
9

General discussion

Discoveries made in the 20th century had a major impact in genetics. In less than 60 years, the field has moved from discovering the structure of DNA to potentially identifying all the variants in an individual's genome. Current (diagnostic) research focuses on identifying the genetic basis of disorders, both at the fine (sequence variants) and large scale (structural variants). While Sanger sequencing has been the golden standard for detection of sequence variants for many years, replacement by whole exome sequencing (WES) is already taking place, soon to be followed by whole genome sequencing (WGS). WES now merely focuses on detection of sequence variation but, although limited, allows detection of structural variation as well. For detection of structural variation, conventional karyotyping has been replaced by microarray platforms, and the resolution of the technique has increased from the 5-10 Mb level towards the kilobase level. With NGS, even smaller structural variation will be detected. For balanced rearrangements and complex rearrangements, although WGS can identify breakpoints, conventional karyotyping and FISH are still valuable techniques.

Several techniques described in this thesis have made their way into clinical diagnostic laboratories. Chapters in this thesis show that each technique has specific applications, and which technical approach to use depends on the resolution and throughput of the technique in combination with the strategy for the identification of the disease gene. While chapter 2 and chapter 4 describe more methodological work on techniques to detect sequence variation and copy number variation, chapters 3 and 5 describe the application of these techniques to the identification of pathogenic variants. Chapter 6 and 7 reflect the rapid introduction of NGS and the successful application of exome sequencing in clinical genetics.

Technological innovations are ongoing. For some of the research published in this thesis we would now already apply other techniques. For instance, the development of a 1400-plex bead assay to screen for CNVs in patients with intellectual disability (chapter 2) is now obsolete, as microarrays made their way into diagnostic laboratories. However, this targeted approach can still be considered useful to circumvent dilemmas like the occurrence of unsolicited findings connected to genome-wide screening. Detailed delineation of microdeletion/microduplication syndromes (chapter 3) is time consuming, and exome sequencing will probably aid in establishing genotype-phenotype correlations within these patients. Exome sequencing of mentally retarded patients with unknown aetiology may reveal pathogenic variants in genes previously described in microdeletion/microduplication intervals. This has been the case for Coffin-Siris syndrome (chapter 8) where pathogenic variants in *ARID1B* were identified in patients using exome sequencing. This gene was previously described in patients carrying a larger deletion affecting *ARID1B,* indicating haploinsuffiency as the causative effect. Additionally, in some of the cases, exome sequencing will allow detection of pathogenic variants on the other allele in genes within

microdeletion/microduplication intervals, confirming a recessive mode of inheritance.

Gene identification in Mendelian disorders

In general, knowledge of advantages and disadvantages of different laboratory techniques will aid in setting up experiments, and allows a critical assessment of subsequent findings. The identification of genes responsible for Mendelian disorders enables molecular diagnosis of patients, as well as testing gene carriers and prenatal testing. Uncovering genetic defects underlying monogenic inherited diseases (this thesis) requires a broad knowledge of genetics. Strategies to identify pathogenic variants include assumptions on inheritance patterns (dominant, *de novo* dominant, recessive, X-linked), careful selection of cases based on phenotype, collection of families and choosing the best experimental approach.

Several strategies for gene identification can be chosen, such as a straightforward candidate gene approach or positional cloning. Previous strategies such as linkage analysis combined with segregation studies and homozygosity mapping have proven successful. Candidate genes within identified genomic loci can then be studied using Sanger sequencing. The combination of classic strategies with novel techniques has proved to be valuable in KFSD research (chapter 4). Data from previously performed linkage analysis were combined with SNP array data to reduce the critical interval of the locus. HRMA was then used as a screening method to identify the pathogenic variant in the candidate genes. There is no doubt that, in this case, X-exome sequencing would have performed equally in identifying the pathogenic variant and would have been faster. This strategy was applied in TOD (Chapter 6), where a splice-site variant was identified in the linkage interval [1]. Remarkably, the variant was missed using other techniques. The effect of the variant on gene splicing could only be shown on patient-derived tumour tissue that had been stored for 15 years. This story illustrates several problems related to sequence variation. Not only can a variant remain undetected due to technical failure, it is also quite possible that a variant is indeed detected, but not recognized as being pathogenic.

The underlying gene defect is still unknown for many rare monogenic diseases but the development of WES has allowed us to resolve many more cases. In particular, *de novo* variants for sporadic cases of previously unknown genetic disorders were identified [2-6].

Genome-wide techniques such as SNP arrays have already changed our approach to genetic disorders and this will continue as the use of WES/WGS will eventually become a standard technique in clinical diagnostic laboratories. The identification of causative variants in patients moved from clinical recognition of a syndrome and collecting patients with overlapping clinical features (phenotype-driven research), towards collecting patients with similar genomic aberrations and, secondly, delineating the phenotypic overlap (genotype-

driven research). The research on Coffin-Siris syndrome (chapter 8) reflects the success of accurate phenotyping of several sporadic patients and searching for a shared defective gene under assumption of different inheritance patterns. The fact that this syndrome is now proven to be *de novo* instead of, as assumed, following autosomal recessive inheritance has a major impact on recurrence risk for parents. The identification allowed us to reduce this risk from 25% to 1-2%. The success of identifying pathogenic variants in a single gene, in multiple unrelated patients with a similar phenotype, is highly dependent on the involvement of a clinical geneticist. Therefore, selection of patients for WES should be performed in close collaboration between molecular and clinical geneticists.

The discovery of causal variants and candidate genes responsible for Mendelian disorders and complex disorders will also help in understanding their biological function.
In addition to WES, experiments can be performed to provide more insight into disease mechanisms. For instance, KFSD (chapter 5), TOD (chapter 6), and AAS (chapter 7) all show X-linked inheritance. In principle, females have random X-inactivation. Thus, for some X-linked disorders, carrier females may show variable and usually milder symptoms of the disease, based on the pattern of X-inactivation. The X-inactivation studies performed for KFSD and TOD showed imbalances in allelic expression, perfectly matched with skewed levels of X-inactivation. In addition, for KFSD, functional *in vitro* assays were developed to prove that the identified variant leads to loss of proteolytic function of the MBTPS2 protein. For AAS, the putative branch point variant was proven to have an effect on splicing by performing RNA analysis. Together, these experiments were critical for understanding the disease mechanism and explaining the severity of phenotype in these disorders. For Coffin-Siris syndrome, functional assays are being developed to provide more insight into epigenetic modifications in the SWI/SNF complex.

Advantages and disadvantages in variant detection with WES

Currently, high-throughput sequencing of the entire genome is still too expensive to be applied in a clinical setting, so a targeted approach such as exome sequencing is the most practical approach towards finding causative variants for genetic disorders. However, NGS is becoming cheaper at an accelerating rate and whole genome sequencing will soon become affordable.
Exome sequencing comes with several disadvantages. It is labour intensive, and includes many steps, with risks of sample swaps and pipetting errors. There is significant variability in capturing efficiency, leading to variability in coverage across the genome. Low input yield and preferential amplification of shorter sequences in the PCR step leads to overrepresentation of one allele, and some reads will be exact copies of each other (PCR

duplicates). This will result in problems regarding variant detection in samples. Removal of these duplicates is possible during data analysis, but insufficient coverage due to capture inefficiency cannot be disregarded. An adequate depth is especially critical for identifying heterozygote variants in *de novo* dominant disorders, or recessive disorders caused by compound heterozygosity. Standards for generation of high-quality exome sequence data are rapidly emerging [7, 8] but there are no clear guidelines yet on the number of reads needed for identification of heterozygote SNV calls, and thresholds are determined by the user when performing data analysis.

Another challenge of NGS is read length. Current NGS machines produce shorter read lengths than Sanger sequencing, which makes the assembly of repeat-rich sequences more difficult [9]. Together with mapping difficulties, this leads to missed variants or an excess of variant calls [10]. Paired-end sequencing (where sequences are retrieved from both ends of the same molecule) helps in reducing mapping errors although indels and structural variants remain challenging. To solve this, other computational methods (read-pair, read-depth, split-read) have been developed [11, 12]. False positives SNV calls most often arise from incorrect mapping and sequencing errors. Sequencing errors can be removed by comparing the test sample to previously sequenced samples, stored in an in-house-database (INHDB). Although detection of structural variants with WES is possible [13], conventional techniques such as karyotyping, FISH, array-CGH and SNP-arrays will still provide useful information on structural variation until WES or WGS becomes a routine diagnostic approach.

The human reference genome plays an important role in several steps of WES. It is used to design chemically synthesized nucleic acid molecules and to design probes to capture regions of interest in the pool of nucleic acids (DNA or RNA). This means that targeted assays, such as WES, do not cover unknown or yet-to-be-identified exons, regulatory sequences or evolutionary conserved coding regions [10]. Moreover, if causal variants lie within exons that are not targeted, they will not be identified. This shortcoming of incomplete capture has occurred in the first attempts by several groups, including our own, to identify the genetic cause for Kabuki syndrome [2].

WES is not targeted towards intronic sequences, promoter or enhancer elements. In routine diagnostics, apart from the direct splice sites, causative variants outside coding regions are not studied. Therefore the true fraction of variants disrupting splicing remains largely unknown. However, the research that was performed for Aarskog-Scott syndrome (Chapter 7) shows that WES can be used to identify intronic variants up to -200 bp from the splice site with sufficient coverage, although it is not specifically designed to do so. A branch point variant was identified that was missed with Sanger sequencing. Detection of branch site variants is possible with WES due to probe overhang. Future probe design for targeted WES approaches may take advantage of this finding. On the other hand, if there is

a continuous demand to increase the coverage or size of targeted regions, whole genome sequencing is probably the best option in the long run. The disadvantage of extending into the non-coding sequences will be the large number of variants detected, and the need to verify aberrant splicing using RNA analysis which is not always possible if the gene is not expressed in available cells.

Whole genome sequencing (WGS) using a paired-end approach should eventually allow the identification of all types of currently known genomic variation in a single experiment, and will gradually replace WES. Improving NGS techniques and lowering the costs, together with the development of affordable analysis software to cope with the millions of variants called per genome, is essential to pave the way for diagnostic application of WGS. Development of software that can combine data from several sources containing information on reported variants has a high priority.

Interpretation of variants using NGS

The more we learn about the human genome, the more we are confronted with the need to separate 'the wheat from the chaff' i.e. distinguishing pathogenic variants from neutral variants. In NGS experiments, individual sequences are aligned to a reference genome. The choice for a specific reference genome (for instance Caucasian or Yoruba) will significantly influence variant detection in the sample of interest. It is important to realize that the human reference genome does not represent any one person's genome, but merely comprises of a mix of sequences from different chromosomes as well as individuals. This means that the human reference genome also contains variations from several donors, including somatic variants. Moreover, the haploid reference sequence has a bias toward a European population background (International Human Genome Consortium, 2004).

In general, variant positions are compared between the sample of interest and the reference, followed by a number of filtering steps to separate benign variants from potential pathogenic variants. The main strategy employed to identify causative variants in WES is to focus on variants with clear functional consequences (nonsense variants, splice-site variants, frameshift, conserved missense and indels). Depending on the enrichment kit, the sequencing platform and the pipeline used for mapping, alignment and variant calling, approximately 20000-50000 variants are identified per exome. After filtering noncoding and synonymous variants, the number of variants remaining is ~5000 [14]. For Mendelian disorders, most of the common and less likely causal variants can be efficiently removed using data from control databases and prediction tools, and between 150-500 non-synonymous and splice site variants remain to be prioritized as potentially pathogenic [15]. Downstream analysis is based on certain assumptions of inheritance (e.g. dominant, recessive, *de novo*) Further filtering strategies are needed to find the causative variant. The

success rate also depends on prior genetic mapping approaches (linkage, homozygosity mapping, CNV analysis) and availability of other patients and/or family members. This means that exome sequencing is highly suitable for cases where linkage intervals had been established, but the causal variant could not be found.

For some Mendelian disorders where the classic linkage study design was unsuitable and a candidate gene approach was not possible or unsuccessful (in very small families, in a few unrelated cases from different families, or even in sporadic cases), exome sequencing also offers opportunities. As there is no candidate genomic locus known, choices have to be made regarding the variants that deserve follow-up work, and proof of pathogenicity has to be obtained. This will require even closer collaboration between clinical and molecular geneticists. Segregation studies are often used to establish co-inheritance with the putative causal variant. For non-coding genomic variants, genomic information alone is not enough to prove pathogenicity. Sequencing RNA isolated from the same blood samples used to extract DNA may provide an attractive approach. Functional studies on RNA and protein (animal models, *in vitro* assays) will become even more important.

Interpreting new findings and translating these to practical healthcare remains a challenge. In addition to an increase in sequence variants, there is currently little experience in interpreting small structural variants, and the interpretation of somatic changes on a large scale. A number of studies have successfully applied WES to study effects of somatic mutations and/or epigenetic regulation and mosaicism, especially in cancer development [16-18].

With the discovery of new SNVs and CNVs in the DNA diagnostic laboratories, the need for practical guidelines on interpretation of genomic variation increased. Both molecular geneticist, cytogeneticists and clinical geneticists are learning to deal with the benefits and drawbacks of applying new techniques in daily clinical practice. Communicating the consequences of unclassified variants is often ambiguous, and efforts should be made to obtain reasonable certainty on the clinical consequences of the variant. Database information (e.g. LOVD, OMIM, Decipher, DGV [19-23]) outcomes, or even suggestions for further studies (e.g. functional studies on RNA or protein, segregation studies) and *in silico* predictions on functional consequences of the variant are useful information to include in the analysis report. Communication of unclassified variants requires special attention because the information given is often not completely understood or misinterpreted by patients and their family [24].

Unanticipated chance findings that are medically relevant (incidental findings) will provide further challenges for counselling. In general, it is very difficult to discuss the disclosure of particular findings to the counselee beforehand, as the spectrum of possibilities is immense. However, as exome sequencing is making its way in the diagnostic field, improved guidelines

for proper informed consent are being developed. One way to minimize these challenges is to analyse only known genes or a disease-specific gene set. This, however, means missing an opportunity to find novel disease loci or make a different diagnosis, especially in clinically and genetically heterogeneous disorders.

An additional formidable challenge is correlating low frequency sequence variants and CNVs to multifactorial diseases such as intellectual disability, autism, schizophrenia, and many other disorders. For complex disorders, Genome-Wide Association Studies (GWAS) have aimed to associate common variants to common diseases by studying large case-control cohorts. However, most of these associated SNVs have a very small effect, and confer a small risk to disease [25]. Identification of all variants in an individual with a common disease, for example autism, may reveal more rare functional variants that have a strong impact on disease, sometimes within a protein network [26]. On the other hand, many of these risk variants are not recurrent but *de novo*. Thus, both *de novo* and extremely rare inherited SNVs and CNVs contribute together to the overall genetic risk. This means that disorders with extreme genetic heterogeneity require the analysis of very large numbers of clinically well-defined patients and extensive sequence data for validation, because the study of individual families will not answer which combinations may or may not be pathogenic [26, 27].
It is currently impossible to predict the outcome of a pregnancy of a foetus carrying a CNV/SNV of this type. Since microarray analysis is increasingly used also in the prenatal setting, such findings create difficult problems in genetic counselling.
In addition, by extending family studies of these apparent strong risk factors for congenital malformation and intellectual disability we will identify young healthy carriers who will want to know the risk of disability in their offspring, and will ask for accurate prenatal diagnosis. The knowledge gained from studying Mendelian disorders and complex diseases will eventually complement each other and aid in genetic counselling.

Databases

Studies on healthy and disease cohorts have identified many variants, which are stored in several databases to enhance clinical interpretation of unclassified variants (SNVs, CNVs). Accumulation of unclassified variants led to the need of dedicated locus specific databases (LSDBs) to enhance clinical interpretation. For all disorders studied in this thesis (KFSD-*MBTPS2*, TOD-*FLNA*, Aarskog-Scott syndrome-*FGD1*, Coffin Siris syndrome-*ARID1B*), LSDBs have been set-up and are publicly available. Variant information can be stored and easily queried by users worldwide, and are subject to editorial screening by curators. In theory, storage in databases available on the internet is effective. In practice, however, a large number of different databases have emerged, leading to the lack of centralised

data storage. Data on genomic variation becomes scattered, depending on where the user decides to upload their experimental data. This could be solved by reaching definite (international) agreements on sharing of variants by laboratories, and honouring the existing commitments. Larger and more general data repositories (including NCBI and UCSC) have requested of locus-specific variant databases (LSDBs) to share their data [28]. The HGMD represents an attempt to collate known gene lesions responsible for human inherited disease [29]. HGMD also includes some variant data from those LSDBs that are in the public domain. Researchers and diagnosticians, genetic counsellors and physicians can use this information of practical diagnostic importance. The need for large, trustworthy databases will only increase when the use of next generation sequencing for diagnostic purposes gains in popularity.

Conclusion

This thesis outlines the rapid development of molecular techniques for detecting genomic variation. It describes advantages and disadvantages of specific techniques, and shows their application in the search for pathogenic variants in patients with genetic disorders. Next generation sequencing brings genetics into a new era of identifying sequence and structural variants in the context of disease. Both the techniques itself and the associated analysis tools will continue to improve in the future, eventually resulting in an approach that can detect all genomic variation in one assay. Every technical advance in genetic analysis has revealed a new level of variation within our genome i.e. initially chromosomal variation, then copy number changes of decreasing size, down to nucleotide level. As further methodological developments can be anticipated, it is reasonable to assume that there are additional levels of genomic complexity still to be revealed. Through identification of new disease genes, the NGS techniques will increase our knowledge of the molecular pathogenesis of genetic disorders. Interpreting these data and translating findings to improve healthcare will be a great challenge. The availability of new techniques for genomic analysis, and the associated benefits for patients, should be balanced against each patient's right not to know what their genome contains. Collaboration between research and diagnostics is pivotal for determining the most appropriate protocols for returning information to counselees. With NGS, the focus in clinical genetics the coming years will shift from identification to interpretation of variants, requiring more follow-up work. Further technical advances are necessary to enable investigation of other genetic mechanisms, including somatic mutations, epigenetic dysregulation and polygenic disruption, each of which is likely to account for a significant proportion of disease.

1. Sun Y, Almomani R, Aten E, Celli J, van der HJ, Venselaar H, Robertson SP, Baroncini A, Franco B, Basel-Vanagaite L, Horii E, Drut R, Ariyurek Y, den Dunnen JT, and Breuning MH (2010) Terminal osseous dysplasia is caused by a single recurrent mutation in the FLNA gene. Am J Hum Genet 87 (1):146-153

2. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, Lee C, Turner EH, Smith JD, Rieder MJ, Yoshiura K, Matsumoto N, Ohta T, Niikawa N, Nickerson DA, Bamshad MJ, and Shendure J (2010) Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet 42 (9):790-793

3. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, and Bamshad MJ (2010) Exome sequencing identifies the cause of a mendelian disorder. Nat Genet 42 (1):30-35

4. Gilissen C, Arts HH, Hoischen A, Spruijt L, Mans DA, Arts P, van LB, Steehouwer M, van RJ, Kant SG, Roepman R, Knoers NV, Veltman JA, and Brunner HG (2010) Exome sequencing identifies WDR35 variants involved in Sensenbrenner syndrome. Am J Hum Genet 87 (3):418-423

5. Hoischen A, van Bon BW, Gilissen C, Arts P, van LB, Steehouwer M, de VP, de RR, Wieskamp N, Mortier G, Devriendt K, Amorim MZ, Revencu N, Kidd A, Barbosa M, Turner A, Smith J, Oley C, Henderson A, Hayes IM, Thompson EM, Brunner HG, de Vries BB, and Veltman JA (2010) De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. Nat Genet 42 (6):483-485

6. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, and Shendure J (2009) Targeted capture and massively parallel sequencing of 12 human exomes. Nature 461 (7261):272-276

7. Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, McGee S, Do R, Liu X, Jun G, Kang HM, Jordan D, Leal SM, Gabriel S, Rieder MJ, Abecasis G, Altshuler D, Nickerson DA, Boerwinkle E, Sunyaev S, Bustamante CD, Bamshad MJ, and Akey JM (2012) Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science 337 (6090):64-69

8. Do R, Kathiresan S, and Abecasis GR (2012) Exome sequencing and complex disease: practical aspects of rare variant association studies. Hum Mol Genet 21 (R1):R1-R9

9. Bolouri A (2009) personal genomics and personalized medicine.

10. Majewski J, Schwartzentruber J, Lalonde E, Montpetit A, and Jabado N (2011) What can exome sequencing do for you? J Med Genet 48 (9):580-589

11. Albers CA, Lunter G, Macarthur DG, McVean G, Ouwehand WH, and Durbin R (2011) Dindel: accurate indel calls from short-read data. Genome Res 21 (6):961-973

12. Ye K, Schulz MH, Long Q, Apweiler R, and Ning Z (2009) Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics 25 (21):2865-2871

13. Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, Kim PM, Palejev D, Carriero NJ, Du L, Taillon BE, Chen Z, Tanzer A, Saunders AC, Chi J, Yang F, Carter NP, Hurles ME, Weissman SM, Harkins TT, Gerstein MB, Egholm M, and Snyder M (2007) Paired-end mapping reveals extensive structural variation in the human genome. Science 318 (5849):420-426

14. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, and Bamshad MJ (2010) Exome sequencing identifies the cause of a mendelian disorder. Nat Genet 42 (1):30-35

15. Gilissen C, Hoischen A, Brunner HG, and Veltman JA (2012) Disease gene identification strategies for exome sequencing. Eur J Hum Genet 20 (5):490-497

16. Yan XJ, Xu J, Gu ZH, Pan CM, Lu G, Shen Y, Shi JY, Zhu YM, Tang L, Zhang XW, Liang WX, Mi JQ, Song HD, Li KQ, Chen Z, and Chen SJ (2011) Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. Nat Genet 43 (4):309-315

17. Grossmann V, Tiacci E, Holmes AB, Kohlmann A, Martelli MP, Kern W, Spanhol-Rosseto A et al (2011) Whole-exome sequencing identifies somatic mutations of BCOR in acute myeloid leukemia with normal karyotype. Blood 118 (23):6153-6163

18. Vissers LE, Fano V, Martinelli D, Campos-Xavier B, Barbuti D, Cho TJ, Dursun A, Kim OH, Lee SH, Timpani G, Nishimura G, Unger S, Sass JO, Veltman JA, Brunner HG, Bonafe L, Dionisi-Vici C, and Superti-Furga A (2011) Whole-exome sequencing detects somatic mutations of IDH1 in metaphyseal chondromatosis with D-2-hydroxyglutaric aciduria (MC-HGA). Am J Med Genet A 155A (11):2609-2616

19. Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, Van VS, Moreau Y, Pettett RM, and Carter NP (2009) DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am J Hum Genet 84 (4):524-533

20. Swaminathan GJ, Bragin E, Chatzimichali EA, Corpas M, Bevan AP, Wright CF, Carter NP, Hurles ME, and Firth HV (2012) DECIPHER: web-based, community resource for clinical interpretation of rare variants in developmental disorders. Hum Mol Genet 21 (R1):R37-R44

21. Database of Genomic Variants. Http://projects.tcag.ca/variation/

22. Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, and den Dunnen JT (2011) LOVD v.2.0: the next generation in gene variant databases. Hum Mutat 32 (5):557-563

23. Online Mendelian Inheritance in Man. www.omim.org

24. Vos J, Gomez-Garcia E, Oosterwijk JC, Menko FH, Stoel RD, van Asperen CJ, Jansen AM, Stiggelbout AM, and Tibben A (2010) Opening the psychological black box in genetic counseling. The psychological impact of DNA testing is predicted by the counselees' perception, the medical impact by the pathogenic or uninformative BRCA1/2-result. Psychooncology
25. Cirulli ET and Goldstein DB (2010) Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet 11 (6):415-425

26. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, Levy R, Ko A, Lee C, Smith JD, Turner EH, Stanaway IB, Vernot B, Malig M, Baker C, Reilly B, Akey JM, Borenstein E, Rieder MJ, Nickerson DA, Bernier R, Shendure J, and Eichler EE (2012) Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 485 (7397):246-250

27. Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, Chu SH et al (2011) Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron 70 (5):863-885

28. den Dunnen JT, Sijmons RH, Andersen PS, Vihinen M, Beckmann JS, Rossetti S, Talbot CC, Jr., Hardison RC, Povey S, and Cotton RG (2009) Sharing data between LSDBs and central repositories. Hum Mutat 30 (4):493-495

29. Stenson PD, Ball EV, Howells K, Phillips AD, Mort M, and Cooper DN (2009) The Human Gene Mutation Database: providing a comprehensive central mutation database for molecular diagnostics and personalized genomics. Hum Genomics 4 (2):69-72


T C C G A G T T C C C T G G A
T C C G A G T T C C C T G G A
G T T C C T T C C G A G T T C
A A T G A G A A T C C G C C G
G T G A G A G C C C C G T C T
G T A C C T A C T G A G T T C
G A G T T A T G G T T C C T T
T C C G A G T T G T A A T T T
A A A T T T G A A A T C T G G
T G C C T A C T G A G T T C C
G T G C C T A C T G A G T T C
G A G C C C C G T C T G G T A
G T T C C T T C C G A G T T C
G G T T C C T T C C G A G T T
T T C C T T C C G A C T T C C