Abstract

OBJECTIVES:
Schizophrenia is a severe psychiatric disease affecting about 1% of the general population. The relative contribution of genetic factors has been estimated to be up to 80%. The mode of inheritance is complex, non-Mendelian, and in most cases involving the combined action of large numbers of genes. This consensus paper of the WFSBP task force on genetics addresses the current knowledge on the molecular genetics of schizophrenia.

METHODS:
This review summarizes recent efforts to identify genetic variants associated with schizophrenia detected e.g. through genome wide association studies, studies on copy number variants or next generation sequencing. The contribution of these findings to the understanding of pathobiology is discussed.

RESULTS:
The last years accumulated a large new body of evidence on genetics of schizophrenia. Many new robustly associated genetic loci have been detected. Furthermore, there is consensus that at least a dozen microdeletions and microduplications contribute to the disease. Genetic overlap between schizophrenia, other psychiatric disorders, and neurodevelopmental syndromes raised new questions regarding the current classification of psychiatric and neurodevelopmental diseases. Also next generation sequencing shows an overlap between schizophrenia and other neurodevelopmental disorders even at the level of a specific highly penetrant recurrent mutation in a single gene.

CONCLUSIONS:
Future studies will address especially the functional characterization of genetic variants. This will hopefully open the doors to our understanding of the pathophysiology of schizophrenia and other related diseases. Complementary, integrated systems biology approaches to genomics, transcriptomics, proteomics and metabolomics may also play crucial roles in enabling a precision medicine approach to the treatment of individual patients.
Heritability of Schizophrenia
Evidence for a strong genetic component to the etiology of schizophrenia was first demonstrated by classical genetic epidemiology in the form of family, twin, and adoption studies (Cardno & Gottesman 2000; Sullivan et al. 2003; Wray & Gottesman 2012) and has more recently been confirmed by molecular genetics (Gusev et al. 2014). While the lifetime risk in the population for developing schizophrenia varies within countries, the overall rate is estimated at around 0.5-1%. Risk for an individual increases non-linearly with the degree of genetic relatedness to a person suffering from the disorder; for third-degree relatives it is approximately 2% rising to approximately 9% for first-degree relatives, or, in the case of the children of two affected parents, around 27% (Gottesman et al. 2010). For monozygotic (MZ) twins, the concordance rate is approximately 50% (Lichtenstein et al. 2006; 2009). For adopted children with a biological parent with schizophrenia the risk of developing schizophrenia is 6 to 10 times higher than in the general population (Shih et al. 2004).

Altogether, the heritability of schizophrenia is estimated to be between 64–81% (Sullivan et al. 2003; Lichtenstein et al. 2009). Large amounts of genetic variations contribute to the disorder making the mode of inheritance complex. Only a small proportion of these DNA variants have been identified so far and it is now clear that genetic susceptibility is the result of many common variants of low penetrance, some rare variants of moderately high penetrance (Kendler 2014; Schizophrenia working group of the Psychiatric Genomics Consortium 2011; 2014) as well as presumably uncommon and rare variants of small-moderately high effect that cannot be detected individually in sample sizes studied to date. Progress in understanding the genetics of complex disease is closely tied to technological developments. The potential of those technologies which include arrays capable of detecting millions of common variants as well as large structural variations, and newer sequencing technologies, has led to an explosion of findings in human genetics as a field. In parallel with the technological capabilities, approaches to complex disease genetics has extended from linkage studies to candidate association studies, genome wide association studies and currently, to exome or whole genome sequencing.

Linkage Studies
Linkage studies are based on the fact that genetic traits located closely together on a chromosome are more likely to segregate together, that is, be co-inherited in families, than are genetic variants located further apart. Linkage is most powerful under models of genetic transmission whereby genes of major effect cause disease in all affected family members. For common diseases this may not be the case and multiply affected families may mimic Mendelian architecture. Early linkage studies suggested involvement of a major risk allele on chromosome 5 [5q11.2 to 5q13.3] (Bassett et al. 1988; Sherrington et al. 1988), but that finding has not been extensively replicated (Kennedy et al. 1988; St Clair et al. 1989). One of the largest meta-analyses was provided by Ng et al. (2009) including 32 independent genome wide linkage scans and a total of 3,255 pedigrees and 7,413 genotyped cases affected with schizophrenia or related disorders. The results pointed to evidence for linkage on 5q (142–168 Mb) and 2q (103–134 Mb). A secondary analysis restricted to the studies based on families of European-ancestry provided suggestive evidence for linkage on chromosome 8p (16–33 Mb) (Ng et al. 2009). So far these linkage studies failed to conclusively identify any susceptibility gene although some latest meta-analyses showed partial overlap for 5p14.1 and 10q26.12 (Vieland et al. 2014).

Candidate Gene Studies
As methods of DNA amplification were developed, association studies became more feasible (Mullis et al. 1986; Mullis et al. 1994). These were mainly based on case control analysis of SNPs (single nucleotide polymorphisms) within candidate genes derived from neurobiological hypotheses. Gatt et al. (2015) presented a review of meta-analyses based on candidate genes. From a total of 97 variants reported in schizophrenia, the strongest evidence (p < .001) was reported for genes involved in the modulation of dopamine (e.g.,
COMT, DRD2, DRD3, DRD4), glutamate (e.g., DAOA, GABRB2, NRG1), neuronal development and function (e.g., AH11, MTHFR, RELN, TRKA), serotonin neurotransmission (HTR2A, SLC6A4, TPH1) or the immune system (IL1B). The authors focused not only on schizophrenia but on specific mental disorders that have core disruptions to emotional and cognitive function and contribute most to burden of illnesses including major depressive disorder (MDD), anxiety disorders (including panic disorder and obsessive compulsive disorder), schizophrenia (SZ), bipolar disorder (BD) and attention deficit hyperactivity disorder (ADHD). A total of 1,519 meta-analyses were conducted across 157 studies reporting multiple genes implicated in one or more of the five disorders. A total of 134 genes (206 variants) were identified as significantly associated risk variants. 13 genetic variants were shared in common between two or more disorders (APOE, ACE, BDNF, COMT, DAOA, DAT1, DRD4, SLC6A4, HTR1A, MTHR, MTHR, and TPH1) demonstrating evidence for pleiotrophy (Gatt et al. 2015). Nevertheless, the significance threshold was not as strict as applied for genome wide studies indicating that there could be spurious associations among these genes. Another drawback is that there tends to be a publication bias towards original articles reporting on positive findings, and only these could be used for meta-analyses.

**Genome Wide Association Studies**

Due to technical developments in array genotyping, research is no longer limited to testing a small number of candidate variants in sample sizes that are underpowered to detect the small effect sizes of common variants we now know to typically associate with complex diseases. Array genotyping instead makes it possible to conduct genome wide association (GWA) studies which extract information for more than 10 million SNPs per sample in a single experiment. That implies that stringent correction for multiple testing has to be performed and as discussed later, power to detect associations at the genome wide significance threshold is incomplete even in the largest studies to date and therefore many of the hits not reaching genome wide significance may turn out to be true associations in larger samples. From a technical point of view, association with a locus not necessarily implicates a particular gene but rather a region, each with a small effect size.

The first GWAs studies on schizophrenia performed in 2007/2008 were clearly underpowered including 178 cases, 144 controls (Lencz et al. 2007) and 738 cases, 733 controls, respectively (Sullivan et al. 2008). The problem of insufficient power to detect genome wide significant markers directly was addressed by O'Donovan (2008) by using top markers (p<10^{-5}) from an initial GWAS on 479 cases and 2,937 controls, as candidates to be followed up in an independent replication cohort of up to 6,829 cases and 9,897 controls. Three out of 12 loci were shown to have a strong independent support, and the overall pattern of replication was unlikely to occur by chance. Most interestingly, evidence for association at the top SNP which maps to a locus containing the gene ZNF804A, strengthened when the affected phenotype included bipolar disorder, which challenges the traditional diagnostic boundaries (O'Donovan et al. 2008). The study of Stefansson et al. (2009) used basically the same approach. GWAS performed with 2,663 cases and 13,498 controls from eight European locations within the SGENE+ consortium revealed no genome wide significant signal. The findings for the top 1,500 SNPs were combined with results for these SNPs (or proxies) from two independent cohorts (2,602 cases/2,885 controls from the International Schizophrenia Consortium - ISC, and 2,687 cases/2,656 controls from the European-American portion of the Molecular Genetics of Schizophrenia Consortium - MGS). The most significant association signals were followed up in 5,013 cases and 15,559 controls from four additional samples sets from Europe, leading to the identification of three novel candidate schizophrenia loci: NRGRN (coding for neurogranin), TCF4 (coding for transcription factor 4), and the MHC (major histocompatibility complex) region (Stefansson et al. 2009). At the same time ISC and MGS reported genome wide associations to the MHC region (International Schizophrenia Consortium 2009; Shi et al. 2009).
As power and therefore sample size is the most limiting factor in genome wide association studies, a new consortium, the Psychiatric Genome Wide Association Study Consortium (PGC) started a highly successful initiative to collect existing GWAS data from as many schizophrenia cases and controls as possible. The first publication in 2011 contained a total sample of more than 51,000 individuals of European ancestry. Genome wide significant association was identified for seven loci, five representing novel loci (1p21, 2q32.3, 8p23.2, 8q21.3 and 10q24.32-q24.33) and two with previous implication (6p21.32-p22.1 and 18q21.2). The strongest new finding was for rs1625579, located within an intron of a putative primary transcript of microRNA 137 (MIR137) which is involved in regulation of neuronal development. Supporting a role of MIR137 dysregulation as a previously unknown etiologic mechanism in schizophrenia was the finding that additional markers in four predicted targets of MIR137 also showed genome wide significant p-values. Combining the schizophrenia with a bipolar disorder sample (16,374 cases and 14,044 controls) led to the identification of three loci that confer susceptibility to both bipolar disorder and schizophrenia, i.e. CACNA1C, ANK3 and the ITIH3-ITIH4 region (The Schizophrenia Psychiatric Genome Wide Association Study (GWAS) Consortium 2011). Given the large sample, the yield of 5 new loci might be considered disappointing, but a follow up study by Hamshere and colleagues (2013) demonstrated in an independent sample that the vast majority of the 81 associations in the PGC study that met an intermediate threshold of significance (P<10^{-5}) represented true associations, providing a robust rationale for expanding GWAS to larger samples. Moreover, when they combined their small (in current GWAS terms) sample with those of the PGC, they identified three additional genome wide significant loci, stressing the cumulative value of building up the large datasets.

In 2013, Ripke et al. re-analyzed the PGC schizophrenia data using 1000 Genomes imputation and conducted a study comprising an initial analysis based on 5,001 cases and 6,243 controls followed by meta-analysis with previous schizophrenia GWAS (8,832 cases and 12,067 controls) and finally by replication of SNPs in up to 168 genomic regions in independent samples (7,413 cases, 19,762 controls and 581 parent-offspring trios). Of the 22 loci associated at genome wide significance level, previous reports had been published for 7 loci with schizophrenia alone (MHC, WBP1L/C10orf26, DPYD-MIR137, SDCCAG8 and MMP16, CACNA1C, and ITIH3-ITIH4), one with a combined phenotype consisting of schizophrenia and bipolar disorder (CACNB2), and one with bipolar disorder (NCAN). The remaining 13 newly identified loci consisted of regions including MAD1L1, TSNARE1, SNX19, QPCT, SLC06A1, ZEB2, FONG, C2orf82, AKT3, C12orf65 as well as loci near GRIA1, TCF4, and ZSWIM6. Examination of candidate regions suggested the involvement of neuronal calcium signaling, the MHC region, which showed the highest significance, MIR137 (dys)regulation via target sites and epigenetic regulation, and development through long intergenic non-coding RNAs (Ripke et al. 2013).

The latest and largest GWAS on schizophrenia was published in 2014 by the schizophrenia PGC team featuring a multi-stage GWA study of 36,989 cases and 113,075 controls. 128 associations in 108 independent loci were identified with strongest associated locus being an extended region of chromosome 6 containing a large number of genes including the MHC region (p=3.48x10^{-31}) (The Schizophrenia Psychiatric Genome Wide Association Study (GWAS) Consortium 2014; bioinformatic data for the 108 loci is found in the Supplementary Information). Novel associations were observed for 83 loci not necessarily implicating a gene but rather a region. Associations were enriched among genes showing epigenetic markers indicative of expression in brain. Interestingly, a number of loci containing former known candidate genes like DRD2 and several genes involved in glutamatergic neurotransmission (GRM3, GRIN2A, SRR, GRIA1) reached genome wide significance, supporting gene products of known and potential therapeutic relevance to schizophrenia, and broadly supporting leading pathophysiological hypotheses of the disorder. The PGC study also reported evidence that associations were enriched in genes expressed in tissues relevant to immunity, although a more sophisticated reanalysis of the data failed to support this, confirming only the enrichment for brain expressed genes (Finucane et al. 2015).
The most consistent result throughout GWAS of schizophrenia is the association to the extended MHC region. This was found in several independent samples of moderate (International Schizophrenia Consortium 2009, Stefansson et al. 2009, Shi et al. 2009, Irish Schizophrenia Genomics Consortium and the Wellcome Trust Case Control Consortium 2 2012) to large sizes (Ripke et al. 2013, The Schizophrenia Psychiatric Genome Wide Association Study (GWAS) Consortium 2014). Interestingly, allelic variation at the Complement component C4A locus, which is located in the MHC region, explains part of the heritability of schizophrenia. There is evidence that C4A is involved in synaptic plasticity, providing a potential functional explanation for the involvement of the immune system in schizophrenia pathogenesis (Sekar et al. 2016).

Polygenic Risk Scores

Genome wide association studies have provided strong evidence that psychiatric disorders are highly polygenic, with a genetic architecture consisting of many common genetic variants. As discussed later, rare risk variants are also involved. To achieve genome wide significance, p-values have to pass a stringent threshold. It is estimated that in common variant GWAS, around 1 million independent association tests are conducted and the threshold for significance after multiple testing is therefore 0.05/1 million or 5 x 10^-8 (NCI-NHGRI Working Group on Replication in Association Studies 2007). Attaining this for alleles of small effect requires very large sample sizes of many ten-thousands of patients and controls. Interestingly, although the total number of loci surpassing the genome-wide significance threshold is often small (or even zero in small studies), GWAS of schizophrenia typically yield many more associations with small p-values than expected by chance. This finding is consistent with a polygenic genetic architecture, and provided the impetus for new statistical methods (Wray et al. 2014). Two frequently applied methods are “polygenic risk scoring”, and “estimating variance explained by all SNPs” (for details see Wray et al. 2014).

The first genomic profile risk scoring for schizophrenia was reported by the International Schizophrenia Consortium (2009). This study provides evidence for a polygenic component to risk of schizophrenia involving at least a thousand common alleles of very small effect (Nagelkerke R^2=0.032; a measure of variance in case-control status explained). Furthermore, the risk profile score was also associated with risk of bipolar disorder, but not to multiple non-psychiatric diseases.

Within the much larger PGC-Schizophrenia sample the R^2 increased to 0.184. Under the assumption of a liability-threshold model, a lifetime risk of 1%, independent SNP effects, and adjusting for case-control ascertainment, the polygenic risk score explained about 7% of variation on the liability scale to schizophrenia across the samples, about half of which (3.4%) is explained by genome wide significant loci in this study (Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014). Many other studies using polygenic risk scores have been conducted including cross-disorder overlaps, schizophrenia symptoms or phenotypes (for review see Wray et al. 2014). Given that polygenic scores represent biomarkers that tap into liability to the disorder even in unaffected people, we anticipate they will become more and more important in research, although to date, they do not capture enough risk to be clinically useful.

Identification of Rare Risk Variants before Genome Wide Association Studies

Up till the mid 2000s, the only genetic variant definitively known to contribute to schizophrenia was a microdeletion at chromosome 22q11.2. The deletion region contains over 40 genes including a number of plausible candidates like COMT (catechol-O-methyltransferase), ProDH, TBX1, and GNB1L, which are known to be involved in brain function and neurodevelopment (Raux et al. 2007; Schneider et al. 2014; Squarcione et al. 2013). However, despite of several studies that have tried to do so, it has not proven possible to
identify which gene(s) within 22q11.2 explain the link between deletion of this region and schizophrenia. Also worth noting from the pre-GWAS era is Disrupted-in-Schizophrenia-1 (DISC1), another key historical candidate gene that has been widely researched. DISC1 was identified as being disrupted by one of two chromosomal breakpoints (Millar et al. 2000) involved in a balanced translocation between chromosome 1 and chromosome 11 (1;11)(q42.1;q14.3) that co-segregates with multiple forms of psychiatric disorder in a single large Scottish pedigree (St Clair et al. 1990; Muir et al. 2008). Although the eponymous, but given the range of phenotypes exhibited in family members somewhat inaccurately named DISC1 gene has plausibility given it has been shown to disrupt numerous important functions of the brain (Randall et al. 2014; Porteous et al. 2014), additional evidence beyond co-segregation is required to decisively implicate this gene in major mental illness.

Copy Number Variants identified through Genome Wide Association Studies

It was originally possible to link the 22q11.2 deletion to schizophrenia because the same mutation also causes a recognizable genomic syndrome called velo-cardiofacial syndrome (VCFS) or DiGeorge syndrome, and it had been observed that those with VCFS had high rates of psychosis (Shprintzen et al. 1992; Pulver et al. 1994). In the mid 2000s, technological developments, particularly those that allow GWAS studies, started to fill the resolution gap between the traditional cytogenetic analysis (>2 Mb) and mutation analysis by DNA sequencing (<1 kb), making it possible to perform systematic scans of the genome for chromosomal microdeletions and duplications (collectively known as copy number variants or CNVs) without any prior knowledge of phenotypic co-morbidity between schizophrenia and a particular genomic syndrome.

A number of investigators (Kirov et al. 2008; Xu et al. 2008; Walsh et al. 2008) were prompted to screen proband-parent trios for newly occurring or de novo mutations, the idea being that the loss of schizophrenia risk alleles resulting from the reduced fecundity of people with the disorder must be counter-balanced by the occurrence of new mutations unless the disorder was vanishing in the population. Early support for the involvement of CNVs was obtained in all these studies, but none was powered to convincingly implicate any specific CNVs due to insufficient sample size. However, these were rapidly followed by two larger studies, one by the SGENE+ consortium and one by the International Schizophrenia Consortium (ISC). In a first step the SGENE+ consortium analyzed 9,878 individuals for transmission from parents to descendant (Stefansson et al. 2008). The study identified 66 novel CNVs which were tested for association in a subsequent case-control study. Three deletions (1q21.1, 15q11.2, and 15q13.3) showing nominal association in the first sample were followed-up in a second sample of 3,285 cases and 7,951 controls. The combined sample showed enrichment of all three microdeletions. Two of these, deletions at 15q13.3 (containing the nicotinergic alpha 7 gene) and 1q21.1, were independently identified in the ISC study which analyzed 3,391 schizophrenia patients and 3,181 controls (International Schizophrenia Consortium 2008). Moreover, in the ISC study, compared to controls, patients with schizophrenia had a 1.15-fold increase of rare CNVs (frequency less than 1% in this study) of more than 100kb in length. Together, these studies decisively confirmed a collective enrichment of rare CNVs in schizophrenia patients and identified specific CNVs associated with risk of the disorder. The deletion identified by SGENE+ but not the ISC, that at 15q11.2, was subsequently confirmed in the ISC and other data by Kirov et al. (2009a), and overall evidence for all 3 loci is beyond doubt in a recent meta-analysis (Rees et al. 2014).

Additional deletions significantly associated with schizophrenia at 3q29 (Mulle et al. 2010; Levinson et al. 2011), 16p11.2 (Guha et al. 2013), 17p12 (Kirov et al. 2009a), and 17q12 (Moreno-De-Luca et al. 2010) were replicated in the meta-analysis by Rees et al. (2014), with 3q29 showing highest ORs (OR=57.65). Additional CNV studies suggested a role for the neurexin 1 gene (NRXN1; 2p16.3). Deletions within this gene previously associated with autism have also been reported in two families with schizophrenia (Kirov et al. 2008; Walsh et al. 2008). In a study conducted by the SGENE+ consortium using 2,977 schizophrenia
patients and 33,746 controls. NRXN1 CNVs (>100 kb) disrupting exons were significantly enriched in cases (0.24% vs. 0.015% in controls) with an odds ratio of 8.97 (Rujescu et al. 2009). A meta-analysis in 8,789 cases and 42,054 controls further supports these results (Kirov et al. 2009b). Significant associations with deletions were also reported by Levinson et al. (2011) and in a recent meta-analysis by Rees et al. (2014). Several studies supported not only the role of deletions but also of duplications in the development of schizophrenia. Within the SGENE+ consortium a 15q11-q13 (Ingason et al. 2011a) and a 16p13.11 duplication (Ingason et al. 2011b) were identified. Both as well as a duplication at 16p11.2 (McCarthy et al. 2009; Levinson et al. 2011; Steinberg et al. 2014) were confirmed in the Rees et al. (2014) meta-analysis. Several other loci could not be confirmed by this meta-analysis and are subject of further investigation: 1q21.1 (Levinson et al. 2011), 7q11.23 associated with William-Beuren-Syndrome (Kirov et al. 2012; Mulle et al. 2014) and 7q36.3 containing VIPR2, a neuropeptide receptor gene (Levinson et al. 2011; Vacic et al. 2011).

CNVs enriched in psychosis are also involved in the pathology of other neurodevelopmental and neurological disorders, including intellectual disability, autism, and seizures (Grozdeva et al. 2012; Grayton et al. 2012). Within the group of 1q21.1 deletion carriers psychiatric and behavioral abnormalities can include epilepsy, autism, or attention deficit hyperactivity disorder (de Kovel et al. 2010; Haldeman-Englert & Jewett 2015; Sanders et al. 2015). Neurexin 1 deletions are also associated with autism spectrum disorders, intellectual disability, language delays (Schaaf et al. 2012) developmental delay, or intellectual disability (Jenkins et al. 2018). Very similar overlaps are seen for the 15q11.2 region including intellectual disability, developmental delay, neurological problems, autism, attention problems, and speech delay (Abdelmoity et al. 2012; Derks et al. 2013). Similarly the 15q13.3 CNVs are present in multiple neurodevelopmental syndromes which causes intellectual disability, epilepsy and variable facial and digital dysmorphisms. Interestingly, the discovered de novo deletions in this region have also been shown to give rise to a number of other phenotypes, including abnormal EEG, significant expressive language deficits, and a spectrum of neuropsychiatric impairments that include autism, ADHD, anxiety disorder, mood disorder and cognitive impairment (Williams et al. 2012; Helbig et al. 2009; Pagnamenta et al. 2009; Isles et al. 2016; Adams et al. 2012).

Other studies have continued to pursue de novo mutations in schizophrenia (e.g. Kirov et al. 2012; Malhotra & Sebat 2012). Both provide an estimated rate for large de novo CNVs in schizophrenia of about 5%, 2-3 times higher than that of controls. Based on the meta-analysis of Rees et al. (2014) the question arises concerning the implications of the CNV findings for genetic counselling. Out of 15 previously implicated CNV loci, 11 were strongly associated in this meta-analysis. The evidence for the remaining four loci is still equivocal and requiring further investigation. These findings indicate that ca. 2.5% of individuals with schizophrenia carry at least one known pathogenic CNV. The odds ratios for schizophrenia range between ca. 2 and more than 50. As noted above, most are also found in a range of other neurodevelopmental disorders, including epilepsy (15q11.2 and 15q13.3), autism and intellectual disability. As Rees et al. (2014) postulated, a number of the pathogenic CNVs are associated with particular physical disease phenotypes including congenital heart disease (1q21.1 and 22q11.2), microcephaly (1q21.1, 3q29 and 16p11.2) and obesity (16p11.2 distal). These findings suggest that screening for CNVs should be considered at least in the case of non-brain-related physical comorbidity. These results have the potential to directly guide everyday clinical patient care (Rees et al. 2014).

Kirov et al. (2012) noted multiple de novo CNVs spanned components of the post-synaptic density (PSD). This was largely explained by enrichment for members of the N-Methyl-D-Aspartate receptor (NMDAR) and neuronal activity-regulated cytoskeleton-associated protein (ARC) postsynaptic signaling complexes; both known to be important in synaptic plasticity and cognition. They also found recurrent CNVs affecting EHMT1, a histone methyl transferase known to directly regulate DLG family members and other components of the PSD. Large scale follow up case-control analyses (Pocklington et al. 2015) have confirmed
the enrichments for schizophrenia related CNVs in PSD, ARC and NMDAR complexes and have extended this to implicate inhibitory post-synaptic GABAergic complexes which has strong functional links in regulating synaptic plasticity with the major excitatory glutamatergic system.

As overlapping genes affected by rare variants and those localized within the associated GWAS loci also show convergence in terms of functional clustering it seems likely common and rare variant studies are complementary rather than antagonistic, and that mechanistic studies driven by rare genetic variation will be informative for schizophrenia (Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014).

**Next Generation Sequencing**

Major developments in sequencing technology have taken place in the last years enabling rapid and increasingly economical whole exome or whole genome sequencing. These sequencing methods can e.g. be applied for case control studies or for the detection of de novo mutations by sequencing both unaffected parents and the affected child. A study by Xu et al. (2011) sequenced the exomes of 53 trios with sporadic cases and their unaffected parents, as well as 22 unaffected controls trios. They identified 40 de novo mutations in 27 cases affecting 40 genes, including a potentially disruptive mutation in DGCR2, a gene located in the schizophrenia-predisposing 22q11.2 microdeletion region. A follow-up study by the same group sequenced 795 exomes from 231 parent-proband trios enriched for sporadic schizophrenia cases, as well as 34 unaffected trios. They observed an excess of de novo nonsynonymous single-nucleotide variants as well as a higher prevalence of gene-disruptive de novo mutations relative to controls. They found four genes (LAMA2, DPYD, TRRAP and VPS39) affected by recurrent de novo events, which is unlikely to have occurred by chance (Xu et al. 2012). Moreover, a de novo deletion of DPYD had first been found in autism (Marshall et al. 2008).

Interestingly, a de novo mutation in another member of the laminin gene family, LAMA1 was described in another schizophrenia sequencing study reported by Girard et al. (2011). Exome sequencing of genomic DNA carried out for 399 persons, including 105 probands with schizophrenia, 84 unaffected sibs, and their 210 unaffected parents was performed by Gulsuner et al. (2013). The results suggest that disruptions of fetal prefrontal cortical neurogenesis may be critical to the pathophysiology of schizophrenia (Gulsuner et al. 2013). McCarthy et al. (2014) carried out exome sequencing on 57 trios with sporadic or familial schizophrenia. In sporadic trios, they observed a ~3.5-fold increase in the proportion of nonsense de-novo mutations. Genes at these loci overlapped with genes implicated in autism (e.g., AUTS2, CHD8 and MECP2) and intellectual disability (e.g., HUWE1 and TRAPPC9), supporting a shared genetic etiology between these disorders. Functionally, CHD8, MECP2 and HUWE1 converge on epigenetic regulation of transcription suggesting that this may be an important risk mechanism.

The largest individual de novo trios and case-control sequencing studies of schizophrenia so far were those respectively of Fromer et al. (2014) and Purcell et al. (2014). Fromer et al. (2014) obtained further evidence for shared genetic aetiology between schizophrenia and both intellectual disability and ASD by testing for overlap of genes affected by de novo loss-of-function mutations in schizophrenia, ASD and intellectual disability. The study shows an overlap between schizophrenia, ASD and intellectual disability at the resolution not just of loci or even individual genes, but even of mutations with similar functional (loss-of-function) effects.

Although Fromer et al. (2014) were unable to confirm the findings from some of the smaller studies that at genome wide, cases are enriched for small de novo mutations, they did find that this type of mutation was overrepresented among the same complexes implicated by de novo CNVs, specifically the ARC and NMDAR complexes. Mutations were additionally enriched in messenger RNAs which are targets of fragile X mental retardation protein (FMRP). Purcell et al. (2014) identified a polygenic burden arising from very rare (frequency less than 1 in 1000 chromosoms) mutations and that once again, these were enriched
among ARC, NMDAR, and PSD protein complexes as well as FMRP targets and calcium channel complexes. Despite their size, neither of these studies was able to implicate specific genes. However, very recently, the UK10K group in collaboration with many of the other groups, has done so by through meta-analysis of their own case-control and de novo mutation data sets with much of the data referred to above in this section. This study (Singh et al. 2016), which included a total of 4,264 SZ cases, 9,343 controls, and 1,077 parent-proband trios, obtained genome wide significant association (p=5.6x10^{-9}) between schizophrenia and rare loss-of-function variant in SETD1A a histone methyl-transferase which the group also showed was implicated in other severe developmental disorders. The finding establishes histone methylation pathways in the pathogenesis of the disorder, and establishes overlap between SZ and other neurodevelopmental disorders at the level of a specific highly penetrant recurrent mutation in a single gene.

**Gene x Environment**

Beside genetics of schizophrenia, gene-environment interactions and epigenetic studies seem promising (van Dongen & Boomsma 2013; Owen et al. 2016; Agerbo et al. 2015). A gene-environment interaction (GxE) represents a genetic influence on vulnerability to environmental factors (Rutter et al. 2006). A comprehensive review of gene-environment interactions in schizophrenia was presented by van Os et al. (2008). However, so far, thorough replication of findings is rare and GxE research still faces several conceptual and methodological challenges (EU-GEI 2014). Published environmental exposures for psychosis for which GxE has been suggested include complications of pregnancy, paternal age, urban environment, cannabis use, migration, and childhood maltreatment. Urbanicity, migration, a lack of social support, and negative expressed emotions are considered as proxies for environmental stressors. Van Winkel et al. (2008) reviewed the role of psychosocial stress in the development of positive psychotic symptoms (hallucinations or delusions). The authors asserted the concept of „behavioral sensitization“, meaning that repetitive stress progressively increases the biological and behavioral (psychopathological) response to subsequent exposure. The substrate for this effect was postulated to be a dysregulation of the hypothalamic-pituitary-adrenal (HPA) axis, namely elevated plasma cortisol contributing to the hypothesized final common pathway of dopamine release and sensitization in the mesolimbic brain area. Polymorphisms of the genes related to catecholamine neurotransmission, neuroplasticity and stress activation (e.g. COMT, DA receptors, BDNF, HPA axis) may influence the extent of the sensitization process though the genetic findings concerning GxE in schizophrenia are not yet widely considered to be robust. A review and meta-analysis of the role of life events in psychosis was carried out by Beards et al. (2013). The meta-analysis yielded OR of 3.19. Miller et al. (2013) reviewed relevant human studies related to prenatal inflammation and risk of schizophrenia. The authors concluded that inflammation may be associated with abnormal neurodevelopment. Epigenetics (DNA methylation, histone/chromatin modifications, non-coding RNA) is related to potentially heritable or acquired changes in gene expression and function that are not caused by variation in the DNA sequence (Szyf 2014; Perkins et al. 2005). Epigenetic DNA status may be partially heritable, but during an individual’s life may also be modified by a broad spectrum of environmental stimuli (e.g. alimentation, toxins, stress or medication) (Waterland & Jirtle 2003). They may be different in different tissues or even different regions in the brain in the same subject. They may contribute to phenotypic differences in monozygotic twins (Wong et al. 2005). The first studies on epigenetic mechanisms in schizophrenia etiopathogenesis have already been performed (e.g. for review see Akbarian 2014). Based on a review, Maric and Svrakic (2012) suggest that epigenetic misregulation of the genome and direct CNS injury are probably the main mechanisms to mediate prenatal environmental effects (e.g. viruses, ethanol, or nutritional deficiency) whereas postnatal risk factors (e.g. stress, urbanicity, cannabis use) may also affect risk via potentiation of vulnerable CNS pathways implicated in schizophrenia (Pelayo-Teran et al. 2012). A review of empirical GxE studies using candidate genes in schizophrenia was published by Modinos et al. (2013).
For future GxE research in schizophrenia, it would be better to assess environmental variables quantitatively in a prospective way. The timing of environmental events is also important. Dimensions of schizophrenia should be monitored quantitatively. Important genes previously found in the GWA schizophrenia studies should be applied in GxE research into schizophrenia as well as schizophrenia endophenotypes. Epigenetic variables should likewise be studied (Moffitt et al. 2005; Jaffee & Price 2007; van Os et al. 2008; van Os & Rutten 2009; Modinos et al. 2013; Svrakic et al. 2013; Uher 2014; EU-GEI 2014).

In summary, well powered and designed studies of GxE interactions and epigenetic studies are still missing.

Conclusion/Outlook
The last few years resulted in the establishment of major new contributions to our knowledge of schizophrenia. The systematic identification of genetic variants throughout the genome opened new avenues to the biology of the disease. Genome wide association studies identified new loci containing genes involved in pathophysiology. One interesting aspect is that the genetic overlap between schizophrenia and bipolar disorder which has been debated upon for decades could be demonstrated on the molecular level. Many studies are under way to characterize these genetic variants and to find pathways of disease. A next major milestone was the discovery of multiple microdeletions and microduplications involved in schizophrenia. It became clear that also structural genetic variants contribute to the disease making them highly interesting targets for new pharmaceuticals. Especially interesting is the major overlap regarding the expressivity and thus clinical presentation or particular deletions or duplications. There is a whole neurodevelopmental spectrum including beside schizophrenia e.g. autism, intellectual disability, developmental delay, attention problems, and speech delay sometimes in combination with other physical manifestations. This observation raised many new questions regarding diagnostic criteria of psychiatric diseases and the current classification in general. Although still in their infancy whole genome sequencing of large samples will become the standard in genetic research in the very near future. The hope is to detect even very rare variants maybe responsible for subsamples or subtypes of the diseases. Furthermore, integrated systems biology approaches integrating genomics, epigenomics, transcriptomics, proteomics and metabolomics may further contribute to the identification of the pathways contributing to schizophrenia enabling a precision medicine approach to the treatment of individual patients. Whole genome sequencing of very large samples and system biology applications have the potential to lead to the desperately needed precision medicine in individual patients.


Jenkins AK, Paterson C, Wang Y, Hyde TM, Kleinman JE, Law AJ. Neurexin 1 (NRXN1) splice isoform expression during


Wray NR, Gottesman II. Using summary data from the danish national registers to estimate heritabilities for schizophrenia, bipolar disorder, and major depressive disorder. Front Genet. 2012 Jul 2;3:118.


Figure 1: Risk factors of schizophrenia

genome-wide association studies
CNV analysis
new genomic disorders
linkage analysis
sequencing
expression analysis
epigenetics

Intermediate phenotypes
imaging parameters
oculomotor function
cognitive impairment
processing speed
verbal memory
visual memory
executive function
vigilance
fluency
sensory motor gating
© Rujescu

birth cohorts, national biobanks
prenatal
(paternal age;
maternal malnutrition,
iron deficiency, bacterial infection,
obesity, lead exposure,
severe adverse life events)

perinatal
(obstetric and perinatal complications, winter birth,
reduced breastfeeding,
early maternal separation)

social factors
(unemployment/social contacts,
death/separation from parent before age 16)

demographic
(sex)

drug related
(cannabis, tobacco)

generic
(migration, urbanicity,
latitude, ethnicity)
Table 1: Microdeletions and microduplications involved in schizophrenia

<table>
<thead>
<tr>
<th>Region</th>
<th>Authors</th>
<th>Replication*</th>
<th>Meta-Analysis*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deletions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1q21.1**</td>
<td>Stefansson et al. 2008; ISC 2008; Levinson et al. 2011</td>
<td>0.0027</td>
<td>4.1 x 10-13</td>
</tr>
<tr>
<td>2p16.3</td>
<td>Kirov et al. 2008; Walsh et al. 2008; Rujescu et al. 2009; Kirov et al. 2009b</td>
<td>7.7 x 10-04</td>
<td>1.3 x 10-11</td>
</tr>
<tr>
<td>3q29</td>
<td>Mulle et al. 2010; Levinson et al. 2011</td>
<td>0.074</td>
<td>1.5 x 10-09</td>
</tr>
<tr>
<td>15q11.2**</td>
<td>Stefansson et al. 2008; Kirov et al. 2009a</td>
<td>0.046</td>
<td>2.5 x 10-10</td>
</tr>
<tr>
<td>15q13.3**</td>
<td>Stefansson et al. 2008; ISC 2008; Levinson et al. 2011</td>
<td>0.38</td>
<td>4.0 x 10-10</td>
</tr>
<tr>
<td>16p11.2**</td>
<td>Walsh et al. 2008; Guha et al. 2013</td>
<td>1.0</td>
<td>0.017</td>
</tr>
<tr>
<td>17p12</td>
<td>Kirov et al. 2009a</td>
<td>0.55</td>
<td>0.0012</td>
</tr>
<tr>
<td>17q12</td>
<td>Moreno-De-Luca et al. 2010</td>
<td>0.52</td>
<td>0.0072</td>
</tr>
<tr>
<td>Duplications</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1q21.1**</td>
<td>Levinson et al. 2011</td>
<td>0.35</td>
<td>9.9 x 10-05</td>
</tr>
<tr>
<td>7q11.23</td>
<td>Kirov et al. 2012; Mulle et al. 2014</td>
<td>0.35</td>
<td>6.9 x 10-05</td>
</tr>
<tr>
<td>7q36.3</td>
<td>Levinson et al. 2011; Vacic et al. 2011</td>
<td>0.99</td>
<td>0.27</td>
</tr>
<tr>
<td>15q11-13**</td>
<td>Kirov et al. 2008; Ingason et al. 2011a</td>
<td>0.0055</td>
<td>5.6 x 10-06</td>
</tr>
<tr>
<td>16p11.2**</td>
<td>Walsh et al. 2008; McCarthy et al. 2009; Levinson et al. 2011: Steinberg et al. 2014</td>
<td>2.3 x 10-08</td>
<td>2.9 x 10-24</td>
</tr>
<tr>
<td>16p13.11</td>
<td>Kirov et al. 2009a; Ingason et al. 2011b</td>
<td>0.056</td>
<td>5.7 x 10-05</td>
</tr>
</tbody>
</table>

* Rees et al. 2014, replication study: cases n=6.882, controls n=6.316

** region with both deletions and duplications