Susceptibility genes for schizophrenia and their functional relationships
Luan, Zhilin

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2017

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
General Introduction
1 Schizophrenia

“The psychiatrist who knows schizophrenia knows psychiatry.”

—David A. Lewis, M.D.[1]

Schizophrenia is the most impressive, complicated and thus, challenging mental disorder, which psychiatrists and mental health caregivers confront. It is a heavy burden for the patients and usually occurs during adulthood. Many cases of schizophrenia run a recurrent and chronic course, which very possibly leads to severe disability [2-6]. Schizophrenia is considered as the prototypic mental disorder since it has a devastating impact on the human essence i.e. personality and intellect. However, schizophrenia is also the most widely misunderstood mental disorder. Common misconceptions involve the belief that the patients possess a split or multiple personalities, or apprehend reality with unfettered creativity and originality. These erroneous perceptions reflect the relatively low concern and importance that our society has rendered to mental disorders in general and schizophrenia in particular [7-9].

1.1 Concept and history

Schizophrenia is a long-term psychotic disorder which impairs individuals’ thinking, visual, auditory and behavioral abilities and in turn disrupts the reality perception of the patients [10]. For instance, some people suffering from schizophrenia may believe that others are reading their minds, controlling their thoughts, or trying to harm them. With a blurred border between the real and the imaginary, it becomes, sometimes terrifyingly difficult for schizophrenia patients to make decisions, manage emotions and relate to others. In some cases, even energy, motivation and expressional ability are drained away by the disorder. With a peak onset at the age of 18-25 years, schizophrenia results in the loss of productivity, poses a considerable burden on the relatives, and causes high medical and social costs [11, 12]. Some fortunate individuals manage to recover to a certain degree and reintegrate into society. However, only 14% of the patients maintain sustained recovery five years after the initial episode [13].

Emil Kraeplin first identified schizophrenia in 1893 as an independent mental illness called dementia praecox that was considered a type of dementia with onset at a young age [14]. Afterwards, the term ‘Schizophrenia’ was proposed/introduced by Eugen Bleuler in 1911, derived from the Greek roots ‘schizo (split)’ and ‘phrene (mind)’ pointing to the patients’ fragmented mind [15]. The split mind referred to the dissociation of various mental functions, such as emotions and drives disconnected from thoughts. Yet, if we divert our attention to the primordial ‘psycho-diagnostic soup’ out of which schizophrenia was later to evolve, the illness itself seems to have affected mankind throughout history. Written material with brief notes mentioning irrational, unintelligible or uncontrolled behavior that may imply schizophrenia can be traced back as far as the second millennium before Christ in ancient Egypt [16]. However, objections against these very early ‘cases’ of schizophrenia have been raised by pointing out that although psychosis was described, there was no true account of a condition meeting the present criteria for schizophrenia. Descriptions pointing to schizophrenia were rare before 1800 but the prevalence of insanity rose during the nineteenth century in Western countries. Remarkably, schizophrenia or schizophrenia-like insanity stayed low in the non-Western countries until the twentieth century suggesting that schizophrenia might be considered as an unfortunate by-product of modern ‘western’ civilization. Concomitant with profound social changes
that arose in the Age of Enlightenment, the radical change in the descriptions of mental disorders led to the recognition of a mental condition or a group of mental conditions that heretofore not seemed to have existed [17]. The historical complexity of schizophrenia has been replaced by the current consensus about the identification of schizophrenia based on deliberate diagnostic criteria.

1.2 Diagnosis and symptoms

So far, no valid biological marker has been found and the clinical symptoms are still the only basis for schizophrenia diagnosis [18, 19]. Currently, schizophrenia has been classified into four phases: 1) Risk phase - this mainly includes some well-acknowledged genetic, immune, obstetric and neurochemical risks. 2) Prodromal phase - this may include social isolation, declined performance in daily issues and odd (but not psychotic) thinking. 3) Psychotic phase - this is the ‘acute period’ in which clear psychotic symptoms appear. 4) Chronic disability phase - this may include some psychotic symptoms, but include in particular ‘negative’ ones (loss of motivation and emotion, for example) [20, 21]. Brain changes have been defined after clear symptoms appear, therefore, early detection and prevention has recently become a research interest.

The psychotic and chronic disability phases manifest a series of symptoms sorted into different groups. For the moment, symptoms of schizophrenia have been admittedly sub-grouped into two main categories, positive and negative [18, 22-24]. The positive symptoms are phenomena excessive to normal experience and are remarkable features of the psychotic phase, including hallucinations, delusions, disorganization, formal thought disorders, inappropriate affect and bizarre behavior [25, 26]. These positive symptoms may occur also in the chronic phase, but usually decrease or have a lesser impact on the patients’ life. The negative symptoms of schizophrenia represent the apparent loss of important personality features and functional abilities, prominent during the chronic phase, and include poverty of thought and speech, social withdrawal, decreased spontaneous movement, lack of initiative, poor self-care and unreliability. The positive symptoms can be recognized and measured more easily than the negative symptoms [26]. However, the latter are much more persistent and important for prognosis, and cause more severe disability [27]. Besides the positive and negative symptoms, cognitive deficits are widely proposed in almost all patients [28, 29]. The profile of cognitive deficits is severe and broad, including impairment in attention, abstraction, verbal memory, language and executive functions [30-33]. Patients with schizophrenia tend to score in the lowest 5-10% of the general population on cognitive assessments. Marked cognitive deficits may be detected in at-risk and first-episode studies indicating that impairment grows at onset, and even prior to symptom onset [34, 35].

To achieve international consensus for schizophrenia research, standardized diagnostic criteria for mental disorder assessment have been developed and introduced [36, 37]. The two dominating systems in use nowadays [7] are the Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV, the American Psychiatric Association) and The International Classification of Disease, tenth edition (ICD-10, the World Health Organization), Classification of Mental and Behavioral Disorders. These two criteria adopt patients’ self-reported experience and observed behavioral abnormalities, followed by a professional clinical assessment. Schizophrenia-associated symptoms occur along a continuum in schizophrenic patients and must increase to a certain level before a diagnosis can be made [38]. The ICD-10 system is typically used in European countries, while DSM-IV used in the United States and the rest of the world, and prevailing in research studies.
[39]. Although widely different in the past, the current editions of these two systems are fairly similar. The main difference between DSM-IV and ICD-10 is the duration of certain symptoms before diagnosing schizophrenia: in DSM-IV, 6 months, in ICD-10, one month. Otherwise, six months of social or occupational dysfunction is required according to DSM-IV to obtain the diagnosis of schizophrenia but this is not specified in ICD-10 (Table 1). However, the definition of schizophrenia has evolved through different editions. The DSM-V was delivered with modest changes in comparison to DSM-IV in 2013 and the ICD-11 has been released in 2015 [40].

Table 1. Comparison of the DSM-IV and ICD-10 diagnostic criteria for schizophrenia.

<table>
<thead>
<tr>
<th>DSM-IV</th>
<th>ICD-10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Symptom Duration</strong></td>
<td>≥1 month</td>
</tr>
<tr>
<td>Deterioration should be present from a premorbid level in functioning</td>
<td>≥1 month</td>
</tr>
<tr>
<td>Bizarre delusions (primary or secondary delusions)</td>
<td>Thought echo/insertion/withdrawal/broadcasting</td>
</tr>
<tr>
<td>Voice commenting/conversing</td>
<td>Delusions of control, delusional perception</td>
</tr>
<tr>
<td>Other less specific symptoms: two present</td>
<td>Other persistent hallucinations</td>
</tr>
<tr>
<td>Other delusions</td>
<td>Thought-form disorder</td>
</tr>
<tr>
<td>Disorganized speech</td>
<td>Catatonia</td>
</tr>
<tr>
<td>Grossly disorganized or catatonic behavior</td>
<td>Negative symptoms</td>
</tr>
<tr>
<td>Negative symptoms</td>
<td>Significant personality change</td>
</tr>
<tr>
<td>Affective symptoms</td>
<td>Affective symptoms, if present, must follow psychotic symptoms</td>
</tr>
<tr>
<td>Affective syndrome included if relatively brief</td>
<td></td>
</tr>
<tr>
<td>Social/occupational dysfunction</td>
<td>One or more major areas of dysfunction (work, occupation, interpersonal)</td>
</tr>
<tr>
<td>Exclusion Criteria</td>
<td>Schizoaffective or mood disorder</td>
</tr>
<tr>
<td>Schizoaffective or mood disorder</td>
<td>Schizoaffective or mood disorder</td>
</tr>
<tr>
<td>Substance use or general medical condition</td>
<td>Drug intoxication or withdrawal</td>
</tr>
<tr>
<td>Pervasive developmental disorder</td>
<td>Overt brain disease</td>
</tr>
</tbody>
</table>

Until now, no single symptom turned out to be specifically associated with schizophrenia. Moreover, there is a remarkable variation of schizophrenia among patients as far as symptom character, intensity and frequency is concerned. This clinical heterogeneity indicates that schizophrenia may be a combination of a number of different disorders [41-44]. Therefore a differential and skeptic attitude towards diagnosis is important.

1.3 Treatment and prognosis

There is hitherto no cure for schizophrenia. However, the symptoms of schizophrenia are treatable. Antipsychotic medication treatments are the most commonly applied and effective therapy in controlling positive symptoms in nearly 90% of acute cases. Antipsychotic medicines are small molecules having predominant serotonin-dopamine antagonistic functions. To date, two main classifications of antipsychotic drugs are used; typical- and atypical antipsychotics (Table 2) [45]. Typical antipsychotics sometimes referred to as ‘first generation’ antipsychotics, are potent antagonists of dopamine receptors, which can effectively reduce the severity of positive symptoms. However, they also give rise to extra-pyramidal side effects, characterized by unsteady movements, body stiffness or involuntary tremor. Unlike typical antipsychotics, the newer ‘second-generation’ atypical antipsychotics are comparatively weaker antagonists of dopamine receptors and potent serotonin antagonists with a relative lack of extra-pyramidal side effects. The main side effect of atypical antipsychotics is weight gain. Some atypical antipsychotics seem to have an appetite stimulating effect and cases of diabetes have also been reported. Large clinical studies have compared typical and atypical antipsychotics and it was concluded that newer atypical antipsychotics work no better than older drugs such as haloperidol, at least for the initial treatment of first-episode schizophrenia. However, there is marked enthusiasm for the fact that atypical antipsychotics have less side effects, and therefore are more acceptable to patients for admitting a better improving quality of their life. However, the most severe psychotic patients often fail to fully respond to newer antipsychotics, and older drugs seem more effective in such cases. Many patients respond well to lower doses of typical antipsychotics with minor side-effects. For this reason, in clinical practice, the risks and benefits between the typical and atypical antipsychotics should be carefully balanced.

Table 2. Typical and atypical antipsychotics in common use

<table>
<thead>
<tr>
<th>Typical antipsychotics</th>
<th>Atypical antipsychotics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clozapine</td>
<td>Olanzapine</td>
</tr>
<tr>
<td>Chlorpromazine</td>
<td>Quetiapine</td>
</tr>
<tr>
<td>Fluphenazine</td>
<td>Asenapine</td>
</tr>
<tr>
<td>Haloperidol</td>
<td>Sulpiride</td>
</tr>
<tr>
<td>Mesoridazine</td>
<td>Amisulpiride</td>
</tr>
<tr>
<td>Perphenazine</td>
<td>Aripiprazole</td>
</tr>
<tr>
<td>Thioridazine</td>
<td>Melperone</td>
</tr>
<tr>
<td>Trifluoperazine</td>
<td>Risperidone</td>
</tr>
</tbody>
</table>
The prognosis of schizophrenia is variable and hard to predict. Symptoms of schizophrenia improve after medication in most cases. But nearly 75% of schizophrenia patients continue to suffer from chronic symptoms and may experience intermittent acute relapses with a lifetime treatment becoming inevitable. Many factors contribute to poor prognosis; such as poor social adjustment, onset early in life, predominance of negative symptoms, and delay between onset of disease and medication therapy. Schizophrenia also results in reduced life expectancy of patients, mostly as a consequence of suicide. Community support programs are required to assist the patients to achieve relative social recovery.

1.4 Genetics of schizophrenia

Since long, it is well known that schizophrenia ‘runs in families’ [46, 47]. There is no doubt that the disease has a strong genetic component, which has been consistently indicated by genetic epidemiological studies. In the general population worldwide, the lifetime risk for schizophrenia is about 1% [48, 49]. The first-degree relatives of an individual with schizophrenia have a risk of around 6-13%. The risk rises further to 17% in dizygotic twins and reaches 48% in monozygotic twins. The risk for relatives of a schizophrenia patient to develop schizophrenia is determined by their kinship distance from the patient [50] (Figure 1). The closer the kinship of a subject to the patient, the higher the risk of developing schizophrenia would be for the subject. Despite high genetic risk, genetic influence alone does not fully determine the etiology of schizophrenia; otherwise 100% concordance between monozygotic twins would arise. The heritability (the proportion of the variance in a population related to inherited factors) of schizophrenia has been estimated between 66 and 85 percent, leaving space for the contribution of environmental factors [51]. The inheritance pattern of schizophrenia suggests a non-Mendelian mode of transmission and makes simple major gene dependency impossible. Instead, a polygenic model seems to provide a better explanation [52].

It has been thought that schizophrenia arises from combinations of numerous genetic risk variants each having only modest effects. The schizophrenia risk variant carriers need not manifest the disorder, while affected individuals need not have the verified risk variants [53].

Figure 1. Average risks for developing schizophrenia in different groups of people

Any individual genetic risk variant acts in a probabilistic rather than deterministic fashion. Interactions between the susceptibility genes containing the risk variants and among their products, as well as their interaction with environmental factors are likely attributed to schizophrenia risk and useful for clinical prediction [54]. However, as a polygenic complex disorder, the genetic etiology of schizophrenia is difficult to identify due to intrinsic attributes, such as genetic heterogeneity, threshold inheritance, variable penetrance, epigenetics, epistasis, gene-environment interactions. The potential effects of these complexities need to be incorporated into analysis of schizophrenia and considered when interpreting results [55]. Balanced genetic approaches including traditional and newly-developed techniques are required to yield substantial insights into the molecular genetic mechanisms of schizophrenia (Table 3) [56]. Nevertheless, whatever technique is used, the primary requirement for genetic studies is a reasonably powered dataset with reliable diagnoses, which will decrease the probability of a false-negative result and increase the chance of obtaining positive results when linkage or association to a genetic marker presents.

<table>
<thead>
<tr>
<th>Method</th>
<th>Basis</th>
<th>Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical surveillance</td>
<td>The clinical process of physicians taking family histories. No DNA analysis.</td>
<td>Over past 100 years, no pedigrees with unequivocal Mendelian inheritance reported</td>
</tr>
<tr>
<td>Genetics epidemiology</td>
<td>Diagnosis of SCZ in various types of relatives (family, twin, adoption). No DNA analysis.</td>
<td>Family history an important SCZ risk factor</td>
</tr>
<tr>
<td>Segregation analysis</td>
<td>What inheritance models are statistically consistent with observed diagnoses of SCZ in pedigrees? No DNA analysis.</td>
<td>Uninformative, many different models consistent with data</td>
</tr>
<tr>
<td>Cytogenetics</td>
<td>Search for genomic changes in SCZ cases using karyotyping (3 Mb or larger)</td>
<td>Deletion of 22q11.2 is a rare, potent, and nonspecific risk factor for SCZ</td>
</tr>
<tr>
<td>Genome-wide linkage</td>
<td>Attempt at an unbiased genome search for regions whose inheritance in pedigrees is correlated with SCZ</td>
<td>No findings meeting modern criteria for significance and replication</td>
</tr>
<tr>
<td>Candidate gene association (CGAS)</td>
<td>Evaluation of frequencies of genetic variation in SCZ cases vs controls for genes selected using prior hypotheses about SCZ etiology</td>
<td>Major research focus but no findings meeting modern criteria for significance and replication, most studies with serious methodological issues</td>
</tr>
</tbody>
</table>
The two major research paradigms for population-based genetic association studies are the candidate gene association study (CGAS) [57] and the genome-wide association study (GWAS) [58], both of which are on the basis of single nucleotide polymorphism (SNP) genotyping and have been effectively applied to define genotype associations with certain disorders [59]. With the advantage of easy recruitment of independent cases and controls, during the last decades the case-control association study format has been adopted by both CGAS and GWAS as a preferential research design. A case-control association study is an investigation comparing specific genetic markers (e.g. SNPs) between a group of people with a disease or condition and another group of healthy people. By investigating specific genetic markers in linkage disequilibrium (LD), susceptibility loci for certain disorders can be revealed. However, the replication of susceptibility loci has been a major problem for interpreting and applying the results obtained from these studies. Even though some inherent features of the association studies are invincible, practical considerations for optimizing study design might improve the credibility and application of association studies.

Based on the trend of schizophrenia coming up in families, it has long been noted that schizophrenia has a genetic basis. The schizophrenia heritability is estimated to be in the range of 64-81% [60]. Historically, multiple genetic methods have remarkably contributed to the gradually in-depth understanding about schizophrenia genetics. Genetic studies have evolved from the initial cytogenetic method evaluating large chromosomal change to high throughput genome-wide genetic approaches [56]. CGAS and GWAS have been the two primary research methods for schizophrenia genetics. Thousands of CGASs have been recorded in the SZGene database since 1965 with more than one thousand genes as potential candidate genes for schizophrenia [61]. Since the last decade, as technology developing, GWAS has become applicable to scan huge amount of SNPs quickly and effectively by DNA probes embedded arrays and numerous GWASs for schizophrenia have been published.

<table>
<thead>
<tr>
<th>Genome-wide association (GWAS)</th>
<th>Attempt at unbiased genome search for loci with differing frequencies in SCZ cases vs controls</th>
<th>Multiple regions meeting modern criteria for significance and replication</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNVs</td>
<td>Search for genomic changes in SCZ cases vs controls (resolution variable but can be 100 kb or smaller)</td>
<td>Evidence that SCZ is highly polygenic No common variants of strong effect</td>
</tr>
<tr>
<td>Resequencing</td>
<td>Use high-throughput methods to resequence regions, exomes, or genomes to identify SNPs, indels, and CNVs with differing frequencies (individually or in aggregate) in SCZ cases vs controls</td>
<td>Initial studies in progress</td>
</tr>
</tbody>
</table>

1.4.1 Hypothesis-based CGAS for schizophrenia

CGAS is a hypothesis-based approach [57, 62-64], which concerns specific genes on genomic regions of interest a priori implicated by pathophysiology or susceptible gene locations with linkage peaks. With CGAS, selected alleles of presuming functional SNPs within candidate genes relevant to the hypotheses have been purposefully evaluated in targeted populations [65-68]. Hypothesis-based CGAS is an effective epidemiologic tool for abstracting relations between genetic loci and disease, especially under the conditions of low allele frequencies, small effect sizes and limited or unique study populations [69]. In spite of the many advantages, the positive results obtained from CGAS occasionally lack reliability due to either poor study design or suboptimal statistical analysis. Aimed at optimizing CGAS application, researchers may consider several issues including well founded hypothesis-based selection of candidate pathways and genes, logical SNP selection along the whole gene region, proper quality control of sample selection and statistical analysis procedures [70-75].

For schizophrenia, there are several classical neurochemical pathophysiology hypotheses: 1) The dopamine hypothesis [76-79]-the oldest and most established schizophrenia hypothesis supported by evident alterations in dopamine signaling pathways of patients with schizophrenia, by effective pharmacological interventions, and by evidence obtained from postmortem- and imaging studies; 2) The glutamate hypothesis [80-82]-a, more proximal hypothesis to the root of schizophrenia, based on clinical, neuropathological and genetic findings indicating a hypofunction of glutamatergic NMDA receptor signaling; 3) The GABA (γ-amino-butyric acid) hypothesis [83-85]-a hypothesis proposing that reduced neuronal GABA levels and decreased neurotransmission lead to cognitive impairment in patients with schizophrenia; 4) The serotonin hypothesis [86-88]-a hypothesis proposing that enhanced serotonergic neurotransmission in the cerebral cortex of patients with schizophrenia, especially in the anterior cingulate cortex and dorsolateral frontal lobe is a primary cause of schizophrenia. In addition to changes in neurochemical signalling, hypotheses including hypotheses on synaptic editing and neurodevelopmental defects provide a more comprehensive vision on the pathology of schizophrenia. The hypothesis on synaptic editing/pruning suggests a convergence of factors leading to the dysfunction of synaptic connectivity in patients with schizophrenia [89-93].

The neurodevelopmental hypothesis suggests that schizophrenia originates from neurodevelopmental defects induced by both genetic and environmental factors [94]. Diverse and circumstantial evidence has accumulated to support this theory and have made it a dominant focus point in schizophrenia research recently [95-98]. Other than genes involved in these hypotheses addressed in CGAS design, genes relating to physiological neural functions are also potential influential factors for schizophrenia.

In addition to the choice of hypothesis-driven candidate genes, there are other factors worth considering in hypothesis-based CGAS, including maximizing strategies for gaining information by effective selection of SNPs, quality control of the study sample-set and genomic data, and appropriate analysis approaches to dig out the disease-related nature of the selected SNPs. Carefully management of these issues may contribute to increase statistical power and strengthen the validity of the conclusions obtained from CGAS.

As the human genome contains a vast amount of SNPs, the selection of SNPs along the candidate gene region requires logical and delicate consideration [99-102]. The most favored genetic markers for CGAS are functional SNPs (non-synonymous and splice junction SNPs) that cause amino acid changes within proteins and thus potentially alter regular functions of proteins. Besides
functional SNPs, synonymous and non-coding SNPs may influence protein function indirectly by affecting transcription rates or bioactivity and stability. However, because adjacent SNPs are likely in high linkage disequilibrium (LD) with each other, a SNP associated with a certain disease may represent a tag for an unmeasured adjacent functional SNP in the same LD block. Therefore, apparently non-functional SNPs could be as informative as functional SNPs. LD analysis can also be performed to extract inferences about any non-measured SNP in the LD block for more valid disease comprehension. For the human genome, all common SNPs can be covered by genotyping relatively few tag SNPs. The International HapMap project (http://www.hapmap.org/) measured the LD between SNPs in numerous common racial and ethnic populations[103]. While selecting tag SNPs, minor allele frequency (MAF) is a critical consideration. Using a minimum MAF, typically 0.05 can confer a greater statistical power for detecting associations. However, It is unadvisable to omit rarer SNPs for the following reasons[104, 105]: 1) omitting rarer SNPs may lead to a counterproductive outcome due to potential larger effect sizes of these SNPs; 2) MAFs are sometimes not properly analyzed statistically; 3) MAFs of the actual study population are different from the reference population.

The selection of tag SNPs is usually based on LD data from the international haplotype database. However, the LD data defined by this database is often limited or fragmented, which results in multiple fragmented haplotype blocks consuming disproportionate tag SNPs and requires many tag SNPs to cover the whole gene region of the interested candidate gene. Therefore, it may be prudent to consider the average SNP distance of a specific ethnic population and drop redundant tag SNPs in favor of gathering more validated functional SNPs. High genotyping call rates are also important for SNP selection and essential for successful haplotype analysis [106, 107].

Quality control of the study sample-set and genomic data is critical for data validity and statistical inference. Poor DNA quality often results in undesirable failed genotyping reactions for SNPs of certain DNA sample-sets, increased genotype uncertainty scores, or heterozygote genotype excess of a contaminated sample-set. Multiple methods can be applied to improve DNA quality [108-111]: 1) Identity-by-state (IBS) method for relatedness between individuals; 2) Deleting cases or using statistical tests for correcting dependence to avoid violation of the independence assumption; 3) Exploratory data analysis tools for detecting differences in genetic background and mixture. As mentioned above, SNPs with low MAFs could not be excluded at liberty, which might lead to some non-informative SNPs with little or no variation in target sample-set. In addition, population stratification and genotyping errors can lead to significant deviations of SNPs from Hardy-Weinberg equilibrium in control and/or case groups, which might cause spurious association [112]. Therefore, drawing inferences from SNPs with a critical departure from Hardy-Weinberg equilibrium requires caution. SNPs which are difficult to genotype and have no sufficient overall genotype quality should be excluded from the association study. By imputation methods or appropriate statistical tests, genotype uncertainty and missing may be solved to quantify information of the selected SNPs and obtain logical inferences [113].

Primary interest of CGAS is to assess the associations between individual SNPs and the specific condition. Multiple comparisons can lead to type I error inflation and is principally managed by means controlling family-wise error rate (e.g., Bonferroni correction) [114]. Permutation tests can assess the overall significance of CGAS and contain SNP LD structure intact, which consequently may have less conservative results than Bonferroni correction, but achieve more power. Some SNPs have been reported associated with a certain condition previously, and thus may be selected in a new
CGAS for further validation, which allows drawing inferences separately from SNPs included for new discovery. However, for these validation SNPs, control of the family-wise error rate should be also applied, which would make association confidence of validation SNPs greater. It should be noted that a SNP without reaching significance after controlling family-wise error rate simply means that it only failed to validate. Validation SNPs are in small number and often from various genomic regions, requiring moderate overall significance and independent test statistics [115, 116]. Besides tag-SNP strategy, other multilocus strategies also exist, however, for practical reasons, tag-SNP strategy is the most favored in CGAS.

1.4.2 GWAS for schizophrenia

CGAS can only consider a quite limited proportion of human genome on the account of time-consuming genotyping [117]. Developments in microarray- and DNA sequencing-based technology allow a quick and effective genotyping of large numbers of SNPs for GWAS at a quite larger scale than CGAS [118, 119]. GWAS proposes an unbiased assessment of common genetic variants across the entire human genome in a large number of affected cases to healthy controls to estimate whether an association with a certain disease or condition exists. Quality control of the study sample-set and genomic data and selection of appropriate analysis approaches require more deliberate considerations for GWAS. Although it seems reasonable that a high heritability of a certain disease or condition provides a relatively easy gene-discovering capacity, a high heritability only means a large genetic contribution on the disease, but not necessarily large effect sizes for specific genes, which is an important factor for discovering risk variants. It is apparently valid that increasing the sample size may raise statistical power to detect associations, which has been approved for schizophrenia GWAS [120-123]. Scanning the entire human genome and including large sample-sets is expensive and hindered by inflated false positive results caused by multiple statistical testing. To control the rate of false positives, the threshold for statistical significance has been defined at a rigorous level. The p-value criteria are often reduced from 0.05 to 5.0 ×10^-8 in GWAS[124], which demands large amounts of cases and controls to regain the chance of observing the genome regions genuinely involved in disease [125]. Based on linkage disequilibrium (LD), SNPs genotyped by commercial microarrays and sequencing for GWAS are selected by their ability to represent the information of the neighboring alleles. Genetic correlation makes it viable to use a million elaborately selected SNPs to undertake about 90% known genetic variants other than investigate all the discovered polymorphic positions. Consequently, the significant SNPs in a GWAS have a great chance for not being the biologically functional variants, but only the phenotypic representatives. Understanding this is crucial in interpreting the clinical significance of the positive results.

More than 30 GWAS’s of schizophrenia have been performed and published so far with the hope to identify novel, consistent and influential risk genetic factors. SNPs and genes reported by these studies are different from each other, and only few studies have detected a SNP with genome-wide significance (p<5.0×10^-8). However, whereas most early GWASs seemed large at their publishing times, they are underpowered to discover genome-wide associations with low effect size of risk genes [126]. Recently, multi-centered GWAS and meta-analysis of several GWAS samples have provided more valid and reliable results for further insight into schizophrenia pathophysiology [127]. However, the most consistent results come from GWAS’s with overlapping samples. The technology used in GWAS can also be used to detect copy number variations (CNVs), submicroscopic duplications or
deletions of DNA segments, ranging from kilobases (kb) to megabases (Mb). About one-quarter of the human genome harbors CNVs. Despite the fact that the majority of CNVs do not contribute to altered phenotypes, genome wide SNP arrays have suggested multiple aberrations in CNVs associated with schizophrenia.

1.4.3 Combined two-stage association study of GWAS and CGAS

As the two major genetic research paradigms, CGAS and GWAS both have their own advantages and disadvantages [53, 128]. CGAS has a noted advantage in comparison with GWAS that it is a targeted screening strategy, which enhances its power on studying lower-frequency SNPs or when study population is small and unique [57]. Additionally, CGAS is a cost-efficient and time-saving application for investigating genetic characteristics of human diseases. However, for most biomedical diseases, including schizophrenia, CGAS results were inconsistent or confusing, caused by potential poor choices of candidate genes or inadequate assessment [129]. Thus GWAS is hitherto the most comprehensive strategy to discover a genetic background of complex genetic diseases such as schizophrenia [130]. Currently, some inherent characteristics of GWAS make the original objective of identifying novel susceptibility genes difficult [131]: 1) no convergence of risk SNPs and genes; 2) no apparent relationships with functions involved certain diseases in a great portion of the genes reported by GWASs and 3) small effect sizes for reported genes. Moreover, for both CGAS and GWAS, rare variants with a low frequency in general population usually would be excluded, which may largely affect the studied diseases. Overcoming these disadvantages requires an objective study design, novel analytic technology development, expansion of the study scope, and downstream pathway studies applying the obtained data.

Nowadays, GWAS and CGAS are often combined for multistage genetic study designs [132-134]. While evaluating a great amount of SNPs scattered along the entire human genome, large sample sizes of GWAS may guarantee sufficient statistical power to discover modest associations, which motivates the development and application of further validation studies. In general, firstly, a sample-set is analyzed at a discovery stage of GWAS to obtain hundreds or thousands of associated SNPs. Then, most significant associated SNPs or SNPs involved in a pathway previously suspected in targeted disease are genotyped with a less-expensive CGAS in an independent sample-set. At the second CGAS stage, subsequent verification studies are performed to restrain the false positives, however, if the significant SNPs reported in the initial GWAS display a nominal significance (p<0.05) after controlling family-wise error rate, they may be regarded as significant[126]. In a two-stage study, the follow-up study is mostly considered as a replication study for confirming the initial findings. However, due to the modest associations and multiple comparisons, to obtain results with sufficient power is reasonably to combine data from the first GWAS stage in a joint analysis viewed as a single-stage GWAS with a less expense. Screening, replication and joint analysis of the most promising SNPs can facilitate and strengthen additional pathophysiological function studies, overcoming the three drawbacks of GWAS.

2 Outline of the thesis

The general objective of population-based genetic association studies for schizophrenia is to identify risk loci with high statistical confidence, provide insight into further exploration of biological mechanisms by which the discovered potential genetic factors influence schizophrenia onset and
psychosis progression, and ultimately facilitate therapy development and prevention options. In this thesis, we aimed to analyze the associations of several new interesting candidate genes with schizophrenia and explore their roles in schizophrenia. The candidate genes were selected either based on previous implications of their involvement in specific functions in the development or maintenance of neuronal circuitry or by a combination of our previous GWAS results and existing schizophrenia hypotheses. Chapter 1 provides an elaborate review on the current understanding of schizophrenia and its genetics. Next, we investigated the potential involvement of the human STON2 gene in schizophrenia in a Chinese Han case-control sample-set (Chapter 2) and found positive associations between schizophrenia and STON2 gene polymorphisms including two functional SNPs rs3813535 (Pro307Ser) and rs2241621 (Ala851Ser). We selected the human STON2 gene as a candidate gene since it plays a role in the synaptic endocytic machinery in developing and mature neurons. In Chapter 3, we reported associations between schizophrenia and several well-characterized functional polymorphisms of the human NAT2 gene, which encodes a specific type of N-acetyltransferase. This N-acetyltransferase has been shown to play an essential role in the metabolism and detoxification of many compounds, including drugs and environmental toxins that are linked to brain pathology. The functional polymorphisms of NAT2 can cause aberrant enzyme activity resulting in too slow or too fast acetylation both associated with a variety of complex human disorders. Our results showed a significant association between NAT2 polymorphisms and risk for schizophrenia in a Northern Chinese sample-set. Based on our earlier schizophrenia GWAS applied in a Chinese Han population, we designed, in Chapter 4, a replication association study in an independent sample-set to investigate the association between schizophrenia and human MSI2 gene, a member of the Musashi gene family encoding an evolutionarily conserved neural RNA-binding protein playing a prominent role in neural stem cell differentiation. We verified the associations of several intronic SNPs of MSI2 with schizophrenia, and identified linkage SNPs for the three schizophrenia-associated SNPs in five East Asian populations (CHB, JPT, CHS, CDX, KHV) from the 1000 Genome Project. In Chapter 5, we examined the association of the human SOX11 gene with susceptibility for schizophrenia in an independent case-control sample-set, and explored the role of SOX11 in neurodevelopment in vitro and in vivo (Chapter 6). SOX11 belongs to the C subfamily of the developmentally essential SOX (Sry-related HMG box) transcription factor gene family. It showed a minor significance in our previous GWAS. SOX11 has been indicated as a potential regulator of various aspects of neuronal development in vitro and in vivo experiments. Our results showed that several schizophrenia-associated SNPs located in the SOX11 flanking gene region are functional, and SOX11 is required in multiple phases of neurodevelopment, such as neural proliferation, neuronal differentiation, neuronal maturation, neurite growth and neuronal migration. In a separate small additional study we investigated the downstream targets of SOX2 and its underlying molecular mechanisms, and found a novel Sox2/survivin pathway regulating NSC survival and homeostasis, implicating a new mechanism in the development, maturation and plasticity of neuronal circuitry (Chapter 7). Finally, we summarized and discussed the major findings of this thesis in Chapter 8 and provided future outlook.
References

33. Green, M.F., et al., Neurocognitive deficits and functional outcome in schizophrenia: are we measuring the "right stuff"? Schizophr Bull, 2000. 26:119-36
General Introduction


41. Bleuler, E., [Dementia praecox or the group of schizophrenias]. Vertex, 2010. 21:394-400


General Introduction

32. Wellcome Trust Case Control, C., Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature, 2007. 447:661-78


