Research in Pharmacy 2023, 13: 8-15
doi: 10.25081/rip.2023.v13.8124
https://updatepublishing.com/journal/index.php/rip

Review Article

ISSN: 2231-539X

Missing heritability paradox in schizophrenia: hypothesis and plausible clues

Thyloor Rama, Tandaga Satyanarayana Nagamani, Hittanahallikoppal Gajendramurthy Gowtham, Halagur Bhogegowda Kiran Kumar*

Department of Biotechnology, Nrupathunga University, Bengaluru-560 001, Karnataka, India

ABSTRACT

Genetic research on schizophrenia, a common psychiatric disease with complex etiology and high (56-80%) heritability, has failed to identify causal genes, variants or causative mechanisms. Given the extensive effort and limited success to date, it is imperative to review potential reasons for this missing heritability. We argue that a successful elucidation of hereditary mechanisms in schizophrenia will likely involve; the identification of discrete endophenotypes; attention to the role of neurodevelopment and cell differentiation; consideration of the genome structure including temporal and spatial patterns, accommodation of environmental effects at the level of gene expression including any sex differences and pattern of mutations including de novo events and the use of analytic techniques that go beyond genome wide association studies. Identification of the heritable component of schizophrenia and sources of "missing heritability" is needed to understand the cause/s of the disorder and to facilitate the development of effective corrective and possibly preventive measures.

Received: November 29, 2022
Revised: February 20, 2023
Accepted: February 22, 2023
Published: March 09, 2023

***Corresponding author:**
Halagur Bhogegowda Kiran Kumar
kirankumarhb@gmail.com

KEYWORDS: Copy number variation, Genome-wide association studies, Missing heritability, Schizophrenia, Single nucleotide polymorphism

INTRODUCTION

Missing Heritability

Over the last four decades, the search for causal genes for a large number of Mendelian disorders has been remarkably successful and has led to a better understanding of the biological causes of these disorders. This understanding has been instrumental in developing several preventive and corrective measures, some currently in practice with still others under development. It has launched a new era in diagnosis, treatment and prediction of the prognosis based on genetic mutations and their functional consequences and has fuelled hopes that similar strategies will be effective in localizing genes and mutations underlying susceptibility to common and complex diseases. Advances in molecular technology have led to a shift away from DNA marker-based strategies towards genome-wide association studies (GWAS). These assess the potential association of hundreds of thousands to millions of genetic markers to the disease phenotype in a single experiment. GWAS approaches are not hypothesis-based and therefore are suitable where we have little understanding of the biological pathways involved. In contrast, candidate gene studies are entirely dependent on an accurate understanding of the underlying biological mechanisms: a factor that may account for the failure to replicate the results of candidate gene studies. The underlying rationale for GWAS is the 'common disease, common variant' hypothesis (Schork *et al.*, 2009).

Early GWAS on complex phenotypes were inadequate in two ways. First, they used too few (thousands) genetic markers and second, they studied a relatively small number (hundreds) of patients. Other limitations included imprecise phenotyping and control groups of questionable comparability. Today GWAS have been aided by comprehensive 'SNP chips' with millions of markers that are able to capture most (if not all) common variation along the genome. They also use thousands of patients and some have moved to international collaborative experiments. However >400 GWAS studies (http://www.genome.gov/gwastudies/) published to date have failed to offer as persuasive an explanation as anticipated. With a few exceptions, the most common variants identified confer relatively small increments in risk (1.1-1.5-fold). Also, the identification of causal variations to date cannot account for estimates of heritability. As an example, over 40 genes identified for human height account for only 5% of the heritability (Visscher, 2008) while only five loci identified for macular degeneration explain over 50% of its heritability estimate

(Maller *et al.*, 2006) This begs the question, where are the genes and genetic mutations that contribute to the high heritability estimates of most complex phenotypes and disorders?

A review by Manolio *et al.* (2009) has termed this phenomenon missing heritability. Manolio argues that a search for missing heritability is important because understanding this genetic variation will contribute to better diagnosis, prevention, and treatment of diseases which are common and cause major health and societal burdens. It appears that there are many contributors to missing heritability. They range from experimental efficiency to biological assumptions regarding the genetic and environmental architecture of the disorder. Predisposition may involve a large number of variants of small effects and/or rare variants with larger effects.

Also, current studies are still plagued by low power to detect potential gene-gene and gene-environment interactions; inadequate accounting for a shared environment among relatives and still unknown and poorly understood genetic alterations during development, differentiation and ageing. Furthermore, some complex disorders, such as neuropsychiatric disorders have no biological diagnostic tests and are especially prone to diagnostic uncertainty and error. This is particularly problematic in situations where subjects are assessed by different diagnosticians such as when samples are pooled in order to achieve sufficient numbers for analysis. Finally, most GWAS have been undertaken on European populations rather than using a representative sample of humans.

Each complex disorder poses a unique set of challenges in exposing heritability at the level of genetic/epigenetic variation. In this paper we will focus specifically on the missing heritability in schizophrenia, a disease where attempts thus far have not led to expected success.

Schizophrenia: As a Complex Disorder

Schizophrenia (SCZ) (MIM 1815000) is a severe neuropsychiatric disorder characterized by abnormalities in the perception, and experience of reality and social dysfunction. Individuals with SCZ present with auditory hallucinations, paranoid or bizarre delusions, and disorganization. SCZ is a complex disorder with overlapping yet diverse symptoms. Diagnosis is based on the symptoms outlined in the diagnostic and statistical manual of mental disorders DSM IV rather than on any biological markers and thus relies on the patient's self-reported experiences and observed behaviour. Furthermore, there is uncertainty about whether the diagnosis represents a single disorder or a number of discrete syndromes with different causation.

Studies suggest that genetics and early environmental and social processes are important factors contributing to causation (Pedersen & Mortensen, 2006). Recreational and some prescription drugs can cause or worsen symptoms (Donoghue *et al.*, 2009). Given its worldwide and lifetime prevalence of 1% (Perala, 2007), it is a major health and societal burden. It is notable that monozygotic (MZ) twins, who apparently share 100% of their genetic makeup, is concordant in only ~50% (Gottesman, 1991) of cases. Moreover, the degree of manifestation in terms of timing and severity may vary among individuals affected with this disease, even in the same family. The disease clusters in some families and genetic relatedness to an affected individual is one of the greatest risk factors (Sullivan, 2005). Sullivan *et al.* (2003) used meta-analysis to assess the heritability of SCZ from pooled twin data. They found evidence for substantial additive genetic effects, with an 81% (95% confidence interval, 73%-90%) estimate of heritability in liability to SCZ. These results are similar to most other published results. Recently, Lichtenstein *et al.* (2009) used familial relationships to assess the heritability of SCZ and bipolar disorder. They found that the heritability for SCZ and bipolar disorder was 64% and 59%, respectively. Shared environmental effects were modest (SCZ: 4.5%, 4.4%, 7.4%; bipolar disorder: 3.4%, 2.3%, 6.2%) for both disorders. The co-morbidity between disorders was mainly (63%) due to additive genetic effects common to both disorders. The results are consistent with a view of SCZ as a complex trait that results from genetic and environmental etiological influences. It suggests that components of genetic variation, particularly additive genetic variation may account for up to 80% of the phenotypic variability in the population and some causal genes may have pleiotropic effects, yielding endophenotypes that may be common to neuropsychiatric disorders beyond SCZ.

What is Heritability?

The heritability of a trait is defined as the proportion of phenotypic variance in a population attributable to genetic factors (broad sense heritability, H) or additive genetic factors (narrow-sense heritability, h^2). This ratio of variances is used to assess if genetic variations play a significant role in explaining the observed phenotypic variation of a trait. For most traits, heritability will vary across environments and populations and over time depending on the manifestation and degree of genetic variation. In humans, it is typically estimated from the discordance of MZ and dizygotic twins (DZ) [broad sense heritability as (DZ-MZ)/DZ] or familial relatedness [narrow-sense heritability as additive genetic variance (VA)/phenotypic variance (VP)]. Here, the estimate of narrow-sense heritability is based on presumed familial relatedness which may not be realistic. For example, full sibs are expected to share on average half their genetic complement, but this proportion can vary in one large study it ranged from 0.37 to 0.6212. The underlying assumptions permit over/under estimates of heritability, but more often such estimates are missinterpreted as transmissibility. In future, it will be possible to refine these estimates using actual genetic relatedness based on the degree of genetic sharing from genome-wide assessments.

What Do We Know about the Genetics of SCZ?

Over 40 years of genetic research on SCZ have established that the disease is multifactorial and genetically heterogeneous. A number

of genes predispose to the development of schizophrenia (http://www.schizophreniaforum.org/res/sczgene/default.asp) and most have relatively small effects (Allen *et al.*, 2008). Recent studies of CNV indicate the involvement of a large number of rare/very rare variants of major effects (The International Schizophrenia Consortium, 2008; Walsh *et al.*, 2008). Such results are not compatible with the common disease-common variant hypothesis. However, these results will accommodate the role of background genotype (epistasis). Further, some of the SCZ-associated variants argue for a pleiotropic effect as some genetic variants are seen in a number of related disorders (Lichtenstein *et al.*, 2009). Once again, consistent evidence for any single gene in a causal role for SCZ has not been unequivocally established. This dilemma is exemplified by the results of a number of more recent and comprehensive studies that have concentrated on GWAS. This approach is considered highly effective for three reasons. First, analysis using high-density arrays with markers covering the complete genome is now financially practical. A large amount of data can be gathered for a relatively small investment. Second, an ever-increasing number of patients can be included in a meta-analysis that will increase the likelihood of identifying variants of small effects. Thus, the trend is to develop international collaborations which pool different collections of patient DNAs. Finally, the results are generated hypotheses free and used for testing different and unrelated genetic models. Genetic studies on SCZ have particularly relied on GWAS.

There are two features of this dataset that deserve elaboration. First, these reports have, to date, identified over 30 risk markers. Second, most but not all markers have been identified in a single study. Collectively these results appear to argue that SCZ is characterized by rare mutations at a large number of loci and that no single gene/mutation is either necessary or sufficient for the development of this complex disease. While GWAS for complex diseases or traits has now been with us for a number of years, mental disorders have proven particularly resistant to this approach.

Thanks to the sharing of data between three large consortia, a meta-analysis was carried out on GWAS results of European descent (Shi *et al.*, 2009; Stefansson *et al.*, 2009; The International Schizophrenia Consortium, 2009). All three resulting papers report that they were able to see associations between SCZ and SNPs in or close to the major histocompatibility complex (MHC) on chromosome 6. The results have rekindled speculation that SCZ may be a response to infection. This is not a new hypothesis (Crow, 1988; O'Reilly & Singh, 1996; Yolken *et al.*, 2000; Brown, 2006). It is known that some events early in life, such as perinatal infections, maternal malnutrition and birth in winter or spring are also risk factors for SCZ. Each of these factors may allow for exposure to infectious agents which in turn could provoke the development of SCZ as adults. So far, hard evidence has been lacking. However, some evidence suggests a role for retroviruses exists. Deb-Rinker *et al.* (1999, 2002) isolated and identified human endogenous retrovirus (HERV) sequences from the genomic DNA of affected members of three MZ twin pairs discordant for SCZ. This was followed by a report of the HERV-W family of endogenous retroviruses in the cerebrospinal fluid individuals without any evidence of neurological or psychiatric diseases (Karlsson *et al.*, 2001). Similar results were also obtained on brain tissues (Weis *et al.*, 2007) and blood samples (Huang *et al.*, 2006). All provide evidence of endogenous retrovirus activity in cerebrospinal fluids, brains and blood of affected individuals. The delineation of a role for retroviruses or other infectious agents potentially interacting with MHC genes now offers a novel lead in the pathogenesis of this multifactorial disease. It is also noteworthy that, using GWAS, other diseases have shown 'hits' in the MHC, including neurological conditions (Alzheimer's disease) (Bryan *et al.*, 2008) and autoimmune disorders (Crohn's disease, both types of diabetes, and multiple sclerosis) (Bruder *et al.*, 2008).

Recent Novel Insights Through Twin Studies in SCZ

Since the genomes of MZ twins are, at a minimum, almost identical, cases of discordance provide opportunities to identify factors that influence phenotypic differences. Most complex disorders display less than 100% concordance between MZ twins. SCZ is no different in that MZ twins show a concordance rate of ~50% and DZ twins are concordant in only 16% of the cases (Gottesman, 1991). The reduction in concordance from the genetic relatedness (100% for MZ and 50% for DZ twins) has often been explained by reduced penetrance and/or variable expressivity, which assumes that only a proportion of individuals with the causal genotype actually manifest any or even some of the SCZ symptoms. Not surprisingly, epigenetic mechanisms have been implicated in the discordance of MZ twins for SCZ (Singh *et al.*, 2002) and a number of reports have identified gene specific DNA methylation differences between MZ twins discordant for the disease (Petronis *et al.*, 2003; Zhang *et al.*, 2007). Epigenetic differences (loss or gain of CpG dinucleotide methylation) in promoter region methylation have been observed between MZ twins discordant for some diseases (Singh *et al.*, 2004). The significance of somatic mutations or epigenetic changes that create genetic diversity among developing neurons has not been investigated but may explain the discordance of MZ twins for a variety of behaviours including psychosis.

Recent studies have begun to challenge some of the genetic assumptions regarding MZ twins. For example, evidence is accumulating that MZ twins are not only different in epigenetic features but may also be different genetically. Bruder *et al.* (2008) reported that all of the 19 pairs of MZ twins he assessed were different for structural variations commonly referred to as copy number variation (CNV). This led Singh *et al.* (2002) to propose that such structural *de novo* changes operating somatically may account for the discordance of MZ twins for SCZ. Further, using genomic DNA hybridization to the Affymetrix 6.0 array, Singh identified an 11 Kb deletion in the cluster of cadherin genes (CDH 12 and CDH 18) on 5p14 in an affected member of a discordant MZ twin pair (Petronis *et al.*, 2003). The genomic region missing in the affected twin belongs to the type II cadherin gene family which is involved in neuronal cell adhesion (Zhang *et al.*, 2007).

Various neurodevelopmental events (Singh *et al*., 2004) may cause genetic mosaics. For example, Westra *et al*. (2008) showed that aneuploid cells from mitotic nondisjunction are present in the proliferating cerebellum. In fact, they may account for 20% of the cells in different brain regions, particularly neuronal and glial cells. Although not assessed, such mechanisms may cause MZ twins to be discordant for behavioural abnormalities including SCZ. Further, Coufal *et al*. (2009) demonstrated that neural progenitor cells undergo LINE1 (L1) retrotransposition in the hippocampus and several other regions of the brain contributing to individual somatic mosaicism. The results suggest that *de novo* L1 retrotransposition events in the human brain may be programmed and not just an accidental occurrence. It is also likely that some genetic variants can lead to stochastic variation in epigenetic status that in turn causes increased variability for a phenotype (Irizarry *et al*., 2009). The relative frequency of such events leading to differential mosaicism in developing brains will likely differ in MZ twins, adding yet another source of genetic individuality. Unfortunately, this form of genetic variation will not be detected by traditional genetic studies that rely on DNA from blood or buccal cells. We propose that incidental neurodevelopmental episodes (Singh *et al*., 2004) leading to somatic mosaicism via aneuploidy, retrotransposition, CNV, DNA methylation and other mechanisms affecting neurodevelopment explain at least some of the missing heritability in SCZ. A review by Kumar *et al*. (2013) summarises the current status of SCZ genetics and various approaches.

Explaining Missing Heritability of SCZ

Genetic relatedness is the most important risk factor for SCZ, and the heritability estimates for the disease have consistently been high (56-81%). However, an extensive search for the causal genes, mutations and mechanisms using traditional methods has been hampered by poorly understood genetic architecture, including epigenetics and the environment. As discussed above, the missing heritability for SCZ could be attributed to heterogeneous sources and traditional methods (which have been successful in the search for causal genes in Mendelian disorders) have not been effective in identifying it. Could it be that some of the critical assumptions in these experiments have not been fulfilled? The answer to this question must be yes, as the inconclusive and hard to replicate results attest. Could it be that the estimate of heritability is unreliable and represents an overestimate? For a variety of reasons, the answer to this question is probably yes. Further, fundamental gaps in the molecular interpretations of estimates of quantitative genetics and the underlying heterogeneity in SCZ exist. For example, all genetic relatedness (sibs, MZ twins etc.) used in the calculation of heritability are presumed and not determined experimentally. Could it be that the diagnosis of SCZ has been problematic? The answer again is, yes. It is based on a heterogeneous set of symptoms rather than a precise test for an entity called SCZ. Furthermore, not all individuals included in all DNA studies have been assessed by the same criteria or confirmed by following them over time. This may overestimate heritability and the results generated on such samples will be difficult to replicate.

An overestimation of heritability for SCZ will also result from the assumed genetic relatedness of individuals including the genetic identity of MZ twins. The diagnostic heterogeneity and assumptions about the degree of genetic relatedness may also confound the biological hypothesis to be tested. In this case, the additive model involving multiple genes, each with the relatively small effect that forms the foundation for the CD-CV hypothesis will require larger and more genetically heterogeneous samples to test in the future. Quantifying linkage disequilibrium variation between populations will be particularly relevant when amalgamating findings from multiple GWAS (Teo *et al*., 2009).

SCZ appears to be a neurodevelopmental disorder (Rapoport *et al*., 2005). Neurodevelopment is poorly understood but includes neuron formation, migration, synaptogenesis, pruning, apoptosis and activity related changes. These processes determine anatomical structures; establish neural connectivity and communication that maintains cognitive processes (attention, memory, language and emotions). The neurodevelopmental process is often assumed to follow the developmental programs common to other organs and tissues. The prominent feature of these processes is that mitosis follows differentiation (due to differential gene expression), where all daughter cells are expected to be genetically identical. As discussed above, this may not always be the case. In fact, mitosis during neuronal differentiation may yield genetically discordant daughter cells. This could be due to variable and individual specific mosaicism caused by mitotic aneuploidy (Westra *et al*., 2008), CNV, DNA methylation (Fraga *et al*., 2005) or L1 transposition (Coufal *et al*., 2009) among other mechanisms. Many of these mechanisms may be sensitive to environmental factors. Once again, such indirect contributory features of brain function are beyond the scope of traditional genetic studies. Those traditional strategies have been at least partially successful. A large number of alleles, with weak individual association to SCZ, points to a polygenic model. These alleles are not shared with any non-psychiatric disorders but overlap with other psychiatric conditions (The International Schizophrenia Consortium, 2009) and contribute to at ~3% of the heritability of SCZ. A finding of considerable interest is the increased mutational burden in SCZ as conferred by rare CNV (The International Schizophrenia Consortium 2008; Walsh *et al*., 2008). Studies have also revealed an up to an eightfold increase in the rate of *de novo* CNV in sporadic cases compared with familial cases and control subjects (Xu *et al*., 2008). This suggests that there may be distinct differences in the genetic determinations of sporadic versus familial cases of SCZ, which goes beyond the common disease-common variant model, the dictum that has inspired many SNP association studies in SCZ. Such variants may be fully penetrant or involve allelic heterogeneity. It is thought that these rare alleles could be recurrent in the population (either inherited or sporadic) but would remain at low frequency due to the selective disadvantage they confer.

Finally, the parent of origin effect has not been included in any GWAS on SCZ. Recently, Kong *et al*. (2009) genotyped Icelanders and used detailed genealogical information together with the long-range phasing of haplotypes, to determine the

parent of origin for the vast majority of SNPs. More importantly, they identified a set of SNPs whose effect on the phenotype was dependent on whether they were inherited from the mother or the father.

Re-appraisal of the Common Disease-Common Variant (CD-CV) Model

The CD-CV model proposes that common alleles with small to moderate disease risks may have an additive or multiplicative effect on SCZ; this universally accepted dictum has inspired many SNP association studies investigating common polymorphisms in SCZ (Reich & Lander, 2001). However, the CD-CV model alone, has not been able to explain the genetic architecture of SCZ. Another proposed theory that has received increasing attention in the past decade is the CD-RV (common disease-rare variant) model, which suggests heterogeneity of SCZ comes from multiple rare variants (Pritchard, 2001; McClellan *et al*., 2007; Cirulli & Goldstein, 2010; Robinson, 2010).

Recent large-scale association studies have successfully identified numerous risk variants. Recent CNV studies corroborate that rare risk loci, individually or collectively, could predispose to SCZ (The International Schizophrenia Consortium, 2009) supporting the CD-RV model. Apart from locus heterogeneity, CNV studies showed that even within the same disease-associated genomic locus there could be allelic heterogeneity (The International Schizophrenia Consortium, 2008). Some of these rare alleles could be recurrent in the population (either inherited or sporadic) but would remain at low frequency due to the selective disadvantage they confer, while others may be private events found only in singleton individuals or individual families.

Population Genetic as a Confounding Factor in Heritability Estimates

Reproducing any novel discoveries from these genome-wide scans in independent studies is now a prerequisite for the putative findings to be accepted. So, the current trend in genome-wide surveys of common diseases and complex traits fundamentally aims to detect indirect associations where the single nucleotide polymorphisms (SNPs) carrying the association signals are not biologically active but are in linkage disequilibrium (LD) with some unknown functional polymorphisms. The problem of portability of phenotypic associations in replicative studies or meta-analyses is in the patterns of LD between populations. Genome-wide analyses of LD variations between populations that allow the identification of candidate regions with a different pattern of LD are the new way to identify a disease association in a disparate population. Meta-analyses of genome scans for the same diseases across different cohorts, in particular across different ethnic populations, will be greatly enhanced by first understanding the extent of LD differences in candidate regions between these cohorts, since true association signals can be weakened in the presence of significant variations in regional LD when the underlying causal variants exist on different haplotypic backgrounds. The next phase of genetic studies will aim to progress beyond identifying associations to establishing the causal mechanisms of disease. This is where the ability to quantify LD variation between populations will be particularly relevant when amalgamating findings from multiple genome-wide scans of the same disease.

Other Probable Sources of Missing Heritability

a. Genetic interactions as a contributing factor

Quantitative geneticists have long known that genetic interactions can affect heritability calculations. Biological processes often depend on the rate-limiting value among multiple inputs, such as the levels of components of a molecular complex required in stoichiometric ratios. Geneticists have harnessed the genetic interactions model and have identified several loci with additive effects in several human diseases, viz., Hirschsprung's disease, ankylosing spondylitis, psoriasis and type I diabetes. The potential magnitude of missing heritability is well illustrated in Crohn's disease, for which GWAS have so far identified 71 risk associated loci (13). Under the usual assumption that the disease arises from a strictly additive genetic architecture, these loci explain only 21.5% of the estimated heritability.

b. Epigenetics

The determinants of this phenomenon are hypothesized to include environmental factors and still-unknown epigenetic mechanisms. Epigenetics refers to modifications to DNA, histones and nucleosomes which are reversible, affected by the environment and may persist over generations (Bird, 2007). Similarly, some heritable genetic variants can lead to stochastic variation in epigenetic status (DNA methylation) (Javierre *et al*., 2009) and might contribute to the low heritability of some phenotypic effects.

c. Other forms of variation could account for missing heritability

In the last few years, research has focused on single nucleotide polymorphisms (SNPs) and recently copy number variation (CNV) in the search for underlying genetic aetiology of complex disorders enabling interrogation of hundreds of thousands of SNPs/CNVs in one assay. However, SNPs are not the only form of genetic variation other forms of variation viz., Variable number tandem repeats (VNTR) polymorphisms, and insertion/deletion (indels) have been consistently shown to be associated with complex disorders including SCZ and ignored as 'junk' DNA, viz., 1. VNTR lies within the third exon of the dopamine receptor D4 (*DRD4*) gene. 2. VNTR polymorphisms within the serotonin transporter (*5HTT*) gene and the monoamine oxidase A (*MAOA*) gene (Brookes, 2013). In summary, the investigations involving VNTR and other forms of repetitive DNA still hold substantial potential for a role in complex disorders via possible functional properties. These variations could perhaps account for the missing heritability.

d. Gender difference as contributors of missing heritability

There is now ample evidence of gender differences in basic neural processes and behaviours. Behaviours in mammals can

be considered sexually dimorphic. It is proposed to be the end result of reciprocal influences between genes, and activational effects of neuroactive hormones and steroid receptors on the brain, learning, social and other environmental influences. Gender differences in central 5-HT neurotransmission appear to depend partly on sex-related variation mechanisms (Loucif *et al*., 2006). Sex differences could result from parent-of-origin effects, linkage to sex chromosomes, genetic interaction, or from differences arising from sex-specific hormonal environments.

e. Parent-of-origin effect

Although many associations between sequence variants and human traits have been discovered through genome-wide associations, the impact of parental origin has largely been ignored in complex disorders. Effects of susceptibility variants may depend on from which parent they are inherited. The allele that confers risk when paternally inherited is protective when maternally transmitted. The most obvious scheme is imprinting in which the effect is limited to the allele inherited from a parent of a specific sex (Kim *et al*., 2013). Most reports of genome wide association studies have treated the paternal and maternal alleles as exchangeable because the information required is often unavailable. As a result, it reduces the power of such studies to discover some susceptibility variants and underestimates the effects of others, contributing to unexplained heritability. Even when association can be established, the true effect is underestimated e.g., Imprinting studies in SCZ (Gulyás-Kovács *et al*., 2018). Although many mechanisms can lead to parental-origin-specific association with a phenotype, sequence variants located close to imprinted genes are more likely to exhibit such behaviour (Maja *et al*., 2014). The complexity of imprinted regulation in brain, and its central roles in neural processes are becoming increasingly appreciated (Perez *et al*., 2015). Determination of the parental origin of alleles in large samples opens new avenues to study associations between sequence variants and human traits.

f. Genetic drift

Genetic drift is the fortuitous occurrence due to bottleneck and/or founder effects that may result in the fixation of rare harmful alleles modifying the genome of populations. Further, the Hardy–Weinberg principle states that within sufficiently large populations, the allele frequencies remain constant over generations unless the equilibrium is disturbed by migration, genetic drift, genetic mutations, gene flow, or selection. The effect of genetic drift is more notable when few copies of an allele exist, and the effect is less notable when many copies exist. Ashraf *et al*. (2014) using a 1000 genome dataset demonstrates that genetic drift can strongly affect the joint distribution of effect size and SNP frequency and that the bias can be positive or negative depending on subtle details in complex traits.

g. Epigenetics

Epigenetics, the determinants of this phenomenon are hypothesized to include environmental factors and still-unknown epigenetic mechanisms. Epigenetics refers to modifications to DNA, histones and nucleosomes which are reversible, affected by the environment and may persist over generations (Bird, 2007). Similarly, some heritable genetic variants can lead to stochastic variation in epigenetic status (DNA methylation) and might contribute to the low heritability of some phenotypic effects (Javierre *et al*., 2009).

Consistent components of complex traits, such as those linked to human stature/height, fertility, food metabolism and hereditary defects have been shown to respond to environmental or nutritional condition and to be epigenetically inherited (Tripaldi *et al*., 2013). To this end growth anomaly, viz., Prader–Willi, Angelman's, and Rett syndrome are shown to be under heritable epigenetic control (Simeone & Alberti, 2014). Further, altered balance of epigenetic networks has been reported to cause major pathologies, in complex phenotype syndromes (Esteller *et al*., 2001; Melanie *et al*., 2019). Another unique feature of epigenetics is its transgenerational inheritance either in male or female gametes (Anway *et al*., 2005). Thus, epigenetic programs contribute to the "missing heritability" in studies of complex traits.

h. Modifier Genes/Intermediate phenotypes

In complex diseases, many variants are expected to influence several intermediate traits termed as modifiers (Riordan & Nadeau, 2017). Several such examples include Age-of Onset (AAO) LRRK2^{G2019S} mutation in Parkinson (Trinh *et al*., 2016). Some individuals that carry disease alleles are nevertheless healthy despite affected family members in the same environment (example-Dtnbp1 haplotypes in SCZ) (Bergen, 2014). Thus, identifying disease-associated modifier genes/genetic variants and the underlying genetic background, their mode of action is essential to advance our understanding of complex traits and could provide clues to missing heritability.

CONCLUSION

In conclusion, the missing heritability in SCZ appears to be accounted for by a number of factors not all have been addressed in the review. These include imprecise diagnosis of the disease and/or its specific endophenotypes, limited understanding of neurodevelopment including the factors which may affect it, imprecision in the determination of genetic relatedness, the inclusion of patients and controls representing diverse groups of humans and the lack of consideration of parental origin of genetic markers. Future models must take into consideration the relatively small and variable effects of rare and singleton alleles and other forms of variation, viz., VNTR. Also, incorporating epigenetics and parent of origin will add value to the gene hunting exercise.

REFERENCES

Allen, N. C., Bagade, S., McQueen, M. B., Ioannidis, J. P., Kavvoura, F. K., Khoury, M. J., Tanzi, R. E., & Bertram, L. (2008). Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nature Genetics, 40*, 827-834. https://doi.org/10.1038/ng.171

Anway, M. D., Cupp, A. S., Uzumcu, M., & Skinner, M. K. (2005). Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science, 308*(5727), 1466-1469. https://doi.org/10.1126/science.1108190TI

Ashraf, B. H., Jensen, J., Asp, T., & Janss, L. L. (2014). Association studies using family pools of outcrossing crops based on allele-frequency

estimates from DNA sequencing. *Theoretical and Applied Genetics, 127*, 1331-1341. https://doi.org/10.1007/s00122-014-2300-4

Bergen, S. E. (2014). Genetic Modifiers and Subtypes in Schizophrenia. *Current Behavioral Neuroscience Reports, 1*, 197-205. https://doi.org/10.1007/s40473-014-0025-y

Bird, A. (2007). Perceptions of Epigenetics. *Nature, 447*, 396-398. https://doi.org/10.1038/nature05913

Brookes, K. J. (2013). The VNTR in Complex Disorders: The Forgotten Polymorphisms? A Functional Way Forward? *Genomics*, *101*(5), 273-281. https://doi.org/10.1016/j.ygeno.2013.03.003

Brown, A. S. (2006). Prenatal infection as a risk factor for schizophrenia. *Schizophrenia Bulletin, 32*(2), 200-202. https://doi.org/10.1093/schbul/sbj052

Bruder, C. E. G., Piotrowski, A., Gijsbers, A. A. C. J., & Anderson, R., Erickson, S., de Stahl, T. D., Menzel, U., Sandgren, J., von Tell, D., Poplawski, A., Crowley, M., Crasto, C., Partridge, E. C., Tiwari, H., Allison, D. B., Komorowski, J., van Ommen, G.-J., Boomsma, D. I., Pedersen, N. L.,…Dumanski, J. P. (2008). Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *American Journal of Human Genetics, 82*(3), 763-771. https://doi.org/10.1016/j.ajhg.2007.12.011

Bryan, K. J., Zhu, X., Harris, P. L., Perry, G., Castellani, R. J., Smith, M. A., & Casadesus, G. (2008). Expression of CD74 is increased in neurofibrillary tangles in Alzheimer's disease. *Molecular Neurodegeneration, 3*, 13. https://doi.org/10.1186/1750-1326-3-13

Cirulli, E. T., & Goldstein, D. B. (2010). Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature Reviews Genetics, 11*, 415-425. https://doi.org/10.1038/nrg2779

Coufal, N. G., Garcia-Perez, J. L., Peng, G. E., Yeo, G. W., Mu, Y., Lovci, M. T., Morell, M., O'Shea, K. S., Moran, J. V., & Gage, F. H. (2009). L1 retrotransposition in human neural progenitor cells. *Nature, 460*, 1127-1131. https://doi.org/10.1038/nature08248

Crow, T. J. (1998). The viral theory of schizophrenia. *The British Journal of Psychiatry, 153*(4), 564-578.

Deb-Rinker, P., Klempan, T. A., O'Reilly, R. L., Torrey, E. F., & Singh, S. M. (1999). Molecular characterization of a MSRV-like sequence identified by RDA from monozygotic twin pairs discordant for schizophrenia. *Genomics, 61*(2), 133-44. https://doi.org/10.1006/geno.1999.5946

Deb-Rinker, P., O'Reilly, R. L., Torrey, E. F., & Singh, S. M. (2002). Molecular characterization of a 2.7-kb, 12q13-specific, retroviral-related sequence isolated by RDA from monozygotic twin pairs discordant for schizophrenia. *Genome, 45*(2), 381-390. https://doi.org/10.1139/g01-152

Donoghue, K., Medley, I., Brewin, J., Glazebrook, C., Mason, P., Cantwell, R., Jones, P.B., Harrison, G., & Doody, G. A. (2009). The association between substance misuse and first-episode psychosis in a defined UK geographical area during the 1990s. *Social Psychiatry & Psychiatric Epidemiology, 46*, 137-142. https://doi.org/10.1007/s00127-009-0175-5

Esteller, M., Corn, P. G., Baylin S. B., & Herman J. G. (2001) A gene hypermethylation profile of human cancer. *Cancer Research, 61*(8), 3225-3229.

Fraga, M. F., Ballestar, E., Paz, M. F., Ropero, S., Seiten, F., Ballester, M. L., Heine-Suñer, L., Cigudosa, J. C., Urioste, M., Benitez, J., Boix-Chornet, M., Sanchez-Aguilera, A., Ling, C., Carlsson, E., Poulsen, P., Vaag, A., Stephan, Z., Spector, T. D., Wu, Y.-Z.,…Esteller, M. (2005). Epigenetic differences arise during the lifetime of monozygotic twins. *Proceedings of the National Academy of Sciences, 102*(30), 10604-10609. https://doi.org/10.1073/pnas.0500398102

Gottesman, I. I. (1991). Schizophrenia genesis: The Origins of Madness. New York, US: W. H. Freeman and Company.

Gulyás-Kovács, A., Keydar, I., Xia, E., Fromer, M., Hoffman, G., Ruderfer, D., Sachidanandam, R., & Chess, A. (2018). Unperturbed expression bias of imprinted genes in schizophrenia. *Nature Communications, 9*, 2914. https://doi.org/10.1038/s41467-018-04960-9

Huang, W.-J., Liu, Z.-C., Wei, W., Wang, G.-H., Wu, J.-G., & Zhu, F. (2006). Human endogenous retroviral pol RNA and protein detected and identified in the blood of individuals with schizophrenia. *Schizophrenia Research*, *83*(2-3), 193-199. https://doi.org/10.1016/j.schres.2006.01.007

Irizarry, R. A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., Ji, H., Potash, J. B., Sabunciyan, S., & Feinberg, A. P. (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature Genetics, 41*, 178-186. https://doi.org/10.1038/ng.298

Javierre, B. M., Fernandez, A. F., Richter, J., Al-Shahrour, F., Martin-Subero, J. I., Rodriguez-Ubreva, J., Berdasco, M., Fraga, M. F., O'Hanlon, T. P., Rider, L. G., Jacinto, F. V., Lopez-Longo, F. J., Dopazo, J., Forn, M., Peinado, M. A., Carreño, L., Sawalha, A. H., Harley, J. B., Seibert, R.,…Ballestar, E. (2009). Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. *Genome Research, 20*(2), 170-179. https://doi.org/10.1101/gr.100289.109

Karlsson, H., Bachmann, S., Schroder, J., McArthur, J., Torrey, E. F., & Yolken, R. H. (2001). Retroviral RNA identified in the cerebrospinal fluids and brains of individuals with schizophrenia. *Proceedings of the National Academy of Sciences, 98*(8), 4634-4639. https://doi.org/10.1073/pnas.061021998

Kim, Y., Ripke, S., Kirov, G., Sklar, P., Purcell, S. M., Owen, M. J., O'Donovan, M. C., & Sullivan, P. F. (2013). Non-random mating, parent-of-origin, and maternal-fetal incompatibility effects in schizophrenia. *Schizophrenia Research, 143*(1), 11-17. https://doi.org/10.1016/j.schres.2012.11.002

Kong, A., Steinthorsdottir, V., Masson, G., Thorleifsson, G., Sulem, P., Besenbacher S., Jonasdottir, A., Sigurdsson, A., Kristinsson, K. Th., Jonasdottir, A., Frigge, M. L., Gylfason, A., Olason, P. I., Gudjonsson, S. A., Sverrisson, S., Stacey, S. N., Sigurgeirsson, B., Benediktsdottir, K. R., Sigurdsson, H.,…Stefansson, K. (2009). Parental origin of sequence variants associated with complex diseases. *Nature, 462*, 868-874. https://doi.org/10.1038/nature08625

Kumar, H. B. K., Castellani, C., Maiti, S., O' Reilly, R., & Singh, S. M. (2013). Search for missing schizophrenia genes will require a new developmental neurogenomic perspective. *Journal of Genetics, 92*, 335-340. https://doi.org/10.1007/s12041-013-0262-y

Lichtenstein, P., Yip, B. H., Björk, C., Pawitan, Y., Cannon, T. D., Sullivan, P. F., & Hultman, C. M. (2009). Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet, 373*(9659), 234-239. https://doi.org/10.1016/S0140-6736(09)60072-6

Loucif, A. J. C., Bonnavion, P., Macri, B., Golmard, J.-L., Boni, C., Melfort, M., Leonard, G., Lesch, K.-P., Adrien, J., & Jacquin, T. D. (2006). Gender-dependent regulation of G-protein-gated inwardly rectifying potassium current in dorsal raphe neurons in knock-out mice devoid of the 5-hydroxytryptamine transporter. *Journal of Neurobiology, 66*(13), 1475-1488. https://doi.org/10.1002/neu.20321

Maja, K., Dorota, F., Tomasz, A., & Blazej, M. (2014). From Prader–Willi syndrome to psychosis: translating parent-of-origin effects into schizophrenia research. *Epigenomics, 6*(6). https://doi.org/10.2217/epi.14.52

Maller, J., George, S., Purcell, S., Fagerness, J., Altshuler, D., Daly, M. J., & Seddon, J. M. (2006). Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. *Nature Genetics, 38*(9), 1055-1059. https://doi.org/10.1038/ng1873

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S.,…Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature, 461*, 747-753. https://doi.org/10.1038/nature08494

McClellan, J. M., Susser, E., & King, M.-C. (2007). Schizophrenia: a common disease caused by multiple rare alleles. *The British Journal of Psychiatry, 190*, 194-199. https://doi.org/10.1192/bjp.bp.106.025585

Melanie, F., Benjamin, D., Nayla, M., Eugene, T. D., David, C., & Gerard, C. (2019). Epigenetic Factors in Schizophrenia: Mechanisms and Experimental Approaches. *Molecular Neuropsychiatry, 5*(1), 6-12. https://doi.org/10.1159/000495063

O'Reilly, R. L., & Singh, S. M. (1996). Retroviruses and schizophrenia revisited. *American Journal of Medical Genetics, 67*(1), 19-24. https://doi.org/10.1002/(sici)1096-8628(19960216)67:1%3C19::aid-ajmg3%3E3.0.co;2-n

Pedersen, C. B., & Mortensen, P. B. (2006). Urbanization and traffic related exposures as risk factors for schizophrenia. *BMC Psychiatry, 6*, 2. https://doi.org/10.1186/1471-244x-6-2

Perala, J., Suvisaari, J., Saarni, S. I., Kuoppasalmi, K., Isometsa, E., Pirkola, S., Partonen, T., Tuulio-Henriksson, A., Hintikka, J., Kieseppä, T., Härkänen, T., Koskinen, S., Lönnqvist, J. (2007). Lifetime prevalence of psychotic and bipolar I disorder in a general population. *Archives of*

General Psychiatry, 64(1), 19-28. https://doi:10.1001/archpsyc.64.1.19
Perez, J. D., Rubinstein, N. D., Fernandez, D. E., Santoro S. W., Needleman, L. A., Ho-Shing, O., Choi, J. J., Zirlinger, M., Chen, S.-K., Liu, J. S., & Dulac, C. (2015). Quantitative and functional interrogation of parent-of-origin allelic expression biases in the brain. *eLife, 4*, e07860. https://doi.org/10.7554/elife.07860
Petronis, A., Gottesman, I. I., Kan, P., Kennedy, J. L., Basile, V. S., Paterson A. D., & Popendikyte, V. (2003). Monozygotic twins exhibit numerous epigenetic differences: clues to twin discordance? *Schizophrenia Bulletin, 29*(1), 169-178. https://doi.org/10.1093/oxfordjournals.schbul.a006988
Pritchard, J. K. (2001). Are rare variants responsible for susceptibility to complex diseases? *The American Journal of Human Genetics, 69*, 124-137. https://doi.org/10.1086/321272
Rapoport, J. L., Addington, A., & Frangou, S. (2005). The neurodevelopmental model of schizophrenia: what can very early onset cases tell us? *Current Psychiatry Reports, 7*, 81-82. https://doi.org/10.1007/s11920-005-0001-z
Reich, D. E., & Lander, E. S. (2001). On the allelic spectrum of human disease. *Trends in Genetics, 17*(9), 502-510. https://doi.org/10.1016/s0168-9525(01)02410-6
Riordan, J. D., & Nadeau, J. H. (2017). From peas to disease: modifier genes, network resilience, and the genetics of health. *American Journal of Human Genetics, 101*(2), 177-191. https://doi.org/10.1016/j.ajhg.2017.06.004
Robinson, R. (2010). Common disease, multiple rare (and distant) variants. *PLOS Biology*, *8*(1), e1000293. https://doi.org/10.1371/journal.pbio.1000293
Schork, N. J., Murray, S. S., Frazer, K. A., & Topol, E. J. (2009). Common vs. rare allele hypotheses for complex diseases. *Current Opinion in Genetics & Development, 19*(3), 212-219. https://doi.org/10.1016/j.gde.2009.04.010
Shi, J., Levinson, D. F., Duan, J., Sanders, A. R., Zheng, Y., Pe'er, I., Dudbridge, F., Holmans, P. A., Whittemore, A. S., Mowry, B. J., Olincy, A., Amin, F., Cloninger, C. R., Silverman, J. M., Buccola, N. G., Byerley, W. F., Black, D. W., Crowe, R. R., Oksenberg, J. R.,… Gejman, P. V. (2009). Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature, 460*, 753-757. https://doi.org/10.1038/nature08192
Simeone, P., & Alberti, S. (2014). Epigenetic heredity of human height. *Physiological Reports, 2*(6), e12047. https://doi.org/10.14814/phy2.12047
Singh, S. M., McDonald, P., Murphy, B., & O'Reilly, R. (2004). Incidental neurodevelopmental episodes in the etiology of schizophrenia: an expanded model involving epigenetics and development. *Clinical Genetics, 65*(6), 435-440. https://doi.org/10.1111/j.1399-0004.2004.00269.x
Singh, S. M., Murphy, B., & O'Reilly, R. (2002). Epigenetic contributors to the discordance of monozygotic twins. *Clinical Genetics, 62*(2), 97-103. https://doi.org/10.1034/j.1399-0004.2002.620201.x
Stefansson, H., Ophoff, R. A., Steinberg, S., Andreassen, O. A., Cichon, S., Rujescu, D., Werge, T., Pietiläinen, O. P. H., Mors, O., Mortensen, P. B., Sigurdsson, E., Gustafsson, O., Nyegaard, M., Tuulio-Henriksson, A., Ingason, A., Hansen, T., Suvisaari, J., Lonnqvist, J., Paunio, T.,…Collier, D. A. (2009). Common variants conferring risk of schizophrenia. *Nature, 460*, 744-747. https://doi.org/10.1038/nature08186
Sullivan, P. F. (2005). The genetics of schizophrenia. *PLOS Medicine, 2*(7), e212. https://doi.org/10.1371/journal.pmed.0020212
Sullivan, P. F., Kendler, K. S., & Neale, M. C. (2003). Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Archives of General Psychiatry, 60*(12), 1187-1192. https://doi.org/10.1001/archpsyc.60.12.1187
Teo, Y. Y., Fry, A. E., Bhattacharya, K., Small, K. S., Kwiatkowski, D. P., & Clark, T. G. (2009). Genome-wide comparisons of variation in linkage disequilibrium. *Genome Research, 19*(10), 1849-1860. https://doi.org/10.1101/gr.092189.109
The International Schizophrenia Consortium. (2008). Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature, 455*, 237-241. https://doi.org/10.1038/nature07239
The International Schizophrenia Consortium. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature, 460*, 748-752. https://doi.org/10.1038/nature08185
Trinh, J., Gustavsson, E. K., Vilariño-Güell, C., Bortnick S., Latourelle, J., McKenzie, M. B., Tu, C. S., Nosova, E., Khinda, J., Milnerwood, A., Lesage, S., Brice, A., Tazir, M., Aasly, J. O., Parkkinen, L., Haytural, H., Foroud, T., Myers, R. H., Sassi, S. B.,… Farrer, M. J. (2016). DNM3 and genetic modifiers of age of onset in LRRK2 Gly2019Ser parkinsonism: a genome-wide linkage and association study. *Lancet Neurology, 15*(12), 1248-1256. https://doi.org/10.1016/s1474-4422(16)30203-4
Tripaldi, R., Stuppia, L., & Alberti, S. (2013). Human height genes and cancer. *Biochemica et Biophysics Acta (BBA) - Reviews on Cancer, 1836*(1), 27-41. https://doi.org/10.1016/j.bbcan.2013.02.002
Visscher, P. M. (2008). Sizing up human height variation. *Nature Genetics, 40*, 489-490. https://doi.org/10.1038/ng0508-489
Walsh, T., McClellan, J. M., McCarthy, S. E., Addington, A. M., Pierce, S. B., Cooper, G. M., Nord, A. S., Kusenda, M., Malhotra, D., Bhandari, A., Stray, S. M., Rippey, C. F., Roccanova, P., Makarov, V., Lakshmi, B., Findling, R. L., Sikich, L., Stromberg, T., Merriman, B.,… Sebat, J. (2008). Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science, 320*(5875), 539-543. https://doi.org/10.1126/science.1155174
Weis, S., Llenos, I. C., Sabunciyan, S., Dulay, J. R., Isler, L., Yolken, R., & Perron, H. (2007). Reduced expression of human endogenous retrovirus (HERV)-W GAG protein in the cingulate gyrus and hippocampus in schizophrenia, bipolar disorder, and depression. *Journal of Neural Transmission, 114*, 645-655. https://doi.org/10.1007/s00702-006-0599-y
Westra, J. W., Peterson, S. E., Yung, Y. C., Mutoh, T., Barral, S., & Chun, J. (2008). Aneuploid mosaicism in the developing and adult cerebellar cortex. *Journal of Comparative Neurology, 507*(6), 1944-1951. https://doi.org/10.1002/cne.21648
Xu, B., Roos, J. L., Levy, S., van Resenburg, E. J., Gogos, J. A., & Karayiogou, M. (2008). Strong association of de novo copy number mutations with sporadic schizophrenia. *Nature Genetics, 40*, 880-885. https://doi.org/10.1038/ng.162
Yolken, R. H., Karlsson, H., Yee, F., Johnton-Wilson, N. L., & Torrey, E. F. (2000). Endogenous retroviruses and schizophrenia. *Brain Research Reviews, 31*(2-3), 193-199. https://doi.org/10.1016/S0165-0173(99)00037-5
Zhang, A.-P., Yu, J., Liu, J.-X., Zhang, H.-Y., Du, Y.-Y., Zhu, J.-D., He, G., Li, X.-W., Gu, N.-F., Feng, G.-Y., & He, L.(2007). The DNA methylation profile within the 5'-regulatory region of DRD2 in discordant sib pairs with schizophrenia. *Schizophrenia Research, 90*(1-3), 97-103. https://doi.org/10.1016/j.schres.2006.11.007