



UPPSALA
UNIVERSITET

*Digital Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Medicine 1231*

Sequence based analysis of neurodevelopmental disorders

JONATAN HALVARDSON



ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2016

ISSN 1651-6206
ISBN 978-91-554-9597-8
urn:nbn:se:uu:diva-287407

Dissertation presented at Uppsala University to be publicly examined in C2:305, BMC, Husargatan 3, Uppsala, Tuesday, 14 June 2016 at 09:30 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English. Faculty examiner: Professor Alexandre Reymond (Université de Lausanne).

Abstract

Halvardson, J. 2016. Sequence based analysis of neurodevelopmental disorders. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* 1231. 62 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-554-9597-8.

In this thesis the main focus is the use of methods and applications of next generation sequencing in order to study three of the most common neurodevelopmental disorders: intellectual disability, epilepsy and schizophrenia. A large fraction of the genes in our genome produce several distinct transcript isoforms through the process of splicing and there is an increasing amount of evidence pinpointing mutations affecting splicing as a mechanism of disease. In **Paper I** we used exome capture of RNA in combination with sequencing in order to enrich for coding sequences. We show that this approach enables us to detect lowly expressed transcript and splice events that would have been missed in regular RNA sequencing using the same coverage. In **Paper II** we selectively depleted the different transcripts of *Quaking (QKI)*, a gene previously associated to schizophrenia. Using RNA sequencing we show that the effects of depletion differ between transcripts and that the *QKI* gene is a potential regulator of the *Glial Fibrillary Acidic Protein (GFAP)*, a gene implicated in several diseases in the central nervous system.

De-novo mutations are frequently reported to be causative in neurodevelopmental disorders with a strong genetic component, such as epilepsy and intellectual disability. In **Paper III** we used exome sequencing in family trios where the child was diagnosed with both intellectual disability and epilepsy, focusing on finding *de-novo* mutations. We identified several previously unknown disease causing mutations in genes previously known to cause disease and used previously published interaction and mutation data to prioritize novel candidate genes. The most interesting result from this study are the implication of the *HECW2* gene as a candidate gene in intellectual disability and epilepsy. In **Paper IV** we used RNA sequencing of post mortem brain tissue in a large cohort of schizophrenics and controls. In this study we could show that the immune system and more specifically the complement system was dysregulated in a large fraction of patients. Further, using co-expression network we also found some evidence suggesting genes involved in axon development and maintenance.

Jonatan Halvardson, Department of Immunology, Genetics and Pathology, Rudbecklaboratoriet, Uppsala University, SE-751 85 Uppsala, Sweden.

© Jonatan Halvardson 2016

ISSN 1651-6206

ISBN 978-91-554-9597-8

urn:nbn:se:uu:diva-287407 (<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-287407>)

List of Papers

This thesis is based on the following Papers, which are referred to in the text by their Roman numerals.

- I **Halvardson J***, Zaghlool A*, Feuk L. (2013) Exome RNA sequencing reveals rare and novel alternative transcripts. *Nucleic Acids Research* 41.1:e6-e6.
- II Radomska KJ, **Halvardson J**, Reinius B, Lindholm Carlström E, Emilsson L, Feuk L, Jazin E. (2013) RNA-binding protein QKI regulates Glial fibrillary acidic protein expression in human astrocytes. *Human Molecular Genetics* 22.7:1373-1382.
- III **Halvardson J**, Zhao J, Zaghlool A, Wentzel C, Georgii-Hemming P, Månsson E, Ederth Sävmarker H, Brandberg G, Soussi Zander C, Thuresson A & Feuk L, (2016) Identification of new candidate genes in intellectual disability and epilepsy, *Journal of medical genetics*, Submitted
- IV Carlström E*, **Halvardsson J***, Etermadikhah M, Rajkowska G, Stockmeier C. A, Feuk L. (2016) Transcriptome sequencing implicates alterations of complement factors levels in schizophrenia, *Manuscript*

*Equal contribution

Related publications

Ameur A, Zaghlool A, **Halvardson J**, Wetterbom A, Gyllensten U, Cavelier L & Feuk L. (2011) Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nature Structural & Molecular Biology* 18(12), pp.1435–1440.

Berger I, Dor T, **Halvardson J**, Edvardson S, Shaag A, Feuk L & Elpeleg O. (2012) Intractable epilepsy of infancy due to homozygous mutation in the EFHC1 gene. *Epilepsia* 53(8), pp.1436–40.

Spiegel R, Pines O, Ta-Shma A, Burak E, Shaag A, **Halvardson J**, Edvardson S, Mahajna M, Zenvirt S, Saada A, Shalev S, Feuk L & Elpeleg O, (2012) Infantile cerebellar-retinal degeneration associated with a mutation in mitochondrial aconitase, ACO2. *American journal of human genetics*, 90(3), pp.518–23.

Zaghlool A, Ameur A, Nyberg L, **Halvardson J**, Grabherr M, Cavelier L & Feuk L. (2013) Efficient cellular fractionation improves RNA sequencing analysis of mature and nascent transcripts from human tissues. *BMC biotechnology*, 13(1), p.99.

Spiegel R, Saada A, **Halvardson J**, Soiferman D, Shaag A, Edvardson S, Horovitz Y, Khayat M, Shalev S.A, Feuk L & Elpeleg O, (2014) Deleterious mutation in FDX1L gene is associated with a novel mitochondrial muscle myopathy. *European journal of human genetics: EJHG*, 22(7), pp.902–6.

Schuster J, **Halvardson J**, Pilar Lorenzo L, Ameur A, Sobol M, Raykova D, Annerén G, Feuk L & Dahl N, (2015) Transcriptome Profiling Reveals Degree of Variability in Induced Pluripotent Stem Cell Lines: Impact for Human Disease Modeling. *Cellular reprogramming*, 17(5), pp.327–37.

Johansson M.M, Lundin E, Qian X, Mirzazadeh M, **Halvardson J**, Darj E, Feuk L, Nilsson M & Jazin E. (2016) Spatial sexual dimorphism of X and Y homolog gene expression in the human central nervous system during early male development. *Biology of sex differences*, 7, p.5.

Contents

| | |
|---|----|
| Introduction..... | 9 |
| From DNA to transcription..... | 9 |
| Variation in the human genome..... | 10 |
| Genetic variation in human disease..... | 12 |
| Inheritance in genetic disease..... | 13 |
| Regulation of expression as a mechanism in disease..... | 14 |
| Tools used in the study of genetic disorders..... | 16 |
| Linkage and genome wide association studies..... | 16 |
| Next generation sequencing..... | 16 |
| Whole exome sequencing..... | 18 |
| Whole genome sequencing..... | 18 |
| Sequence data analysis..... | 19 |
| Transcriptome sequencing..... | 20 |
| Network analysis..... | 21 |
| Intellectual disability, epilepsy and Schizophrenia..... | 23 |
| Schizophrenia..... | 23 |
| Genetics of Schizophrenia..... | 23 |
| Intellectual disability..... | 24 |
| Genetics of intellectual disability..... | 24 |
| Epilepsy..... | 26 |
| Shared genetics of ID, epilepsy and schizophrenia..... | 26 |
| Methods..... | 28 |
| SOLiD sequencing..... | 28 |
| Whole exome sequencing..... | 28 |
| Transcriptome sequencing..... | 30 |
| Concluding remarks and future perspectives..... | 31 |
| Summary of included Papers..... | 34 |
| Paper I. Exome RNA sequencing reveals rare and novel alternative transcripts..... | 34 |
| Paper II: RNA-binding protein QKI regulates <i>Glial fibrillary acidic protein</i> expression in human astrocytes..... | 37 |
| Paper III: Identification of new candidate genes in intellectual disability and epilepsy..... | 39 |

| | |
|---|-----------|
| Paper IV: Transcriptome sequencing implicates alterations of complement factors levels in schizophrenia..... | 41 |
| Acknowledgements..... | 43 |
| References..... | 44 |

Abbreviations

| | |
|--------|--|
| A | Adenine |
| Bp | Base pair |
| C | Cytosine |
| CADD | Combined annotation dependent depletion |
| CNV | Copy number variant |
| dbSNP | Database of single nucleotide polymorphism |
| DGV | Database of genomic variants |
| DNA | Deoxyribonucleic acid |
| eQTL | expression Quantitative Trait Locus |
| ExAC | Exome aggregation consortium |
| FATHMM | Functional Analysis through Hidden Markov Models |
| G | Guanine |
| GO | Gene Ontology |
| GWAS | Genome Wide Association Study |
| HGP | Human Genome Project |
| ID | Intellectual Disability |
| Indel | Insertion or deletion |
| Kb | Kilobases (1000 base pairs) |
| LOD | Logarithm of the odds |
| mRNA | Messenger RNA |
| NMD | Nonsense Mediated Decay |
| PCR | Polymerase chain reaction |
| RNA | Ribonucleic acid |
| SNV | Single Nucleotide Variant |
| T | Thymine |
| U | Uracil |
| WES | Whole exome sequencing |
| WGS | Whole genome sequencing |

Introduction

From DNA to transcription

The human genome contains all biological information necessary for a human to live, grow and develop. The genetic instructions are encoded by Deoxyribonucleic acid (DNA) molecules. These molecules consist of four distinct bases Adenine (A), Guanine (G), Thymine (T) and Cytosine (C), chained together via a phosphate backbone. Most cellular DNA is double stranded, meaning that two DNA molecules are bound together creating a double helix structure (*Figure 1*). Due to the chemical properties of each base Adenine can only create a pair with Thymine and Guanine with Cytosine. This specific base pairing forces paired DNA molecules to be complementary, and from the order of bases of either strand, the bases on the second strand can easily be determined. The sequence of bases along the DNA strand constitute the genetic code that contain the instructions for the creation of every molecule the cell produce. Despite extensive efforts, there are still many aspects of this code that are not understood.

In the cell the DNA occurs as long molecules called chromosomes, and a normal human cell contains 46 chromosomes. The 46 chromosomes include 22 pairs of homologous chromosomes called autosomes, as well as the sex determining chromosomes, the X and Y chromosomes. We inherit 23 chromosomes from each parent. In addition, the cell contains DNA in the mitochondria, which are exclusively inherited from the mother.

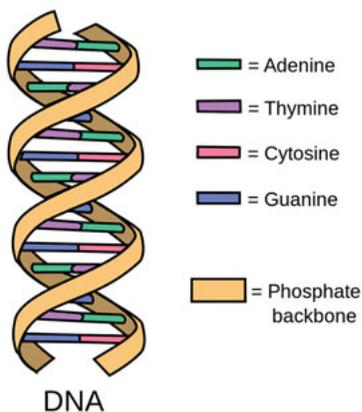


Figure 1: The structure of the DNA molecule

The chromosomes contain genes, which correspond to stretches along the chromosomes that are frequently expressed, meaning that the DNA is used as a template to create an RNA molecule. Like DNA, RNA consists of a chain of nucleotides. However, instead of Cytosine the RNA introduces a Uracil (U) and it also differs from DNA in that it is a single

stranded molecule. The RNA has many important roles in the cell. The classic role of RNA is to function as a carrier of the information required to create a specific protein. However, RNA can also be functional in itself as non-protein coding RNA, a class of molecules that are crucial for controlling or performing many cellular processes. The exact number of genes in our genome is still debated, however with the completion of the Human Genome Project (HGP) in 2004, the number of protein coding genes was reported to be approximately 20,000 – 25,000 (Human Genome Sequencing Consortium 2004).

Most human protein coding genes contain introns and exons, and specific sequences of regulatory DNA inside or around the gene determine when and how much the gene will be expressed (Maston et al. 2006). When a protein coding gene is expressed both the exons and introns are transcribed to RNA, however the resulting transcript is usually processed. In the processing step the introns are removed, and modifications are added to the ends of the transcript. The mature processed transcript is called messenger RNA (mRNA) and contains the information necessary to create a protein.

The removal of introns from a transcript is called splicing. The process of splicing does not only remove introns, but in some cases it may also remove exons in a process referred to as alternative splicing. Alternative splicing gives the genome the ability to create many different transcripts from the same gene and plays an important role for the diversity of functions for genes in the human genome (Kelemen et al. 2013). There are several observations of protein isoforms produced from different splicing events in the same gene that show distinctly different properties with respect to structure, function and localization (Yang et al. 2016). Furthermore, it has been estimated that 92-94% of the human multi-exon genes produce more than one distinct protein isoform via alternative splicing (Wang et al. 2008; Pan et al. 2008).

Variation in the human genome

The ability of the genome to mutate, i.e. to change the order, content or organization of DNA, is the foundation of the evolutionary process and of our phenotypic diversity. Mutations occur spontaneously in every generation and can be either neutral, increase fitness or decrease fitness and in some cases cause disease (Fay et al. 2001). Each human carries their own range of inherited and acquired mutations and these mutations are known as genetic variation. There are several mechanisms that give rise to genetic mutations, including damage to the DNA which is not properly repaired, errors in cellular replication, or insertions of mobile genetic elements (Malkova & Haber 2012; Kaer & Speek 2013).

The most common types of mutations are single nucleotide variants (SNVs) and small insertions or deletions (indels). The SNVs are positions in the genome where one base has been exchanged for another. Indels are positions in our genome where a few bases have been either deleted, inserted or where a deletion and an insertion have occurred in the same position (Mullaney et al. 2010). It has been estimated that each individual carries an average of 38 - 73 SNVs not inherited from the parents (Besenbacher et al. 2015; Kong et al. 2012). These variants are called *de-novo* mutations, as they are new to the genome of the specific carrier and the overall mutation rate of single nucleotides in the human genome has been calculated to be $1.27e-8$ per base and generation (Besenbacher et al. 2015). The absolute majority of the variation in our genome is considered to be made up by rare variants (minor allele frequency $< 1\%$) (Tennessen et al. 2012; Marth et al. 2011). Furthermore, rare variants have been indicated to more frequently disrupt protein coding genes compared to variants with a higher minor allele frequency, suggesting that these variants are more probable to have biological or medical consequences (Nelson et al. 2012).

Another common type of mutations are large rearrangements such as copy number variation (CNVs) and translocations. CNVs are large deletions or duplications of genomic regions (Scherer et al. 2007). Translocations are mutations caused by the rearrangements of chromosomal regions between nonhomologous chromosomes (Roukos et al. 2013). The number of *de-novo* CNVs has been estimated to 1.2×10^{-2} CNVs per genome per generation making them common in our genome (Itsara et al. 2010). There is an inverse correlation between size and frequency for CNVs, with CNVs $>1\text{Mb}$ rarely seen in the normal population, indicating that longer CNVs have a greater impact on our genome than shorter ones and are therefore under negative selection (Itsara et al. 2009).

There have been significant efforts to catalogue genetic variation in the human genome resulting in the development of several large databases such as the dbSNP, DGV and ExAC (Sherry 2001; Exome Aggregation Consortium et al. 2015; Wheeler et al. 2007; MacDonald et al. 2014). These databases have acquired data from large-scale projects such as the 1000 Genomes Project, mapping single nucleotide variants (SNVs) and structural variation in 2504 individuals from 26 populations. Interestingly despite previous variation discovery efforts the 1000 Genomes Project identified 84.7 million SNVs where a large fraction of the variants had not been previously observed, underlining the vast amount of variability existing between individuals and populations (Auton et al. 2015).

These results show that each population and individual carry a wide array of inherited variants, giving each population a unique genomic background.

Genetic variation in human disease

The introduction of new mutations in the genome is a normal process contributing to diversification of phenotypes and it is one of the driving forces of evolution. However, mutations will in some cases be pathogenic and cause disease. Identifying the variants that are associated with disease is one of the main goals in genetics research and novel mutations associated with disease are now discovered at a rapid pace. Currently, the database Online Mendelian Inheritance in Man (OMIM), a highly updated database specializing in genetic disease, lists more than 3000 genes reported to induce a distinct phenotype if disrupted by a mutation (www.omim.org).

There are multiple mechanisms by which a specific mutation may cause disease, and even when the disease mutation is identified the etiology of the disease often remains elusive. For example pathological CNVs deleting or duplicating several genes can change the expression of one or several proteins involved in disease. As CNVs in many cases contain several genes it is often challenging to measure the exact contribution made by each gene to the resulting phenotype (Henrichsen, Chaignat, et al. 2009). In addition, larger structural variants may change the genomic landscape in ways that also affect the expression of genes outside the region directly affected by the aberration (Henrichsen, Vinckenbosch, et al. 2009).

Single point mutations in coding exons may have different effect depending on the type of substitution. They can either cause a premature stop codon, a non-synonymous change or a synonymous change, as well as the loss of a stop codon or cause changes in splicing. The effect of the gain of a stop codon may either lead to a truncated protein or, more commonly, to an unstable mRNA product degraded by the nonsense mediated decay pathway (NMD). Both of these mechanisms have been shown to cause disease (Holbrook et al. 2004). However, while the decay of the mRNA will lead to different protein expression levels, a truncated protein can in some cases cause disease through reduced function or changed properties (Inoue et al. 2004). Indels in coding regions often create frameshift mutations, making them similar to stop gains in terms of the effect on the protein. Both indels and the gain of a stop codon are considered to have highly disruptive effects on the gene, making them more likely to cause disease compared to non-synonymous mutations (Thusberg et al. 2011).

Disease causing non-synonymous mutations can change the properties of the protein, for example making an enzyme bind stronger to its substrate. Another way these mutations may cause disease is to make the resulting transcript enter the NMD pathway. A series of studies have observed that certain protein domains are overrepresented as carriers of disease associated de-novo mutations (Wang et al. 2012; David et al. 2012). These studies show that not only inherent capabilities of the mutated DNA, but also factors affecting protein-protein and protein-DNA interaction are important for the impact of a SNV.

As it is generally challenging to determine the impact of non-synonymous SNVs, computational methods to predict the effect of mutations, such as MutationTaster and PolyPhen2, have been developed (Schwarz et al. 2010; Adzhubei et al. 2010). These methods estimate the severity of a mutation and are based on inherent characteristics of the genome such as sequence, mRNA structure, conservation, splice sites and amino acid change. Furthermore, in recent years, with increasing access to genomic annotation data and large catalogues of SNVs, scores based on machine learning have been developed (e.g. CADD and FATHMM) (Shihab et al. 2014; Kircher et al. 2014). These methods use known causative and benign SNVs together with genomic annotations (e.g. gene lists, genetic motifs, conservation and transcription factor binding sites) to train the underlying algorithm to score the severity of mutations in the human genome. To assess the full effect of a non-synonymous mutation these methods might however not be sufficient and further evidence are usually needed to establish an SNV as causative.

Inheritance in genetic disease

Pathogenic mutations can either be inherited and segregate in a family, or they can be a new mutation in the affected carriers giving rise to sporadic cases of disease. As the human genome is diploid, all genes on the autosomal chromosomes exist in two copies and specific mutations can effect either one or both copies of the same gene. Therefore, pathogenic mutations can exist in two forms, either dominant where mutation in one of the two copies of a gene is sufficient to cause disease, or recessive where both gene copies need to be affected. Both dominant and recessive disorders can be inherited and it is common for them to segregate within families. The same is true for disorders involving genes on the sex chromosomes or in the mitochondrial DNA. However, as males carry one X and one Y chromosome and as the mitochondrial DNA is inherited exclusively from the mother, mutations in these parts of the genome follow their own specific inheritance patterns. Diseases that are caused by a single mutation are called Mendelian diseases and their pattern of inheritance can easily be spotted in a pedigree.

For Mendelian diseases with a high overall severity and early onset such as ID, reproductive fitness is decreased and therefore there might be low or no inheritance. In contrast, late onset diseases such as Alzheimer's have less effect on reproduction (Steele 2000). This indicates that severe conditions with early onset would be rare in the population compared to less severe or late onset conditions. Curiously, this is challenged by the observation that some severe conditions are relatively common in the population with e.g. ID occurring in approximately one percent of the population (Maulik et al. 2011). The high prevalence of these conditions can partly be explained by *de-novo* mutations. This is supported by several recent investigations that reveal a large

fraction of *de-novo* mutations in ID and similar disorders (Helbig et al. 2016; Erickson 2016). For de-novo mutations to explain the high prevalence of a severe disease, at least two conditions need to be met. Firstly, the number of causative genes needs to be relatively large and secondly the disease must have the potential to be triggered by a single mutation.

However, some diseases are not triggered by a single mutation, but instead they are triggered by a combination of different genetic and environmental factors. These types of disorders are called complex diseases and include many common diseases such as diabetes, arthritis and schizophrenia (Karalliedde & Gnudi 2016; Kavanagh et al. 2015). Many occurrences of complex disease are sporadic but a familial aggregation not following mendelian inheritance is commonly seen (Motulsky 2006).

Regulation of expression as a mechanism in disease

Genetic variation can have a direct effect on gene expression and an extreme example of this is the deletion of a genetic region. However, mutations can also influence expression in less direct ways and it has been shown that SNVs or indels can affect the expression of one or several genes (Stranger et al. 2005; Lappalainen et al. 2013). Loci in the genome containing genetic variants that affect gene expression are called expression quantitative traits loci (eQTL) (Schadt et al. 2003). Variants found inside of eQTLs can either act in *cis* or *trans*. Variants acting in *cis* affect nearby genes on the same physical chromosome while variants acting in *trans* can affect distant genes and genes on both alleles of a target gene (Becker et al. 2012). In a recent study using GWAS in 11 human complex diseases, hits mapping to regulatory regions were found at a higher rate than would be expected by chance (Gusev et al. 2014). These results indicate that mutations affecting regulatory regions are an important factor in the development of complex disease. This conclusion is further supported by several studies showing an overlap between eQTLs and GWAS hits in human complex disease (Zou et al. 2012; Yin et al. 2014).

Apart from quantitative changes in gene expression mutations may also induce changes in the expression of different mRNA splice isoforms produced from a gene. Mutations in splice sites have the ability to restrict or prevent the expression of specific mRNA isoforms. Alternative splicing is especially common in the nervous system (Castle et al. 2008), and changes in alternative splicing have been observed in different stages of brain development (Mazin et al. 2013). These observations suggest that alternative splicing is an important factor in normal neuronal function and disruptions of alternative splicing has been indicated as a component in neurologic disease (Lieve Claes et al. 2001; Rosewich et al. 2012; Paulussen et al. 2011; Licatalosi & Darnell 2006; Strong 2010). In addition to mutations in splice sites, mutations in genes encoding splicing factors, involved in directing the splicing of target genes,

have been shown to change splicing patterns (Korir et al. 2014; Zhang et al. 2015).

Another important process by which the cell can control the expression of specific genes is by controlling the level of condensation of the chromatin. A more condensed chromatin tends to inhibit gene expression while more loosely ordered chromatin allows expression factors to bind and increase expression of a gene (Kornberg & Lorch 1995). Proteins affecting the condensation of protein have been indicated to have important roles in regulating gene expression. Interestingly one class of these proteins, the chromatin-remodeling complexes has also been indicated to have an important task in mammalian neural system development (Bultman et al. 2000; Pereira et al. 2010). This is highlighted by the identification of causative mutations in chromatin remodeling factors in several studies investigating ID, autism, and related disorders (Gibson et al. 2012; Santen et al. 2012; Tsurusaki et al. 2012).

Tools used in the study of genetic disorders

Linkage and genome wide association studies

In linkage analysis specific genetic markers, usually SNVs or microsatellites, are used to identify genetic regions co-segregating with a certain trait or disease. The method relies on estimation of recombination frequencies and is based on the assumption that the pathogenic mutation will be inherited together with nearby markers. To calculate the probability of a certain marker co-segregating with a disease causing mutation the logarithm of the odds (LOD) score is used (Ott 1974). In large pedigrees linkage studies can severely decrease both the number and size of genetic regions of interest when searching for a causative gene.

Another widely used method is genome wide association studies (GWAS). In these studies SNVs are typically used as markers. However, unlike linkage analysis a GWAS does not require pedigrees, but instead rely on cohorts where individuals with a specific phenotype are collected and compared to controls. For each marker the allele frequency is measured and the odds ratios are calculated (Clarke et al. 2011). Markers having odds ratios significantly different from one are said to be associated with the trait studied.

Next generation sequencing

Before the advent of next generation sequencing (NGS) the capabilities of the available technology to investigate the genomic sequence of any organism was effectively restricted to targeted Sanger sequencing (Sanger et al. 1977). The Sanger method was later adapted for whole genome shotgun sequencing, a method to split longer DNA fragments into small fragments that are randomly sequenced and subsequently assembled (Staden 1979). However, the method was limited to the sequencing of small genomes (4000 – 7000 bp). In the beginning of this millennium several technologies were developed for parallel sequencing of DNA fragments, dramatically increasing the throughput compared to sanger sequencing (Brenner et al. 2000; Shendure et al. 2005). Combined with shotgun sequencing these methods gave the ability to sequence large and complex genomes (Venter et al. 2001). However, it was not until the introduction of the first commercially available NGS machine from Roche in 2005 that NGS could be used on a larger scale (Margulies et al. 2005).

Since then several innovative sequencing techniques have been developed (and in some cases disappeared from the market) and NGS techniques have been adapted to not only study our DNA but a wide array of different processes including transcription, and methylation (Buermans & den Dunnen 2014). Today several platforms are available for NGS and amongst these the Illumina, LifeTechnologies Semiconductor sequencing (Ion Proton) and PacBio are widely used.

The Illumina systems measure the fluorescence from labeled nucleotides to detect the order of bases in each DNA fragment sequenced. In a first step each DNA fragment is attached to a solid surface and a clonal amplification of the fragments are done. This results in clusters of cloned single stranded DNA fragments attached to the surface. Primers are then attached to the DNA and fluorescently tagged nucleotides are added one by one. Each nucleotide added creates a unique emission and the combined emission from each cluster is captured using digital imaging at the end of each cycle. Currently used Illumina systems produce reads spanning 50 to 300 bp (depending on the machine and protocols used in the library prep) (Mardis 2008)

In contrast to Illumina sequencing, the Ion Proton technology does not require an imaging step, and the sequencing runs are therefore considerably faster. As a first preparatory step the ion proton technologies creates an oil-water emulsion, and a clonal amplification of DNA fragments takes place inside of each oil vesicle. Following this, vesicles containing amplified DNA fragments are transferred to microwells on a semiconductor chip. The sequencing is performed by measuring changes to pH induced by the release of hydrogen ions when complementary bases are bound to the amplified single stranded DNA fragments. The reads generated by this technology do not have a fixed length, instead a distribution of read lengths are created with the longest spanning over 600 bp (Merriman & Rothberg 2012). The decreased runtime of this system however comes at the price of an excess of sequencing artifacts in the results (Buermans & den Dunnen 2014).

While both technologies described above require a clonal amplification step the single molecule real time sequencing introduced by Pacific Biosciences does not. The sensors in the PacBio machine are sensitive enough to detect the order of bases in a single DNA fragment. Similar to the Illumina system, the PacBio machine detects fluorescence to determine the order of bases in a DNA fragment. Unlike Illumina however, individual DNA fragments are placed at the bottom of small wells, and the fluorescent signals are recorded in real time. The main advantage of the PacBio system is the read length. The instrument is able to generate sequence reads that are 10 – 15 thousand bases long. However, the technology currently produces a limited number of total reads per run, making the cost per sequenced base notably higher than both the Illumina and Ion Proton technologies (Quail et al. 2012).

Whole exome sequencing

When studying genetic disease one of the key factors investigated is mutations in coding regions of genes. To this end whole exome sequencing (WES) was developed. This method introduces a capture step to enrich for exonic sequences prior to sequencing (Mamanova et al. 2010). This enables the sequencing of a large part of the coding sequences in the human genome while at the same time restricting the total length of DNA investigated.

However, exome sequencing suffers from some drawbacks. Firstly, not all exons are captured, and secondly, non-exonic causative mutations have a severely limited probability of being detected (Bamshad et al. 2011). Despite these disadvantages exome sequencing has been successfully used in finding causative *de-novo* mutations at a higher rate than previous non-NGS methods (Yang et al. 2014)

Whole genome sequencing

In whole genome sequencing (WGS) all chromosomal and mitochondrial DNA of an organism's genome are sequenced. In theory, raw data from each of the 6 billion bases in the human genome should be retrieved using WGS, however the sequencing depth (meaning the number of sequence reads covering each position in the genome) is a vital factor determining the ability of this technology to detect genetic aberrations (Meynert et al. 2014). For example it has been estimated that a genome needs to be sequenced to an average depth of 30 – 50 reads to retrieve a majority of SNVs and indels in a human genome (Ajay et al. 2011; Bentley et al. 2008). Furthermore, repetitive genomic sequences can create ambiguities in the mapping, which can complicate the interpretation of the data (Treangen & Salzberg 2012).

WGS is not only used to identify small variation but it can also be used for detection of structural variation. Several different analysis strategies have been developed to achieve variation calling across the full sizespectrum. One popular method is to use information gained from paired end reads (reads produced by sequencing both ends of a DNA fragment). As these reads are generally produced from DNA fragments of the same length they have a specific insert size and deviations from this can be used to pinpoint deletions and duplications (Korbel et al. 2007). Another approach to identify structural variations is by using split read analysis. This method assumes that reads not mapping to the genome may contain a CNV breakpoint. By remapping these reads to the genome and allowing for parts of the read to map to different positions it is possible to map these breakpoints (Zhang et al. 2011). Another method is to simply measure the read depth, assuming a correlation between read depth and CNVs (Yoon et al. 2009).

All of these methods have advantages and disadvantages. Measuring read depth provides good estimates of absolute copy number, but paired end and split reads methods perform better when detecting short CNVs (Alkan et al. 2011; Medvedev et al. 2009; Bellos et al. 2012). Because of this several methods using combined approaches, such as LUMPY, have been developed (Layer et al. 2014). Taken together, the possible applications of WGS make it an excellent technique for identifying pathogenic mutations. However, the downside of WGS is that compared to WES a significantly larger amount of reads needs to be sequenced to achieve the same coverage over exon regions. Furthermore, using WGS to catalogue the full spectrum of mutations is in many cases highly practical, but WGS is prohibitively expensive for large studies.

Sequence data analysis

After sequencing, whether using WGS or WES, the data produced needs to be analyzed and interpreted. In a first step, the reads must be mapped to a reference genome (unless a *de-novo* assembly approach is used) and after this, variants are identified. For experiments where the main aim is to identify variants and mutations several different softwares are available, e.g GATK ,SAMtools or FreeBayes (McKenna et al. 2010; Li et al. 2009; Garrison & Marth 2012).

A typical workflow when identifying causative SNVs and indels in WGS and WES data is to first filter the identified SNVs on quality, e.g. the number and mapping quality of reads covering the SNV, scores assigned to the probability of the sequencer to observe the correct base etc. Following this, known benign variants (common in the population and not disease associated) are removed. After an introductory filtering a large number of SNVs typically remain and further filtering is usually required. To achieve a list of candidate variants that is short enough for validation experiments to be feasible several strategies can be used.

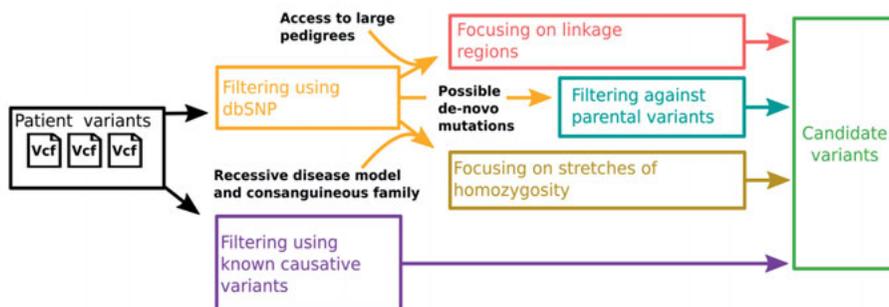


Figure 2: A sample workflow showing different filtering strategies that can be applied for exome sequencing data.

For recessive diseases in consanguineous families, homozygosity mapping can be used to decrease search space (Lander & Botstein 1987). Similarly if a large pedigree is available linkage analysis can be used to pinpoint regions containing variants of interest (Pulst 1999). When suspecting dominant disease caused by *de-novo* mutations, the full variation in the parents can be effectively used for filtering (Iossifov et al. 2014) (*Figure 2*). Careful planning of the study, analysis steps and selection of patients is crucial to maximize the chances of identifying causative mutations.

Transcriptome sequencing

Measuring gene activity by quantifying the abundance of expressed RNA is an important approach to investigate cells. To this end NGS sequencing technologies were early on adapted to sequence RNA. As sequencing machines were developed to sequence DNA, extracted RNA is commonly transformed to cDNA using reverse transcriptase before sequencing (Wang et al. 2009). Transcriptome sequencing is mainly used for two things, transcript identification and quantification of gene expression.

There are two general strategies for transcript identification, either the reads are mapped to a reference genome and transcripts are identified based on read alignments (Roberts et al. 2011). Alternatively, a *de-novo* transcript assembly may be performed; this method uses overlapping reads to identify expressed stretches of RNA (Robertson et al. 2010). The identification of transcripts in the human genome has important implications. For example, an accurate annotation of exons will increase the chances of correctly identify mutations affecting coding regions.

Quantification of gene expression is commonly used when investigating differences in global expression between two groups. The method is applicable when studying the effect of a specific mutation, but also when studying the effect of an administered substance. As a transcriptome sequencing experiment will produce slightly different numbers of reads between samples normalization is generally required before performing any comparative analysis. There are several techniques that are used to normalize transcriptome sequence data, and they can roughly be divided into two subgroups. One group of methods are based on a distribution adjustment of read counts, but can also take read length into account, e.g. reads per kilobase per million reads (RPKM) (Mortazavi et al. 2008). These methods are based on the assumption that there is some similarity between the read count distributions that are compared and may therefore under or over-estimate expression differences between transcripts if this assumption is not true. Another group of methods are based on the assumption that most regions of the genome in any given experiment are not differentially expressed. Popular methods belonging to this group are the normalizations introduced in DESeq and Trimmed Mean of M-

values (TMM) normalization that is used in the popular software edgeR (Hansen et al. 2010; Robinson & Oshlack 2010; Anders & Huber 2010; Trapnell et al. 2013). After normalization, gene expression can be tested on a global scale. Popular tools for this are the DESeq2 and edgeR packages.

Network analysis

For highly penetrant genetic variants creating a clear phenotype, models measuring only a few features are usually sufficient to identify the biological processes involved. However, in complex diseases, variants contributing to disease interact in highly complex biological pathways and systems. In order to understand the genetic contribution to these phenotypes several features need to be measured and understood. During recent years Network analysis has become a popular way of shedding light on interactions between genes, and has in many cases been successfully used to identify biological pathways and candidate genes in disease (Hormozdiari et al. 2015). A network analysis can be performed using several different types of data as input but the main theoretical framework remains the same. Briefly, each molecule (gene or protein) is modeled as a node in the network. After this the relationship between nodes is calculated. Relationships between nodes are called edges, often visualized as lines between gene names. There is no need for the edges to show a direct or physical interaction, and in many cases edges reflect statistical similarity (e.g. correlation) or inference but can also be a combination of any of these measurements. The information gained from calculating the edges can be used to organize the network and to create a structure from which it is possible to draw biological conclusions.

There are two main sources from which networks can be created, either literature curated data (e.g. protein interaction databases), or data derived directly from experiments such as expression levels. While the use of curated data from the literature gives some strength in facilitating the full use of previous knowledge it also bears the risk of containing bias. Highly studied genes will contribute more information to the network than less studied genes, making the network focus on what is already known. Furthermore, relationships identified may only be true for certain tissues and cell states and may lack connections present in the specific tissue or cell type studied. Despite this, these networks are still valuable for evaluating connections between specific genes and pathways and for producing lists of gene or protein candidates for further experiments.

Another type of network is gene co-expression network. These networks are typically created using microarray or RNA sequencing data as input. There are several algorithms used to create gene co-expression networks e.g.

SPACE, GeneNet, WGCNA and Aracne using either partial correlation, correlation, information theory or Bayesian Network as statistical method (Allen et al. 2012).

It is common to use data from a specific RNA sequencing experiment but there are examples where data from databases such as Allen Brain Atlas (<http://www.brain-map.org/>) are used (Ben-David & Shifman 2012). Co-expression networks have the strength of being specific to the tissue used and reflect another level of cellular organization than for example protein interaction networks. This is highlighted in several studies where gene co-expression networks created from patient data have successfully been used to identify groups of dysregulated genes defining coherent biological processes relevant to the disease (Voineagu et al. 2011; Gupta et al. 2014; Torkamani et al. 2010; Chen et al. 2013). Another approach has been to create co-expression networks from healthy controls, and identify co-expressed modules enriched for genes associated with a specific disease. This method has successfully identified modules of interest in several diseases, including neurodevelopmental disorders (Bettencourt et al. 2016; Li et al. 2015).

Intellectual disability, epilepsy and Schizophrenia

Schizophrenia

Schizophrenia is a common neuropsychiatric disorder, with a prevalence of 0.5 - 1% (Tandon et al. 2008). Disease symptoms typically occur in the mid-twenties, however the disease may manifest at any age (Bergen et al. 2014). Schizophrenia is highly heritable and the genetic component contributing to the disease has been estimated to be 64% - 81% (Lichtenstein et al. 2009; Sullivan et al. 2003). The incidence of schizophrenia is higher in males than in females (incidence of 10 females and 15 males per 100 000 people and year) (McGrath et al. 2008). Typical symptoms of schizophrenia include delusions, hallucinations, impaired motivation, social withdrawal and cognitive impairment and studies show that schizophrenia patients generally have a lower quality of life compared to the general population (Solanki et al. 2008; Owen et al. 2016). Schizophrenia is a heterogeneous disease and despite extensive study, no physiological changes common to all or even a majority of patients have been observed (Linden 2012). However, certain pathophysiological changes have been noted in a fraction of the patients, including thinning of grey matter and white matter abnormalities (Haijma et al. 2013). The indication of brain abnormality as a pathophysiological effect in schizophrenia is further underlined by several studies locating the disease effect to the prefrontal cortex (Barch & Ceaser 2012; Lewis 2012). Although there has been limited progress in understanding the detailed mechanisms of pathology, there has recently been important progress in understanding of the genetic contribution to schizophrenia.

Genetics of Schizophrenia

Genome wide association studies (GWAS) has led to the identification of more than 100 distinct genetic loci contributing to disease susceptibility in schizophrenia (Sullivan et al. 2012; Ripke et al. 2014). Furthermore, several inherited and *de-novo* CNVs have been identified as risk factors in schizophrenia, significantly increasing the risk of disease. Additionally, studies using exome sequencing in trio families have showed that *de-novo* SNVs are

contributing to schizophrenia prevalence (Fromer et al. 2014). It is also interesting to note that an increase in schizophrenia prevalence has been associated with paternal age (Malaspina et al. 2001). Taken together these facts suggest that schizophrenia is a highly complex disease with many distinct genetic factors contributing to disease development.

When investigating the evidence from genetic studies it becomes clear that certain classes of genes are frequently associated with schizophrenia. In exome sequencing studies *de-novo* mutations in genes coding for synaptic proteins have been extensively reported in patients (Hall et al. 2015). Furthermore, GWAS studies have implicated calcium-channels, glutamate receptors, dopamine receptors and genes in the MHC locus as associated with the disease (Ripke et al. 2014). These results correlate well with previous hypotheses about disease mechanisms in schizophrenia, including roles for autoimmunity and the dopamine system (Benros et al. 2014; Brisch et al. 2014).

Several environmental factors have also been reported to increase the susceptibility to schizophrenia, for example environmentally induced brain damage occurring during the fetal stage or early infancy (Owen 2012). The contribution of the environment to the risk of developing schizophrenia has to some extent also been confirmed in animal trials (Meyer & Feldon 2010). These observations support the idea that the development of schizophrenia is affected by gene-environment interactions and makes it a future challenge to separate nature from nurture.

Intellectual disability

ID affects approximately 1% of the population and is defined as a condition where the affected patients have impaired intelligence with an IQ under 70 and impaired social functioning (Maulik et al. 2011). The symptoms manifest early in life and have a remaining effect on development. ID is often associated with several secondary clinical findings such as malformations, metabolic diseases and epilepsy (Roselló et al. 2014; Arvio & Sillanpää 2003). Furthermore, ID often co-segregates with other psychiatric disorders, such as schizophrenia and autism (Duong et al. 2012; Berkel et al. 2010).

Genetics of intellectual disability

ID is highly heritable with an overall recurrence in families of 8.5%, making it clear that a large fraction of cases have genetic causes (Van Naarden Braun et al. 2005). The first identified genetic causes of ID were large chromosomal abnormalities detected by karyotyping and today it is estimated that approximately 4% of ID is caused by these types of mutations (Curry et al. 1997). With the development of microarrays several smaller CNVs were identified

as causative in ID. Taken together CNVs and large chromosomal aberrations are considered causative in approximately 14-18% of ID patients (Cooper et al. 2011; Hochstenbach et al. 2011).

ID has a high prevalence of sporadic and unexplained cases and it has long been hypothesized that a large fraction of these cases might be due to *de-novo* mutations. In recent years this have been confirmed by several large scale WES projects focused on the sequencing of family trios (Wright et al. 2014; Epi4K consortium 2012; de Ligt et al. 2012).

It is challenging to estimate the exact fraction of ID cases where *de-novo* mutations are causative, however in a recent large-scale study including 1133 families *de-novo* SNVs were considered causative in 19% of cases (Wright et al. 2014). In the same study inherited SNVs and indels were considered causative in 9% of cases. With access to WES, WGS and Array technologies novel ID genes are now detected at a rapid pace. ID occurs both as a recessive and dominant disease and today we know of more than 300 recessive and about one hundred dominant genes causing ID (Vissers et al. 2016). Furthermore it has long been known that males are overrepresented amongst ID patients. This is explained by mutations in genes on the X-chromosome where mutations have more severe effect in males compared to females, as males carry only one copy (Ropers & Hamel 2005). These conditions are referred to as X-linked ID and today we know of approximately one hundred X-linked ID genes, explaining less than 10% of the ID cases (Lubs et al. 2012).

Despite recent technological advances the number of ID patient's receiving a confirmed diagnosis is low and a large fraction of patients have no known molecular cause for their condition (Karam et al. 2015). The reasons for this are most probably a combination of several factors. The genetic screening may fail to detect causative mutations. For example, microarrays have a limited ability to detect short CNVs and exome sequencing does not report mutations outside of coding regions. Both of these problems could be reduced using WGS, however this technology is more costly and not commonly used in the clinic (Gilissen et al. 2014). Furthermore, our knowledge of genes involved in ID is limited, complicating the assessment of mutations found in patients (Vissers et al. 2016). It is also known that ID may develop as a consequence of environmental factors, e.g. by neurological damage during pregnancy or early development and in these cases a genetic diagnosis is impossible (Karam et al. 2015). Furthermore, several specific CNVs associated with ID have been shown to induce both heterogenous phenotypes and difference in disease severity (Girirajan et al. 2012; Mefford et al. 2008; Girirajan et al. 2010). This supports a model where background variation contributes to the disease in a complex manner.

The catalogue of known ID genes contains genes involved in several different biological processes and pathways, with examples including transcriptional and translational control, protein modification, chromatin remodeling, and centrosome function (Tsurusaki et al. 2012; Santen et al. 2012; Tanaka et

al. 2016; de Ligt et al. 2012; Najmabadi et al. 2011). This suggests that ID is not a single distinct disorder, but rather a large and diverse collection of disorders that have impaired cognitive abilities as a common trait. Despite this heterogeneity, biological processes commonly involved in ID are starting to emerge. These includes mutations in genes in the RAS-MAPK pathway with a distinct group of known ID disorders (Schubbert et al. 2007). Furthermore, mutations in the RHO GTPase pathway have showed to be associated with several forms of ID (Ba et al. 2013). An overrepresentation of mutations in several classes of proteins controlling expression of genes, such as chromatin remodeling factors, have also been frequently associated with ID (Sanchez-Mut et al. 2012).

Epilepsy

Epilepsy is defined by having recurrent seizures that are due to abnormal neuronal activity in the brain. The prevalence of epilepsy is difficult to estimate, however an incidence of 5.8 per 1000 people in developed countries and 10.3 to 15.4 per 1000 in developing countries has been reported (Bell et al. 2014). Furthermore, epilepsy is subdivided into several different classes based on seizure type, age of onset, co-morbid features and etiology (Berg et al. 2010). It is estimated that 20 – 30% of epilepsy cases are caused by acquired conditions, and the remaining 70 – 80% of cases are believed to be caused by genetic factors (Hildebrand et al. 2013). Epilepsy can occur both as a mendelian and a complex trait, however many cases occur sporadically with no family history of disease (Ottman & Risch 2012). Recent WES studies have revealed that *de-novo* mutations explain a fraction of sporadic cases, especially in cases with more severe epilepsy (Epi4K consortium 2012). Early genetic findings in epilepsy led to the hypothesis that epilepsy is primarily caused by mutations leading to dysfunction of ion channels (Singh et al. 1998; L Claes et al. 2001; Steinlein et al. 1995). Large scale studies using WES have added further support to this hypothesis but also identified causative genes implicating several other biological processes (Appenzeller et al. 2014; Epi4K consortium 2012; Helbig et al. 2016). Amongst these, genes involved in chromatin remodelling and transcriptional regulation are a large and growing group (Carvill et al. 2013; Allen et al. 2013).

Shared genetics of ID, epilepsy and schizophrenia

It has long been known that there is a strong comorbidity between neurodevelopmental disorders and it has been shown that there is a 3 -5 fold increase in ID amongst individuals diagnosed with schizophrenia compared to the general population (Hemmings 2006; Morgan et al. 2008). Furthermore, recent

studies show that a large fraction of ID cases also have autism or schizophrenia (Fombonne 2003). Epilepsy has been reported to be present in 26% of patient with ID and individuals with epilepsy have an ~8.5 fold increased risk of developing schizophrenia (McGrother et al. 2006; Arshad et al. 2010).

Interestingly, analysis of results from large-scale sequencing projects have identified several genes that are associated with both ID, schizophrenia, epilepsy and autism spectrum disorders (Li et al. 2015). Adding to this, over 50 CNVs have been associated with ID, schizophrenia and epilepsy or any combination of two of these conditions ID (McGillivray et al. 1990; Pescosolido et al. 2013). There is also emerging evidence using genetic networks that genes contributing to neurodevelopmental disorders are closely connected through protein interaction networks (Hormozdiari et al. 2015).

In some cases mutations in the same gene causing different disease may be explained by different functional impacts (e.g. disrupting specific splicing or protein interaction domains). Furthermore it is possible that the effect of a mutation is modified by a secondary inherited or *de-novo* mutation. It is also possible that environmental factors contribute to the development of differing disease phenotypes, especially in less severe conditions. Taken together the co-morbidity and shared genetics of ID, epilepsy and schizophrenia implicate an at least partially shared genetic etiology between these conditions.

Methods

The methods used in the Papers presented in this thesis are outlined below

SOLiD sequencing

SOLiD sequencing was used in **Paper I**, **Paper II** and **Paper III**. A short description of the SOLiD sequencing technology is outlined below. In a first step DNA or cDNA fragments are attached to beads and an emulsion PCR reaction is conducted. The beads are after this covalently attached to a glass slide and a sequencing-by-ligation reaction takes place. Briefly, a mixture of fluorescently labeled di-nucleotide probes and a universal primer are added to the reaction chamber. Numbered from the 3' end of the probe the first two bases are specific followed by degenerate positions. The probes are hybridized to the template and the fluorescent markers are cleaved off, generating a signal captured by digital imaging. This is repeated 5–7 cycles (depending on the readlength). After this the annealed template are melted away, and a new primer 1 bp shorter than the first are added. The cycle is repeated until every position of the template has been interrogated.

Whole exome sequencing

WES was used in both **Paper I** and **Paper III**. The Agilent Sure Select 50 Mb exome enrichment kit was used to enrich for the exonic regions prior to sequencing. In this enrichment method DNA or cDNA is hybridized to biotinylated RNA probes. In a following step the biotinylated RNA is captured using streptavidin coated magnetic beads and all unbound genetic material is washed away. The captured sequences are then amplified before sequencing.

In **Paper I** we investigated the possible applications of exome capture on RNA extracted from human tissues. Our primary focus was to investigate how well this method could capture lowly expressed transcripts compared to traditional RNA sequencing. We included brain tissue as one of the tissues analyzed (in addition to liver), as it is known that brain has a large number of expressed genes and considerable amount of alternative splicing (Wang et al. 2008; de la Grange et al. 2010). As it has been shown that there are considerable differences in expression between different developmental stages we extracted RNA from

both adult and fetal samples for both tissues (Kumar et al. 2000). Extracted RNA was transformed to double stranded cDNA and we successfully showed that this could be used as starting material for in solution capture. Sequencing was performed using a SOLiD4 instrument and sequenced reads were mapped to the hg19 version of the human reference genome using the program TopHat. The program Cufflinks was then used to identify and quantify transcripts (Roberts et al. 2011). Splice sites were compared to several genome annotations, including known genes, expressed sequence tags and RNA sequencing results from the exact same tissues.

In **Paper III** our main aim was to identify *de-novo* mutations in trios where the patient was diagnosed with ID and epilepsy. Here a more traditional approach to WES was adopted, and capture was performed on genomic DNA extracted from blood. As this study spanned several years, the machines available to us for sequencing were exchanged which forced us to use different sequencing technologies in different parts of the investigation. The three technologies used were SOLiD, Illumina and the Ion Proton systems. These technologies create slightly different outputs and therefore we used different programs to map the data generated by each machine (LifeScope for SOLiD, torrent-suite for Ion Proton and BWA for Illumina). To identify SNVs in sequenced samples the GATK toolkit was used. This program is a collection of tools designed to facilitate variant discovery. After mapping and removal of duplicated reads the GATK was in a first step used to locally realign reads in positions with a large number of mismatches compared to the reference genome. As reads mapping to positions containing indels tend to create several misalignments this step facilitates the detection of indels by creating a more exact alignment in these regions. After this a base recalibration were performed, this step uses empirical data and machine learning to recalculate the quality score assigned to each base. This is done as the original scores may be subjected to systematic as well as technical biases. After this variant calling can be performed, and this is done by in a first step determining which regions of the genome that contain evidence of significant variation. The possible haplotypes in each region are calculated and possible variant sites are identified. The likelihood for each haplotype is then calculated by realignment of all reads in the region in order to determine which model that is best explained by the data. In a last step the previously calculated likelihoods are used to calculate the posterior probabilities of each genotype and the most likely genotype is assigned to each site (*figure 3*). The resulting list

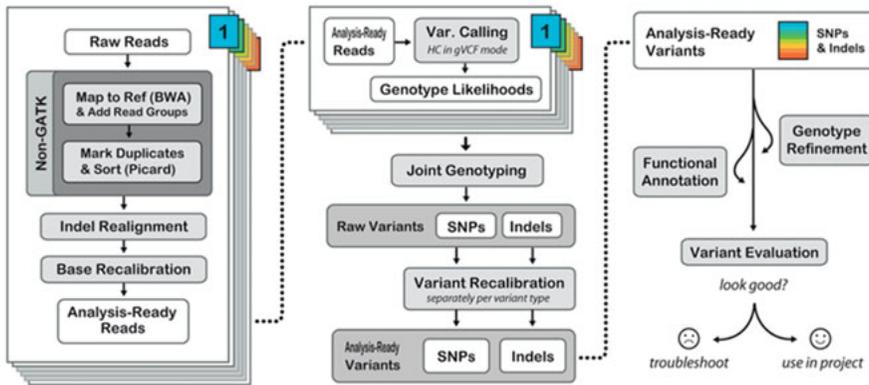


Figure 3: The recommended GATK workflow. Reads are mapped and duplicates are marked. Indels are realigned and bases recalibrated. After this variants are called and recalibrated. "Best Practices for Variant Calling with the GATK, source: Broad Institute, <http://www.broadinstitute.org/gatk/>"

of variants was first filtered using our in-house database of variants from previous exome studies, together with the dbSNP database. Remaining variants were then filtered against the variants identified in the parents. Candidate variants were then validated using sequencing.

Transcriptome sequencing

Transcriptome sequencing (also called RNA sequencing) was used in **Paper II** and it is also the main method in **Paper IV**. In **Paper II** our main interest was to investigate the genetic impact of different mRNA isoforms of the *QKI* gene. We decided to use human astrocytes as starting material as these cells are known to express *QKI*. The different splice isoforms of *QKI* were selectively silenced using siRNA and a SOLiD 5500XL instrument was used for transcriptome sequencing. Global differential expression analysis was performed using the R program DESeq resulting in a small list of genes that were differentially expressed.

In **Paper IV** RNA was extracted from frontal cortex and inferior frontal gyrus of schizophrenics and controls. The main reasoning behind using transcriptome sequencing in this Paper is that schizophrenia is a complex disease and even if the causative mechanism are unknown, changes in gene expression in patients may give valuable clues regarding disease etiology. RNA sequencing was performed using an Illumina HiSeq 2500 instrument, producing 100 bp paired-end reads and the software TopHat2 was used to align reads to the hg19 version of the human reference genome. The alignment was performed in a way so that reads were aligned to the known transcriptome (known expressed sequences) in a first step, and reads not aligned in this step were

aligned to full genome in a second step. Counting of reads was performed using the program HTSeq, and reads mapping to positions where different genes on the same strand overlapped was determined to be ambiguous and were excluded. The R program DESeq2 was used to identify genes showing a significant differential expression between cases and controls. To further analyse the sequence data a gene co-expression analysis was performed using the R library WGCNA. This resulted in several modules, of which three could be associated with the patient group.

Concluding remarks and future perspectives

Since the discovery of the structure of DNA in 1953 the field of human genetics has developed at a rapid pace. New technologies have readily been adapted to be used as tools in the investigations of the human genome, leading to increased capabilities and a successive broadening in the field of research. The human genome is complex and even though the instruments used to observe it reach unparalleled levels of accuracy and specificity, it can be considered to be only partially explored. Nonetheless, by adopting new technologies, tools and applications and studying specific biological phenomena, we fill these gaps in our knowledge piece by piece.

In **Paper I** we combined exome capture with RNA sequencing in an effort to identify lowly expressed transcripts and splice sites. We show that this method can detect transcripts and splice junctions below the detection threshold of traditional RNA sequencing at a comparable number of sequenced reads. From this, we conclude that this method is a cost-effective alternative to RNA sequencing for genome wide discovery of lowly expressed transcripts. A similar method was used by Mercer et al. to discover a large number of transcripts at low levels, further strengthening this conclusion (Mercer et al. 2012). The functional impact of low level transcripts can be questioned and the biological effects of different expression levels are not well investigated for most genes. Despite this several arguments strengthening the relevance of investigating low level transcripts exists. A human tissue sample generally contains several cell types in different fractions meaning that a low level transcript might just reflect a transcript expressed in a low fraction of the cell types sequenced. If so, the transcript might have a relevant functional impact in this specific cell type. Hence, further experiments are needed to pinpoint the source of low level transcripts found in human tissues. Furthermore, identifying lowly expressed transcripts in one tissue is in no way proof that it is lowly expressed in all human tissues. Therefore the detection of unannotated exons and transcripts on a global scale still facilitates the detection of genetic regions of interest e.g. when investigating a disease associated with a certain gene.

In **Paper II** we used RNA sequencing in combination with selective silencing to investigate the different splice isoforms of the QKI gene. The QKI gene

has previously been associated with schizophrenia and is involved in central nervous system development (Aberg et al. 2006). Interestingly, RNA sequencing data revealed a 5th previously unknown QKI transcript variant showing the benefits of carefully investigating results from this technology. The analysis of differential expression gave sparse results. This is not surprising, as it is generally not expected that the silencing of a single gene will result in expression changes of a magnitude that can be detected in a small number of replicate experiments. Amongst the most interesting results was the significant change in expression of *GFAP* making it a potential mRNA target of QKI. The *GFAP* gene has previously been implicated to have a major function in synaptic plasticity (Middeldorp & Hol 2011). Future studies regarding the QKI and GFAP interactions are needed to better understand their effects on neural development and disease.

In **Paper III** we identify *de-novo* mutations in patient parent trios, where the patients are diagnosed with both epilepsy and ID. We identify 29 *de-novo* mutations in 39 trio families and one inherited homozygous mutation in one family. Using the American College of Medical Genetics and Genomics standard guidelines for interpretation of coding variants we identified causative mutations in 11 families giving us a diagnostic yield of 28.2%. The most exciting finding in this study was the identification of a non-synonymous *de-novo* mutation in the *HECW2* gene. This gene was previously not associated with disease, however when studying results from previous large scale sequencing projects we found another five *de-novo* mutations in this gene in patients with phenotypes similar to our patient. Furthermore, five of the six mutations (including ours) was localized to the HECT domain of the HECW2 protein, implying a gain of function mechanism. The candidacy of the *HECW2* gene was further strengthened using network analysis. It remains a future challenge to investigate the exact biological impact the mutations in the *HECW2* gene. Using animal models such as zebrafish by introducing mutations in the homologous gene and affecting the same domain would be one exciting way to further investigate the effect of these mutations.

In **Paper IV** we investigated schizophrenia using RNA sequencing. Primary brain tissue was acquired from 65 patients and 47 controls. Most interestingly the results from the differential expression analysis showed a strong enrichment of genes involved in the complement system. These results are in line with a recent study pinpointing the *C4* gene as involved in schizophrenia (Sekar et al. 2016). To further investigate differences in expression between patients and controls a gene co-expression analysis was performed. This resulted in three modules associated with schizophrenia, with one module containing several genes associated with axon development. These results show a correlation between schizophrenia, the complement system and axon development. However, to add further support to these conclusions and to understand the mechanisms and function underlying these correlations, additional experiments are needed. Using WGS in order to identify relevant mutations

would be our next step to in more detail investigate possible mechanisms driving the observed differences in expression. This would also be important as it would clarify the level at which environmental effects contribute to the disease process. Furthermore, analysis and comparison of chromatin states between patients and controls would give further insights into the impact of possible epigenetic effects in schizophrenia.

In conclusion, this thesis contributes to the knowledge of specific genes and pathways involved in schizophrenia, ID with epilepsy as well as genes involved in the development of the central nervous system. Within this thesis several candidate genes for ID and epilepsy are presented, adding information to the knowledge of these disorders. It further expands the knowledge of methods used for transcript identification, adding to the current catalogue of available and tested methods. Furthermore, genes and mechanisms possibly contributing to the development of schizophrenia are revealed, adding proof to previous hypotheses.

Summary of included Papers

Paper I. Exome RNA sequencing reveals rare and novel alternative transcripts

Aim:

To investigate the use of exome capture on mRNA

Methods:

In solution capture of cDNA
RNA sequencing

RNA sequencing is a routinely used technique in the analysis of gene expression and splicing. When preparing RNA for sequencing two strategies are mainly used; either all polyadenylated transcripts are selected or amplified (poly(A) sequencing), or RNA is amplified using random hexamers, enriching for all RNAs in the sample before sequencing (total RNA sequencing). Using either of these methods will result in the sequencing of regions where no transcripts are annotated, representing a mix of real unannotated transcripts and background noise. Here we aimed to investigate expression of annotated transcripts using RNA sequencing in combination with in solution exome target capture. We hypothesized that targeted enrichment of exonic RNA before sequencing would increase the fraction of reads covering genes and exons thus facilitating the discovery of rare splicing events transcripts expressed at low levels.

To sequence captured RNA we first produced cDNA from total RNA. After this the Agilent SureSelect 50 Mb capture kit was used for capture of cDNA sequences. The captured cDNA was then sequenced using standard protocols on a SOLiD4 instrument. As gene expression and splicing differ extensively between tissues, or even in the same tissue in different stages of development, we tested this approach on four human tissue types (adult cortex, fetal cortex, adult liver and fetal liver).

After sequencing we first compared the read distributions found in our data to previously sequenced total RNA sequencing data from the same tissues. Not

surprisingly, we observed a higher number of reads mapped to exonic regions in our data confirming that the capture strategy indeed produced a higher amount of exonic reads. We then continued with measuring relative expression over exons and compared to total RNA sequencing. This resulted in only a weak correlation suggesting that the methods differ in their capability to estimate expression. To further investigate the ability to quantify expression we used qrtPCR as an independent method to get expression estimates. The results showed that total RNA sequencing more closely matched the qrtPCR results leading us to conclude that our method had a lower accuracy when estimating relative expression. This difference may reflect a difference in capture affinity of enrichment probes for specific exons. Another explanation may be that different exons are captured by different numbers of overlapping probes, which may have the consequence that certain exons are captured more frequently than others.

One common application of RNA sequencing is the detection of transcript isoforms. In order to evaluate how our method performed in this respect we counted the number of splice junctions detected in each tissue and compared it to the number of splice junctions in the total RNA sequencing. For all tissues the results showed that our method generated a significantly greater number of splice junctions as well as a higher fraction of spliced reads compared to total RNA. To further investigate this we also compared to a sample prepared using poly(A) sequencing, and this sample showed a slightly increase in splice junction compared to total RNA sequencing, however the number of splice junctions identified using our capture method was still significantly higher.

To further evaluate our method we investigated how well it identified annotated splice junctions. This was done by comparing the splice junctions identified in each sample to the University of California Santa Cruz (UCSC) annotated genes database. The results showed that we could identify a large overlap between splice junctions found by our method and the gene database. Interestingly, we also found that our data contained splice junctions describing splicing events not present in the gene database. These novel splice junctions did in some cases correspond to a novel splice junction between known exons, or in other cases suggest the presence of a novel exon. In both of these cases our data suggested the presence of a new and in many cases lowly expressed transcript. When comparing these novel splice junctions to total RNA sequencing we could observe an eightfold increase in novel splice junctions in our data. To investigate the false positive ratio of our method we compared our identified novel splice junctions to the expressed sequence tags (EST) UCSC database. The fractions of splice junctions mapping to the EST database was similar in both total RNA sequencing and our method suggesting that the fraction of false positive findings was also similar between these two approaches.

In conclusion this study shows that using an exome capture strategy in combination with RNA sequencing enriches for the detection of lowly expressed transcripts and splice junctions. However, this comes at the cost of a significant loss of accuracy in quantification of expression. The study shows that whole exome capture is a viable technology for discovery of novel exons and alternatively spliced transcript isoforms.

Paper II: RNA-binding protein QKI regulates *Glial fibrillary acidic protein* expression in human astrocytes.

Aim:

To investigate the function of the *QKI* gene and its transcripts in primary astrocytes

Methods:

RNA sequencing

In this Paper we aimed to investigate the impact of selectively depleting the transcripts of the Quaking (*QKI*) gene, a gene previously associated with schizophrenia (Aberg et al. 2006). We did the analysis in primary astrocytes as these cells are known to readily express QKI and astrocyte abnormalities have been observed in QKI deficiency.

There are four known QKI transcripts (QKI5, QKI6, QKI7 and QKI7b). As a first step in this study we measured the cellular localization of the isoforms using immunostaining (except for QKI7b for which there was no specific antibody at that time). The results showed that QKI7 and QKI6 proteins are localized both in the nucleus and the cytoplasm, while QKI5 is present only in the nucleus. The difference in subcellular localization suggests that these isoforms have distinct and different functions.

To further investigate the function of QKI a loss of function approach was adopted, selectively silencing each of the QKI splice forms using siRNA. As this approach included targeting small unique regions of the RNA differing between splice isoforms, this could not be done for QKI6 where no unique regions could be identified. In order to assess the contribution of QKI6 to total QKI expression a total silencing of the QKI gene was also performed. Furthermore, silencing of QKI5 not only decreased QKI5 protein and mRNA but significantly decreased expression of all three isoforms, making the effect similar to the effect of the total silencing of the QKI gene. From the resulting data we further estimated that QKI6 is the most highly expressed splice isoform, with higher expression compared to QKI7. Interestingly the distributions of reads also suggested the existence of a previously unknown transcript, which was named QKI6b.

Next we performed differential expression analysis investigating the effect of silencing of QKI-tot and QKI7 on global gene expression in the astrocytes. We found that silencing of QKI seemed to have only a minor effect on gene expression with only a few genes showing up as significantly differentially

expressed. When investigating the differentially expressed genes in the cells with the QKI7 transcript silenced we identified the gene *Glial fibrillary acidic protein (GFAP)* as significantly differentially expressed. We further observed that differential expression of *GFAP* was also present in the total silencing of QKI compared to controls, although it failed to reach significance after adjustment for multiple testing. The *GFAP* gene has been associated with several disorders of the CNS and, perhaps most intriguingly, low expression of *GFAP* has been associated with schizophrenia. Taken together our results suggest that the QKI7 isoform of the *QKI* gene functions as a regulator of the *GFAP* gene, connecting *QKI* to *GFAP*.

Paper III: Identification of new candidate genes in intellectual disability and epilepsy

Aim:

To identify causative genes in conditions in patients with ID in combination with Epilepsy

Methods:

Exome sequencing

During recent years with the development of exome sequencing strategies *de-novo* mutations have, in addition to large scale copy number variation, been implicated as a major contributor in disease development in both intellectual disability and epilepsy. In this study we applied exome sequencing in the investigation of 39 trio families where the affected child is co-diagnosed with both ID and Epilepsy. Sequencing both parents in addition to the proband was advantageous in this study as it supplied us with the means to identify *de-novo* mutations at a high throughput. Identification of *de-novo* mutations in the patients was the primary focus of the study.

Following exome sequencing 29 *de-novo* mutations in 22 of the families were identified and validated using Sanger sequencing. Our first goal after identifying the *de-novo* mutations was to investigate if these mutations were in anyway more damaging than mutations found in the general population. To this end the Combined Annotation Dependent Depletion (CADD) score was calculated for each mutation and the resulting distribution was compared to the distribution of CADD scores from randomly chosen SNPs from the Exome Aggregation Consortium (ExAC) database. The results showed that on average the identified *de-novo* mutations were more damaging than variants found in the ExAC database. Here it should be mentioned that some *de-novo* mutations also showed a low impact on the genome. This is perhaps not a surprising result as we would expect that *de-novo* mutations would not explain the disease in all patients.

Each *de-novo* mutation found was carefully investigated and the American College of Medical Genetics and Genomics guidelines (ACMG) were used as an aid to determine whether variants were causative. This resulted in causative *de-novo* mutations identified in 10 families. In addition, a causative homozygous inherited recessive mutation was identified in one patient. Together, these results gave a total clinical yield of 28.2%. Interestingly, we also identified several *de-novo* mutations in genes previously not associated with ID or Epilepsy. As these genes were not previously known to cause disease we were

faced with the challenge to prioritize candidate genes for which we have identified a *de-novo* mutations in only a single patient.

With the aim to further investigate the candidate genes and evaluate the potential of them being involved in the disease phenotype, we compiled a list of *de-novo* mutations identified in previous exome sequencing studies. The mutations were collected from several sequencing projects and the list was restricted to include results from studies of ID, epilepsy and autism spectrum disorder, as these conditions have been shown to be genetically overlapping. Control data was also collated from published studies. As expected, many of the genes with mutations determined to be pathogenic in our study were frequently carrying *de-novo* mutations also in previous studies. However, the most interesting result from this comparison showed that the gene *HECW2*, previously not associated with any disease, was found to carry *de-novo* mutations in 6 patients and 0 controls. To further study these *de-novo* mutations we investigated their positioning over the gene, concluding that all but one clustered inside or close to the HECT domain. This result lead us to investigate the overall mutation rate over the exons of the *HECW2* gene using data from the ExAC database. In line with our previous observation we can report a decrease of non-synonymous mutations over parts of the HECT domain in the *HECW2* gene, suggesting that this domain is especially sensitive to protein altering mutations.

We then created a genetic interaction network using all genes where we identified *de-novo* mutations, together with genes where previously causative mutations had been identified in exome sequencing of trio families. The interaction data used to build the network was supplied by public databases. The results show a tightly interconnected network where the *HECW2* gene is connected to several of the genes known to cause ID. This result should be interpreted with caution as the catalogue of genetic interactions in the human genome is likely to be incomplete. In spite of this the results add evidence to the connection between *HECW2* and ID.

The results of our study confirm the importance of *de-novo* mutations in ID and epilepsy. We further explore several approaches in order to prioritize candidate genes. Overall, we build upon the knowledge of known causative genes as well as suggest *HECW2* as a novel candidate gene in ID and epilepsy.

Paper IV: Transcriptome sequencing implicates alterations of complement factors levels in schizophrenia

Aim:

To identify transcriptomic changes between patients diagnosed with schizophrenia and controls

Methods:

RNA sequencing

Differential expression analysis

Weighted gene co-expression network analysis

Schizophrenia is a heterogeneous heritable disease with a population prevalence of 0.05-1% and an early onset, typically in the range of 20-30 years. In this paper we perform RNA sequencing of post mortem brain tissue from 65 patients diagnosed with schizophrenia and 43 controls. Our main hypothesis was that there would be observable differences in the transcriptome between patients and controls and that these differences would provide valuable clues regarding the mechanisms and genes involved in disease progression.

We sequenced patients and controls to a total depth of approximately 60 million reads per samples. Using differential expression analysis we report the presence of 74 genes showing differential expression between patients and controls. When examining the results we conclude that a surprisingly large number of the genes over-expressed in the patients were involved in immunity and inflammatory processes. To confirm this a GO analysis including all significantly differentially expressed genes was performed. The results of this analysis confirmed our previous observation showing several categories involved in inflammation and immunity as significantly enriched, and that the category complement activation showed the lowest p-value in the analysis. Specifically calculating the mean expression for the genes involved in the Complement activation category showed a strongly increased expression in a subset of patients and only a few controls. Taken together these results add to previous observation of immune dysregulation in schizophrenia as well as the results of a recent GWAS study associating the minor histocompatibility complex (MHC) to schizophrenia (Ripke et al. 2014).

In order to investigate the enrichment of complement genes detected in our results we calculated the mean expression of all differentially expressed genes in the complement activation GO category for all patients and controls. The resulting distribution revealed a strong mean expression of complement genes

in a large fraction of the patients and only a few of the controls, suggesting that activation of the complement system is linked to disease progression in schizophrenia.

It has previously been shown that patients with schizophrenia harbor a significantly higher amount of *de-novo* mutations than the general population. This would imply that a fraction of the *de-novo* mutations found in schizophrenia contribute to the genetic basis of disease. To this end, we collected a list of genes with identified *de-novo* mutations from several large studies in schizophrenia and compared these to our results. Two of our differentially expressed genes, *transglutaminase 2 (TGM2)* and *laminin 2 (LAMA2)*, have independently and significantly been associated with schizophrenia based on *de-novo* mutations. It is also relevant to point out that both of these genes are highly expressed in the CNS. We further investigated the overlap between our results and genes found in recent GWAS studies however no overlap was found.

In order to further investigate any differences between patients and controls a weighted differential co-expression network was created. The network resulted in three nodes significantly associated with schizophrenia. Following this GO analysis of the genes in each node was performed leading to the identification of the significantly enriched GO category Axon in one of the nodes. Interestingly, several studies have observed a thinning of grey matter in schizophrenic patients and it has been hypothesized that this thinning is due to a lower number of axons. When reviewing the genes in these nodes another interesting result emerged as one of the driving genes in this node was determined to be the *NEFH* gene. The NEFH protein is commonly used as a marker of neuronal damage in the autoimmune disease ALS (Al-Chalabi et al. 1999).

Acknowledgements

I would first and foremost like to thank my supervisor **Lars Feuk** for helping and inspiring me and for his great patience and leadership. Thank you for encouraging and challenging me, and helping me to evolve. Your great knowledge are a true inspiration. Furthermore I would like to thank my co-supervisor **Ulf Gyllensten**, for his support and help.

I would further like to thank my collaborator **Lotta Thuresson** for her valuable help and knowledge. Lotta, you have been a great help in my research and your input have been invaluable for me.

Furthermore, to my wonderful group **Eva Carlström**, **Ammar Zhaghlool**, **Jin Zhao**, **Adnan Niazi**, and **Mitra Etemadikhah** I could not ask for better colleagues. I would also like to thank **Anna Johansson** former member of our group, your knowledge and support really helped me.

And **Ignas Bunikis** even though we never did science together you are a great friend and colleague. Your support and help during the years have helped me, as well as our scientific, and sometimes not so scientific discussions. You are a true friend.

Furthermore I would like to thank **Lars Forsberg** for always answering my math question, you have been a great help throughout the years.

And of course I would like to thank my fiancé **Ida Karlsson** for supporting me during the time I wrote my thesis. I asked the right woman to marry me!

I would also like to thank my parents for being supportive and always allowed me to follow my dreams. Without you I would not be where I am today.

References

- Aberg, K., Saetre, P., Lindholm, E., Ekholm, B., Pettersson, U., Adolfsson, R. & Jazin, E., 2006. Human QKI, a new candidate gene for schizophrenia involved in myelination. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, 141B(1), pp.84–90. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16342280> [Accessed April 24, 2016].
- Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S. & Sunyaev, S.R., 2010. A method and server for predicting damaging missense mutations. *Nature methods*, 7(4), pp.248–9. Available at: <http://dx.doi.org/10.1038/nmeth0410-248> [Accessed December 13, 2013].
- Ajay, S.S., Parker, S.C.J., Abaan, H.O., Fajardo, K.V.F. & Margulies, E.H., 2011. Accurate and comprehensive sequencing of personal genomes. *Genome research*, 21(9), pp.1498–505. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3166834&tool=pmc&rendertype=abstract> [Accessed April 23, 2016].
- Al-Chalabi, A., Andersen, P.M., Nilsson, P., Chioza, B., Andersson, J.L., Russ, C., Shaw, C.E., Powell, J.F. & Leigh, P.N., 1999. Deletions of the heavy neurofilament subunit tail in amyotrophic lateral sclerosis. *Human molecular genetics*, 8(2), pp.157–64. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9931323> [Accessed March 21, 2016].
- Alkan, C., Coe, B.P. & Eichler, E.E., 2011. Genome structural variation discovery and genotyping. *Nature reviews. Genetics*, 12(5), pp.363–76. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4108431&tool=pmc&rendertype=abstract> [Accessed July 10, 2014].
- Allen, A.S., Berkovic, S.F., Cossette, P., Delanty, N., Dlugos, D., Eichler, E.E., Epstein, M.P., Glauser, T., Goldstein, D.B., Han, Y., Heinzen, E.L., Hitomi, Y., Howell, K.B., Johnson, M.R., Kuzniecky, R., Lowenstein, D.H., Lu, Y.-F., Madou, M.R.Z., Marson, A.G., et al., 2013. De novo mutations in epileptic encephalopathies. *Nature*, 501(7466), pp.217–21. Available at: <http://dx.doi.org/10.1038/nature12439> [Accessed May 13, 2015].
- Allen, J.D., Xie, Y., Chen, M., Girard, L. & Xiao, G., 2012. Comparing statistical methods for constructing large scale gene networks. *PloS one*, 7(1), p.e29348. Available at: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0029348> [Accessed March 18, 2016].
- Anders, S. & Huber, W., 2010. Differential expression analysis for sequence count data. *Genome biology*, 11(10), p.R106. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3218662&tool=pmc&rendertype=abstract> [Accessed December 11, 2013].

- Appenzeller, S., Balling, R., Barisic, N., Baulac, S., Caglayan, H., Craiu, D., De Jonghe, P., Depienne, C., Dimova, P., Djémié, T., Gormley, P., Guerrini, R., Helbig, I., Hjalgrim, H., Hoffman-Zacharska, D., Jähn, J., Klein, K.M., Koeleman, B., Komarek, V., et al., 2014. De novo mutations in synaptic transmission genes including DNMI cause epileptic encephalopathies. *American journal of human genetics*, 95(4), pp.360–70. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4185114&tool=pmc-entrez&rendertype=abstract> [Accessed August 16, 2015].
- Arshad, S., Winterhalder, R., Underwood, L., Kelesidi, K., Chaplin, E., Kravariti, E., Anagnostopoulos, D., Bouras, N., McCarthy, J. & Tsakanikos, E., 2010. Epilepsy and intellectual disability: does epilepsy increase the likelihood of co-morbid psychopathology? *Research in developmental disabilities*, 32(1), pp.353–7. Available at: http://www.researchgate.net/publication/47811504_Epilepsy_and_intellectual_disability_does_epilepsy_increase_the_likelihood_of_co-morbid_psychopathology [Accessed June 23, 2015].
- Arvio, M. & Sillanpää, M., 2003. Prevalence, aetiology and comorbidity of severe and profound intellectual disability in Finland. *Journal of Intellectual Disability Research*, 47(2), pp.108–112. Available at: <http://doi.wiley.com/10.1046/j.1365-2788.2003.00447.x> [Accessed April 24, 2016].
- Auton, A., Abecasis, G.R., Altshuler, D.M., Durbin, R.M., Bentley, D.R., Chakravarti, A., Clark, A.G., Donnelly, P., Eichler, E.E., Flicek, P., Gabriel, S.B., Gibbs, R.A., Green, E.D., Hurles, M.E., Knoppers, B.M., Korb, J.O., Lander, E.S., Lee, C., Lehrach, H., et al., 2015. A global reference for human genetic variation. *Nature*, 526(7571), pp.68–74. Available at: <http://dx.doi.org/10.1038/nature15393> [Accessed September 30, 2015].
- Ba, W., van der Raadt, J. & Nadif Kasri, N., 2013. Rho GTPase signaling at the synapse: implications for intellectual disability. *Experimental cell research*, 319(15), pp.2368–74. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23769912> [Accessed April 24, 2016].
- Bamshad, M.J., Ng, S.B., Bigham, A.W., Tabor, H.K., Emond, M.J., Nickerson, D.A. & Shendure, J., 2011. Exome sequencing as a tool for Mendelian disease gene discovery. *Nature reviews. Genetics*, 12(11), pp.745–55. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21946919> [Accessed July 10, 2014].
- Barch, D.M. & Ceaser, A., 2012. Cognition in schizophrenia: core psychological and neural mechanisms. *Trends in cognitive sciences*, 16(1), pp.27–34. Available at: <http://www.sciencedirect.com/science/article/pii/S1364661311002488> [Accessed April 4, 2016].
- Becker, J., Wendland, J.R., Haenisch, B., Nöthen, M.M. & Schumacher, J., 2012. A systematic eQTL study of cis-trans epistasis in 210 HapMap individuals. *European journal of human genetics: EJHG*, 20(1), pp.97–101. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3234520&tool=pmc-entrez&rendertype=abstract> [Accessed April 23, 2016].
- Bell, G.S., Neligan, A. & Sander, J.W., 2014. An unknown quantity--the worldwide prevalence of epilepsy. *Epilepsia*, 55(7), pp.958–62. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24964732> [Accessed April 24, 2016].
- Bellos, E., Johnson, M.R. & Coin, L.J.M., 2012. cnvHiTSeq: integrative models for high-resolution copy number variation detection and genotyping using population sequencing data. *Genome biology*, 13(12), p.R120. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4056371&tool=pmc-entrez&rendertype=abstract> [Accessed April 23, 2016].

- Ben-David, E. & Shifman, S., 2012. Networks of Neuronal Genes Affected by Common and Rare Variants in Autism Spectrum Disorders G. Gibson, ed. *PLoS Genetics*, 8(3), p.e1002556. Available at: <http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1002556> [Accessed January 11, 2016].
- Benros, M.E., Pedersen, M.G., Rasmussen, H., Eaton, W.W., Nordentoft, M. & Mortensen, P.B., 2014. A nationwide study on the risk of autoimmune diseases in individuals with a personal or a family history of schizophrenia and related psychosis. *The American journal of psychiatry*, 171(2), pp.218–26. Available at: <http://ajp.psychiatryonline.org/doi/abs/10.1176/appi.ajp.2013.13010086> [Accessed April 23, 2016].
- Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R., Boutell, J.M., Bryant, J., Carter, R.J., Keira Cheetham, R., Cox, A.J., Ellis, D.J., Flatbush, M.R., Gormley, N.A., Humphray, S.J., et al., 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456(7218), pp.53–9. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2581791&tool=pmc&rendertype=abstract> [Accessed July 9, 2014].
- Berg, A.T., Berkovic, S.F., Brodie, M.J., Buchhalter, J., Cross, J.H., van Emde Boas, W., Engel, J., French, J., Glauser, T.A., Mathern, G.W., Moshé, S.L., Nordli, D., Plouin, P. & Scheffer, I.E., 2010. Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005-2009. *Epilepsia*, 51(4), pp.676–85. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20196795> [Accessed July 9, 2014].
- Bergen, S.E., O’Dushlaine, C.T., Lee, P.H., Fanous, A.H., Ruderfer, D.M., Ripke, S., Sullivan, P.F., Smoller, J.W., Purcell, S.M. & Corvin, A., 2014. Genetic modifiers and subtypes in schizophrenia: investigations of age at onset, severity, sex and family history. *Schizophrenia research*, 154(1-3), pp.48–53. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4422643&tool=pmc&rendertype=abstract> [Accessed March 23, 2016].
- 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. & Rappold, G.A., 2010. Mutations in the SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. *Nature genetics*, 42(6), pp.489–91. Available at: <http://dx.doi.org/10.1038/ng.589> [Accessed December 16, 2013].
- Besenbacher, S., Liu, S., Izarzugaza, J.M.G., Grove, J., Belling, K., Bork-Jensen, J., Huang, S., Als, T.D., Li, S., Yadav, R., Rubio-García, A., Lescai, F., Demontis, D., Rao, J., Ye, W., Mailund, T., Friborg, R.M., Pedersen, C.N.S., Xu, R., et al., 2015. Novel variation and de novo mutation rates in population-wide de novo assembled Danish trios. *Nature Communications*, 6, p.5969. Available at: <http://www.nature.com/ncomms/2015/150119/ncomms6969/full/ncomms6969.html> [Accessed January 19, 2015].
- Bettencourt, C., Forabosco, P., Wiethoff, S., Heidari, M., Johnstone, D.M., Botía, J.A., Collingwood, J.F., Hardy, J., Milward, E.A., Ryten, M. & Houlden, H., 2016. Gene co-expression networks shed light into diseases of brain iron accumulation. *Neurobiology of disease*, 87, pp.59–68. Available at: <http://www.sciencedirect.com/science/article/pii/S0969996115301066> [Accessed February 2, 2016].

- Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., Roth, R., George, D., Eletr, S., Albrecht, G., Vermaas, E., Williams, S.R., Moon, K., Burcham, T., Pallas, M., et al., 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature biotechnology*, 18(6), pp.630–4. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10835600> [Accessed April 19, 2016].
- Brisch, R., Saniotis, A., Wolf, R., Bielau, H., Bernstein, H.-G., Steiner, J., Bogerts, B., Braun, K., Braun, A.K., Jankowski, Z., Kumaratilake, J., Kumaritlake, J., Henneberg, M. & Gos, T., 2014. The role of dopamine in schizophrenia from a neurobiological and evolutionary perspective: old fashioned, but still in vogue. *Frontiers in psychiatry*, 5, p.47. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4032934&tool=pmc-entrez&rendertype=abstract> [Accessed April 23, 2016].
- Buermans, H.P.J. & den Dunnen, J.T., 2014. Next generation sequencing technology: Advances and applications. *Biochimica et biophysica acta*, 1842(10), pp.1932–1941. Available at: <http://www.sciencedirect.com/science/article/pii/S092544391400180X> [Accessed July 10, 2014].
- Bultman, S., Gebuhr, T., Yee, D., La Mantia, C., Nicholson, J., Gilliam, A., Randazzo, F., Metzger, D., Chambon, P., Crabtree, G. & Magnuson, T., 2000. A Brg1 Null Mutation in the Mouse Reveals Functional Differences among Mammalian SWI/SNF Complexes. *Molecular Cell*, 6(6), pp.1287–1295. Available at: <http://www.sciencedirect.com/science/article/pii/S1097276500001271> [Accessed December 16, 2013].
- Carvill, G.L., Heavin, S.B., Yendle, S.C., McMahon, J.M., O’Roak, B.J., Cook, J., Khan, A., Dorschner, M.O., Weaver, M., Calvert, S., Malone, S., Wallace, G., Stanley, T., Bye, A.M.E., Bleasel, A., Howell, K.B., Kivity, S., Mackay, M.T., Rodriguez-Casero, V., et al., 2013. Targeted resequencing in epileptic encephalopathies identifies de novo mutations in CHD2 and SYNGAP1. *Nature genetics*, 45(7), pp.825–30. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3704157&tool=pmc-entrez&rendertype=abstract> [Accessed June 25, 2015].
- Castle, J.C., Zhang, C., Shah, J.K., Kulkarni, A. V, Kalsotra, A., Cooper, T.A. & Johnson, J.M., 2008. Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines. *Nature genetics*, 40(12), pp.1416–25. Available at: <http://dx.doi.org/10.1038/ng.264> [Accessed December 13, 2013].
- Chen, C., Cheng, L., Grennan, K., Pibiri, F., Zhang, C., Badner, J.A., Gershon, E.S. & Liu, C., 2013. Two gene co-expression modules differentiate psychotics and controls. *Molecular psychiatry*, 18(12), pp.1308–14. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4018461&tool=pmc-entrez&rendertype=abstract> [Accessed April 23, 2016].
- Claes, L., Del-Favero, J., Ceulemans, B., Lagae, L., Van Broeckhoven, C. & De Jonghe, P., 2001. De Novo Mutations in the Sodium-Channel Gene SCN1A Cause Severe Myoclonic Epilepsy of Infancy. *The American Journal of Human Genetics*, 68(6), pp.1327–1332. Available at: <http://www.sciencedirect.com/science/article/pii/S000292970761043X> [Accessed December 16, 2013].

- Claes, L., Del-Favero, J., Ceulemans, B., Lagae, L., Van Broeckhoven, C. & De Jonghe, P., 2001. De novo mutations in the sodium-channel gene SCN1A cause severe myoclonic epilepsy of infancy. *American journal of human genetics*, 68(6), pp.1327–32. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1226119&tool=pmc&rendertype=abstract> [Accessed April 24, 2016].
- Clarke, G.M., Anderson, C.A., Pettersson, F.H., Cardon, L.R., Morris, A.P. & Zondervan, K.T., 2011. Basic statistical analysis in genetic case-control studies. *Nature protocols*, 6(2), pp.121–33. Available at: <http://dx.doi.org/10.1038/nprot.2010.182> [Accessed October 2, 2014].
- 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., et al., 2011. A copy number variation morbidity map of developmental delay. *Nature genetics*, 43(9), pp.838–46. Available at: <http://dx.doi.org/10.1038/ng.909> [Accessed December 16, 2013].
- Curry, C.J., Stevenson, R.E., Aughton, D., Byrne, J., Carey, J.C., Cassidy, S., Cunniff, C., Graham, J.M., Jones, M.C., Kaback, M.M., Moeschler, J., Schaefer, G.B., Schwartz, S., Tarleton, J. & Opitz, J., 1997. Evaluation of mental retardation: recommendations of a Consensus Conference: American College of Medical Genetics. *American journal of medical genetics*, 72(4), pp.468–77. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9375733> [Accessed January 7, 2014].
- David, A., Razali, R., Wass, M.N. & Sternberg, M.J.E., 2012. Protein-protein interaction sites are hot spots for disease-associated nonsynonymous SNPs. *Human mutation*, 33(2), pp.359–63. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22072597> [Accessed December 16, 2013].
- Duong, L., Klitten, L.L., Møller, R.S., Ingason, A., Jakobsen, K.D., Skjødt, C., Didriksen, M., Hjalgrim, H., Werge, T. & Tommerup, N., 2012. Mutations in NRXN1 in a family multiply affected with brain disorders: NRXN1 mutations and brain disorders. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, 159B(3), pp.354–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22337556> [Accessed December 16, 2013].
- Epi4K consortium, 2012. Epi4K: gene discovery in 4,000 genomes. *Epilepsia*, 53(8), pp.1457–67. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3418423&tool=pmc&rendertype=abstract> [Accessed December 3, 2015].
- Erickson, R.P., 2016. The importance of de novo mutations for pediatric neurological disease-It is not all in utero or birth trauma. *Mutation research. Reviews in mutation research*, 767, pp.42–58. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/27036065> [Accessed April 23, 2016].
- Exome Aggregation Consortium, E.A., Lek, M., Karczewski, K., Minikel, E., Samocha, K., Banks, E., Fennell, T., O'Donnell-Luria, A., Ware, J., Hill, A., Cummings, B., Tukiainen, T., Birnbaum, D., Kosmicki, J., Duncan, L., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., et al., 2015. *Analysis of protein-coding genetic variation in 60,706 humans*, Cold Spring Harbor Labs Journals. Available at: <http://biorxiv.org/content/early/2015/10/30/030338.abstract> [Accessed October 31, 2015].
- Fay, J.C., Wyckoff, G.J. & Wu, C.-I., 2001. Positive and Negative Selection on the Human Genome. *Genetics*, 158(3), pp.1227–1234. Available at: <http://www.genetics.org/content/158/3/1227> [Accessed April 23, 2016].

- Fombonne, E., 2003. Epidemiological Surveys of Autism and Other Pervasive Developmental Disorders: An Update. *Journal of Autism and Developmental Disorders*, 33(4), pp.365–382. Available at: <http://link.springer.com/article/10.1023/A%3A1025054610557> [Accessed December 16, 2013].
- 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., et al., 2014. De novo mutations in schizophrenia implicate synaptic networks. *Nature*, 506(7487), pp.179–84. Available at: <http://dx.doi.org/10.1038/nature12929> [Accessed August 25, 2015].
- Garrison, E. & Marth, G., 2012. Haplotype-based variant detection from short-read sequencing. , p.9. Available at: <http://arxiv.org/abs/1207.3907> [Accessed April 23, 2016].
- Gibson, W.T., Hood, R.L., Zhan, S.H., Bulman, D.E., Fejes, A.P., Moore, R., Mungall, A.J., Eydoux, P., Babul-Hirji, R., An, J., Marra, M.A., Chitayat, D., Boycott, K.M., Weaver, D.D. & Jones, S.J.M., 2012. *Mutations in EZH2 Cause Weaver Syndrome*, Available at: <http://www.sciencedirect.com/science/article/pii/S0002929711004964> [Accessed December 16, 2013].
- Gilissen, C., Hehir-Kwa, J.Y., Thung, D.T., van de Vorst, M., van Bon, B.W.M., 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.A., Kleefstra, T., Brunner, H.G., et al., 2014. Genome sequencing identifies major causes of severe intellectual disability. *Nature*, 511(7509), pp.344–7. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24896178> [Accessed August 26, 2015].
- Girirajan, S., Rosenfeld, J.A., Coe, B.P., Parikh, S., Friedman, N., Goldstein, A., Filipink, R.A., McConnell, J.S., Angle, B., Meschino, W.S., Nezarati, M.M., Asamoah, A., Jackson, K.E., Gowans, G.C., Martin, J.A., Carmany, E.P., Stockton, D.W., Schnur, R.E., Penney, L.S., et al., 2012. Phenotypic heterogeneity of genomic disorders and rare copy-number variants. *The New England journal of medicine*, 367(14), pp.1321–31. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3494411&tool=pmc-entrez&rendertype=abstract> [Accessed April 19, 2016].
- Girirajan, S., Rosenfeld, J.A., Cooper, G.M., Antonacci, F., Siswara, P., Itsara, A., Vives, L., Walsh, T., McCarthy, S.E., Baker, C., Mefford, H.C., Kidd, J.M., Browning, S.R., Browning, B.L., Dickel, D.E., Levy, D.L., Ballif, B.C., Platky, K., Farber, D.M., et al., 2010. A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nature genetics*, 42(3), pp.203–9. Available at: <http://dx.doi.org/10.1038/ng.534> [Accessed December 16, 2013].
- Gupta, S., Ellis, S.E., Ashar, F.N., Moes, A., Bader, J.S., Zhan, J., West, A.B. & Arking, D.E., 2014. Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. *Nature communications*, 5, p.5748. Available at: <http://www.nature.com/ncomms/2014/141210/ncomms6748/full/ncomms6748.html> [Accessed February 23, 2016].

- Gusev, A., Lee, S.H., Trynka, G., Finucane, H., Vilhjálmsson, B.J., Xu, H., Zang, C., Ripke, S., Bulik-Sullivan, B., Stahl, E., Kähler, A.K., Hultman, C.M., Purcell, S.M., McCarroll, S.A., Daly, M., Pasaniuc, B., Sullivan, P.F., Neale, B.M., Wray, N.R., et al., 2014. Partitioning Heritability of Regulatory and Cell-Type-Specific Variants across 11 Common Diseases. *The American Journal of Human Genetics*, 95(5), pp.535–552. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4225595&tool=pmc&rendertype=abstract> [Accessed November 6, 2014].
- Haijma, S. V., Van Haren, N., Cahn, W., Koolschijn, P.C.M.P., Hulshoff Pol, H.E. & Kahn, R.S., 2013. Brain volumes in schizophrenia: a meta-analysis in over 18 000 subjects. *Schizophrenia bulletin*, 39(5), pp.1129–38. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3756785&tool=pmc&rendertype=abstract> [Accessed March 29, 2016].
- Hall, J., Trent, S., Thomas, K.L., O'Donovan, M.C. & Owen, M.J., 2015. Genetic Risk for Schizophrenia: Convergence on Synaptic Pathways Involved in Plasticity. *Biological Psychiatry*, 77(1), pp.52–58. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25152434> [Accessed February 16, 2016].
- Hansen, K.D., Brenner, S.E. & Dudoit, S., 2010. Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic acids research*, 38(12), p.e131. Available at: <http://nar.oxfordjournals.org/content/early/2010/04/14/nar.gkq224> [Accessed February 24, 2016].
- Helbig, K.L., Farwell Hagman, K.D., Shinde, D.N., Mroske, C., Powis, Z., Li, S., Tang, S. & Helbig, I., 2016. Diagnostic exome sequencing provides a molecular diagnosis for a significant proportion of patients with epilepsy. *Genetics in medicine: official journal of the American College of Medical Genetics*. Available at: <http://dx.doi.org/10.1038/gim.2015.186> [Accessed April 6, 2016].
- Hemmings, C.P., 2006. Schizophrenia spectrum disorders in people with intellectual disabilities. *Current opinion in psychiatry*, 19(5), pp.470–4. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16874118> [Accessed April 24, 2016].
- Henrichsen, C.N., Chaignat, E. & Reymond, A., 2009. Copy number variants, diseases and gene expression. *Human molecular genetics*, 18(R1), pp.R1–8. Available at: <http://hmg.oxfordjournals.org/content/18/R1/R1.abstract> [Accessed April 23, 2016].
- Henrichsen, C.N., Vinckenbosch, N., Zöllner, S., Chaignat, E., Pradervand, S., Schütz, F., Ruedi, M., Kaessmann, H. & Reymond, A., 2009. Segmental copy number variation shapes tissue transcriptomes. *Nature genetics*, 41(4), pp.424–9. Available at: <http://dx.doi.org/10.1038/ng.345> [Accessed April 21, 2016].
- Hildebrand, M.S., Dahl, H.-H.M., Damiano, J.A., Smith, R.J.H., Scheffer, I.E. & Berkovic, S.F., 2013. Recent advances in the molecular genetics of epilepsy. *Journal of medical genetics*, 50(5), pp.271–9. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23468209> [Accessed April 24, 2016].
- Hochstenbach, R., Buizer-Voskamp, J.E., Vorstman, J.A.S. & Ophoff, R.A., 2011. Genome arrays for the detection of copy number variations in idiopathic mental retardation, idiopathic generalized epilepsy and neuropsychiatric disorders: lessons for diagnostic workflow and research. *Cytogenetic and genome research*, 135(3-4), pp.174–202. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22056632> [Accessed April 23, 2016].
- Holbrook, J.A., Neu-Yilik, G., Hentze, M.W. & Kulozik, A.E., 2004. Nonsense-mediated decay approaches the clinic. *Nature genetics*, 36(8), pp.801–8. Available at: <http://dx.doi.org/10.1038/ng1403> [Accessed March 18, 2016].

- Hormozdiari, F., Penn, O., Borenstein, E. & Eichler, E.E., 2015. The discovery of integrated gene networks for autism and related disorders. *Genome research*, 25(1), pp.142–54. Available at: <http://genome.cshlp.org/content/early/2014/11/05/gr.178855.114.abstract> [Accessed April 23, 2016].
- Human Genome Sequencing Consortium, 2004. Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011), pp.931–45. Available at: <http://dx.doi.org/10.1038/nature03001> [Accessed July 15, 2014].
- Inoue, K., Khajavi, M., Ohyama, T., Hirabayashi, S., Wilson, J., Reggin, J.D., Mancias, P., Butler, I.J., Wilkinson, M.F., Wegner, M. & Lupski, J.R., 2004. Molecular mechanism for distinct neurological phenotypes conveyed by allelic truncating mutations. *Nature genetics*, 36(4), pp.361–9. Available at: <http://dx.doi.org/10.1038/ng1322> [Accessed December 16, 2013].
- Iossifov, I., O’Roak, B.J., Sanders, S.J., Ronemus, M., Krumm, N., Levy, D., Stessman, H.A., Witherspoon, K.T., Vives, L., Patterson, K.E., Smith, J.D., Paepfer, B., Nickerson, D.A., Dea, J., Dong, S., Gonzalez, L.E., Mandell, J.D., Mane, S.M., Murtha, M.T., et al., 2014. The contribution of de novo coding mutations to autism spectrum disorder. *Nature*, 515(7526), pp.216–21. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4313871&tool=pmc&rendertype=abstract> [Accessed October 29, 2014].
- Itsara, A., Cooper, G.M., Baker, C., Girirajan, S., Li, J., Absher, D., Krauss, R.M., Myers, R.M., Ridker, P.M., Chasman, D.I., Mefford, H., Ying, P., Nickerson, D.A. & Eichler, E.E., 2009. Population analysis of large copy number variants and hotspots of human genetic disease. *American journal of human genetics*, 84(2), pp.148–61. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2668011&tool=pmc&rendertype=abstract> [Accessed March 5, 2016].
- Itsara, A., Wu, H., Smith, J.D., Nickerson, D.A., Romieu, I., London, S.J. & Eichler, E.E., 2010. De novo rates and selection of large copy number variation. *Genome research*, 20(11), pp.1469–81. Available at: <http://genome.cshlp.org/content/20/11/1469.short> [Accessed December 16, 2013].
- Kaer, K. & Speek, M., 2013. Retroelements in human disease. *Gene*, 518(2), pp.231–41. Available at: <http://www.sciencedirect.com/science/article/pii/S0378111913000462> [Accessed April 13, 2016].
- Karalliedde, J. & Gnudi, L., 2016. Diabetes mellitus, a complex and heterogeneous disease, and the role of insulin resistance as a determinant of diabetic kidney disease. *Nephrology, dialysis, transplantation: official publication of the European Dialysis and Transplant Association - European Renal Association*, 31(2), pp.206–13. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25550448> [Accessed March 7, 2016].
- Karam, S.M., Riegel, M., Segal, S.L., Félix, T.M., Barros, A.J.D., Santos, I.S., Matijasevich, A., Giugliani, R. & Black, M., 2015. Genetic causes of intellectual disability in a birth cohort: a population-based study. *American journal of medical genetics. Part A*, 167(6), pp.1204–14. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25728503> [Accessed April 24, 2016].
- Kavanagh, D.H., Tansey, K.E., O’Donovan, M.C. & Owen, M.J., 2015. Schizophrenia genetics: emerging themes for a complex disorder. *Molecular psychiatry*, 20(1), pp.72–6. Available at: <http://dx.doi.org/10.1038/mp.2014.148> [Accessed March 21, 2016].

- Kelemen, O., Convertini, P., Zhang, Z., Wen, Y., Shen, M., Falaleeva, M. & Stamm, S., 2013. Function of alternative splicing. *Gene*, 514(1), pp.1–30. Available at: <http://www.sciencedirect.com/science/article/pii/S0378111912009791> [Accessed April 23, 2016].
- Kircher, M., Witten, D.M., Jain, P., O’Roak, B.J., Cooper, G.M. & Shendure, J., 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature genetics*, 46(3), pp.310–5. Available at: <http://dx.doi.org/10.1038/ng.2892> [Accessed July 9, 2014].
- 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.W., Sigurdsson, G., Walters, G.B., Steinberg, S., Helgason, H., Thorleifsson, G., Gudbjartsson, D.F., Helgason, A., Magnusson, O.T., et al., 2012. Rate of de novo mutations and the importance of father’s age to disease risk. *Nature*, 488(7412), pp.471–5. Available at: http://www.nature.com/nature/journal/v488/n7412/full/nature11396.html%3FWT.ec_id=NATURE-20120823 [Accessed December 13, 2013].
- Korbel, J.O., Urban, A.E., Affourtit, J.P., Godwin, B., Grubert, F., Simons, J.F., Kim, P.M., Palejev, D., Carriero, N.J., Du, L., Taillon, B.E., Chen, Z., Tanzer, A., Saunders, A.C.E., Chi, J., Yang, F., Carter, N.P., Hurler, M.E., Weissman, S.M., et al., 2007. Paired-end mapping reveals extensive structural variation in the human genome. *Science (New York, N.Y.)*, 318(5849), pp.420–6. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2674581&tool=pmc-entrez&rendertype=abstract> [Accessed March 31, 2016].
- Korir, P.K., Roberts, L., Ramesar, R. & Seoighe, C., 2014. A mutation in a splicing factor that causes retinitis pigmentosa has a transcriptome-wide effect on mRNA splicing. *BMC research notes*, 7(1), p.401. Available at: <http://bmcresearchnotes.biomedcentral.com/articles/10.1186/1756-0500-7-401> [Accessed May 1, 2016].
- Kornberg, R.D. & Lorch, Y., 1995. Interplay between chromatin structure and transcription. *Current Opinion in Cell Biology*, 7(3), pp.371–375. Available at: <http://www.sciencedirect.com/science/article/pii/0955067495800921> [Accessed December 16, 2013].
- Kumar, D., Whiteside, T.L. & Kasid, U., 2000. Identification of a novel tumor necrosis factor-alpha-inducible gene, SCC-S2, containing the consensus sequence of a death effector domain of fas-associated death domain-like interleukin-1beta-converting enzyme-inhibitory protein. *The Journal of biological chemistry*, 275(4), pp.2973–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10644768> [Accessed March 1, 2016].
- de la Grange, P., Gratadou, L., Delord, M., Dutertre, M. & Auboeuf, D., 2010. Splicing factor and exon profiling across human tissues. *Nucleic acids research*, 38(9), pp.2825–38. Available at: <http://nar.oxfordjournals.org/content/38/9/2825.long> [Accessed April 24, 2016].
- Lander, E. & Botstein, D., 1987. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science*, 236(4808), pp.1567–1570. Available at: <http://science.sciencemag.org/content/236/4808/1567.abstract> [Accessed April 23, 2016].

- Lappalainen, T., Sammeth, M., Friedländer, M.R., 't Hoen, P.A.C., Monlong, J., Rivas, M.A., González-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G., Barann, M., Wieland, T., Greger, L., van Iterson, M., Almlöf, J., Ribeca, P., Pulyakhina, I., Esser, D., Giger, T., et al., 2013. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*, 501(7468), pp.506–11. Available at: <http://dx.doi.org/10.1038/nature12531> [Accessed April 12, 2016].
- Layer, R.M., Chiang, C., Quinlan, A.R. & Hall, I.M., 2014. LUMPY: a probabilistic framework for structural variant discovery. *Genome biology*, 15(6), p.R84. Available at: <http://genomebiology.biomedcentral.com/articles/10.1186/gb-2014-15-6-r84> [Accessed October 15, 2014].
- Lewis, D.A., 2012. Cortical circuit dysfunction and cognitive deficits in schizophrenia--implications for preemptive interventions. *The European journal of neuroscience*, 35(12), pp.1871–8. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3383640&tool=pmc-entrez&rendertype=abstract> [Accessed February 2, 2016].
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, 25(16), pp.2078–9. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2723002&tool=pmc-entrez&rendertype=abstract> [Accessed July 9, 2014].
- Li, J., Cai, T., Jiang, Y., Chen, H., He, X., Chen, C., Li, X., Shao, Q., Ran, X., Li, Z., Xia, K., Liu, C., Sun, Z.S. & Wu, J., 2015. Genes with de novo mutations are shared by four neuropsychiatric disorders discovered from NPdenovo database. *Molecular psychiatry*, 21(2), pp.290–297. Available at: <http://dx.doi.org/10.1038/mp.2015.40> [Accessed January 20, 2016].
- Licatalosi, D.D. & Darnell, R.B., 2006. Splicing Regulation in Neurologic Disease. *Neuron*, 52(1), pp.93–101. Available at: <http://www.sciencedirect.com/science/article/pii/S0896627306007240> [Accessed December 16, 2013].
- Lichtenstein, P., Yip, B.H., Björk, C., Pawitan, Y., Cannon, T.D., Sullivan, P.F. & Hultman, C.M., 2009. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet (London, England)*, 373(9659), pp.234–9. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3879718&tool=pmc-entrez&rendertype=abstract> [Accessed March 2, 2016].
- de Ligt, J., Willemsen, M.H., van Bon, B.W.M., Kleefstra, T., Yntema, H.G., Kroes, T., Vulto-van Silfhout, A.T., Koolen, D.A., de Vries, P., Gilissen, C., del Rosario, M., Hoischen, A., Scheffer, H., de Vries, B.B.A., Brunner, H.G., Veltman, J.A. & Vissers, L.E.L.M., 2012. Diagnostic Exome Sequencing in Persons with Severe Intellectual Disability. *New England Journal of Medicine*, 367(20), pp.1921–1929. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23033978> [Accessed July 6, 2015].
- Linden, D.E.J., 2012. The challenges and promise of neuroimaging in psychiatry. *Neuron*, 73(1), pp.8–22. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22243743> [Accessed April 10, 2016].
- Lubs, H.A., Stevenson, R.E. & Schwartz, C.E., 2012. Fragile X and X-Linked Intellectual Disability: Four Decades of Discovery. *The American Journal of Human Genetics*, 90(4), pp.579–590. Available at: <http://www.sciencedirect.com/science/article/pii/S0002929712001036> [Accessed December 16, 2013].

- MacDonald, J.R., Ziman, R., Yuen, R.K.C., Feuk, L. & Scherer, S.W., 2014. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic acids research*, 42(Database issue), pp.D986–92. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3965079&tool=pmc-entrez&rendertype=abstract> [Accessed March 5, 2016].
- Malaspina, D., Harlap, S., Fennig, S., Heiman, D., Nahon, D., Feldman, D. & Susser, E.S., 2001. Advancing paternal age and the risk of schizophrenia. *Archives of general psychiatry*, 58(4), pp.361–7. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11296097> [Accessed March 21, 2016].
- Malkova, A. & Haber, J.E., 2012. Mutations arising during repair of chromosome breaks. *Annual review of genetics*, 46, pp.455–73. Available at: <http://www.annualreviews.org/doi/abs/10.1146/annurev-genet-110711-155547> [Accessed April 23, 2016].
- Mamanova, L., Coffey, A.J., Scott, C.E., Kozarewa, I., Turner, E.H., Kumar, A., Howard, E., Shendure, J. & Turner, D.J., 2010. Target-enrichment strategies for next-generation sequencing. *Nature methods*, 7(2), pp.111–8. Available at: <http://dx.doi.org/10.1038/nmeth.1419> [Accessed December 11, 2013].
- Mardis, E.R., 2008. Next-generation DNA sequencing methods. *Annual review of genomics and human genetics*, 9, pp.387–402. Available at: <http://www.annualreviews.org/doi/abs/10.1146/annurev.genom.9.081307.164359> [Accessed July 9, 2014].
- Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bembien, L.A., Berka, J., Braverman, M.S., Chen, Y.-J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X. V, Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Ho, C.H., et al., 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057), pp.376–80. Available at: <http://dx.doi.org/10.1038/nature03959> [Accessed December 11, 2013].
- Marth, G.T., Yu, F., Indap, A.R., Garimella, K., Gravel, S., Leong, W.F., Tyler-Smith, C., Bainbridge, M., Blackwell, T., Zheng-Bradley, X., Chen, Y., Challis, D., Clarke, L., Ball, E. V, Cibulskis, K., Cooper, D.N., Fulton, B., Hartl, C., Koboldt, D., et al., 2011. The functional spectrum of low-frequency coding variation. *Genome biology*, 12(9), p.R84. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3308047&tool=pmc-entrez&rendertype=abstract> [Accessed April 24, 2016].
- Maston, G.A., Evans, S.K. & Green, M.R., 2006. Transcriptional Regulatory Elements in the Human Genome. Available at: <http://www.annualreviews.org/doi/abs/10.1146/annurev.genom.7.080505.115623> [Accessed April 24, 2016].
- Maulik, P.K., Mascarenhas, M.N., Mathers, C.D., Dua, T. & Saxena, S., 2011. Prevalence of intellectual disability: a meta-analysis of population-based studies. *Research in developmental disabilities*, 32(2), pp.419–36. Available at: <http://www.sciencedirect.com/science/article/pii/S0891422210003082> [Accessed March 24, 2016].
- Mazin, P., Xiong, J., Liu, X., Yan, Z., Zhang, X., Li, M., He, L., Somel, M., Yuan, Y., Phoebe Chen, Y.-P., Li, N., Hu, Y., Fu, N., Ning, Z., Zeng, R., Yang, H., Chen, W., Gelfand, M. & Khaitovich, P., 2013. Widespread splicing changes in human brain development and aging. *Molecular systems biology*, 9, p.633. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3564255&tool=pmc-entrez&rendertype=abstract> [Accessed April 23, 2016].

- McGillivray, B.C., Bassett, A.S., Langlois, S., Pantzar, T. & Wood, S., 1990. Familial 5q11.2---q13.3 segmental duplication cosegregating with multiple anomalies, including schizophrenia. *American journal of medical genetics*, 35(1), pp.10–3. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3142272&tool=pmc-entrez&rendertype=abstract> [Accessed December 16, 2013].
- McGrath, J., Saha, S., Chant, D. & Welham, J., 2008. Schizophrenia: a concise overview of incidence, prevalence, and mortality. *Epidemiologic reviews*, 30, pp.67–76. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18480098> [Accessed July 14, 2014].
- McGrother, C.W., Bhaumik, S., Thorp, C.F., Hauck, A., Branford, D. & Watson, J.M., 2006. Epilepsy in adults with intellectual disabilities: prevalence, associations and service implications. *Seizure*, 15(6), pp.376–86. Available at: <http://www.sciencedirect.com/science/article/pii/S1059131106000677> [Accessed June 23, 2015].
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. & DePristo, M.A., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, 20(9), pp.1297–303. Available at: <http://genome.cshlp.org/content/20/9/1297.short> [Accessed December 11, 2013].
- Medvedev, P., Stanciu, M. & Brudno, M., 2009. Computational methods for discovering structural variation with next-generation sequencing. *Nature methods*, 6(11 Suppl), pp.S13–20. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19844226> [Accessed April 13, 2016].
- 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.H.F., de Leeuw, N., Reardon, W., Gimelli, S., et al., 2008. Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *The New England journal of medicine*, 359(16), pp.1685–99. Available at: <http://www.nejm.org/doi/full/10.1056/NEJMoa0805384> [Accessed May 1, 2016].
- Mercer, T.R., Gerhardt, D.J., Dinger, M.E., Crawford, J., Trapnell, C., Jeddloh, J.A., Mattick, J.S. & Rinn, J.L., 2012. Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nature biotechnology*, 30(1), pp.99–104. Available at: <http://dx.doi.org/10.1038/nbt.2024> [Accessed May 1, 2016].
- Merriman, B. & Rothberg, J.M., 2012. Progress in ion torrent semiconductor chip based sequencing. *Electrophoresis*, 33(23), pp.3397–417. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23208921> [Accessed February 9, 2016].
- Meyer, U. & Feldon, J., 2010. Epidemiology-driven neurodevelopmental animal models of schizophrenia. *Progress in neurobiology*, 90(3), pp.285–326. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19857543> [Accessed April 23, 2016].
- Meynert, A.M., Ansari, M., FitzPatrick, D.R. & Taylor, M.S., 2014. Variant detection sensitivity and biases in whole genome and exome sequencing. *BMC bioinformatics*, 15(1), p.247. Available at: <http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-15-247> [Accessed April 23, 2016].
- Middeldorp, J. & Hol, E.M., 2011. GFAP in health and disease. *Progress in neurobiology*, 93(3), pp.421–43. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21219963> [Accessed March 8, 2016].

- Morgan, V.A., Leonard, H., Bourke, J. & Jablensky, A., 2008. Intellectual disability co-occurring with schizophrenia and other psychiatric illness: population-based study. *The British journal of psychiatry : the journal of mental science*, 193(5), pp.364–72. Available at: <http://bjp.rcpsych.org/content/193/5/364> [Accessed April 24, 2016].
- Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. & Wold, B., 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods*, 5(7), pp.621–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18516045> [Accessed July 11, 2014].
- Motulsky, A.G., 2006. Genetics of complex diseases. *Journal of Zhejiang University. Science. B*, 7(2), pp.167–8. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1363767&tool=pmc-entrez&rendertype=abstract> [Accessed April 23, 2016].
- Mullaney, J.M., Mills, R.E., Pittard, W.S. & Devine, S.E., 2010. Small insertions and deletions (INDELs) in human genomes. *Human molecular genetics*, 19(R2), pp.R131–6. Available at: <http://hmg.oxfordjournals.org/content/19/R2/R131.short> [Accessed February 23, 2016].
- Van Naarden Braun, K., Autry, A. & Boyle, C., 2005. A population-based study of the recurrence of developmental disabilities--Metropolitan Atlanta Developmental Disabilities Surveillance Program, 1991-94. *Paediatric and perinatal epidemiology*, 19(1), pp.69–79. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15670112> [Accessed December 16, 2013].
- Najmabadi, H., Hu, H., Garshasbi, M., Zemojtel, T., Abedini, S.S., Chen, W., Hosseini, M., Behjati, F., Haas, S., Jamali, P., Zecha, A., Mohseni, M., Püttmann, L., Vahid, L.N., Jensen, C., Moheb, L.A., Bienek, M., Larti, F., Mueller, I., et al., 2011. Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature*, 478(7367), pp.57–63. Available at: <http://dx.doi.org/10.1038/nature10423> [Accessed December 12, 2013].
- Nelson, M.R., Wegmann, D., Ehm, M.G., Kessner, D., St Jean, P., Verzilli, C., Shen, J., Tang, Z., Bacanu, S.-A., Fraser, D., Warren, L., Aponte, J., Zawistowski, M., Liu, X., Zhang, H., Zhang, Y., Li, J., Li, Y., Li, L., et al., 2012. An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science (New York, N.Y.)*, 337(6090), pp.100–4. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4319976&tool=pmc-entrez&rendertype=abstract> [Accessed March 18, 2016].
- Ott, J., 1974. Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies. *American journal of human genetics*, 26(5), pp.588–97. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1762722&tool=pmc-entrez&rendertype=abstract> [Accessed April 23, 2016].
- Ottman, R. & Risch, N., 2012. Genetic Epidemiology and Gene Discovery in Epilepsy. Available at: <http://www.ncbi.nlm.nih.gov/books/NBK98200/> [Accessed April 24, 2016].
- Owen, M.J., 2012. Implications of genetic findings for understanding schizophrenia. *Schizophrenia bulletin*, 38(5), pp.904–7. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3446223&tool=pmc-entrez&rendertype=abstract> [Accessed March 1, 2016].
- Owen, M.J., Sawa, A. & Mortensen, P.B., 2016. Schizophrenia. *Lancet (London, England)*. Available at: <http://www.thelancet.com/article/S0140673615011216/fulltext> [Accessed January 22, 2016].

- Pan, Q., Shai, O., Lee, L.J., Frey, B.J. & Blencowe, B.J., 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature genetics*, 40(12), pp.1413–5. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18978789> [Accessed July 9, 2014].
- Paulussen, A.D.C., Stegmann, A.P.A., Blok, M.J., Tserpelis, D., Posma-Velter, C., Detisch, Y., Smeets, E.E.J.G.L., Wagemans, A., Schrandt, J.J.P., van den Boogaard, M.-J.H., van der Smagt, J., van Haeringen, A., Stolte-Dijkstra, I., Kerstjens-Frederikse, W.S., Mancini, G.M., Wessels, M.W., Hennekam, R.C.M., Vreeburg, M., Geraedts, J., et al., 2011. MLL2 mutation spectrum in 45 patients with Kabuki syndrome. *Human mutation*, 32(2), pp.E2018–25. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21280141> [Accessed December 16, 2013].
- Pereira, J.D., Sansom, S.N., Smith, J., Dobenecker, M.-W., Tarakhovskiy, A. & Livesey, F.J., 2010. Ezh2, the histone methyltransferase of PRC2, regulates the balance between self-renewal and differentiation in the cerebral cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 107(36), pp.15957–62. Available at: <http://www.pnas.org/content/107/36/15957.short> [Accessed December 16, 2013].
- Pescosolido, M.F., Gamsiz, E.D., Nagpal, S. & Morrow, E.M., 2013. Distribution of disease-associated copy number variants across distinct disorders of cognitive development. *Journal of the American Academy of Child and Adolescent Psychiatry*, 52(4), pp.414–430.e14. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3774163&tool=pmc-entrez&rendertype=abstract> [Accessed April 24, 2016].
- Pulst, S.M., 1999. Genetic Linkage Analysis. *Archives of Neurology*, 56(6), p.667. Available at: <http://archneur.jamanetwork.com/article.aspx?articleid=775035> [Accessed March 22, 2016].
- Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A., Swerdlow, H.P. & Gu, Y., 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC genomics*, 13, p.341. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3431227&tool=pmc-entrez&rendertype=abstract> [Accessed July 9, 2014].
- Ripke, S., Neale, B.M., Corvin, A., Walters, J.T.R., Farh, K.-H., Holmans, P.A., Lee, P., Bulik-Sullivan, B., Collier, D.A., Huang, H., Pers, T.H., Agartz, I., Agerbo, E., Albus, M., Alexander, M., Amin, F., Bacanu, S.A., Begemann, M., Belliveau Jr, R.A., et al., 2014. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, 511(7510), pp.421–427. Available at: <http://dx.doi.org/10.1038/nature13595> [Accessed July 22, 2014].
- Roberts, A., Pimentel, H., Trapnell, C. & Pachter, L., 2011. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics (Oxford, England)*, 27(17), pp.2325–9. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21697122> [Accessed July 14, 2014].
- Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S.D., Mungall, K., Lee, S., Okada, H.M., Qian, J.Q., Griffith, M., Raymond, A., Thiessen, N., Cezard, T., Butterfield, Y.S., Newsome, R., Chan, S.K., She, R., Varhol, R., et al., 2010. De novo assembly and analysis of RNA-seq data. *Nature methods*, 7(11), pp.909–12. Available at: <http://dx.doi.org/10.1038/nmeth.1517> [Accessed March 14, 2016].

- Robinson, M.D. & Oshlack, A., 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome biology*, 11(3), p.R25. Available at: <http://genomebiology.biomedcentral.com/articles/10.1186/gb-2010-11-3-r25> [Accessed January 7, 2015].
- Ropers, H.-H. & Hamel, B.C.J., 2005. X-linked mental retardation. *Nature reviews. Genetics*, 6(1), pp.46–57. Available at: <http://dx.doi.org/10.1038/nrg1501> [Accessed December 12, 2013].
- Roselló, M., Martínez, F., Monfort, S., Mayo, S., Oltra, S. & Orellana, C., 2014. Phenotype profiling of patients with intellectual disability and copy number variations. *European journal of paediatric neurology : EJPN : official journal of the European Paediatric Neurology Society*, 18(5), pp.558–66. Available at: <http://www.sciencedirect.com/science/article/pii/S1090379814000622> [Accessed April 23, 2016].
- Rosewich, H., Thiele, H., Ohlenbusch, A., Maschke, U., Altmüller, J., Frommolt, P., Zirn, B., Ebinger, F., Siemes, H., Nürnberg, P., Brockmann, K. & Gärtner, J., 2012. Heterozygous de-novo mutations in ATP1A3 in patients with alternating hemiplegia of childhood: a whole-exome sequencing gene-identification study. *The Lancet Neurology*, 11(9), pp.764–773. Available at: <http://www.sciencedirect.com/science/article/pii/S1474442212701825> [Accessed December 16, 2013].
- Roukos, V., Voss, T.C., Schmidt, C.K., Lee, S., Wangsa, D. & Misteli, T., 2013. Spatial dynamics of chromosome translocations in living cells. *Science (New York, N.Y.)*, 341(6146), pp.660–4. Available at: <http://science.sciencemag.org/content/341/6146/660.abstract> [Accessed March 11, 2016].
- Sanchez-Mut, J. V, Huertas, D. & Esteller, M., 2012. Aberrant epigenetic landscape in intellectual disability. *Progress in brain research*, 197, pp.53–71. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22541288> [Accessed April 24, 2016].
- Sanger, F., Nicklen, S. & Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12), pp.5463–5467. Available at: <http://www.pnas.org/content/74/12/5463.short> [Accessed December 16, 2013].
- Santen, G.W.E., Aten, E., Sun, Y., Almomani, R., Gilissen, C., Nielsen, M., Kant, S.G., Snoeck, I.N., Peeters, E.A.J., Hilhorst-Hofstee, Y., Wessels, M.W., den Hollander, N.S., Ruivenkamp, C.A.L., van Ommen, G.-J.B., Breuning, M.H., den Dunnen, J.T., van Haeringen, A. & Kriek, M., 2012. Mutations in SWI/SNF chromatin remodeling complex gene ARID1B cause Coffin-Siris syndrome. *Nature genetics*, 44(4), pp.379–80. Available at: <http://dx.doi.org/10.1038/ng.2217> [Accessed December 12, 2013].
- Schadt, E.E., Monks, S.A., Drake, T.A., Lusk, A.J., Che, N., Colino, V., Ruff, T.G., Milligan, S.B., Lamb, J.R., Cavet, G., Linsley, P.S., Mao, M., Stoughton, R.B. & Friend, S.H., 2003. Genetics of gene expression surveyed in maize, mouse and man. *Nature*, 422(6929), pp.297–302. Available at: <http://dx.doi.org/10.1038/nature01434> [Accessed April 23, 2016].
- Scherer, S.W., Lee, C., Birney, E., Altshuler, D.M., Eichler, E.E., Carter, N.P., Hurles, M.E. & Feuk, L., 2007. Challenges and standards in integrating surveys of structural variation.
- Schubert, S., Bollag, G. & Shannon, K., 2007. Deregulated Ras signaling in developmental disorders: new tricks for an old dog. *Current opinion in genetics & development*, 17(1), pp.15–22. Available at: <http://europemc.org/abstract/med/17208427> [Accessed April 24, 2016].

- Schwarz, J.M., Rödelsperger, C., Schuelke, M. & Seelow, D., 2010. MutationTaster evaluates disease-causing potential of sequence alterations. *Nature methods*, 7(8), pp.575–6. Available at: <http://dx.doi.org/10.1038/nmeth0810-575> [Accessed December 13, 2013].
- Sekar, A., Bialas, A.R., de Rivera, H., Davis, A., Hammond, T.R., Kamitaki, N., Tooley, K., Presumey, J., Baum, M., Van Doren, V., Genovese, G., Rose, S.A., Handsaker, R.E., Daly, M.J., Carroll, M.C., Stevens, B. & McCarroll, S.A., 2016. Schizophrenia risk from complex variation of complement component 4. *Nature*, 530(7589), pp.177–183. Available at: <http://dx.doi.org/10.1038/nature16549> [Accessed January 27, 2016].
- Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D., Zhang, K., Mitra, R.D. & Church, G.M., 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science (New York, N.Y.)*, 309(5741), pp.1728–32. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16081699> [Accessed June 25, 2015].
- Sherry, S.T., 2001. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research*, 29(1), pp.308–311. Available at: <http://nar.oxfordjournals.org/content/29/1/308.long> [Accessed May 20, 2015].
- Shihab, H.A., Gough, J., Mort, M., Cooper, D.N., Day, I.N.M. & Gaunt, T.R., 2014. Ranking non-synonymous single nucleotide polymorphisms based on disease concepts. *Human genomics*, 8, p.11. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4083756&tool=pmc-entrez&rendertype=abstract> [Accessed April 23, 2016].
- Singh, N.A., Charlier, C., Stauffer, D., DuPont, B.R., Leach, R.J., Melis, R., Ronen, G.M., Bjerre, I., Quattlebaum, T., Murphy, J. V., McHarg, M.L., Gagnon, D., Rosales, T.O., Peiffer, A., Anderson, V.E. & Leppert, M., 1998. A novel potassium channel gene, KCNQ2, is mutated in an inherited epilepsy of newborns. *Nature genetics*, 18(1), pp.25–9. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9425895> [Accessed April 24, 2016].
- Solanki, R.K., Singh, P., Midha, A. & Chugh, K., 2008. Schizophrenia: Impact on quality of life. *Indian journal of psychiatry*, 50(3), pp.181–6. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2738356&tool=pmc-entrez&rendertype=abstract> [Accessed April 23, 2016].
- Staden, R., 1979. A strategy of DNA sequencing employing computer programs. *Nucleic acids research*, 6(7), pp.2601–10. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=327874&tool=pmc-entrez&rendertype=abstract> [Accessed February 9, 2016].
- Steele, C.D., 2000. The genetics of Alzheimer disease. *The Nursing clinics of North America*, 35(3), pp.687–94. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10957683> [Accessed December 16, 2013].
- Steinlein, O.K., Mulley, J.C., Propping, P., Wallace, R.H., Phillips, H.A., Sutherland, G.R., Scheffer, I.E. & Berkovic, S.F., 1995. A missense mutation in the neuronal nicotinic acetylcholine receptor alpha 4 subunit is associated with autosomal dominant nocturnal frontal lobe epilepsy. *Nature genetics*, 11(2), pp.201–3. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/7550350> [Accessed April 24, 2016].
- Stranger, B.E., Forrest, M.S., Clark, A.G., Minichiello, M.J., Deutsch, S., Lyle, R., Hunt, S., Kahl, B., Antonarakis, S.E., Tavaré, S., Deloukas, P. & Dermitzakis, E.T., 2005. Genome-wide associations of gene expression variation in humans. *PLoS genetics*, 1(6), p.e78. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1315281&tool=pmc-entrez&rendertype=abstract> [Accessed April 23, 2016].

- Strong, M.J., 2010. The evidence for altered RNA metabolism in amyotrophic lateral sclerosis (ALS). *Journal of the Neurological Sciences*, 288(1), pp.1–12. Available at: <http://www.sciencedirect.com/science/article/pii/S0022510X09008892> [Accessed January 7, 2014].
- Sullivan, P.F., Daly, M.J. & O’Donovan, M., 2012. Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nature reviews. Genetics*, 13(8), pp.537–51. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4110909&tool=pmc-entrez&rendertype=abstract> [Accessed March 9, 2016].
- Sullivan, P.F., Kendler, K.S. & Neale, M.C., 2003. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Archives of general psychiatry*, 60(12), pp.1187–92. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/14662550> [Accessed February 26, 2015].
- Tanaka, A.J., Cho, M.T., Retterer, K., Jones, J.R., Nowak, C., Douglas, J., Jiang, Y.-H., McConkie-Rosell, A., Schaefer, G.B., Kaylor, J., Rahman, O.A., Telegrafi, A., Friedman, B., Douglas, G., Monaghan, K.G. & Chung, W.K., 2016. De novo pathogenic variants in CHAMP1 are associated with global developmental delay, intellectual disability, and dysmorphic facial features. *Molecular Case Studies*, 2(1), p.a000661. Available at: <http://molecularcasestudies.cshlp.org/content/2/1/a000661.full> [Accessed April 24, 2016].
- Tandon, R., Keshavan, M.S. & Nasrallah, H.A., 2008. Schizophrenia, “just the facts” what we know in 2008. 2. Epidemiology and etiology. *Schizophrenia research*, 102(1-3), pp.1–18. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18514488> [Accessed July 18, 2014].
- Tennessen, J.A., Bigham, A.W., O’Connor, T.D., Fu, W., Kenny, E.E., Gravel, S., McGee, S., Do, R., Liu, X., Jun, G., Kang, H.M., Jordan, D., Leal, S.M., Gabriel, S., Rieder, M.J., Abecasis, G., Altshuler, D., Nickerson, D.A., Boerwinkle, E., et al., 2012. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science (New York, N.Y.)*, 337(6090), pp.64–9. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3708544&tool=pmc-entrez&rendertype=abstract> [Accessed March 4, 2016].
- Thusberg, J., Olatubosun, A. & Vihinen, M., 2011. Performance of mutation pathogenicity prediction methods on missense variants. *Human mutation*, 32(4), pp.358–68. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21412949> [Accessed December 16, 2013].
- Torkamani, A., Dean, B., Schork, N.J. & Thomas, E.A., 2010. Coexpression network analysis of neural tissue reveals perturbations in developmental processes in schizophrenia. *Genome research*, 20(4), pp.403–12. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2847743&tool=pmc-entrez&rendertype=abstract> [Accessed April 2, 2016].
- Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L. & Pachter, L., 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature biotechnology*, 31(1), pp.46–53. Available at: <http://dx.doi.org/10.1038/nbt.2450> [Accessed July 9, 2014].
- Treangen, T.J. & Salzberg, S.L., 2012. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nature reviews. Genetics*, 13(1), pp.36–46. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3324860&tool=pmc-entrez&rendertype=abstract> [Accessed April 7, 2016].

- Tsurusaki, Y., Okamoto, N., Ohashi, H., Kosho, T., Imai, Y., Hibi-Ko, Y., Kaname, T., Naritomi, K., Kawame, H., Wakui, K., Fukushima, Y., Homma, T., Kato, M., Hiraki, Y., Yamagata, T., Yano, S., Mizuno, S., Sakazume, S., Ishii, T., et al., 2012. Mutations affecting components of the SWI/SNF complex cause Coffin-Siris syndrome. *Nature genetics*, 44(4), pp.376–8. Available at: <http://dx.doi.org/10.1038/ng.2219> [Accessed December 16, 2013].
- Wang, E.T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P. & Burge, C.B., 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature*, 456(7221), pp.470–6. Available at: <http://dx.doi.org/10.1038/nature07509> [Accessed December 12, 2013].
- Wang, X., Wei, X., Thijssen, B., Das, J., Lipkin, S.M. & Yu, H., 2012. Three-dimensional reconstruction of protein networks provides insight into human genetic disease. *Nature biotechnology*, 30(2), pp.159–64. Available at: <http://dx.doi.org/10.1038/nbt.2106> [Accessed December 16, 2013].
- Wang, Z., Gerstein, M. & Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics*, 10(1), pp.57–63. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2949280&tool=pmc-entrez&rendertype=abstract> [Accessed July 10, 2014].
- 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., et al., 2001. The sequence of the human genome. *Science (New York, N.Y.)*, 291(5507), pp.1304–51. Available at: <http://www.sciencemag.org/content/291/5507/1304.abstract> [Accessed December 11, 2013].
- Wheeler, D.L., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M., Edgar, R., Federhen, S., Geer, L.Y., Kapustin, Y., Khovayko, O., Landsman, D., Lipman, D.J., Madden, T.L., Maglott, D.R., Ostell, J., Miller, V., et al., 2007. Database resources of the National Center for Biotechnology Information. *Nucleic acids research*, 35(Database issue), pp.D5–12. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1781113&tool=pmc-entrez&rendertype=abstract> [Accessed June 24, 2015].
- Vissers, L.E.L.M., Gilissen, C. & Veltman, J.A., 2016. Genetic studies in intellectual disability and related disorders. *Nature reviews. Genetics*, 17(1), pp.9–18. Available at: <http://dx.doi.org/10.1038/nrg3999> [Accessed April 24, 2016].
- Voineagu, I., Wang, X., Johnston, P., Lowe, J.K., Tian, Y., Horvath, S., Mill, J., Cantor, R.M., Blencowe, B.J. & Geschwind, D.H., 2011. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature*, 474(7351), pp.380–4. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3607626&tool=pmc-entrez&rendertype=abstract> [Accessed March 15, 2016].
- Wright, C.F., Fitzgerald, T.W., Jones, W.D., Clayton, S., McRae, J.F., van Kogelenberg, M., King, D.A., Ambridge, K., Barrett, D.M., Bayzietinova, T., Bevan, A.P., Bragin, E., Chatzimichali, E.A., Gribble, S., Jones, P., Krishnappa, N., Mason, L.E., Miller, R., Morley, K.I., et al., 2014. Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *The Lancet*, 385(9975), pp.1305–1314. Available at: <http://europepmc.org/articles/PMC4392068> [Accessed December 17, 2014].

- Yang, X., Coulombe-Huntington, J., Kang, S., Sheynkman, G.M., Hao, T., Richardson, A., Sun, S., Yang, F., Shen, Y.A., Murray, R.R., Spirohn, K., Begg, B.E., Duran-Frigola, M., MacWilliams, A., Pevzner, S.J., Zhong, Q., Trigg, S.A., Tam, S., Ghamsari, L., et al., 2016. Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing. *Cell*, 164(4), pp.805–817. Available at: <http://www.sciencedirect.com/science/article/pii/S0092867416300435> [Accessed February 13, 2016].
- Yang, Y., Muzny, D.M., Xia, F., Niu, Z., Person, R., Ding, Y., Ward, P., Braxton, A., Wang, M., Buhay, C., Veeraraghavan, N., Hawes, A., Chiang, T., Leduc, M., Beuten, J., Zhang, J., He, W., Scull, J., Willis, A., et al., 2014. Molecular Findings Among Patients Referred for Clinical Whole-Exome Sequencing. *JAMA*, 312(18), pp.1870–1879. Available at: <http://jama.jamanetwork.com/article.aspx?articleid=1918774> [Accessed October 18, 2014].
- Yin, X., Cheng, H., Lin, Y., Fan, X., Cui, Y., Zhou, F., Shen, C., Zuo, X., Zheng, X., Zhang, W., Yang, S. & Zhang, X., 2014. Five regulatory genes detected by matching signatures of eQTL and GWAS in psoriasis. *Journal of dermatological science*, 76(2), pp.139–42. Available at: <http://www.sciencedirect.com/science/article/pii/S0923181114001807> [Accessed April 24, 2016].
- Yoon, S., Xuan, Z., Makarov, V., Ye, K. & Sebat, J., 2009. Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome research*, 19(9), pp.1586–92. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2752127&tool=pmc-entrez&rendertype=abstract> [Accessed April 23, 2016].
- Zhang, J., Lieu, Y.K., Ali, A.M., Penson, A., Reggio, K.S., Rabadan, R., Raza, A., Mukherjee, S. & Manley, J.L., 2015. Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities. *Proceedings of the National Academy of Sciences of the United States of America*, 112(34), pp.E4726–34. Available at: <http://www.pnas.org/content/112/34/E4726.abstract> [Accessed March 24, 2016].
- Zhang, Z.D., Du, J., Lam, H., Abyzov, A., Urban, A.E., Snyder, M. & Gerstein, M., 2011. Identification of genomic indels and structural variations using split reads. *BMC genomics*, 12, p.375. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3161018&tool=pmc-entrez&rendertype=abstract> [Accessed April 23, 2016].
- Zou, F., Chai, H.S., Younkin, C.S., Allen, M., Crook, J., Pankratz, V.S., Carrasquillo, M.M., Rowley, C.N., Nair, A.A., Middha, S., Maharjan, S., Nguyen, T., Ma, L., Malphrus, K.G., Palusak, R., Lincoln, S., Bisceglia, G., Georgescu, C., Kouri, N., et al., 2012. Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. *PLoS genetics*, 8(6), p.e1002707. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3369937&tool=pmc-entrez&rendertype=abstract> [Accessed April 21, 2016].

Acta Universitatis Upsaliensis

*Digital Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Medicine 1231*

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine”.)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-287407



ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2016