Abstract. A comprehensive review of the body of genetic studies on schizophrenia seems even more daunting than the battle a psychiatrist wages daily in the office with her archenemy of a thousand faces. The following article reunites some genetic, epigenetic and environmental factors of schizophrenia from revered and vast studies in a chronological and progressive fashion. Twin studies set the basics of heritability and a particular study by Davis and Phelps considers the widely ignored influence of prenatal environment in the development of schizophrenia. Mostly ignited by linkage studies, candidate gene studies explore further by fine-mapping the hypothesized variants [mostly in the forms single nucleotide polymorphisms (SNPs) and less but with greater impact copy number variations (CNVs)] associated with the disease. Genome-wide association studies (GWAS) increase considerably the sample sizes and thus the validity of the results, while the next-generation sequencing (NGS) attain the highest yet unreplicated level of validity results.

Contents
1. Introduction
2. Heritability and twin-studies
3. Linkage studies
4. The search for candidate genes in schizophrenia
5. Genome-wide association studies: A quantum leap
6. Copy number variants: Rare but pivotal alleles
7. Next generation sequencing
8. Conclusions

1. Introduction

Schizophrenia is a common yet devastating mental disorder characterized by a series of cognitive, behavioral and emotional dysfunctions, none of which are pathognomonic for the disease. It includes both positive symptoms, mainly represented by hallucinations and delusions and negative ones, such as blunted affect, avolition and social isolation, along with disturbed attention, executive function and working memory (1). Structural changes in the brain (white matter, grey matter, size) have been shown in controlled neuropathology leading to the idea that schizophrenia might be a true psychosomatic disorder (2). The onset is either abrupt or insidious, anywhere between late adolescence and mid of the 4th decade, followed by an episodic and deteriorating course, with every new episode worsening the prognosis (3,4). Genetic epidemiological studies suggest that the life time risk of developing schizophrenia in general population is 0.5-1%, although it increases considerably when relatives with the disorder are present (5). Even though the established genetic component of schizophrenia is high (heritability of ~0.8), much of its genetic architecture remains unknown (6). As many other complex psychiatric disorders, for example the intrapsychic dissociative phenomenon (also genetically conditioned) (7), schizophrenia is a multifactorial disorder, which encompasses the interplay of multiple susceptibility genes, epigenetic processes and environmental factors (8). Large studies on de novo, common and rare variants pointed to many causal aspects from the N-methyl-D-aspartate receptor signaling and postsynaptic density (PSD), calcium channels, targets of micro-RNA miR-137, glutamate pathways, processes related to neurogenesis and synaptic integrity and yet there are many unknown pathogenic pathways in the etiology of schizophrenia (9-13) as well as linking neuroendocrine aspects during pregnancy and postpartum life of the female patient affecting the new born development (14). We explored the contributions to the understanding of schizophrenia genetics of the classical twin and linkage studies, as well as the newer candidate gene and genome-wide association studies (GWAS), with the implied risk loci and single nucleotide polymorphism (SNP) and the critical, so-called copy number variation (CNV).

2. Heritability and twin-studies

Heritability is a concept that estimates in a population the fraction of differences within a trait that is due to genetic
variability. It can vary from 0 to 1, where 0 means no influence of genetic variation, while 1 shows 100% genetic contribution to the variation. Twin and adoption studies are the most frequently used study designs to portray heritability. The most cited study on heritability is the meta-analysis by Sullivan et al which includes 12 studies of European and Northern American origins and applies a multi-group twin model to conclude a heritability of up to 81% as well as clear evidence for shared environmental influences on schizophrenia (6).

An important discovery of family, twin and adoption studies is the proportional increase in risk for disease with the degree of genetic relationship to a person suffering from schizophrenia. Thus, the risk is approximated at 2% for third-degree relatives, 9% for first-degree relatives, 27% for children of two affected parents and 50% for monozygotic twins (15). As for an adopted child with a schizophrenic biological parent, the risk for disease proves to be 6 to 10 times higher than the general population (16). The nature of heredity in schizophrenia goes beyond the Mendelian rules and percentages of risk, in a complex landscape of different genetic polymorphisms with low penetration and sometimes rare genetic variants of high risk (17).

The evidence that monozygotic (MZ) twins have high concordance of schizophrenia is taken as evidence for genetic influences in the majority of studies in this area (18). Moreover, a study by Waller et al (19) on monozygotic twins reared apart (MZA) concludes that sharing no environmental elements makes the MZA correlation a direct estimate of heritability. Still, a very important aspect is left out of all of the studies, that is, the prenatal environment in which the twins develop and the placentation variants, when considering placenta as the direct physical and psychological link between the mother and child, through which nutrients, drugs, toxins are shared and can affect the neurodevelopment of the embryos. In this case, a maternal infection with MZ twins will most likely affect both of them, while an infection of a DC twin mother is far more likely to pass to one but not the other.

A study by Davis and Phelps (18) tries to investigate this unaccounted for prenatal environment in twins, hypothesizing that monochorionic monozygotic (MC-MZ) twins are more concordant for schizophrenia than dichorionic monozygotic (DC-MZ) twins. Their hypothesis is in accord with the latest studies that show an increased risk of schizophrenia for children exposed to infectious disease especially in the second trimester (20-22). Because there are no studies investigating placenta in twins and the risk for schizophrenia, the authors used mirrored handedness as a retrospective marker for placenta. If placenta occurs within 4 days of fertilization, the twins will be DC-MZ, developing separate placentas and chorions and almost always separate circulations with the mother (23). However, if twinning occurs after day 4, the twins will be MC-MZ and will share fetal circulation in 90% of cases. Twinning after day 4 will result in mirror imaging limited to the ectodermal layer and observed phenotypically as mirror hair swirls, dermal ridge patterns on hands and feet or hand preferences (24,25). Only mirror hand preference was used as placenta marker in the study of 71 pairs of twins. The retrospective marker of mirror handedness for placenta has a few shortcomings that decrease its statistical power: handedness is a characteristic that could be caused by other factors than late twinning such as learning, brain pathology or perinatal stressors and MC twins with same-handedness could be considered DC twins with same handedness or vice versa confounding the study groups. Nevertheless, within the 71 case-study there were far more opposite-hand twin pairs that developed schizophrenia or psychosis (9 out of 15 pairs, 60%) compared with the same hand preference group (18 out of 56, 32.1%) (18). The low number of opposite-hand cases and the rate of MC twinning of generally 60% could mean that the opposite-hand group is in fact MC twins. Likewise, the relatively high concordance (32.1%) for schizophrenia in the DC-MZ twins' group would imply that many of the second study group are MC-MZ same-handed twins (18). A way of validating these results would be a study with clear data for placentation, obtained post-partum; nonetheless, this study covers a mainly ignored possible source of schizophrenia in identical twins.

3. Linkage studies

Coupling studies was one of the first molecular genetic approaches. It resided in the notion that genetic traits located close to each other were more prone to be inherited together compared with traits farther apart (17). Although early linkage studies were inconsistent and hardly replicable, a large meta-analysis found a genome-wide result on 2q, fact confirmed partially five years later in another meta-analysis (26,27).

A more recent and relevant large-scale linkage study has reignited the interest in the disrupted in schizophrenia 1 (DISC1) gene (associated with a large number of cytoskeletal proteins) resulted from a 4q24 translocation and previously described as segregating with psychopathology in a large Scottish family (28).

Other linkage studies using schizophrenia pedigrees of European and African American-decent has identified a larger region on 6q (6q13-q26) as implicated in the disease (29). Also, G72 on chromosome 13q32-34, Epsin 4 on chromosome 5q33 and other genes originally identified in linkage studies have follow up systematic association studies and fine-mapping of the regions (28).

4. The search for candidate genes in schizophrenia

Although gene identification represents an important step to uncovering the complex pathophysiology of schizophrenia, the road to the genetic origins of this disease is strewn with difficulties. It would have been ideal if the binary diagnosis (‘affected’/‘unaffected’) employed in many genetic studies for schizophrenia were directly conclusive, but the vast phenotypic heterogeneity of the disease makes it difficult to account for the differences in representation throughout samples of various components of the illness. No single gene is necessary or sufficient to determine the disease, rather a combination of risk genes with small effects describe the highly heterogeneous genetic basis of schizophrenia (28). Other confounding variables of association studies of schizophrenia using the case-control design include small sample sizes (making the results possible false positive/negative), the unequal representation of allele frequencies in the study groups or the publication bias that results in
negative studies being rarely published in the journals where
the original discovery was published (28-30). The discovery
of susceptibility genes has been made successful as a result
of the completion of sequencing the human genome. With
the new DNA amplification methods, the huge number of
genetic variants in schizophrenia is now easily processed
in association studies (17). The most often studied variants
are SNPs, that are substitutions of single bases within the
genome. Methodologically, the allele frequencies of a group
with a phenotype (e.g. schizophrenia) are compared with a
control group (a population without the disease) (17). Another
variation used in studies is CNV, which is characterized by
duplications or deletions of DNA sequence. This variation is
much less common than SNPs and because it affects a larger
part of a chromosome the risk ration of disease is higher.

A review by Karayiogou and Gorgos (28) brings together
a series of important candidate genes for schizophrenia,
which were identified through systematic positional cloning
in regions of linkage and linkage disequilibrium (LD) mapping
methods. They have highlighted the consistent linkage signal
of the chromosome 13q32-34 and also the implication of a
broad region of the chromosome 8p12-21.

The gene for proline dehydrogenase maps on the 22q11
chromosome and association between hemizygous deletion
of the 22q11 locus and schizophrenia has been previously
stated (28). The gene codes for an enzyme that metabolizes
L-proline, an amino acid that may be directly involved in
glutamatergic transmission, one of the core pathways implied
in schizophrenia (32,33). Through fine-mapping of the
22q11 locus, overexpression of haplotype variants at the 3' end
of the gene has been identified (31,34). This finding has been
confirmed in two independent studies evaluating a large (528)
group of Chinese families, as well as 274 Ashkenazi Jewish
origin families, even though a negative study was also
published (35-37).

In another important study on 360 Iranian subjects
(175 schizophrenic and 185 controls) 3 polymorphisms of the
PRODH gene (757C/T, 1766 A/G and 1852 G/A) were associ-
ated with an increased risk of schizophrenia (38).

Rare variants of the gene have shown to reduce the
activity of the enzyme (31). In an animal-model study on mice
involving the PRODH gene variants, abnormal plasticity of
glutamatergic synapses and dopamine dysregulation in the
frontal cortex have been identified (39,40). The dysregulation
of dopamine generates increased levels of transcripts of the
catecol-O-methyltransferase (COMT) gene, also located on
the 22q11 chromosome. This seems to be a triggered compen-
satory response to the glutamate dysregulation (28).

Another gene of the 22q11 locus is the ZDHHC8, identified
in the same LD screen of the PRODH gene previously presented.
Of the five SNPs identified, one of them (rs175174) was associ-
ated with a 1.5-fold increase risk of the disease (34,41).

Thirdly, the COMT gene is also located in 22q11 region,
somewhere in between the 2 anterior genes described. Its
codes for an enzyme involved in dopamine breakdown and a
variant that modulates enzyme activity (Val, high activity; Met, low activity in the 158 codon) has been especially studied.
The high activity (Val) allele appears to increase the risk of schizophrenia and affect executive function, a
domain dysregulated in the disease, despite the unsubstantial
results (42-45). On animal models, the low activity variant
also proved to increase the risk for disease, the insufficient
enzyme inappropriately metabolizing the increased dopa-
mine (46). The result of the low activity enzyme was replicated
in a follow-up study of children with 22q11 microdeletion
that showed decreased volume of the prefrontal cortex and
levels of cognition, as well as the debut of psychotic episode
in adolescence (47).

Dystrobrevin-binding protein 1 (DTNBP1) or dysbindin
is another gene identified by fine-mapping of 6p24-22 locus,
previously cited in linkage studies. Dysbindin is part of the
dystrophin protein complex (DPC), as well as the biogenesis
of lysosome-related organelle complex (BLOC) (48,49). Two
studies showcased the decrease of DTNBP1 mRNA in the
dorsolateral prefrontal cortex and hippocampus of schizo-
phrenic patients compared with controls (50,51).

The same approach of fine-mapping of 8p12-21 chromo-
some locus, previously identified in linkage studies has given
the Neuroregulin 1 (NRG1) gene. Many differences in the
frequency of haplotypes throughout the samples could indicate
important heterogeneity of the LD structure of NRG1 locus or
the coexistence of several alleles of risk there (52,53). Because
these haplotypes proved no functional relevance, the knockout
NRG1 mice studied cannot be used to model the implication of
the gene in determining schizophrenia (54,55).

The follow-up systematic association studies for the
disrupted in schizophrenia 1 (DISC1) gene identified in
linkage studies, identified a positional candidate on 1q42. Even
tough negative studies were reported, the gene showed asso-
ciation with schizophrenia for allelic heterogeneity (56-58).
The complexity of the gene is supported by different studies
showing variants of the gene involved in altered hippocampal
structure and functions in healthy individuals or visual
working memory performance (59,60). Furthermore, the gene
has unclear implications in development and plasticity, being
involved in a series of cellular functions such as cell migra-
tion, microtubule function, membrane trafficking of receptors,
neurite outgrowth, mitochondrial function and phosphodies-
terase signaling (61).

Also mentioned in the linkage studies chapter is the
trace amine receptor 4 (TAAR6) gene that was subsequently
fine-mapped and identified as a positional candidate for
schizophrenia on the 6q23.2 (62). It is considered a GPCR
extensively expressed in the brain (63).

Another solid linkage signal for both bipolar disorder
and schizophrenia has led to the fine-mapping of the 13q32-34
locus. The results included several SNPs and haplotypes of the
locus associated with schizophrenia in French-Canadian
studies and replicated in Russian ones (60). Through enzym-
ic studies it was shown that G72 modulates the activity of
D-amino acid oxidase (DAO) and in turn affects glutama-
tergic signaling (64).

Similar implication in both schizophrenia and bipolar
disorder were suggested for the abnormal expression of the
carboxyl-terminal PDZ ligand of neuronal nitric oxide
synthase (CAPON) gene (65). CAPON was also shown to
be involved in NMDA receptor-coupled nitric oxide
signaling (66).

Located on chromosome 5q33, Epsin 4 gene was
fine-mapped and 4 haplotypes showed evidence of LD with
schizophrenia (67). The gene codes for a protein involved in the transport and stability of neurotransmitter vesicles at synapses.

Finally, the GABA subunit gene cluster showed in sequential studies evidence for the implication of GABA transmission in schizophrenia (68).

To complement these findings, a separate meta-analysis identified other associated genes at the dopaminergic (DRD2, DRD3 and DRD4) and serotoninergic (HTR2A, SLC6A4 and TPH1) systems, as well as genes affecting neuro development (AHII, MTHFR, RELN and TRKA) (69).

Aside from the described candidate genes, another 12 regions (2p, 5q, 3p, 11q, 2q, 1q, 22q, 8p, 6p, 20p, 13q and 14q) of the human genome were identified to likely contain susceptibility genes for schizophrenia (69,70).

Taking the research results to the next level would mean to identify and characterize the susceptibility genes for schizophrenia in vivo, allowing for development of mechanism-targeted therapies.

5. Genome-wide association studies: A quantum leap

The leap in GWAS studies consists in the number of SNPs tested, which could go up to 10 million in a single experiment compared with only a few SNPs analyzed in candidate gene studies (17). This significant quantitative improvement raises the significance threshold, reducing the chance of false-positives.

Chronologically, in the first major genome-wide study by O’Donovan et al a polymorphism in the zinc finger protein 804 A (ZNF804) gene was associated with schizophrenia (71). The next important study by Stefansson et al (72) brought about associations in the transcription factor 4 (TCF) gene, Neurogranin (NRGRN) gene and the MHC region, the last association being convincingly replicated in large consortia (72-74).

Psychiatric Genome Wide Association Study Consortium (PGC) represents the joint forces of many specialists world-wide and was created out of the need for greater samples, reaching the level of thousands (11). The latest and most impressing study of the Consortium resulted in 128 associations or 108 independent loci with genome-wide relevance, obtained on a remarkable sample of 36,989 patients and 113,075 controls (75).

However, impressing the results of GWAS studies compared with those of the previous era, the case control numbers remain too low to encircle the entire amount of genetic variability that accounts for susceptibility to schizophrenia.

6. Copy number variants: Rare but pivotal alleles

The possibility of assessing millions of SNPs at once introduced by GWAS generated enthusiasm of resolving the genetic basis of schizophrenia. Although considerable genome wide meta-analyses have singled-out many promising candidate genes, they only account for ~2 of 80% heritability in schizophrenia (76,77).

With increasing evidence showing that common variants cover a tiny fraction of heritability, the possible implication of rare variants (both SNPs and CNVs) becomes higher. Moreover, rare de novo CNVs with high penetrance could account for an important piece of the heritability puzzle (78).

By definition, CNVs are structural genomic variants that comprise of duplications, deletions, insertions and translocations of varying sizes from 1 kilobase to several megabase pairs (79). A well-known CNV is the 22q11 deletion, 20-30% of people with it having schizophrenia (80). Concomitantly, the 22q11.2 deletion causes the velocardiofacial syndrome (Di George syndrome), a multiphenotypical severe disorder (81).

In a study by Luo et al employing integration of prioritization data with genetic association and PPI interaction (protein-protein interaction) the following 8 genes were identified to be frequently disrupted by CNVs in schizophrenia cases: NRXN1, CHRNA7, BCL9, CYFIP1, GJA8, NDE1, SNAP29 and GJA5 (79).

7. Next generation sequencing

New advances in sequencing technologies come in the form of cost-effective ‘whole genome sequencing’ and ‘well exome sequencing’. An initial pilot study applying NGS to a trio generated intriguing, yet unsustain results that await to be confirmed (82).

Three other larger studies lead the way in the NGS era. In a study by Gulsuner et al (13) genes involved in the fetal prefrontal cortex neurogenesis were identified after the exome sequencing of 105 patients with schizophrenia, 84 unaffected siblings and 210 unaffected parents. Fromer et al (83), aside from reinstating the genetic overlap of schizophrenia and neurodevelopmental disorders (autism and mental retardation), also signaled the presence of de novo CNVs in the activity-regulated cytoskeleton-associated protein (ARC) or N-methyl-D-aspartate receptor complex (NMDAR). Lastly, Purcell et al (84) confirmed the presence of CNVs in ARC and NMDAR complex as well as in Fragile-X-mental retardation-protein-protein complex (FMRP) and calcium channel complexes.

8. Conclusions

The level and scale of testing the genetic fabric of schizophrenia has come a long way from the first twin-pair studies to the meta-analyses of GWAS involving simultaneous testing of millions of SNPs from thousands of subjects or ‘the whole genome sequencing’ applied in tris or case-controls, courtesy of the NGS. This urge for covering exponentially more genome at once, illustrated even by the creation of Psychiatric Genome Wide Association Study Consortium (PGC) highlights once for all the enormous genetic heterogeneity and still unchartered interdependences of genes in schizophrenia. From the impressing 80% heritability stated by Sullivan, very little has been pinpointed and confirmed in live models or come even close to the ultimate goal of targeted therapy. The closest to this much desired goal are maybe the prioritized CNVs of genes like NRXN1, CHRNA7, BCL9 or CYFIP1. Many confounding variables still plague all levels of testing starting from sample sizes, absence of negative studies of insufficient follow-up for high credibility.

Acknowledgements

Not applicable.
Funding

No funding was received.

Availability of data and materials

Not applicable.

Authors’ contributions

SCT contributed in all the stages of the article preparation, designed the study and revised the manuscript for important intellectual content. BK and AV acquired the data by screening the studies identified on PubMed and drafted the manuscript. BEP contributed to the conception of the study and revised the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References


Evidence for association between Bipolar I disorder and schizophrenia: A highly significant association between a COMT haplotype and schizophrenia. Am J Hum Genet 71: 1296‑1302, 2002.


