that some of them are false positives. The PRS thus models the aggregate effect of SNPs (alleles) associated with disease status present in each individual and allows us to utilize the power of large GWAS to be applied robustly in small samples (Ferreira et al., 2008; Purcell et al., 2009; Sklar et al., 2011; Dudbridge et al., 2013). For example the variance explained by the current schizophrenia PRS is 18%, while the variance explained by the CACNA1C gene is around 0.005% (Ripke et al., 2014). In Figure 1 we show that even with a sample size of
>150,000 we only have 60% power to detect the impact of the CACNA1C gene. With the statistical power of the PRS being exponentially better than that of a single SNP it points to a paradigm shift in the field of imaging genetics.

The first application of the PRS method was in schizophrenia, a study that supported the polygenic component of schizophrenia and its similarities with bipolar disorder (Purcell et al., 2009). These results were replicated in a second larger sample that doubled the proportion of variance explained by common variants to 6% in schizophrenia (Ripke et al., 2011). The polygenic model has also been successfully used in other traits like the body mass index (BMI; Speliotes et al., 2010), where the BMI PRS has shown positive association with disease expression (Peterson et al., 2011). Furthermore, the PRS method has been used successfully for the association of personality traits and mood disorders (Middeldorp et al., 2011; Luciano et al., 2012), depression and anxiety (Demirkan et al., 2011) and the five major psychiatric disorders (GROUP et al., 2013). It has also been used to differentiate genetically, for example, schizophrenia and autism (Vorstman et al., 2013) or schizophrenia and intelligence (van Scheltinga et al., 2013). The schizophrenia PRS has been correlated with quantitative measures of psychosis in terms of symptoms scales (Derks et al., 2012) and neuroimaging measures (Walton et al., 2013). Specifically the PRS for
schizophrenia was associated positively with reduced left DLPFC during a WM processing paradigm (Walton et al., 2013).

Recently, the bipolar disorderPRS has been used to examine its relationship with brain function and structure in individuals at familial risk for a mood disorder (Whalley et al., 2012; 2013). Whalley et al. (2012) showed a positive correlation between bipolar disorderPRS and activation of the ACC, amygdala and other limbic regions of the brain, areas that have been impaired in bipolar disorder expression. In the same cohort, the bipolar disorderPRS was significantly associated with decreased white matter integrity in the right superior longitudinal fasciculus (Whalley et al., 2013). Lastly, a study found an association between the PRS for major depressive disorder and reduced cortical thickness in the left amygdala-medial prefrontal circuitry in healthy young adults (Holmes et al., 2012).

Clinical relevance

Where does the PRS fit in clinical practice? In a large Swedish population-based study research showed that adding a PRS on a prediction model for prostate biopsies, would cut the biopsies from 12% to 5%, without missing less than 1% of the aggressive cancers in the population (Aly et al., 2011). Marking SNPs in patients have advantages that other biological tests currently lack: they are cheap, easy to analyse and most
importantly stay stable throughout the lifespan of someone life. As sample sizes increase exponentially in the field of psychiatry one can assume that more genetic risk variants will be identified in the near future, thus adding to the usefulness of genetic profiles substantially. The hope is that through polygenic scoring, especially in common diseases that seem to be affected by a large number of genes with small effects sizes, as is the case with schizophrenia and bipolar disorder, the road will lead to personalized medicine. Personalized genetics information is hopefully going to lead from interventional medicine to preventive medicine, which will help cut down health care costs. Another area of medicine where genetic information can influence greatly is pharmacogenomics, by keeping in mind that around 100,000 people die each year in the USA due to adverse drug reactions (Lazarou et al., 1998) and despite efforts to improve patient safety in the past few years it has not been successful (Landrigan et al., 2010; Phillips and Barker, 2010). Although until now there have been no robust findings of common genetic variants associated to drug response that could be translated in clinical practice in psychiatry the road has been paved for the identification of genetic determinants to personalized psychiatric treatment (Kirchheiner et al., 2005; Kim et al., 2006; Kato et al., 2010; Mrazek et al., 2011; Tansey et al., 2012). The efficacy of polygenic models for risk prediction, intervention and
personalized medicine can only benefit with the inclusion of other traditional risk factors such as family history and age of onset (Chatterjee et al., 2013).


**Acknowledgements**

This work was funded in part by the National Institute for Health Research (NIHR) Biomedical Research Centre for Mental Health at South London and Maudsley NHS Foundation Trust and [Institute of Psychiatry] King’s College London. This article/paper/report presents independent research in part funded by the National Institute for Health Research (NIHR). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

---

**Figure 1.** Power calculations for sample size for the CACNA1C gene and the schizophrenia PRS according to variance explained.
References


*Arch Gen Psychiatry* 68:1207-1217.


