INVITED REVIEW

The emerging molecular architecture of schizophrenia, polygenic risk scores and the clinical implications for GxE research

Conrad Iyegbe · Desmond Campbell ·
Amy Butler · Olesya Ajnakina · Pak Sham

Abstract Schizophrenia is a devastating mental disorder. The level of risk in the general population is sustained by the persistence of social, environmental and biological factors, as well as their interactions. Socio-environmental risk factors for schizophrenia are well established and robust. The same can belatedly be said of genetic risk factors for the disorder. Recent progress in schizophrenia genetics is primarily fuelled by genome-wide association, which is able to leverage substantial proportions of additional explained variance previously classified as 'missing'. Here, we provide an outline of the emerging genetic landscape of schizophrenia and demonstrate how this knowledge can be turned into a simple empirical measure of genetic risk, known as a polygenic risk score. We highlight the statistical framework used to assess the clinical potential of the new score and finally, draw relevance to and discuss the clinical implications for the study of gene–environment interaction.

Introduction

Schizophrenia is a devastating mental illness characterised by psychotic features, of which delusions, hallucinations and thought disorder are hallmarks. The prognosis of schizophrenia is poor in at least 20 % of cases [1]; sufferers of schizophrenia have a higher risk of comorbidity, they face social exclusion and the mortality rate is more than twice that of the general population [2].

Schizophrenia is estimated to be the most financially costly of psychiatric illnesses [3]. The global economic impact of schizophrenia can be measured in terms of healthcare expenditure and lost economic productivity [2–4]. The social impact of this disorder on sufferers, families, carers and the community is harder to measure but equally devastating [4, 5]. These issues give justifiable cause to develop strategies that may help identify high risk individuals and avert the poor outcomes associated with the subsequent onset of illness.

The baseline risk found in the general population is sustained by the persistence of social, environmental and biological factors, the study of which spans the two divergent disciplines of social science and genetics. The large population fractions attributed to socio-environmental risk factors for schizophrenia suggest that addressing these factors collectively would significantly alleviate the global burden of this illness [6]. However, without genetics, it is impossible to know for whom these risks can be averted.

A recent surge of new genetic data suggests that individually targeted prevention strategies are a realistic goal

C. Iyegbe (✉) · O. Ajnakina
Department of Psychosis Studies, Institute of Psychiatry,
King's College, London, UK
e-mail: c.iyegbe@iop.kcl.ac.uk

C. Iyegbe
Department of Bioinformatics, Social Genetic and
Developmental Psychiatry Centre, Institute of Psychiatry,
London, UK

D. Campbell · A. Butler · P. Sham
Department of Psychiatry, Li Ka Shing Faculty of Medicine,
The University of Hong Kong, Pokfulam, Hong Kong

P. Sham
The State Key Laboratory of Brain and Cognitive Sciences,
The University of Hong Kong, Pokfulam, Hong Kong

for schizophrenia [7]. Therefore, we begin this review by outlining an emerging genetic landscape of schizophrenia and how such knowledge can be converted into a simple risk score metric. This can be used to provide answers for questions concerning the genetic architecture of schizophrenia and related phenotypes. We outline the statistical infrastructure now allowing the clinical potential of this risk score to be assessed. Finally, we draw relevance to and discuss implications for GxE research.

Susceptibility to schizophrenia incorporates genetic and environmental risk

The scientific rationale for the genetic interrogation of a phenotype comes from the prior certainty of its genetic origin. Establishing that this is the case, by demonstrating the heritability of a trait, is the foundation for the subsequent exploration of genetic association.

Heritability expresses the proportion of phenotypic variance ascribed to genetic variance. Estimations can be derived in multiple ways. The archetypal approach uses twins and other family based designs [8]. Modern approaches to heritability estimation are a product of huge technological strides made in the 'post-genomic' era, a term used to describe the period since the human genome was first sequenced [9, 10]. These methods rely on information obtained from a DNA 'chip' to estimate the heritability associated with millions of genetic markers located throughout the genome (i.e. the entirety of one's DNA sequence). The most common class of marker interrogated by DNA chips is the single nucleotide polymorphism (SNP): a single subunit change in the DNA sequence. Typically, the information contained by these markers creates a characteristic genetic profile of each individual within a study. DNA chips allow efficient screening of disease genomes, within a genetic study design known as a Genome-wide Association Study (GWAS) [10].

A divergence exists between heritability estimates derived using family based study designs and those derived from population cohorts [11]. Heritability estimations from twins average around 0.8 [12], while population-based estimates are usually lower at around 0.65 [13]. As estimates of heritability from national records may be more representative of the samples comprising the large international GWAS consortia [14], it is conceivable that the upper estimated heritability of schizophrenia just mentioned may be methodologically enhanced. From this perspective, it is all too easy to overlook the importance of the contribution made by the environment. Its importance is underlined by the fact that not all carriers of highly penetrant genetic risk factors succumb to schizophrenia (or any other psychiatric illness) [15]. Key to understanding the

expression of genetic risk, therefore, is an appreciation of contextual influences arising from the environment (as well as other background genetic modifiers).

The environment and schizophrenia

Environmental risk factors for schizophrenia include prenatal maternal nutrition, maternal infection during pregnancy, obstetric complications, season of birth, urbanicity, migration, socio-economic status and cannabis. Table 1 lists these and other important environmental influences on risk of schizophrenia.

One advantage environmental factors have over genetic variables is their estimated effects: for complex diseases such as schizophrenia genetic risk factors tend to have very modest effects, while effects of environmental risk factors tend to be much larger. The typical odds ratios associated with the risk factors in Table 1 range between 1.5 and 11.0. In contrast, common genetic risk factors for schizophrenia are much smaller, with typical odds ratios of between 1.1 and 1.4. Apart from the subject-specific literature cited in Table 1, the wider range of non-genetic influences on risk is covered elsewhere [44] and hence is not discussed in depth here.

Debate about the relative importance of genes versus environment is never far away in schizophrenia and seems to be perpetuated by the zeal with which both sides cling to optimistic estimated contributions to population risk. Often, these may be flawed. For instance, a study that recently estimated the proportion of schizophrenia cases avoidable through social interventions [6] does not control for the influence of a positive family history (of schizophrenia), which is used as a fairly reliable proxy for genetic influence. Genetic studies, on the other hand, seldom acknowledge the potential for confounding due to hidden discrepancies in the environmental profile of comparison groups. Thus, in truth, the methodological biases present in both fields mean that the distinction between genetic and environmental influences may be less clear than it first seems. This means that a disorder resulting from GxE will more often be attributed to genes by geneticists and to the environment by social scientists.

The emerging molecular landscape of schizophrenia

A disorder of rare and common genetic effects

The infrastructure that has enabled GWAS to flourish was developed in the aftermath of the first human genome sequence [9, 45, 46]. GWAS was originally designed to test the Common Disease Common Variant hypothesis, a

Table 1 Environmental risk factors for Schizophrenia

Context	Environmental risk factor	Review	Recent meta-analysis
Social	Urban–rural dwelling	Heinz et al. [16]; Kelly et al. [17]	Vassos et al. [18]
	Social context—neighbourhood effects	Brown 2011 [19]	–
	Social discrimination–discrimination	Veling and Susser [20]	Allardyce et al. [21]
	Migration	Cantor-Graae and Selten [22]; Morgan et al. [23]	DeAlberto et al. [24]
Environmental	Cannabis smoking	Henquet et al. [25]	Semple et al. [26]; Arsenault et al. [27]; Moore et al. [28]
	Chemical pathogens	Parakh and Basu [29]	–
	Famine	Brown [19]	–
Familial	Childhood trauma	Schäfer and Fisher [30]	–
	Childhood adversity	Morgan and Fisher [31]	Varese et al. [32]
		–	
	Advanced paternal age	Miller et al. [33]	Miller et al. [34]
Neurodevelopmental	Seasonal birth	Davies et al. [35]	Davies et al. [35]
	Birth defects/obstetric complications	Palmer [36]	–
		Scott [37]	
	Vitamin D	McGrath et al. [38]	–
	Prenatal maternal infections	Khandaker et al. [39]; Brown [40]	–
Economic	Developed vs developing country	–	Saha et al. [2]
	Socio-economic status	Cohen et al. [41]	–
Other	Gender	–	Aleman et al. [42]; McGrath et al. [43]

theory which explains the high burden of cases in a population through the joint effect of thousands of common risk alleles present at >5 % frequency in the general population. Both in theory and in practice, common variants individually explain only tiny amounts of actual risk.

After a slow start, the genetic landscape of schizophrenia has steadily been transformed by GWAS [47]. The total number of genome-wide 'hits' for schizophrenia currently stands at 128 [50]. In comparative terms, schizophrenia genetics outperforms all other psychiatric disorders [50].This, in part, reflects the different architectural structure of each disorder. For example, the heritability of major depression ($\sim$0.40 [51]), a disorder in which few GWAS hits have been identified, is approximately half that of schizophrenia's. Hence, a 4- to 5-fold increase in current sample size would be required to provide for a GWAS of depression the statistical power as schizophrenia currently has [50].

In the time since it was initially developed, the GWAS hardware (a 'chip' which contains thousands of biological features that facilitate whole-genome scanning) has become laden with newer features which allow rarer genetic novelties to be interrogated, either through imputation methods [52] or targeted assays [53, 54]. The most ready examples of this are copy number variants (CNVs).

The term describes the duplication and deletion events that occur at hotspots present throughout the genome. This type of variation is consistent with the 'Multiple Rare Variant' theory of causality, which defines schizophrenia as a disorder resulting from a high burden of these structural events in the population. The rarity of these variants individually (they occur at frequencies much lower than 1 %) is explained by large effects on risk and an associated negative impact on fecundity [55], which results in purifying selection against these variants. A recognised source of the rare variants found in schizophrenia is spontaneous (or 'de novo') mutations [56]. This coincides with the observation that the fathers of individuals with schizophrenia tend to be older [57, 58]. Some of these resist negative selection pressure to persist in the population, by segregating within families. But because de novo events do not contribute to the phenotypic similarity between twins concordant for the disorder, such events fail to contribute to estimated heritability.

On rare occasions, the co-transmission of genetic markers with illness has allowed the bounds surrounding causal regions to be narrowed enough for schizophrenia genes to be identified. This has been done using methods purposefully designed to capture genetic signals of segregation within families [59, 60]. It is important to note the

high level of consistency between current and previous efforts to target such effects. A comparison of previous meta-analytic evidence [61, 62] and current genomic findings [63–67] reveals a reassuring level of overlap between previous candidate regions and structural mutations on chromosomes 1q, 2p, 2q, 3q, 4q, 5q, 11q, 15q, 17q and 22q.

It is because the allelic spectrum of schizophrenia demonstrably ranges from common [49] to rare [48] that it is inaccurate to assume a delineation between common and rare genetic architectures [14]. While the catalogue of rare genomic events in schizophrenia is not currently extensive enough to tally with the genetic classification of schizophrenia as a common disorder, the initial screens for rare variants have been far from exhaustive and thus will have to wait for more of the sequencing initiatives that are currently underway (e.g. http://www.uk10k.org/studies/neurodevelopment.html) to run their course.

The broader common architecture can be purposefully leveraged

The key to the recent success of GWAS is a large international cohort which continues to expand. The total number of schizophrenia cases acquired by the international Psychiatric Genomic Consortium (PGC) has reached 35,000 [50]. In comparison, the earliest cohorts used for GWAS purposes were more than ten times smaller [68, 69]. The figure is projected to reach 60,000 cases by 2015. One hopes that the level of success (measurable by the count of new discoveries and the total variance explained by genome-wide data) will continue to match the effort required to assemble such cohorts. Current figures would suggest that the international movement has acquired a sufficient level of momentum to sustain these discoveries; at present, every 300 cases genotyped by the PGC generate a new GWAS finding (Gerome Breen, pers commun). Individually, typical effect sizes do not exceed an OR of 1.2, while explained variances range between 0.05 % (*Znf804A*) and 0.67 % (*RELN*) [14]. Although CNVs are much more likely to be disruptive, they are also much less frequent and, therefore, account for similar proportions (e.g. 17p12: 0.02 %; 22q11: 0.21 %) of the overall variance.

A landmark study of schizophrenia demonstrated that a large chunk of heritability previously assumed to be 'missing' [70] has been hiding all along on the same chips used to reach this conclusion. This 'dark' variance may be harnessed by tethering annotated genetic risk to a single quantitative 'polygenic' risk score. The score may incorporate sub-threshold effects, which GWAS is underpowered to detect at the conventional threshold of $P = 5 \times 10^{-8}$ [47]. The additional variance that can now be recovered has been analysed in relation to the clinical phenotype, as well as against quantitative traits related to the disorder (known as 'endophenotypes'), with varying success [71, 72].

The polygenic score is derived and constructed in a 'training' sample before implementation in a separate independent dataset. Deriving the score involves the use of permissive P value thresholds to aid the discovery of signals in the training sample. The P value thresholds used occupy a sliding scale ranging from $P < 0.01$ to $P < 0.5$ [47]. Genetic signals that pass these benchmarks will incorporate many true underlying associations initially missed because of low power. The polygenic risk score is calculated as follows [73]:

$$\text{Polygenic risk score} = \sum x_i \times \log(\text{OR}_i) \quad (1)$$

where OR_i is the allelic odds ratio as estimated in the discovery dataset and x_i is the number of risk alleles present at a single bi-allelic locus (i.e. 0, 1 or 2) within a subject. The allelic odds ratio for the association of SNP_i vs. trait is multiplied by the number of risk alleles. The procedure is repeated for each SNP falling within a specified P value range. The aggregate count of weighted risk alleles is calculated to create a per person polygenic score. The distribution of score values is normalised by fitting to a standard normal distribution curve, which helps with the interpretation of the score in downstream analyses (for example, the per standard deviation increase on the corresponding risk scale). The polygenic score suffers from the mass infiltration of null GWAS effects. The negative impact on the statistical power of the score was demonstrated graphically both by Purcell et al. [47] and by Dudbridge [74]. As the composition of the score at a given inclusion threshold is a balance between true and null effects, it is common practice to present the change in the association statistic as the inclusion threshold is varied. Recent work by the PGC suggests that around 25 % of additive genetic liability to schizophrenia is explained by the information derived from its cohort [75]. The high proportion of heritability captured by the polygenic score explains the additional leverage it provides for association analyses.

New insights from polygenic risk score analysis of schizophrenia

Demographic differences in genetic risk are currently estimated to be small

Very few high-powered genetic studies have been performed in samples of African American, or other African populations to date. It is reasonable to expect that the

genetic contribution to schizophrenia will vary by population on account of population genetic differences. Even if the genetic risk architecture is the same for Caucasian and non-Caucasian populations, difficulties associated with confirming this should be anticipated in practice. This is because the genomes of these two populations have fundamentally different correlation structures [76]. This correlation, known as linkage disequilibrium (LD), allows a pair of neighbouring (physically close) SNPs to be genetically close in one population but distant and uncorrelated in another. Because the genetic proximity between two loci is the basis of genetic association (between marker and disease locus), the efficiency with which disease loci are captured (or 'tagged') will also vary in different populations. Given these considerations, the level of correlation found between the two population risk profiles is surprisingly high so far; the heritability associated with the current genomic profile of schizophrenia is similar regardless of whether it is studied in African Americans (0.24, SE = 0.09) or in European ancestral groups ($\sim$0.28, SE = 0.03) [77]. The similarity in genetic risk architecture is also confirmed by the high degree of genetic correlation between European and African American schizophrenia genomes (i.e. African American vs. European: $r = 0.61$–0.66).

Gender-specific GWAS such as Shifman et al. [68] has only been performed on one occasion. This is curious, given that almost all features of schizophrenia, including prevalence, incidence, age of onset, presentation, course and treatment response, demonstrate a gender difference [78]. The correlation between male and female schizophrenia genomes was recently examined on the basis of the polygenic risk score for schizophrenia. Genomic 'partitioning' methods were used to evaluate a combined discovery/replication sample of 9,087 cases and 9,343 controls from the PGC-SCZ cohort. The estimated overlap found ($r = 0.89$, SE = 0.06) indicates that the majority of the genetic variance captured by chips is shared between sexes [75]. Also determined was that the variance explained by SNPs on the X chromosome is in proportion to its length. Neither piece of evidence is consistent with a skew in the distribution of genetic risk between genders; hence there is no genetic basis for the gender differences widely observed in schizophrenia.

Cross-disorder overlap of common genetic influences

Mental disorders have traditionally been viewed as distinct, categorical conditions. This view is being challenged by emerging evidence that suggests psychiatric traits may fit into a single broader framework [79]. The organisation of this framework may be underpinned by genetic correlation [80]. For example, the common genetic architecture of schizophrenia is highly correlated with both bipolar disorder ($r = 0.68 \pm 0.04$) and major depression ($r = 0.43 \pm 0.06$) [80]. Beyond this trio of disorders, a correlation between schizophrenia and autism is also significant ($r = 0.16 \pm 0.06$) [80].

Attempts have recently been made to pinpoint the molecular sources of cross-disorder overlap [81]. Four SNPs with genome-wide significant influence ($P < 5 \times 10^{-8}$) on cross-disorder risk were identified, implicating the genes *ITIH3*, *AS3MT*, *CACNA1C* and *CACNB2*. Only the effect of *CACNA1C* was not fully evident across all the disorders examined (schizophrenia, major depression, bipolar disorder, autism and attention-deficit/hyperactivity disorder). The SNP effect detected in this gene was restricted to bipolar disorder and schizophrenia (at $P = 1.8\text{x}10^{-8}$). When the search for cross-disorder effects was broadened to include previous GWAS hits for both schizophrenia ($n = 10$) and bipolar disorder ($n = 4$), a disorder-specific model was found to provide the best-fit for 7 of the 14 hits. The remaining 7 SNPs (6 schizophrenia and 1 bipolar disorder) showed genome-wide evidence ($P < 5 \times 10^{-8}$) of cross-disorder effects. Findings from these recent cross-disorder analyses confirm long-held suspicions of biological overlap between psychiatric traits that date back to the pre-GWAS era [82].

Clinical applications of the schizophrenia polygenic score

The schizophrenia polygenic score may have beneficial implications for future clinical decision making, and may form the basis for a coherent public health strategy for schizophrenia. Many options exist for evaluating the clinical efficacy of risk scores. The area under the receiver–operator characteristic (ROC) curve is the most widely recognised [7, 83, 84]. The ROC is obtained by ranking a set of individuals with known disease status according to their risk for disease. The curve is derived by plotting sensitivity (each correctly ranked case) against 1-sensitivity (each instance when a control is incorrectly ranked above a case, by the score). The AUC reflects the area under this curve and expresses the accuracy with which high- and low-risk groups can be distinguished according to a bivariate outcome (such as case–control status). Uninformative prediction corresponds to an AUC of 0.5. Classifiers with AUCs >0.75 are useful for identifying high-risk groups who might benefit from screening [85]. Classifiers exceeding an AUC of 0.99 can be used to diagnose a disease reliably in the general population [85]. The AUC has recently been formularised for genetic purposes [7, 83, 84] to take account of extrinsic influences, such as disease prevalence and heritability, which vary

with trait architecture. It expresses risk using the liability threshold model, which assumes that the individuals of a population possess a naturally varying liability to disease, with clinical illness only developing in those whose excess risk exceeds a certain threshold [11].

With a genomic profile accounting for roughly a quarter (23 %) of trait variance, the polygenic risk score for schizophrenia is estimated to be consistent with a genetic AUC of 0.75. If this is true, then the score could potentially already have clinical applications in undefined groups at high risk of schizophrenia [85]. However, even under the idealised assumption that all the heritability of schizophrenia can readily be translated into genetic findings, the AUC for schizophrenia is predicted to fall short of the level required for screening the general population (i.e. >0.99 [85]). Some of the extensions made to the AUC mean that the clinical utility of polygenic risk scores can be improved by allowing clinical questions to be refined [84]; however, ultimately the current GWAS strategy will need to be broadened to accommodate other complementary approaches. The effect of methodological constraints on current prediction power is not factored into these estimates. The following section explores these in greater detail.

Constraints on risk prediction accuracy

One key factor limiting the accuracy of polygenic risk prediction is the size of the discovery sample used to calculate the risk score weights. At limiting sample sizes, sampling error allows null SNPs unassociated with schizophrenia into the risk score. This comes at the expense of prediction power. When the purpose of deriving the score is to test for association, the impact of sampling error can be offset by replication in independent datasets. Risk prediction, however, is an exact science that penalises for every false discovery included in the final genetic model. It is similarly undone when true associations are omitted. Prediction of risk by genetic scores has failed when the sample size used to generate risk scores has been suboptimal [86], whereas the sample size improvements needed to increase the chances of successful outcome may be relatively modest [74]. The high heritability explained by the polygenic score for schizophrenia ensures it can provide a convenient preview of future GWAS hits. The impending arrival of genomics to clinical settings, as well as ongoing initiatives in biobanking, will have a vital role to play in the large-scale delivery of future samples for GWAS [87, 88].

GWAS chips only explain around 80 % of the known SNP variance in the Caucasian genome [89]. Increasing the level of coverage is, therefore, critical for unlocking the untapped sources of genetic variance. Current evidence

suggests broad convergence of the genetic profile of schizophrenia in Caucasians and Africans. But only 23 % of liability variance has been explained so far. Therefore, it should not be assumed that a single genomic profile will satisfy the purposes of risk prediction in the global community; potentially some reconfiguration of the score will be needed to maximise the efficiency and accuracy. Much of this reconfiguration may be driven by new sequence-level insights from different populations.

Future strategies for enhancing prediction

The incorporation of family history

Family history of schizophrenia still remains the most reliable predictive instrument currently available to the clinician, but is of demonstrably limited use when implemented on its own. Optimistic simulations suggest that a family history of schizophrenia equates to an explained liability variance (or genetic AUC) of between 0.56 and 0.66 [7]. One of the practical drawbacks of family history is that while positive scores can be helpful and informative of genetic status, little can be inferred from a negative family history [90], as recall bias, varying family size and knowledge of family history all play a part in making the interpretation of a negative response not so straightforward. The interpretation of family history is further undermined by the assumption that it only reflects genetic liability, when in fact the familial environment contributes 6–11 % of this effect [12, 13].

Risk prediction can be boosted by supplementing the polygenic score profile with information on family history [7, 83]. Chatterjee et al. [83] were able to boost the AUC for polygenic scores for a number of GWAS disorders that encompass different prevalences, heritabilities and familial loading. Therefore, this work has potential relevance to schizophrenia. Despite the added efficacy of including family history, AUC scores remained below 0.9 for all but the rarest and most familial disorders (Type-1 diabetes and Crohn's disease) in the analytical models explored by Chatterjee et al. Also it is necessary to check that family history and the polygenic score are not actually correlated (e.g. Belskey et al. [91]), because this will overstate the advantage gained by combining both sets of information.

Environmentally informed genetic risk prediction

The strongest evidence for the existence of GxE comes from empirical studies which use family history as a proxy tool to assess the joint effects of environmental and genetic factors on risk. Family history has been shown to augment the psychotogenetic effects of cannabis [92]. Another study

compared and found an identical rate of schizophrenia family history in cannabis-induced cases of psychosis and schizophrenia. This suggests that sensitivity to the psychotogenic effects of cannabis is modified by genetic liability to schizophrenia [93]. Similar GxE effects can be demonstrated for urbanicity [94] and prenatal infection [95], but seemingly not for obstetric complications [96]. The adoption study design is a convenient way to disentangle the genetic and environmental components of familial risk. It has been used to demonstrate the interaction between social determinants such as parental separation and heritable risk [97]. If these molecular effect modifiers could be captured and tethered to the existing polygenic score, then prediction performance could be boosted to make population screening feasible and relevant to the exposure profile of the individual. This point underlies why the modification of environmental effects by polygenic scores is poised to become a major paradigm for translational research in schizophrenia.

Gene–environment interactions in Schizophrenia

The central paradigms

Typically, statistical interaction is estimated via inclusion of an interaction term in a regression model. A wider debate surrounds how these regression models should be scaled [98, 99], as the presence or absence of statistical interaction depends on how one expects risk factors to combine. For multifactorial disease, researchers typically use logistic regression to look for disease risk factors. Under this model, risk factors are assumed to be additive on the log odds ratio scale. This is the same as saying risk factor odds ratios are multiplicative. For rare diseases, this is equivalent to saying relative risks are multiplicative. Logistic regression models multifactorial disease architecture well and typically there is no interaction on this scale. This lack of interaction has perplexed many because it flies in the face of knowledge that interaction is endemic in biological systems, but this stems from their lack of appreciation that additivity in logistic regression actually models a particular type of interaction (multiplicative odds ratios).

An alternative view of disease aetiology comes from the counterfactual model or potential outcomes model of disease [100]. This is perhaps the most coherent definition of cause for multifactorial phenomena. Rothman's Causal Pies [100] conceptualise this; a sufficient cause is a set of criteria (component causes) which when satisfied guarantee the onset of the disease. There may be several ways for a person to get a disease (i.e. there may be several causal pies), and these may differ between individuals, so a risk factor for one person may not be relevant for another. This provides a framework for prediction of population health intervention efficacy. If the proportion of persons with both necessary risk factors 'A and B' is large, then removal of either risk factor A or B will save many from disease, whereas if small and the proportion of 'A or B' persons is large then maybe only removal of A and B will be efficacious. Estimation of population proportions of these person types is, therefore, desirable. Unfortunately, it is confounded because the number of possible risk exposure strata (from which one might estimate person type proportions) is less than that of possible person types. For example, for exposure A, we have 2 strata—exposed or not, but four possible person types—causal, protective, immune, doomed. Nevertheless, some restrictions can be put on the ranges for person type proportions and a useful statistic has been derived (with some additional assumptions) for the interaction of two risk factors. This interaction statistic, synergy, also known as biological interaction, asserts the risk difference scale as the correct scale for measuring interaction.

The traditional GxE approach

The key to recent triumphs of the GWAS approach in schizophrenia is the fact that its implementation has been more rigorous and methodologically homogeneous than the previous generation of candidate-gene studies [101]. This contrasts with GxE research, where a diversity of methods and standards still exists. (A summary of studies that have explored GxE interactions in psychosis and related phenotypes can be found in Table 2, and have recently been reviewed by Modinos et al. [102]). This diversity has overshadowed some of the better study templates for this model [103] and has even precluded meta-analytical evaluation of the current GxE evidence [102]. This appears to be the legacy of a time in which geneticists and social scientists competed to demonstrate which science contributed most to schizophrenia liability. In doing so, opportunities to lay foundations for future GxE research were missed. The result today is a relative paucity of datasets able to adequately assess the effect of joint exposure to genes and environment.

Longitudinal studies sit at the top of a complex methodological hierarchy of epidemiological study designs. But these too have failed to bring resolution to ongoing controversies, such as the controversial interaction between cannabis and the *COMT* gene in schizophrenia [104, 105]. One fundamental problem that classical approaches to GxE face is that the sample sizes required to find GXE effects are several times larger than are needed to detect the main effect of a gene [106]. Some estimates suggest that the average GxE effect would need to be ten times larger than

Table 2 Published GxE studies in psychosis research

References	Sample	Candidate G (and locus name)	Candidate E	Outcome variable	Association result
Di Forti et al. [107]	489 first-episode psychosis patients, 278 Healthy controls	AKT1 (rs2494732)	Cannabis (lifetime history of use)	Psychotic disorder	Positive
		AKT1 (rs2494732)	Cannabis (lifetime frequency of use)	Psychotic disorder	Positive
Bhattacharyya et al. [108]	35 healthy controls	DAT1 (3′ UTR VNTR, AKT1 rs1130233)	Cannabis (delta-9-THC)	Delta-9-THC-induced psychotic experiences	Positive
Zammit et al. [104]	2630 healthy controls	COMT (rs4680, rs4818, rs6269, rs737865, rs2097603, rs165599)	Cannabis	Psychotic experiences	Negative
van Winkel et al. [109]	810 SZ, 740 siblings, 419 healthy controls	152 SNPs in 42 genes	Cannabis	Psychotic disorder	Positive
Kantrowitz et al. [110]	92 SZ patients (33 white, 46 African–American)	COMT (rs4680)	Cannabis	Cannabis use in Caucasians and African–Americans	Negative
Henquet et al. [111]	31 psychotic disorder, 25 healthy controls	COMT (rs4680)	Cannabis	Psychotic experiences (ESM)	Positive
Zammit et al. [112]	750 SZ, 688 healthy controls	CNR1 (rs1049353)	Cannabis	Psychotic disorder	Negative
		COMT (rs4680, rs737865 and rs165599)	Cannabis	Psychotic disorder	Negative
Henquet et al. [113]	30 psychotic disorder, 12 relatives, 32 healthy controls	COMT (rs4680)	Cannabis (delta-9–THC)	Delta-9-THC-induced psychotic experiences	Positive
Caspi et al. [105]	803 healthy controls	COMT (rs4680)	Cannabis	Schizophreniform disorder	Positive
Alemany et al. [114]	533 healthy controls	BDNF (rs6265)	Childhood abuse	Psychotic experiences	Positive
		BDNF (rs6265)	Childhood neglect	Psychotic experiences	Negative
Muntjewerff et al. [115]	742 SZ, 884 healthy controls	MTHFR (rs1801133)	Winter birth	Schizophrenia	Negative
Chotai et al. [116]	954 UPAD, BPAD, and SZ 395 healthy controls	TPH (rs1800532)	Seasonality of birth	Season of birth variations in UPAD, BPAD, and SZ	Positive
		5-HTTLPR	Seasonality of birth	Season of birth variations in UPAD, BPAD, and SZ	Positive
		DRD4	Seasonality of birth	Season of birth variations in UPAD, BPAD, and SZ	Positive
Tochigi et al. [117]	110 SZ, 493 healthy controls	HLA (HLA-A24, HLA-A26)	Seasonality of birth	Association between HLA-A and winter birth in SZ	Negative
Narita et al. [118]	60 SZ + HLA-DR1, 307 SZ no HLA-DR1	HLA (HLA-DR1)	Seasonality of birth	Association between HLA-DR1 and winter birth in SZ	Positive
Nicodemus et al. [119]	116 SZ spectrum disorders, 134 Healthy controls	AKT1	Obstetric complications	Schizophrenia	Positive

Table 2 continued

References	Sample	Candidate G (and locus name)	Candidate E	Outcome variable	Association result
		BDNF	Obstetric complications	Schizophrenia	Positive
		DTNBP1	Obstetric complications	Schizophrenia	Positive
		GRM3	Obstetric complications	Schizophrenia	Positive
Peerbooms et al. [120]	98 psychotic disorder, 118 healthy controls	*COMT* (rs4680), *MTHFR* (rs1801133), *MTHFR* (rs1801131)	Stress (ESM)	Psychotic experiences (ESM)	Positive
		COMT (rs4680), *MTHFR* (rs1801131)	Stress (ESM)	Psychotic experiences (ESM)	Negative
Collip et al. [121]	86 psychotic disorder, 109 healthy controls	*COMT* (rs4680)	Stress (ESM)	Psychotic experiences (ESM)	Positive
Kéri et al. [122]	200 SZ	*NRG1* (rs6994992, rs10954867 and rs7005288) The latter two were used as control SNPs	Psychosocial stress (neutral or conflict-related family interactions)	Unusual thoughts	Positive
Simons et al. [123]	579 young adult female twins (general population)	*COMT* (rs4680)	Stress (ESM)("event" and "social" stress)	Feelings of paranoia (ESM)	Positive
		BDNF (rs6265)	Stress (ESM) ("event" and "social" stress)	Feelings of paranoia (ESM)	Positive
van Winkel et al. [124]	31 psychotic disorder + cannabis, 25 healthy controls + cannabis	*COMT* (rs4680)	Stress (ESM)	Psychotic experiences (ESM)	Positive
Stefanis et al. [125]	306 healthy controls	*COMT* (rs4680)	Stress	Psychotic symptoms	Positive

Adapted from Modinos et al. [102]

SZ patients with schizophrenia, *UPAD* unipolar affective disorder, *BPAD* bipolar affective disorder, *ESM* experience sampling method, *delta-9 THC* delta-9-tetrahydrocannabinol: the main psychoactive ingredient of cannabis

those identified by GWAS, for the typical sample sizes used in candidate-gene research (typically in the hundreds) to be relevant [103]. This problem is compounded by the issue of publication bias. To refute a GxE discovery, a study must be, on average, six times larger than those which support the finding, to make it into the public domain [103].

Genome-wide approaches: GWEIS

An alternative to candidate GxE studies is to consider GxE on a genome-wide basis. A methodological approach that allows this to be done efficiently is the genome-wide environment interaction study design (GWEIS) [126]. The number of possible pairwise and higher-order interactions is vast. This necessitates simplifications in modelling which may be better and more flexibly expressed through machine learning and data mining approaches. Under most models, interaction terms contribute to marginal effects [127–129]. Some have recommended consolidating

genome-wide detection power by exploring the marginal and interaction properties of the genome simultaneously [130]. Others seek to increase effect size and decrease multiple testing by considering gene by E, or pathway by E interaction. Two-stage GWEIS methods have been developed which increase power by reducing the multiple testing burden. Genetic loci are first selected by their associations with disease and environment separately, and then subsequently tested for interaction [131]. Power considerations also underlie the case-only study design which derives power from the assumption that G and E are independent in the population.

GWEIS methods are yet to take off in schizophrenia, thus the current crop of GxE findings (reviewed in Modinos et al. [102]) for schizophrenia is still yet to face the same acid test used to put the previous generation of genetic association candidates on trial [132, 133]. Just one GWEIS survey has been conducted to date in schizophrenia [134]. The study of genome-wide interaction with maternal cytomegalovirus (CMV) infection, returned one genome-

wide significant hit, an SNP within the gene *CTNNA3* (P value = 7.3×10^{-7}), which had not previously been implicated in schizophrenia.

In theory, the same polygenic principle described earlier can be applied in the context of GWEIS. Therefore, GWEIS may have an important part to play in deriving an environmentally informed genetic risk profile for schizophrenia (one whose power can be enhanced by selective application to individuals that have the corresponding exposure profile. However, the same constraints outlined earlier (power and coverage) will be limiting here also.

Systematic approaches could also be used to identify novel environmental risk factors. The potential to recover interactions between weak risk environments and rarer genetic variants has been demonstrated in theory [135], but in practice can only be tested in the largest schizophrenia datasets, in which rarer risk variants will be better represented. As the frequency of scores in the extreme upper portion of the risk score distribution may be similar (in frequency and penetrance) to rarer structural mutations, the polygenic score for schizophrenia could also be leveraged to identify new environmental factors that contribute to GxE.

Future parallel strategies for improving the power of GxE

The polygenic risk score provides the ultimate genetic tool for detecting GxE. But further gains in detection power can be made by increasing the sensitivity with which environmental measures are captured. Simulated studies of measurement error suggest that an increase in correlation with true exposure values from 0.4 to 0.7 can equate to as much as a 20-fold gain in sample size [136]. Purposefully, written tools have since been developed that allow the precision of exposure measurement to be factored into power calculations (see http://www.hsph.harvard.edu/faculty/peter-kraft/software/ or the ESPRESSO power calculator at http://www.p3gobservatory.org/powercalculator.htm).

The use of endophenotypes will be beneficial in this regard and is consistent with the P factor hypothesis [79], which proposes that all psychiatric traits map onto a single underlying psychopathological profile.

Conclusion

GWAS of schizophrenia can be considered to have delivered a satisfactory return on the initial capital outlay. The resulting genomic profile has already replaced family history as the most reliable genetic prediction tool in schizophrenia research. We have seen that while this tool has considerable clinical potential, it clearly has some way

to go before it can be considered valid for widespread clinical use.

Hence, the ability to provide an accurate discrimination of genetic risk, based on a case-by-case understanding of historical environmental footprints, is an important clinical objective, and one which is consistent with the goal of personalised medicine. The polygenic score provides considerable scope to improve the detection of GxE at the statistical level.

The option to extend the search for meaningful interaction into other biological domains (e.g. gene expression) is attractive but currently under-resourced in the field of schizophrenia. Such approaches will help to build an understanding of the biological architecture that underpins schizophrenia in its various environments of origin.

References

1. Harrison G, Hopper K, Craig T et al (2001) Recovery from psychotic illness: a 15- and 25-year international follow-up study. Br J Psychiatry 178:506–517
2. Saha S, Chant D, Welham J, McGrath J (2005) A systematic review of the prevalence of schizophrenia. PLoS Med 2(5):e141
3. O'Malley AJ, Frank RG, Normand SL (2011) Estimating cost-offsets of new medications: use of new antipsychotics and mental health costs for schizophrenia. Stat Med 30(16):1971–1988
4. Awad AG, Voruganti LN (2008) The burden of schizophrenia on caregivers: a review. Pharmacoeconomics 26(2):149–162
5. Chan SW (2011) Global perspective of burden of family caregivers for persons with schizophrenia. Arch Psychiatr Nurs 25(5):339–349
6. Kirkbride J, Coid JW, Morgan C et al (2010) Translating the epidemiology of psychosis into public mental health: evidence, challenges and future prospects. J Public Ment Health 9(2):4–14
7. Wray NR, Yang J, Goddard ME, Visscher PM (2010) The genetic interpretation of area under the ROC curve in genomic profiling. PLoS Genet 6(2):e1000864
8. Cardno AG, Marshall EJ, Coid B et al (1999) Heritability estimates for psychotic disorders: the Maudsley twin psychosis series. Arch Gen Psychiatry 56(2):162–168
9. Lander ES, Linton LM, Birren B et al (2001) Initial sequencing and analysis of the human genome. Nature 409(6822):860–921
10. Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88(1):76–82
11. Wray NR, Visscher PM (2010) Narrowing the boundaries of the genetic architecture of schizophrenia. Schizophr Bull 36(1):14–23

12. Sullivan PF, Kendler KS, Neale MC (2003) Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Arch Gen Psychiatry 60(12):1187–1192

13. Lichtenstein P, Yip BH, Bjork C et al (2009) Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. Lancet 373(9659):234–239

14. Visscher PM, Goddard ME, Derks EM, Wray NR (2012) Evidence-based psychiatric genetics, AKA the false dichotomy between common and rare variant hypotheses. Mol Psychiatry 17(5):474–485

15. Vassos E, Collier DA, Holden S et al (2010) Penetrance for copy number variants associated with schizophrenia. Hum Mol Genet 19(17):3477–3481

16. Heinz A, Deserno L, Reininghaus U (2013) Urbanicity, social adversity and psychosis. World Psychiatry 12(3):187–197

17. Kelly BD, O'Callaghan E, Waddington JL et al (2010) Schizophrenia and the city: a review of literature and prospective study of psychosis and urbanicity in Ireland. Schizophr Res 116(1):75–89

18. Vassos E, Pedersen CB, Murray RM, Collier DA, Lewis CM (2012) Meta-analysis of the association of urbanicity with schizophrenia. Schizophr Bull 38(6):1118–1123

19. Brown AS (2011) The environment and susceptibility to schizophrenia. Prog Neurobiol 93(1):23–58

20. Veling W, Susser E (2011) Migration and psychotic disorders. Expert Rev Neurother 11(1):65–76

21. Allardyce J, Boydell J (2006) Review: the wider social environment and schizophrenia. Schizophr Bull 32(4):592–598

22. Cantor-Graae E, Selten JP (2005) Schizophrenia and migration: a meta-analysis and review. Am J Psychiatry 162(1):12–24

23. Morgan C, Charalambides M, Hutchinson G, Murray RM (2010) Migration, ethnicity, and psychosis: toward a sociodevelopmental model. Schizophr Bull 36(4):655–664

24. Dealberto MJ (2010) Ethnic origin and increased risk for schizophrenia in immigrants to countries of recent and long-standing immigration. Acta Psychiatr Scand 121(5):325–339

25. Henquet C, Murray R, Linszen D, van Os J (2005) The environment and schizophrenia: the role of cannabis use. Schizophr Bull 31(3):608–612

26. Semple DM, McIntosh AM, Lawrie SM (2005) Cannabis as a risk factor for psychosis: systematic review. J Psychopharmacol 19(2):187–194

27. Arseneault L, Cannon M, Witton J, Murray RM (2004) Causal association between cannabis and psychosis: examination of the evidence. Br J Psychiatry 184:110–117

28. Moore TH, Zammit S, Lingford-Hughes A et al (2007) Cannabis use and risk of psychotic or affective mental health outcomes: a systematic review. Lancet 370(9584):319–328

29. Parakh P, Basu D (2013) Cannabis and psychosis: have we found the missing links? Asian J Psychiatr 6(4):281–287

30. Schafer I, Fisher HL (2011) Childhood trauma and psychosis—what is the evidence? Dialogues Clin Neurosci 13(3):360–365

31. Morgan C, Fisher H (2007) Environment and schizophrenia: environmental factors in schizophrenia: childhood trauma—a critical review. Schizophr Bull 33(1):3–10

32. Varese F, Smeets F, Drukker M et al (2012) Childhood adversities increase the risk of psychosis: a meta-analysis of patient-control, prospective- and cross-sectional cohort studies. Schizophr Bull 38(4):661–671

33. Miller B, Pihlajamaa J, Haukka J et al (2010) Paternal age and mortality in nonaffective psychosis. Schizophr Res 121(1–3):218–226

34. Miller B, Messias E, Miettunen J et al (2011) Meta-analysis of paternal age and schizophrenia risk in male versus female offspring. Schizophr Bull 37(5):1039–1047

35. Davies G, Welham J, Chant D, Torrey EF, McGrath J (2003) A systematic review and meta-analysis of Northern Hemisphere season of birth studies in schizophrenia. Schizophr Bull 29(3):587–593

36. Palmer CG (2010) Evidence for maternal-fetal genotype incompatibility as a risk factor for schizophrenia. J Biomed Biotechnol 2010:576318

37. Scott J, McNeill Y, Cavanagh J, Cannon M, Murray R (2006) Exposure to obstetric complications and subsequent development of bipolar disorder: systematic review. Br J Psychiatry 189:3–11

38. McGrath JJ, Eyles DW, Pedersen CB et al (2010) Neonatal vitamin D status and risk of schizophrenia: a population-based case-control study. Arch Gen Psychiatry 67(9):889–894

39. Khandaker GM, Zimbron J, Lewis G, Jones PB (2013) Prenatal maternal infection, neurodevelopment and adult schizophrenia: a systematic review of population-based studies. Psychol Med 43(2):239–257

40. Brown AS (2012) Epidemiologic studies of exposure to prenatal infection and risk of schizophrenia and autism. Dev Neurobiol 72(10):1272–1276

41. Cohen A, Patel V, Thara R, Gureje O (2008) Questioning an axiom: better prognosis for schizophrenia in the developing world? Schizophr Bull 34(2):229–244

42. Aleman A, Kahn RS, Selten JP (2003) Sex differences in the risk of schizophrenia: evidence from meta-analysis. Arch Gen Psychiatry 60(6):565–571

43. McGrath J, Saha S, Welham J, El Saadi O, MacCauley C, Chant D (2004) A systematic review of the incidence of schizophrenia: the distribution of rates and the influence of sex, urbanicity, migrant status and methodology. BMC Med 2:13

44. Tandon R, Keshavan MS, Nasrallah HA (2008) Schizophrenia, "just the facts" what we know in 2008. 2. Epidemiology and etiology. Schizophr Res 102(1–3):1–18

45. Gunderson KL, Steemers FJ, Lee G, Mendoza LG, Chee MS (2005) A genome-wide scalable SNP genotyping assay using microarray technology. Nat Genet 37(5):549–554

46. Wang DG, Fan JB, Siao CJ et al (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280(5366):1077–1082

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

48. Walsh T, McClellan JM, McCarthy SE et al (2008) Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science 320(5875):539–543

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

50. Psychiatric Genetics Consortium (2014) in preparation

51. Sullivan PF, Neale MC, Kendler KS (2000) Genetic epidemiology of major depression: review and meta-analysis. Am J Psychiatry 157(10):1552–1562

52. Abecasis GR, Altshuler D, Auton A et al (2010) A map of human genome variation from population-scale sequencing. Nature 467(7319):1061–1073

53. Sebat J, Lakshmi B, Troge J et al (2004) Large-scale copy number polymorphism in the human genome. Science 305(5683):525–528

54. Iafrate AJ, Feuk L, Rivera MN et al (2004) Detection of large-scale variation in the human genome. Nat Genet 36(9):949–951

55. Uher R (2009) The role of genetic variation in the causation of mental illness: an evolution-informed framework. Mol Psychiatry 14(12):1072–1082

56. Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M (2008) Strong association of de novo copy number mutations with sporadic schizophrenia. Nat Genet 40(7):880–885

57. Jaffe AE, Eaton WW, Straub RE, Marenco S, Weinberger DR (2013) Paternal age, de novo mutations and schizophrenia. Mol Psychiatry [Epub ahead of print]

58. Malaspina D, Corcoran C, Fahim C et al (2002) Paternal age and sporadic schizophrenia: evidence for de novo mutations. Am J Med Genet 114(3):299–303

59. Stefansson H, Sigurdsson E, Steinthorsdottir V et al (2002) Neuregulin 1 and susceptibility to schizophrenia. Am J Hum Genet 71(4):877–892

60. Blackwood DH, Fordyce A, Walker MT, St Clair DM, Porteous DJ, Muir WJ (2001) Schizophrenia and affective disorders—cosegregation with a translocation at chromosome 1q42 that directly disrupts brain-expressed genes: clinical and P300 findings in a family. Am J Hum Genet 69(2):428–433

61. Lewis CM, Levinson DF, Wise LH et al (2003) Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: schizophrenia. Am J Hum Genet 73(1):34–48

62. Ng MY, Levinson DF, Faraone SV et al (2009) Meta-analysis of 32 genome-wide linkage studies of schizophrenia. Mol Psychiatry 14(8):774–785

63. Magri C, Sacchetti E, Traversa M et al (2010) New copy number variations in schizophrenia. PLoS One 5(10):e13422

64. Rees E, Walters JT, Georgieva L et al (2013) Analysis of copy number variations at 15 schizophrenia-associated loci. Br J Psychiatry [Epub ahead of print]

65. International Schizophrenia Consortium (2008) Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature 455(7210):237–241

66. Levinson DF, Duan J, Oh S et al (2011) Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. Am J Psychiatry 168(3):302–316

67. Mulle JG, Dodd AF, McGrath JA et al (2010) Microdeletions of 3q29 confer high risk for schizophrenia. Am J Hum Genet 87(2):229–236

68. Shifman S, Johannesson M, Bronstein M et al (2008) Genome-wide association identifies a common variant in the reelin gene that increases the risk of schizophrenia only in women. PLoS Genet 4(2):e28

69. O'Donovan MC, Craddock N, Norton N et al (2008) Identification of loci associated with schizophrenia by genome-wide association and follow-up. Nat Genet 40(9):1053–1055

70. Maher B (2008) Personal genomes: the case of the missing heritability. Nature 456(7218):18–21

71. Derks EM, Vorstman JA, Ripke S, Kahn RS, Ophoff RA (2012) Investigation of the genetic association between quantitative measures of psychosis and schizophrenia: a polygenic risk score analysis. PLoS One 7(6):e37852

72. Terwisscha van Scheltinga AF, Bakker SC, van Haren NE et al (2013) Genetic schizophrenia risk variants jointly modulate total brain and white matter volume. Biol Psychiatry 73(6):525–531

73. Evans DM, Visscher PM, Wray NR (2009) Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. Hum Mol Genet 18(18):3525–3531

74. Dudbridge F (2013) Power and predictive accuracy of polygenic risk scores. PLoS Genet 9(3):e1003348

75. Lee SH, DeCandia TR, Ripke S et al (2012) Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. Nat Genet 44(3):247–250

76. International HapMap Consortium (2005) A haplotype map of the human genome. Nature 437(7063):1299–1320

77. de Candia TR, Lee SH, Yang J et al (2013) Additive genetic variation in schizophrenia risk is shared by populations of African and European descent. Am J Hum Genet 93(3):463–470

78. Abel KM, Drake R, Goldstein JM (2010) Sex differences in schizophrenia. Int Rev Psychiatry 22(5):417–428

79. Avshalom Caspi RMH, Belsky DW et al (2013) The p factor: one general psychopathology factor in the structure of psychiatric disorders? Clin Psychol Sci [Epub ahead of print]

80. Lee SH, Ripke S, Neale BM et al (2013) Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet 45(9):984–994

81. Smoller JW, Craddock N, Kendler K et al (2013) Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet 381(9875):1371–1379

82. Craddock N, O'Donovan MC, Owen MJ (2006) Genes for schizophrenia and bipolar disorder? Implications for psychiatric nosology. Schizophr Bull 32(1):9–16

83. Chatterjee N, Wheeler B, Sampson J, Hartge P, Chanock SJ, Park JH (2013) Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. Nat Genet 45(4):400–405 (pp 405e401–403)

84. So HC, Sham PC (2010) A unifying framework for evaluating the predictive power of genetic variants based on the level of heritability explained. PLoS Genet 6(12):e1001230

85. Janssens AC, Moonesinghe R, Yang Q, Steyerberg EW, van Duijn CM, Khoury MJ (2007) The impact of genotype frequencies on the clinical validity of genomic profiling for predicting common chronic diseases. Genet Med 9(8):528–535

86. Machiela MJ, Chen CY, Chen C, Chanock SJ, Hunter DJ, Kraft P (2011) Evaluation of polygenic risk scores for predicting breast and prostate cancer risk. Genet Epidemiol 35(6):506–514

87. Harris JR, Burton P, Knoppers BM et al (2012) Toward a roadmap in global biobanking for health. Eur J Hum Genet 20(11):1105–1111

88. Iyegbe C, Modinos G, Rivera M (2012) Old obstacles on new horizons: The challenge of implementing gene X environment discoveries in schizophrenia research. In: Maddock J (ed) Public Health: Methodology, Environmental and Systems Issues, vol 1. InTech, Croatia, pp 77–106

89. Bhangale TR, Rieder MJ, Nickerson DA (2008) Estimating coverage and power for genetic association studies using near-complete variation data. Nat Genet 40(7):841–843

90. Kendler KS (1987) Sporadic vs familial classification given etiologic heterogeneity: I. Sensitivity, specificity, and positive and negative predictive value. Genet Epidemiol 4(5):313–330

91. Belsky DW, Moffitt TE, Baker TB et al (2013) Polygenic risk and the developmental progression to heavy, persistent smoking and nicotine dependence: evidence from a 4-decade longitudinal study. JAMA Psychiatry 70(5):534–542

92. McGuire PK, Jones P, Harvey I, Williams M, McGuffin P, Murray RM (1995) Morbid risk of schizophrenia for relatives of patients with cannabis-associated psychosis. Schizophr Res 15(3):277–281

93. Arendt M, Mortensen PB, Rosenberg R, Pedersen CB, Waltoft BL (2008) Familial predisposition for psychiatric disorder: comparison of subjects treated for cannabis-induced psychosis and schizophrenia. Arch Gen Psychiatry 65(11):1269–1274

94. van Os J, Hanssen M, Bak M, Bijl RV, Vollebergh W (2003) Do urbanicity and familial liability coparticipate in causing psychosis? Am J Psychiatry 160(3):477–482

95. Clarke MC, Tanskanen A, Huttunen M, Whittaker JC, Cannon M (2009) Evidence for an interaction between familial liability and prenatal exposure to infection in the causation of schizophrenia. Am J Psychiatry 166(9):1025–1030

96. Margari F, Petruzzelli MG, Lecce PA et al (2011) Familial liability, obstetric complications and childhood development abnormalities in early onset schizophrenia: a case control study. BMC Psychiatry 11:60

97. Wicks S, Hjern A, Dalman C (2010) Social risk or genetic liability for psychosis? A study of children born in Sweden and reared by adoptive parents. Am J Psychiatry 167(10):1240–1246

98. Kendler KS, Gardner CO (2010) Interpretation of interactions: guide for the perplexed. Br J Psychiatry 197(3):170–171

99. Zammit S, Lewis G, Dalman C, Allebeck P (2010) Examining interactions between risk factors for psychosis. Br J Psychiatry 197(3):207–211

100. Rothman KJ (1976) Causes. Am J Epidemiol 104(6):587–592

101. Little J, Higgins JP, Ioannidis JP et al (2009) STrengthening the REporting of Genetic Association Studies (STREGA): an extension of the STROBE statement. PLoS Med 6(2):e22

102. Modinos G, Iyegbe C, Prata D et al (2013) Molecular genetic gene-environment studies using candidate genes in schizophrenia: a systematic review. Schizophr Res 150(2–3):356–365

103. Duncan LE, Keller MC (2011) A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. Am J Psychiatry 168(10):1041–1049

104. Zammit S, Owen MJ, Evans J, Heron J, Lewis G (2011) Cannabis, COMT and psychotic experiences. Br J Psychiatry 199:380–385

105. Caspi A, Moffitt TE, Cannon M et al (2005) Moderation of the effect of adolescent-onset cannabis use on adult psychosis by a functional polymorphism in the catechol-O-methyltransferase gene: longitudinal evidence of a gene X environment interaction. Biol Psychiatry 57(10):1117–1127

106. Smith PG, Day NE (1984) The design of case-control studies: the influence of confounding and interaction effects. Int J Epidemiol 13(3):356–365

107. Di Forti M, Iyegbe C, Sallis H et al (2012) Confirmation that the AKT1 (rs2494732) genotype influences the risk of psychosis in cannabis users. Biol Psychiatry 72(10):811–816

108. Bhattacharyya S, Atakan Z, Martin-Santos R et al (2012) Preliminary report of biological basis of sensitivity to the effects of cannabis on psychosis: AKT1 and DAT1 genotype modulates the effects of delta-9-tetrahydrocannabinol on midbrain and striatal function. Mol Psychiatry 17(12):1152–1155

109. van Winkel R (2011) Family-based analysis of genetic variation underlying psychosis-inducing effects of cannabis: sibling analysis and proband follow-up. Arch Gen Psychiatry 68(2):148–157

110. Kantrowitz JT, Nolan KA, Sen S, Simen AA, Lachman HM, Bowers MB Jr (2009) Adolescent cannabis use, psychosis and catechol-O-methyltransferase genotype in African Americans and Caucasians. Psychiatr Q 80(4):213–218

111. Henquet C, Rosa A, Delespaul P et al (2009) COMT ValMet moderation of cannabis-induced psychosis: a momentary assessment study of 'switching on' hallucinations in the flow of daily life. Acta Psychiatr Scand 119(2):156–160

112. Zammit S, Spurlock G, Williams H et al (2007) Genotype effects of CHRNA7, CNR1 and COMT in schizophrenia: interactions with tobacco and cannabis use. Br J Psychiatry 191:402–407

113. Henquet C, Rosa A, Krabbendam L et al (2006) An experimental study of catechol-O-methyltransferase Val158Met moderation of delta-9-tetrahydrocannabinol-induced effects on psychosis and cognition. Neuropsychopharmacology 31(12):2748–2757

114. Alemany S, Arias B, Aguilera M et al (2011) Childhood abuse, the BDNF-Val66Met polymorphism and adult psychotic-like experiences. Br J Psychiatry 199(1):38–42

115. Muntjewerff JW, Ophoff RA, Buizer-Voskamp JE, Strengman E, den Heijer M (2011) Effects of season of birth and a common MTHFR gene variant on the risk of schizophrenia. Eur Neuropsychopharmacol 21(4):300–305

116. Chotai J, Serretti A, Lattuada E, Lorenzi C, Lilli R (2003) Gene-environment interaction in psychiatric disorders as indicated by season of birth variations in tryptophan hydroxylase (TPH), serotonin transporter (5-HTTLPR) and dopamine receptor (DRD4) gene polymorphisms. Psychiatry Res 119(1–2):99–111

117. Tochigi M, Ohashi J, Umekage T et al (2002) Human leukocyte antigen-A specificities and its relation with season of birth in Japanese patients with schizophrenia. Neurosci Lett 329(2):201–204

118. Narita K, Sasaki T, Akaho R et al (2000) Human leukocyte antigen and season of birth in Japanese patients with schizophrenia. Am J Psychiatry 157(7):1173–1175

119. Nicodemus KK, Marenco S, Batten AJ et al (2008) Serious obstetric complications interact with hypoxia-regulated/vascular-expression genes to influence schizophrenia risk. Mol Psychiatry 13(9):873–877

120. Peerbooms O, Rutten BP, Collip D et al (2012) Evidence that interactive effects of COMT and MTHFR moderate psychotic response to environmental stress. Acta Psychiatr Scand 125(3):247–256

121. Collip D, van Winkel R, Peerbooms O et al (2011) COMT Val158Met-stress interaction in psychosis: role of background psychosis risk. CNS Neurosci Ther 17(6):612–619

122. Keri S, Kiss I, Seres I, Kelemen O (2009) A polymorphism of the neuregulin 1 gene (SNP8NRG243177/rs6994992) affects reactivity to expressed emotion in schizophrenia. Am J Med Genet B Neuropsychiatr Genet 150B(3):418–420

123. Simons CJ, Wichers M, Derom C et al (2009) Subtle gene-environment interactions driving paranoia in daily life. Genes Brain Behav 8(1):5–12

124. van Winkel R, Henquet C, Rosa A et al (2008) Evidence that the COMT(Val158Met) polymorphism moderates sensitivity to stress in psychosis: an experience-sampling study. Am J Med Genet B Neuropsychiatr Genet 147B(1):10–17

125. Stefanis NC, Henquet C, Avramopoulos D et al (2007) COMT Val158Met moderation of stress-induced psychosis. Psychol Med 37(11):1651–1656

126. Aschard H, Lutz S, Maus B et al (2012) Challenges and opportunities in genome-wide environmental interaction (GWEI) studies. Hum Genet 131(10):1591–1613

127. Clayton D, McKeigue PM (2001) Epidemiological methods for studying genes and environmental factors in complex diseases. Lancet 358(9290):1356–1360

128. Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 6(2):95–108

129. Wang WY, Barratt BJ, Clayton DG, Todd JA (2005) Genome-wide association studies: theoretical and practical concerns. Nat Rev Genet 6(2):109–118

130. Kraft P, Yen YC, Stram DO, Morrison J, Gauderman WJ (2007) Exploiting gene-environment interaction to detect genetic associations. Hum Hered 63(2):111–119

131. Ege MJ, Strachan DP (2013) Comparisons of power of statistical methods for gene-environment interaction analyses. Eur J Epidemiol 28(10):785–797

132. Sanders AR, Duan J, Levinson DF et al (2008) No significant association of 14 candidate genes with schizophrenia in a large European ancestry sample: implications for psychiatric genetics. Am J Psychiatry 165(4):497–506

133. Collins AL, Kim Y, Sklar P, O'Donovan MC, Sullivan PF (2012) Hypothesis-driven candidate genes for schizophrenia compared to genome-wide association results. Psychol Med 42(3):607–616

134. Borglum AD, Demontis D, Grove J et al (2013) Genome-wide study of association and interaction with maternal cytomegalo-virus infection suggests new schizophrenia loci. Mol Psychiatry. doi:10.1038/mp.2013.2

135. Bermejo JL, Hemminki K (2007) Gene-environment studies: any advantage over environmental studies? Carcinogenesis 28(7):1526–1532

136. Wong MY, Day NE, Luan JA, Wareham NJ (2004) Estimation of magnitude in gene-environment interactions in the presence of measurement error. Stat Med 23(6):987–998