RARE GENOMIC DELETIONS UNDERLYING SCHIZOPHRENIA AND RELATED NEURODEVELOPMENTAL DISORDERS

Olli Pietiläinen

ACADEMIC DISSERTATION

To be presented with the permission of the Faculty of Medicine, University of Helsinki, for public examination in lecture hall 2, Haartman Institute, on 19 December, 2014 at 1 pm

Institute for Molecular Medicine Finland (FIMM)
Hjelt Institute, Faculty of Medicine
Doctoral School in Health Sciences
Doctoral Programme Brain and Mind
University of Helsinki, Finland
and
Public Health Genomics Unit
National Institute for Health and Welfare, Helsinki Finland

Helsinki 2014

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

FiMM
Institute for Molecular Medicine Finland
Nordic EMBL Partnership for Molecular Medicine

TERVEYDEN JA
HYVINVOINNIN LAITOS
National Institute for Health and Welfare

Supervisors

Professor Aarno Palotie, MD, PhD
Institute for Molecular Medicine Finland FIMM
Finland
The Broad Institute of MIT and Harvard, Boston, MA, USA
USA

Research Professor Jaana Suvisaari, MD, PhD
National Institute for Health and Welfare
Department of Mental Health and Substance Abuse Services
Finland

Academician of Science, Professor Leena Peltonen (Palotie), MD, PhD
Until 2010

Reviewers
Håkan Karlsson, PhD
Senior Scientist
Karolinska Institutet
Department of Neuroscience
Stockholm, Sweden

Helena Kääriäinen, MD, PhD
Research Professor
National Institute for Health and Welfare
Helsinki, Finland

Opponent
Cathryn Lewis, PhD
Professor
King's College London
Statistical Genetics Unit
London, United Kingdom

"In the field of scientific observation, chance favors the prepared mind"
-Louis Pasteur-

Dedicated to my family and friends

Abstract

Severe mental disorders including schizophrenia often segregate within the same families. Twin and family studies suggest that this co-occurrence is largely genetic, which implies that the different mental disorders have a shared genetic background. Some symptomatic features, such as cognitive impairment also manifest to a variable degree in the majority of severe mental disorders. Cognitive impairment occurs already before the onset of the disease and healthy family members of patients perform worse in cognitive tests than do the general population, which suggests that the cognitive impairment is indicative of genetic loading of the disease. Furthermore, the cognitive impairment persists throughout the disease and is associated with poorer outcome. This led us to hypothesize that the genetic architecture of schizophrenia is more similar to developmental disorders than had been considered earlier. Specifically, we hypothesized that rare high impact genetic variants play a role in the genetic risk for schizophrenia. Rare recurrent large-scale structural variation has long known to cause developmental syndromes, such as Prader-Willi syndrome or Velocardiofacial syndrome. In this study we investigated the role of large-scale chromosomal copy number variants in the genetic background of schizophrenia and other traits hypothesized to reflect abnormal neuronal development.

In this study four chromosomal deletions on 1q21, 15q11.2, 15q13.3 and 22q11.2 were identified to be associated with schizophrenia. Three of the deletions occurred recurrently, whereas the deletion on 22q11.22 was a founder mutation enriched especially in the North-East region of Finland. On a population level, carriers of large deletions were found to have more intellectual disability or sub-normality (IQ<85) than non-carriers. Also milder learning difficulties as measured by repeated grades in school were more common among carriers of large deletions. The four deletions specifically identified as associating with schizophrenia are linked to variable phenotypes with the strongest effect manifesting in intellectual disabilities. The regional enrichment of the deletion on 22q11.22 also enabled the assessment of recessive effects related to the deletion. Four individuals, all presenting with a neurodevelopmental phenotype and/or schizophrenia, were identified as homozygous for the deletion. This deletion overlaps one gene encoding for topoisomerase 3 beta (TOP3β) that forms a protein complex with FMRP, the fragile X mental retardation protein, via tudor domain containing 3 (TDRD3) protein.

The results of this study imply that rare high risk variants are present in a sub set of schizophrenia patients and that these variants are shared with developmental disorders. The study also demonstrates that special populations such as population

isolates can provide useful study designs in identifying rare genetic risk variants, especially with recessive effects for complex traits.

Tiivistelmä

Monet vakavat mielenterveydenhäiriöt, kuten skitsofrenia, kasaantuvat usein samoihin perheisiin. Kaksos- ja perhetutkimusten perusteella tämä yhteisesiintyminen selittyy suurelta osin geneettisillä tekijöillä, mikä viittaisi siihen, että useat diagnostisesti erotellut mielenterveyden häiriöt jakavat osan geneettisestä taustasta keskenään. Tätä tukee myös se, että osa sairauksien oireista, kuten kognitiivisten eli tiedonkäsittelyyn liittyvien toimintojen häiriöt, ovat samankaltaisia joskin vaikeusasteeltaan vaihtelevia useimmissa vakavissa mielenterveyshäiriöissä. Kognitiivisten toimintojen häiriöt ovat erityisen tavallisia skitsofreniaa sairastavilla. Ne ilmaantuvat potilailla jo ennen varsinaisen taudin puhkeamista, ja niitä esiintyy myös terveillä perheenjäsenillä. Kognitiivisten toimintojen häiriöt saattavatkin olla merkki geneettisestä alttiudesta sairauteen. Ne säilyvät läpi sairauden ja ennustavat usein huonompaa hoitotulosta. Tämä oli lähtökohtana oletukselle, että skitsofrenian geneettinen arkkitehtuuri olisi samankaltaisempi kehityksellisten sairauksien, kuten älyllisen kehitysvammaisuuden, kanssa kuin mitä aiemmin on oletettu. Tässä tutkimuksessa oletettiin erityisesti, että harvinaiset geneettiset muutokset, joihin liittyy suuri sairastumisriski, olisivat tärkeitä skitsofrenian geneettisessä etiologiassa. Harvinaisten suurikokoisten perimän rakenteellisten muutosten on pitkään tiedetty aiheuttavan normaalin kehityksen häiriöitä ja oireyhtymiä, kuten Prader-Willin tai velokardiofasiaalisen oireyhtymän. Tässä tutkimuksessa selvitettiin isokokoisten perimän kopiolukumuutosten osuutta skitsofrenian geneettisessä taustassa ja ominaisuuksissa, joiden oletimme heijastelevan hermoston kehityksen häiriötä.

Tutkimuksessa tunnistettiin neljä perimän poistumaa eli deleetiota kromosomeissa 1q21, 15q11.2, 15q13.3 ja 22q11.22, jotka assosioituivat skitsofreniaan. Kolme deleetioista ilmaantuvat toistuvasti, kun taas 22q11.22-deleetio on perustajamutaatio, joka on rikastunut väestöön erityisesti Suomessa Koillismaalla. Väestötasolla isojen deleetioiden kantajissa yleisesti havaittiin enemmän älyllisen kehityksen häiriöitä (älykkyysosamäärä alle 85) kuin henkilöissä, joilla ei havaittu isokokoisia deleetioita. Myös lievemmät oppimisvaikeudet, määriteltynä henkilön luokalle jääntinä olivat yliedustettuina suurien deleetioiden kantajissa. Kaikki neljä tunnistettua deleetiota assosioituvat vaihtelevaan ilmiasuun, jossa suurin vaikutus näyttäisi olevan juuri älyllisiin kykyihin. 22q11.22- deleetion alueellinen rikastuminen mahdollisti siihen liittyvien peittyvien ominaisuuksien tutkimisen. Neljä henkilöä kantoi kromosomin 22q11.22 deleetioita homotsygoottisesti. Heillä kaikilla on diagnosoitu skitsofrenia tai älyllisen kehityksen häiriö. Deleetio poistaa topoisomeraasi 3 beeta proteiinia koodittavan geenin (TOP3β). TOP3β proteiinin havaittiin muodostavan proteiinikompleksin FMRP:n (fragile X mental retardation protein) kanssa TDRD3 (Tudor domain containing 3) proteiinin välityksellä.

Tutkimuksen tulokset osoittavat, että pieni osajoukko skitsofreniaa sairastavista kantaa harvinaisia suuren sairastumisriskin geneettisiä muutoksia. Nämä muutokset on lisäksi jaettuja kehityksellisten häiriöiden kanssa. Tutkimus osoittaa, että geneettisesti eristäytyneet väestöt voivat tarjota edullisia tutkimusasetelmia erityisesti harvinaisten haitallisten geneettisten muutosten tunnistamisessa monitekijäisten sairauksien taustalla. Varsinkin muutoksiin liittyvien peittyvien vaikutusten havaitseminen voi olla helpompaa näissä väestöissä.

Avainsanat: skitsofrenia, kognitiiviset toiminnot, kopiolukuvariaatiot, Perimänlaajuinen assosiaatiotutkimus

Contents

List of original publications

I   Stefansson H*, Rujescu D*, Cichon S*, Pietiläinen OP, Ingason A, Steinberg S, Fossdal R, Sigurdsson E, Sigmundsson T, Buizer-Voskamp JE, Hansen T, Jakobsen KD, Muglia P, Francks C, Matthews PM, Gylfason A, Halldorsson BV, Gudbjartsson D, Thorgeirsson TE, Sigurdsson A, Jonasdottir A, Jonasdottir A, Bjornsson A, Mattiasdottir S, Blondal T, Haraldsson M, Magnusdottir BB, Giegling I, Möller HJ, Hartmann A, Shianna KV, Ge D, Need AC, Crombie C, Fraser G, Walker N, Lonnqvist J, Suvisaari J, Tuulio-Henriksson A, Paunio T, Toulopoulou T, Bramon E, Di Forti M, Murray R, Ruggeri M, Vassos E, Tosato S, Walshe M, Li T, Vasilescu C, Mühleisen TW, Wang AG, Ullum H, Djurovic S, Melle I, Olesen J, Kiemeney LA, Franke B; GROUP, Sabatti C, Freimer NB, Gulcher JR, Thorsteinsdottir U, Kong A, Andreassen OA, Ophoff RA, Georgi A, Rietschel M, Werge T, Petursson H, Goldstein DB, Nöthen MM, Peltonen L, Collier DA, St Clair D, Stefansson K.(2008) Large recurrent microdeletions associated with schizophrenia. Nature, 455 (7210) 232-6.

II   Pietiläinen OP, Rehnström K, Jakkula E, Service SK, Congdon E, Tilgmann C, Hartikainen AL, Taanila A, Heikura U, Paunio T, Ripatti S, Jarvelin MR, Isohanni M, Sabatti C, Palotie A, Freimer NB, Peltonen L (2011) Phenotype mining in CNV carriers from a population cohort. Human Molecular Genetics, 20(13):2686-95.

III Stoll G*, Pietiläinen OP*, Linder B*, Suvisaari J, Brosi C, Hennah W, Leppä V, Torniainen M, Ripatti S, Ala-Mello S, Plöttner O, Rehnström K, Tuulio-Henriksson A, Varilo T, Tallila J, Kristiansson K, Isohanni M, Kaprio J, Eriksson JG, Raitakari OT, Lehtimäki T, Jarvelin MR, Salomaa V, Hurles M, Stefansson H, Peltonen L, Sullivan PF, Paunio T, Lönnqvist J, Daly MJ, Fischer U, Freimer NB, Palotie A. (2013) Deletion of TOP3β, a component of FMRP-containing mRNPs, contributes to neurodevelopmental disorders. Nat Neurosci. 16(9):1228-37.

     *These authors contributed equally

Abbreviations

A   Adenine

ADHD   Attention deficit and hyper activity disorder

ALS   Amyotrophic lateral sclerosis

ARC   Activity regulated cytoskeleton associated scaffold protein

AS   Angelman syndrome

BAF   B-allele frequency

C   Cytosine

CEU   Central Europeans in Utah

CGH   Comparative Genomic Hybridization

CHB   Han Chinese from Beijing

CI   Confidence Interval

CNP   Copy number polymorphism

CN   Copy Number

CNV   Copy number variation/variant

CNVR   CNV- region

CVLT   California Verbal Learning Test

Dilgom   Dietary, Lifestyle and Genetic determinants of Obesity and Metabolic syndrome in the Helsinki region

DNA   Deoxyribonucleic acid

DSM   Diagnostic and Statistical Manual of Mental Disorders

DZ   Dizygotic

EJC   Exon junction complex

ENCODE	Encyclopedia of DNA elements
eQTL	Expression quantitative trait loci
FDH	Finnish disease heritage
FRAX	Fragile X syndrome
FSFS	Finnish schizophrenia family sample
FTC	Finnish Twin Cohort
G	Guanine
GO	Gene ontology
GWAS	Genome-wide association study
H2000	Health 2000
HBCS	Helsinki Birth Cohort Study
HEK293	Human embryonic kidney 293 cells
HWE	Hardy-Weinberg equilibrium
IBD	Identity by descent
ICD	International Classification of Diseases and Related Health problems
IDF	International Diabetes Federation
IQ	Intelligence quotient
ISC	the International Schizophrenia Consortium
JPT	Japanese from Tokyo
kb	Kilobase
LCR	Low copy repeat
LD	Linkage disequilibrium

LoF	Loss of function
LRR	Normalized Total Intensity Log R ratio
MAF	Minor allele frequency
MHC	Major histocompatibility complex
MMRBI	Microhomology-mediated break-induced replication
mRNA	Messenger ribonucleic acid
mRNP	Messenger ribonucleoproteins
mtDNA	Mitochondrial DNA
MZ	Monozygotic
NAHR	Nonallelic homologous recombination
NCBI	National Center for Biotechnology Information
NFBC1966	Northern Finland 1966 Birth Cohort
NFBC 1986	Northern Finland 1986 Birth Cohort
NHEJ	Nonhomologous end joining
NMDA	N-methyl-D-aspartate
NMDAR	N-methyl-D-aspartate Receptor Complex
OMIM	Online Mendelian Inheritance in Man
OPCRIT	Operational Criteria Checklist
OR	Odds ratio
PCR	Polymerase chain reaction
PWS	Prader-Willi Syndrome
QC	Quality control

QRT-PCR	Quantitative Real-time Polymerase Chain Reaction
RNA	Ribonucleic acid
SANS	Scale for the Assessment of Negative Symptoms
SAPS	Scale for the Assessment of Positive Symptoms
SCID	Structured Clinical Interview for DSM-IV
SNP	Single nucleotide polymorphism
SSC	Swedish Schizophrenia Cohort
T	Thymine
T2D	Type II diabetes
THL	The National Institute for Health and Welfare
VCFS	Velocardiofacial syndrome
WAIS-R	Wechsler Adult Intelligence Scale – Revised
WBS	Williams-Beuren Syndrome
WES	Whole exome sequencing
WMS-R	Wechsler Memory Scale – Revised
WTSI	The Wellcome Trust Sanger Institute
YFS	the Cardio Vascular Risk in Young Finns Study
YRI	Yorubans of Nigeria

1 Introduction

In addition to its vital life sustaining role, the higher level functions of the brain are fundamental to human behavior, intellect, cognition and are core features of the concept of humanity. Therefore disorders affecting these brain functions are so devastating and compromise the very existence of the person affected. Mental disorders have puzzled humanity for millennia, but their fundamental causes and underlying biology have remained a mystery. Historically, mental disorders have been separated from other medical illnesses because they affect the mind (Andreasen 1997). Schizophrenia was recognized as its own disease entity and separated from manic-depressive illness and dementia in the elderly (later named as Alzheimer's disease) 120 years ago by Emil Kraepelin (Andreasen 1995). However, since then the causes for schizophrenia and other mental disorders have remained unknown and until the discovery of first effective antipsychotic drugs in the 1950's, people suffering from schizophrenia could not be treated efficiently. In the history of mankind, people with severe mental disorders have faced cruelty due to the lack of understanding and unjustified assumptions of the disease etiology (Brown 1997). Currently, even though some symptoms can be treated, other core features of the disorders remain, and in case of schizophrenia, individuals affected by it rarely achieve full recovery (Harrison *et al.* 2001, Robinson *et al.* 2004, Jaaskelainen *et al.* 2013). Since schizophrenia onsets at a young age also the economic impact on society is enormous (Knapp *et al.* 2004).

Today, 120 years after the clinical definition of schizophrenia and 60 years after the introduction of first efficient neuroleptic drugs the biological background of schizophrenia and other mental disorders is finally emerging. Based on evidence from pharmacological studies, schizophrenia has been considered as both a "dopamine disorder" and more recently a "glutamate disorder", yet it has remained unclear how these are biologically related to the underlying causes of the disorder (Insel 2010). The heritability of severe mental disorders, including schizophrenia is among the highest for complex human diseases, and undisputedly refers to the biological nature of the disease (Sullivan *et al.* 2012). Consequently, one of the major drivers for the new biological insights has been the rapid development of genetic research technologies. These technologies have enabled the investigation of genetic variants genome-wide in high resolution and with adequate sample sizes without an *a priori* hypothesis of biological network or genes, but rather allowing the generation of novel hypothesis based on the genes and the networks identified. The governing idea underlying genome-wide association studies is the common disease common variant hypothesis. This hypothesis postulates that an individual's

genetic risk for a disease is attributable to multiple genetic variants each contributing only a minor increase to the disease risk. According to this hypothesis majority of the genetic risk for common complex diseases is due to common ancestral variants that are shared across populations. That is, the genetic risk for these diseases is also shared across populations (Lander 1996, Risch *et al.* 1996). After the first successful genome-wide association studies (GWAS) published in 2005 for macula degeneration (Klein *et al.* 2005), thousands of genomic loci have been identified underlying various complex traits (www.genome.gov/GWAStudies).

The first large-scale GWAS of schizophrenia were published in 2009 each including roughly 3000 patients (Purcell *et al.* 2009, Shi *et al.* 2009, Stefansson *et al.* 2009). These revealed associations with the major histocompatibility complex (MHC) on chromosome six. In addition, a polygenic score that consisted of multiple nominally associated common variants was found to underlie a significant portion of the genetic risk for schizophrenia. Supported by the polygenic risk score, the role of common low impact variants in schizophrenia seemed evident also in the genetic level and GWAS meta-analyses with larger sample sizes were conducted (Purcell *et al.* 2009). Currently, 108 loci for schizophrenia have been identified and one third of the variance in liability is estimated to be accounted for by mostly common genetic variants (Ripke *et al.* 2013, Schizophrenia Working Group of the Psychiatric Genomics 2014). Taken together the genetic studies support schizophrenia as a highly multifactorial disease resulting from the interplay between both genes and environment (Howes *et al.* 2014).

Alongside the common disease common variant hypothesis, a major role for rare genetic variants has been suggested in schizophrenia. The common disease rare variant hypothesis argues that multiple rare genetic variants, each having relatively high penetrance, are major contributors for the genetic susceptibility to common diseases (Schork *et al.* 2009). The common disease rare variant hypothesis in mental disorders was strengthened by earlier genetic findings in other severe early onset developmental disorders including intellectual disability and autism. Even in schizophrenia the Velocardiofacial Syndrome (VCFS), caused by a micro-deletion on chromosome 22q11.2, is one of the highest known singular risk factors (Karayiorgou *et al.* 1995). There was also evidence for family specific high risk variants such as the balanced translocation disrupting *DISC1* (Millar *et al.* 2000). Furthermore, the existence of rare high risk *de novo* variants is compatible with the evolutionary paradigm of reduced fecundity, increased mortality, and large portion of sporadic cases observed for schizophrenia. Even prior to the first GWAS results for common Single Nucleotide Polymorphisms (SNPs), rare recurrent large scale genomic Copy Number Variants (CNVs) were found to associate with a markedly

increased risk for schizophrenia (International Schizophrenia Consortium 2008, Stefansson *et al.* 2008). The early findings have been since then replicated in multiple studies and today at least 15 rare CNVs are consistently associated with schizophrenia (Malhotra *et al.* 2012). Although important contributors to carrier's individual genetic risk, the global impact of rare high risk CNVs is limited to a small percentage of patients.

Although still in its early days, common biological themes have started to emerge from the genetic findings of schizophrenia and other mental disorders. Most importantly, schizophrenia appears to be as biological in origin as any somatic disorders and has a strong genetic component. Genetic architecture of schizophrenia follows a complex inheritance pattern that is governed by thousands of common variants: a pattern that is similar to many other complex traits including adult height and susceptibility to other complex diseases (Sullivan *et al.* 2012). Yet, also rare and *de novo* variants contribute to schizophrenia at least in some sub-groups of the patients (Owen *et al.* 2010). The schizophrenia associated genes are overrepresented with neuronal expressed genes and include synaptic gene-networks related to plasticity and synaptic transmission (Kirov *et al.* 2012, Fromer *et al.* 2014, Purcell *et al.* 2014, Schizophrenia Working Group of the Psychiatric Genomics 2014). Despite the clinical distinction between psychiatric disease entities, following Kraepelin's idea from over century ago, the genetic evidence supports shared biological background across mental disorders, including schizophrenia, bipolar disorder, major depressive disorder, autism, attention deficit and hyper activity disorder (ADHD), and intellectual disability (Sullivan *et al.* 2012). Furthermore, the growing body of biological knowledge of not only severe mental disorders but also neurological disorders has begun to eliminate classical distinctions between psychiatric and neurological disease entities. This may have a profound impact in our way of conceptualizing and treating severe mental disorders in the future.

2 Review of the Literature

2.1 Schizophrenia and related severe mental disorders

2.1.1 Clinical features of schizophrenia

There are currently no biomarkers for schizophrenia or for most other mental illnesses. They are syndromes and their diagnoses are based on combination of signs, symptoms, and outcome. The fact that most symptoms form a continuum with normality adds an extra layer of complexity for diagnosis, which is why the diagnostic criteria include thresholds for severity and duration. Schizophrenia was originally called dementia praecox, and it was separated from manic depressive illness and from dementia in elderly (Alzheimer's disease) in the late 19th century by Emil Kraepelin. The name schizophrenia, introduced by Bleuler in 1911, emphasizes severe fragmentation of thinking and personality associated with the disorder (Andreasen 1995). Schizophrenia has been considered a disorder of executive functioning of the brain that affects multiple separate domains (perception, inferential thinking, language, behavioral monitoring, conceptual fluency, emotional expression, and avolition). Schizophrenia is a multifactorial disease that is defined by symptoms that are commonly divided to positive, disorganized and negative symptoms. Positive symptoms are exaggerations of normal functions, such as hallucinations and delusions. Disorganized symptoms include bizarre behavior, disorganized speech and inappropriate affect. Negative symptoms are loss of normal functions, such as affective blunting, social anhedonia, and poverty of speech (Andreasen 1995).

The diagnosis of schizophrenia is usually defined in the international research practice by the criteria specified in the Diagnostic and Statistical Manual of Mental Disorders. In this study the fourth edition (DSM-IV) has primarily been used to define schizophrenia (Table 1) (American Psychiatric Association 1994). In DSM-IV the psychotic disorders are distinguished from each other by the duration, dysfunction, type of delusions and hallucinations, presence of affective symptoms, and associated substance use or medical condition. In clinical practice schizophrenia is diagnosed according to the criteria specified on the Tenth Revision of the International Classification of Diseases and Related Health Problems (ICD-10). The diagnostic criteria of ICD-10 and DSM-IV overlap on most part, but the criteria differ for example in the required duration of symptoms (one month in ICD-10 versus six months in DSM-IV), emphasis given for the auditory hallucinations,

Table 1. DSM-IV diagnostic criteria for schizophrenia (American Psychiatric Association 1994)

Criterion A. Characteristic symptoms: Two (or more) of the following, each present for a significant portion of time during a 1-month period (or less if successfully treated):

Positive symptoms:

A1: Delusions

A2: Hallucinations

A3: Disorganized speech

A4: Grossly disorganized or catatonic behavior

A5: Negative symptoms:

- Affective flattening
- Alogia
- Avolition

Only one Criterion A symptom is required if delusions are bizarre or hallucinations consist of a voice keeping up a running commentary on the person's behavior or thoughts, or two or more voices conversing with each other.

Criterion B. Social/occupational dysfunction: For a significant portion of the time since the onset of the disturbance, one or more major areas of functioning such as work, interpersonal relations, or self-care are markedly below the level achieved prior to the onset (or when the onset is in childhood or adolescence, failure to achieve expected level of interpersonal, academic, or occupational achievement).

Criterion C. Duration: Continuous signs of the disturbance persist for at least 6 months. This 6-month period must include at least 1 month of symptoms (or less if successfully treated) that meet Criterion A (i.e., active-phase symptoms) and may include periods of prodromal or residual symptoms. During these prodromal or residual periods, the signs of the disturbance may be manifested by only negative symptoms or two or more symptoms listed in Criterion A present in an attenuated form (e.g. odd beliefs, unusual perceptual experiences).

Criterion D. Schizoaffective and Mood Disorder exclusion: Schizoaffective Disorder and Mood Disorder With psychotic Features have been ruled out because either (1) no Major Depressive, Manic, or Mixed Episodes have occurred concurrently with the active-phase symptoms; or (2) if mood episodes have occurred during active-phase symptoms, their total duration has been brief relative to the duration of the active and residual periods.

Criterion E. Substance /general medical condition exclusion: The disturbance is not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition.

Criterion F. Relationship to a pervasive Developmental Disorder: If there is a history of Autistic Disorder or another Pervasive Developmental Disorder, the additional diagnosis of Schizophrenia is made only if prominent delusions or hallucinations are also present for at least a month (or less if successfully treated).

---

distinction between schizoaffective disorder, and emphasis on the requirement for poor outcome in DSM-IV (Perälä 2013).

There is significant heterogeneity in the combinations of symptoms in schizophrenia, and none of them are specific for schizophrenia (van Os *et al.* 2009). Although schizophrenia has been considered a unique entity, its diagnostic criteria have varied throughout the six editions of DSMs. The DSM-IV was recently updated to DSM-5 (Tandon *et al.* 2013). The core of the six diagnostic criteria (A-F, table 1) for schizophrenia have remained virtually the same and most people who met criteria for schizophrenia in DSM-IV will do so also in DSM-5. In DSM-5 at least two of criteria A need to be present in a patient, while prioritizing criteria A1-A3, and the emphasis on the bizarreness of the delusions has been omitted. In addition the definition of negative symptoms has been clarified to "Negative symptoms, i.e. diminished emotional expression or avolition". Also the criteria F has been slightly modified from DSM-IV and states in DSM-5: " If there is a history of autism spectrum disorder or other communication disorder of childhood onset, the additional diagnosis of schizophrenia is made only if prominent delusions or hallucinations are also present for at least 1 month (or less if successfully treated)."(Tandon *et al.* 2013).

The symptoms of psychotic disorders can be divided into five dimensions based on psychopathology: psychosis (positive symptoms), negative symptoms, cognitive impairment, and affective dysregulation including mania and depression dimensions. The dimensions for positive and negative symptoms seem to follow different paths over time and only the negative dimension is associated with neurocognitive alterations. The symptoms of different psychotic disorders, including schizophrenia, schizoaffective disorder, and bipolar disorder overlap in the five dimensions. Schizophrenia diagnosis is used for psychosis that is associated with few affective (related to mood) symptoms. In contrast, psychosis with fewer negative symptoms that is preceded by high level of affective symptoms is diagnosed as bipolar disorder or depression. Schizoaffective disorder lies between affective and non-affective psychosis (Figure 1)(van Os *et al.* 2009).
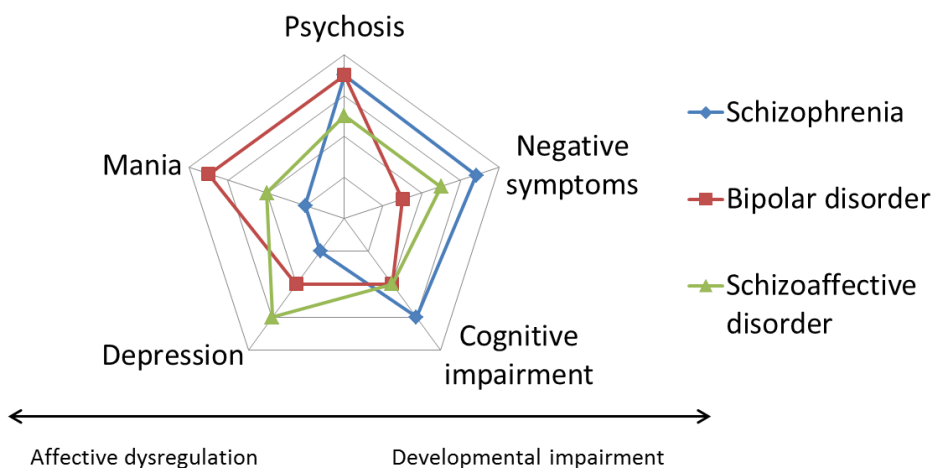
Figure 1.   Diagnostic dimension of schizophrenia, schizoaffective disorder, and bipolar disorder. Modified from (van Os *et al.* 2009)

DSM-5 has excluded the DSM-IV subtypes of schizophrenia and replaced them with dimensional assessment of eight psychopathological dimensions of schizophrenia. The aim in DSM-5 has been to emphasize the high level of heterogeneity in the combination and severity of symptoms, which are also known to vary over time among patients. Also, particularly supported by recent genetic evidence, the overlap and comorbidity between traditionally distinct diagnostic entities, is taken better into account with the dimensionality. In addition to the five criteria A of DSM-IV for schizophrenia (Table 1), the eight domains of dimensional assessment in DSM-5 include depression, mania and impaired cognition. The domains are measured in a scale of zero to four: 0 (not present), 1 (equivocal), 2 (present, but mild), 3 (present and moderate), and 4 (present and severe). In comparison to DSM-IV, when an individual's primary symptoms of psychosis (criteria A) score higher than two it can be considered to fulfill the DSM-IV criteria. In addition, the DSM-5 includes a dimension of cognitive functioning in the diagnostic assessment; however it is not included as a differential diagnostic symptom of schizophrenia. This is because, although, a marked portion of individuals with psychotic disorders have impairments over a range of cognitive domains, the impairments are similar in both affective (related to mood) and non-affective (not related to mood) psychoses, even though the impairment seems more severe in non-affective psychosis (Barch *et al.* 2013, Tandon *et al.* 2013).

2.1.2 Cognitive functioning in schizophrenia

Schizophrenia is associated with generalized impairment of various cognitive functions, and patients perform on average one standard deviation below unaffected controls (Heinrichs *et al.* 1998, Mesholam-Gately *et al.* 2009, Fioravanti *et al.* 2012). Cognitive functions are defined as the processes of acquiring, processing, storing of, and acting upon information. They are assessed by standardized neuropsychological test methods that measure, for example, memory, global cognitive abilities, verbal skills, and executive functioning. The impairments in schizophrenia patients are already broadly present during the premorbid period and the first episode and they remain similar throughout the later phases of illness (Woodberry *et al.* 2008, Mesholam-Gately *et al.* 2009, Reichenberg *et al.* 2010). Cognitive impairments are more severe among patients with an early onset disease (Rajji *et al.* 2009). Relatives of schizophrenia patients perform intermediate between schizophrenia patients and controls in neurocognitive measures, suggesting that these deficits indicate a genetic predisposition to the disease (Faraone *et al.* 2000, Sitskoorn *et al.* 2004, Toulopoulou *et al.* 2010). Although there is significant heterogeneity in neurocognitive dysfunction among patients (Kremen *et al.* 2004), many consider it as a core feature of schizophrenia (Elvevag *et al.* 2000). However, it is still not included as a characteristic feature even in the DSM-5 diagnostic criteria. This is mainly because, although cognitive impairment is acknowledged to be present in large portion of cases, it does not distinguish between schizophrenia and other psychiatric disorders (Barch *et al.* 2013, Tandon *et al.* 2013). Some studies have reported weaker school performance and delays in early developmental mile stones in children who later develop schizophrenia (Jones *et al.* 1994, Sorensen *et al.* 2010). Patients with schizophrenia in Finland have not been shown to deviate in academic factors of school performance (Cannon *et al.* 1999). However, there is a paucity in progressing to high school and individuals who are diagnosed with a psychiatric illness have more often repeated grades in school (Isohanni *et al.* 1998, Cannon *et al.* 1999).

2.1.3 Neurodevelopmental model of schizophrenia

The prevailing and widely accepted hypothesis is that schizophrenia is a neurodevelopmental disorder that results as an end state of abnormal developmental processes that have started long before the onset of the disease. A significant body of evidence supports the gradual development of schizophrenia, although the risk factors are often nonspecific for schizophrenia but are rather associated with risk to a broad range of psychotic and developmental disorders (Rapoport *et al.* 2005, Insel 2010, Rapoport *et al.* 2012). Individuals who later develop schizophrenia have lower cognitive and motor performance, and will already have had more psychotic symptoms in childhood. Schizophrenia is associated with pre- and perinatal risk factors, including prenatal infections, famine during gestational period and placental

pathology. Consistently, also low birth weight is a significant risk factor for schizophrenia. Other premorbid risk factors include urban environment, ethnic minority, and childhood trauma, but as is the case for other risk factors these are diagnostically non-specific. The developmental model of schizophrenia is further supported by evidence from longitudinal neuroimaging studies that suggest progressive grey and white matter abnormalities occurring particularly in the prefrontal and temporal cortices of patients with schizophrenia. Finally the developmental model is supported by recent genetic findings that link schizophrenia with other developmental disorders. The premorbid indicators of schizophrenia are subtle and non-specific, but the consistency of the findings support that psychosis does not occur in a completely healthy brain (Rapoport *et al.* 2005, Insel 2010, Rapoport *et al.* 2012).

2.1.4 Prevalence of schizophrenia

Schizophrenia, as most mental disorders, has a peak age of onset in the second and third decades of life (Andreasen 1995, Pedersen *et al.* 2014). The prevalence and incidence estimates for schizophrenia vary between studies depending on what diagnostic criteria were used and whether the study was conducted as a register or community survey study (Pedersen *et al.* 2014). The global incidence for schizophrenia is relatively low, with only 15.2 per 100 000 individuals per year reported (McGrath *et al.* 2004). The median global lifetime prevalence of schizophrenia is 0.4% but the prevalence estimates vary over six-fold across studies (Saha *et al.* 2005, McGrath *et al.* 2008). A recent register based study conducted on a Danish population concluded a lifetime cumulative incidence (i.e. the probability of being treated before 50 years of age) of 1.73% for males and 1.24% for females, and a lifetime prevalence of 1.93% for males and 1.56% for females. Over 3.5% of population was treated because of schizophrenia and related disorders (Pedersen *et al.* 2014). The prevalence for schizophrenia in Finland is high (0.87%-2.2%) and the lifetime prevalence for all psychotic disorders is over 3% (Lehtinen *et al.* 1990, Hovatta *et al.* 1997, Cannon *et al.* 1998, Arajarvi *et al.* 2005, Perala *et al.* 2007). The risk for the disorder varies regionally (Hovatta *et al.* 1997, Haukka *et al.* 2001, Perala *et al.* 2007). In general there is a south-north and west-east gradient with odds ratios (ORs) of 4.0 and 7.2, respectively (Perala *et al.* 2007). Among the highest risks is observed in the North-Eastern isolate of Kuusamo with a three-fold age corrected lifetime risk for schizophrenia compared to rest of Finland (Hovatta *et al.* 1997). Globally, the life time morbid risk for schizophrenia is 0.72% (Saha *et al.* 2005). The regional variation in the life time risk for schizophrenia in Finland may reflect regional genetic, environmental, and sociodemographic differences (Perala *et al.* 2008).

2.1.5 Genetic background of schizophrenia

Schizophrenia is a highly heritable disorder with heritability estimates in twins ranging up to 80% (Sullivan *et al.* 2003). In Finland the corresponding estimate in twins is 83% (Cannon *et al.* 1998). In fact, family history is the strongest single risk factor for schizophrenia (Mortensen *et al.* 1999, Sullivan 2005). The population attributable risk is still modest, however, and a marked portion of schizophrenia patients do not have a family history of schizophrenia (Mortensen *et al.* 1999). Schizophrenia is associated with a broad range of mental disorders in first degree relatives, which suggests that the genetic factors contributing to these disorders are also shared (Lichtenstein *et al.* 2009, Mortensen *et al.* 2009). Recent genetic studies including those results that are specifically presented in this thesis directly support there being a shared genetic etiology between a broad range of mental and developmental illnesses. Schizophrenia heritability accounted by common SNPs in case control setting is estimated to be close to 32% (Ripke *et al.* 2013). Furthermore, there is a significant genetic correlation calculated from common SNPs between schizophrenia and bipolar disorder (0.68), major depressive disorder (0.43), and autism spectrum disorders (0.16) (Cross-Disorder Group of the Psychiatric Genomics *et al.* 2013). In terms of co-heritability with schizophrenia, this would respectively translate into 15%, 8.7%, and 3%. (Cross-Disorder Group of the Psychiatric Genomics *et al.* 2013) While the strongest correlation with common variants in schizophrenia is with affective psychosis, rare genetic variants associated with schizophrenia are often shared with developmental disorders, including intellectual disability and autism spectrum disorders (Malhotra *et al.* 2012). Moreover, both in schizophrenia and autism the association with potentially high impact *de novo* variants is driven by cases with cognitive deficits, whereas cases with normal cognitive abilities do not differ from controls (Fromer *et al.* 2014).

2.1.6 Genetic background of Intellectual impairment

The impairment in cognitive functions is a distinctive feature shared between schizophrenia and some developmental disorders. Considering the genetic architecture of intellectual deficit or learning disabilities more broadly, the cumulative evidence would support these disorders as complex and genetically heterogeneous. Cases with severe intellectual disability commonly do not have family history of low intelligence quotient (IQ) and their close relatives do not differ from general population in their cognitive abilities, suggesting that there is heterogeneity in the genetic architecture of intellectual disability with the most severe forms having *de novo* or recessive origins. On the other hand, the normal range of intelligence has a significant genetic component that consists of multiple genes with small effects (Davies *et al.* 2011). Cases with mild intellectual disability

have more affected relatives, which implies that mild intellectual disability is part of the main IQ distribution and is genetically driven by a polygenic signal (Nichols 1984). A study on Swedish school children with mild intellectual disability of unknown cause, reported that half of the children had at least one first degree relative with borderline or subnormal intelligence according to school history and the condition was associated with neurological and psychiatric abnormalities (Hagberg *et al.* 1981). Similarly, standardized mathematics and reading scores correlate with each other, and specific learning difficulties defined as test scores below one standard deviation in arithmetic, reading, and spelling disorders often coexist in individuals suggesting common biological background of specific scholastic skill levels (Landerl *et al.* 2010, Davis *et al.* 2014). These scholastic skill levels are highly heritable and approximately one half of their correlation is accounted by shared genetic effects referred to as "generalized genes" (Kovas *et al.* 2007, Davis *et al.* 2014). Collectively, it seems that genes affecting normal variation in learning abilities also affect learning disabilities and these effects are shared across different scholastic domains (e.g. mathematical and verbal skills) (Plomin *et al.* 2005). Consistently, mild intellectual disability, but not severe, is associated with lower educational and socioeconomic level of parents of patients (Camp *et al.* 1998, Stromme *et al.* 2000, Chapman *et al.* 2002, Heikura *et al.* 2008).

The prevalence of both mild (IQ < 70) and severe (IQ < 50) intellectual disability in northern Finland is approximately twice the values measured in Sweden with a combined prevalence of 1.2% (Rantakallio *et al.* 1986). During 20 years the prevalence have remained same, although there has been a shift from severe towards mild intellectual disability (Heikura *et al.* 2003). Also the parental socioeconomic and educational level is associated with intellectual disability in Northern Finland (Heikura *et al.* 2008). Intellectual sub-normality (IQ: 71-85) has a prevalence of 1.4% in Northern Finland. Of children born in 1966 3.8% were below the appropriate school grade for their age and of these 66% had IQ above 85 (Rantakallio *et al.* 1986). Furthermore, a possible causal prenatal, perinatal or postnatal factor could be assigned less often for mild (IQ<70) and borderline (IQ< 86) intellectual disability than for severe intellectual disability (IQ<50) (Rantakallio *et al.* 1985, von Wendt *et al.* 1987). When taken together the evidence would support a complex polygenic architecture for intellectual disability in Finland that might also be shared with other severe mental disorders, such as schizophrenia.

2.2 Human genome

2.2.1 Structure

The genetic code is encrypted to an alternating sequence of four organic bases: Adenine (A), Cytosine (C), Guanine (G), and Thymine (T) that are joined together by 2'-deoxyribose diphosphoester chain to form deoxyribonucleic acid (DNA). Two complementary DNA strands, running in opposite directions, are bound by hydrogen bonds, generating a double helix, where A pairs only with T and C only with G (Figure 2) (Watson *et al.* 1953). A diploid human genome consists of approximately 6.5 billion base pairs that are organized into 23 pairs of nuclear chromosomes and into 16.5 kb circular mitochondrial DNA (mtDNA). Of the nuclear chromosomes, 22 pairs are called autosomes and the 23rd pair comprises the sex chromosomes which occur either as two X chromosomes (females) or an X and a Y (males). Half of individual's diploid genome is inherited maternally and the other half paternally.
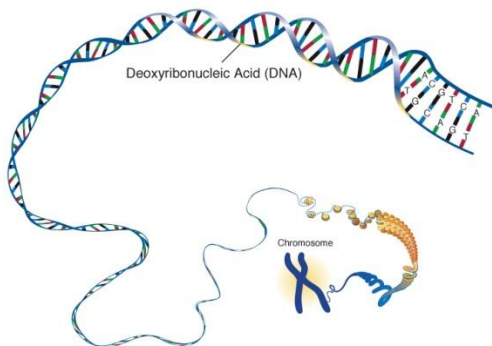
Figure 2.   The structure of DNA and its organization to chromosomes (modified from www.genome.gov/Glossary).

The first draft of the human genome was published in 2001 (Lander *et al.* 2001, Venter *et al.* 2001), and the first release of the completed reference sequence of 2.85 billion nucleotides in 2004 (NCBI Human Genome Build 35) (International Human Genome Sequencing Consortium 2004). The definition of a gene has varied over time from a unit of inheritance that carries a characteristic from parent to child into being associated with instructions for building proteins or as being discrete physical parts of a sequence transcribed into messenger ribonucleic acid (mRNA) (Pearson 2006). Currently a more broad view of the gene has been adopted and it can be

defined as "a union of genomic sequences encoding a coherent set of potentially overlapping functional products" (Gerstein *et al.* 2007). Approximately only 1.2% of the genome encodes for proteins, it is termed the exome and it is distributed into roughly 20,000 protein coding genes (Ng *et al.* 2009, Durbin *et al.* 2011, Encode Project Consortium *et al.* 2012, Harrow *et al.* 2012). The protein coding genes are fragments of DNA that contain the information for the amino acid sequence of a protein. The non-coding regions between the coding exons within a gene are termed introns. Intronic sequences are removed from mRNA in a process known as splicing that takes place prior to protein translation. Multiple different transcripts called isoforms can be formed from one gene by alternative splicing. The number of protein-coding transcripts is over four-times the number of protein-coding genes (Gencode, version 19, July 2013, http://www.gencodegenes.org). The remaining 99% of the genome was previously considered to be largely nonfunctional. However, the bulk of this so called "junk DNA" is now known to have wide spread functional elements with important biological roles delineated in the Encyclopedia of DNA elements (ENCODE) project (Encode Project Consortium *et al.* 2012). In addition to the protein coding genes, thousands of non-coding RNA genes (Djebali *et al.* 2012, Harrow *et al.* 2012), regulatory(Sanyal *et al.* 2012, Thurman *et al.* 2012), and protein interacting elements (Neph *et al.* 2012) exist in the human genome. It is estimated that at least 80% of the genome participates in at least one biochemical function (Encode Project Consortium *et al.* 2012).

Any two individuals are 99.9% identical with each other in their DNA (Przeworski *et al.* 2000, Reich *et al.* 2002). Still both sequence and structural variation exists in the genome and results in much of the differences seen between individuals. These genetic variants can also be used as genetic markers in the mapping of genomic regions that contribute to the phenotypic variation (Altshuler *et al.* 2008).

2.2.2 Single nucleotide polymorphisms

SNPs have been the most accessible and well characterized form of genetic variation. In parallel with the Human Genome Project, a project to identify and catalog common human variation was initiated (Sachidanandam *et al.* 2001). The goal of the SNP Consortium was to achieve a comprehensive view of variation that could then be systematically used to explore associations for human traits. The number of SNPs with minor allele frequency of over 1% is around 10 million (Reich *et al.* 2003). Consequently, the majority of heterozygous sites in each individual are due to common variants even though collectively most varying sites are rare (Sachidanandam *et al.* 2001, International HapMap Consortium 2005, Durbin *et al.* 2011). Re-sequencing of thousands of individuals has uncovered a multitude of new,

mostly rare, variants (Wheeler *et al.* 2008, Durbin *et al.* 2011, Genomes Project Consortium *et al.* 2012). The current dbSNP release includes over 60 million variants (dbSNP Build 138 for Human, Date: 23 July 2013, http://www.ncbi.nlm.nih.gov/projects/SNP/).

The substitution mutation rate in the human genome is estimated to be ~1.2×10^{-8} per base per generation and correlates strongly with paternal age (Kong *et al.* 2012). The majority of *de novo* single nucleotide mutations are paternal in origin (Kong *et al.* 2012). SNPs have a low rate of recurrent mutation, which is why common ancestral SNPs have generally singular origin in the ancestral human populations and are shared between populations. The frequencies can however vary substantially across populations (International HapMap Consortium 2003, Hinds *et al.* 2005, International HapMap Consortium 2005). In contrast to common variants, low-frequency variants show significant geographic differentiation (Genomes Project Consortium *et al.* 2012). Rare variants are enriched with functional variants that directly change the amino acid sequence (Genomes Project Consortium *et al.* 2012). Still, each individual carries approximately 100 Loss of Function (LoF) variants, of which 20 are in the homozygous state (MacArthur *et al.* 2012).

2.2.3 Structural variation in human genome

Genetic variants that alter genetic structure are broadly defined as structural variation. They can be further categorized into balanced changes, such as inversions and translocations or as changes that alter the copy number of the DNA, termed copy number variants (CNVs). CNVs are in turn further categorized as deletions and duplications (Figure 3). The size variation of structural variants forms a continuum in the genome ranging from single bases to whole chromosomes, but the term is usually reserved for variation of over 1 kb in size (Feuk *et al.* 2006, Hurles *et al.* 2008). CNVs vary significantly in size ranging up to several mega bases (Iafrate *et al.* 2004). The larger (>100 kb) CNVs tend to be rare, although around half of them are still observed in multiple individuals (Iafrate *et al.* 2004, Sebat *et al.* 2004, Sharp *et al.* 2006). In deed, it has been estimated that any two individuals differ by 11 CNVs of over 100 kb in size (Sebat *et al.* 2004). Large scale CNVs, as a class of variants, account for most bases that vary in the genome (Wheeler *et al.* 2008).
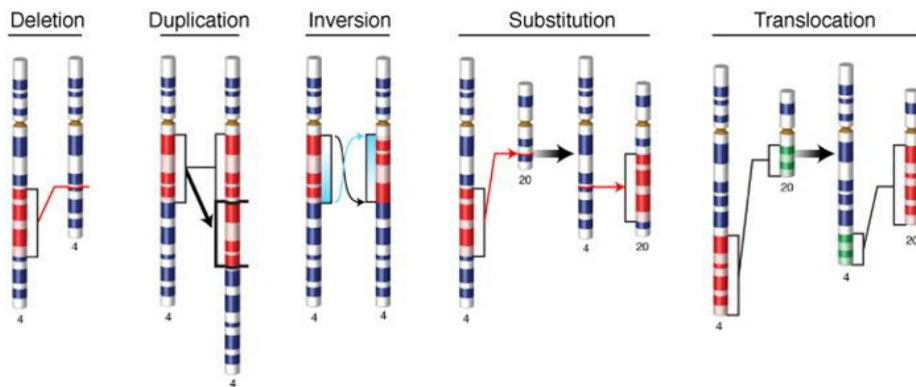
Figure 3. Description of types of large scale structural mutations in the genome (modified from www.genome.gov/Glossary).

The extent of large-scale CNVs in the human genome outside the context of cancer and chromosomal abnormalities began to be realized in 2004 (Fredman *et al.* 2004, Iafrate *et al.* 2004, Sebat *et al.* 2004). A first comprehensive map of CNVs in the human genome was generated in 2006 with over 1,447 sites varying in copy number (Redon *et al.* 2006). Currently, thousands of CNV regions have been recognized and they encompass a marked portion of the genome ranging from early estimates of 12% to current 3% (Redon *et al.* 2006, Conrad *et al.* 2009). Similarly to sequence variants common CNVs are shared across populations whereas the rare ones are often observed in a single population (Mills *et al.* 2011). CNVs have been demonstrated to have impact on biological processes including gene expression and disease susceptibility (Stranger *et al.* 2007, Hurles *et al.* 2008). While numerous functional sequences including genes and conserved sequences are overlapped by CNVs (Iafrate *et al.* 2004, Tuzun *et al.* 2005, Redon *et al.* 2006, Conrad *et al.* 2009), both deletions and duplications are generally biased away from genes (Conrad *et al.* 2006, Redon *et al.* 2006, Conrad *et al.* 2009). Moreover, this paucity is stronger among deletions and common CNVs (Conrad *et al.* 2009). Deletions are on average also shorter than duplications (Redon *et al.* 2006).

The CNV mutation rate is estimated to be 3×10^{-2} per haploid genome per generation (Conrad *et al.* 2009). CNVs tend to cluster in genomic regions containing interspersed duplications, called segmental duplications, or low copy repeats (LCR) (Fredman *et al.* 2004, Iafrate *et al.* 2004, Sebat *et al.* 2004, Sharp *et al.* 2005, Tuzun *et al.* 2005, McCarroll *et al.* 2006, Redon *et al.* 2006). The LCR size, sequence identity, and orientation, as well as the distance between two LCRs affect the

probability for miss-alignment during meiosis via nonallelic homologous recombination (NAHR) (Stankiewicz *et al.* 2002, Redon *et al.* 2006). Intrachromosomal segmental duplications are also more likely to be associated with CNVs than interchromosomal segmental duplications (Redon *et al.* 2006). Approximately 5% of the human genome contains LCRs. LCRs span from one to hundreds of kilobases and they share a high (>90%) sequence identity (Bailey *et al.* 2001, Cheung *et al.* 2001, Eichler 2001, Lander *et al.* 2001, International Human Genome Sequencing Consortium 2004). LCRs contain both coding and non-coding sequences and are dispersed into the entire genome, but are enriched to pericentromeric and telomeric regions (Bailey *et al.* 2001). LCRs are known mediators of clinically significant chromosomal rearrangements (Lupski 1998, Ji *et al.* 2000, Stankiewicz *et al.* 2002), including Prader-Willi/Angelman syndrome (PWS/AS) (Amos-Landgraf *et al.* 1999, Christian *et al.* 1999), VCFS (Edelmann *et al.* 1999, Shaikh *et al.* 2000), and Williams-Beuren syndrome (Francke 1999, Peoples *et al.* 2000). Breakpoint analysis of 1000 genomes sequence data suggests that large scale CNVs are primarily formed via NAHR (Mills *et al.* 2011). The majority of smaller scale CNVs are formed by non-homolougs end joining (NHEJ) or microhomology-mediated break-induced replication (MMRBI), while small insertions are attributable for transposition activity (Mills *et al.* 2011).

Several lines of evidence support deletions as a class of variants that are enriched with disruptive mutations, and are under negative selection. A marked portion of detected deletions are rare (Conrad *et al.* 2006, Conrad *et al.* 2009). Furthermore, the deletion size is negatively correlated with its frequency and the majority of common deletions are small (<10kb) (Conrad *et al.* 2006, Locke *et al.* 2006, McCarroll *et al.* 2006). Also the length and frequency of deletions varies in populations of different ages. Even though there are more deletions in older populations (African ancestry) they tend to be shorter compared to younger populations (European ancestry), (Conrad *et al.* 2006, Conrad *et al.* 2009). Also, compared to duplications, deletions are markedly (one-third) shorter (Redon *et al.* 2006). Fitted statistical models suggest that especially exonic deletions are subject to purifying selection (Conrad *et al.* 2009). Consistently, particularly low frequency deletions appear in multiple haplotypes suggesting that the deletion events have occurred recurrently as to the deletions inherited from a singular ancestral event (Conrad *et al.* 2006). High frequency deletions are clustered away from coding regions (Mills *et al.* 2011). Furthermore, compared to duplications, a lower portion of deletions involve genes (Redon *et al.* 2006, Conrad *et al.* 2009). For duplications the evolutionary constraint is understood to a lesser extent. The existence of gene families, and based on genomic comparisons between species, such as humans and great apes (hominoids), it is evident that gene duplications have provided a substrate for the evolution of

new gene functions and have been subject to positive selection pressure (Fortna *et al.* 2004, Hurles 2004).

2.2.4 The extent of linkage disequilibrium

During meiosis chromosomes exchange genetic material. This genetic recombination creates new combinations of alleles in an individual, called haplotypes. When two nearby variants are inherited together more often than what would be expected by chance, they are said to be in linkage disequilibrium (LD). That is, the two co-inherited variants are not segregated independently in a population and LD exists because of the shared ancestry of contemporary chromosomes. In general, the likelihood for recombination increases with the distance between two loci. However, the recombination rate is unevenly distributed in the genome with some regions being more prone for recombination than others, consequently called recombination hotspots (Jeffreys *et al.* 2001). This results in a block like structure of 1-100 kilobase (kb) segments of DNA (termed haplotype blocks or haploblocks) each with limited allelic diversity and high LD separated by recombination hotspots of low LD (Daly *et al.* 2001). Although the genome is fragmented into short LD blocks by recombination, weaker long range LD also exists between the blocks weakening with distance. Within the haplotype block, LD is disrupted primarily by new mutation events, which is why they typically include only few haplotypes (Figure 4) (Daly *et al.* 2001, Patil *et al.* 2001, Reich *et al.* 2001, Gabriel *et al.* 2002).

LD can initially extend over larger genomic distances, but decays gradually over time by recombination. LD depends on demographic history, stochastic events and functional constraint (Hinds *et al.* 2005). The haploblocks are approximately 50% shorter in populations of African ancestry compared to European and Asian populations, and haplotype diversity is greatest in African populations (Reich *et al.* 2001, Gabriel *et al.* 2002, De La Vega *et al.* 2005). However, the block boundaries as well as the underlying haplotypes are largely shared across populations with those populations of African ancestry having more private haplotypes than others (Gabriel *et al.* 2002, De La Vega *et al.* 2005). A direct consequence of LD is that it enables the prediction of an allele of a variant based on the genotype of another without the need to genotype them both. In other words most of the haplotype diversity and genetic variation can be tagged by selecting a group of informative SNP markers, called tag SNPs (Hinds *et al.* 2005). Most of the genomic variation can be captured with common haplotypes and the incomplete correlation with missed variance can be overcome by increasing sample size (Gabriel *et al.* 2002). Two commonly used measures for pairwise LD are squared correlation coefficient between two alleles (r^2)

and D'-value. The D'-value is a normalized measure of linkage disequilibrium (D), it is defined as 1 in the absence of recombination and it declines only due to recombination and recurrent mutation (Lewontin 1964, Daly *et al.* 2001, International HapMap Consortium 2005).
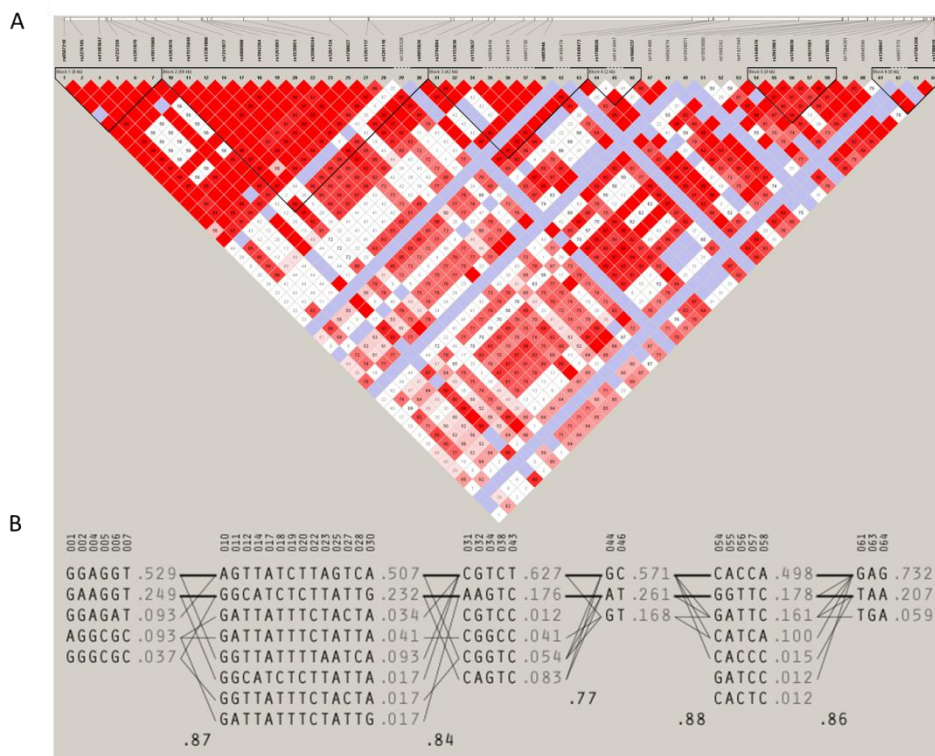
Figure 4. Haplotype block structure (A), and haplotypes (B) for the gene encoding for TCF4 (chr18: 51.05-51.41 Mb, NCBI build 36) in HapMap CEU-individuals, for 64 randomly selected SNPs from HapMap phase II and III, release 28, a view from Haploview software (Barrett *et al.* 2005). A) The relative position for each SNP is presented atop. The LD is colored according to D' estimates and blocks are defined according to that described by Gabriel et al (Gabriel *et al.* 2002). B) Haplotypes with corresponding frequencies are presented for each block. Thick line connections of > 10%, and thin lines connections of > 1%. In the crossing area multiallelic D' is shown.

An effort to characterize the pattern of common sequence variation in the human genome was undertaken by the International HapMap project (http://hapmap.ncbi.nlm.nih.gov/)(International HapMap Consortium 2003). In the first phase of the project approximately one million SNPs were characterized, and their frequencies and the co-inheritance between them were determined in four different populations (Caucasian represented by central Europeans in Utah-CEU;

Asian represented by Han Chinese from Beijing-CHB and Japanese from Tokyo-JPT, and African represented by the Yoruban from Nigeria-YRI) (International HapMap Consortium 2005). In phase II the HapMap project included 3.1million, roughly one third of all common (minor allele frequency, MAF above 0.05) SNPs, improving the original resolution of the map (Frazer *et al.* 2007). In the latest version, Phase III HapMap sample sizes have been increased and additional seven populations were included into the study (International HapMap *et al.* 2010). The HapMap project has provided the scientific community a publically available resource of validated SNPs and with information of the genomic landscape in different populations that can be used as a tool in genetic mapping studies of complex traits.

The extent of LD in a given population depends on the effective population size and the age of the population. Recombination is the major source for decay of LD, thus populations of recent ancestry and small founder population have longer stretches of LD. However, the LD maps are generally very similar between populations (De La Vega *et al.* 2005, Service *et al.* 2006). In a study by Service and coworkers (Service *et al.* 2006) a young subisolate of Kuusamo was found to have the most extensive LD among 11 European descendent isolates studied. The authors concluded that isolate populations could be a benefit to GWAS studies because less assayed SNPs are needed to tag the allelic diversity of the isolate population.

The evolutionary history of common CNVs, sometimes referred to as copy number polymorphisms (CNPs), is similar to that of common SNPs (Conrad *et al.* 2009). Common, generally small scale (< 10 kb) deletions, located outside LCRs often descend from a single ancestral event, are shared between populations, are in LD with nearby SNPs, and exist also in a homozygous state (Hinds *et al.* 2006, Locke *et al.* 2006, McCarroll *et al.* 2006, Redon *et al.* 2006, Mills *et al.* 2011). Furthermore, similarly to SNPs, the extent of LD is lower in African ancestry population compared non-African populations (Redon *et al.* 2006). Even so, the extent of LD is still however lower between CNVs and SNPs than for two neighboring SNPs (Redon *et al.* 2006). Rare CNVs are also less well tagged by SNPs (Conrad *et al.* 2009). Less than 5% of SNPs that associate with complex traits tag a known CNV. Moreover, since the majority of common CNVs can be measured indirectly with tag SNPs, common CNVs are not likely to provide a significant additional contributions to explain the heritability of complex traits other than by SNP association studies (Conrad *et al.* 2009), as will be discussed in the next section.

2.3 Genetic mapping of complex traits

2.3.1 Estimating the contribution of genetic factors in a trait

An obvious cornerstone of genetic analysis is the hypothesis that variance within a given trait is influenced by genetic factors. If a trait has a heritable component, individuals of close kinship, i.e. those who share the largest part of their DNA, resemble each other more closely than individuals with more distant kinship or no familial relation. An example is the genetic risk for a given disorder. If genetic factors do play a role in the vulnerability to a given disroder, close relatives of a patient would be at increased risk and the risk would decrease towards the general risk in a population for the more distant relatives as a function of decrease in genetic sharing. Family studies alone are not, however, sufficient to determine if the familial aggregation is due to shared genetic or shared environmental factors. For dissecting the genetic component of a trait, twin and adoption studies are needed. The difference in the concordance rate between monozygotic (MZ) and dizygotic (DZ) twin pairs provides an estimate of the extent of the genetic component in a trait. Adoption studies on the other hand, enable the estimation of the difference in risk between biological and non-biological siblings of both affected and unaffected parents, which provides an estimate of the relative contributions of shared environmental and genetic factors (Boomsma *et al.* 2002).

2.3.2 Searching for gene variants contributing to complex traits

The allelic spectrum of genetic variants in the human genome ranges from singletons to very common variants with a minor allele-frequency of ~50%. Different frequency bins are likely to be very different with regard to an individual variant's impact on explaining variation in human traits and disorders (Figure 5). Natural selection controls the frequency of mutations that cause severely deleterious diseases, which is why severe Mendelian disorders tend to be caused by rare variants (Altshuler *et al.* 2008). Common variants with strong deleterious effects are unlikely to exist in complex diseases because if they did they would have most likely been identified by now (Manolio *et al.* 2009). On the other hand, common diseases are typically late onset diseases and have low impact on reproductive fitness, which is why variants causing them can increase to higher frequencies (Altshuler *et al.* 2008). Some current disease causing alleles could have been advantageous previously in human history, such as some autoimmune predisposing alleles, or balancing selection that can act on the causative alleles, such in sickle-cell anemia and malaria resistance (Jostins *et al.* 2012). Depending on the frequency bin and assumed impact, different methods and study designs are adopted in genetic studies. While the rare variants are tackled systematically by sequencing, common variants can be efficiently interrogated by genome-wide SNP arrays (GWAS).
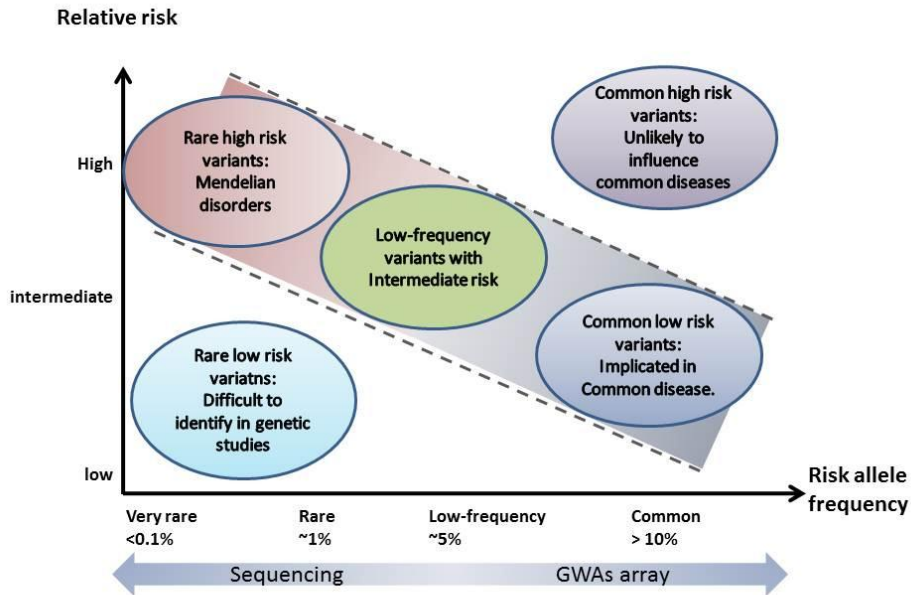
Figure 5. Expected relative risks associated with genetic variants with different allele frequencies. The dashed lines confine the targets of genetic studies (Modified from(McCarthy *et al.* 2008, Manolio *et al.* 2009)

The search for causal variants that influence a trait is termed genetic mapping. In principle genetic mapping can be conducted either directly by testing each potentially causative variant against a trait, or indirectly by relying on LD between nearby variants (Collins *et al.* 1997). A systematic approach that utilizes naturally occurring DNA variations as generic markers to build up a linkage map of the human genome and systematically trace inheritance in families was proposed in the 1980s (Botstein *et al.* 1980). Linkage studies proved early on to be successful in identifying gene loci for Mendelian diseases, such as Huntington's disease (Gusella *et al.* 1983). Currently, thousands of phenotype-causing mutations are known (http://www.ncbi.nlm.nih.gov/omim). Linkage studies have been largely unsuccessful, however, in identifying gene variants for common forms of common traits in humans (Altshuler *et al.* 2008).

Candidate gene driven association studies emerged alongside genome-wide linkage studies. In association studies, rather than studying the transmission in pedigrees,

differences in allele frequencies are compared between cases and controls. If an allele or a genotype has a higher frequency among cases this can be interpreted as evidence that the allele is associated with the trait studied. Association analysis is carried out for sample sets that have been ascertained from a population. For this reason, it requires denser map of genetic markers to interrogate haplotype diversity compared to those used in linkage studies tracing transmission within families. Association studies were limited to small genomic regions and candidate genes for technical reasons for long time. Although there are examples in which the candidate gene approach led to the identification of confirmed associations, most of the hundreds of associations reported are not robust and consistently replicated. The inability to replicate early genetic associations has been affected by weak genetic effects, small sample sizes, and population stratification (Hirschhorn *et al.* 2002). Exceptions include associations, between APOE and Alzheimer's disease (Strittmatter *et al.* 1993), Factor V Leiden and deep vein thrombosis (Bertina *et al.* 1994), and PPARγ and type II diabetes (T2D) (Altshuler *et al.* 2000) as examples.

2.3.3 Genome-wide association studies of complex traits

The majority of common complex disorders have significant genetic components (Lander *et al.* 1994). During the past decade GWAS studies have revolutionized genetic mapping and hundreds of loci for multiple common human disorders and complex traits have been identified (Figure 6), such as for Crohn's disease (Jostins *et al.* 2012), T2D (Morris *et al.* 2012), adult height (Lango Allen *et al.* 2010), and schizophrenia (Schizophrenia Working Group of the Psychiatric Genomics 2014). The grounding hypothesis of GWAS is the common disease common variant hypothesis which postulates that the genetic risk in an individual is attributable to the combination of many high frequency variants, each of which only marginally increase the risk for the disorder (Risch *et al.* 1996). A significant portion of common variation (~10 million SNPs) across populations can be captured by assaying several hundreds of thousands to a few millions of variants and testing them for association with a condition indirectly in a large case-control setting (Gabriel *et al.* 2002, Frazer *et al.* 2007). One of the early findings obtained from GWAS studies was that the effect sizes contributing to the associations are typically small (OR < 1.5) (Altshuler *et al.* 2008). A consequence of this is that large sample sizes are required to compensate for the low effect sizes. Since large numbers of variants are tested and there is a low prior probability for any one marker to be associated, this adds up to a required genome-wide significance threshold of p= 5×10^{-8}. This corresponds to the number of independent loci in the genome.
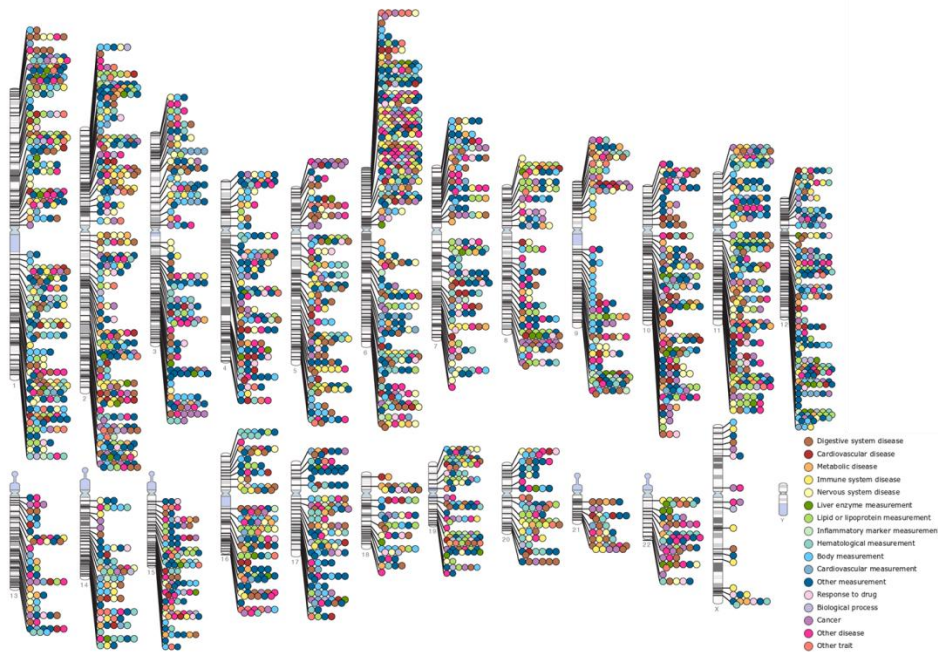
Figure 6. Published Genome-Wide Associations with p ≤ 5×10-8 up to December 2012 for 17 trait categories. (Modified from www.genome.gov/GWAStudies)

Even though multiple genetic risk loci have been identified for various human traits and diseases, a significant portion of the heritable susceptibility still remains to be assigned for specific gene variants. Although the genetic architecture likely varies from trait to trait, much of this so called "missing heritability" is likely to be explained by as yet undiscovered common variants that have only modest effects and which can only be detected in ever increasing sample sizes. It is also plausible that rare variants, gene-gene and gene-environment interactions account for some portion of the undefined genetic risk (McCarthy *et al.* 2008, Manolio *et al.* 2009).

2.3.4 Rare variants in complex diseases

Collectively, rare variants outnumber common variants in the human genome (Genomes Project Consortium *et al.* 2012). Rare variants are enriched with non-neutral functional and deleterious variants supportive of them having a significant role also in explaining the variation of complex diseases (Li *et al.* 2010, Kiezun *et al.* 2012). Analysis of rare variants is hindered exactly because of the rarity of the variants, which limits statistical power (Kiezun *et al.* 2012). Rare sequence variants are associated with high base-line variation of a mutation rate of ~10^{-8} per base per

generation (Genomes Project Consortium *et al.* 2012). The baseline variation is high even if only considering LoF variants. Each individual carries between 60-100 LoF variants, from which 20 in homozygous state (Genomes Project Consortium *et al.* 2012, MacArthur *et al.* 2012). Unlike some structural variants, sequence variants have low recurrence rate, so that any two individuals are unlikely to harbor exactly the same *de novo* sequence variant. Prior to the emergence of the new sequencing technologies, systematic genome-wide analysis of rare sequence variants has been impossible. During the past years, genome-wide sequencing studies have focused on the exome, hypothesized to tackle the highest rate of functional variants with the cost of sequencing. Whole exome sequencing (WES) is used particularly to identify rare coding variants that are not tagged by SNP array based GWAS studies (Teer *et al.* 2010). Compared to the sample sizes used in GWAS studies, the sample sizes in WES studies have still been small, which is likely to be one of the reasons why these studies have not yet reached their full potential in revealing functional variants. However, in Mendelian diseases WES studies have already proven to be an efficient approach (Kiezun *et al.* 2012).

2.4 Population structure in genetic studies of complex traits

A systematic ancestry difference between cases and controls, called population stratification, is a well characterized confounder of association analysis. Population stratification can easily remain unnoticed in candidate gene studies where only handful of SNPs are tested. Ancestry differences between cases and controls in GWAS can be detected as inflation of the expected distribution of test statistics and be appropriately handled as part of the analysis (Price *et al.* 2010). Population differences, however, can also be used to advantage in genetic studies. Although population differences for common variants have proven to be of little benefit in genetic studies, the situation for rare variants is different and utilization of special populations and consanguineous families has led to the identification of multiple causative variants for Mendelian disorders.

2.4.1 Utilizing founder populations in disease mapping

A significant reduction in population sizes in genetic terms referred to as a population bottleneck, followed by a rapid population growth reduces allelic diversity in the founder population and allows genetic drift to act on the remaining genetic variants. This leads to the relative enrichment of some alleles at the cost of losing others (Figure 7). These genealogical events can have far-reaching consequences arising as local difference in disease prevalence and the genetic makeup underlying them (Peltonen *et al.* 2000). Rare variants are more prone to

stochastic alteration in their allele frequencies and are thus a major substrate for genetic drift in founder populations.
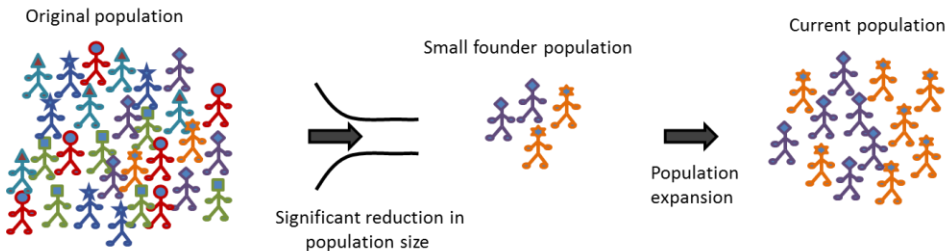

Figure 7. Schematic presentation of a population bottleneck and genetic drift shaping the genetic makeup of a founder population. During the bottleneck and the following population expansion the genetic diversity of the founder population declines and some variants become enriched in the subsequent founder population.

Population isolates have been useful in identifying genetic causes for rare Mendelian diseases, such as the diseases of Finnish disease heritage - a constellation of 36 rare, mostly autosomally recessive disorders that are more common in Finland than elsewhere in the world (http://www.findis.org/) (Norio *et al.* 1973, Peltonen *et al.* 1999). The isolate design has also been a benefit to the identification of genetic variants of more complex traits, such as lactose intolerance, which shows significant geographical variation in prevalence likely due to positive selection (Enattah *et al.* 2002, Enattah *et al.* 2007). Other Mendelian variants of complex diseases identified in other isolate populations include Hirschsprung disease in the Amish and nonsyndromic hearing loss in the Bedouin populations (Puffenberger *et al.* 1994, Scott *et al.* 1995). One of the most impressive examples of the use of isolate populations in genetic studies is from Iceland, where the resourceful use of known genealogy, extended pedigrees and detailed medical data has been successfully used to identify a plethora of genetic variants to complex traits, such as pigmentation, myocardial infarction, T2D, and many others (Helgadottir *et al.* 2004, Grant *et al.* 2006, Sulem *et al.* 2007). An extreme example of simplistic genetic architecture of complex traits in isolate populations is the identification of nonsynonymous mutation in Exon 2 of *TYRP1* gene. The variant explains 46% of variation in blond hair color among Solomon islanders, whereas the mutation is absent in other populations (Kenny *et al.* 2012). A GWAS of amyotrophic lateral sclerosis (ALS), carried out in Finland required 10% of the number of familial cases in Finland compared to a study conducted in sporadic patients from United Kingdom (Laaksovirta *et al.* 2010, Shatunov *et al.* 2010). Indeed the success here has been mostly driven by the tendency of the affected individuals to carry the same allele

and also the relative simplicity in the genetic makeup underlying the traits among the specific populations.

2.4.2 Finland: a north eastern isolate in Europe

Finland has been inhabited since the end of the last glacial age around 10000 years ago. During the prehistoric era only a few thousand hunter gatherers lived on the Finnish peninsula. According to archeological, linguistic, and genetic evidence there was an influx of small immigrant groups mainly from south and east, but also from the west during this time, 6000-4000 years ago. Based on studies on Y-chromosome lineages, a dual origin of Finns has been hypothesized to result from migration from east and south 4000 and 2000 years ago, respectively. However, all migrations can be considered to have mixed with the native Finnish population that had been inhabiting the area since the continental ice sheet melted (Varilo 1999, Norio 2003).

The founding of the current Finnish population has continued for thousands of years. The first inhabitants settled in the coastal region (south and west), consequently termed the "early settlement" and for centuries only this part of the land remained populated. Starting in the 16th century, the inhabitation of the inland known as the "Late Settlement" of Finland began from the South Savo region (Figure 8). At this time the population in Finland was 250 000 inhabitants. Both the pressure to cultivate more land and political decisions including broadening the country's borders for economic purposes and tax benefits drove the migration movement. By the end of the 17th century the inhabited area of Finland had doubled and most of the land had become inhabited albeit sparsely. Subsequent genetic isolation of the Finnish population followed due to famine and epidemics. During the 1696 and 1698 period, famine and following epidemics killed one third of the population of 400 000 at that time. During the last 300 years the Finnish population has grown from 250 000 to the 5.5 million inhabitants today (Norio *et al.* 1973, Varilo 1999, Norio 2003).
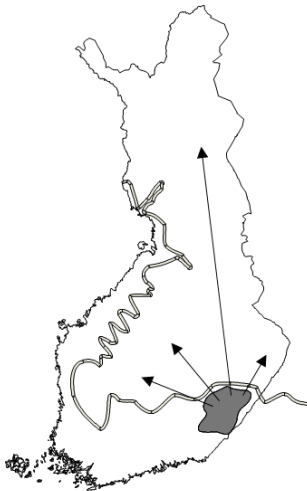
Figure 8. The internal migration from South-Savo (grey shaded area) towards the western, central and north of Finland (Teppo Varilo).

2.4.3 Population genetics of Finland

Significant effort has been invested into studies of the genetic origins of Finns. The uniparentally inherited mitochondrial and Y-chromosomal markers are useful for population genetics due to their special modes of inheritance. The Y-chromosome is inherited paternally as a haploid chromosome. It undergoes recombination only in the telomeric parts, which is why new haplotypes are introduced to it only as a consequence of new mutations. Mitochondrial DNA (mtDNA) is inherited maternally. It has a higher rate of mutation than autosomes and it does not undergo recombination.

The paternally inherited Y-chromosomes show reduced genetic diversity in Finns compared to rest of Europe consistent with a tight population bottleneck followed by rapid population expansion (Sajantila *et al.* 1996, Kittles *et al.* 1998, Lahermo *et al.* 1999, Palo *et al.* 2009). There are two major Y-chromosome lineages among Finns, one of eastern and the other of western (Scandinavian) origin (Kittles *et al.* 1998, Lahermo *et al.* 1999, Lappalainen *et al.* 2006). Similar reduction in genetic diversity is not seen for mtDNA as for Y-chromosome in Finns (Lahermo *et al.* 1996, Sajantila *et al.* 1996, Hedman *et al.* 2007, Palo *et al.* 2009). Earlier, the higher mutation rate of mtDNA has been considered to cause these differences between Y-chromosomal and mtDNA lineages (Sajantila *et al.* 1996). However, since the mitochondrial markers used represent old stable mutations, more likely explanation seems to be a more recent male specific gene flow from Scandinavia to the west-

coast of Finland (Lappalainen *et al.* 2006, Palo *et al.* 2009). Overall, both mtDNA and Y-chromosomal data suggest European ancestry of the Finnish population. In addition, the Y-chromosomal gene pool shows significant similarities with more Eastern populations - a tendency that is not shared with the mtDNA, which is indistinguishable from other European populations (Sajantila *et al.* 1995, Lahermo *et al.* 1996, Kittles *et al.* 1998, Lahermo *et al.* 1999, Finnila *et al.* 2001).

The hallmarks of the recent founding bottleneck also manifest in the autosomal genetic fingerprint of Finns. In general Finns have longer overall LD compared to rest of Europe (Varilo *et al.* 2003, Service *et al.* 2006).There is an enrichment of rare and low frequency variants in Finland. Moreover, the haplotypes that carry these variants are on average longer among Finns than among other European populations, all of which is suggestive of a recent shared ancestry among Finns(Genomes Project Consortium *et al.* 2012). The multiple consecutive bottlenecks have resulted in the reduction of genetic diversity in Finland compared to the rest of Europe. Despite this phenomenom, Finns are not genetically homogenous, but internal migration has also caused significant genetic substructure within the country. The genetic variation of the Y-chromosome in the Finnish population demonstrates significant east-west differences, supporting that there has been a later Scandinavian gene flow to the population of Western-Finland, but not into Eastern-Finland (Lappalainen *et al.* 2006, Palo *et al.* 2009). Autosomal SNP allele information also demonstrates clear east-west and north-south gradients, as well as regional genetic differences within the country (Jakkula *et al.* 2008). Illustrative of the uniqueness of the genetic makeup of Finland is the Finnish disease heritage diseases that are more common in Finland than anywhere else in the world. The majority of the diseases are autosomally recessive and one major mutation is identified for each of them. Depending on the age of the founder mutation, some of the diseases are scattered around the country whereas others demonstrate regional clustering of patients (Peltonen *et al.* 1999). Similarly, risk for some more complex disorders, including schizophrenia and intellectual disability, varies regionally in Finland following the pattern of internal migration (Hovatta *et al.* 1997, Haukka *et al.* 2001, Perala *et al.* 2008).

2.4.4 The North-Eastern internal isolate of Kuusamo

Kuusamo is located in the north-eastern border area of Finland. It is the geographical region of one of the most studied internal genetic isolates in Finland. The area of Kuusamo was inhabited by Finns towards the end of the internal migration wave in the 17th century. The first Finnish settler moved to Kuusamo in 1676 and was followed by 34 families of 194 individuals mainly from Ostrobothnia and South

Kainuu in 1685. The populations of both regions had previously descended from earlier migration from South Savo. Up to 60 to 70 families lived in Kuusamo by 1687 and during the great famine in 1696-1698 half of the Finnish population and nearly all of the native Saami inhabitants died. Since then the Saami people have gradually disappeared by emigration and by assimilation into the agricultural Finnish population. In 1718, 615 individuals inhabited Kuusamo. The population grew rapidly and reached 2000 individuals in 1760 and it had further doubled by 1842, it eventually reached 10 000 individuals by 1910. The current population of Kuusamo is roughly 16 000 inhabitants. The population of Kuusamo remained almost completely genetically isolated until the Second World War (Varilo 1999). Consistent with its population history, Kuusamo demonstrates even longer LD compared to other regions of Finland (Varilo *et al.* 2000, Jakkula *et al.* 2008). The recent genealogical history of Kuusamo (Figure 9) is such that the majority of patients with schizophrenia in that region are part of one extended pedigree (Hovatta *et al.* 1997, Hovatta *et al.* 1999).
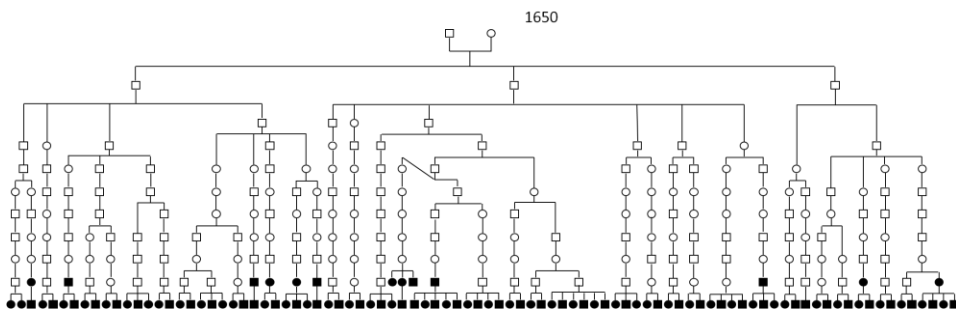
Figure 9.   Extended pedigrees in the internal isolate of Kuusamo links schizophrenia patients in the modern population (Courtesy of Teppo Varilo).

2.5 Genetic architecture of schizophrenia and related mental disorders

Both common and rare genetic variants contribute to the genetic risk of schizophrenia (Figure 10). Schizophrenia has been demonstrated to have an indisputable polygenic component that consists mostly of common low impact variants that make up the majority of the population genetic risk (Purcell *et al.* 2009, Ripke *et al.* 2013). In addition to common variants, rare high impact variants are also important in the etiology of schizophrenia. Although the rare variants have low attributable risk at the population level, individually they can confer substantial risks. Suggested already by epidemiological studies, the genetic evidence strongly

supports shared genetic risk across psychiatric and developmental disorders, suggesting common biological grounds for these diagnostically different disease entities (Sullivan *et al.* 2012).
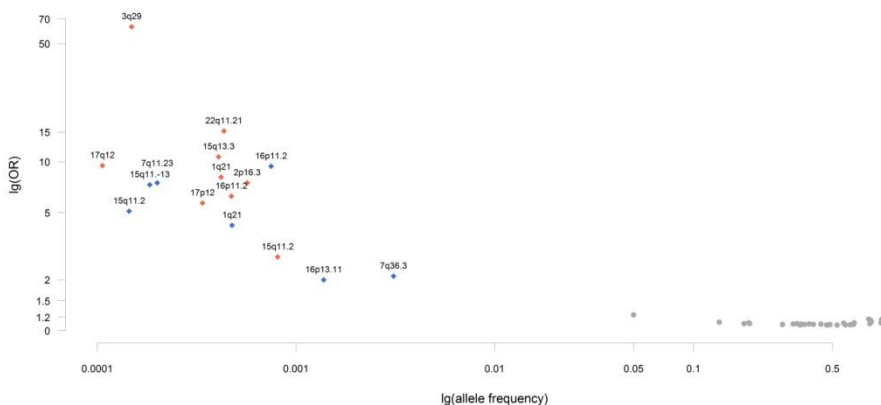
Figure 10. The allelic spectrum of schizophrenia. Chromosomal deletions shown in red, duplications in blue and SNPs in grey. The x-axis the $\log_{10}$ allele frequency and y-axis is the $\log_{10}$ of the Odds Ratio (OR). The values are taken from.(Stefansson *et al.* 2008, Ingason *et al.* 2011, Levinson *et al.* 2011, Ripke *et al.* 2011, Malhotra *et al.* 2012, Guha *et al.* 2013, Ripke *et al.* 2013)

2.5.1 Common genetic variants in schizophrenia

Schizophrenia is a highly polygenic disorder, with thousands of genetic variants contributing to the risk, each variant having only minor effects (Purcell *et al.* 2009, Ripke *et al.* 2013). The cumulative common genetic variation is estimated to account for over 30% of the variance in the liability (Ripke *et al.* 2013). However, the majority of the risk accounted for by genetic variants still remains to be assigned to specific gene variants.

As in the case of many other complex diseases, genetic studies that used family based linkage and candidate gene approaches were largely unsuccessful in identifying gene variants underlying schizophrenia (Lewis *et al.* 2003, Allen *et al.* 2008, Ng *et al.* 2009). In a meta-analysis of over 1000 independent association studies in schizophrenia, 16 genes (*APOE, COMT, DAO, DRD1, DRD2, DRD4, DTNBP1, GABRB2, GRIN2B, HP, IL1B, MTHFR, PLXNA2, SLC6A4, TP53, TPH1*) showed suggestive evidence of association with average odds ratio of 1.23 (Allen *et al.* 2008). Of these 16 genes, *DRD2* is the main target for antipsychotic medication and has been confirmed to convey a risk for schizophrenia in a recent large scale

GWAS study. (Schizophrenia Working Group of the Psychiatric Genomics 2014). The genetic risk for schizophrenia started to emerge with the advent of SNP array technology and large scale case-control GWAS studies. The first GWAS studies were published in 2009 and revealed a substantial association with the MHC region on chromosome 6 (Purcell *et al.* 2009, Shi *et al.* 2009, Stefansson *et al.* 2009). These studies were followed by two larger meta-analyses with the latter combining 13 833 cases and 18 310 controls and identified 22 distinct loci for schizophrenia. Of these, 13 associations were novel and nine had reached genome-wide significance in previous studies (Purcell *et al.* 2009, Shi *et al.* 2009, Stefansson *et al.* 2009, Ripke *et al.* 2011). Of the 22 associated loci, six (6p21.333-p21.32, 1p31.1, 5q33.1-q33.2, 2q22.3, 2q33.1, 3p22.2) overlapped with regions that had previously been reported to show nominal evidence for linkage (Lewis *et al.* 2003, Ng *et al.* 2009). However, the linkage regions spanned tens of megabases and the statistical evidence has been at best only suggestive, thus the evidence prior GWAS for these regions and schizophrenia were inconclusive to say the least. In addition, none of the assigned probable candidate genes in the 22 regions had been implicated in schizophrenia in candidate gene studies prior to December 2011 (http://www.szgene.org/) with the exception of *CACNAC1* (Green *et al.* 2010, Nyegaard *et al.* 2010). *CACNAC1* was initially implicated in bipolar disorder (Sklar *et al.* 2008), which also triggered its study in other psychiatric disorders including schizophrenia and depression (Green *et al.* 2010, Nyegaard *et al.* 2010). Previously, suggestive association between *ZNF804A*, located several hundred kb distal to 2q32.1, and schizophrenia, cognitive performance in schizophrenia paitents, in addition to bipolar disorder had also been reported (O'Donovan *et al.* 2008, Riley *et al.* 2010, Steinberg *et al.* 2011). Several studies have also provided suggestive evidence for *TCF4* on 18q21.2 and *NRGN*, located 700 kb away from the GWAS association on 11q24.2, for schizophrenia prior to the latest GWAS studies (Ruano *et al.* 2008, Stefansson *et al.* 2009, Li *et al.* 2010).

The largest GWAS conducted for schizophrenia to date included 37 000 cases and 113 000 controls and revealed 128 distinct associations in 108 genetic loci for schizophrenia: 83 of which were novel. The large number of associations enabled for the first time convincingly to speculate the biological pathways involved in the etiology of schizophrenia. First, the association is not randomly distributed across genes, but is enriched in those genes that are expressed in the brain and neurons, particularly those of the cortical and striatal neuronal lineages, in addition to immunological genes. Moreover, these associations included genes related to dopaminergic and glutaminergic neurotransmission and synaptic plasticity, systems that had previously been implicated in schizophrenia by pharmacological studies. The latest GWAS results are consistent with and extend the previous GWAS

findings of genes encoding for voltage gated calcium channel subunit. As with other complex traits, the major associations do not seem to affect amino-acid sequences encoded by the genes but rather highlight regulatory mechanisms. The multiple loci included gene variants with regulatory roles in gene expression, as defined by the expression quantitative trait loci (eQTL) and enhancers that are active in the brain and in tissues with immune functions (Schizophrenia Working Group of the Psychiatric Genomics 2014).

2.5.2 Rare and *de novo* sequence variants in schizophrenia

Similar to that found for common variants, the cumulative polygenic burden of rare variants (i.e. variants with allele frequency of < 0.0001) influences the genetic susceptibility to schizophrenia. The contribution of the rare variants to schizophrenia remains, however, still smaller on a population level compared to the susceptibility explained by common genetic variants (Purcell *et al.* 2014). Although the studies of rare sequence variants in schizophrenia have not unequivocally pinpointed specific risk variants, a burden of rare variants in functional gene groups including voltage gated calcium ion channels, activity regulated cytoskeleton associated scaffold protein (ARC) of postsynaptic density, and targets of Fragile X Mental Retardation Protein (FMRP) are enriched with rare disruptive mutations in schizophrenia patients (Purcell *et al.* 2014). Moreover, the schizophrenia GWAS results overlap with rare CNVs and *de novo* nonsynonymous mutations that have been identified in schizophrenia, autism, and intellectual disability (Schizophrenia Working Group of the Psychiatric Genomics 2014).

The hypothesis of significant impact of rare variants in mental disorders is well justified. Epidemiological studies suggest increased mortality and reduced fecundity in severe mental disorders that raise an evolutionary paradox of a heritable genetic predisposition for the disorders (Nordentoft *et al.* 2013, Power *et al.* 2013). Despite high heritability estimates, a significant portion of schizophrenia patients are sporadic and do not have a family history of mental disorders (Mortensen *et al.* 1999). The correlation of paternal age with risks for mental disorders (D'Onofrio *et al.* 2014, McGrath *et al.* 2014), and increased mutation rates (Kong *et al.* 2012), in addition to the importance of recurrent and *de novo* mutations in early onset mental disorders (autism and intellectual disability) gave rise to the *de novo* hypothesis in mental disorders. *De novo* mutations seem especially important in severe intellectual disability, where patients constantly show higher rates of *de novo* LoF mutations compared to controls and these mutations cluster in known intellectual disability causing genes (Vissers *et al.* 2010, Rauch *et al.* 2012, Gilissen *et al.* 2014). In addition, the somewhat disappointing results obtained from the first GWAS studies

on schizophrenia led to even more enthusiasm for the hypothesis of the importance of rare variants in the etiology of schizophrenia.

Indeed, some studies have reported elevated mutation rates in schizophrenia (2.59×10^{-8} per nucleotide per generation) (Girard *et al.* 2011) and in autism (2.17×10^{-8} per nucleotide per generation) (O'Roak *et al.* 2012). However, collectively, *de novo* sequence variants seem less significant for schizophrenia compared to that of intellectual disability or even autism (Fromer *et al.* 2014). Patients with schizophrenia have been reported to have slightly more nonsynonymous mutations compared to controls. This enrichment has been reported to be biased towards fetal expressed genes that are also found to carry excess *de novo* mutation in autism (Xu *et al.* 2011, Xu *et al.* 2012). These findings are not, however, supported across all studies (Fromer *et al.* 2014). The strongest impact of *de novo* LoF mutations seems to be among schizophrenia patients with poor school performance (Fromer *et al.* 2014). The *de novo* nonsynonymous mutations are enriched in synaptic genes, specifically in genes associated with N-methyl-D-aspartate receptor (NMDAR) complex and the ARC complex. In addition, FMRP targets are reported to be enriched for *de novo* nonsynonymous mutations (Fromer *et al.* 2014).

The analysis of rare variants is statistically challenging due to their rarity which puts severe limitations on statistical power. The sample sizes of current sequencing studies have been one-tenth to one-hundredth compared to those of GWAS (and CNV studies). The human genome has a high background rate of rare variation with a mutation rate of 10^{-8} per nucleotide per generation, corresponding to over 60 new mutations. This would translate into on average 0.86 amino-acid altering mutations per generation in the general population (Lynch 2010). Even variations that are predicted to be protein disruptive are abundant in the general "healthy" population (MacArthur *et al.* 2012). This is reflected in the inability to identify specific rare sequence variants at an appropriate statistical threshold. The largest study of *de novo* mutations in schizophrenia to date (623 trios) reported two *de novo* LoF mutations in the gene that encodes for *TAF13*, suggesting this as a candidate gene for schizophrenia (Fromer *et al.* 2014).

Patients with autism seem to have more potentially disruptive *de novo* nonsynonymous mutations similarly to schizophrenia patients (Iossifov *et al.* 2012, Neale *et al.* 2012, Sanders *et al.* 2012). Recurrent mutations for autism in *CHD8*, *KATNAL2* and *SCN2A* identified in WES studies suggest these genes are genuine risk genes for autism (Neale *et al.* 2012, O'Roak *et al.* 2012, Sanders *et al.* 2012). Emerging evidence for the enrichment of disruptive *de novo* mutations support the

role of brain expressed, synaptic genes and FMRP target mRNAs in autism (Iossifov *et al.* 2012, Neale *et al.* 2012, O'Roak *et al.* 2012, Sanders *et al.* 2012). Furthermore, among autism patients rare mutations in genes related to NMDAR complex, including S*HANK2*, *SYNGAP1*, and *DLGAP2*, or mutations in *SHANK3*, *NLGN3*, and *NLGN4X,* indicate shared biological pathways between autism and schizophrenia (Jamain *et al.* 2003, Durand *et al.* 2007, Berkel *et al.* 2010, Pinto *et al.* 2010).

2.5.3 Rare CNVs in schizophrenia

Structural variation has long been acknowledged as a risk factor for psychiatric illnesses including schizophrenia. Even to date, one of the highest known singular risk factors for schizophrenia is the VCFS, which was described over 20 years ago. Approximately one-third of VCFS patients develop a psychotic disorder and over 20% fulfill diagnostic criteria for schizophrenia (Pulver *et al.* 1994, Murphy *et al.* 1999). The systematic genome-wide assessment of microscopic and submicroscopic CNVs has become possible with the development of array based methods involving both comparative genomic hybridization (CGH) and oligonucleotide SNP arrays (Pinkel *et al.* 1998, Bignell *et al.* 2004). Both methods were used early on for studies on intellectual disability and autism due to the strong *a priori* hypothesis of the prominent role of CNVs in these conditions (Vissers *et al.* 2003, Friedman *et al.* 2006, Marshall *et al.* 2008). Cumulative evidence from the first CNV scans in schizophrenia suggested that there is an increased burden of rare (<1%) large CNVs in schizophrenia, although the magnitudes varied extensively depending on the study (International Schizophrenia Consortium 2008, Walsh *et al.* 2008, Xu *et al.* 2008, Kirov *et al.* 2009). Although these studies indicated a handful of individual plausible risk CNVs including deletions on 22q11.2, and *NRXN1*, they were unable to provide required statistical evidence for involvement of specific variants. Only two of the first phase CNV scans, including that carried out in study I of this thesis, had the statistical power to identify individual risk deletions at 1p21.2, 15q11.2, 15q11.3, and 22q11.2 with confidence (International Schizophrenia Consortium 2008, Stefansson *et al.* 2008).

To date, at least 15 CNVs have been repeatedly associated with increased risk for schizophrenia by independent studies in different populations. The CNVs include roughly equal number of duplications and deletions (Table 2). Combined, over 2.5% of schizophrenia patients carry these CNVs though their frequency in controls is only around 0.5% (Stefansson *et al.* 2008, Ingason *et al.* 2011, Levinson *et al.* 2011, Malhotra *et al.* 2012, Guha *et al.* 2013). As a consequence of the reduced fecundity associated with severe mental disorders including schizophrenia, CNVs associated

with large effect sizes are expected to rapidly disappear from the population. Accordingly, especially the sporadic cases of both schizophrenia and autism have an elevated rate of *de novo* CNVs and many of the reported CNVs represent *de novo* mutations from recurrent CNV events (Sebat *et al.* 2007, Stefansson *et al.* 2008, Xu *et al.* 2008, Kirov *et al.* 2012). Overall, the mutation rate for the CNVs associated with schizophrenia is high affecting one in a few thousand to one in several tens of thousands individuals. However, negative selection pressure operates on the recurrent CNVs in such a way that following the mutation event each mutation exists only for only a few generations in the population (Rees *et al.* 2011).

Besides the large sample sizes, the analysis of rare large-scale CNVs has benefitted from the low background rate of large scale structural variation in general populations. In general populations 11% of individuals are reported to carry >500kb CNVs (Pietilainen *et al.* 2011). The variations that have been identified in schizophrenia have been associated with very high risks (OR: 2.5-60.0) and they have been mostly recurrent with biased mutation events in specific regions of the genome. This has enabled multiple replications of the findings across populations in sample sets of tens of thousands individuals (Levinson *et al.* 2011, Malhotra *et al.* 2012, Kirov *et al.* 2014). For rare sequence variations, similar recurrence for a single mutation at a given site is less probable. Much more variation is expected for sequence variants compared to large CNVs as basically any of the 6.5 billion bases can mutate. Even if the microhomology of the breakpoint were to vary from individual to individual, the main chunk of the CNV remains usually the same.

As with *de novo* sequence variants, schizophrenia linked CNVs are enriched with genes encoding for postsynaptic proteins and genes related to postsynaptic density. Specifically NMDAR and ARC postsynaptic signaling complexes are enriched among the CNVs associated with schizophrenia. These complexes are important for synaptic plasticity and in cognition, and their functions are also now linked to schizophrenia (Kirov *et al.* 2012).

Table 2.   Rare CNVs associated with schizophrenia

Locus	type	start (Mb)	end (Mb)	genes	OR	Association to Aut/ID	Assocaition to BPD	OMIM #
1q21	del	146.57	147.39	multiple	8.1	yes		612474
1q21	dup	146.57	147.39	multiple	4.2	yes	yes	612475
2p16.3	del	50.14	51.26	NRXN1	7.5	yes		614332
3q29	del	195.73	197.34	multiple	63.0	yes	yes	609425
7q11.23	dup	72.74	74.14	multiple	7.5	yes		609757
7q36.3	dup	158.8	158.9	VIPR2	2.1	yes		613959
15q11.2	del	22.8	23.09	multiple	2.73	yes		615656
15q11.2-13.1	dup	20.8	26.2	multiple	5.1	yes		176270
15q13.3	del	30.9	33.5	multiple	10.7	yes		612001
16p13.11	dup	15.4	16.3	multiple	2	yes		na
16p11.2	distal del	28.82	29.05	multiple	6.25	yes		611913
16p11.2	dup	29.5	30.2	multiple	9.4	yes	yes	614671
17q12 del	del	34.81	36.2	multiple	9.5	yes		614527
17p12/HNPP	del	14.0	15.4	multiple	5.7	no		162500
22q11.21	del	18.7	21.8	multiple	15.21	yes	yes	192430

Source: (Stefansson *et al.* 2008, Levinson *et al.* 2011, Malhotra *et al.* 2012, Guha *et al.* 2013)

Since most of the CNVs associated with schizophrenia are large and span multiple genes, it is difficult to make a clear inference about the pathological mechanism related to them. The clearest exception here is the 2p16.3 deletion that involves only the *NRXN1* gene (Kirov *et al.* 2008, Rujescu *et al.* 2008, Walsh *et al.* 2008). The NRXN1 protein is a synaptic cell adhesion molecule that is found presynaptically. It is thought to interact with postsynaptic neuroligin together with other NRXNs to connect presynaptic and postsynaptic neurons at synapses in excitatory and inhibitory neurons (Sudhof 2008). Deletions and disruptive sequence variants in NRXN1 have also been linked with autism and intellectual disability (Feng *et al.* 2006, Friedman *et al.* 2006, Autism Genome Project *et al.* 2007, Kim *et al.* 2008, Marshall *et al.* 2008). In addition, mutations in genes encoding for neuroligins are associated with autism (Thomas *et al.* 1999, Jamain *et al.* 2003). Another example of a single gene involving CNV is the duplication in 7q36.3 that involves the *VIPR2* gene (Vacic *et al.* 2011). *VIPR2* encodes for vasoactive intestinal peptide receptor that serves as a neuropeptide receptor. It is a widely expressed G-coupled receptor that is believed to have a role in regulating synaptic transmission (Cunha-Reis *et al.* 2005, Yang *et al.* 2009).

The mechanisms leading to pleiotropic effects observed in genetic mutations associated with mental disorders are yet to be discovered. For example mutations in neuronal adhesion molecules neurexins and neuroligins are linked with schizophrenia, autism, Tourette's syndrome and learning disability and the phenotypes can vary even within families. It may be the case that cognitive diseases, including schizophrenia, arise from subtle changes in a subset of synapses in a neural circuit in comparison with general impairment of all synapses in all circuits. A consequence of this would be that the same molecular alteration could produce different circuit changes, which would result in different ranges of neurological symptoms that are then classified as different disorders (Sudhof 2008). To date no monogenic form of schizophrenia has been identified (Sullivan 2012). Even the rare high penetrant variations identified in schizophrenia are associated with variable phenotype and are usually more strongly associated with cognitive performance, autism, and intellectual disability (Kirov *et al.* 2014, Stefansson *et al.* 2014). Severe disrupting mutations in genes such as *CACNA1C* and *TCF4* with common variation linked with schizophrenia do not actually cause schizophrenia, but instead syndromes such as Timothy syndrome and Pitt-Hopkins syndrome, respectively (Splawski *et al.* 2004, Amiel *et al.* 2007).

The genetic evidence for schizophrenia clearly supports a highly polygenic model for schizophrenia. It may be that the risk for schizophrenia would be mediated

through a pathway rather than direct aggregation of risk variants (Sullivan 2012). This hypothesis is supported by 1) the evident polygenicity of the disorder; 2) the emerging evidence from genetic studies that suggest the existence of specific pathways, such as MIR-137 regulated network; and 3) the lack of evidence for Mendelian forms of schizophrenia. Most high impact variants associated with schizophrenia are associated with a heterogeneous phenotype, with a cognitive domain being the main driver of the associations. Moreover, severe protein destroying mutations in genes, for which common variants are associated with schizophrenia do not result in a schizophrenia diagnosis but rather in intellectual disability. For many other complex diseases such as Alzheimer's or T2D, Mendelian forms exist (Sullivan 2012).

2.5.4 Genetic sharing across mental disorder groups

Epidemiological analyses of family, adoption, and twin studies suggest that the diagnostically distinct mental disorders, such as bipolar disorder, major depressive disorder, autism, and intellectual disability to co-exist in same families (Mortensen *et al.* 2009). Since these psychiatric illnesses are highly heritable, this coexistence is thought to be at least partially due to shared genetic risk factors. The evidence for a shared genetic component is particularly strong for schizophrenia and bipolar disorder. These two disorders have a high genetic correlation as estimated by common SNPs (68%) and occurrence within families (63%) (Lichtenstein *et al.* 2009, Purcell *et al.* 2009, Cross-Disorder Group of the Psychiatric Genomics *et al.* 2013). Moreover, this genetic risk has started to unfold with the discovery of specific gene variants such as the individual and combined associations with *CACNAC1*, *NCAN*, *ANK3*, *DGKH*, *ITIH3-ITIH4* for both schizophrenia and bipolar disorder (Ferreira *et al.* 2008, Sklar *et al.* 2008, Cichon *et al.* 2011, Psychiatric 2011, Ripke *et al.* 2011, Zeng *et al.* 2011, Muhleisen *et al.* 2012, Ripke *et al.* 2013). *CACNAC1* is also independently associated with major depressive disorder (Green *et al.* 2010). A significant genetic correlation calculated from common SNPs exists also between schizophrenia and major depressive disorder (43%) and between autism spectrum disorders (16%), as well as between bipolar disorder and major depressive disorder (47%), and ADHD and major depressive disorder (32%) (Cross-Disorder Group of the Psychiatric Genomics *et al.* 2013). A combined association analysis with the five disorders revealed shared genomic loci that probably involve genes encoding for *ITIH3*, *AS3MT*, *CACNAC1*, and *CACNB2*, suggesting that some of the independent associations seen previously for schizophrenia are also shared with the other disorders (Cross-Disorder Group of the Psychiatric Genomics *et al.* 2013). The sharing of rare variants across diagnostic boundaries seems to be more common in schizophrenia and the early onset disorders according to evidence

obtained from structural, and *de novo* sequence variants (Cooper *et al.* 2011, Malhotra *et al.* 2012, Xu *et al.* 2012, Fromer *et al.* 2014, Kirov *et al.* 2014). Rare sequence variants, include functional entities that are seen also in studies of *de novo* variants, CNVs, and common variants (Purcell *et al.* 2014).

3 Aims of the study

The aim of this thesis was to investigate the role of large scale CNVs in schizophrenia and other related neurodevelopmental traits. The aim was to illuminate the genetic relationship between disorders that are traditionally considered different entities. We focused on rare CNVs, with particular emphasis on large scale deletions for three reasons. First, at the time of conducting these studies it was feasible to identify such variants in the existing genotype data obtained from genome-wide association studies that had been conducted in schizophrenia case-control samples and in Finnish population and family cohorts. Second, large genomic variations, particularly deletions, are predicted to have substantial effect on phenotypes. It was therefore plausible to detect association with the data sets available for the studies. Third, the rare disruptive mutations fit the evolutionary paradigm of severe mental disorders there being a role of recurrent structural variation to developmental disorders that has been known for decades.

The specific aims of this study were

I.   To investigate if large recurrent deletions predispose to schizophrenia (Study I).

II.   To study the global impact of large (>500 kb) CNVs on a population level for seven traits hypothesized to reflect abnormal neuronal development (Study II).

III.   To utilize young population isolate in identifying ancestral genomic deletions that would predispose to schizophrenia and related neurodevelopmental phenotypes (Study III).

4 Materials and methods

4.1 Study subjects

The Finnish schizophrenia patients were identified from a Finnish nationwide study sample of families with schizophrenia, from birth cohorts and from population-based health surveys. Finnish population and control samples were ascertained from nationwide collections, including participants to prospective birth cohorts and population-based health surveys. The cohorts have detailed clinical phenotype information collected from Hospital Discharge Register, questionnaires, and clinical examinations available. The clinical data were utilized in studies II and III. The international samples included schizophrenia patients and controls combined in the international schizophrenia consortia, including, SGENE (Stefansson *et al.* 2009), the International Schizophrenia Consortium (ISC) (Stone *et al.* 2008), and the Swedish Schizophrenia Consortium (SSC) (Bergen *et al.* 2012). All of the multicenter studies have been reviewed and approved by their respective appropriate local ethical committees and all participants had signed a written informed consent. The study subjects and the GWAS arrays used are listed in table 4 in the method section.

4.1.1 Finnish study samples

FINNISH SCHIZOPHRENIA FAMILY SAMPLE (FSFS)

Individuals from the FSFS were included in studies I and III. The schizophrenia family sample collection was initiated in 1988 by the National Public Health Institute (since 2009 the National Institute for Health and Welfare (THL)). Three national registers: the Hospital Discharge Register, the Pension Register, and the Populations Register were used to identify all individuals born between 1940 and 1976, and who had been hospitalized, had received disability pension, or had been granted entitlement to free out-patient antipsychotic medication for the treatment of schizophrenia between 1969-1998 according to the International Classification of Diseases ICD-8, ICD-10, and DSM-III-R. (N=33,731 individuals) (Wegelius *et al.* 2013). First degree relatives (i.e. parents, siblings, and offspring) of patients were identified through the Population Register Centre and pedigrees of the families were constructed based on this information (Figure 11) (Hovatta *et al.* 1997). Information of psychiatric morbidity and diagnoses for the family members was obtained through the health care registers.

Probands were contacted by their treating physician and additional family members were contacted only upon approval by the proband. In general, families with at least two affected individuals were contacted. However, families that originated from the internal isolate that had a high risk for schizophrenia, also families with one affected individual were contacted. Clinical data were collected from all mental health treatment contacts.

A DSM-IV lifetime diagnosis of each patient was assigned independently by two psychiatrists based on the patient's medical records from all lifetime in- and outpatient psychiatric treatment contacts, completed by information from the clinical interview when available. In the case of discrepancy of diagnosis, a third psychiatrist was consulted and a best estimate for life time diagnosis was assigned. In general the concordance between the psychiatrist was high (>95%) (Paunio *et al.* 2001, Tuulio-Henriksson *et al.* 2002). The Operational Criteria Checklist (OPCRIT) was also completed for the affected individuals (McGuffin *et al.* 1991). All patients with a subset of 929 family members (340 affected with schizophrenia or schizoaffective disorder) had been interviewed using the Structured Clinical Interview for DSM-IV (SCID-I and SCID II) (First *et al.* 1997) and have been tested using a comprehensive neuropsychological test battery (Tuulio-Henriksson *et al.* 2002). The test battery included measures from the Wechsler Memory Scale-Revised (WMS-R) (Wechsler 1987), the California Verbal Learning Test (CVLT) (Delis *et al.* 1987), Wechsler Adult Intelligence Scale-Revised (WAIS-R) (Wechsler 1981) and covers the central cognitive functions impaired in schizophrenia (Heinrichs *et al.* 1998). In addition, the interview included the Scale for Assessment of Positive Symptoms (SAPS) (Andreasen 1984) and the Scale for the Assessment of Negative Symptoms (SANS) (Andreasen 1983).

DNA samples were collected from 2868 study participants in 670 families (N=983 patients with schizophrenia or schizoaffective disorder). One index case for schizophrenia or schizoaffective disorder from each family was selected for genome-wide association study (GWAS) on two occasions. A set of 191 patients were included in an international EU-funded consortium called the SGENE (www.sgene.eu, Study I, and III). Here individuals with a neurocognitive test data available and a subset of the schizophrenia family collection originating from the internal north eastern isolate of Kuusamo with approximately three-fold lifetime risk of schizophrenia were prioritized. An index case from each of the remaining families (N=375) were genotyped genome-wide in study III. Of the 566 patients that were included in study III, 171 originated from the internal isolate of Kuusamo.

The rest of the Finnish schizophrenia patients (N=86) were identified among the participants of the Health 2000 nationwide health survey (H2000) (Perala *et al.* 2007) and Northern Finland 1966 Birth Cohort (NFBC 1966) (Moilanen *et al.* 2003). These patients had been diagnosed with DSM III-R and DSM IV schizophrenia or schizoaffective disorder by using multiple sources of information, including structured clinical interviews with SCID-I and medical records. The cohorts are described in more detail below.
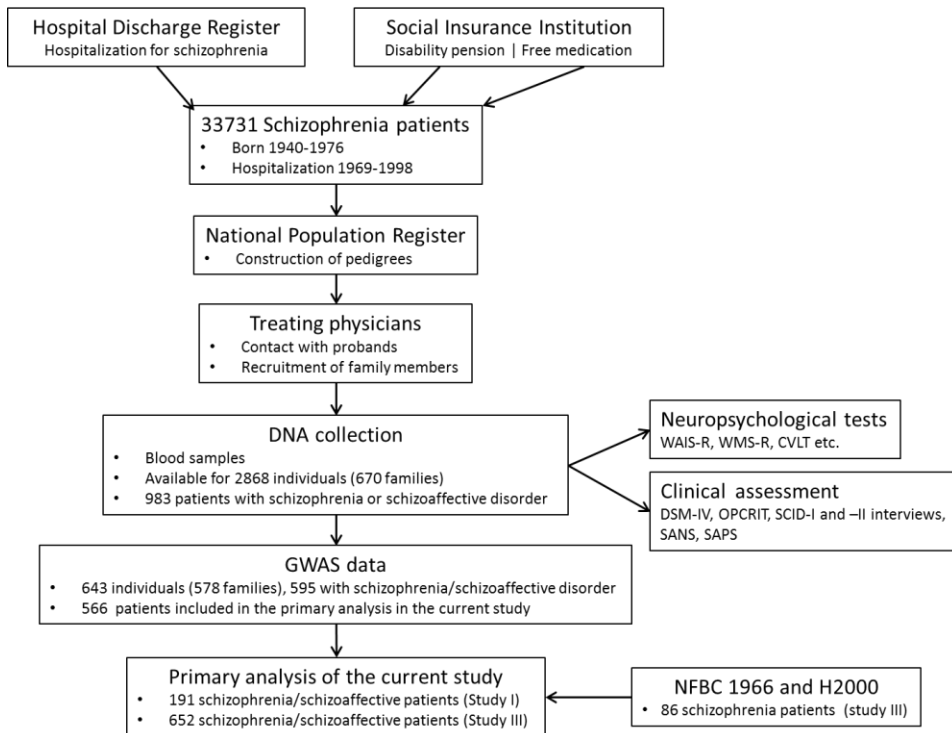

Figure 11.   Collection scheme of the schizophrenia family sample.

THE NORTHERN FINLAND 1966 AND 1986 BIRTH COHORTS (NFBC 1966 AND 1986)

All individuals born in 1966 in the two most Northern provinces of Finland were recruited to participate in an unselected geographically based prospective birth cohort called the Northern Finland 1966 Birth Cohort (NFBC 1966, N=12 058) (Rantakallio 1988). The cohort began with the recording of gestational information from pregnant mothers and continued with follow ups at birth, at one year old, at 14 years, at 31 years, and the currently ongoing 46 years old follow up (Sovio *et al.* 2007) (http://www.oulu.fi/nfbc/). The follow ups include clinical examinations,

questionnaires and information obtained from official registers, combined in a detailed phenotype database of the cohort participants. During the last follow-up also DNA samples were collected from a representative subset of the cohort that was included in studies II and III (N=4932).

Seven phenotypes recorded in NFCB 1966, postulated to relate to central nervous system development, were utilized in studies II and III. They were cognitive defect (IQ<85), a standardized measure of poor school performance, psychosis, epilepsy, neonatal seizures, cerebral palsy/perinatal brain damage, and impaired hearing (Table 3). The sample included 59 schizophrenia patients (DSM-III-R) that were included as cases for the association analysis of schizophrenia in study II. The rest of the cohort (N=488) were used as controls. Study III utilized the division between the Kuusamo isolate and rest of Finland. Of the NFBC 1966, 450 individuals (13 schizophrenia patients) originated from the isolate.

A sample set of 212 individuals included as controls from the internal isolate of Kuusamo in study III were ascertained from the NFBC 1986. As with the NFBC 1966 cohort, NFBC 1986 is a representative prospective birth cohort of all individuals born 20 years later in the same geographical region (Jarvelin *et al.* 1997). The NFBC 1986 cohort was founded on 9340 live born children that have been prospectively followed up since the 12th gestational week. The last follow-up to date at age 16 years (years 2001-2002) included 6645 individuals who gave consent to participate in the study (Jarvelin *et al.* 1993, Jaaskelainen *et al.* 2005). Individuals with grandparents originating from the isolate were included in study III.

Table 3. Seven Neurodevelopmental phenotypes recorded for NFBC 1966 included in studies II and III.

IQ < 85 (Rantakallio *et al.* 1985)
- Questionnaire filled in by the midwife at birth
- questionnaire administered for all children admitted to a children's hospital during the first 28 days of life
- diagnoses on admission to a children's hospital during 1966–1972
- a questionnaire given to parents when the child was 1 year old, relating to his or her health and development
- hospital records and special forms for children who visited neurological outpatient clinics required because of their symptoms.
- all existing protocols for IQ tests and psychologist's evaluations from child guidance centers, hospitals and institutions for mentally retarded children, information from national registers of death certificates, hospital discharge registers and child subsidies for chronically sick and mentally retarded children

Repeated grades in school (Isohanni *et al.* 1998)
- Questionnaire at the age of 14 years from the participants

Psychosis (Moilanen *et al.* 2003)
- DSM-III-R

Epilepsy (Lofgren *et al.* 2009)
- Hospital Discharge Register 1966-2004
- Social Insurance Institution of Finland 1966-2004

Neonatal Seizures and Cerebral Palsy/Perinatal Brain Damage (von Wendt *et al.* 1985, Jarvelin *et al.* 1997, Jones *et al.* 1998)
- Hospital Discharge Register
- Hospital charts

Impaired hearing (Sorri *et al.* 1985, Jarvelin *et al.* 1997)
- Air-conduction pure tone thresholds exceeded 20 dB at any of the frequencies of 0.25, 0.5, 1, 2, 3, 4, 6, 8 kHz

---

HELSINKI BIRTH COHORT STUDY

Helsinki Birth Cohort study (HBCS) was used as a whole Finland population sample for study III (N=1586) (Eriksson *et al.* 2001, Barker *et al.* 2005). The cohort includes 8760 individuals born in Helsinki Central Hospital between 1934 and 1944. The participants had attended the child welfare clinics in Helsinki and were resident in Finland in 1971 when a unique personal identification number was allocated to each resident living in Finland including every member of Finnish population (Eriksson *et al.* 2001). A subset of the cohort (N=2003) participated in a clinical study between 2001 and 2004 that also involved collection of DNA samples from the participants (Barker *et al.* 2005).

THE HEALTH 2000 SURVEY (H2000)

A total of 2 402 individuals were ascertained on two occasions from a nationally representative cross-sectional health survey called the Health 2000 survey and were included in studies I and III (Aromaa *et al.* 2004, Pirkola *et al.* 2005). The H2000 health survey,designed to give information on health of the Finnish population, was carried out in 2000 and 2001. It included a nationwide representative sample of 8028 individuals. A subset of 200 individuals was included as the Finnish control sample set in a large European schizophrenia consortium called the SGENE (study I) (Stefansson *et al.* 2008). The 200 individuals were matched with Finnish schizophrenia cases with regard to sex, age (± 5 years) and place of birth, excluding persons with a lifetime history of any psychotic disorders. Persons who had participated in a sub-study in which neuropsychological testing had been conducted were prioritized (Perala *et al.* 2007), as well as participants from the high risk isolate. The majority of the H2000 participants used in study III (N=2212) were part of a sub-study of H2000 that investigated genetic risk factors for metabolic syndrome. All individuals with International Diabetes Federation (IDF) definition of metabolic syndrome were identified from H2000 and matched controls were selected for those individuals (Pajunen *et al.* 2010, Kristiansson *et al.* 2012). From these individuals, 27 met the DSM-IV diagnosis for schizophrenia or schizoaffective disorder, diagnosed in a sub-study focusing on psychotic disorders (Perala *et al.* 2007), and were included as cases in study III.

THE CARDIOVASCULAR RISK IN YOUNG FINNS STUDY (YFS)

The Cardiovascular Risk in Young Finns Study (YFS) is a follow-up study of cardiovascular risk factors from childhood to adulthood (http://youngfinnsstudy.utu.fi/). The study participants were randomly chosen from the Population Register and recruited from five university cities in Finland. The baseline study was launched in 1980 and included 3596 individuals. Follow-ups have taken place on average every three years with the last one to date being in 2007 at 30 years old. DNA was extracted from the study participants in 2001(Raitakari *et al.* 2003, Raitakari *et al.* 2008, Smith *et al.* 2010). The YFS individuals were included as controls in study III and included 2308 individuals from YFS with GWAS data available.

THE NATIONAL FINRISK HEALTH STUDY (FINRISK)

Participants of two population cohorts were selected to investigate the enrichment of rare deletions between two geographical regions in Finland of differing schizophrenia prevalence. All individuals who had both parents originating from the internal north-eastern isolate with a high relative risk for schizophrenia were

selected among the voluntary participants of a nationwide health cohort Finrisk (http://www.thl.fi/fi/tutkimus-ja-asiantuntijatyo/vaestotutkimukset/finriski-tutkimus/the-national-finrisk-study). The Finrisk cohort is a national survey on risk factors of chronic and non-communicable diseases in Finland. The survey has been conducted every five years since year 1972 using randomly selected, representative population samples from different parts of the country (Vartiainen *et al.* 2010). The 173 individuals from the isolate, included in the present study, were selected from among the participants of Finrisk surveys 1997 to 2007 (Vartiainen *et al.* 2010). A subset of individuals, aged between 25 and 74, from the Finrisk 2007 health survey participated a sub study called Dietary, Lifestyle and Genetic determinants of Obesity and Metabolic syndrome in the Helsinki region (Dilgom) (Inouye *et al.* 2010). The present study included 612 individuals from Dilgom who were treated as population controls.

FINNISH TWIN COHORT (FTC)

A sample set of 1492 individuals was ascertained from the study participants of the Finnish Twin Cohort (FTC) (one individual per pair) (Kaprio *et al.* 2002, Kaprio *et al.* 2002). The FTC includes both monozygotic and dizygotic twins, collected in several separate phases. The older cohort was initiated in 1975 and included 13888 same-sex pairs with known zygosity. The cohort was followed up in 1996 with a collection of an additional 8000 pairs with opposite sex born between 1938-1957 (Kaprio *et al.* 2002). The younger Finnish twin cohorts, referred to as FinnTwin12 and FinnTwin16, are longitudinal studies of behavioral development and health habits of Finnish twins enrolled at ages 11-12 and 16, respectively. The FinnTwin12 cohort included five consecutive birth cohorts of Finnish twins born in the 1983-1987 period and included 2295 complete pairs of monozygotic and dizygotic twins. In the FinnTwin16 cohort twin pairs born between 1975 and 1979 were identified and the cohort included 2618 complete pairs of monozygotic and dizygotic twins (Kaprio *et al.* 2002, Jelenkovic *et al.* 2011).

4.1.2 International schizophrenia cohorts

SGENE

The phase I sample set of the SGENE consortium (study I) included 1433 cases diagnosed with schizophrenia spectrum psychosis and 33,250 controls from six countries: Iceland (N=646 cases, 32442 controls), Scotland (N=211 cases, 229 controls), Germany (N=195 cases, 192 controls), England (N=105cases, 96 controls), Italy (N=85 cases, 91 controls), and Finland (N=191 cases from the schizophrenia family sample and 200 controls, from the H2000 cohort). The phase II sample used for replication attempt of the preliminary findings included 3285 cases

and 7,951 controls from Scotland (N=451 cases, 441 controls) Germany (Munich, N=420 cases, 422 controls), Germany (Bonn, N=491 cases, 875 controls), The Netherlands (N=806 cases, 4,039 controls), Denmark (N=442 cases and 1,439 controls), Norway (N=237 cases, 272 controls), and China (N=438 cases, 463 controls)

THE INTERNATIONAL SCHIZOPHRENIA CONSORTIUM (ISC)

The ISC consisted of 3391 schizophrenia patients diagnosed with DSM-IV or ICD 10 and 3181controls of European origin and was utilized as a replication sample in study III. The CNV calls for the cohort were obtained from the intensity data of the genotype probes on the Affymetrix 5.0 and 6.0 genotyping arrays by using Birdseye (Korn *et al.* 2008). A detailed description of the data set together with the CNV calls has been previously published (Stone *et al.* 2008).

THE SWEDISH SCHIZOPHRENIA COHORT (SSC)

A total of 5785 schizophrenia patients and 6348 controls were selected from the Swedish Schizophrenia cohort. The study subjects were identified from among all individuals hospitalized at least twice for schizophrenia since 1973 in Sweden. The controls had no hospitalization due to psychiatric disorders and were matched with the cases for age, sex, and county of residence (Bergen *et al.* 2012). All subjects were over 18 years old and gave their written informed consent to participate in the study.

4.2 Methods

4.2.1 Genome wide SNP arrays

Data obtained from the genome-wide SNP genotyping chip arrays were used to identify copy number variants in all studies. In addition, the allelic information from the Genome-wide arrays was used to construct common SNP haplotypes in studies II and III. The GWAS data were generated at several time points for the study samples as a part of multiple research projects in different institutes using SNP arrays that were available at each time point (Table 4).

Table 4.   Genome-wide genotyping arrays used for each sample set

Study sample	No. of Individuals	GWAS array	genotyping center	Study number
SGENE Phase I	33683	Illumina HumanHap300	deCode genetics, Iceland	I
SGENE Phase II	11234	Illumina HumanHap300, HumanHap550, HumanCNV370, Affymetrix GeneChip® GenomeWide 6.0 SNP Array, TaqMan	Duke University, Bonn University, deCode Genetics	I
NFBC 1966	4932	Illumina Infinium 370cnvDuo	The Broad Institute, USA	II, III
HBCS	1586	Human670K customBeadChip	WTSI	III
Finrisk	173	Human670K customBeadChip	WTSI	III
H2000	2402	Human610-Quad BeadChip, HumanHap 300	WTSI	III
NFBC 1986	212	Human670K customBeadChip	WTSI	III
YFS	2308	Human670K customBeadChip	WTSI	III
FSFS	566	Human670K customBeadChip, Illumina HumanHap300	WTSI, DeCode Genetics	III
SSC	12133	Affymetrix 5.0 SNP array	The Broad Institute, USA	III
ISC	6572	Affymetrix 5.0 SNP array	The Broad Institute, USA	III
FTC	1492	Human670K customBeadChip	WTSI	III

4.2.2 Genotyping CNVs with Sequenom MassARRAY and Real-time quantitative PCR

Verification of accurate CNV calling was conducted in a subset of 258 individuals (19 carrying a 22q11.22 deletion) who were genotyped using 31 SNPs selected from the GWAS arrays (Study III). Of the 31 SNPs, 26 overlapped the 22q11.22 deletion and the remaining five flanked it on both sides. The genotyping was performed

using Sequenom MassARRAY platform in 384-well plates with one iPLEX reaction using standard protocols specified by the manufacturer (Sequenom Inc., San Diego, CA, USA). Automated allele calling was manually verified using the MassARRAY Typer 4 software. The 384-well plate included eight water controls and eight inter-plate duplicate samples for quality control.

Individuals that had a homozygous stretch that corresponded to the deletion were further genotyped with Quantitative Real-time Polymerase Chain Reaction (QRTPCR). For this two probes were designed to both ends of the deletion area. The gene encoding for B-globin was used as a reference. The reactions were done in single plexes using SYBR$^®$ GREEN PCR Master Mix on the ABI Prism® 7900 Sequence Detection System (Applied Biosystems, Foster City, CA, USA). The reactions for each probe were carried out in three repeats to control for sporadic variation. Each 384-well plate included six water controls for quality control. In addition eight individuals with known 22q11.22 deletion status were used to validate the measurements.

4.2.3 Gene Expression measurement

Genome-wide expression data were generated for 65 individuals of 18 families belonging to the Finnish schizophrenia family sample using Illumina HT-12 v4 Expression BeadChip. Among these 65 individuals 12 were heterozygous carriers of the 22q11.22 deletion, 2 had a homozygous deletion and 51 were non-carriers.

4.2.4 Detection of CNVs

The CNVs were genotyped using normalized total intensity Log R ratio (LRR) and allelic intensity ratios, termed B-allele frequency (BAF) obtained from the probes on the Illumina genotyping arrays using PennCNV and QuantiSNP (Colella *et al.* 2007, Wang *et al.* 2007). Both software utilize a hidden Markov model for predicting the likelihood for a CNV. The model assumes that the observed intensities from an assumed Gaussian emission distribution are related to an unobserved copy number state at each locus. In addition, the neighboring loci are assumed to have similar copy number states and the probability of transition in copy number state is dependent on the distance between two markers (Wang *et al.* 2007, Wang *et al.* 2008). The CNV calling was done independently for each cohort and the quality control (QC) was carried out by graphical clustering of LRR or BAF standard deviations and removing outlier samples from each cohort. In all cohorts, individuals were excluded if their LRR or BAF standard deviations were over 0.35 and 0.07, respectively; the waviness factor was outside the range -0.04 -0.04; and

BAF drift exceeded 0.002, according to the software recommendations. Only those that exceeded the likelihood estimate of log Bayes factor > 10 in the CNV calls from QuantiSNP were included in the analysis. To control for false positive calls, each CNV was required to be detected with at least 10 probes. In each study the predicted CNV genotypes were confirmed by visual inspection of the LRR and BAF plots in Illumina Genome Studio or in R (Figure 12).
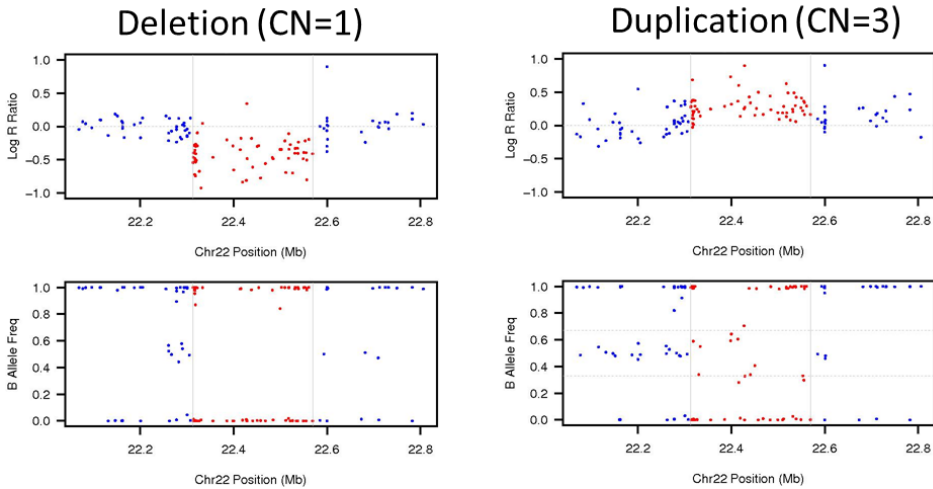

Figure 12.   Example of LogR Ratio and B-allele Frequency distribution for deletion and duplication. Each dot in the plot represents one marker on the GWAS array. Markers within the CNV are in red. For deletion the LogR Ratio reduces to -0.5 and no heterozygous genotypes (B allele frequency ~0.5) are observed. For duplication the Log R Ratio increases to 0.5 and the heterozygous genotype band is devided according to allele distribution to AAB (BAF~0.3) and ABB (BAF ~0.6) genotypes.

4.2.5 Statistical testing and haplotype construction

The statistical testing for association included Cochran-Mantel-Haenszel test, Fisher exact test, χ^2-test, and t-test, analyses were conducted using PLINK software and R software (version 2.0.2) (Purcell *et al.* 2007). A statistical correction for multiple comparisons was included in all studies. In studies I and II a conservative Bonferroni correction (α_B) was applied that divides the test-wise significance (α) level by the number of tests (m) (Equation 1).

$$\alpha_B = \frac{\alpha}{m} \qquad (1)$$

The statistical significance in study III was achieved through 10 000 permutations.

The gene expression analyses were conducted using R (MA-ANOVA software package). The ancestral haplotypes were constructed from common SNPs by phasing with Beagle software (v3.1) (Browning *et al.* 2007). The SNP QC was performed according to standard procedures and included by-marker and by-sample success rates of >95%, and tests for the correct sex. Deviation from Hardy-Weinberg equilibrium (HWE) can be indicative of problems in genotyping. Each marker was tested for HWE using the exact test implemented in PLINK (Wigginton *et al.* 2005). Markers deviating from Hardy-Weinberg equilibrium with $p < 1 \times 10^{-6}$ were removed. Kinship among study participants was calculated from genome-wide SNP data by estimating genome-wide sharing by identity by descent (IBD) likewise implemented in PLINK (Milligan 2003). Stretches of extended homozygosity were searched with PLINK to identify individuals that potentially carry deletions at a given site. For a stretch to be considered homozygous maximum two heterozygous genotypes were allowed per region. The pathway analysis for genes showing nominal statistical differences of mRNA levels ($p < 0.05$) between deletion carriers and non-carriers was conducted with DAVID bioinformatics resources 6.7 (Huang da *et al.* 2009).

5 Results and Discussion

The role of large genomic CNVs in schizophrenia and neurodevelopmental phenotypes was studied in this thesis. The combined association findings obtained from studies I, II and III are presented in table 5. Specifically three deletions at genomic regions prone for recurrent CNV events were associated with schizophrenia (Study I). Carriers of large (>500kb) deletions were additionally found to have more often phenotypes related to abnormal neuronal development, including intellectual impairment (IQ<85) and repeated grades in school, as well as impaired hearing compared to non-carriers (Study II). Utilizing the unique genealogy of Finns, a deletion on chromosome 22q11.22 presenting as a founder mutation, was identified to associate with schizophrenia. Moreover, the non-schizophrenic carriers of the 22q11.22 deletion had intellectual deficit and repeated grades in school (Study III) more often than non-carriers. The results combined in this thesis reveal a role for large genomic deletions in the pathogenesis of schizophrenia and intellectual abilities. The results further provide genetic evidence that schizophrenia is indeed a neurodevelopmental disorder and some part of the genetic predisposition at least in a subset of patients is non-specific to schizophrenia but is shared with other psychiatric disorders.

Table 5. Deletions associated with schizophrenia and phenotypes of neuronal development combined from studies I, II, and III.

CNV-locus	phenotype	N cases	N CTRLs	p-value	OR	Study
1q21.2	SZ	4,718	41,199	2.9×10^{-5}	14.83; 3.55-60.40	I
15q12.1	SZ	4,718	41,199	6.0×10^{-4}	2.73; 1.50-4.89	I
15q13.3	SZ	4,718	41,199	5.3×10^{-4}	11.54; 2.53-49.58	I
> 500 kb deletions	IQ<85	68	4,473	0.002	3.79; 1.54-8.14	II
> 500 kb deletions	Repeated grades	189	5,352	0.0009	2.7; 1.47-4.67	II
> 500 kb deletions	Impaired hearing	151	4,390	0.002	2.72; 1.38-4.95	II
22q11.22 (TOP3β)	SZ	9,828	21,400	0.007	2.17 ; 1.18-2.87	III
22q11.22 (TOP3β)	Repeated grades	197	4,675	0.0034	3.99; 1.78-8-94	III
22q11.22 (TOP3β)	IQ<85	69	4,781	0.033	4.60; 1.41-14.96	III

5.1 Recurrent microdeletions increase risk for schizophrenia (Study I)

To identify genomic regions susceptible for recurrent CNV events, 9878 parent-offspring transmission were scrutinized for *de novo* CNVs in an Icelandic discovery sample of 2160 trios (mother, father, and child) and 5558 parent-offspring pairs. The sample was ascertained from the general population but did not include any schizophrenia patients. The hypothesis was that if the identified *de novo* CNVs occur at a low frequency in a population despite a relatively high mutation rate (~1/10000 meiosis) they are likely to be under negative selection pressure. Thus, they could also have a role in severe mental disorders. The transmission analysis yielded 66 plausible *de novo* CNVs (15 duplications and 51 deletions in 62 non-overlapping genomic regions) that were present in the offspring but not in the parents. Of the *de novo* CNVs, 32 were flanked by LCRs at least on one side (23 from both sides). Of the 23 CNVs, flanked by LCRs on both sides, 18 (78%) were observed in more than one individual. The remaining 34 CNVs were not flanked by LCRs and 27 (79%) of these were identified only in a single discovery trio.

The 66 plausible recurrent CNV sites that were identified to harbor *de novo* CNVs in the discovery population were further tested for association to schizophrenia in a multicenter sample of 1433 patients and 33 250 controls (SGENE phase I sample). Large deletions in three regions, 1q21.1 (OR=8.68; 95%-CI: 1.02-49.76), 15q11.2 (OR=3.90; 95%-CI: 1.42-9.37) and 15q13.3 (OR=8.94; 95%-CI: 0.79-58.15) showed a nominal association for schizophrenia (p<0.05). From the Finnish sample set, only one control individual carried the deletion on chromosome 15q11.2. In the second phase the preliminary association for all three loci was replicated in a multicenter sample of 3285 cases and 7951 controls (SGENE Phase II sample) (p<0.05). In a combined analysis of phase I and phase II sample sets of all three deletions provided significant associations to schizophrenia (p<0.001) with large ORs (1q21.1: OR=14.83; 95%-CI: 3.55-60.40; 15q11.2: OR=2.73, 95%-CI: 1.50-4.89; 15q13.3: OR=11.54, 95%-CI: 2.53-49.58).

The 15q11.2 and 15q13.3 deletions that were associated with schizophrenia flanked the PWS region on 15q11-q13. The 15q11.2 deletion is included in the PWS type 2 deletions (Figure 13) (Cassidy *et al.* 2012). PWS results typically from a 4-5Mb paternal deletion spanning 15q11.2-q13 or maternal uniparental disomy. The region contains imprinted uniparentally expressed genes that contribute to the phenotype. A maternally inherited deletion results in Angelman syndrome (AS). PWS is a

multisystem disorder with symptoms that include intellectual disability, behavioral phenotypes, obesity and characteristic dysmorphic features. Some of the PWS behavioral characteristics are similar to autism. In addition, psychosis is present in 10-20% of PWS patients (Cassidy *et al.* 2012). Symptoms of AS typically include developmental delay, characteristic behavioral phenotype, dysmorphic features, and speech impairment (Williams *et al.* 2010). Reciprocal duplications of the PWS/AS region are phenotypically distinct from either PWS or AS. The duplication carriers lack the facial dimorphisms, but have varying learning deficits and autistic features (Boyar *et al.* 2001).
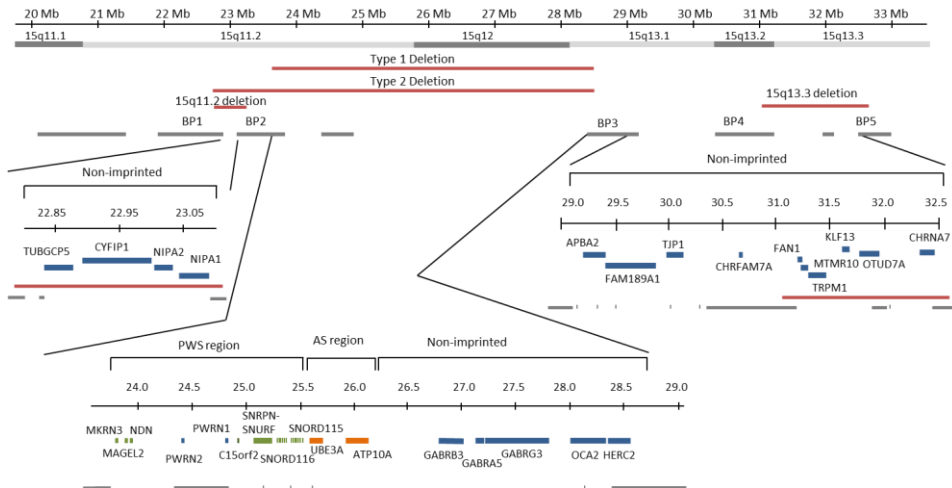

Figure 13. Prader-Willi syndrome (PWS) showing the region on 15q11.1-q13.3. The horizontal bold red lines show the deleted regions that are associated with psychiatric disorders in the region. The type 1 and type 2 deletions indicate deletions that underlie PWS. The low-copy repeats in the regions are shown as bold grey lines and the common break points (BP1-BP5) that are associated with the deletions is also shown. Three regions are zoomed in: BP1-BP2, BP2-BP3, and BP3-BP4. Genes that underlie these regions are presented as bold blue lines (non-imprinted), bold green lines (paternally expressed), bold orange lines (maternally expressed, related to Angelman Syndrome). Note that these regions can contain additional transcripts that are not presented in the figure and that the figure is not fully in scale. The figure is adapted from (Cassidy *et al.* 2012).

The most palusbile candidate gene in the 15q11.2 region is the cytoplasmic FMR1 interacting protein 1 (CYFIP1). In addition to interacting with FMRP (Schenck *et al.* 2001), CYFIP1 interacts with Rho GTPase Rac1. Rac1 is important for regulating axonal and dendritic outgrowth through actin reorganization and is necessary for the development and maintenance of neuronal structures (Luo 2000). In addition to

FMR1, other X-linked mental retardation genes have been implicated in the Rho/Rac pathway (Chelly *et al.* 2001).

The 15q13.3 deletion disrupts a gene that encodes for CHRNA7. It is an α7 nicotinic acetylcholine receptor subunit, which is part of a ligand-gated ion channel that mediates fast signal transmissions at the synapses. It is expressed mostly presynaptically in the brain and has a role in modulating glutamate release at the synapses. Nicotine acetylcholine receptors had earlier been suggested to play a role in different diseases of the central nervous system, including Alzheimer's disease, Parkinson's disease, Tourette's syndrome, ADHD, autism and schizophrenia (Gotti *et al.* 2004, Gotti *et al.* 2006). Evidence from GWAS also support a role for common genetic variants of cholinergic receptor genes in schizophrenia (Schizophrenia Working Group of the Psychiatric Genomics 2014). The 15q13.3 deletion, identified in the present study has specifically been implicated in an increased risk for autism, intellectual disability, and epilepsy (Sharp *et al.* 2008, Ben-Shachar *et al.* 2009).

Deletion of the corresponding α nicotinic acetylcholine receptor in mice causes impaired memory and learning (Fernandes *et al.* 2006, Hellier *et al.* 2012). These defects are suggested to originate from impaired cortical development of parvalbumin containing GABAergic interneurons (Lin *et al.* 2014). Deficits in parvalbumin-containing GABAergic interneurons have also been suggested to play role in cognitive deficits that present in schizophrenia (Lewis *et al.* 2012). Furthermore, nicotinic acetylcholine receptors, including the α7-receptor, have emerged as potential therapeutic targets for treating neurocognitive impairment in schizophrenia (Freedman 2014).

The 1q21.1 deletion that was associated with schizophrenia overlaps several genes including two gap junction protein coding genes *GJA5* and *GJA8* that respectively encode for connexin 40 and connexin 50. Gap junction proteins enable direct metabolic and electrical communication between cells in the brain and other tissues (Sohl *et al.* 2005). Mutations in *GJA5* and *GJA8* have been associated with cardiac and eye phenotypes, respectively (Jiang 2010, Delmar *et al.* 2012). The *GJA8* gene has earlier been associated with schizophrenia (Ni *et al.* 2007). The 1.4Mb 1q21.1 deletion is also referred to as 1q21.1 deletion syndrome (OMIM#612474). The 1q21.1 deletion syndrome is associated with high level of phenotypic variation, including a variable degree of intellectual disability, behavioral abnormalities, dysmorphic features, cardiac abnormalities, and cataracts. The deletion is reported to be *de novo* and also being inherited from unaffected or mildly affected parent. Duplications in the region are also associated with intellectual disability and autistic

features (Brunetti-Pierri *et al.* 2008, Mefford *et al.* 2008). In the present study, one Icelandic schizophrenia patients and 13 controls (12 from Iceland and one from Finland) had the corresponding duplication of 1q21.1. The 1q21.1 deletion has been associated repeatedly with schizophrenia and autism (Autism Genome Project *et al.* 2007, International Schizophrenia Consortium 2008, Walsh *et al.* 2008, Sahoo *et al.* 2011).

5.2 Large deletions associate with impaired cognitive functioning (Study II)

To study the impact of large over 500 kb deletions on neuronal development, a method termed phenotype mining that involves search of rich phenotype databases for measures correlated with genetic variation was used. A subset of seven pre-selected phenotypes postulated to relate to abnormal neuronal development were studied in an unselected population sample of 4932 individuals ascertained from the NFBC 1966. These seven phenotypes included cognitive defect (IQ<85), a standardized measure of poor school performance, psychosis, epilepsy, neonatal seizures, cerebral palsy/perinatal brain damage, and impaired hearing (Table 3). A genome-wide scan for CNVs over 500 kb in the NFBC 1966 resulted in 634 CNVs in 529 individuals. A majority of the carriers (83%) had only a single CNV. The CNVs had a mean size of 991 kb and they spanned 165 non-overlapping CNV-regions (CNVRs). A majority (65%) of the identified CNVRs included only a single CNV. Deletions were significantly underrepresented compared to duplications (p=8.6×10^{-28}). Also no homozygous deletions were observed. Five individuals with trisomy 21 and three with trisomy X were identified and removed from subsequent analysis with the phenotypes.

Individuals with large deletions (N=160) had higher frequency of intellectual defect, which is defined as IQ<85 (OR=3.79, 95%-CI: 1.54-8.14), and learning difficulties indicated as repeated grades in school (OR=2.7, 95%-CI: 1.47-4.67) compared to individuals with no CNVs (N=4381). All individuals with intellectual defect had also repeated grades. However, even after excluding these individuals, learning difficulties were twice as common among deletion carriers compared to non-carriers (OR=2.02, 95%-CI: 0.84-4.24). In addition, impaired hearing was more frequent among carriers of large deletions compared to individuals with no CNVs (OR=2.72, 95%-CI 1.38-4.95). However, the frequency of psychosis did not differ significantly between the deletion carriers and non-carriers (OR=1.25, 95%-CI 0.33-3.38). Further, none of the phenotype categories were significantly overrepresented collectively among individuals with duplications (p<0.05).

A marked co-occurrence of the three overrepresented phenotype categories of IQ<85, repeated grades, and impaired hearing was observed among the carriers of large deletions. Of the 160 deletion carries, 24 had at least one of the phenotypes, 10 at least two phenotypes, and three individuals had all three phenotypes. The overrepresentation of the three phenotypes with deletion carriers was accounted for by 18 distinct deletions. The deletion carriers of 11 of the 18 deletions had at least two of the phenotypes (Table 6).

In additional analysis, we investigated also whether the age of achieving early developmental milestones would differ between carriers of large deletions and non-carriers. The deletion carriers were less likely to know their own name at the age of 3 (p=0.03, OR = 0.33; 95%-CI: 0.09-0.93) and more likely to have had abnormal hearing at age of four (p=0.0012, OR3.54; 95%-CI:2.58-7.24). However none of these differences remained significant after Bonferroni correction for 44 tests. The nominal association with impaired hearing at the age of four was not overlapping with the association detected for impaired hearing measured at 14 years old.

Table 6. The 18 deletions that underlie the associations with IQ<85, repeated grades, and impaired hearing in the NFBC 1966.

Chr	Start	End	Tot obs.	IQ < 85	Repeated grades	Impaired hearing
2p16.3	50.0	51.6	1	ID25	ID25	–
4q24–q25	104.5	109.4	1	–	–	ID26
4q35.2	188.4	189.5	1	–	ID27	ID27
4q35.2	189.5	190.7	3	ID28	ID28	–
6p22.3	22.2	22.8	1	–	–	ID30
6q16.1	94.4	95.0	4	ID31	ID31	–
7p12.1	53.1	53.8	13	–	ID34, ID35	ID32, ID33
7q11.21	64.2	64.8	1	–	ID36	–
8p23.1	7.2	7.8	17	–	–	ID37
10q11.22	46.2	46.8	25	ID16	ID16	ID14, ID15, ID16
10q26.2–q26.3	128.3	132.7	1	–	–	ID18
11p14.3	21.7	22.3	1	ID19	ID19	–
15q11.2	21.9	22.7	13	–	ID8, ID12	–
15q13.1–q13.2	28.7	30.3	1	ID20	ID20	ID20
16p11.2	29.5	30.1	3	ID21	ID21	–
17q21.31–q21.32	41.5	42.0	1	–	ID22	ID22
22q11.21	17.2	19.8	1	ID23	ID23	ID23
22q11.22–22q11.23	21.3	22.0	3	–	ID24	–

The associations between large deletions and intellectual phenotypes were accounted for by deletions in 14 genomic regions (16 individuals), from which eight were associated only with intellectual disability. These include the 15q13.3 deletion found earlier to be associated with schizophrenia (study I). The 14 regions included two other schizophrenia associated deletions at 2p16.3 and 22q11.2 (International Schizophrenia Consortium 2008, Stefansson *et al.* 2008, Kirov *et al.* 2009). All of the carriers of these deletions had IQ<85 and also had repeated grades in school, but did not have diagnosis for psychosis in the birth cohort. One individual likewise with IQ<85 and repeated grades in school had a 16p11.21 deletion. Reciprocal duplications in this region are associated with schizophrenia and autism (Weiss *et al.* 2008, McCarthy *et al.* 2009).

The enrichment of large deletions among individuals that have repeated grades in school suggests that large deletions affect cognitive skills on a general level in

population. It has been obligatory since 1921 that children aged seven attend a public elementary school in Finland. A postponement or exemption can be granted if the child has intellectual disability or delayed development. The compulsory public elementary education enables the systematic assessment of a child's school performance. Compulsory schooling is generally nine years and lasts until age of 16. Among all children born in 1966 in Northern Finland, 2.2% did not attend regular school and 3.8% were at a lower grade than expected for that at the age of 14. Among those children in lower grade 44% of cases had an IQ that was below 85 (Rantakallio *et al.* 1986). Individuals who have repeated grades in school have been reported to have an increased risk for psychiatric disorders including schizophrenia (Isohanni *et al.* 1998).

The results suggest that large deletions have a stronger cumulative impact on traits related to early onset of intellectual disabilities, including milder subclinical features, than they do for either psychosis or schizophrenia. This observation is in line with findings from other studies that demonstrate that those CNVs that underlie schizophrenia are associated with a broad range of clinical variability. Even though conferring a marked risk for schizophrenia they are still associated with incomplete penetrance of generally less than 10% for schizophrenia (Vassos *et al.* 2010). Control individuals that carry CNVs associated with neuropsychiatric disorders perform in measures of cognitive ability intermediate to schizophrenia patients that carry the CNVs and non-carrier controls (Stefansson *et al.* 2014). Moreover, the risk conferred by the CNVs tends to increase towards earlier onset disorders including intellectual disability, congenital malformations, and autism spectrum disorders (Cooper *et al.* 2011, Malhotra *et al.* 2012, Kirov *et al.* 2014). Consistently, the burden of pathogenic CNVs is even higher among patients of childhood onset of schizophrenia compared to adult onset schizophrenia (Ahn *et al.* 2014).

5.3 Rare deletions in young isolate population represent founder mutations (Study II)

The enrichment of neurodevelopmental phenotypes among carriers of large deletions in the NFBC 1966 could be attributed to 18 distinct deletions (Table 6). Of these 18 deletions, eight were observed in more than one individual. Motivated by the Finnish population history, the possibility that some of these deletions would be founder mutations of a singular ancestral was investigated. Allelic haplotypes were constructed from the common SNPs included in the SNP array. Seven of the eight deletions that had multiple carriers were nested in a single ancestral haplotype. In addition, the parental birthplaces of the carriers of each deletion demonstrated

regional clustering further supporting common ancestry for the carriers and that the deletions were shared IBD between the carriers.

Genetic variants with strong adverse effects are likely to be infrequent in the population. An extreme example of such a variant are the disruptive recurrent *de novo* deletions associated with schizophrenia in study I. However, even these deletions are occasionally inherited. The rarity of these variants and the high level of baseline variation make it difficult to accrue statistical evidence that any particular variant is associated with a given trait. In recently expanded population isolates, however, the high levels of genetic drift that occur could potentially generate relatively high frequencies for high impact variants, and thereby create an opportunity to identify associations of particular phenotypes for single variants (Peltonen *et al.* 2000). The prevalence of certain complex neurodevelopmental phenotypes, including schizophrenia and intellectual disability in Finland shows a marked gradient from the Southwest to Northeast of Finland (Figure 14) (Rantakallio *et al.* 1986, Hovatta *et al.* 1997, Heikura *et al.* 2003, Arajarvi *et al.* 2005, Perala *et al.* 2008). This prompted us to pose the following question in study III: Could the increased prevalence of common disorders in particular Finnish regions be at least partially due to a similar local enrichment of rare risk variants? (Varilo *et al.* 2004)?
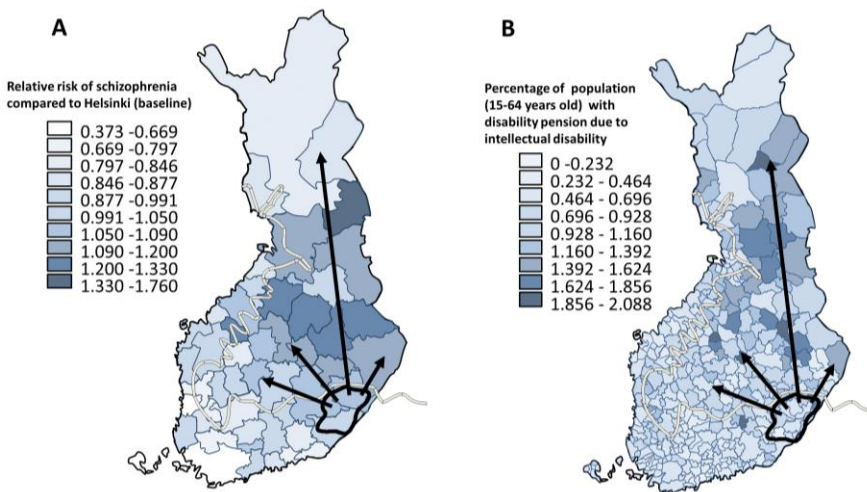

A

Relative risk of schizophrenia compared to Helsinki (baseline)

- 0.373 -0.669
- 0.669 -0.797
- 0.797 -0.846
- 0.846 -0.877
- 0.877 -0.991
- 0.991 -1.050
- 1.050 -1.090
- 1.090 -1.200
- 1.200 -1.330
- 1.330 -1.760

B

Percentage of population (15-64 years old) with disability pension due to intellectual disability

- 0 -0.232
- 0.232 - 0.464
- 0.464 - 0.696
- 0.696 - 0.928
- 0.928 - 1.160
- 1.160 - 1.392
- 1.392 - 1.624
- 1.624 - 1.856
- 1.856 - 2.088

Figure 14. Regional variation seen A) in the relative risk for schizophrenia and B) the frequency of intellectual disability as measured by the rate of issuing disability pensions. The pattern follows also the internal migration movement that occurred in Finland starting in the 16th century. The early and late settlements are separated by the grey border.

5.4 Utilizing population isolates in identifying rare variants underlying complex traits (study III)

To identify deletions that because of recent bottlenecks would have become enriched to regions with high relative risk for schizophrenia in Finland, a genome-wide scan for rare (frequency <5%) genomic deletions of over 20 kb was performed in two geographically based Finnish population samples. The two samples were assembled from the capital region which represents the mean relative risk of schizophrenia (N=1586, HBCS) and from the North-East Isolate with a high relative risk of schizophrenia (N=173, Finrisk). The scan resulted in 5313 putative rare deletions with a mean size of 89.4 kb and which comprised 1041 non-overlapping genomic loci.

Initial statistical comparisons between the sub-isolate population sample and the "whole Finland population sample" (from the capital region) for all 1041 putative deletion loci were conducted. For statistical significance, it was required that the difference in frequency between the sub-isolate sample and the whole Finland sample would occur by chance less frequently than once in 1000 permutations, corresponding roughly to the number of loci tested. Of the 1041 putative deletion loci, three loci displayed a significantly higher frequency of deletions in the sub-isolate population sample (empirical p-value <0.001 after 1000 permutations). To confirm the deletion loci for further analyses, a manual inspection of the deletion genotypes using intensity distribution plots in the three regions including testing for Hardy-Weinberg equilibrium was conducted. Finally, two deletion loci on chromosomes 4q12 and 22q11.22 passed the QC. These two deletions were found to be significantly more common in the general population originating from the "high risk" region compared to the capital area "low risk" population. The 4q12 deletion spanned 181.7 kb and was carried by six individuals from the isolate and none from the whole Finland sample (p=8.7×10^{-7}). The 22q11.22 deletion spanned 243.9 kb and was carried by 18 individuals from the sub isolate sample and one from whole Finland sample (p=8.55×10^{-18}).

A single ancestral origin was evident for both deletions, as determined by shared SNP haplotypes observed in both the sub-isolate population sample and the whole Finland population sample (total N=1793). For the 4q12 deletion, carriers shared a single 2.25 Mb haplotype (248 kb proximal and 2.0 Mb distal to the deletion), while carriers of the 22q11.22 deletion shared a 289 kb haplotype (167 kb proximal and 122 kb distal to the deletion). Both deletions must have arisen relatively recently. We deduced this possibility as both the 4q and 22q haplotypes were present in non-carriers of the deletions, with population frequencies of 1.1% and 13.0%, respectively. In arriving at these values, we assumed full LD between their proximal

and distal portions. Based on the haplotype sizes and frequencies, the 4q12 deletion probably arose more recently than the 22q11.22 deletion.

Next, the two deletions, enriched in the sub-isolate sample, were tested for association with schizophrenia in a sample of 185 patients and 747 controls, that originated from the isolate. Of the two deletions, the one on 22q11.22 showed significant association with schizophrenia within the isolate sample (OR=1.84, 95%-CI: 1.05-3.23). The deletion on 4q12 showed no such difference (p=0.14) however. The initial association of the 22q11.22 deletion with schizophrenia was further followed up in the whole Finland and non-Finnish European schizophrenia case-control samples of 467 and 9176 patients, and 11 124 and 9529 controls, respectively. The deletion showed nominal association in the whole Finland sample (OR=2.63 95%-CI: 1.28-5.59) but did not reach statistical significance in the non-Finnish samples (OR=2.17, 95%-CI: 0.81-5.80). The combined estimate for association with schizophrenia was with OR 1.84 (95%-CI 1.18-2.87, p=0.007).

The relationship between the 4q12 and 22q11.22 deletions and seven neurodevelopmental phenotypes utilized in Study II was analyzed in 4872 non-schizophrenic individuals of the NFBC 1966. Carriers of the 22q11.22 deletion had significantly more intellectual disability compared to non-carriers (OR=4.6, 95%-CI: 1.41-14.96, p=0.03), and an even greater overrepresentation of milder learning difficulties (OR= 5.9, 95%-CI: 1.78-8.94, p=0.003). However, although the deletion showed an increased risk for psychosis (excluding a diagnosis of schizophrenia), the test for association did not reach statistical significance (OR=4.1, 95%-CI: 0.55-30.17, p=0.2). No phenotypes were significantly overrepresented among the 4q12 deletion carriers.

Cognitive functioning was also evaluated among 19 schizophrenia patients of Finnish schizophrenia families that carried the 22q11.22 deletion and 547 patients without the deletion. The carriers performed slightly worse in tests of CVLT assessing verbal memory after short delay (B=-1.5, p=0.03), and showed similar tendency in other measures including processing speed and executive functioning, attention and working memory, although not reaching statistical significance. Four individuals were found to be homozygous for the 22q11.22 deletion. All four individuals displayed cognitive impairment that ranged from poor performance on tests of executive function and information processing to moderate intellectual disability. Two of them also had a diagnosis of schizophrenia.

The 22q11.2 region has not been implicated with schizophrenia in previous linkage studies conducted in the isolate or other Finnish schizophrenia families (Hovatta *et*

al. 1998, Hovatta *et al.* 1999, Ekelund *et al.* 2000, Paunio *et al.* 2001). The identified deletion spanned 240 kb and was ~500 kb distal to the most commonly deleted region in VCFS, overlapping the distal 22q11.2 deletion syndrome region (Figure 15) (Tan *et al.* 2011). The deletion flanked two genes, from which the other one was an immunoglobulin variant and the other one was a gene encoding for topoisomerase 3 β (*TOP3β*). The phenotypic impact of the deletion reported here likely derives from its effect on *TOP3β*. The deletion directly disrupts this gene, down regulates its mRNA levels, and does not alter transcript levels of other genes in the deletion region. Recent surveys of *de novo* mutations have yielded both direct and indirect (through biological networks) suggestive genetic evidence connecting *TOP3β* with neurodevelopmental phenotypes (O'Roak *et al.* 2012, Xu *et al.* 2012). A recent report described a single individual, of unspecified ethnic origin, who had a 240 kb deletion that appears to be identical to the one presented here, which demonstrated speech delay and minor physical abnormalities (Tan *et al.* 2011). More suggestively, Xu *et al.* recently reported a 22q11.22 deletion covering the same region, as well as a *de novo TOP3β* missense mutation in Afrikaner schizophrenia patients (Xu *et al.* 2012). Finally, an autistic patient with macrocephaly, cognitive impairment, and defective social behavior and language development carried truncating mutations in three genes (*RCAN1, CHD8* and *CUBN*) that are members of a β-catenin-linked pathway that also includes *TOP3β* (O'Roak *et al.* 2012).
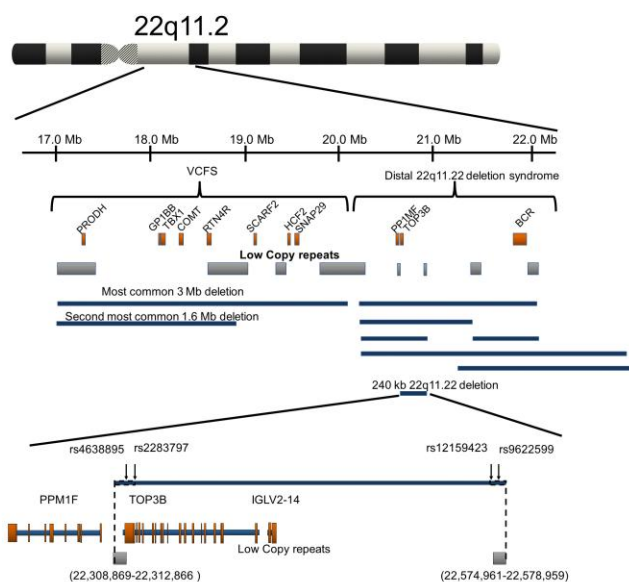
Figure 15.   Schematic presentation of the 22q11.22 deletion region. Deletions that are associated with VCFS and distal 22q11.2 deletion syndrome are shown in bold blue lines and the LCRs facilitating these mutation events are presented in grey. Known disease related OMIM genes are shown in bold orange lines. The break point associated SNPs are marked above the 240 kb deletion. The genes overlapped by the 240 kb 22q11.22 deletion are presented below the deletion, with exons presented as bold orange lines.

The deletion had a single dosage dependent association with the *TOP3β* expression in lymphoblastoid cell lines of 65 individuals (51 non-carriers, 12 heterozygous carriers, and 2 homozygous carriers) of 18 families ascertained from the Finnish schizophrenia family study. The expression level of *TOP3β* in non-carriers (269.99, 95%-CI=249.18-290.79) was twice that of heterozygotes (127.61, 95%-CI=103.92-151.30), while homozygous deletion carriers had no detectable transcript. The deletion had no genome-wide significant effect on mRNA levels of any other gene. However, 813 genes (858 of 9872 probes) showed nominal differences (p<0.05) in their respective mRNA levels between deletion carriers and non-carriers. From this set of 813 genes, we evaluated whether any functional gene category, plausibly affected by the down-regulation of *TOP3β* would be enriched among deletion carriers compared to non-carriers. Genes involved in two related categories, translation and ribosomal complexes displayed a 3.1 and a 2.8-fold enrichment, respectively, (p=6.2×10^{-8} and 2.8×10^{-10}), which could imply that the 22q11.22 deletion exerts its phenotypic effect via disturbed translational regulation (Table 7)

Table 7. Functional annotation of genes that differ in mRNA levels between the 22q11.22 deletion carriers and non-carriers.

Term	GO ID	%-of genes	Fold-enrichment	p-value
Biological process				
Translation	0006412	4.8	3.1	0.00014
Cellular component				
Ribonucleoprotein complex	0030529	7.0	2.8	1.10×10^{-7}
Cytosol	0005829	12.1	1.9	1.80×10^{-5}
Ribosome	0005840	3.1	3.1	0.0061
Ribosomal subunit	0044391	2.2	3.6	0.024
Molecular function				
mRNA binding	0003723	7.0	2	0.0037
Structural constituent of ribosome	0003735	2.7	3.4	0.013

Biochemical investigations revealed functions for the TOP3β protein that could explain the phenotypic effect of its deletion. The TOP3β protein formed a protein complex with FMRP via Tudor domain–containing protein 3 (TDRD 3) through the immunoprecipitation from Human Embryonic Kidney 293 cells (HEK293). FMRP is a polyribosome associated RNA-binding protein that inhibits the translation of pre- and post-synaptic RNAs (Laggerbauer *et al.* 2001, Darnell *et al.* 2011). The dysregulation of translation resulting from the inactivation of FMRP causes fragile X syndrome (FRAX, OMIM #300624) (Bassell *et al.* 2008). FRAX is the second most common form of intellectual disability after Down Syndrome and results from $(CGG)_n$ repeat mutations at the 5´UTR of *FMR1*, which results in hyper methylation of the repeat and the following CpG island. The hypermethylation further results in transcriptional silencing of the *FMR1* encoding for FMRP (Fu *et al.* 1991, Pieretti *et al.* 1991, Devys *et al.* 1993, Rousseau *et al.* 1995).

TOP3β was found to bind mRNA, which suggests that it is a component of cytosolic messenger ribonucleoproteins (mRNPs). The protein complex was also present in the neuronal cells of mouse brain lysates and TOP3β had similar intracellular localization and trafficking as the nucleo-cytoplasmic shuttling proteins FMRP and TDRD3. TOP3β is a type IA topoisomerase that alters DNA topology (Viard *et al.* 2007), However, it was found to be catalytically active also on RNA. Furthermore, the TOP3β-TDRD3-FMRP- complex was found to be a component of translating mRNPs, reported already earlier for FMRP (Siomi *et al.* 1996).

TDRD is multidomain protein that binds to the exon junction complex (EJC), a set of proteins that are deposited as a consequence of splicing upstream of mRNA exon-

exon junctions, and likely mediates the interaction between TOP3β and FMRP (Tange *et al.* 2004). The interaction of TDRD3 with EJC was not disrupted even though binding sites for FMRP or TOP3β were deleted, suggesting that it mediates the concomitant integration of TOP3β and FMRP into mRNPs. Furthermore, TDRD3 deficient cells also prevent TOP3β recruitment to polysomes, but had not effect on FMRP. Together these findings of topoisomerase activity on RNA substrates suggest that it is involved in the metabolism of FMRP-bound mRNAs.

The relevance of TOP3β in neurodevelopmental disorders were further supported by an accompanying study that also concluded that TOP3β is an RNA topoisomerase that interacts with FMRP. A Missensense mutation in FMR1 identified in a patient with the FRAX phenotype also eliminates the TDRD3 mediated TOP3β interaction with FMRP: a finding that further supports the role of this complex in the pathogenesis of mental disorders. Moreover, TOP3β was found to bind to mRNAs of multiple neuronal genes, also linked with schizophrenia and autism. In addition, synapse formation is defective in *Top3β* mutant flies and mice, just as it is defective in FMRP mutant files and mice, suggesting that TOP3β acts in concert with FMRP and they both are important for normal neuronal development (Xu 2013). In conclusion, TOP3β is an RNA topoisomerase that seems important for neurodevelopment via systems that involve interaction with FMRP, translational regulation and possibly RNA topoisomerase activity.

5.5 Limitations of the study

The most significant limitations of the studies described in this thesis are the limited sample sizes for complex disease genetics, lack of replication in study (II) and the technical challenges related to reliably identifying and genotyping CNVs from SNP array data.

5.5.1 Sample size and statistical power

Statistical confidence for deducting association signals is proportional to the sample size, allele frequency, disease prevalence, and the risk associated with the variants studied (Psychiatric Gwas Consortium Coordinating Committee *et al.* 2009). Due to limited samples sizes, the studies combined in this thesis were limited to studying variants of large effects (OR>2). Genotypic relative risks associated to common variants (frequency >5%) in GWAS have been commonly found to be below 1.2 and thus those variants were beyond the statistical power of the studies included here. On the other hand the rarity of the high risk variants introduces a major caveat for the statistical power and the genome-wide significance level ($P = 5\times10^{-8}$), which

could not been reached in any of the studies in this thesis. This is despite that the risks associated with the variants were markedly larger than those generally expected for complex traits (OR up to 14). In order to limit the number of statistical comparisons and need for statistical correction the studies were limited to investigating a small proportion of the total variation at each time. In study I only *de novo* CNV sites were included which limited the number of comparisons to 66, corresponding to Bonferroni corrected p-value of 7.6×10^{-4}. Similarly, the number of tests was controlled in studies II and III. In study II a global burden of only very large CNVs was estimated rather than studying individual variants. Study III was limited to rare relatively large CNVs that were more common in the isolate population.

5.5.2 Replication

Due to the stochastic correlations, findings should be replicated in additional sample sets to confirm the preliminary association result. In studies I and III replications for the preliminary findings are provided in larger sample sets. However, a lack of within-study replication is a shortage in study II. The results of study II, therefore must be considered within the context of the body of evidence emerging in the field.

5.5.3 Technical limitations

The quantitative measuring of DNA copy number is still challenging due to the high levels of stochastic variation in the marker intensities on a SNP array. The stochastic error can be reduced by requiring multiple probes (e.g. 10 or more) to assign each CNV and also by validating the CNV genotypes by using complimentary methods, such as quantitative real-time PCR (used in studies I and III), or FISH (used in study I). Nonetheless many of these complimentary methods are likewise unreliable and laborious to conduct. Moreover, requirement of multiple probes on a GWAS array to define a CNV will limit the detection of smaller CNVs. If an average SNP density in a GWAS array would be 1 SNP/kb, when requiring more than 10 probes to assign a valid CNV call this will already limit the detection of CNVs below 10-20 kb. Also genomic regions enriched with CNVs are prone to false positive calls due to repetitive sequences and are also less well covered in the genotyping arrays used. One way to control for false calls is by manually inspecting the intensity distribution plots.

5.5.4 Phenotyping

Genetic studies often start by defining a phenotype and then searching genotypes that account for some part of the variation observed for the defined phenotype.

However, when considering psychiatric traits, they are often imprecisely defined in a biologically meaningful way and difficult to assess objectively. This can restrict the accuracy by which they can be designated as being shared between individuals. Study II approached this question by using the genotype (carrier of large deletions) as a starting point in the analysis and then looking what phenotypic features the carriers share with each other. The idea being that in contrast to the phenotype, the genotype can be objectively and precisely measured. This "phenotype mining" method requires consideration of evaluating the significance of associations detected between traits and genotypes. While the genetic space is well defined, the total number of phenotypes cannot be generically described and nor can their degree of independence from each other be easily defined. For these reasons, the analysis in study II was limited to seven phenotype categories postulated to be relevant for neuronal development.

5.5.5 Population stratification

Population stratification is a major confounder of association analysis and affects analyses of both rare and common variants. Study I focused on *de novo* variants and although the study sample consisted of multiple populations, the mutation rates can be assumed to be similar across populations. In addition, variants associated with very large dominant effects, such as identified in the study I, are not likely to grow in frequency even in young founder populations, such as in the Finnish population. Study II was a geographically based population study, which does not involve stratification caused by sampling to cases and controls. Study III utilizes internal genetic structure to identify variants possibly contributing to schizophrenia. In this study the stratification was controlled by using parental birth places of the participants and using MHC statistics designed for stratified analysis. It is, however, still likely that the associations observed, particularly involving the whole Finland sample are also affected by population sub-structure.

5.5.6 Clinical and biological interpretation of the findings

The genetic findings of this study remain clinically unspecific. An OMIM ID has been granted for the deletions identified in study I and in some cases these may be of diagnostic importance. However, the clinical picture related to the carriers of these deletions varies substantially and they also include incomplete penetrance. For these reasons, clinical findings involving these deletions should still be interpreted with caution. In study III, a fully penetrant recessive effect of *TOP3β* deletion was observed. Yet, this finding was observed with four individuals and was associated to heterogenic phenotype, which is why clinical implications should also in this case be

interpreted cautiously. The exact biological mechanisms related to the genetic findings in this thesis likewise remain incompletely understood. In many cases the associated deletions overlap multiple genes and the causative gene or genes are not known. Study III provides most detailed investigation of the biological relevance of the finding. However, the direct mechanism related to the *TOP3β* deletion and the neurodevelopmental phenotypes also remains uncertain.

6 Conclusions

We have identified large genomic deletions that predispose to a set of phenotypes, that are hypothesized to reflect abnormal neuronal development, including schizophrenia, intellectual disability, and milder learning difficulties. In addition to the global burden of large deletions in neuronal development, we specifically identified four genomic deletions at 1q21.1, 15q11.2, 15q13.3 and 22q11.22 (*TOP3β*) that were associated with schizophrenia and impaired intellectual functioning. Of the deletions three (1q21.1, 15q12.1, 15q13.3) represent recurrent mutations, and even though they can also be inherited, as in the case of 1q21.1 deletion (Mefford *et al.* 2008), they often occur *de novo* in patients. The fourth deletion, overlapping *TOP3β* is a founder mutation that because of recent population bottlenecks has become enriched in Finland. In contrast to the other three deletions, the *TOP3β* deletion was found to have a particularly strong recessive effect. Biologically, the genes overlapped by the four deletions suggest the involvement of synaptic genes with prominent roles in the development and maintenance of neuronal networks to be relevant in schizophrenia. Furthermore, the regulation of synaptic translation is suggested to be of relevance in the etiology of schizophrenia and this implication is especially supported by the deletion of *TOP3β* and its association with FMRP. The role of the FMRP pathway has been further supported by findings from other studies of CNVs, as well as rare and *de novo* sequence variants in schizophrenia (Kirov *et al.* 2012, Ripke *et al.* 2013, Fromer *et al.* 2014, Purcell *et al.* 2014).

Three major themes emerge from the results of this study. First, a small portion of the genetic risk of schizophrenia is due to rare high risk variants, namely large chromosomal deletions. The second theme is the links between developmental phenotypes, namely cognitive functions and mental disorders. All of the deletions associated with schizophrenia described in this thesis have also been linked with intellectual functioning by the present study and also by studies conducted by other groups (Malhotra *et al.* 2012, Stefansson *et al.* 2014). Importantly, the results imply the role of large deletions, including those associated with schizophrenia, and also in milder intellectual deficits that can manifest as learning difficulties. These observations agree well with decreased cognitive performance reported in schizophrenia (Heinrichs *et al.* 1998, Mesholam-Gately *et al.* 2009, Fioravanti *et al.* 2012). Third, the genetic studies are moving towards the systematic assessment of rare variants in complex traits, thus different study designs may be adopted. For example, in the present study we have demonstrated the potential usefulness of large population samples in studying the relationship of rare genetic variants across

different diagnostic entities – a method we have named phenotype mining. In addition, as demonstrated earlier for Mendelian disorders, our findings of the *TOP3β* deletion support that isolated founder populations can also assist studies of rare variants in complex traits, such as schizophrenia and intellectual disability.

As a final comment, the synthesis from the wealth of genetic evidence from both common and rare variants accumulated during the last 10 years including the results presented in this thesis indisputably demonstrate that genetic susceptibility is shared across psychiatric diagnostic boundaries (Figure 16). The strongest correlations arises between schizophrenia, bipolar disorder, and major depressive disorder primarily accounted for by common variants (Cross-Disorder Group of the Psychiatric Genomics *et al.* 2013). The findings from rare variants support shared genetic factors particularly between schizophrenia, autism, and intellectual disability (Fromer *et al.* 2014, Kirov *et al.* 2014, Purcell *et al.* 2014). However, less evidence supports a major role for rare high risk CNVs in affective psychosis including bipolar disorder, and the impact of the CNVs studied in this thesis seems to be strongest in the early onset cases (Zhang *et al.* 2009, Grozeva *et al.* 2010, Malhotra *et al.* 2011, McQuillin *et al.* 2011, Malhotra *et al.* 2012, Priebe *et al.* 2012, Grozeva *et al.* 2013).
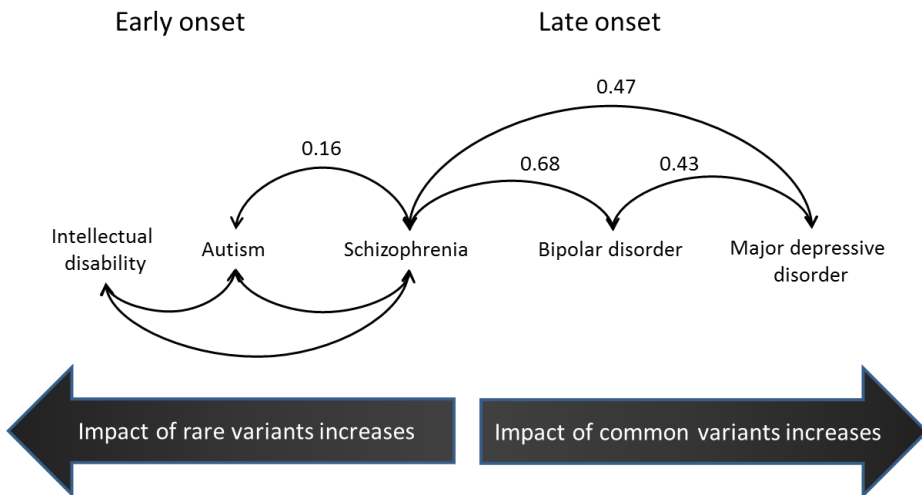

Figure 16.   Genetic sharing between psychiatric disorders. Evidence from GWAs studies, studies of rare and de novo CNV studies and sequencing studies.

The question arises: Why do the same genetic variants lead to such variable phenotypic presentations and significantly variable age of onset? One obvious answer is that studies have mostly estimated the impact of gene variants to

psychiatric end-diagnosis, but not to those biological processes more directly affected by the gene variants. Moreover, individual gene variants do not act in a vacuum but in combination with other possibly modifying genetic factors and environment. A plethora of environmental factors including viral infections, birth complications, early life insults, famine, etc. have been associated with increased risk for schizophrenia (Sullivan 2005). It has been speculated that the genetic risk for schizophrenia could emerge from a cumulative risk of thousands of low risk variants on a pathway rather than from a single node on the pathway (Sullivan 2012). This suggestion is consistent with the fact that no Mendelian form of schizophrenia has been identified – even in childhood onset forms of schizophrenia. Moreover, rare disruptive mutations in genes with common variants associating with schizophrenia such as *NRXN1, CACNA1C*, and *TCF4*, do not cause schizophrenia but rather they cause intellectual disability or autism (Sullivan 2012).

Considering the genetic evidence to date for psychiatric disorders, the common variants seem to play a more prominent role in later (early adulthood) onset disorders, whereas rare high risk variants are identified more often in patients with early onset disorders. For example, the load of rare disruptive CNVs is even higher in childhood onset of schizophrenia compared to the general form of schizophrenia (Ahn *et al.* 2014). This seems to be evolutionarily consistent with the associated reduced fecundity with the most severe forms of mental disorders. Current genetic evidence suggest that the effect of one highly pathogenic variant, such as a high risk CNV, is not biologically equal to the cumulative sum of multiple less pathogenic variants, such as common SNPs with small effects. This is analogous to the situation whereby driving a car into a wall once at 100 km/h is not same as driving a car 10 times into a wall at 10 km/h. The consequences are quite different. When translating this analogy and applying it to the genetic risk for neurodevelopmental disorders, it is tempting to speculate that if a variant exceeds a certain threshold of pathogenicity, it will exert its effect at an early stage of "crude" neurodevelopment resulting more often in early onset disorders. However, the cumulative sum of variants with low pathogenicity (polygenic score of common variants) will not exceed this threshold (at least not very often) and would thus express its effect during a later developmental "fine-tuning" stage, which will more often result in later onset disorders.

7 Concluding remarks and future prospects

What have we learned from the genetic studies of schizophrenia? First of all, we now know that schizophrenia is a biological disorder to begin with. Schizophrenia is as biological as any somatic disorders that are generally perceived as being biological in nature. In fact schizophrenia and other severe mental disorders have generally the highest heritability estimates compared to other complex diseases and these have now started to be formulated into specific biological pathways. It should, however, be emphasized that this does not mean that schizophrenia is *solely* a genetic disorder; just as any other complex diseases are not solely genetic for that matter. In addition to the genetic predisposition for schizophrenia, a number of environmental factors, including pre and postnatal, social, childhood and later life events are associated with an increased risk for schizophrenia. The currently widely accepted model perceives schizophrenia to be a neurodevelopmental disorder where the risk formulates gradually in concert with genetic makeup together with social and environmental factors (Rapoport *et al.* 2012, Howes *et al.* 2014). The second lesson is that the genetic architecture of schizophrenia is very similar to other complex traits, including human height and many complex diseases that pile up mostly from a multitude of risk variants; each of which has a modest effect. Third, genetic associations enriched in genes expressed in the brain and neurons, demonstrated that schizophrenia can truly be considered as a neuronal disorder, not merely a disorder that affects the brain. With the discovery of neuroleptic drugs a view emerged of schizophrenia being a disorder of dopaminergic system, and then later on the glutaminergic dysfunction hypothesis emerged (Coyle 1996). Genetic evidence further supports that these pathways are of etiological relevance for the disease. Further, since these systems are known therapeutic targets, it highlights the possibility to extract also novel therapeutic insights from the present genetic discoveries. Finally, the genetic findings support shared biological etiology for different psychiatric disorders, and challenges the current categorized clinical view.

Due to the technical advances, the field of genetic research has evolved rapidly from genotyping several tens or hundreds of genetic markers to genotyping millions of them simultaneously. Practically all common variation in the human genome can be interrogated efficiently in one GWAS experiment. Population level sequence data obtained from large sequencing studies, such as 1000 genomes (Durbin *et al.* 2011) and UK10K (http://www.uk10k.org/), both of which also include Finnish samples, can be used to impute the SNP array data to include also rare sequence variants.

Here the unique population history and reduced genetic diversity of Finns can be of great scientific value. Thousands of individuals from the Finnish population cohorts have GWAS data available and a large reference sequence data, gathered from different research projects, collectively called the Sequencing Intiative Suomi (SISu-project) is already available (http://www.nationalbiobanks.fi/ , http://sisu.fimm.fi/). The current trend in genetics has moved beyond GWAS to high coverage exome and whole genome sequencing of even larger sample sets. The accumulation of genetic data in the future will provide grounds for systematically addressing the impact of especially rare variants in various phenotypes including those related to severe mental disorders. For studies of common variants, special populations have provided limited benefit, but for rare variants this is likely to change.

Genetics has opened up new avenues in studies of mental health that allows the postulation of novel hypotheses about the biological pathophysiology of severe mental disorders, which has previously been largely unattained. Yet genetics alone is insufficient but rather it contributes another stream of convergent evidence to merge with those of from multiple disciplines, including molecular biology and neuroscience, which will be crucial in illuminating the biological mechanism leading to these disorders.

8 Acknowledgements

on the way. I have been extremely lucky to have known and worked with such a remarkable person as Leena and I miss her dearly.

A huge thank you to Professor Cathryn Lewis for accepting the role of the Opponent of my defense despite of her busy schedule. Also, I want to thank Håkan Karlsson and Professor Helena Kääriäinen for their valuable comments during the revision of this thesis. Alisdair Mclean is acknowledged for the language editing of this thesis.

I have been extremely fortunate to have been given a chance to work in a network of world-class scientists both in Finland and abroad. Professor Nelson Freimer is warmly thanked for taking such a huge interest on my work and helping me countless times throughout the many years. Nelson's support and rationality has been invaluable for finishing this work. I want to express my gratitude to all the co-authors in the publications included in this thesis. Especially I wish to express my appreciation and gratitude to Professors Jouko Lönnqvist, Mark Daly, Jaakko Kaprio, Tiina Paunio, and Matti Isohanni, and Adjunct Professors Annamari Tuulio-Henriksson and Teppo Varilo. Professor Samuli Ripatti is thanked for all his help and support particularly in statistical questions and for accepting the role of Kustos of the dissertation. I want to thank all our clinical collaborators. Without their excellent and persistent work this study would not have been possible.

I have been privileged to work with amazing group of people with whom I have shared wonderful times in and outside the office. People from the former Psycho group back in KTL: Anu L., Emma N., Emma P., Juho, Katri, Liisa, Marika, and William are thanked for the fun and inspirational working environment. A special thanks to Joni, who laid the foundation for my interest in psychiatric genetics and research.

I wish to acknowledge all the wonderful people who are responsible for the invigorating and supportive atmosphere at work that has made it so pleasant to come to work every day: Alfredo, Anja, Anne, Anni, Anniina, Ansku, Antti-Pekka, Anu, Annu, Christian, Diana, Elmo, Elisbaeth, Emmi, Eveliina, Elli, Esther, Hanna, Hannele, Henna, Himanshu, Huei-yi, Ida, Jaakko, Joel, Johannes, Jussi, Kaisa, Karola, Katta, Liisa A., Liisa T., Liisa X. Luca, Lucy, Mari, Marjis, Marjo, Matti, Marine, Markus P., Meri, Miisa, Minttu, Minna, Mikko, Mitja, Mervi, Outi, Padhdraig, Perttu, Priit, Pekka, both Peter W., Pietari, PP, Päivi, Perttu, Siv, Suvi, Tea, Tero, Tintti, Tiia, Virpi, Verneri and many others I am bound to have forgotten to mention by name. A special thanks to Karola, Jonna, and Helena from the fantastic scientific and not so scientific discussions and endless planning of new projects over the years. Jarkko, Annu, Miisa, and Taru are especially thanked for their friendship and support.

9 References

Ahn, K., Gotay, N., Andersen, T. M., Anvari, A. A., Gochman, P., Lee, Y., Sanders, S., Guha, S., Darvasi, A., Glessner, J. T., Hakonarson, H., Lencz, T., State, M. W., Shugart, Y. Y. and Rapoport, J. L. (2014). High rate of disease-related copy number variations in childhood onset schizophrenia. *Mol Psychiatry*. 19, 568-72

Allen, N. C., Bagade, S., McQueen, M. B., Ioannidis, J. P., Kavvoura, F. K., Khoury, M. J., Tanzi, R. E. and Bertram, L. (2008). Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat Genet*. 40, 827-34

Altshuler, D., Hirschhorn, J. N., Klannemark, M., Lindgren, C. M., Vohl, M. C., Nemesh, J., Lane, C. R., Schaffner, S. F., Bolk, S., Brewer, C., Tuomi, T., Gaudet, D., Hudson, T. J., Daly, M., Groop, L. and Lander, E. S. (2000). The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet*. 26, 76-80

Altshuler, D., Daly, M. J. and Lander, E. S. (2008). Genetic mapping in human disease. *Science*. 322, 881-8

American Psychiatric Association (1994). DSM IV: diagnostic and statistical manual of mental disorders.

Amiel, J., Rio, M., de Pontual, L., Redon, R., Malan, V., Boddaert, N., Plouin, P., Carter, N. P., Lyonnet, S., Munnich, A. and Colleaux, L. (2007). Mutations in TCF4, encoding a class I basic helix-loop-helix transcription factor, are responsible for Pitt-Hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction. *Am J Hum Genet*. 80, 988-93

Amos-Landgraf, J. M., Ji, Y., Gottlieb, W., Depinet, T., Wandstrat, A. E., Cassidy, S. B., Driscoll, D. J., Rogan, P. K., Schwartz, S. and Nicholls, R. D. (1999). Chromosome breakage in the Prader-Willi and Angelman syndromes involves recombination between large, transcribed repeats at proximal and distal breakpoints. *Am J Hum Genet*. 65, 370-86

Andreasen, N. (1983). The Scale for the Assessment of Negative Symptoms (SANS).University of Iowa, Iowa City, IA, USA

Andreasen, N. (1984). The scale for the assessment of positive symptoms (SAPS). *University of Iowa, Iowa City*.

Andreasen, N. C. (1995). Symptoms, signs, and diagnosis of schizophrenia. *Lancet*. 346, 477-81.

Andreasen, N. C. (1997). Linking mind and brain in the study of mental illnesses: a project for a scientific psychopathology. *Science*. 275, 1586-93

Arajarvi, R., Suvisaari, J., Suokas, J., Schreck, M., Haukka, J., Hintikka, J., Partonen, T. and Lonnqvist, J. (2005). Prevalence and diagnosis of schizophrenia based on register, case record and interview data in an isolated Finnish birth cohort born 1940-1969. *Soc Psychiatry Psychiatr Epidemiol*. 40, 808-16

Aromaa, A. and Koskinen, S. (2004). Health and functional capacity in Finland. Baseline Results of the Health 2000 Health Examination Survey.Publications of the National Public Health Institute

Autism Genome Project, C., Szatmari, P., Paterson, A. D., Zwaigenbaum, L., Roberts, W., Brian, J., Liu, X. Q., Vincent, J. B., Skaug, J. L., Thompson, A. P., Senman, L., Feuk, L., Qian, C., Bryson, S. E., Jones, M. B., Marshall, C. R., Scherer, S. W., Vieland, V. J., Bartlett, C., Mangin, L. V., Goedken, R., Segre, A., Pericak-Vance, M. A., Cuccaro, M. L., Gilbert, J. R., Wright, H. H., Abramson, R. K., Betancur, C., Bourgeron, T., Gillberg, C., Leboyer, M., Buxbaum, J. D., Davis, K. L., Hollander, E., Silverman, J. M., Hallmayer, J., Lotspeich, L., Sutcliffe, J. S., Haines, J. L., Folstein, S. E., Piven, J., Wassink, T. H., Sheffield, V., Geschwind, D. H., Bucan, M., Brown, W. T., Cantor, R. M., Constantino, J. N., Gilliam, T. C., Herbert, M., Lajonchere, C., Ledbetter, D. H., Lese-Martin, C., Miller, J., Nelson, S., Samango-Sprouse, C. A., Spence, S., State, M., Tanzi, R. E., Coon, H., Dawson, G., Devlin, B., Estes, A., Flodman, P., Klei, L., McMahon, W. M., Minshew, N., Munson, J., Korvatska, E., Rodier, P. M., Schellenberg, G. D., Smith, M., Spence, M. A., Stodgell, C., Tepper, P. G., Wijsman, E. M., Yu, C. E., Roge, B., Mantoulan, C., Wittemeyer, K., Poustka, A., Felder, B., Klauck, S. M., Schuster, C., Poustka, F., Bolte, S., Feineis-Matthews, S., Herbrecht, E., Schmotzer, G., Tsiantis, J., Papanikolaou, K., Maestrini, E., Bacchelli, E., Blasi, F., Carone, S., Toma, C., Van Engeland, H., de Jonge, M., Kemner, C., Koop, F., Langemeijer, M., Hijmans, C., Staal, W. G., Baird, G., Bolton, P. F., Rutter, M. L., Weisblatt, E., Green, J., Aldred, C., Wilkinson, J. A., Pickles, A., Le Couteur, A., Berney, T., McConachie, H., Bailey, A. J., Francis, K., Honeyman, G., Hutchinson, A., Parr, J. R., Wallace, S., Monaco, A. P., Barnby, G., Kobayashi, K., Lamb, J. A., Sousa, I., Sykes, N., Cook, E. H., Guter, S. J., Leventhal, B. L., Salt, J., Lord, C., Corsello, C., Hus, V., Weeks, D. E., Volkmar, F., Tauber, M., Fombonne, E., Shih, A. and Meyer, K. J. (2007). Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet*. 39, 319-28

Bailey, J. A., Yavor, A. M., Massa, H. F., Trask, B. J. and Eichler, E. E. (2001). Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res*. 11, 1005-17

Barch, D. M., Bustillo, J., Gaebel, W., Gur, R., Heckers, S., Malaspina, D., Owen, M. J., Schultz, S., Tandon, R., Tsuang, M., Van Os, J. and Carpenter, W. (2013). Logic and justification for dimensional assessment of symptoms and related clinical phenomena in psychosis: relevance to DSM-5. *Schizophr Res*. 150, 15-20

Barker, D. J., Osmond, C., Forsen, T. J., Kajantie, E. and Eriksson, J. G. (2005). Trajectories of growth among children who have coronary events as adults. *The New England journal of medicine*. 353, 1802-9

Barrett, J. C., Fry, B., Maller, J. and Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 21, 263-5

Bassell, G. J. and Warren, S. T. (2008). Fragile X syndrome: loss of local mRNA regulation alters synaptic development and function. *Neuron*. 60, 201-14

Ben-Shachar, S., Lanpher, B., German, J. R., Qasaymeh, M., Potocki, L., Nagamani, S. C., Franco, L. M., Malphrus, A., Bottenfield, G. W., Spence, J. E., Amato, S., Rousseau, J. A., Moghaddam, B., Skinner, C., Skinner, S. A., Bernes, S., Armstrong, N., Shinawi, M., Stankiewicz, P., Patel, A., Cheung, S. W., Lupski, J. R., Beaudet, A. L. and Sahoo, T. (2009). Microdeletion 15q13.3: a locus with incomplete penetrance for autism, mental retardation, and psychiatric disorders. *J Med Genet*. 46, 382-8

Bergen, S. E., O'Dushlaine, C. T., Ripke, S., Lee, P. H., Ruderfer, D. M., Akterin, S., Moran, J. L., Chambert, K. D., Handsaker, R. E., Backlund, L., Osby, U., McCarroll, S., Landen, M., Scolnick, E. M., Magnusson, P. K., Lichtenstein, P., Hultman, C. M., Purcell, S. M., Sklar, P. and Sullivan, P. F. (2012). Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder. *Mol Psychiatry*. 17, 880-6

Berkel, S., Marshall, C. R., Weiss, B., Howe, J., Roeth, R., Moog, U., Endris, V., Roberts, W., Szatmari, P., Pinto, D., Bonin, M., Riess, A., Engels, H., Sprengel, R., Scherer, S. W. and Rappold, G. A. (2010). Mutations in the SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. *Nat Genet*. 42, 489-91

Bertina, R. M., Koeleman, B. P., Koster, T., Rosendaal, F. R., Dirven, R. J., de Ronde, H., van der Velden, P. A. and Reitsma, P. H. (1994). Mutation in blood coagulation factor V associated with resistance to activated protein C. *Nature*. 369, 64-7

Bignell, G. R., Huang, J., Greshock, J., Watt, S., Butler, A., West, S., Grigorova, M., Jones, K. W., Wei, W., Stratton, M. R., Futreal, P. A., Weber, B., Shapero, M. H. and Wooster, R. (2004). High-resolution analysis of DNA copy number using

oligonucleotide microarrays. *Genome Res*. 14, 287-95

Boomsma, D., Busjahn, A. and Peltonen, L. (2002). Classical twin studies and beyond. *Nat Rev Genet*. 3, 872-82

Botstein, D., White, R. L., Skolnick, M. and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet*. 32, 314-31.

Boyar, F. Z., Whitney, M. M., Lossie, A. C., Gray, B. A., Keller, K. L., Stalker, H. J., Zori, R. T., Geffken, G., Mutch, J., Edge, P. J., Voeller, K. S., Williams, C. A. and Driscoll, D. J. (2001). A family with a grand-maternally derived interstitial duplication of proximal 15q. *Clinical genetics*. 60, 421-30

Brown, S. (1997). Excess mortality of schizophrenia. A meta-analysis. *British Journal of Psychiatry*. 171, 502-508

Browning, S. R. and Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet*. 81, 1084-97

Brunetti-Pierri, N., Berg, J. S., Scaglia, F., Belmont, J., Bacino, C. A., Sahoo, T., Lalani, S. R., Graham, B., Lee, B., Shinawi, M., Shen, J., Kang, S. H., Pursley, A., Lotze, T., Kennedy, G., Lansky-Shafer, S., Weaver, C., Roeder, E. R., Grebe, T. A., Arnold, G. L., Hutchison, T., Reimschisel, T., Amato, S., Geragthy, M. T., Innis, J. W., Obersztyn, E., Nowakowska, B., Rosengren, S. S., Bader, P. I., Grange, D. K., Naqvi, S., Garnica, A. D., Bernes, S. M., Fong, C. T., Summers, A., Walters, W. D., Lupski, J. R., Stankiewicz, P., Cheung, S. W. and Patel, A. (2008). Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nat Genet*. 40, 1466-71

Camp, B. W., Broman, S. H., Nichols, P. L. and Leff, M. (1998). Maternal and neonatal risk factors for mental retardation: defining the 'at-risk' child. *Early human development*. 50, 159-73

Cannon, M., Jones, P., Huttunen, M. O., Tanskanen, A., Huttunen, T., Rabe-Hesketh, S. and Murray, R. M. (1999). School performance in Finnish children and later development of schizophrenia: a population-based longitudinal study. *Arch Gen Psychiatry*. 56, 457-63

Cannon, T. D., Kaprio, J., Lonnqvist, J., Huttunen, M. and Koskenvuo, M. (1998). The genetic epidemiology of schizophrenia in a Finnish twin cohort. A population-based modeling study. *Arch Gen Psychiatry*. 55, 67-74

Cassidy, S. B., Schwartz, S., Miller, J. L. and Driscoll, D. J. (2012). Prader-Willi syndrome. *Genet Med*. 14, 10-26

Chapman, D. A., Scott, K. G. and Mason, C. A. (2002). Early risk factors for mental retardation: role of maternal age and maternal education. *Am J Ment Retard*. 107, 46-59

Chelly, J. and Mandel, J. L. (2001). Monogenic causes of X-linked mental retardation. *Nat Rev Genet*. 2, 669-80

Cheung, V. G., Nowak, N., Jang, W., Kirsch, I. R., Zhao, S., Chen, X. N., Furey, T. S., Kim, U. J., Kuo, W. L., Olivier, M., Conroy, J., Kasprzyk, A., Massa, H., Yonescu, R., Sait, S., Thoreen, C., Snijders, A., Lemyre, E., Bailey, J. A., Bruzel, A., Burrill, W. D., Clegg, S. M., Collins, S., Dhami, P., Friedman, C., Han, C. S., Herrick, S., Lee, J., Ligon, A. H., Lowry, S., Morley, M., Narasimhan, S., Osoegawa, K., Peng, Z., Plajzer-Frick, I., Quade, B. J., Scott, D., Sirotkin, K., Thorpe, A. A., Gray, J. W., Hudson, J., Pinkel, D., Ried, T., Rowen, L., Shen-Ong, G. L., Strausberg, R. L., Birney, E., Callen, D. F., Cheng, J. F., Cox, D. R., Doggett, N. A., Carter, N. P., Eichler, E. E., Haussler, D., Korenberg, J. R., Morton, C. C., Albertson, D., Schuler, G., de Jong, P. J., Trask, B. J. and Consortium, B. A. C. R. (2001). Integration of cytogenetic landmarks into the draft sequence of the human genome. *Nature*. 409, 953-8

Christian, S. L., Fantes, J. A., Mewborn, S. K., Huang, B. and Ledbetter, D. H. (1999). Large genomic duplicons map to sites of instability in the Prader-Willi/Angelman syndrome chromosome region (15q11-q13). *Hum Mol Genet*. 8, 1025-37

Cichon, S., Muhleisen, T. W., Degenhardt, F. A., Mattheisen, M., Miro, X., Strohmaier, J., Steffens, M., Meesters, C., Herms, S., Weingarten, M., Priebe, L., Haenisch, B., Alexander, M., Vollmer, J., Breuer, R., Schmal, C., Tessmann, P., Moebus, S., Wichmann, H. E., Schreiber, S., Muller-Myhsok, B., Lucae, S., Jamain, S., Leboyer, M., Bellivier, F., Etain, B., Henry, C., Kahn, J. P., Heath, S., Bipolar Disorder Genome Study, C., Hamshere, M., O'Donovan, M. C., Owen, M. J., Craddock, N., Schwarz, M., Vedder, H., Kammerer-Ciernioch, J., Reif, A., Sasse, J., Bauer, M., Hautzinger, M., Wright, A., Mitchell, P. B., Schofield, P. R., Montgomery, G. W., Medland, S. E., Gordon, S. D., Martin, N. G., Gustafsson, O., Andreassen, O., Djurovic, S., Sigurdsson, E., Steinberg, S., Stefansson, H., Stefansson, K., Kapur-Pojskic, L., Oruc, L., Rivas, F., Mayoral, F., Chuchalin, A., Babadjanova, G., Tiganov, A. S., Pantelejeva, G., Abramova, L. I., Grigoroiu-Serbanescu, M., Diaconu, C. C., Czerski, P. M., Hauser, J., Zimmer, A., Lathrop, M., Schulze, T. G., Wienker, T. F., Schumacher, J., Maier, W., Propping, P., Rietschel, M. and Nothen, M. M. (2011). Genome-wide association study identifies genetic variation in neurocan as a susceptibility factor for bipolar disorder. *Am J Hum Genet*. 88, 372-81

Colella, S., Yau, C., Taylor, J. M., Mirza, G., Butler, H., Clouston, P., Bassett, A. S., Seller, A., Holmes, C. C. and Ragoussis, J. (2007). QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res*. 35, 2013-25

Collins, F. S., Guyer, M. S. and Charkravarti, A. (1997). Variations on a theme: cataloging human DNA sequence variation. *Science*. 278, 1580-1

Conrad, D. F., Andrews, T. D., Carter, N. P., Hurles, M. E. and Pritchard, J. K. (2006). A high-resolution survey of deletion polymorphism in the human genome. *Nat Genet*. 38, 75-81

Conrad, D. F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T. D., Barnes, C., Campbell, P., Fitzgerald, T., Hu, M., Ihm, C. H., Kristiansson, K., Macarthur, D. G., Macdonald, J. R., Onyiah, I., Pang, A. W., Robson, S., Stirrups, K., Valsesia, A., Walter, K., Wei, J., Tyler-Smith, C., Carter, N. P., Lee, C., Scherer, S. W. and Hurles, M. E. (2009). Origins and functional impact of copy number variation in the human genome. *Nature*.

Cooper, G. M., Coe, B. P., Girirajan, S., Rosenfeld, J. A., Vu, T. H., Baker, C., Williams, C., Stalker, H., Hamid, R., Hannig, V., Abdel-Hamid, H., Bader, P., McCracken, E., Niyazov, D., Leppig, K., Thiese, H., Hummel, M., Alexander, N., Gorski, J., Kussmann, J., Shashi, V., Johnson, K., Rehder, C., Ballif, B. C., Shaffer, L. G. and Eichler, E. E. (2011). A copy number variation morbidity map of developmental delay. *Nat Genet*. 43, 838-46

Coyle, J. T. (1996). The glutamatergic dysfunction hypothesis for schizophrenia. *Harvard review of psychiatry*. 3, 241-53

Cross-Disorder Group of the Psychiatric Genomics, C. and Genetic Risk Outcome of Psychosis, C. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet*. 381, 1371-9

Cross-Disorder Group of the Psychiatric Genomics, C., Lee, S. H., Ripke, S., Neale, B. M., Faraone, S. V., Purcell, S. M., Perlis, R. H., Mowry, B. J., Thapar, A., Goddard, M. E., Witte, J. S., Absher, D., Agartz, I., Akil, H., Amin, F., Andreassen, O. A., Anjorin, A., Anney, R., Anttila, V., Arking, D. E., Asherson, P., Azevedo, M. H., Backlund, L., Badner, J. A., Bailey, A. J., Banaschewski, T., Barchas, J. D., Barnes, M. R., Barrett, T. B., Bass, N., Battaglia, A., Bauer, M., Bayes, M., Bellivier, F., Bergen, S. E., Berrettini, W., Betancur, C., Bettecken, T., Biederman, J., Binder, E. B., Black, D. W., Blackwood, D. H., Bloss, C. S., Boehnke, M., Boomsma, D. I., Breen, G., Breuer, R., Bruggeman, R., Cormican, P., Buccola, N. G., Buitelaar, J. K., Bunney, W. E.,

Buxbaum, J. D., Byerley, W. F., Byrne, E. M., Caesar, S., Cahn, W., Cantor, R. M., Casas, M., Chakravarti, A., Chambert, K., Choudhury, K., Cichon, S., Cloninger, C. R., Collier, D. A., Cook, E. H., Coon, H., Cormand, B., Corvin, A., Coryell, W. H., Craig, D. W., Craig, I. W., Crosbie, J., Cuccaro, M. L., Curtis, D., Czamara, D., Datta, S., Dawson, G., Day, R., De Geus, E. J., Degenhardt, F., Djurovic, S., Donohoe, G. J., Doyle, A. E., Duan, J., Dudbridge, F., Duketis, E., Ebstein, R. P., Edenberg, H. J., Elia, J., Ennis, S., Etain, B., Fanous, A., Farmer, A. E., Ferrier, I. N., Flickinger, M., Fombonne, E., Foroud, T., Frank, J., Franke, B., Fraser, C., Freedman, R., Freimer, N. B., Freitag, C. M., Friedl, M., Frisen, L., Gallagher, L., Gejman, P. V., Georgieva, L., Gershon, E. S., Geschwind, D. H., Giegling, I., Gill, M., Gordon, S. D., Gordon-Smith, K., Green, E. K., Greenwood, T. A., Grice, D. E., Gross, M., Grozeva, D., Guan, W., Gurling, H., De Haan, L., Haines, J. L., Hakonarson, H., Hallmayer, J., Hamilton, S. P., Hamshere, M. L., Hansen, T. F., Hartmann, A. M., Hautzinger, M., Heath, A. C., Henders, A. K., Herms, S., Hickie, I. B., Hipolito, M., Hoefels, S., Holmans, P. A., Holsboer, F., Hoogendijk, W. J., Hottenga, J. J., Hultman, C. M., Hus, V., Ingason, A., Ising, M., Jamain, S., Jones, E. G., Jones, I., Jones, L., Tzeng, J. Y., Kahler, A. K., Kahn, R. S., Kandaswamy, R., Keller, M. C., Kennedy, J. L., Kenny, E., Kent, L., Kim, Y., Kirov, G. K., Klauck, S. M., Klei, L., Knowles, J. A., Kohli, M. A., Koller, D. L., Konte, B., Korszun, A., Krabbendam, L., Krasucki, R., Kuntsi, J., Kwan, P., Landen, M., Langstrom, N., Lathrop, M., Lawrence, J., Lawson, W. B., Leboyer, M., Ledbetter, D. H., Lee, P. H., Lencz, T., Lesch, K. P., Levinson, D. F., Lewis, C. M., Li, J., Lichtenstein, P., Lieberman, J. A., Lin, D. Y., Linszen, D. H., Liu, C., Lohoff, F. W., Loo, S. K., Lord, C., Lowe, J. K., Lucae, S., Macintyre, D. J., Madden, P. A., Maestrini, E., Magnusson, P. K., Mahon, P. B., Maier, W., Malhotra, A. K., Mane, S. M., Martin, C. L., Martin, N. G., Mattheisen, M., Matthews, K., Mattingsdal, M., McCarroll, S. A., McGhee, K. A., McGough, J. J., McGrath, P. J., McGuffin, P., McInnis, M. G., McIntosh, A., McKinney, R., McLean, A. W., McMahon, F. J., McMahon, W. M., McQuillin, A., Medeiros, H., Medland, S. E., Meier, S., Melle, I., Meng, F., Meyer, J., Middeldorp, C. M., Middleton, L., Milanova, V., Miranda, A., Monaco, A. P., Montgomery, G. W., Moran, J. L., Moreno-De-Luca, D., Morken, G., Morris, D. W., Morrow, E. M., Moskvina, V., Muglia, P., Muhleisen, T. W., Muir, W. J., Muller-Myhsok, B., Murtha, M., Myers, R. M., Myin-Germeys, I., Neale, M. C., Nelson, S. F., Nievergelt, C. M., Nikolov, I., Nimgaonkar, V., Nolen, W. A., Nothen, M. M., Nurnberger, J. I., Nwulia, E. A., Nyholt, D. R., O'Dushlaine, C., Oades, R. D., Olincy, A., Oliveira, G., Olsen, L., Ophoff, R. A., Osby, U., Owen, M. J., Palotie, A., Parr, J. R., Paterson, A. D., Pato, C. N., Pato, M. T., Penninx, B. W., Pergadia, M. L., Pericak-Vance, M. A., Pickard, B. S., Pimm, J., Piven, J., Posthuma, D.,

Potash, J. B., Poustka, F., Propping, P., Puri, V., Quested, D. J., Quinn, E. M., Ramos-Quiroga, J. A., Rasmussen, H. B., Raychaudhuri, S., Rehnstrom, K., Reif, A., Ribases, M., Rice, J. P., Rietschel, M., Roeder, K., Roeyers, H., Rossin, L., Rothenberger, A., Rouleau, G., Ruderfer, D., Rujescu, D., Sanders, A. R., Sanders, S. J., Santangelo, S. L., Sergeant, J. A., Schachar, R., Schalling, M., Schatzberg, A. F., Scheftner, W. A., Schellenberg, G. D., Scherer, S. W., Schork, N. J., Schulze, T. G., Schumacher, J., Schwarz, M., Scolnick, E., Scott, L. J., Shi, J., Shilling, P. D., Shyn, S. I., Silverman, J. M., Slager, S. L., Smalley, S. L., Smit, J. H., Smith, E. N., Sonuga-Barke, E. J., St Clair, D., State, M., Steffens, M., Steinhausen, H. C., Strauss, J. S., Strohmaier, J., Stroup, T. S., Sutcliffe, J. S., Szatmari, P., Szelinger, S., Thirumalai, S., Thompson, R. C., Todorov, A. A., Tozzi, F., Treutlein, J., Uhr, M., van den Oord, E. J., Van Grootheest, G., Van Os, J., Vicente, A. M., Vieland, V. J., Vincent, J. B., Visscher, P. M., Walsh, C. A., Wassink, T. H., Watson, S. J., Weissman, M. M., Werge, T., Wienker, T. F., Wijsman, E. M., Willemsen, G., Williams, N., Willsey, A. J., Witt, S. H., Xu, W., Young, A. H., Yu, T. W., Zammit, S., Zandi, P. P., Zhang, P., Zitman, F. G., Zollner, S., International Inflammatory Bowel Disease Genetics, C., Devlin, B., Kelsoe, J. R., Sklar, P., Daly, M. J., O'Donovan, M. C., Craddock, N., Sullivan, P. F., Smoller, J. W., Kendler, K. S. and Wray, N. R. (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet*.

Cunha-Reis, D., Ribeiro, J. A. and Sebastiao, A. M. (2005). VIP enhances synaptic transmission to hippocampal CA1 pyramidal cells through activation of both VPAC1 and VPAC2 receptors. *Brain research*. 1049, 52-60

D'Onofrio, B. M., Rickert, M. E., Frans, E., Kuja-Halkola, R., Almqvist, C., Sjolander, A., Larsson, H. and Lichtenstein, P. (2014). Paternal Age at Childbearing and Offspring Psychiatric and Academic Morbidity. *JAMA psychiatry*.

Daly, M. J., Rioux, J. D., Schaffner, S. F., Hudson, T. J. and Lander, E. S. (2001). High-resolution haplotype structure in the human genome. *Nat Genet*. 29, 229-32

Darnell, J. C., Van Driesche, S. J., Zhang, C., Hung, K. Y., Mele, A., Fraser, C. E., Stone, E. F., Chen, C., Fak, J. J., Chi, S. W., Licatalosi, D. D., Richter, J. D. and Darnell, R. B. (2011). FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell*. 146, 247-61

Davies, G., Tenesa, A., Payton, A., Yang, J., Harris, S. E., Liewald, D., Ke, X., Le Hellard, S., Christoforou, A., Luciano, M., McGhee, K., Lopez, L., Gow, A. J., Corley, J., Redmond, P., Fox, H. C., Haggarty, P., Whalley, L. J., McNeill, G., Goddard, M. E., Espeseth, T., Lundervold, A. J., Reinvang, I., Pickles, A., Steen, V. M., Ollier, W.,

Porteous, D. J., Horan, M., Starr, J. M., Pendleton, N., Visscher, P. M. and Deary, I. J. (2011). Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Mol Psychiatry*. 16, 996-1005

Davis, O. S., Band, G., Pirinen, M., Haworth, C. M., Meaburn, E. L., Kovas, Y., Harlaar, N., Docherty, S. J., Hanscombe, K. B., Trzaskowski, M., Curtis, C. J., Strange, A., Freeman, C., Bellenguez, C., Su, Z., Pearson, R., Vukcevic, D., Langford, C., Deloukas, P., Hunt, S., Gray, E., Dronov, S., Potter, S. C., Tashakkori-Ghanbaria, A., Edkins, S., Bumpstead, S. J., Blackwell, J. M., Bramon, E., Brown, M. A., Casas, J. P., Corvin, A., Duncanson, A., Jankowski, J. A., Markus, H. S., Mathew, C. G., Palmer, C. N., Rautanen, A., Sawcer, S. J., Trembath, R. C., Viswanathan, A. C., Wood, N. W., Barroso, I., Peltonen, L., Dale, P. S., Petrill, S. A., Schalkwyk, L. S., Craig, I. W., Lewis, C. M., Price, T. S., Wellcome Trust Case Control, C., Donnelly, P., Plomin, R. and Spencer, C. C. (2014). The correlation between reading and mathematics ability at age twelve has a substantial genetic component. *Nature communications*. 5, 4204

De La Vega, F. M., Isaac, H., Collins, A., Scafe, C. R., Halldorsson, B. V., Su, X., Lippert, R. A., Wang, Y., Laig-Webster, M., Koehler, R. T., Ziegle, J. S., Wogan, L. T., Stevens, J. F., Leinen, K. M., Olson, S. J., Guegler, K. J., You, X., Xu, L. H., Hemken, H. G., Kalush, F.,

Itakura, M., Zheng, Y., de The, G., O'Brien, S. J., Clark, A. G., Istrail, S., Hunkapiller, M. W., Spier, E. G. and Gilbert, D. A. (2005). The linkage disequilibrium maps of three human chromosomes across four populations reflect their demographic history and a common underlying recombination pattern. *Genome Res*. 15, 454-62

Delis, D. C., Kramer, J. H., Kaplan, E. and Ober, B. (1987). California Verbal Learning Test: Manual, Reserch Edition.Psychological Corporation, Harcourt Brace & Company, San Anotnio.

Delmar, M. and Makita, N. (2012). Cardiac connexins, mutations and arrhythmias. *Current opinion in cardiology*. 27, 236-41

Devys, D., Lutz, Y., Rouyer, N., Bellocq, J. P. and Mandel, J. L. (1993). The FMR-1 protein is cytoplasmic, most abundant in neurons and appears normal in carriers of a fragile X premutation. *Nat Genet*. 4, 335-40

Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C., Marinov, G. K., Khatun, J., Williams, B. A., Zaleski, C., Rozowsky, J., Roder, M., Kokocinski, F., Abdelhamid, R. F., Alioto, T., Antoshechkin, I., Baer, M. T., Bar, N. S., Batut, P., Bell, K., Bell, I., Chakrabortty, S., Chen, X., Chrast, J., Curado, J., Derrien, T., Drenkow, J., Dumais, E., Dumais, J., Duttagupta, R., Falconnet, E., Fastuca, M., Fejes-Toth,

K., Ferreira, P., Foissac, S., Fullwood, M. J., Gao, H., Gonzalez, D., Gordon, A., Gunawardena, H., Howald, C., Jha, S., Johnson, R., Kapranov, P., King, B., Kingswood, C., Luo, O. J., Park, E., Persaud, K., Preall, J. B., Ribeca, P., Risk, B., Robyr, D., Sammeth, M., Schaffer, L., See, L. H., Shahab, A., Skancke, J., Suzuki, A. M., Takahashi, H., Tilgner, H., Trout, D., Walters, N., Wang, H., Wrobel, J., Yu, Y., Ruan, X., Hayashizaki, Y., Harrow, J., Gerstein, M., Hubbard, T., Reymond, A., Antonarakis, S. E., Hannon, G., Giddings, M. C., Ruan, Y., Wold, B., Carninci, P., Guigo, R. and Gingeras, T. R. (2012). Landscape of transcription in human cells. *Nature*. 489, 101-8

Durand, C. M., Betancur, C., Boeckers, T. M., Bockmann, J., Chaste, P., Fauchereau, F., Nygren, G., Rastam, M., Gillberg, I. C., Anckarsater, H., Sponheim, E., Goubran-Botros, H., Delorme, R., Chabane, N., Mouren-Simeoni, M. C., de Mas, P., Bieth, E., Roge, B., Heron, D., Burglen, L., Gillberg, C., Leboyer, M. and Bourgeron, T. (2007). Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nat Genet*. 39, 25-7

Durbin, R. M., Abecasis, G. R., Altshuler, D. L., Auton, A., Brooks, L. D., Gibbs, R. A., Hurles, M. E. and McVean, G. A. (2011). A map of human genome variation from population-scale sequencing. *Nature*. 467, 1061-73

Edelmann, L., Pandita, R. K. and Morrow, B. E. (1999). Low-copy repeats mediate the common 3-Mb deletion in patients with velo-cardio-facial syndrome. *Am J Hum Genet*. 64, 1076-86

Eichler, E. E. (2001). Recent duplication, domain accretion and the dynamic mutation of the human genome. *Trends Genet*. 17, 661-9

Ekelund, J., Lichtermann, D., Hovatta, I., Ellonen, P., Suvisaari, J., Terwilliger, J. D., Juvonen, H., Varilo, T., Arajarvi, R., Kokko-Sahin, M. L., Lonnqvist, J. and Peltonen, L. (2000). Genome-wide scan for schizophrenia in the Finnish population: evidence for a locus on chromosome 7q22. *Hum Mol Genet*. 9, 1049-57

Elvevag, B. and Goldberg, T. E. (2000). Cognitive impairment in schizophrenia is the core of the disorder. *Critical reviews in neurobiology*. 14, 1-21

Enattah, N. S., Sahi, T., Savilahti, E., Terwilliger, J. D., Peltonen, L. and Jarvela, I. (2002). Identification of a variant associated with adult-type hypolactasia. *Nat Genet*. 30, 233-7

Enattah, N. S., Trudeau, A., Pimenoff, V., Maiuri, L., Auricchio, S., Greco, L., Rossi, M., Lentze, M., Seo, J. K., Rahgozar, S., Khalil, I., Alifrangis, M., Natah, S., Groop, L., Shaat, N., Kozlov, A., Verschubskaya, G., Comas, D., Bulayeva, K., Mehdi, S. Q., Terwilliger, J. D., Sahi, T., Savilahti, E., Perola, M.,

Sajantila, A., Jarvela, I. and Peltonen, L. (2007). Evidence of still-ongoing convergence evolution of the lactase persistence T-13910 alleles in humans. *Am J Hum Genet*. 81, 615-25

Encode Project Consortium, Bernstein, B. E., Birney, E., Dunham, I., Green, E. D., Gunter, C. and Snyder, M. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*. 489, 57-74

Eriksson, J. G., Forsen, T., Tuomilehto, J., Osmond, C. and Barker, D. J. (2001). Early growth and coronary heart disease in later life: longitudinal study. *BMJ*. 322, 949-53

Faraone, S. V., Seidman, L. J., Kremen, W. S., Toomey, R., Pepple, J. R. and Tsuang, M. T. (2000). Neuropsychologic functioning among the nonpsychotic relatives of schizophrenic patients: the effect of genetic loading. *Biol Psychiatry*. 48, 120-6

Feng, J., Schroer, R., Yan, J., Song, W., Yang, C., Bockholt, A., Cook, E. H., Jr., Skinner, C., Schwartz, C. E. and Sommer, S. S. (2006). High frequency of neurexin 1beta signal peptide structural variants in patients with autism. *Neurosci Lett*. 409, 10-3

Fernandes, C., Hoyle, E., Dempster, E., Schalkwyk, L. C. and Collier, D. A. (2006). Performance deficit of alpha7 nicotinic receptor knockout mice in a delayed matching-to-place task suggests a mild impairment of working/episodic-like memory. *Genes, brain, and behavior*. 5, 433-40

Ferreira, M. A., O'Donovan, M. C., Meng, Y. A., Jones, I. R., Ruderfer, D. M., Jones, L., Fan, J., Kirov, G., Perlis, R. H., Green, E. K., Smoller, J. W., Grozeva, D., Stone, J., Nikolov, I., Chambert, K., Hamshere, M. L., Nimgaonkar, V. L., Moskvina, V., Thase, M. E., Caesar, S., Sachs, G. S., Franklin, J., Gordon-Smith, K., Ardlie, K. G., Gabriel, S. B., Fraser, C., Blumenstiel, B., Defelice, M., Breen, G., Gill, M., Morris, D. W., Elkin, A., Muir, W. J., McGhee, K. A., Williamson, R., MacIntyre, D. J., MacLean, A. W., St, C. D., Robinson, M., Van Beck, M., Pereira, A. C., Kandaswamy, R., McQuillin, A., Collier, D. A., Bass, N. J., Young, A. H., Lawrence, J., Ferrier, I. N., Anjorin, A., Farmer, A., Curtis, D., Scolnick, E. M., McGuffin, P., Daly, M. J., Corvin, A. P., Holmans, P. A., Blackwood, D. H., Gurling, H. M., Owen, M. J., Purcell, S. M., Sklar, P., Craddock, N. and Wellcome Trust Case Control, C. (2008). Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat Genet*. 40, 1056-8

Feuk, L., Carson, A. R. and Scherer, S. W. (2006). Structural variation in the human genome. *Nat Rev Genet*. 7, 85-97

Finnila, S., Lehtonen, M. S. and Majamaa, K. (2001). Phylogenetic network for European mtDNA. *Am J Hum Genet*. 68, 1475-84

Fioravanti, M., Bianchi, V. and Cinti, M. E. (2012). Cognitive deficits in schizophrenia: an updated metanalysis of the scientific evidence. *BMC psychiatry*. 12, 64

First, M. B., Spitzer, R. L. and Gibbon, M. (1997). Structured Clinical Interview for DSM–IV Axis I Disorders – Clinician Version (SCID–CV).American Psychiatric Press, Inc., Washington DC, USA

Fortna, A., Kim, Y., MacLaren, E., Marshall, K., Hahn, G., Meltesen, L., Brenton, M., Hink, R., Burgers, S., Hernandez-Boussard, T., Karimpour-Fard, A., Glueck, D., McGavran, L., Berry, R., Pollack, J. and Sikela, J. M. (2004). Lineage-specific gene duplication and loss in human and great ape evolution. *PLoS biology*. 2, E207

Francke, U. (1999). Williams-Beuren syndrome: genes and mechanisms. *Hum Mol Genet*. 8, 1947-54

Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., Belmont, J. W., Boudreau, A., Hardenbol, P., Leal, S. M., Pasternak, S., Wheeler, D. A., Willis, T. D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhou, J., Gabriel, S. B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R. C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Sun, W., Wang, H., Wang, Y., Xiong, X., Xu, L., Waye, M. M., Tsui, S. K., Xue, H., Wong, J. T., Galver, L. M., Fan, J. B., Gunderson, K., Murray, S. S., Oliphant, A. R., Chee, M. S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J. F., Phillips, M. S., Roumy, S., Sallee, C., Verner, A., Hudson, T. J., Kwok, P. Y., Cai, D., Koboldt, D. C., Miller, R. D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L. C., Mak, W., Song, Y. Q., Tam, P. K., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A., Tanaka, T., Tsunoda, T., Deloukas, P., Bird, C. P., Delgado, M., Dermitzakis, E. T., Gwilliam, R., Hunt, S., Morrison, J., Powell, D., Stranger, B. E., Whittaker, P., Bentley, D. R., Daly, M. J., de Bakker, P. I., Barrett, J., Chretien, Y. R., Maller, J., McCarroll, S., Patterson, N., Pe'er, I., Price, A., Purcell, S., Richter, D. J., Sabeti, P., Saxena, R., Schaffner, S. F., Sham, P. C., Varilly, P., Stein, L. D., Krishnan, L., Smith, A. V., Tello-Ruiz, M. K., Thorisson, G. A., Chakravarti, A., Chen, P. E., Cutler, D. J., Kashuk, C. S., Lin, S., Abecasis, G. R., Guan, W., Li, Y., Munro, H. M., Qin, Z. S., Thomas, D. J., McVean, G., Auton, A., Bottolo, L., Cardin, N., Eyheramendy, S., Freeman, C., Marchini, J., Myers, S., Spencer, C., Stephens, M., Donnelly, P., Cardon, L. R., Clarke, G., Evans, D. M., Morris, A. P., Weir, B. S., Mullikin, J. C., Sherry, S. T., Feolo, M., Skol, A.,

Zhang, H., Matsuda, I., Fukushima, Y., Macer, D. R., Suda, E., Rotimi, C. N., Adebamowo, C. A., Ajayi, I., Aniagwu, T., Marshall, P. A., Nkwodimmah, C., Royal, C. D., Leppert, M. F., Dixon, M., Peiffer, A., Qiu, R., Kent, A., Kato, K., Niikawa, N., Adewole, I. F., Knoppers, B. M., Foster, M. W., Clayton, E. W., Watkin, J., Muzny, D., Nazareth, L., Sodergren, E., Weinstock, G. M., Yakub, I., Birren, B. W., Wilson, R. K., Fulton, L. L., Rogers, J., Burton, J., Carter, N. P., Clee, C. M., Griffiths, M., Jones, M. C., McLay, K., Plumb, R. W., Ross, M. T., Sims, S. K., Willey, D. L., Chen, Z., Han, H., Kang, L., Godbout, M., Wallenburg, J. C., L'Archeveque, P., Bellemare, G., Saeki, K., An, D., Fu, H., Li, Q., Wang, Z., Wang, R., Holden, A. L., Brooks, L. D., McEwen, J. E., Guyer, M. S., Wang, V. O., Peterson, J. L., Shi, M., Spiegel, J., Sung, L. M., Zacharia, L. F., Collins, F. S., Kennedy, K., Jamieson, R. and Stewart, J. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature*. 449, 851-61

Fredman, D., White, S. J., Potter, S., Eichler, E. E., Den Dunnen, J. T. and Brookes, A. J. (2004). Complex SNP-related sequence variation in segmental genome duplications. *Nature Genetics*. 36, 861-866

Freedman, R. (2014). alpha7-nicotinic acetylcholine receptor agonists for cognitive enhancement in schizophrenia. *Annual review of medicine*. 65, 245-61

Friedman, J. M., Baross, A., Delaney, A. D., Ally, A., Arbour, L., Armstrong, L., Asano, J., Bailey, D. K., Barber, S., Birch, P., Brown-John, M., Cao, M., Chan, S., Charest, D. L., Farnoud, N., Fernandes, N., Flibotte, S., Go, A., Gibson, W. T., Holt, R. A., Jones, S. J., Kennedy, G. C., Krzywinski, M., Langlois, S., Li, H. I., McGillivray, B. C., Nayar, T., Pugh, T. J., Rajcan-Separovic, E., Schein, J. E., Schnerch, A., Siddiqui, A., Van Allen, M. I., Wilson, G., Yong, S. L., Zahir, F., Eydoux, P. and Marra, M. A. (2006). Oligonucleotide microarray analysis of genomic imbalance in children with mental retardation. *Am J Hum Genet*. 79, 500-13

Fromer, M., Pocklington, A. J., Kavanagh, D. H., Williams, H. J., Dwyer, S., Gormley, P., Georgieva, L., Rees, E., Palta, P., Ruderfer, D. M., Carrera, N., Humphreys, I., Johnson, J. S., Roussos, P., Barker, D. D., Banks, E., Milanova, V., Grant, S. G., Hannon, E., Rose, S. A., Chambert, K., Mahajan, M., Scolnick, E. M., Moran, J. L., Kirov, G., Palotie, A., McCarroll, S. A., Holmans, P., Sklar, P., Owen, M. J., Purcell, S. M. and O'Donovan, M. C. (2014). De novo mutations in schizophrenia implicate synaptic networks. *Nature*. 506, 179-84

Fu, Y. H., Kuhl, D. P., Pizzuti, A., Pieretti, M., Sutcliffe, J. S., Richards, S., Verkerk, A. J., Holden, J. J., Fenwick, R. G., Jr., Warren, S. T. and et al. (1991). Variation of the CGG repeat at the fragile X site results in

genetic instability: resolution of the Sherman paradox. *Cell*. 67, 1047-58

Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordero, S. N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E. S., Daly, M. J. and Altshuler, D. (2002). The structure of haplotype blocks in the human genome. *Science*. 296, 2225-9

Genomes Project Consortium, Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., Handsaker, R. E., Kang, H. M., Marth, G. T. and McVean, G. A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*. 491, 56-65

Gerstein, M. B., Bruce, C., Rozowsky, J. S., Zheng, D., Du, J., Korbel, J. O., Emanuelsson, O., Zhang, Z. D., Weissman, S. and Snyder, M. (2007). What is a gene, post-ENCODE? History and updated definition. *Genome Res*. 17, 669-81

Gilissen, C., Hehir-Kwa, J. Y., Thung, D. T., van de Vorst, M., van Bon, B. W., Willemsen, M. H., Kwint, M., Janssen, I. M., Hoischen, A., Schenck, A., Leach, R., Klein, R., Tearle, R., Bo, T., Pfundt, R., Yntema, H. G., de Vries, B. B., Kleefstra, T., Brunner, H. G., Vissers, L. E. and Veltman, J. A. (2014). Genome sequencing identifies major causes of severe intellectual disability. *Nature*.

Girard, S. L., Gauthier, J., Noreau, A., Xiong, L., Zhou, S., Jouan, L., Dionne-Laporte, A., Spiegelman, D., Henrion, E., Diallo, O., Thibodeau, P., Bachand, I., Bao, J. Y., Tong, A. H., Lin, C. H., Millet, B., Jaafari, N., Joober, R., Dion, P. A., Lok, S., Krebs, M. O. and Rouleau, G. A. (2011). Increased exonic de novo mutation rate in individuals with schizophrenia. *Nat Genet*. 43, 860-3

Gotti, C. and Clementi, F. (2004). Neuronal nicotinic receptors: from structure to pathology. *Progress in neurobiology*. 74, 363-96

Gotti, C., Zoli, M. and Clementi, F. (2006). Brain nicotinic acetylcholine receptors: native subtypes and their relevance. *Trends in pharmacological sciences*. 27, 482-91

Grant, S. F., Thorleifsson, G., Reynisdottir, I., Benediktsson, R., Manolescu, A., Sainz, J., Helgason, A., Stefansson, H., Emilsson, V., Helgadottir, A., Styrkarsdottir, U., Magnusson, K. P., Walters, G. B., Palsdottir, E., Jonsdottir, T., Gudmundsdottir, T., Gylfason, A., Saemundsdottir, J., Wilensky, R. L., Reilly, M. P., Rader, D. J., Bagger, Y., Christiansen, C., Gudnason, V., Sigurdsson, G., Thorsteinsdottir, U., Gulcher, J. R., Kong, A. and Stefansson, K. (2006). Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet*. 38, 320-3

Green, E. K., Grozeva, D., Jones, I., Jones, L., Kirov, G., Caesar, S., Gordon-Smith, K., Fraser, C., Forty, L., Russell, E., Hamshere, M. L., Moskvina, V., Nikolov, I., Farmer, A., McGuffin, P., Wellcome Trust Case Control, C., Holmans, P. A., Owen, M. J., O'Donovan, M. C. and Craddock, N. (2010). The bipolar disorder risk allele at CACNA1C also confers risk of recurrent major depression and of schizophrenia. *Mol Psychiatry*. 15, 1016-22

Grozeva, D., Kirov, G., Ivanov, D., Jones, I. R., Jones, L., Green, E. K., St Clair, D. M., Young, A. H., Ferrier, N., Farmer, A. E., McGuffin, P., Holmans, P. A., Owen, M. J., O'Donovan, M. C., Craddock, N. and Wellcome Trust Case Control, C. (2010). Rare copy number variants: a point of rarity in genetic risk for bipolar disorder and schizophrenia. *Arch Gen Psychiatry*. 67, 318-27

Grozeva, D., Kirov, G., Conrad, D. F., Barnes, C. P., Hurles, M., Owen, M. J., O'Donovan, M. C. and Craddock, N. (2013). Reduced burden of very large and rare CNVs in bipolar affective disorder. *Bipolar Disord*. 15, 893-8

Guha, S., Rees, E., Darvasi, A., Ivanov, D., Ikeda, M., Bergen, S. E., Magnusson, P. K., Cormican, P., Morris, D., Gill, M., Cichon, S., Rosenfeld, J. A., Lee, A., Gregersen, P. K., Kane, J. M., Malhotra, A. K., Rietschel, M., Nothen, M. M., Degenhardt, F., Priebe, L., Breuer, R., Strohmaier, J., Ruderfer, D. M.,

Moran, J. L., Chambert, K. D., Sanders, A. R., Shi, J., Kendler, K., Riley, B., O'Neill, T., Walsh, D., Malhotra, D., Corvin, A., Purcell, S., Sklar, P., Iwata, N., Hultman, C. M., Sullivan, P. F., Sebat, J., McCarthy, S., Gejman, P. V., Levinson, D. F., Owen, M. J., O'Donovan, M. C., Lencz, T., Kirov, G., Molecular Genetics of Schizophrenia, C. and Wellcome Trust Case Control, C. (2013). Implication of a rare deletion at distal 16p11.2 in schizophrenia. *JAMA psychiatry*. 70, 253-60

Gusella, J. F., Wexler, N. S., Conneally, P. M., Naylor, S. L., Anderson, M. A., Tanzi, R. E., Watkins, P. C., Ottina, K., Wallace, M. R., Sakaguchi, A. Y. and et al. (1983). A polymorphic DNA marker genetically linked to Huntington's disease. *Nature*. 306, 234-8

Hagberg, B., Hagberg, G., Lewerth, A. and Lindberg, U. (1981). Mild mental retardation in Swedish school children. II. Etiologic and pathogenetic aspects. *Acta paediatrica Scandinavica*. 70, 445-52

Harrison, G., Hopper, K., Craig, T., Laska, E., Siegel, C., Wanderling, J., Dube, K. C., Ganev, K., Giel, R., an der Heiden, W., Holmberg, S. K., Janca, A., Lee, P. W., Leon, C. A., Malhotra, S., Marsella, A. J., Nakane, Y., Sartorius, N., Shen, Y., Skoda, C., Thara, R., Tsirkin, S. J., Varma, V. K., Walsh, D. and Wiersma, D. (2001). Recovery from psychotic illness: a 15- and 25-year international follow-up study. *Br J Psychiatry*. 178, 506-17

Harrow, J., Frankish, A., Gonzalez, J. M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B. L., Barrell, D., Zadissa, A., Searle, S., Barnes, I., Bignell, A., Boychenko, V., Hunt, T., Kay, M., Mukherjee, G., Rajan, J., Despacio-Reyes, G., Saunders, G., Steward, C., Harte, R., Lin, M., Howald, C., Tanzer, A., Derrien, T., Chrast, J., Walters, N., Balasubramanian, S., Pei, B., Tress, M., Rodriguez, J. M., Ezkurdia, I., van Baren, J., Brent, M., Haussler, D., Kellis, M., Valencia, A., Reymond, A., Gerstein, M., Guigo, R. and Hubbard, T. J. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res*. 22, 1760-74

Haukka, J., Suvisaari, J., Varilo, T. and Lonnqvist, J. (2001). Regional variation in the incidence of schizophrenia in Finland: a study of birth cohorts born from 1950 to 1969. *Psychol Med*. 31, 1045-53

Hedman, M., Brandstatter, A., Pimenoff, V., Sistonen, P., Palo, J. U., Parson, W. and Sajantila, A. (2007). Finnish mitochondrial DNA HVS-I and HVS-II population data. *Forensic science international*. 172, 171-8

Heikura, U., Taanila, A., Olsen, P., Hartikainen, A. L., von Wendt, L. and Jarvelin, M. R. (2003). Temporal changes in incidence and prevalence of intellectual disability between two birth cohorts in Northern Finland. *Am J Ment Retard*. 108, 19-31

Heikura, U., Taanila, A., Hartikainen, A. L., Olsen, P., Linna, S. L., von Wendt, L. and Jarvelin, M. R. (2008). Variations in prenatal sociodemographic factors associated with intellectual disability: a study of the 20-year interval between two birth cohorts in northern Finland. *American journal of epidemiology*. 167, 169-77

Heinrichs, R. W. and Zakzanis, K. K. (1998). Neurocognitive deficit in schizophrenia: a quantitative review of the evidence. *Neuropsychology*. 12, 426-45

Helgadottir, A., Manolescu, A., Thorleifsson, G., Gretarsdottir, S., Jonsdottir, H., Thorsteinsdottir, U., Samani, N. J., Gudmundsson, G., Grant, S. F., Thorgeirsson, G., Sveinbjornsdottir, S., Valdimarsson, E. M., Matthiasson, S. E., Johannsson, H., Gudmundsdottir, O., Gurney, M. E., Sainz, J., Thorhallsdottir, M., Andresdottir, M., Frigge, M. L., Topol, E. J., Kong, A., Gudnason, V., Hakonarson, H., Gulcher, J. R. and Stefansson, K. (2004). The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke. *Nat Genet*. 36, 233-9

Hellier, J. L., Arevalo, N. L., Smith, L., Xiong, K. N. and Restrepo, D. (2012). alpha7-Nicotinic acetylcholine receptor: role in early odor learning preference in mice. *PLoS One*. 7, e35251

Hinds, D. A., Stuve, L. L., Nilsen, G. B., Halperin, E., Eskin, E., Ballinger, D. G., Frazer, K. A. and Cox, D. R. (2005). Whole-genome patterns of common DNA variation in three human populations. *Science*. 307, 1072-9

Hinds, D. A., Kloek, A. P., Jen, M., Chen, X. and Frazer, K. A. (2006). Common deletions and SNPs are in linkage disequilibrium in the human genome. *Nat Genet*. 38, 82-5

Hirschhorn, J. N., Lohmueller, K., Byrne, E. and Hirschhorn, K. (2002). A comprehensive review of genetic association studies. *Genet Med*. 4, 45-61

Hovatta, I., Terwilliger, J. D., Lichtermann, D., Makikyro, T., Suvisaari, J., Peltonen, L. and Lonnqvist, J. (1997). Schizophrenia in the genetic isolate of Finland. *American journal of medical genetics*. 74, 353-60

Hovatta, I., Lichtermann, D., Juvonen, H., Suvisaari, J., Terwilliger, J. D., Arajarvi, R., Kokko-Sahin, M. L., Ekelund, J., Lonnqvist, J. and Peltonen, L. (1998). Linkage analysis of putative schizophrenia gene candidate regions on chromosomes 3p, 5q, 6p, 8p, 20p and 22q in a population-based sampled Finnish family set. *Mol Psychiatry*. 3, 452-7

Hovatta, I., Varilo, T., Suvisaari, J., Terwilliger, J. D., Ollikainen, V., Arajärvi, R., Juvonen, H., Kokko-Sahin,

M. L., Väisänen, L., Mannila, H., Lönnqvist, J. and Peltonen, L. (1999). A genomewide screen for schizophrenia genes in an isolated Finnish subpopulation, suggesting multiple susceptibility loci. *Am J Hum Genet*. 65, 1114-24

Howes, O. D. and Murray, R. M. (2014). Schizophrenia: an integrated sociodevelopmental-cognitive model. *Lancet*. 383, 1677-87

Huang da, W., Sherman, B. T. and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 4, 44-57

Hurles, M. (2004). Gene duplication: the genomic trade in spare parts. *PLoS biology*. 2, E206

Hurles, M. E., Dermitzakis, E. T. and Tyler-Smith, C. (2008). The functional impact of structural variation in humans. *Trends Genet*. 24, 238-45

Iafrate, A. J., Feuk, L., Rivera, M. N., Listewnik, M. L., Donahoe, P. K., Qi, Y., Scherer, S. W. and Lee, C. (2004). Detection of large-scale variation in the human genome. *Nat Genet*. 36, 949-51

Ingason, A., Kirov, G., Giegling, I., Hansen, T., Isles, A. R., Jakobsen, K. D., Kristinsson, K. T., le Roux, L., Gustafsson, O., Craddock, N., Moller, H. J., McQuillin, A., Muglia, P., Cichon, S., Rietschel, M., Ophoff, R. A., Djurovic, S., Andreassen, O. A.,

Pietilainen, O. P., Peltonen, L., Dempster, E., Collier, D. A., St Clair, D., Rasmussen, H. B., Glenthoj, B. Y., Kiemeney, L. A., Franke, B., Tosato, S., Bonetto, C., Saemundsen, E., Hreidarsson, S. J., Investigators, G., Nothen, M. M., Gurling, H., O'Donovan, M. C., Owen, M. J., Sigurdsson, E., Petursson, H., Stefansson, H., Rujescu, D., Stefansson, K. and Werge, T. (2011). Maternally derived microduplications at 15q11-q13: implication of imprinted genes in psychotic illness. *Am J Psychiatry*. 168, 408-17

Inouye, M., Silander, K., Hamalainen, E., Salomaa, V., Harald, K., Jousilahti, P., Mannisto, S., Eriksson, J. G., Saarela, J., Ripatti, S., Perola, M., van Ommen, G. J., Taskinen, M. R., Palotie, A., Dermitzakis, E. T. and Peltonen, L. (2010). An immune response network associated with blood lipid levels. *PLoS Genet*. 6,

Insel, T. R. (2010). Rethinking schizophrenia. *Nature*. 468, 187-93

International HapMap, C., Altshuler, D. M., Gibbs, R. A., Peltonen, L., Altshuler, D. M., Gibbs, R. A., Peltonen, L., Dermitzakis, E., Schaffner, S. F., Yu, F., Peltonen, L., Dermitzakis, E., Bonnen, P. E., Altshuler, D. M., Gibbs, R. A., de Bakker, P. I., Deloukas, P., Gabriel, S. B., Gwilliam, R., Hunt, S., Inouye, M., Jia, X., Palotie, A., Parkin, M., Whittaker, P., Yu, F., Chang, K., Hawes, A., Lewis, L. R., Ren, Y., Wheeler, D., Gibbs, R. A., Muzny, D.

M., Barnes, C., Darvishi, K., Hurles, M., Korn, J. M., Kristiansson, K., Lee, C., McCarrol, S. A., Nemesh, J., Dermitzakis, E., Keinan, A., Montgomery, S. B., Pollack, S., Price, A. L., Soranzo, N., Bonnen, P. E., Gibbs, R. A., Gonzaga-Jauregui, C., Keinan, A., Price, A. L., Yu, F., Anttila, V., Brodeur, W., Daly, M. J., Leslie, S., McVean, G., Moutsianas, L., Nguyen, H., Schaffner, S. F., Zhang, Q., Ghori, M. J., McGinnis, R., McLaren, W., Pollack, S., Price, A. L., Schaffner, S. F., Takeuchi, F., Grossman, S. R., Shlyakhter, I., Hostetter, E. B., Sabeti, P. C., Adebamowo, C. A., Foster, M. W., Gordon, D. R., Licinio, J., Manca, M. C., Marshall, P. A., Matsuda, I., Ngare, D., Wang, V. O., Reddy, D., Rotimi, C. N., Royal, C. D., Sharp, R. R., Zeng, C., Brooks, L. D. and McEwen, J. E. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature*. 467, 52-8

International HapMap Consortium (2003). The International HapMap Project. *Nature*. 426, 789-96

International HapMap Consortium (2005). A haplotype map of the human genome. *Nature*. 437, 1299-320

International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. *Nature*. 431, 931-45

International Schizophrenia Consortium (2008). Rare

chromosomal deletions and duplications increase risk of schizophrenia. *Nature*. 455, 237-41

Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., Yamrom, B., Lee, Y. H., Narzisi, G., Leotta, A., Kendall, J., Grabowska, E., Ma, B., Marks, S., Rodgers, L., Stepansky, A., Troge, J., Andrews, P., Bekritsky, M., Pradhan, K., Ghiban, E., Kramer, M., Parla, J., Demeter, R., Fulton, L. L., Fulton, R. S., Magrini, V. J., Ye, K., Darnell, J. C., Darnell, R. B., Mardis, E. R., Wilson, R. K., Schatz, M. C., McCombie, W. R. and Wigler, M. (2012). De novo gene disruptions in children on the autistic spectrum. *Neuron*. 74, 285-99

Isohanni, I., Jarvelin, M. R., Nieminen, P., Jones, P., Rantakallio, P., Jokelainen, J. and Isohanni, M. (1998). School performance as a predictor of psychiatric hospitalization in adult life. A 28-year follow-up in the Northern Finland 1966 Birth Cohort. *Psychol Med*. 28, 967-74

Jaaskelainen, A., Pussinen, J., Nuutinen, O., Schwab, U., Pirkola, J., Kolehmainen, M., Jarvelin, M. R. and Laitinen, J. (2005). Intergenerational transmission of overweight among Finnish adolescents and their parents: a 16-year follow-up study. *International journal of obesity (2005)*. 35, 1289-94

Jaaskelainen, E., Juola, P., Hirvonen, N., McGrath, J. J., Saha, S., Isohanni, M., Veijola, J. and Miettunen, J. (2013). A systematic review and meta-analysis of recovery in schizophrenia. *Schizophr Bull*. 39, 1296-306

Jakkula, E., Rehnstrom, K., Varilo, T., Pietilainen, O. P., Paunio, T., Pedersen, N. L., deFaire, U., Jarvelin, M. R., Saharinen, J., Freimer, N., Ripatti, S., Purcell, S., Collins, A., Daly, M. J., Palotie, A. and Peltonen, L. (2008). The genome-wide patterns of variation expose significant substructure in a founder population. *Am J Hum Genet*. 83, 787-94

Jamain, S., Quach, H., Betancur, C., Rastam, M., Colineaux, C., Gillberg, I. C., Soderstrom, H., Giros, B., Leboyer, M., Gillberg, C., Bourgeron, T. and Paris Autism Research International Sibpair, S. (2003). Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat Genet*. 34, 27-9

Jarvelin, M. R., Hartikainen-Sorri, A. L. and Rantakallio, P. (1993). Labour induction policy in hospitals of different levels of specialisation. *Br J Obstet Gynaecol*. 100, 310-5

Jarvelin, M. R., Elliott, P., Kleinschmidt, I., Martuzzi, M., Grundy, C., Hartikainen, A. L. and Rantakallio, P. (1997). Ecological and individual predictors of birthweight in a northern Finland birth cohort 1986. *Paediatr Perinat Epidemiol*. 11, 298-312

Jeffreys, A. J., Kauppi, L. and Neumann, R. (2001). Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat Genet*. 29, 217-22

Jelenkovic, A., Ortega-Alonso, A., Rose, R. J., Kaprio, J., Rebato, E. and Silventoinen, K. (2011). Genetic and environmental influences on growth from late childhood to adulthood: a longitudinal study of two Finnish twin cohorts. *Am J Hum Biol*. 23, 764-73

Ji, Y., Eichler, E. E., Schwartz, S. and Nicholls, R. D. (2000). Structure of chromosomal duplicons and their role in mediating human genomic disorders. *Genome Res*. 10, 597-610

Jiang, J. X. (2010). Gap junctions or hemichannel-dependent and independent roles of connexins in cataractogenesis and lens development. *Current molecular medicine*. 10, 851-63

Jones, P., Rodgers, B., Murray, R. and Marmot, M. (1994). Child development risk factors for adult schizophrenia in the British 1946 birth cohort. *Lancet*. 344, 1398-402.

Jones, P. B., Rantakallio, P., Hartikainen, A. L., Isohanni, M. and Sipila, P. (1998). Schizophrenia as a long-term outcome of pregnancy, delivery, and perinatal complications: a 28-year follow-up of the 1966 north Finland general population birth cohort. *American Journal of Psychiatry*. 155, 355-64

Jostins, L., Ripke, S., Weersma, R. K., Duerr, R. H., McGovern, D. P., Hui, K. Y., Lee, J. C., Schumm, L. P., Sharma, Y., Anderson, C. A., Essers, J., Mitrovic, M., Ning, K., Cleynen, I., Theatre, E., Spain, S. L., Raychaudhuri, S., Goyette, P., Wei, Z., Abraham, C., Achkar, J. P., Ahmad, T., Amininejad, L., Ananthakrishnan, A. N., Andersen, V., Andrews, J. M., Baidoo, L., Balschun, T., Bampton, P. A., Bitton, A., Boucher, G., Brand, S., Buning, C., Cohain, A., Cichon, S., D'Amato, M., De Jong, D., Devaney, K. L., Dubinsky, M., Edwards, C., Ellinghaus, D., Ferguson, L. R., Franchimont, D., Fransen, K., Gearry, R., Georges, M., Gieger, C., Glas, J., Haritunians, T., Hart, A., Hawkey, C., Hedl, M., Hu, X., Karlsen, T. H., Kupcinskas, L., Kugathasan, S., Latiano, A., Laukens, D., Lawrance, I. C., Lees, C. W., Louis, E., Mahy, G., Mansfield, J., Morgan, A. R., Mowat, C., Newman, W., Palmieri, O., Ponsioen, C. Y., Potocnik, U., Prescott, N. J., Regueiro, M., Rotter, J. I., Russell, R. K., Sanderson, J. D., Sans, M., Satsangi, J., Schreiber, S., Simms, L. A., Sventoraityte, J., Targan, S. R., Taylor, K. D., Tremelling, M., Verspaget, H. W., De Vos, M., Wijmenga, C., Wilson, D. C., Winkelmann, J., Xavier, R. J., Zeissig, S., Zhang, B., Zhang, C. K., Zhao, H., International, I. B. D. G. C., Silverberg, M. S., Annese, V., Hakonarson, H., Brant, S. R., Radford-Smith, G., Mathew, C. G., Rioux, J. D., Schadt, E. E., Daly, M. J., Franke, A., Parkes, M.,

Vermeire, S., Barrett, J. C. and Cho, J. H. (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*. 491, 119-24

Kaprio, J. and Koskenvuo, M. (2002). Genetic and environmental factors in complex diseases: the older Finnish Twin Cohort. *Twin Res*. 5, 358-65

Kaprio, J., Pulkkinen, L. and Rose, R. J. (2002). Genetic and environmental factors in health-related behaviors: studies on Finnish twins and twin families. *Twin Res*. 5, 366-71

Karayiorgou, M., Morris, M. A., Morrow, B., Shprintzen, R. J., Goldberg, R., Borrow, J., Gos, A., Nestadt, G., Wolyniec, P. S., Lasseter, V. K. and et al. (1995). Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. *Proc Natl Acad Sci U S A*. 92, 7612-6

Kenny, E. E., Timpson, N. J., Sikora, M., Yee, M. C., Moreno-Estrada, A., Eng, C., Huntsman, S., Burchard, E. G., Stoneking, M., Bustamante, C. D. and Myles, S. (2012). Melanesian blond hair is caused by an amino acid change in TYRP1. *Science*. 336, 554

Kiezun, A., Garimella, K., Do, R., Stitziel, N. O., Neale, B. M., McLaren, P. J., Gupta, N., Sklar, P., Sullivan, P. F., Moran, J. L., Hultman, C. M., Lichtenstein, P., Magnusson, P., Lehner, T., Shugart, Y. Y., Price, A. L., de Bakker, P. I., Purcell, S. M. and

Sunyaev, S. R. (2012). Exome sequencing and the genetic basis of complex traits. *Nat Genet*. 44, 623-30

Kim, H. G., Kishikawa, S., Higgins, A. W., Seong, I. S., Donovan, D. J., Shen, Y., Lally, E., Weiss, L. A., Najm, J., Kutsche, K., Descartes, M., Holt, L., Braddock, S., Troxell, R., Kaplan, L., Volkmar, F., Klin, A., Tsatsanis, K., Harris, D. J., Noens, I., Pauls, D. L., Daly, M. J., MacDonald, M. E., Morton, C. C., Quade, B. J. and Gusella, J. F. (2008). Disruption of neurexin 1 associated with autism spectrum disorder. *Am J Hum Genet*. 82, 199-207

Kirov, G., Gumus, D., Chen, W., Norton, N., Georgieva, L., Sari, M., O'Donovan, M. C., Erdogan, F., Owen, M. J., Ropers, H. H. and Ullmann, R. (2008). Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. *Hum Mol Genet*. 17, 458-65

Kirov, G., Grozeva, D., Norton, N., Ivanov, D., Mantripragada, K. K., Holmans, P., Craddock, N., Owen, M. J. and O'Donovan, M. C. (2009). Support for the involvement of large cnvs in the pathogenesis of schizophrenia. *Hum Mol Genet*.

Kirov, G., Rujescu, D., Ingason, A., Collier, D. A., O'Donovan, M. C. and Owen, M. J. (2009). Neurexin 1 (NRXN1) deletions in schizophrenia. *Schizophr Bull*. 35, 851-4

Kirov, G., Pocklington, A. J., Holmans, P., Ivanov, D., Ikeda, M., Ruderfer, D., Moran, J., Chambert, K., Toncheva, D., Georgieva, L., Grozeva, D., Fjodorova, M., Wollerton, R., Rees, E., Nikolov, I., van de Lagemaat, L. N., Bayes, A., Fernandez, E., Olason, P. I., Bottcher, Y., Komiyama, N. H., Collins, M. O., Choudhary, J., Stefansson, K., Stefansson, H., Grant, S. G., Purcell, S., Sklar, P., O'Donovan, M. C. and Owen, M. J. (2012). De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol Psychiatry*. 17, 142-53

Kirov, G., Rees, E., Walters, J. T., Escott-Price, V., Georgieva, L., Richards, A. L., Chambert, K. D., Davies, G., Legge, S. E., Moran, J. L., McCarroll, S. A., O'Donovan, M. C. and Owen, M. J. (2014). The penetrance of copy number variations for schizophrenia and developmental delay. *Biol Psychiatry*. 75, 378-85

Kittles, R. A., Perola, M., Peltonen, L., Bergen, A. W., Aragon, R. A., Virkkunen, M., Linnoila, M., Goldman, D. and Long, J. C. (1998). Dual origins of Finns revealed by Y chromosome haplotype variation. *Am J Hum Genet*. 62, 1171-9

Klein, R. J., Zeiss, C., Chew, E. Y., Tsai, J. Y., Sackler, R. S., Haynes, C., Henning, A. K., SanGiovanni, J. P., Mane, S. M., Mayne, S. T., Bracken, M. B., Ferris, F. L., Ott, J., Barnstable, C. and Hoh, J. (2005). Complement factor H polymorphism in age-related macular degeneration. *Science*. 308, 385-9

Knapp, M., Mangalore, R. and Simon, J. (2004). The global costs of schizophrenia. *Schizophr Bull*. 30, 279-93

Kong, A., Frigge, M. L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., Gudjonsson, S. A., Sigurdsson, A., Jonasdottir, A., Jonasdottir, A., Wong, W. S., Sigurdsson, G., Walters, G. B., Steinberg, S., Helgason, H., Thorleifsson, G., Gudbjartsson, D. F., Helgason, A., Magnusson, O. T., Thorsteinsdottir, U. and Stefansson, K. (2012). Rate of de novo mutations and the importance of father's age to disease risk. *Nature*. 488, 471-5

Korn, J. M., Kuruvilla, F. G., McCarroll, S. A., Wysoker, A., Nemesh, J., Cawley, S., Hubbell, E., Veitch, J., Collins, P. J., Darvishi, K., Lee, C., Nizzari, M. M., Gabriel, S. B., Purcell, S., Daly, M. J. and Altshuler, D. (2008). Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet*. 40, 1253-60

Kovas, Y., Haworth, C. M., Dale, P. S. and Plomin, R. (2007). The genetic and environmental origins of learning abilities and disabilities in the early school years. *Monographs of the Society for Research in Child Development*. 72, vii, 1-144

Kremen, W. S., Seidman, L. J., Faraone, S. V., Toomey, R. and Tsuang, M. T. (2004). Heterogeneity of schizophrenia: a study of individual neuropsychological profiles. *Schizophr Res*. 71, 307-21

Kristiansson, K., Perola, M., Tikkanen, E., Kettunen, J., Surakka, I., Havulinna, A. S., Stancakova, A., Barnes, C., Widen, E., Kajantie, E., Eriksson, J. G., Viikari, J., Kahonen, M., Lehtimaki, T., Raitakari, O. T., Hartikainen, A. L., Ruokonen, A., Pouta, A., Jula, A., Kangas, A. J., Soininen, P., Ala-Korpela, M., Mannisto, S., Jousilahti, P., Bonnycastle, L. L., Jarvelin, M. R., Kuusisto, J., Collins, F. S., Laakso, M., Hurles, M. E., Palotie, A., Peltonen, L., Ripatti, S. and Salomaa, V. (2012). Genome-wide screen for metabolic syndrome susceptibility Loci reveals strong lipid gene contribution but no evidence for common genetic basis for clustering of metabolic syndrome traits. *Circulation*. 5, 242-9

Laaksovirta, H., Peuralinna, T., Schymick, J. C., Scholz, S. W., Lai, S. L., Myllykangas, L., Sulkava, R., Jansson, L., Hernandez, D. G., Gibbs, J. R., Nalls, M. A., Heckerman, D., Tienari, P. J. and Traynor, B. J. (2010). Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: a genome-wide association study. *Lancet neurology*. 9, 978-85

Laggerbauer, B., Ostareck, D., Keidel, E. M., Ostareck-Lederer, A. and Fischer, U. (2001). Evidence that fragile X mental retardation protein is a negative regulator of translation. *Hum Mol Genet*. 10, 329-38

Lahermo, P., Sajantila, A., Sistonen, P., Lukka, M., Aula, P., Peltonen, L. and Savontaus, M. L. (1996). The genetic relationship between the Finns and the Finnish Saami (Lapps): analysis of nuclear DNA and mtDNA. *Am J Hum Genet*. 58, 1309-22

Lahermo, P., Savontaus, M. L., Sistonen, P., Beres, J., de Knijff, P., Aula, P. and Sajantila, A. (1999). Y chromosomal polymorphisms reveal founding lineages in the Finns and the Saami. *Eur J Hum Genet*. 7, 447-58

Lander, E. S. and Schork, N. J. (1994). Genetic dissection of complex traits. *Science*. 265, 2037-48

Lander, E. S. (1996). The new genomics: global views of biology. *Science*. 274, 536-9

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A.,

Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V.,

Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y. J. and International Human Genome Sequencing, C. (2001). Initial sequencing and analysis of the human genome. *Nature*. 409, 860-921

Landerl, K. and Moll, K. (2010). Comorbidity of learning disorders: prevalence and familial transmission. *Journal of child psychology and*

psychiatry, and allied disciplines. 51, 287-94

Lango Allen, H., Estrada, K., Lettre, G., Berndt, S. I., Weedon, M. N., Rivadeneira, F., Willer, C. J., Jackson, A. U., Vedantam, S., Raychaudhuri, S., Ferreira, T., Wood, A. R., Weyant, R. J., Segre, A. V., Speliotes, E. K., Wheeler, E., Soranzo, N., Park, J. H., Yang, J., Gudbjartsson, D., Heard-Costa, N. L., Randall, J. C., Qi, L., Vernon Smith, A., Magi, R., Pastinen, T., Liang, L., Heid, I. M., Luan, J., Thorleifsson, G., Winkler, T. W., Goddard, M. E., Sin Lo, K., Palmer, C., Workalemahu, T., Aulchenko, Y. S., Johansson, A., Zillikens, M. C., Feitosa, M. F., Esko, T., Johnson, T., Ketkar, S., Kraft, P., Mangino, M., Prokopenko, I., Absher, D., Albrecht, E., Ernst, F., Glazer, N. L., Hayward, C., Hottenga, J. J., Jacobs, K. B., Knowles, J. W., Kutalik, Z., Monda, K. L., Polasek, O., Preuss, M., Rayner, N. W., Robertson, N. R., Steinthorsdottir, V., Tyrer, J. P., Voight, B. F., Wiklund, F., Xu, J., Zhao, J. H., Nyholt, D. R., Pellikka, N., Perola, M., Perry, J. R., Surakka, I., Tammesoo, M. L., Altmaier, E. L., Amin, N., Aspelund, T., Bhangale, T., Boucher, G., Chasman, D. I., Chen, C., Coin, L., Cooper, M. N., Dixon, A. L., Gibson, Q., Grundberg, E., Hao, K., Juhani Junttila, M., Kaplan, L. M., Kettunen, J., Konig, I. R., Kwan, T., Lawrence, R. W., Levinson, D. F., Lorentzon, M., McKnight, B., Morris, A. P., Muller, M., Suh Ngwa, J., Purcell, S., Rafelt, S., Salem, R. M., Salvi, E., Sanna, S., Shi, J., Sovio, U., Thompson, J. R., Turchin, M. C., Vandenput, L., Verlaan, D. J., Vitart, V., White, C. C., Ziegler, A., Almgren, P., Balmforth, A. J., Campbell, H., Citterio, L., De Grandi, A., Dominiczak, A., Duan, J., Elliott, P., Elosua, R., Eriksson, J. G., Freimer, N. B., Geus, E. J., Glorioso, N., Haiqing, S., Hartikainen, A. L., Havulinna, A. S., Hicks, A. A., Hui, J., Igl, W., Illig, T., Jula, A., Kajantie, E., Kilpelainen, T. O., Koiranen, M., Kolcic, I., Koskinen, S., Kovacs, P., Laitinen, J., Liu, J., Lokki, M. L., Marusic, A., Maschio, A., Meitinger, T., Mulas, A., Pare, G., Parker, A. N., Peden, J. F., Petersmann, A., Pichler, I., Pietilainen, K. H., Pouta, A., Ridderstrale, M., Rotter, J. I., Sambrook, J. G., Sanders, A. R., Schmidt, C. O., Sinisalo, J., Smit, J. H., Stringham, H. M., Bragi Walters, G., Widen, E., Wild, S. H., Willemsen, G., Zagato, L., Zgaga, L., Zitting, P., Alavere, H., Farrall, M., McArdle, W. L., Nelis, M., Peters, M. J., Ripatti, S., van Meurs, J. B., Aben, K. K., Ardlie, K. G., Beckmann, J. S., Beilby, J. P., Bergman, R. N., Bergmann, S., Collins, F. S., Cusi, D., den Heijer, M., Eiriksdottir, G., Gejman, P. V., Hall, A. S., Hamsten, A., Huikuri, H. V., Iribarren, C., Kahonen, M., Kaprio, J., Kathiresan, S., Kiemeney, L., Kocher, T., Launer, L. J., Lehtimaki, T., Melander, O., Mosley, T. H., Jr., Musk, A. W., Nieminen, M. S., O'Donnell, C. J., Ohlsson, C., Oostra, B., Palmer, L. J., Raitakari, O., Ridker, P. M., Rioux, J. D., Rissanen, A., Rivolta, C., Schunkert, H., Shuldiner, A. R., Siscovick, D. S., Stumvoll, M., Tonjes, A., Tuomilehto, J., van Ommen, G. J., Viikari, J., Heath, A. C., Martin, N. G., Montgomery, G.

W., Province, M. A., Kayser, M., Arnold, A. M., Atwood, L. D., Boerwinkle, E., Chanock, S. J., Deloukas, P., Gieger, C., Gronberg, H., Hall, P., Hattersley, A. T., Hengstenberg, C., Hoffman, W., Lathrop, G. M., Salomaa, V., Schreiber, S., Uda, M., Waterworth, D., Wright, A. F., Assimes, T. L., Barroso, I., Hofman, A., Mohlke, K. L., Boomsma, D. I., Caulfield, M. J., Cupples, L. A., Erdmann, J., Fox, C. S., Gudnason, V., Gyllensten, U., Harris, T. B., Hayes, R. B., Jarvelin, M. R., Mooser, V., Munroe, P. B., Ouwehand, W. H., Penninx, B. W., Pramstaller, P. P., Quertermous, T., Rudan, I., Samani, N. J., Spector, T. D., Volzke, H., Watkins, H., Wilson, J. F., Groop, L. C., Haritunians, T., Hu, F. B., Kaplan, R. C., Metspalu, A., North, K. E., Schlessinger, D., Wareham, N. J., Hunter, D. J., O'Connell, J. R., Strachan, D. P., Wichmann, H. E., Borecki, I. B., van Duijn, C. M., Schadt, E. E., Thorsteinsdottir, U., Peltonen, L., Uitterlinden, A. G., Visscher, P. M., Chatterjee, N., Loos, R. J., Boehnke, M., McCarthy, M. I., Ingelsson, E., Lindgren, C. M., Abecasis, G. R., Stefansson, K., Frayling, T. M. and Hirschhorn, J. N. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*. 467, 832-8

Lappalainen, T., Koivumaki, S., Salmela, E., Huoponen, K., Sistonen, P., Savontaus, M. L. and Lahermo, P. (2006). Regional differences among the Finns: a Y-chromosomal perspective. *Gene*. 376, 207-15

Lehtinen, V., Joukamaa, M., Lahtela, K., Raitasalo, R., Jyrkinen, E., Maatela, J. and Aromaa, A. (1990). Prevalence of mental disorders among adults in Finland: basic results from the Mini Finland Health Survey. *Acta psychiatrica Scandinavica*. 81, 418-25

Levinson, D. F., Duan, J., Oh, S., Wang, K., Sanders, A. R., Shi, J., Zhang, N., Mowry, B. J., Olincy, A., Amin, F., Cloninger, C. R., Silverman, J. M., Buccola, N. G., Byerley, W. F., Black, D. W., Kendler, K. S., Freedman, R., Dudbridge, F., Pe'er, I., Hakonarson, H., Bergen, S. E., Fanous, A. H., Holmans, P. A. and Gejman, P. V. (2011). Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. *Am J Psychiatry*. 168, 302-16

Lewis, C. M., Levinson, D. F., Wise, L. H., DeLisi, L. E., Straub, R. E., Hovatta, I., Williams, N. M., Schwab, S. G., Pulver, A. E., Faraone, S. V., Brzustowicz, L. M., Kaufmann, C. A., Garver, D. L., Gurling, H. M., Lindholm, E., Coon, H., Moises, H. W., Byerley, W., Shaw, S. H., Mesen, A., Sherrington, R., O'Neill, F. A., Walsh, D., Kendler, K. S., Ekelund, J., Paunio, T., Lonnqvist, J., Peltonen, L., O'Donovan, M. C., Owen, M. J., Wildenauer, D. B., Maier, W., Nestadt, G., Blouin, J. L., Antonarakis, S. E., Mowry, B. J., Silverman, J. M., Crowe, R. R., Cloninger, C. R., Tsuang, M. T., Malaspina, D., Harkavy-Friedman, J.

M., Svrakic, D. M., Bassett, A. S., Holcomb, J., Kalsi, G., McQuillin, A., Brynjolfson, J., Sigmundsson, T., Petursson, H., Jazin, E., Zoega, T. and Helgason, T. (2003). Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. *Am J Hum Genet*. 73, 34-48

Lewis, D. A., Curley, A. A., Glausier, J. R. and Volk, D. W. (2012). Cortical parvalbumin interneurons and cognitive dysfunction in schizophrenia. *Trends in neurosciences*. 35, 57-67

Lewontin, R. C. (1964). The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. *Genetics*. 49, 49-67

Li, T., Li, Z., Chen, P., Zhao, Q., Wang, T., Huang, K., Li, J., Li, Y., Liu, J., Zeng, Z., Feng, G., He, L. and Shi, Y. (2010). Common variants in major histocompatibility complex region and TCF4 gene are significantly associated with schizophrenia in Han Chinese. *Biol Psychiatry*. 68, 671-3

Li, Y., Vinckenbosch, N., Tian, G., Huerta-Sanchez, E., Jiang, T., Jiang, H., Albrechtsen, A., Andersen, G., Cao, H., Korneliussen, T., Grarup, N., Guo, Y., Hellman, I., Jin, X., Li, Q., Liu, J., Liu, X., Sparso, T., Tang, M., Wu, H., Wu, R., Yu, C., Zheng, H., Astrup, A., Bolund, L., Holmkvist, J., Jorgensen, T., Kristiansen, K., Schmitz, O., Schwartz, T. W., Zhang, X., Li, R., Yang, H., Wang, J., Hansen, T., Pedersen, O., Nielsen, R. and Wang, J. (2010). Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nat Genet*. 42, 969-72

Lichtenstein, P., Yip, B. H., Bjork, C., Pawitan, Y., Cannon, T. D., Sullivan, P. F. and Hultman, C. M. (2009). Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet*. 373, 234-9

Lin, H., Hsu, F. C., Baumann, B. H., Coulter, D. A., Anderson, S. A. and Lynch, D. R. (2014). Cortical parvalbumin GABAergic deficits with alpha7 nicotinic acetylcholine receptor deletion: implications for schizophrenia. *Molecular and cellular neurosciences*. 61C, 163-175

Locke, D. P., Sharp, A. J., McCarroll, S. A., McGrath, S. D., Newman, T. L., Cheng, Z., Schwartz, S., Albertson, D. G., Pinkel, D., Altshuler, D. M. and Eichler, E. E. (2006). Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome. *Am J Hum Genet*. 79, 275-90

Lofgren, E., Pouta, A., von Wendt, L., Tapanainen, J., Isojarvi, J. I. and Jarvelin, M. R. (2009). Epilepsy in the northern Finland birth cohort 1966 with special reference to fertility. *Epilepsy Behav*. 14, 102-7

Luo, L. (2000). Rho GTPases in neuronal morphogenesis. *Nat Rev Neurosci*. 1, 173-80

Lupski, J. R. (1998). Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet*. 14, 417-22

Lynch, M. (2010). Rate, molecular spectrum, and consequences of human mutation. *Proc Natl Acad Sci U S A*. 107, 961-8

MacArthur, D. G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins, L., Habegger, L., Pickrell, J. K., Montgomery, S. B., Albers, C. A., Zhang, Z. D., Conrad, D. F., Lunter, G., Zheng, H., Ayub, Q., DePristo, M. A., Banks, E., Hu, M., Handsaker, R. E., Rosenfeld, J. A., Fromer, M., Jin, M., Mu, X. J., Khurana, E., Ye, K., Kay, M., Saunders, G. I., Suner, M. M., Hunt, T., Barnes, I. H., Amid, C., Carvalho-Silva, D. R., Bignell, A. H., Snow, C., Yngvadottir, B., Bumpstead, S., Cooper, D. N., Xue, Y., Romero, I. G., Genomes Project, C., Wang, J., Li, Y., Gibbs, R. A., McCarroll, S. A., Dermitzakis, E. T., Pritchard, J. K., Barrett, J. C., Harrow, J., Hurles, M. E., Gerstein, M. B. and Tyler-Smith, C. (2012). A systematic survey of loss-of-function variants in human protein-coding genes. *Science*. 335, 823-8

Malhotra, D., McCarthy, S., Michaelson, J. J., Vacic, V., Burdick, K. E., Yoon, S., Cichon, S., Corvin, A., Gary, S., Gershon, E. S., Gill, M.,

Karayiorgou, M., Kelsoe, J. R., Krastoshevsky, O., Krause, V., Leibenluft, E., Levy, D. L., Makarov, V., Bhandari, A., Malhotra, A. K., McMahon, F. J., Nothen, M. M., Potash, J. B., Rietschel, M., Schulze, T. G. and Sebat, J. (2011). High frequencies of de novo CNVs in bipolar disorder and schizophrenia. *Neuron*. 72, 951-63

Malhotra, D. and Sebat, J. (2012). CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell*. 148, 1223-41

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., Boehnke, M., Clark, A. G., Eichler, E. E., Gibson, G., Haines, J. L., Mackay, T. F., McCarroll, S. A. and Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*. 461, 747-53

Marshall, C. R., Noor, A., Vincent, J. B., Lionel, A. C., Feuk, L., Skaug, J., Shago, M., Moessner, R., Pinto, D., Ren, Y., Thiruvahindrapduram, B., Fiebig, A., Schreiber, S., Friedman, J., Ketelaars, C. E., Vos, Y. J., Ficicioglu, C., Kirkpatrick, S., Nicolson, R., Sloman, L., Summers, A., Gibbons, C. A., Teebi, A., Chitayat, D., Weksberg, R., Thompson, A., Vardy, C., Crosbie, V., Luscombe, S., Baatjes, R., Zwaigenbaum, L., Roberts, W.,

Fernandez, B., Szatmari, P. and Scherer, S. W. (2008). Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet*. 82, 477-88

McCarroll, S. A., Hadnott, T. N., Perry, G. H., Sabeti, P. C., Zody, M. C., Barrett, J. C., Dallaire, S., Gabriel, S. B., Lee, C., Daly, M. J. and Altshuler, D. M. (2006). Common deletion polymorphisms in the human genome. *Nat Genet*. 38, 86-92

McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P. and Hirschhorn, J. N. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet*. 9, 356-69

McCarthy, S. E., Makarov, V., Kirov, G., Addington, A. M., McClellan, J., Yoon, S., Perkins, D. O., Dickel, D. E., Kusenda, M., Krastoshevsky, O., Krause, V., Kumar, R. A., Grozeva, D., Malhotra, D., Walsh, T., Zackai, E. H., Kaplan, P., Ganesh, J., Krantz, I. D., Spinner, N. B., Roccanova, P., Bhandari, A., Pavon, K., Lakshmi, B., Leotta, A., Kendall, J., Lee, Y. H., Vacic, V., Gary, S., Iakoucheva, L. M., Crow, T. J., Christian, S. L., Lieberman, J. A., Stroup, T. S., Lehtimaki, T., Puura, K., Haldeman-Englert, C., Pearl, J., Goodell, M., Willour, V. L., Derosse, P., Steele, J., Kassem, L., Wolff, J., Chitkara, N., McMahon, F. J., Malhotra, A. K., Potash, J. B., Schulze, T. G., Nothen, M. M., Cichon, S., Rietschel, M., Leibenluft, E.,

Kustanovich, V., Lajonchere, C. M., Sutcliffe, J. S., Skuse, D., Gill, M., Gallagher, L., Mendell, N. R., Craddock, N., Owen, M. J., O'Donovan, M. C., Shaikh, T. H., Susser, E., Delisi, L. E., Sullivan, P. F., Deutsch, C. K., Rapoport, J., Levy, D. L., King, M. C. and Sebat, J. (2009). Microduplications of 16p11.2 are associated with schizophrenia. *Nat Genet*. 41, 1223-7

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

McGrath, J., Saha, S., Chant, D. and Welham, J. (2008). Schizophrenia: a concise overview of incidence, prevalence, and mortality. *Epidemiologic reviews*. 30, 67-76

McGrath, J. J., Petersen, L., Agerbo, E., Mors, O., Mortensen, P. B. and Pedersen, C. B. (2014). A Comprehensive Assessment of Parental Age and Psychiatric Disorders. *JAMA psychiatry*.

McGuffin, P., Farmer, A. and Harvey, I. (1991). A polydiagnostic application of operational criteria in studies of psychotic illness. Development and reliability of the OPCRIT system. *Archives of General Psychiatry*. 48, 764-70

McQuillin, A., Bass, N., Anjorin, A., Lawrence, J., Kandaswamy, R., Lydall, G., Moran, J., Sklar, P., Purcell, S. and Gurling, H. (2011). Analysis of genetic deletions and duplications in the University College London bipolar disorder case control sample. *Eur J Hum Genet*. 19, 588-92

Mefford, H. C., Sharp, A. J., Baker, C., Itsara, A., Jiang, Z., Buysse, K., Huang, S., Maloney, V. K., Crolla, J. A., Baralle, D., Collins, A., Mercer, C., Norga, K., de Ravel, T., Devriendt, K., Bongers, E. M., de Leeuw, N., Reardon, W., Gimelli, S., Bena, F., Hennekam, R. C., Male, A., Gaunt, L., Clayton-Smith, J., Simonic, I., Park, S. M., Mehta, S. G., Nik-Zainal, S., Woods, C. G., Firth, H. V., Parkin, G., Fichera, M., Reitano, S., Lo Giudice, M., Li, K. E., Casuga, I., Broomer, A., Conrad, B., Schwerzmann, M., Raber, L., Gallati, S., Striano, P., Coppola, A., Tolmie, J. L., Tobias, E. S., Lilley, C., Armengol, L., Spysschaert, Y., Verloo, P., De Coene, A., Goossens, L., Mortier, G., Speleman, F., van Binsbergen, E., Nelen, M. R., Hochstenbach, R., Poot, M., Gallagher, L., Gill, M., McClellan, J., King, M. C., Regan, R., Skinner, C., Stevenson, R. E., Antonarakis, S. E., Chen, C., Estivill, X., Menten, B., Gimelli, G., Gribble, S., Schwartz, S., Sutcliffe, J. S., Walsh, T., Knight, S. J., Sebat, J., Romano, C., Schwartz, C. E., Veltman, J. A., de Vries, B. B., Vermeesch, J. R., Barber, J. C., Willatt, L., Tassabehji, M. and Eichler, E. E. (2008). Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *The New England journal of medicine*. 359, 1685-99

Mesholam-Gately, R. I., Giuliano, A. J., Goff, K. P., Faraone, S. V. and Seidman, L. J. (2009). Neurocognition in first-episode schizophrenia: a meta-analytic review. *Neuropsychology*. 23, 315-36

Millar, J. K., Wilson-Annan, J. C., Anderson, S., Christie, S., Taylor, M. S., Semple, C. A., Devon, R. S., Clair, D. M., Muir, W. J., Blackwood, D. H. and Porteous, D. J. (2000). Disruption of two novel genes by a translocation co-segregating with schizophrenia. *Hum Mol Genet*. 22, 1415-1423

Milligan, B. G. (2003). Maximum-likelihood estimation of relatedness. *Genetics*. 163, 1153-67

Mills, R. E., Walter, K., Stewart, C., Handsaker, R. E., Chen, K., Alkan, C., Abyzov, A., Yoon, S. C., Ye, K., Cheetham, R. K., Chinwalla, A., Conrad, D. F., Fu, Y., Grubert, F., Hajirasouliha, I., Hormozdiari, F., Iakoucheva, L. M., Iqbal, Z., Kang, S., Kidd, J. M., Konkel, M. K., Korn, J., Khurana, E., Kural, D., Lam, H. Y., Leng, J., Li, R., Li, Y., Lin, C. Y., Luo, R., Mu, X. J., Nemesh, J., Peckham, H. E., Rausch, T., Scally, A., Shi, X., Stromberg, M. P., Stutz, A. M., Urban, A. E., Walker, J. A., Wu, J., Zhang, Y., Zhang, Z. D., Batzer, M. A., Ding, L., Marth, G. T., McVean, G., Sebat, J., Snyder, M., Wang, J., Ye, K., Eichler, E. E., Gerstein, M. B., Hurles, M. E., Lee, C., McCarroll, S. A., Korbel, J. O. and

Genomes, P. (2011). Mapping copy number variation by population-scale genome sequencing. *Nature*. 470, 59-65

Moilanen, K., Veijola, J., Laksy, K., Makikyro, T., Miettunen, J., Kantojarvi, L., Kokkonen, P., Karvonen, J. T., Herva, A., Joukamaa, M., Jarvelin, M. R., Moring, J., Jones, P. B. and Isohanni, M. (2003). Reasons for the diagnostic discordance between clinicians and researchers in schizophrenia in the Northern Finland 1966 Birth Cohort. *Soc Psychiatry Psychiatr Epidemiol*. 38, 305-10

Morris, A. P., Voight, B. F., Teslovich, T. M., Ferreira, T., Segre, A. V., Steinthorsdottir, V., Strawbridge, R. J., Khan, H., Grallert, H., Mahajan, A., Prokopenko, I., Kang, H. M., Dina, C., Esko, T., Fraser, R. M., Kanoni, S., Kumar, A., Lagou, V., Langenberg, C., Luan, J., Lindgren, C. M., Muller-Nurasyid, M., Pechlivanis, S., Rayner, N. W., Scott, L. J., Wiltshire, S., Yengo, L., Kinnunen, L., Rossin, E. J., Raychaudhuri, S., Johnson, A. D., Dimas, A. S., Loos, R. J., Vedantam, S., Chen, H., Florez, J. C., Fox, C., Liu, C. T., Rybin, D., Couper, D. J., Kao, W. H., Li, M., Cornelis, M. C., Kraft, P., Sun, Q., van Dam, R. M., Stringham, H. M., Chines, P. S., Fischer, K., Fontanillas, P., Holmen, O. L., Hunt, S. E., Jackson, A. U., Kong, A., Lawrence, R., Meyer, J., Perry, J. R., Platou, C. G., Potter, S., Rehnberg, E., Robertson, N., Sivapalaratnam, S., Stancakova, A., Stirrups, K., Thorleifsson, G., Tikkanen, E., Wood, A. R., Almgren, P.,

Atalay, M., Benediktsson, R., Bonnycastle, L. L., Burtt, N., Carey, J., Charpentier, G., Crenshaw, A. T., Doney, A. S., Dorkhan, M., Edkins, S., Emilsson, V., Eury, E., Forsen, T., Gertow, K., Gigante, B., Grant, G. B., Groves, C. J., Guiducci, C., Herder, C., Hreidarsson, A. B., Hui, J., James, A., Jonsson, A., Rathmann, W., Klopp, N., Kravic, J., Krjutskov, K., Langford, C., Leander, K., Lindholm, E., Lobbens, S., Mannisto, S., Mirza, G., Muhleisen, T. W., Musk, B., Parkin, M., Rallidis, L., Saramies, J., Sennblad, B., Shah, S., Sigurethsson, G., Silveira, A., Steinbach, G., Thorand, B., Trakalo, J., Veglia, F., Wennauer, R., Winckler, W., Zabaneh, D., Campbell, H., van Duijn, C., Uitterlinden, A. G., Hofman, A., Sijbrands, E., Abecasis, G. R., Owen, K. R., Zeggini, E., Trip, M. D., Forouhi, N. G., Syvanen, A. C., Eriksson, J. G., Peltonen, L., Nothen, M. M., Balkau, B., Palmer, C. N., Lyssenko, V., Tuomi, T., Isomaa, B., Hunter, D. J., Qi, L., Wellcome Trust Case Control, C., Meta-Analyses of, G., Insulin-related traits Consortium, I., Genetic Investigation of, A. T. C., Asian Genetic Epidemiology Network-Type 2 Diabetes, C., South Asian Type 2 Diabetes, C., Shuldiner, A. R., Roden, M., Barroso, I., Wilsgaard, T., Beilby, J., Hovingh, K., Price, J. F., Wilson, J. F., Rauramaa, R., Lakka, T. A., Lind, L., Dedoussis, G., Njolstad, I., Pedersen, N. L., Khaw, K. T., Wareham, N. J., Keinanen-Kiukaanniemi, S. M., Saaristo, T. E., Korpi-Hyovalti, E., Saltevo, J., Laakso, M., Kuusisto, J., Metspalu, A., Collins, F. S., Mohlke, K. L., Bergman, R. N., Tuomilehto, J.,

Boehm, B. O., Gieger, C., Hveem, K., Cauchi, S., Froguel, P., Baldassarre, D., Tremoli, E., Humphries, S. E., Saleheen, D., Danesh, J., Ingelsson, E., Ripatti, S., Salomaa, V., Erbel, R., Jockel, K. H., Moebus, S., Peters, A., Illig, T., de Faire, U., Hamsten, A., Morris, A. D., Donnelly, P. J., Frayling, T. M., Hattersley, A. T., Boerwinkle, E., Melander, O., Kathiresan, S., Nilsson, P. M., Deloukas, P., Thorsteinsdottir, U., Groop, L. C., Stefansson, K., Hu, F., Pankow, J. S., Dupuis, J., Meigs, J. B., Altshuler, D., Boehnke, M., McCarthy, M. I., Replication, D. I. G. and Meta-analysis, C. (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet*. 44, 981-90

Mortensen, P. B., Pedersen, C. B., Westergaard, T., Wohlfahrt, J., Ewald, H., Mors, O., Andersen, P. K. and Melbye, M. (1999). Effects of family history and place and season of birth on the risk of schizophrenia. *The New England journal of medicine*. 340, 603-8

Mortensen, P. B., Pedersen, M. G. and Pedersen, C. B. (2009). Psychiatric family history and schizophrenia risk in Denmark: which mental disorders are relevant? *Psychol Med*. 1-10

Muhleisen, T. W., Mattheisen, M., Strohmaier, J., Degenhardt, F., Priebe, L., Schultz, C. C., Breuer, R., Meier, S., Hoffmann, P., Investigators, G., Rivandeneira, F., Hofman, A., Uitterlinden, A. G., Moebus, S.,

Gieger, C., Emeny, R., Ladwig, K. H., Wichmann, H. E., Schwarz, M., Kammerer-Ciernioch, J., Schlosser, R. G., Nenadic, I., Sauer, H., Mossner, R., Maier, W., Rujescu, D., Lange, C., Ophoff, R. A., Schulze, T. G., Rietschel, M., Nothen, M. M. and Cichon, S. (2012). Association between schizophrenia and common variation in neurocan (NCAN), a genetic risk factor for bipolar disorder. *Schizophr Res*. 138, 69-73

Murphy, K. C., Jones, L. A. and Owen, M. J. (1999). High rates of schizophrenia in adults with velo-cardio-facial syndrome. *Arch Gen Psychiatry*. 56, 940-5.

Neale, B. M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K. E., Sabo, A., Lin, C. F., Stevens, C., Wang, L. S., Makarov, V., Polak, P., Yoon, S., Maguire, J., Crawford, E. L., Campbell, N. G., Geller, E. T., Valladares, O., Schafer, C., Liu, H., Zhao, T., Cai, G., Lihm, J., Dannenfelser, R., Jabado, O., Peralta, Z., Nagaswamy, U., Muzny, D., Reid, J. G., Newsham, I., Wu, Y., Lewis, L., Han, Y., Voight, B. F., Lim, E., Rossin, E., Kirby, A., Flannick, J., Fromer, M., Shakir, K., Fennell, T., Garimella, K., Banks, E., Poplin, R., Gabriel, S., DePristo, M., Wimbish, J. R., Boone, B. E., Levy, S. E., Betancur, C., Sunyaev, S., Boerwinkle, E., Buxbaum, J. D., Cook, E. H., Jr., Devlin, B., Gibbs, R. A., Roeder, K., Schellenberg, G. D., Sutcliffe, J. S. and Daly, M. J. (2012). Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature*. 485, 242-5

Neph, S., Vierstra, J., Stergachis, A. B., Reynolds, A. P., Haugen, E., Vernot, B., Thurman, R. E., John, S., Sandstrom, R., Johnson, A. K., Maurano, M. T., Humbert, R., Rynes, E., Wang, H., Vong, S., Lee, K., Bates, D., Diegel, M., Roach, V., Dunn, D., Neri, J., Schafer, A., Hansen, R. S., Kutyavin, T., Giste, E., Weaver, M., Canfield, T., Sabo, P., Zhang, M., Balasundaram, G., Byron, R., MacCoss, M. J., Akey, J. M., Bender, M. A., Groudine, M., Kaul, R. and Stamatoyannopoulos, J. A. (2012). An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature*. 489, 83-90

Ng, M. Y., Levinson, D. F., Faraone, S. V., Suarez, B. K., DeLisi, L. E., Arinami, T., Riley, B., Paunio, T., Pulver, A. E., Irmansyah, Holmans, P. A., Escamilla, M., Wildenauer, D. B., Williams, N. M., Laurent, C., Mowry, B. J., Brzustowicz, L. M., Maziade, M., Sklar, P., Garver, D. L., Abecasis, G. R., Lerer, B., Fallin, M. D., Gurling, H. M., Gejman, P. V., Lindholm, E., Moises, H. W., Byerley, W., Wijsman, E. M., Forabosco, P., Tsuang, M. T., Hwu, H. G., Okazaki, Y., Kendler, K. S., Wormley, B., Fanous, A., Walsh, D., O'Neill, F. A., Peltonen, L., Nestadt, G., Lasseter, V. K., Liang, K. Y., Papadimitriou, G. M., Dikeos, D. G., Schwab, S. G., Owen, M. J., O'Donovan, M. C., Norton, N., Hare, E., Raventos, H., Nicolini, H., Albus, M., Maier, W., Nimgaonkar, V. L., Terenius, L., Mallet, J., Jay, M., Godard, S., Nertney, D., Alexander, M., Crowe, R. R., Silverman, J. M., Bassett, A. S., Roy, M. A., Merette, C., Pato, C. N., Pato, M. T., Roos, J. L., Kohn, Y., Amann-Zalcenstein, D., Kalsi, G., McQuillin, A., Curtis, D., Brynjolfson, J., Sigmundsson, T., Petursson, H., Sanders, A. R., Duan, J., Jazin, E., Myles-Worsley, M., Karayiorgou, M. and Lewis, C. M. (2009). Meta-analysis of 32 genome-wide linkage studies of schizophrenia. *Mol Psychiatry*. 14, 774-85

Ng, S. B., Turner, E. H., Robertson, P. D., Flygare, S. D., Bigham, A. W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E. E., Bamshad, M., Nickerson, D. A. and Shendure, J. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*. 461, 272-6

Ni, X., Valente, J., Azevedo, M. H., Pato, M. T., Pato, C. N. and Kennedy, J. L. (2007). Connexin 50 gene on human chromosome 1q21 is associated with schizophrenia in matched case control and family-based studies. *J Med Genet*. 44, 532-6

Nichols, P. L. (1984). Familial mental retardation. *Behavior genetics*. 14, 161-70

Nordentoft, M., Wahlbeck, K., Hallgren, J., Westman, J., Osby, U., Alinaghizadeh, H., Gissler, M. and Laursen, T. M. (2013). Excess mortality, causes of death and life expectancy in 270,770 patients with recent onset of mental disorders in

Denmark, Finland and Sweden. *PLoS One*. 8, e55176

Norio, R., Nevanlinna, H. R. and Perheentupa, J. (1973). Hereditary diseases in Finland; rare flora in rare soul. *Annals of Clinical Research*. 5, 109-41

Norio, R. (2003). Finnish Disease Heritage II: population prehistory and genetic roots of Finns. *Human genetics*. 112, 457-69

Nyegaard, M., Demontis, D., Foldager, L., Hedemand, A., Flint, T. J., Sorensen, K. M., Andersen, P. S., Nordentoft, M., Werge, T., Pedersen, C. B., Hougaard, D. M., Mortensen, P. B., Mors, O. and Borglum, A. D. (2010). CACNA1C (rs1006737) is associated with schizophrenia. *Mol Psychiatry*. 15, 119-21

O'Donovan, M. C., Craddock, N., Norton, N., Williams, H., Peirce, T., Moskvina, V., Nikolov, I., Hamshere, M., Carroll, L., Georgieva, L., Dwyer, S., Holmans, P., Marchini, J. L., Spencer, C. C., Howie, B., Leung, H. T., Hartmann, A. M., Moller, H. J., Morris, D. W., Shi, Y., Feng, G., Hoffmann, P., Propping, P., Vasilescu, C., Maier, W., Rietschel, M., Zammit, S., Schumacher, J., Quinn, E. M., Schulze, T. G., Williams, N. M., Giegling, I., Iwata, N., Ikeda, M., Darvasi, A., Shifman, S., He, L., Duan, J., Sanders, A. R., Levinson, D. F., Gejman, P. V., Cichon, S., Nothen, M. M., Gill, M., Corvin, A., Rujescu, D., Kirov, G., Owen, M. J., Buccola, N. G., Mowry, B.

J., Freedman, R., Amin, F., Black, D. W., Silverman, J. M., Byerley, W. F., Cloninger, C. R. and Molecular Genetics of Schizophrenia, C. (2008). Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet*. 40, 1053-5

O'Roak, B. J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B. P., Levy, R., Ko, A., Lee, C., Smith, J. D., Turner, E. H., Stanaway, I. B., Vernot, B., Malig, M., Baker, C., Reilly, B., Akey, J. M., Borenstein, E., Rieder, M. J., Nickerson, D. A., Bernier, R., Shendure, J. and Eichler, E. E. (2012). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature*. 485, 246-50

Owen, M. J., Craddock, N. and O'Donovan, M. C. (2010). Suggestion of roles for both common and rare risk variants in genome-wide studies of schizophrenia. *Arch Gen Psychiatry*. 67, 667-73

Pajunen, P., Rissanen, H., Harkanen, T., Jula, A., Reunanen, A. and Salomaa, V. (2010). The metabolic syndrome as a predictor of incident diabetes and cardiovascular events in the Health 2000 Study. *Diabetes & metabolism*. 36, 395-401

Palo, J. U., Ulmanen, I., Lukka, M., Ellonen, P. and Sajantila, A. (2009). Genetic markers and population history: Finland revisited. *Eur J Hum Genet*. 17, 1336-46

Patil, N., Berno, A. J., Hinds, D. A., Barrett, W. A., Doshi, J. M., Hacker, C. R., Kautzer, C. R., Lee, D. H., Marjoribanks, C., McDonough, D. P., Nguyen, B. T., Norris, M. C., Sheehan, J. B., Shen, N., Stern, D., Stokowski, R. P., Thomas, D. J., Trulson, M. O., Vyas, K. R., Frazer, K. A., Fodor, S. P. and Cox, D. R. (2001). Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science*. 294, 1719-23

Paunio, T., Ekelund, J., Varilo, T., Parker, A., Hovatta, I., Turunen, J. A., Rinard, K., Foti, A., Terwilliger, J. D., Juvonen, H., Suvisaari, J., Arajarvi, R., Suokas, J., Partonen, T., Lonnqvist, J., Meyer, J. and Peltonen, L. (2001). Genome-wide scan in a nationwide study sample of schizophrenia families in Finland reveals susceptibility loci on chromosomes 2q and 5q. *Hum Mol Genet*. 10, 3037-48.

Pearson, H. (2006). Genetics: what is a gene? *Nature*. 441, 398-401

Pedersen, C. B., Mors, O., Bertelsen, A., Waltoft, B. L., Agerbo, E., McGrath, J. J., Mortensen, P. B. and Eaton, W. W. (2014). A comprehensive nationwide study of the incidence rate and lifetime risk for treated mental disorders. *JAMA psychiatry*. 71, 573-81

Peltonen, L., Jalanko, A. and Varilo, T. (1999). Molecular genetics of the Finnish disease heritage. *Human Molecular Genetics*. 8, 1913-1923

Peltonen, L., Palotie, A. and Lange, K. (2000). Use of population isolates for mapping complex traits. *Nat Rev Genet*. 1, 182-90.

Peoples, R., Franke, Y., Wang, Y. K., Perez-Jurado, L., Paperna, T., Cisco, M. and Francke, U. (2000). A physical map, including a BAC/PAC clone contig, of the Williams-Beuren syndrome--deletion region at 7q11.23. *Am J Hum Genet*. 66, 47-68

Perala, J., Suvisaari, J., Saarni, S. I., Kuoppasalmi, K., Isometsa, E., Pirkola, S., Partonen, T., Tuulio-Henriksson, A., Hintikka, J., Kieseppa, T., Harkanen, T., Koskinen, S. and Lonnqvist, J. (2007). Lifetime prevalence of psychotic and bipolar I disorders in a general population. *Arch Gen Psychiatry*. 64, 19-28

Perala, J., Saarni, S. I., Ostamo, A., Pirkola, S., Haukka, J., Harkanen, T., Koskinen, S., Lonnqvist, J. and Suvisaari, J. (2008). Geographic variation and sociodemographic characteristics of psychotic disorders in Finland. *Schizophr Res*. 106, 337-47

Pieretti, M., Zhang, F. P., Fu, Y. H., Warren, S. T., Oostra, B. A., Caskey, C. T. and Nelson, D. L. (1991). Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell*. 66, 817-22

Pietilainen, O. P., Rehnstrom, K., Jakkula, E., Service, S. K., Congdon, E., Tilgmann, C., Hartikainen, A. L., Taanila, A., Heikura, U., Paunio, T., Ripatti, S., Jarvelin, M. R., Isohanni, M., Sabatti, C., Palotie, A., Freimer, N. B. and Peltonen, L. (2011). Phenotype mining in CNV carriers from a population cohort. *Hum Mol Genet*. 20, 2686-95

Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W. L., Chen, C., Zhai, Y., Dairkee, S. H., Ljung, B. M., Gray, J. W. and Albertson, D. G. (1998). High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet*. 20, 207-11

Pinto, D., Pagnamenta, A. T., Klei, L., Anney, R., Merico, D., Regan, R., Conroy, J., Magalhaes, T. R., Correia, C., Abrahams, B. S., Almeida, J., Bacchelli, E., Bader, G. D., Bailey, A. J., Baird, G., Battaglia, A., Berney, T., Bolshakova, N., Bolte, S., Bolton, P. F., Bourgeron, T., Brennan, S., Brian, J., Bryson, S. E., Carson, A. R., Casallo, G., Casey, J., Chung, B. H., Cochrane, L., Corsello, C., Crawford, E. L., Crossett, A., Cytrynbaum, C., Dawson, G., de Jonge, M., Delorme, R., Drmic, I., Duketis, E., Duque, F., Estes, A., Farrar, P., Fernandez, B. A., Folstein, S. E., Fombonne, E., Freitag, C. M., Gilbert, J., Gillberg, C., Glessner, J. T., Goldberg, J., Green, A., Green, J., Guter, S. J., Hakonarson, H., Heron, E. A., Hill, M., Holt, R., Howe, J. L., Hughes, G., Hus, V., Igliozzi, R., Kim, C., Klauck, S. M., Kolevzon, A., Korvatska, O., Kustanovich, V., Lajonchere, C. M., Lamb, J. A., Laskawiec, M., Leboyer, M., Le Couteur, A., Leventhal, B. L., Lionel, A. C., Liu, X. Q., Lord, C., Lotspeich, L., Lund, S. C., Maestrini, E., Mahoney, W., Mantoulan, C., Marshall, C. R., McConachie, H., McDougle, C. J., McGrath, J., McMahon, W. M., Merikangas, A., Migita, O., Minshew, N. J., Mirza, G. K., Munson, J., Nelson, S. F., Noakes, C., Noor, A., Nygren, G., Oliveira, G., Papanikolaou, K., Parr, J. R., Parrini, B., Paton, T., Pickles, A., Pilorge, M., Piven, J., Ponting, C. P., Posey, D. J., Poustka, A., Poustka, F., Prasad, A., Ragoussis, J., Renshaw, K., Rickaby, J., Roberts, W., Roeder, K., Roge, B., Rutter, M. L., Bierut, L. J., Rice, J. P., Salt, J., Sansom, K., Sato, D., Segurado, R., Sequeira, A. F., Senman, L., Shah, N., Sheffield, V. C., Soorya, L., Sousa, I., Stein, O., Sykes, N., Stoppioni, V., Strawbridge, C., Tancredi, R., Tansey, K., Thiruvahindrapduram, B., Thompson, A. P., Thomson, S., Tryfon, A., Tsiantis, J., Van Engeland, H., Vincent, J. B., Volkmar, F., Wallace, S., Wang, K., Wang, Z., Wassink, T. H., Webber, C., Weksberg, R., Wing, K., Wittemeyer, K., Wood, S., Wu, J., Yaspan, B. L., Zurawiecki, D., Zwaigenbaum, L., Buxbaum, J. D., Cantor, R. M., Cook, E. H., Coon, H., Cuccaro, M. L., Devlin, B., Ennis, S., Gallagher, L., Geschwind, D. H., Gill, M., Haines, J. L., Hallmayer, J., Miller, J., Monaco, A. P., Nurnberger, J. I., Jr., Paterson, A. D., Pericak-Vance, M. A., Schellenberg, G. D., Szatmari, P., Vicente, A. M., Vieland,

V. J., Wijsman, E. M., Scherer, S. W., Sutcliffe, J. S. and Betancur, C. (2010). Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*. 466, 368-72

Pirkola, S. P., Isometsa, E., Suvisaari, J., Aro, H., Joukamaa, M., Poikolainen, K., Koskinen, S., Aromaa, A. and Lonnqvist, J. K. (2005). DSM-IV mood-, anxiety- and alcohol use disorders and their comorbidity in the Finnish general population--results from the Health 2000 Study. *Soc Psychiatry Psychiatr Epidemiol*. 40, 1-10

Plomin, R. and Kovas, Y. (2005). Generalist genes and learning disabilities. *Psychological bulletin*. 131, 592-617

Power, R. A., Kyaga, S., Uher, R., MacCabe, J. H., Langstrom, N., Landen, M., McGuffin, P., Lewis, C. M., Lichtenstein, P. and Svensson, A. C. (2013). Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings. *JAMA psychiatry*. 70, 22-30

Price, A. L., Zaitlen, N. A., Reich, D. and Patterson, N. (2010). New approaches to population stratification in genome-wide association studies. *Nat Rev Genet*. 11, 459-63

Priebe, L., Degenhardt, F. A., Herms, S., Haenisch, B., Mattheisen, M., Nieratschker, V., Weingarten, M.,

Witt, S., Breuer, R., Paul, T., Alblas, M., Moebus, S., Lathrop, M., Leboyer, M., Schreiber, S., Grigoroiu-Serbanescu, M., Maier, W., Propping, P., Rietschel, M., Nothen, M. M., Cichon, S. and Muhleisen, T. W. (2012). Genome-wide survey implicates the influence of copy number variants (CNVs) in the development of early-onset bipolar disorder. *Mol Psychiatry*. 17, 421-32

Przeworski, M., Hudson, R. R. and Di Rienzo, A. (2000). Adjusting the focus on human variation. *Trends Genet*. 16, 296-302

Psychiatric, G. C. B. D. W. G. (2011). Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet*. 43, 977-83

Psychiatric Gwas Consortium Coordinating Committee, Cichon, S., Craddock, N., Daly, M., Faraone, S. V., Gejman, P. V., Kelsoe, J., Lehner, T., Levinson, D. F., Moran, A., Sklar, P. and Sullivan, P. F. (2009). Genomewide association studies: history, rationale, and prospects for psychiatric disorders. *Am J Psychiatry*. 166, 540-56

Puffenberger, E. G., Hosoda, K., Washington, S. S., Nakao, K., deWit, D., Yanagisawa, M. and Chakravart, A. (1994). A missense mutation of the endothelin-B receptor gene in multigenic Hirschsprung's disease. *Cell*. 79, 1257-66

Pulver, A. E., Nestadt, G., Goldberg, R., Shprintzen, R. J., Lamacz, M., Wolyniec, P. S., Morrow, B., Karayiorgou, M., Antonarakis, S. E., Housman, D. and et al. (1994). Psychotic illness in patients diagnosed with velo-cardio-facial syndrome and their relatives. *The Journal of nervous and mental disease*. 182, 476-8

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J. and Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 81, 559-75

Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., Sklar, P., Purcell Leader, S. M., Stone, J. L., Sullivan, P. F., Ruderfer, D. M., McQuillin, A., Morris, D. W., O'Dushlaine, C. T., Corvin, A., Holmans, P. A., O'Donovan, M. C., Sklar, P., Wray, N. R., Macgregor, S., Sklar, P., Sullivan, P. F., O'Donovan, M. C., Visscher, P. M., Gurling, H., Blackwood, D. H., Corvin, A., Craddock, N. J., Gill, M., Hultman, C. M., Kirov, G. K., Lichtenstein, P., McQuillin, A., Muir, W. J., O'Donovan, M. C., Owen, M. J., Pato, C. N., Purcell, S. M., Scolnick, E. M., St Clair, D., Stone, J. L., Sullivan, P. F., Sklar Leader, P., O'Donovan, M. C., Kirov, G. K., Craddock, N. J., Holmans, P. A., Williams, N. M., Georgieva, L., Nikolov, I., Norton, N., Williams, H., Toncheva, D., Milanova, V., Owen, M. J., Hultman, C. M., Lichtenstein, P.,

Thelander, E. F., Sullivan, P., Morris, D. W., O'Dushlaine, C. T., Kenny, E., Quinn, E. M., Gill, M., Corvin, A., McQuillin, A., Choudhury, K., Datta, S., Pimm, J., Thirumalai, S., Puri, V., Krasucki, R., Lawrence, J., Quested, D., Bass, N., Gurling, H., Crombie, C., Fraser, G., Leh Kuan, S., Walker, N., St Clair, D., Blackwood, D. H., Muir, W. J., McGhee, K. A., Pickard, B., Malloy, P., Maclean, A. W., Van Beck, M., Wray, N. R., Macgregor, S., Visscher, P. M., Pato, M. T., Medeiros, H., Middleton, F., Carvalho, C., Morley, C., Fanous, A., Conti, D., Knowles, J. A., Paz Ferreira, C., Macedo, A., Helena Azevedo, M., Pato, C. N., Stone, J. L., Ruderfer, D. M., Kirby, A. N., Ferreira, M. A., Daly, M. J., Purcell, S. M., Sklar, P., Purcell, S. M., Stone, J. L., Chambert, K., Ruderfer, D. M., Kuruvilla, F., Gabriel, S. B., Ardlie, K., Moran, J. L., Daly, M. J., Scolnick, E. M. and Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*.

Purcell, S. M., Moran, J. L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., O'Dushlaine, C., Chambert, K., Bergen, S. E., Kahler, A., Duncan, L., Stahl, E., Genovese, G., Fernandez, E., Collins, M. O., Komiyama, N. H., Choudhary, J. S., Magnusson, P. K., Banks, E., Shakir, K., Garimella, K., Fennell, T., DePristo, M., Grant, S. G., Haggarty, S. J., Gabriel, S., Scolnick, E. M., Lander, E. S., Hultman, C. M., Sullivan, P. F., McCarroll, S. A. and Sklar, P. (2014). A polygenic burden of rare disruptive

mutations in schizophrenia. *Nature*. 506, 185-90

Raitakari, O. T., Juonala, M., Kahonen, M., Taittonen, L., Laitinen, T., Maki-Torkko, N., Jarvisalo, M. J., Uhari, M., Jokinen, E., Ronnemaa, T., Akerblom, H. K. and Viikari, J. S. (2003). Cardiovascular risk factors in childhood and carotid artery intima-media thickness in adulthood: the Cardiovascular Risk in Young Finns Study. *Jama*. 290, 2277-83

Raitakari, O. T., Juonala, M., Ronnemaa, T., Keltikangas-Jarvinen, L., Rasanen, L., Pietikainen, M., Hutri-Kahonen, N., Taittonen, L., Jokinen, E., Marniemi, J., Jula, A., Telama, R., Kahonen, M., Lehtimaki, T., Akerblom, H. K. and Viikari, J. S. (2008). Cohort profile: the cardiovascular risk in Young Finns Study. *Int J Epidemiol*. 37, 1220-6

Rajji, T. K., Ismail, Z. and Mulsant, B. H. (2009). Age at onset and cognition in schizophrenia: meta-analysis. *Br J Psychiatry*. 195, 286-93

Rantakallio, P. and von Wendt, L. (1985). Risk factors for mental retardation. *Arch Dis Child*. 60, 946-52

Rantakallio, P. and von Wendt, L. (1986). Mental retardation and subnormality in a birth cohort of 12,000 children in Northern Finland. *Am J Ment Defic*. 90, 380-7

Rantakallio, P. (1988). The longitudinal study of the northern Finland birth cohort of 1966. *Paediatr Perinat Epidemiol*. 2, 59-88.

Rapoport, J. L., Addington, A. M., Frangou, S. and Psych, M. R. (2005). The neurodevelopmental model of schizophrenia: update 2005. *Mol Psychiatry*. 10, 434-49

Rapoport, J. L., Giedd, J. N. and Gogtay, N. (2012). Neurodevelopmental model of schizophrenia: update 2012. *Mol Psychiatry*. 17, 1228-38

Rauch, A., Wieczorek, D., Graf, E., Wieland, T., Endele, S., Schwarzmayr, T., Albrecht, B., Bartholdi, D., Beygo, J., Di Donato, N., Dufke, A., Cremer, K., Hempel, M., Horn, D., Hoyer, J., Joset, P., Ropke, A., Moog, U., Riess, A., Thiel, C. T., Tzschach, A., Wiesener, A., Wohlleber, E., Zweier, C., Ekici, A. B., Zink, A. M., Rump, A., Meisinger, C., Grallert, H., Sticht, H., Schenck, A., Engels, H., Rappold, G., Schrock, E., Wieacker, P., Riess, O., Meitinger, T., Reis, A. and Strom, T. M. (2012). Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet*. 380, 1674-82

Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., Fiegler, H., Shapero, M. H., Carson, A. R., Chen, W., Cho, E. K., Dallaire, S., Freeman, J. L., Gonzalez, J. R., Gratacos, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J. R., Marshall, C. R., Mei,

R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M. J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, J., Zerjal, T., Zhang, J., Armengol, L., Conrad, D. F., Estivill, X., Tyler-Smith, C., Carter, N. P., Aburatani, H., Lee, C., Jones, K. W., Scherer, S. W. and Hurles, M. E. (2006). Global variation in copy number in the human genome. *Nature*. 444, 444-54

Rees, E., Moskvina, V., Owen, M. J., O'Donovan, M. C. and Kirov, G. (2011). De novo rates and selection of schizophrenia-associated copy number variants. *Biol Psychiatry*. 70, 1109-14

Reich, D. E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P. C., Richter, D. J., Lavery, T., Kouyoumjian, R., Farhadian, S. F., Ward, R. and Lander, E. S. (2001). Linkage disequilibrium in the human genome. *Nature*. 411, 199-204

Reich, D. E., Schaffner, S. F., Daly, M. J., McVean, G., Mullikin, J. C., Higgins, J. M., Richter, D. J., Lander, E. S. and Altshuler, D. (2002). Human genome sequence variation and the influence of gene history, mutation and recombination. *Nat Genet*. 32, 135-42

Reich, D. E., Gabriel, S. B. and Altshuler, D. (2003). Quality and completeness of SNP databases. *Nat Genet*. 33, 457-8

Reichenberg, A., Caspi, A., Harrington, H., Houts, R., Keefe, R. S., Murray, R. M., Poulton, R. and Moffitt, T. E. (2010). Static and dynamic cognitive deficits in childhood preceding adult schizophrenia: a 30-year study. *Am J Psychiatry*. 167, 160-9

Riley, B., Thiselton, D., Maher, B. S., Bigdeli, T., Wormley, B., McMichael, G. O., Fanous, A. H., Vladimirov, V., O'Neill, F. A., Walsh, D. and Kendler, K. S. (2010). Replication of association between schizophrenia and ZNF804A in the Irish Case-Control Study of Schizophrenia sample. *Mol Psychiatry*. 15, 29-37

Ripke, S., Sanders, A. R., Kendler, K. S., Levinson, D. F., Sklar, P., Holmans, P. A., Lin, D. Y., Duan, J., Ophoff, R. A., Andreassen, O. A., Scolnick, E., Cichon, S., St Clair, D., Corvin, A., Gurling, H., Werge, T., Rujescu, D., Blackwood, D. H., Pato, C. N., Malhotra, A. K., Purcell, S., Dudbridge, F., Neale, B. M., Rossin, L., Visscher, P. M., Posthuma, D., Ruderfer, D. M., Fanous, A., Stefansson, H., Steinberg, S., Mowry, B. J., Golimbet, V., De Hert, M., Jonsson, E. G., Bitter, I., Pietilainen, O. P., Collier, D. A., Tosato, S., Agartz, I., Albus, M., Alexander, M., Amdur, R. L., Amin, F., Bass, N., Bergen, S. E., Black, D. W., Borglum, A. D., Brown, M. A., Bruggeman, R., Buccola, N. G., Byerley, W. F., Cahn, W., Cantor, R. M., Carr, V. J., Catts, S. V., Choudhury, K., Cloninger, C. R., Cormican, P., Craddock, N., Danoy, P. A., Datta, S., de Haan, L., Demontis, D., Dikeos, D., Djurovic, S., Donnelly, P., Donohoe, G., Duong, L., Dwyer, S., Fink-Jensen,

A., Freedman, R., Freimer, N. B.,
Friedl, M., Georgieva, L., Giegling, I.,
Gill, M., Glenthoj, B., Godard, S.,
Hamshere, M., Hansen, M., Hansen,
T., Hartmann, A. M., Henskens, F. A.,
Hougaard, D. M., Hultman, C. M.,
Ingason, A., Jablensky, A. V.,
Jakobsen, K. D., Jay, M., Jurgens, G.,
Kahn, R. S., Keller, M. C., Kenis, G.,
Kenny, E., Kim, Y., Kirov, G. K.,
Konnerth, H., Konte, B., Krabbendam,
L., Krasucki, R., Lasseter, V. K.,
Laurent, C., Lawrence, J., Lencz, T.,
Lerer, F. B., Liang, K. Y., Lichtenstein,
P., Lieberman, J. A., Linszen, D. H.,
Lonnqvist, J., Loughland, C. M.,
Maclean, A. W., Maher, B. S., Maier,
W., Mallet, J., Malloy, P., Mattheisen,
M., Mattingsdal, M., McGhee, K. A.,
McGrath, J. J., McIntosh, A., McLean,
D. E., McQuillin, A., Melle, I., Michie,
P. T., Milanova, V., Morris, D. W.,
Mors, O., Mortensen, P. B., Moskvina,
V., Muglia, P., Myin-Germeys, I.,
Nertney, D. A., Nestadt, G., Nielsen, J.,
Nikolov, I., Nordentoft, M., Norton,
N., Nothen, M. M., O'Dushlaine, C. T.,
Olincy, A., Olsen, L., O'Neill, F. A.,
Orntoft, T. F., Owen, M. J., Pantelis,
C., Papadimitriou, G., Pato, M. T.,
Peltonen, L., Petursson, H., Pickard,
B., Pimm, J., Pulver, A. E., Puri, V.,
Quested, D., Quinn, E. M., Rasmussen,
H. B., Rethelyi, J. M., Ribble, R.,
Rietschel, M., Riley, B. P., Ruggeri, M.,
Schall, U., Schulze, T. G., Schwab, S.
G., Scott, R. J., Shi, J., Sigurdsson, E.,
Silverman, J. M., Spencer, C. C.,
Stefansson, K., Strange, A.,
Strengman, E., Stroup, T. S., Suvisaari,
J., Terenius, L., Thirumalai, S.,
Thygesen, J. H., Timm, S., Toncheva,

D., van den Oord, E., van Os, J., van
Winkel, R., Veldink, J., Walsh, D.,
Wang, A. G., Wiersma, D.,
Wildenauer, D. B., Williams, H. J.,
Williams, N. M., Wormley, B., Zammit,
S., Sullivan, P. F., O'Donovan, M. C.,
Daly, M. J. and Gejman, P. V. (2011).
Genome-wide association study
identifies five new schizophrenia loci.
Nat Genet. 43, 969-76

Ripke, S., O'Dushlaine, C., Chambert,
K., Moran, J. L., Kahler, A. K., Akterin,
S., Bergen, S. E., Collins, A. L.,
Crowley, J. J., Fromer, M., Kim, Y., Lee,
S. H., Magnusson, P. K., Sanchez, N.,
Stahl, E. A., Williams, S., Wray, N. R.,
Xia, K., Bettella, F., Borglum, A. D.,
Bulik-Sullivan, B. K., Cormican, P.,
Craddock, N., de Leeuw, C., Durmishi,
N., Gill, M., Golimbet, V., Hamshere,
M. L., Holmans, P., Hougaard, D. M.,
Kendler, K. S., Lin, K., Morris, D. W.,
Mors, O., Mortensen, P. B., Neale, B.
M., O'Neill, F. A., Owen, M. J.,
Milovancevic, M. P., Posthuma, D.,
Powell, J., Richards, A. L., Riley, B. P.,
Ruderfer, D., Rujescu, D., Sigurdsson,
E., Silagadze, T., Smit, A. B.,
Stefansson, H., Steinberg, S.,
Suvisaari, J., Tosato, S., Verhage, M.,
Walters, J. T., Multicenter Genetic
Studies of Schizophrenia, C., Levinson,
D. F., Gejman, P. V., Kendler, K. S.,
Laurent, C., Mowry, B. J., O'Donovan,
M. C., Owen, M. J., Pulver, A. E., Riley,
B. P., Schwab, S. G., Wildenauer, D. B.,
Dudbridge, F., Holmans, P., Shi, J.,
Albus, M., Alexander, M., Campion,
D., Cohen, D., Dikeos, D., Duan, J.,
Eichhammer, P., Godard, S., Hansen,
M., Lerer, F. B., Liang, K. Y., Maier, W.,

Mallet, J., Nertney, D. A., Nestadt, G., Norton, N., O'Neill, F. A., Papadimitriou, G. N., Ribble, R., Sanders, A. R., Silverman, J. M., Walsh, D., Williams, N. M., Wormley, B., Psychosis Endophenotypes International, C., Arranz, M. J., Bakker, S., Bender, S., Bramon, E., Collier, D., Crespo-Facorro, B., Hall, J., Iyegbe, C., Jablensky, A., Kahn, R. S., Kalaydjieva, L., Lawrie, S., Lewis, C. M., Lin, K., Linszen, D. H., Mata, I., McIntosh, A., Murray, R. M., Ophoff, R. A., Powell, J., Rujescu, D., Van Os, J., Walshe, M., Weisbrod, M., Wiersma, D., Wellcome Trust Case Control, C., Donnelly, P., Barroso, I., Blackwell, J. M., Bramon, E., Brown, M. A., Casas, J. P., Corvin, A. P., Deloukas, P., Duncanson, A., Jankowski, J., Markus, H. S., Mathew, C. G., Palmer, C. N., Plomin, R., Rautanen, A., Sawcer, S. J., Trembath, R. C., Viswanathan, A. C., Wood, N. W., Spencer, C. C., Band, G., Bellenguez, C., Freeman, C., Hellenthal, G., Giannoulatou, E., Pirinen, M., Pearson, R. D., Strange, A., Su, Z., Vukcevic, D., Donnelly, P., Langford, C., Hunt, S. E., Edkins, S., Gwilliam, R., Blackburn, H., Bumpstead, S. J., Dronov, S., Gillman, M., Gray, E., Hammond, N., Jayakumar, A., McCann, O. T., Liddle, J., Potter, S. C., Ravindrarajah, R., Ricketts, M., Tashakkori-Ghanbaria, A., Waller, M. J., Weston, P., Widaa, S., Whittaker, P., Barroso, I., Deloukas, P., Mathew, C. G., Blackwell, J. M., Brown, M. A., Corvin, A. P., McCarthy, M. I., Spencer, C. C., Bramon, E., Corvin, A. P., O'Donovan, M. C., Stefansson, K., Scolnick, E.,

Purcell, S., McCarroll, S. A., Sklar, P., Hultman, C. M. and Sullivan, P. F. (2013). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Genet*. 45, 1150-9

Risch, N. and Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science*. 273, 1516-7.

Robinson, D. G., Woerner, M. G., McMeniman, M., Mendelowitz, A. and Bilder, R. M. (2004). Symptomatic and functional recovery from a first episode of schizophrenia or schizoaffective disorder. *Am J Psychiatry*. 161, 473-9

Rousseau, F., Rouillard, P., Morel, M. L., Khandjian, E. W. and Morgan, K. (1995). Prevalence of carriers of premutation-size alleles of the FMRI gene--and implications for the population genetics of the fragile X syndrome. *Am J Hum Genet*. 57, 1006-18

Ruano, D., Aulchenko, Y. S., Macedo, A., Soares, M. J., Valente, J., Azevedo, M. H., Hutz, M. H., Gama, C. S., Lobato, M. I., Belmonte-de-Abreu, P., Goodman, A. B., Pato, C., Heutink, P. and Palha, J. A. (2008). Association of the gene encoding neurogranin with schizophrenia in males. *Journal of psychiatric research*. 42, 125-33

Rujescu, D., Ingason, A., Cichon, S., Pietilainen, O. P., Barnes, M. R., Toulopoulou, T., Picchioni, M., Vassos, E., Ettinger, U., Bramon, E., Murray,

R., Ruggeri, M., Tosato, S., Bonetto, C., Steinberg, S., Sigurdsson, E., Sigmundsson, T., Petursson, H., Gylfason, A., Olason, P. I., Hardarsson, G., Jonsdottir, G. A., Gustafsson, O., Fossdal, R., Giegling, I., Moller, H. J., Hartmann, A., Hoffmann, P., Crombie, C., Fraser, G., Walker, N., Lonnqvist, J., Suvisaari, J., Tuulio-Henriksson, A., Andreassen, O. A., Djurovic, S., Hansen, T., Werge, T., Melle, I., Kiemeney, L. A., Franke, B., Buizer-Voskamp, J. E., Ophoff, R. A., Rietschel, M., Nothen, M. M., Stefansson, K., Peltonen, L., St Clair, D., Stefansson, H. and Collier, D. A. (2008). Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum Mol Genet*.

Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., Sherry, S., Mullikin, J. C., Mortimore, B. J., Willey, D. L., Hunt, S. E., Cole, C. G., Coggill, P. C., Rice, C. M., Ning, Z., Rogers, J., Bentley, D. R., Kwok, P. Y., Mardis, E. R., Yeh, R. T., Schultz, B., Cook, L., Davenport, R., Dante, M., Fulton, L., Hillier, L., Waterston, R. H., McPherson, J. D., Gilman, B., Schaffner, S., Van Etten, W. J., Reich, D., Higgins, J., Daly, M. J., Blumenstiel, B., Baldwin, J., Stange-Thomann, N., Zody, M. C., Linton, L., Lander, E. S., Altshuler, D. and International, S. N. P. M. W. G. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*. 409, 928-33

Saha, S., Chant, D., Welham, J. and McGrath, J. (2005). A systematic review of the prevalence of schizophrenia. *PLoS medicine*. 2, e141

Sahoo, T., Theisen, A., Rosenfeld, J. A., Lamb, A. N., Ravnan, J. B., Schultz, R. A., Torchia, B. S., Neill, N., Casci, I., Bejjani, B. A. and Shaffer, L. G. (2011). Copy number variants of schizophrenia susceptibility loci are associated with a spectrum of speech and developmental delays and behavior problems. *Genet Med*. 13, 868-80

Sajantila, A., Lahermo, P., Anttinen, T., Lukka, M., Sistonen, P., Savontaus, M. L., Aula, P., Beckman, L., Tranebjaerg, L., Gedde-Dahl, T., Issel-Tarver, L., DiRienzo, A. and Paabo, S. (1995). Genes and languages in Europe: an analysis of mitochondrial lineages. *Genome Res*. 5, 42-52

Sajantila, A., Salem, A. H., Savolainen, P., Bauer, K., Gierig, C. and Paabo, S. (1996). Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. *Proc Natl Acad Sci U S A*. 93, 12035-9

Sanders, S. J., Murtha, M. T., Gupta, A. R., Murdoch, J. D., Raubeson, M. J., Willsey, A. J., Ercan-Sencicek, A. G., DiLullo, N. M., Parikshak, N. N., Stein, J. L., Walker, M. F., Ober, G. T., Teran, N. A., Song, Y., El-Fishawy, P., Murtha, R. C., Choi, M., Overton, J. D., Bjornson, R. D., Carriero, N. J., Meyer, K. A., Bilguvar, K., Mane, S. M., Sestan, N., Lifton, R. P., Gunel, M.,

Roeder, K., Geschwind, D. H., Devlin, B. and State, M. W. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature*. 485, 237-41

Sanyal, A., Lajoie, B. R., Jain, G. and Dekker, J. (2012). The long-range interaction landscape of gene promoters. *Nature*. 489, 109-13

Schenck, A., Bardoni, B., Moro, A., Bagni, C. and Mandel, J. L. (2001). A highly conserved protein family interacting with the fragile X mental retardation protein (FMRP) and displaying selective interactions with FMRP-related proteins FXR1P and FXR2P. *Proc Natl Acad Sci U S A*. 98, 8844-9

Schizophrenia Working Group of the Psychiatric Genomics, C. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 511, 421-7

Schork, N. J., Murray, S. S., Frazer, K. A. and Topol, E. J. (2009). Common vs. rare allele hypotheses for complex diseases. *Current opinion in genetics & development*. 19, 212-9

Scott, D. A., Carmi, R., Elbedour, K., Duyk, G. M., Stone, E. M. and Sheffield, V. C. (1995). Nonsyndromic autosomal recessive deafness is linked to the DFNB1 locus in a large inbred Bedouin family from Israel. *Am J Hum Genet*. 57, 965-8

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Maner, S., Massa, H., Walker, M., Chi, M., Navin, N., Lucito, R., Healy, J., Hicks, J., Ye, K., Reiner, A., Gilliam, T. C., Trask, B., Patterson, N., Zetterberg, A. and Wigler, M. (2004). Large-scale copy number polymorphism in the human genome. *Science*. 305, 525-8

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J., Leotta, A., Pai, D., Zhang, R., Lee, Y. H., Hicks, J., Spence, S. J., Lee, A. T., Puura, K., Lehtimaki, T., Ledbetter, D., Gregersen, P. K., Bregman, J., Sutcliffe, J. S., Jobanputra, V., Chung, W., Warburton, D., King, M. C., Skuse, D., Geschwind, D. H., Gilliam, T. C., Ye, K. and Wigler, M. (2007). Strong association of de novo copy number mutations with autism. *Science*. 316, 445-9

Service, S., DeYoung, J., Karayiorgou, M., Roos, J. L., Pretorious, H., Bedoya, G., Ospina, J., Ruiz-Linares, A., Macedo, A., Palha, J. A., Heutink, P., Aulchenko, Y., Oostra, B., van Duijn, C., Jarvelin, M. R., Varilo, T., Peddle, L., Rahman, P., Piras, G., Monne, M., Murray, S., Galver, L., Peltonen, L., Sabatti, C., Collins, A. and Freimer, N. (2006). Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. *Nat Genet*. 38, 556-60

Shaikh, T. H., Kurahashi, H., Saitta, S. C., O'Hare, A. M., Hu, P., Roe, B. A., Driscoll, D. A., McDonald-McGinn, D. M., Zackai, E. H., Budarf, M. L. and Emanuel, B. S. (2000). Chromosome 22-specific low copy repeats and the 22q11.2 deletion syndrome: genomic organization and deletion endpoint analysis. *Hum Mol Genet*. 9, 489-501

Sharp, A. J., Locke, D. P., McGrath, S. D., Cheng, Z., Bailey, J. A., Vallente, R. U., Pertz, L. M., Clark, R. A., Schwartz, S., Segraves, R., Oseroff, V. V., Albertson, D. G., Pinkel, D. and Eichler, E. E. (2005). Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet*. 77, 78-88

Sharp, A. J., Hansen, S., Selzer, R. R., Cheng, Z., Regan, R., Hurst, J. A., Stewart, H., Price, S. M., Blair, E., Hennekam, R. C., Fitzpatrick, C. A., Segraves, R., Richmond, T. A., Guiver, C., Albertson, D. G., Pinkel, D., Eis, P. S., Schwartz, S., Knight, S. J. and Eichler, E. E. (2006). Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nat Genet*. 38, 1038-42

Sharp, A. J., Mefford, H. C., Li, K., Baker, C., Skinner, C., Stevenson, R. E., Schroer, R. J., Novara, F., De Gregori, M., Ciccone, R., Broomer, A., Casuga, I., Wang, Y., Xiao, C., Barbacioru, C., Gimelli, G., Bernardina, B. D., Torniero, C., Giorda, R., Regan, R., Murday, V., Mansour, S., Fichera, M., Castiglia, L., Failla, P., Ventura, M.,

Jiang, Z., Cooper, G. M., Knight, S. J., Romano, C., Zuffardi, O., Chen, C., Schwartz, C. E. and Eichler, E. E. (2008). A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. *Nat Genet*. 40, 322-8

Shatunov, A., Mok, K., Newhouse, S., Weale, M. E., Smith, B., Vance, C., Johnson, L., Veldink, J. H., van Es, M. A., van den Berg, L. H., Robberecht, W., Van Damme, P., Hardiman, O., Farmer, A. E., Lewis, C. M., Butler, A. W., Abel, O., Andersen, P. M., Fogh, I., Silani, V., Chio, A., Traynor, B. J., Melki, J., Meininger, V., Landers, J. E., McGuffin, P., Glass, J. D., Pall, H., Leigh, P. N., Hardy, J., Brown, R. H., Jr., Powell, J. F., Orrell, R. W., Morrison, K. E., Shaw, P. J., Shaw, C. E. and Al-Chalabi, A. (2010). Chromosome 9p21 in sporadic amyotrophic lateral sclerosis in the UK and seven other countries: a genome-wide association study. *Lancet neurology*. 9, 986-94

Shi, J., Levinson, D. F., Duan, J., Sanders, A. R., Zheng, Y., Pe'er, I., Dudbridge, F., Holmans, P. A., Whittemore, A. S., Mowry, B. J., Olincy, A., Amin, F., Cloninger, C. R., Silverman, J. M., Buccola, N. G., Byerley, W. F., Black, D. W., Crowe, R. R., Oksenberg, J. R., Mirel, D. B., Kendler, K. S., Freedman, R. and Gejman, P. V. (2009). Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature*.

Siomi, M. C., Zhang, Y., Siomi, H. and Dreyfuss, G. (1996). Specific sequences in the fragile X syndrome protein FMR1 and the FXR proteins mediate their binding to 60S ribosomal subunits and the interactions among them. *Molecular and cellular biology*. 16, 3825-32

Sitskoorn, M. M., Aleman, A., Ebisch, S. J., Appels, M. C. and Kahn, R. S. (2004). Cognitive deficits in relatives of patients with schizophrenia: a meta-analysis. *Schizophr Res*. 71, 285-95

Sklar, P., Smoller, J. W., Fan, J., Ferreira, M. A., Perlis, R. H., Chambert, K., Nimgaonkar, V. L., McQueen, M. B., Faraone, S. V., Kirby, A., de Bakker, P. I., Ogdie, M. N., Thase, M. E., Sachs, G. S., Todd-Brown, K., Gabriel, S. B., Sougnez, C., Gates, C., Blumenstiel, B., Defelice, M., Ardlie, K. G., Franklin, J., Muir, W. J., McGhee, K. A., MacIntyre, D. J., McLean, A., VanBeck, M., McQuillin, A., Bass, N. J., Robinson, M., Lawrence, J., Anjorin, A., Curtis, D., Scolnick, E. M., Daly, M. J., Blackwood, D. H., Gurling, H. M. and Purcell, S. M. (2008). Whole-genome association study of bipolar disorder. *Mol Psychiatry*. 13, 558-69

Smith, E. N., Chen, W., Kahonen, M., Kettunen, J., Lehtimaki, T., Peltonen, L., Raitakari, O. T., Salem, R. M., Schork, N. J., Shaw, M., Srinivasan, S. R., Topol, E. J., Viikari, J. S., Berenson, G. S. and Murray, S. S. (2010). Longitudinal genome-wide association of cardiovascular disease risk factors in the Bogalusa heart study. *PLoS Genet*. 6,

Sohl, G., Maxeiner, S. and Willecke, K. (2005). Expression and functions of neuronal gap junctions. *Nat Rev Neurosci*. 6, 191-200

Sorensen, H. J., Mortensen, E. L., Schiffman, J., Reinisch, J. M., Maeda, J. and Mednick, S. A. (2010). Early developmental milestones and risk of schizophrenia: a 45-year follow-up of the Copenhagen Perinatal Cohort. *Schizophr Res*. 118, 41-7

Sorri, M. and Rantakallio, P. (1985). Prevalence of hearing loss at the age of 15 in a birth cohort of 12 000 children from northern Finland. *Scandinavian audiology*. 14, 203-7

Sovio, U., King, V., Miettunen, J., Ek, E., Laitinen, J., Joukamaa, M., Veijola, J. and Jarvelin, M. R. (2007). Cloninger's Temperament dimensions, socio-economic and lifestyle factors and metabolic syndrome markers at age 31 years in the Northern Finland Birth Cohort 1966. *J Health Psychol*. 12, 371-82

Splawski, I., Timothy, K. W., Sharpe, L. M., Decher, N., Kumar, P., Bloise, R., Napolitano, C., Schwartz, P. J., Joseph, R. M., Condouris, K., Tager-Flusberg, H., Priori, S. G., Sanguinetti, M. C. and Keating, M. T. (2004). Ca(V)1.2 calcium channel dysfunction causes a multisystem disorder including

arrhythmia and autism. *Cell*. 119, 19-31

Stankiewicz, P. and Lupski, J. R. (2002). Genome architecture, rearrangements and genomic disorders. *Trends Genet*. 18, 74-82

Stefansson, H., Rujescu, D., Cichon, S., Pietilainen, O. P., Ingason, A., Steinberg, S., Fossdal, R., Sigurdsson, E., Sigmundsson, T., Buizer-Voskamp, J. E., Hansen, T., Jakobsen, K. D., Muglia, P., Francks, C., Matthews, P. M., Gylfason, A., Halldorsson, B. V., Gudbjartsson, D., Thorgeirsson, T. E., Sigurdsson, A., Jonasdottir, A., Jonasdottir, A., Bjornsson, A., Mattiasdottir, S., Blondal, T., Haraldsson, M., Magnusdottir, B. B., Giegling, I., Moller, H. J., Hartmann, A., Shianna, K. V., Ge, D., Need, A. C., Crombie, C., Fraser, G., Walker, N., Lonnqvist, J., Suvisaari, J., Tuulio-Henriksson, A., Paunio, T., Toulopoulou, T., Bramon, E., Di Forti, M., Murray, R., Ruggeri, M., Vassos, E., Tosato, S., Walshe, M., Li, T., Vasilescu, C., Muhleisen, T. W., Wang, A. G., Ullum, H., Djurovic, S., Melle, I., Olesen, J., Kiemeney, L. A., Franke, B., Sabatti, C., Freimer, N. B., Gulcher, J. R., Thorsteinsdottir, U., Kong, A., Andreassen, O. A., Ophoff, R. A., Georgi, A., Rietschel, M., Werge, T., Petursson, H., Goldstein, D. B., Nothen, M. M., Peltonen, L., Collier, D. A., St Clair, D. and Stefansson, K. (2008). Large recurrent microdeletions associated with schizophrenia. *Nature*. 455, 232-6

Stefansson, H., Ophoff, R. A., Steinberg, S., Andreassen, O. A., Cichon, S., Rujescu, D., Werge, T., Pietilainen, O. P., Mors, O., Mortensen, P. B., Sigurdsson, E., Gustafsson, O., Nyegaard, M., Tuulio-Henriksson, A., Ingason, A., Hansen, T., Suvisaari, J., Lonnqvist, J., Paunio, T., Borglum, A. D., Hartmann, A., Fink-Jensen, A., Nordentoft, M., Hougaard, D., Norgaard-Pedersen, B., Bottcher, Y., Olesen, J., Breuer, R., Moller, H. J., Giegling, I., Rasmussen, H. B., Timm, S., Mattheisen, M., Bitter, I., Rethelyi, J. M., Magnusdottir, B. B., Sigmundsson, T., Olason, P., Masson, G., Gulcher, J. R., Haraldsson, M., Fossdal, R., Thorgeirsson, T. E., Thorsteinsdottir, U., Ruggeri, M., Tosato, S., Franke, B., Strengman, E., Kiemeney, L. A., Group, Melle, I., Djurovic, S., Abramova, L., Kaleda, V., Sanjuan, J., de Frutos, R., Bramon, E., Vassos, E., Fraser, G., Ettinger, U., Picchioni, M., Walker, N., Toulopoulou, T., Need, A. C., Ge, D., Lim Yoon, J., Shianna, K. V., Freimer, N. B., Cantor, R. M., Murray, R., Kong, A., Golimbet, V., Carracedo, A., Arango, C., Costas, J., Jonsson, E. G., Terenius, L., Agartz, I., Petursson, H., Nothen, M. M., Rietschel, M., Matthews, P. M., Muglia, P., Peltonen, L., St Clair, D., Goldstein, D. B., Stefansson, K., Collier, D. A., Kahn, R. S., Linszen, D. H., van Os, J., Wiersma, D., Bruggeman, R., Cahn, W., de Haan, L., Krabbendam, L. and Myin-Germeys, I. (2009). Common variants conferring risk of schizophrenia. *Nature*.

Stefansson, H., Meyer-Lindenberg, A., Steinberg, S., Magnusdottir, B., Morgen, K., Arnarsdottir, S., Bjornsdottir, G., Walters, G. B., Jonsdottir, G. A., Doyle, O. M., Tost, H., Grimm, O., Kristjansdottir, S., Snorrason, H., Davidsdottir, S. R., Gudmundsson, L. J., Jonsson, G. F., Stefansdottir, B., Helgadottir, I., Haraldsson, M., Jonsdottir, B., Thygesen, J. H., Schwarz, A. J., Didriksen, M., Stensbol, T. B., Brammer, M., Kapur, S., Halldorsson, J. G., Hreidarsson, S., Saemundsen, E., Sigurdsson, E. and Stefansson, K. (2014). CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature*. 505, 361-6

Steinberg, S., Mors, O., Borglum, A. D., Gustafsson, O., Werge, T., Mortensen, P. B., Andreassen, O. A., Sigurdsson, E., Thorgeirsson, T. E., Bottcher, Y., Olason, P., Ophoff, R. A., Cichon, S., Gudjonsdottir, I. H., Pietilainen, O. P., Nyegaard, M., Tuulio-Henriksson, A., Ingason, A., Hansen, T., Athanasiu, L., Suvisaari, J., Lonnqvist, J., Paunio, T., Hartmann, A., Jurgens, G., Nordentoft, M., Hougaard, D., Norgaard-Pedersen, B., Breuer, R., Moller, H. J., Giegling, I., Glenthoj, B., Rasmussen, H. B., Mattheisen, M., Bitter, I., Rethelyi, J. M., Sigmundsson, T., Fossdal, R., Thorsteinsdottir, U., Ruggeri, M., Tosato, S., Strengman, E., Kiemeney, L. A., Melle, I., Djurovic, S., Abramova, L., Kaleda, V., Walshe, M., Bramon, E., Vassos, E., Li, T., Fraser, G., Walker, N., Toulopoulou, T., Yoon, J., Freimer, N. B., Cantor, R. M., Murray, R., Kong, A., Golimbet, V., Jonsson, E. G., Terenius, L., Agartz, I., Petursson, H., Nothen, M. M., Rietschel, M., Peltonen, L., Rujescu, D., Collier, D. A., Stefansson, H., St Clair, D. and Stefansson, K. (2011). Expanding the range of ZNF804A variants conferring risk of psychosis. *Mol Psychiatry*. 16, 59-66

Stone, J. L., O'Donovan M, C. and Gurling, H. (2008). Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*. 455, 237-41

Stranger, B. E., Forrest, M. S., Dunning, M., Ingle, C. E., Beazley, C., Thorne, N., Redon, R., Bird, C. P., de Grassi, A., Lee, C., Tyler-Smith, C., Carter, N., Scherer, S. W., Tavare, S., Deloukas, P., Hurles, M. E. and Dermitzakis, E. T. (2007). Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*. 315, 848-53

Strittmatter, W. J., Saunders, A. M., Schmechel, D., Pericak-Vance, M., Enghild, J., Salvesen, G. S. and Roses, A. D. (1993). Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proc Natl Acad Sci U S A*. 90, 1977-81

Stromme, P. and Magnus, P. (2000). Correlations between socioeconomic status, IQ and aetiology in mental retardation: a population-based study of Norwegian children. *Soc Psychiatry Psychiatr Epidemiol*. 35, 12-8

Sudhof, T. C. (2008). Neuroligins and neurexins link synaptic function to cognitive disease. *Nature*. 455, 903-11

Sulem, P., Gudbjartsson, D. F., Stacey, S. N., Helgason, A., Rafnar, T., Magnusson, K. P., Manolescu, A., Karason, A., Palsson, A., Thorleifsson, G., Jakobsdottir, M., Steinberg, S., Palsson, S., Jonasson, F., Sigurgeirsson, B., Thorisdottir, K., Ragnarsson, R., Benediktsdottir, K. R., Aben, K. K., Kiemeney, L. A., Olafsson, J. H., Gulcher, J., Kong, A., Thorsteinsdottir, U. and Stefansson, K. (2007). Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat Genet*. 39, 1443-52

Sullivan, P. F., Kendler, K. S. and Neale, M. C. (2003). Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry*. 60, 1187-92

Sullivan, P. F. (2005). The genetics of schizophrenia. *PLoS medicine*. 2, e212

Sullivan, P. F. (2012). Puzzling over schizophrenia: schizophrenia as a pathway disease. *Nature medicine*. 18, 210-1

Sullivan, P. F., Daly, M. J. and O'Donovan, M. (2012). Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nat Rev Genet*. 13, 537-51

Tan, T. Y., Collins, A., James, P. A., McGillivray, G., Stark, Z., Gordon, C. T., Leventer, R. J., Pope, K., Forbes, R., Crolla, J. A., Ganesamoorthy, D., Burgess, T., Bruno, D. L., Slater, H. R., Farlie, P. G. and Amor, D. J. (2011). Phenotypic variability of distal 22q11.2 copy number abnormalities. *Am J Med Genet A*. 155A, 1623-33

Tandon, R., Gaebel, W., Barch, D. M., Bustillo, J., Gur, R. E., Heckers, S., Malaspina, D., Owen, M. J., Schultz, S., Tsuang, M., Van Os, J. and Carpenter, W. (2013). Definition and description of schizophrenia in the DSM-5. *Schizophr Res*. 150, 3-10

Tange, T. O., Nott, A. and Moore, M. J. (2004). The ever-increasing complexities of the exon junction complex. *Current opinion in cell biology*. 16, 279-84

Teer, J. K. and Mullikin, J. C. (2010). Exome sequencing: the sweet spot before whole genomes. *Hum Mol Genet*. 19, R145-51

Thomas, N. S., Sharp, A. J., Browne, C. E., Skuse, D., Hardie, C. and Dennis, N. R. (1999). Xp deletions associated with autism in three females. *Human genetics*. 104, 43-8

Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E., Sheffield, N. C., Stergachis, A. B., Wang, H., Vernot, B., Garg, K., John, S., Sandstrom, R., Bates, D., Boatman, L., Canfield, T. K., Diegel, M., Dunn, D., Ebersol, A. K.,

Frum, T., Giste, E., Johnson, A. K., Johnson, E. M., Kutyavin, T., Lajoie, B., Lee, B. K., Lee, K., London, D., Lotakis, D., Neph, S., Neri, F., Nguyen, E. D., Qu, H., Reynolds, A. P., Roach, V., Safi, A., Sanchez, M. E., Sanyal, A., Shafer, A., Simon, J. M., Song, L., Vong, S., Weaver, M., Yan, Y., Zhang, Z., Zhang, Z., Lenhard, B., Tewari, M., Dorschner, M. O., Hansen, R. S., Navas, P. A., Stamatoyannopoulos, G., Iyer, V. R., Lieb, J. D., Sunyaev, S. R., Akey, J. M., Sabo, P. J., Kaul, R., Furey, T. S., Dekker, J., Crawford, G. E. and Stamatoyannopoulos, J. A. (2012). The accessible chromatin landscape of the human genome. *Nature*. 489, 75-82

Toulopoulou, T., Goldberg, T. E., Mesa, I. R., Picchioni, M., Rijsdijk, F., Stahl, D., Cherny, S. S., Sham, P., Faraone, S. V., Tsuang, M., Weinberger, D. R., Seidman, L. J. and Murray, R. M. (2010). Impaired intellect and memory: a missing link between genetic risk and schizophrenia? *Arch Gen Psychiatry*. 67, 905-13

Tuulio-Henriksson, A., Haukka, J., Partonen, T., Varilo, T., Paunio, T., Ekelund, J., Cannon, T. D., Meyer, J. M. and Lonnqvist, J. (2002). Heritability and number of quantitative trait loci of neurocognitive functions in families with schizophrenia. *American journal of medical genetics*. 114, 483-90

Tuzun, E., Sharp, A. J., Bailey, J. A., Kaul, R., Morrison, V. A., Pertz, L. M., Haugen, E., Hayden, H., Albertson, D., Pinkel, D., Olson, M. V. and Eichler, E. E. (2005). Fine-scale structural variation of the human genome. *Nature Genetics*. 37, 727-732

Vacic, V., McCarthy, S., Malhotra, D., Murray, F., Chou, H. H., Peoples, A., Makarov, V., Yoon, S., Bhandari, A., Corominas, R., Iakoucheva, L. M., Krastoshevsky, O., Krause, V., Larach-Walters, V., Welsh, D. K., Craig, D., Kelsoe, J. R., Gershon, E. S., Leal, S. M., Dell Aquila, M., Morris, D. W., Gill, M., Corvin, A., Insel, P. A., McClellan, J., King, M. C., Karayiorgou, M., Levy, D. L., DeLisi, L. E. and Sebat, J. (2011). Duplications of the neuropeptide receptor gene VIPR2 confer significant risk for schizophrenia. *Nature*. 471, 499-503

Walsh, T., McClellan, J. M., McCarthy, S. E., Addington, A. M., Pierce, S. B., Cooper, G. M., Nord, A. S., Kusenda, M., Malhotra, D., Bhandari, A., Stray, S. M., Rippey, C. F., Roccanova, P., Makarov, V., Lakshmi, B., Findling, R. L., Sikich, L., Stromberg, T., Merriman, B., Gogtay, N., Butler, P., Eckstrand, K., Noory, L., Gochman, P., Long, R., Chen, Z., Davis, S., Baker, C., Eichler, E. E., Meltzer, P. S., Nelson, S. F., Singleton, A. B., Lee, M. K., Rapoport, J. L., King, M. C. and Sebat, J. (2008). Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science*. 320, 539-43

van Os, J. and Kapur, S. (2009). Schizophrenia. *Lancet*. 374, 635-45

Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S. F., Hakonarson, H. and Bucan, M. (2007). PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res*. 17, 1665-74

Wang, K., Chen, Z., Tadesse, M. G., Glessner, J., Grant, S. F., Hakonarson, H., Bucan, M. and Li, M. (2008). Modeling genetic inheritance of copy number variations. *Nucleic Acids Res*. 36, e138

Varilo, T., Laan, M., Hovatta, I., Wiebe, V., Terwilliger, J. D. and Peltonen, L. (2000). Linkage disequilibrium in isolated populations: Finland and a young sub- population of Kuusamo. *Eur J Hum Genet*. 8, 604-12.

Varilo, T., Paunio, T., Parker, A., Perola, M., Meyer, J., Terwilliger, J. D. and Peltonen, L. (2003). The interval of linkage disequilibrium (LD) detected with microsatellite and SNP markers in chromosomes of Finnish populations with different histories. *Hum Mol Genet*. 12, 51-9.

Varilo, T. and Peltonen, L. (2004). Isolates and their potential use in complex gene mapping efforts. *Current opinion in genetics & development*. 14, 316-23

Vartiainen, E., Laatikainen, T., Peltonen, M., Juolevi, A., Mannisto, S., Sundvall, J., Jousilahti, P., Salomaa, V., Valsta, L. and Puska, P. (2010). Thirty-five-year trends in cardiovascular risk factors in Finland. *Int J Epidemiol*. 39, 504-18

Vassos, E., Collier, D. A., Holden, S., Patch, C., Rujescu, D., St Clair, D. and Lewis, C. M. (2010). Penetrance for copy number variants associated with schizophrenia. *Hum Mol Genet*. 19, 3477-81

Watson, J. D. and Crick, F. H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*. 171, 737-8

Wechsler, D. (1981). Wechsler Adult Intelligence Scale – Revised (WAIS-R), Manual. *The Psychological Corporation. Cleveland: Harcourt Brace Jovanovich, Inc*.

Wechsler, D. (1987). Weschler Memory Scale-Revised (WMS-R), Manual. *The Psychological Corporation, Harcourt Brace Jovanovich Inc., San Antonio*.

Wegelius, A., Pankakoski, M., Lehto, U., Suokas, J., Hakkinen, L., Tuulio-Henriksson, A., Lonnqvist, J., Paunio, T. and Suvisaari, J. (2013). An association between both low and high birth weight and increased disorganized and negative symptom severity in schizophrenia and other psychoses. *Psychiatry Res*. 205, 18-24

Weiss, L. A., Shen, Y., Korn, J. M., Arking, D. E., Miller, D. T., Fossdal, R., Saemundsen, E., Stefansson, H., Ferreira, M. A., Green, T., Platt, O. S., Ruderfer, D. M., Walsh, C. A., Altshuler, D., Chakravarti, A., Tanzi, R. E., Stefansson, K., Santangelo, S. L., Gusella, J. F., Sklar, P., Wu, B. L. and Daly, M. J. (2008). Association between microdeletion and microduplication at 16p11.2 and autism. *The New England journal of medicine*. 358, 667-75

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y. H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigo, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick,

L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y. H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A. and Zhu, X. (2001). The sequence of the human genome. *Science*. 291, 1304-51

Wheeler, D. A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y. J., Makhijani, V., Roth, G. T., Gomes, X., Tartaro, K., Niazi, F., Turcotte, C. L., Irzyk, G. P., Lupski, J. R., Chinault, C., Song, X. Z., Liu, Y., Yuan, Y., Nazareth, L., Qin, X., Muzny, D. M., Margulies, M., Weinstock, G. M., Gibbs, R. A. and Rothberg, J. M. (2008). The complete genome of an individual by massively parallel DNA sequencing. *Nature*. 452, 872-6

Viard, T. and de la Tour, C. B. (2007). Type IA topoisomerases: a simple puzzle? *Biochimie*. 89, 456-67

Wigginton, J. E., Cutler, D. J. and Abecasis, G. R. (2005). A note on exact tests of Hardy-Weinberg equilibrium. *Am J Hum Genet*. 76, 887-93

Williams, C. A., Driscoll, D. J. and Dagli, A. I. (2010). Clinical and genetic aspects of Angelman syndrome. *Genet Med*. 12, 385-95

Vissers, L. E., de Vries, B. B., Osoegawa, K., Janssen, I. M., Feuth, T., Choy, C. O., Straatman, H., van der Vliet, W., Huys, E. H., van Rijk, A., Smeets, D., van Ravenswaaij-Arts, C. M., Knoers, N. V., van der Burgt, I., de Jong, P. J., Brunner, H. G., van Kessel, A. G., Schoenmakers, E. F. and Veltman, J. A. (2003). Array-based comparative genomic hybridization for the genomewide detection of submicroscopic chromosomal abnormalities. *Am J Hum Genet*. 73, 1261-70

Vissers, L. E., de Ligt, J., Gilissen, C., Janssen, I., Steehouwer, M., de Vries, P., van Lier, B., Arts, P., Wieskamp, N., del Rosario, M., van Bon, B. W., Hoischen, A., de Vries, B. B., Brunner, H. G. and Veltman, J. A. (2010). A de novo paradigm for mental retardation. *Nat Genet*. 42, 1109-12

von Wendt, L., Rantakallio, P., Saukkonen, A. L., Tuisku, M. and Makinen, H. (1985). Cerebral palsy and additional handicaps in a 1-year birth cohort from northern Finland--a prospective follow-up study to the age of 14 years. *Ann Clin Res*. 17, 156-61

von Wendt, L. and Rantakallio, P. (1987). Mild mental retardation in northern Finland. *Upsala journal of medical sciences. Supplement*. 44, 47-51

Woodberry, K. A., Giuliano, A. J. and Seidman, L. J. (2008). Premorbid IQ in schizophrenia: a meta-analytic review. *Am J Psychiatry*. 165, 579-87

Xu, B., Roos, J. L., Levy, S., van Rensburg, E. J., Gogos, J. A. and Karayiorgou, M. (2008). Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet*. 40, 880-5

Xu, B., Roos, J. L., Dexheimer, P., Boone, B., Plummer, B., Levy, S., Gogos, J. A. and Karayiorgou, M. (2011). Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nat Genet*. 43, 864-8

Xu, B., Ionita-Laza, I., Roos, J. L., Boone, B., Woodrick, S., Sun, Y., Levy, S., Gogos, J. A. and Karayiorgou, M. (2012). De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nat Genet*. 44, 1365-9

Xu, D. (2013). Top3B is an RNA topoisomerase that works with Fragile X syndrome protein to promote synapse formation. *Nature Neuroscience*. 16,

Yang, K., Trepanier, C. H., Li, H., Beazely, M. A., Lerner, E. A., Jackson, M. F. and MacDonald, J. F. (2009).

Vasoactive intestinal peptide acts via multiple signal pathways to regulate hippocampal NMDA receptors and synaptic transmission. *Hippocampus*. 19, 779-89

Zeng, Z., Wang, T., Li, T., Li, Y., Chen, P., Zhao, Q., Liu, J., Li, J., Feng, G., He, L. and Shi, Y. (2011). Common SNPs and haplotypes in DGKH are associated with bipolar disorder and schizophrenia in the Chinese Han population. *Mol Psychiatry*. 16, 473-5

Zhang, D., Cheng, L., Qian, Y., Alliey-Rodriguez, N., Kelsoe, J. R., Greenwood, T., Nievergelt, C., Barrett, T. B., McKinney, R., Schork, N., Smith, E. N., Bloss, C., Nurnberger, J., Edenberg, H. J., Foroud, T., Sheftner, W., Lawson, W. B., Nwulia, E. A., Hipolito, M., Coryell, W., Rice, J., Byerley, W., McMahon, F., Schulze, T. G., Berrettini, W., Potash, J. B., Belmonte, P. L., Zandi, P. P., McInnis, M. G., Zollner, S., Craig, D., Szelinger, S., Koller, D., Christian, S. L., Liu, C. and Gershon, E. S. (2009). Singleton deletions throughout the genome increase risk of bipolar disorder. *Mol Psychiatry*. 14, 376-80