The genetics of spinocerebellar ataxia and dystonia

Nibbeling, Esther

General Discussion

7

SCA and dystonia are both genetically very heterogeneous disorders for which there is, as yet, no therapy available. Currently, 45 SCA types have been described and mutations found in 31 separate genes (see **chapter 1**). Dystonia is in a similar situation: 21 genes are mutated in 29 dystonia subtypes (see **chapter 1**). Despite the genetic heterogeneity, the symptoms of all SCA types are relatively homogeneous, as all patients exhibit a cerebellar ataxic syndrome accompanied by a wide range of other features. In dystonia, the clinical presentation is more heterogeneous as it can affect multiple body parts; head, trunk, legs or arms, or even fingers or vocal cords. Currently, despite the implementation of so-called ataxia gene panels for diagnostic sequencing, approximately 25% of all SCA patients are still without a genetic diagnosis. For this group it would be beneficial to identify novel disease genes as this would improve the genetic counseling they are given, while knowing the underlying molecular cause of the disease can be the starting point for the development of new therapies. We have shown that it is not easy to pinpoint the disease-causing variant from exome sequencing data. However, using a combined approach of whole exome sequencing and targeted resequencing of likely candidates, we did identify seven putative novel SCA genes (**chapter 2** and **chapter 4**) and one dystonia gene (**chapter 3**). Moreover, the pathogenicity of variants should be validated by functional analysis, as shown in **chapter 5** for a range of potential SCA13 mutations.

The clinical homogeneity between various SCA types, the overlap in symptoms with dystonia, and the role of the cerebellum in both disorders are reasons to search for common pathological mechanisms underlying these disorders. However, the majority of the reported disease genes have completely different biological roles, so that, to date, no singular molecular mechanism has been identified for either SCA or dystonia. Based on co-expression data of known disease genes, the pathway involving most of the disease genes is synaptic signaling, followed by nervous system development (**chapter 1**). This finding shows that not all the known disease genes are involved in a single pathway, which complicates the use of targeted resequencing of all genes in a particular pathway as a gene identification strategy (as shown in **chapter 4**). In this chapter we found that mutations in genes encoding components of the glutamatergic signaling system do not frequently cause SCA, but are more often linked to intellectual disability, which raises questions about the hypothesis of glutamatergic signaling as a common mechanism in SCAs.

The quest for new genes

Not so long ago, the search for causal genes in diseases was mainly carried out by linkage analysis in large families. This analysis yielded large chromosomal areas in which

the disease gene was presumed to be located. However, these intervals commonly contained many genes, and Sanger sequencing all those genes one by one was a laborious job and, in most cases, did not lead to identification of the disease-causing mutation. Nowadays, sequencing multiple genes has become much easier with the introduction of next generation sequencing (NGS) techniques, which are high throughput and therefore time- and cost-effective. Whole exome sequencing (WES), in which the entire coding region of the DNA is sequenced, is now common practice when it comes to identifying genes for Mendelian forms of dystonia and SCA. So far, 20 ataxia and dystonia genes have been identified using exome sequencing, including recently discovered mutations in *COL6A3* (DYT27)[1] and *TRPC3* in SCA41.[2] With the realization that WES yields long lists of candidate variants, linkage analysis is still commonly used to pinpoint the region of interest and to filter exome sequencing results, as described in the identification of *CACNA1G* in SCA42.[3] We have successfully used WES to identify *CACNA1H* as the causal gene in writer's cramp (**chapter 3**). The success rate of WES in the identification of novel disease genes is not 100%, possibly due to several reasons. On the technical side, variants can be missed by lack of sequence coverage, or because of a large deletion or duplication, which is not seen with sequencing. Other variants might be missed by bioinformatic variant-calling issues or by misinterpretation in the variant-filtering steps. Moreover, the causal variant may be located outside the coding regions of the genome.[4]

An alternative for WES is Whole Genome Sequencing (WGS) in which the entire genome is sequenced, including the non-coding regions. WGS does not depend on target enrichment before sequencing and therefore shows a more uniform coverage of the entire genome, independent of the GC percentage. This means that all exons will be covered whereas first exons are regularly missed with WES, and it makes WGS more suitable for copy number analysis.[5] Furthermore, WGS has been shown to be able to identify coding variants in intellectual disability patients that had been missed by WES analysis.[6] However, WGS yields a vast amount of data, i.e. ~4 million variants with WGS compared to 40,000 in WES,[6,7] and the interpretation of non-coding variants is still a huge challenge. Despite the efforts of the Encyclopedia of DNA elements (ENCODE) and the NIH Roadmap Epigenetics project, which showed that non-coding parts of the genome are involved in all kinds of regulatory processes influencing many genes,[8,9] we still do not know what the consequences are for the majority of variants seen in our genome. One way to overcome this issue is to link genetic variation to gene expression, since variants that correlate with specific changes in gene expression are more likely to be disease-causing than variants that have no effect on gene expression. Adding RNA sequencing data to WES data gives insight into the relative expression of the sequenced variants compared to non-mutated genes, and offers the possibility to examine the specific tissue of interest. However, for neurological disorders, this would

imply the need for brain samples which are not easy, if not impossible, to obtain unlike blood samples, for example. Nonetheless, many ataxia genes are expressed in blood, rendering RNA sequencing a promising option in the identification of novel ataxia genes.

One method that does not incorporate the entire genome, but only a predefined subset of genes, is targeted resequencing (TRS). This method reduces the cost and time needed for sequencing and data analysis. TRS is frequently used and is an effective method for screening using so-called gene panels for genetically heterogeneous disorders in diagnostic settings, such as cardiomyopathy,[10] peripheral neuropathies,[11] or focal epilepsy.[12] However, if the disease-causing variant is not found within this predefined set of genes, WES or WGS is still needed, which will double the cost and time, and needs more genetic counseling as WES and WGS entail the risk of unsolicited findings. Therefore, in current diagnostic settings, so called "clinical exomes" are rising in popularity. In these WES is used to screen initially the predefined set of genes for mutations but when the disease-causing variant is not identified in these genes, the rest of the exome is available for analysis. We used TRS in combination with WES (described in **chapter 2)** to identify five novel putative SCA genes, while in **chapter 4** we describe a pathway-oriented gene search using TRS to identify two new SCA genes.

The disadvantages of all NGS methods are that large deletions and duplications cannot be efficiently detected as yet and it remains very difficult to sequence coding or non-coding repetitive sequences, rendering these methods unsuitable for diagnosing multiple common SCA types, such as SCA1, 2, 3, 6, and 7. For disorders that are solely caused by conventional missense or nonsense mutations, NGS can completely replace conventional methods in DNA diagnostics.

The struggles of variant selection

The main challenge in performing any kind of NGS approach is the filtering and selection of long lists of genetic variants for follow-up studies. The first step is to assess the quality of the sequencing reads and include variants with a read depth of at least 20x to reduce the amount of false-positive calls. Furthermore, the zygosity of the variants needs to be in line with the expected mode of disease inheritance in the patient. Moreover, variants outside the coding and splice regions and those that do not result in amino acid changes (synonymous) are usually eliminated from further analysis, since it is assumed that these will have minimal impact on protein function.[4]

The following step in the filtering is to assess the frequency of the variants in common genetic databases, since its frequency is highly predictive of the variant's severity.[4,13,14] As both spinocerebellar ataxia and dystonia are rare disorders with disease frequencies

of 1-3 and 16 in 100,000, respectively, and they show dominant inheritance patterns, we expect the disease-causing variants to be present only at a very low frequency (<0.01%), if at all, in genetic databases like 1000Genomes, dbSNP, EVS, and ExAC.[15–18] However, according to the disease frequency, we might go down to 0.001% for SCA, although this increases the risk of eliminating the disease-causing variant since database frequencies depend highly on their population size and in smaller databases frequencies will thus be higher. The newest database is the Reference Variant Store (RVS), which contains 473 million distinct variants from 82,600 samples and includes resources such as ClinVar, OMIM, COSMIC, and prediction scores from dbNSFP.[19] However, as these databases continue to grow and the disease status of the individuals is not monitored, the chance that a database will house disease-causing mutations grows, and variants should therefore not be excluded from further analysis by only their (rare) presence in databases. We encountered this situation in the work described in **chapter 3,** in which the p.Arg481Cys mutation in *CACNA1H* was found as a cause for writer's cramp. This variant is present twice in the ExAC database, with a frequency of 0.0001134, and we therefore used functional tests to prove its pathogenicity.

Recently, an elegant method linked to genetic databases was published; it labels genes with a gene damage index (GDI). This is based on the notion that 58% of variants in a single patient's exome data are found in 2.4% of all genes, rendering these genes highly unlikely to be the cause of monogenic disorders. The genes with the lowest GDI score are thus more likely to be disease genes for monogenic disorders and, not surprisingly, these genes are mostly highly conserved and indispensable genes involved in ribosomes, chemokine signaling, proteasome, and spliceosome.[20] However, this model does not take gene size into account, which will yield higher GDI scores for larger genes as they have more chance of accumulating genetic variation over time and, therefore, they might be assigned a false-positive status more easy. Not surprisingly, from the novel genes we identified and reported in **chapter 2**, the very large genes, *FAT1* and *FAT2,* were assigned a high mutation level, whereas the other candidates, including *CACNA1H* from **chapter 3**, were assigned a medium mutation level.

To estimate the functional effect of non-synonymous SNVs, multiple computational prediction programs have been developed. These are either based on amino acid conservation such as PolyPhen2,[21] SIFT,[22] MutationAssessor,[23] and MSRV,[24] or on DNA sequence alignments such as PhyloP,[25] GERP,[26] SiPhy,[27] and SInBaD;[28] they differ from each other mainly by their algorithm and training data sets.[29] However, these programs return different outcomes for single variants making it very difficult to interpret the results. Therefore, one should consult multiple programs and set an arbitrary cut-off at which a variant is predicted to be damaging by the majority of programs. In **chapter 2** we report using SIFT, PolyPhen2 and MutationTaster (these are commonly used pro-

grams and available via the Alamut diagnostic software) and our follow-up on variants which all three programs assigned a damaging prediction to.

To overcome the discrepancies in predicted variant effect of the various programs and return a comprehensible output, multiple tools have been developed to integrate several predictions, these are the so-called ensemble tools. For example, combined annotation-dependent depletion (CADD) integrates a range of data types including transcript information, conservation scores, regulatory information, and protein-level prediction scores.[30] This method is suitable for most types of variants, including non-coding variants, unlike PolyPhen and SIFT, for example, that can only be used to predict the effect of missense variants. Moreover, deleterious annotation of genetic variants using neural networks (DANN), a program that is based on CADD, uses another algorithm to capture non-linear relationships among the features and it therefore has a more detailed prediction of the variant's impact.[31] Furthermore, SnvForest integrates eleven functional scores at the variant level and eight association scores at the gene level, which are fewer than CADD or DANN.[32] Intriguingly, in addition to SIFT and PolyPhen2, CADD was one of the functional scores, meaning that SIFT and PolyPhen2 are integrated twice, which leads to a bias towards the SIFT and PolyPhen2 training sets. We did not use any of these ensemble scores in our work, as we had already analyzed our data before these tools were available, however, unsurprisingly, post-selection analysis showed high (>20) CADD-Phred scores for ten out of twelve variants in novel SCA genes reported in **chapter 2**. The ensemble tools are very promising for analyzing single cases and we will use them to further analyze our unsolved families. However, a major drawback of all prediction tools is that they only analyze single nucleotide variants, whereas alterations of two adjacent nucleotides may cause pathogenic amino acid changes that might be missed by filtering and analyzing single nucleotides only.

Other models have been built that integrate phenotypic data into the analysis, enhancing candidate gene prioritization, as they will search for genes involved in phenotypes similar to that of the investigated patient. Examples of these predictors are eXtasy, Phen-Gen, PhenIX, Phevor, and Exomiser, which were reviewed by Smedley and Robinson;[33] they all make use of the Human Phenotype Ontology (HPO) database. PHenotypic Interpretation of Variants in Exomes (PHIVE) performs cross-species analysis approaches with HPO and the Mammalian Phenotype Ontology (MPO) database to identify similarities between human disease characteristics and phenotypes seen in genetically modified model organisms, mainly rodents.[34] This could be really helpful, although of course it relies highly on the number of rodent experiments and which phenotypic traits have been taken into account. Furthermore, the recently published OMIMexplorer makes use of protein interaction networks of genes involved in diseases that are phenotypically closely related to the disease seen in the patient, as determined

with semantic searches in the Online Mendelian Inheritance in Men (OMIM) database.[35] This method comes very close to our approach, described in **chapter 2,** of creating a set of higher prioritized genes based on co-expression with known ataxia genes. However, we did not take other similar phenotypes into account and we think that this tool is very promising for the analysis of single cases.

Although the computational prediction tools are becoming more accurate in discriminating damaging variants from those that are likely benign, the outcome is still a prediction and not 'real biology'. Therefore, functional analysis is highly recommended in order to validate the severity of the variant at the protein-, cell- or even organism-level, although many new disease genes are being published guided only by computational predictions.[36]

The proof of variant causality

Whenever a list of prioritized variants is created, the challenge of proving causality of the variant with the disease starts. MacArthur et al. have established guidelines to describe causality of sequence variants involving assessments on the gene and variant level.[36] The most conventional starting point, that we also used in **chapter 2** and **chapter 3**, is to confirm the presence of the variants in the patient's genome by Sanger sequencing. Thereafter, available family members can be tested for co-segregation, as segregating variants should be correlated with the disease (*i.e.* all affected family members should carry the variant and unaffected family members preferably not). However, this is rather tricky in late-onset disorders such as SCA since currently non-symptomatic family members may become affected in the future. Consequently, we prefer not to include supposedly unaffected family members in our validation, although multiple studies use the absence of a variant in non-affected family members as proof of variant pathogenicity in the affected members. Moreover, low penetrance of the disease mutations is seen in dystonia, indicating that not all mutation carriers go on to develop symptoms.[37,38]

In the situation that a variant is highly likely to be the disease-causing variant in a family, another predicted damaging variant in the same gene should preferentially be present in an independent case or another independent family. If variants in a gene can be found in independent patients with the same phenotype, this makes a much stronger case for that gene being involved in the pathology.[36] However, because of the rarity of SCA and dystonia, finding a second mutation in the same gene in an independent case or family is not always possible, as reported in **chapter 2** and **chapter 3**. Moreover, in **chapter 5**, we show that screening a large cohort of patients (n=848) to find variants in *KCNC3* yielded twelve novel variants, of which only one was validated as a new

mutation after functional studies. In **chapter 6** we show that even for genes that are published as being associated with disease, such as *ARSG* in writer's cramp, finding the disease-causing variant is a tough job. Out of eleven sequenced variants, we were only able to prove association of a single variant to the disease phenotype and we failed to identify a causal variant. However, this could also be due to the fact that the original study was a GWAS and the causal mutation was not necessarily located in *ARSG*.[39]

For the family-based analyses, it is extremely important that the medical doctor makes an accurate clinical characterization of all the available family members. To be able to perform co-segregation analysis we need to know which family members are affected and which ones are certainly not. When establishing a cohort to screen for additional variants in a putative candidate gene, the cases in the cohort should suffer from the same disease as the probands, and the cohort should preferably not include patients who only show part of the expected phenotype. Moreover, these patients should not have a genetic diagnosis for another subtype of the disease or have been referred only for exclusion of the disease: as we have seen in **chapter 4** this can be misleading and we identified a predicted damaging variant in *GRIA3* in a person suffering from intellectual disability instead of cerebellar ataxia. The importance of the clinical characterization also highlights the importance of clear communication between medical doctors and research/diagnostic scientists.

If no additional affected family members are available, the list of validated variants will remain rather extensive and, in this situation, gene prioritization by computational prediction programs that include phenotypic data might be highly beneficial (see section above). Additionally, genes can be manually selected based on their function and expression — although this is not very efficient. However, genes can have multiple functions in different tissues or even within the same tissue and a gene should therefore not be excluded based on only its known function. The prioritization approach that we describe in **chapter 2** included analysis of gene co-expression networks based on the assumption that genes co-expressed with known SCA genes are more likely to be involved in the same biological pathways, and therefore more likely to be novel disease genes.

The next step in variant validation is to determine the true effect of the variant by functional analyses.[36] The most commonly used follow-up at protein level is to determine the effect of the variant on gene function in a cell model. This permits the expression and localization of the mutated protein to be assessed, as changes in these parameters will give clues that the variant is not benign. We did this for the variants in *GRIN3B* and *GRIK1* in **chapter 4** and for variants in several genes in **chapter 2**. However, these results are not conclusive since not all damaging variants affect protein levels, but if the function of the protein is known, the effect of the variant on protein function can be

determined via activity measurements. These could involve electrophysiological experiments for ion channels, for example, as shown in **chapter 2**, **chapter 3** and **chapter 5**, or enzyme activity measurements, as described in **chapter 2**. The proper cell model should, however, be chosen carefully, as easy-to-handle models such as HEK293 or HeLa are not always the best representation of the disease. For brain disorders, brain cells should preferably be used, but these are not often available and more difficult to maintain in culture.

Ultimately, we need to generate *in vivo* models that mimic the patients' phenotype. In the case of an autosomal dominant disease, the putative disease gene can be overexpressed or knockout models can be generated. But the generation and characterization of *in vivo* models, such as mouse models, is very expensive and takes a long time, and therefore they cannot be used in a high-throughput fashion for each variant that is slightly interesting.

CRISPR/Cas9 is a new technique based on clustered, regularly interspaced, short palindromic repeats and their associated protein 9. This method has shown great potential in genome editing in a wide variety of cell types and organisms.[40–42] It is based on RNA-guided, targeted double-strand DNA breaks that are repaired by either non-homologous end joining, inducing insertions or deletions, or homology-directed repair when a donor template is present, leading to induction of specific insertions, deletions or mutations.[42,43] This method allows a disease model to be created that carries the same mutation as patients and the model is therefore a better representation of the disease than the traditional overexpression or knockout models. Moreover, mutations can be generated in a heterozygous state, simplifying the possibility to study dominant negative effects which are probably causing several SCA types, e.g. SCA19.[44] In patient cell lines, this technique can be used to undo the mutation, which will hopefully rescue the phenotype.

NGS beyond gene identification

In addition to the identification of novel disease genes, NGS techniques are suitable for other genetic applications, such as research on putative genetic modifiers of disorders. In many diseases, patients with the same mutation show differences in disease presentation regarding age at onset (AO), disease severity, disease progression, and secondary manifestations.[45] In SCA3, for example, age of onset is mainly determined by the length of the CAG-repeat stretch in *ATXN3,*[46,47] however, not all patients with the same repeat length have the same AO, which suggests there are disease-modifying factors.[45,47,48] We have used WES plus targeted sequencing of the promoter regions of 262 genes of interest to identify potential modifiers in 32 SCA3 patients who showed a deviation of

some 10 years from the expected AO based on their repeat length (data not shown in this thesis). We compared the distribution of variants between 16 relatively early onset cases and 16 relatively late onset cases using Fisher's exact test and we corrected for multiple testing. Unfortunately, our cohort was too small to obtain significant results. The top hit associated with early AO was an intergenic TA-insertion (rs35443981) located between zinc finger genes *ZNF818P* and *ZNF677*, but we did not detect any obvious candidates in the top-10 variants. However, in a large, seven-generation Dutch family we used WES to detect three variants in four early AO cases that were absent in five late AO cases in the same family. Unfortunately, none of the three variants segregated completely with the phenotype in the rest of the family. A major problem in such an analysis is the arbitrary definition of AO and the observational bias that is seen in this, as people are more aware of disease symptoms and onset if a disease runs in the family. In other disorders, such as Huntington's disease, in which the presence of modifiers was investigated, several were identified that all have a very small effect on the AO but their functions remain unexplained.[49] This demonstrates the difficulty of this kind of research, but it presents a great opportunity for using new methods to interpret non-coding variants and for using RNA-sequencing, given that gene expression might well explain a substantial part of the phenotypic variance.

Therapeutic strategies in SCA and dystonia

Despite the efforts to identify more disease genes in SCA and dystonia, both disorders are still incurable. This illustrates the need to elucidate the molecular pathways involved to aid the development of therapies suitable for many, if not all, patients. Common themes that we have identified that involve most of the known SCA and dystonia disease genes are synaptic transmission and development of the nervous system. Neither is easy to target on a general level since there is normally a perfect balance between excitation and inhibition, as well as an intricate brain organization, so any manipulation must be very careful.

In dystonic patients, it is possible to relieve symptoms in three ways: (1) locally by injecting botulinum toxin,[50,51] (2) more generally by oral medications, such as dopaminergic, anticholinergic, dopamine-depleting drugs, and muscle relaxants,[52] or (3) surgically by deep brain stimulation of the globus pallidus internus.[53] However, none of these therapies will cure the underlying pathologies, except in those dystonia types that are due to deficits in the dopaminergic synthesis pathways.

Alongside pharmacological therapies that are now being developed, some genetic therapies are being investigated. For example, in SCA1, RNAi therapies have been shown to reduce the level of mutant ATXN1 protein and this improves the phenotype in

mice.[54] Moreover, in SCA3, promising results are being obtained by skipping of the exon containing the CAG-repeat, in a similar manner to the treatment of Duchenne muscular dystrophy,[55–57] however, no results on the phenotype of the mice have been shown thus far.[55] For conventional mutations it is possible to use antisense oligonucleotides (AONs) to reduce mutant allele expression,[57] or CRISPR/Cas9 technology (see above) to correct the mutation.[42] Both these methods require patient-specific AONs or guidance RNAs for CRISPR/Cas9, demonstrating the need for personalized medicine strategies.[57] Furthermore, delivery of the AONs or guidance RNAs into the target region in the brain is complicated by the blood-brain barrier. Therefore, the most common route is to directly inject the vehicle into target regions of the brain, although this results in a focal effect.[58]

The loss of Purkinje cells in SCA hinders a cure for the patients since these cells will not be replaced, although loss of PCs might be more beneficial than dysfunctioning PCs.[59] Therefore, it would be useful to find a treatment that restores PC function and prevents further loss of PCs, and to start this treatment before clinical symptoms appear. Although this might be possible in families who know their genetic predisposition, in sporadic cases treatment will only be started after symptoms appear - and by then the damage has already been done. Even though it has been shown that human embryonic stem cells and induced pluripotent stem cells are able to differentiate into PCs *in vitro,*[60,61] and beneficial effects of stem cell transplantation have been seen in mice,[62] it is not yet possible to replace degenerated PCs in patients.

References

1. Zech, M., Lam, D.D., Francescatto, L., Schormair, B., Salminen, A. V, Jochim, A., Wieland, T., Lichtner, P., Peters, A., Gieger, C., et al. (2015). Recessive mutations in the α3 (VI) collagen gene COL6A3 cause early-onset isolated dystonia. Am. J. Hum. Genet. *96*, 883–893.
2. Fogel, B.L., Hanson, S.M., and Becker, E.B.E. (2015). Do mutations in the murine ataxia gene TRPC3 cause cerebellar ataxia in humans? Mov. Disord. *30*, 284–286.
3. Coutelier, M., Blesneac, I., Monteil, A., Monin, M., Ando, K., Mundwiller, E., Brusco, A., Le Ber, I., Anheim, M., Castrioto, A., et al. (2015). A recurrent mutation in CACNA1G alters Cav3.1 T-Type calcium-channel conduction and causes autosomal-dominant cerebellar ataxia. Am. J. Hum. Genet. *97*, 726–737.
4. Gilissen, C., Hoischen, A., Brunner, H.G., and Veltman, J.A. (2012). Disease gene identification strategies for exome sequencing. Eur. J. Hum. Genet. *20*, 490–497.
5. Meienberg, J., Bruggmann, R., Oexle, K., and Matyas, G. (2016). Clinical sequencing: is WGS the better WES? Hum. Genet. *135*, 359–362.
6. 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., et al. (2014). Genome sequencing identifies major causes of severe intellectual disability. Nature *511*, 344–347.
7. Vissers, L.E.L.M., de Ligt, J., Gilissen, C., Janssen, I., Steehouwer, M., de Vries, P., van Lier, B., Arts, P., Wieskamp, N., del Rosario, M., et al. (2010). A de novo paradigm for mental retardation. Nat. Genet. *42*, 1109–1112.
8. Consortium, R.E., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330.
9. Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C., and Snyder, M. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.
10. Sikkema-Raddatz, B., Johansson, L.F., de Boer, E.N., Almomani, R., Boven, L.G., van den Berg, M.P., van Spaendonck-Zwarts, K.Y., van Tintelen, J.P., Sijmons, R.H., Jongbloed, J.D.H., et al. (2013). Targeted next-generation sequencing can replace Sanger sequencing in clinical diagnostics. Hum. Mutat. *34*, 1035–1042.
11. Antoniadi, T., Buxton, C., Dennis, G., Forrester, N., Smith, D., Lunt, P., and Burton-Jones, S. (2015). Application of targeted multi-gene panel testing for the diagnosis of inherited peripheral neuropathy provides a high diagnostic yield with unexpected phenotype-genotype variability. BMC Med. Genet. *16,* 84.
12. Hildebrand, M.S., Myers, C.T., Carvill, G.L., Regan, B.M., Damiano, J.A., Mullen, S.A., Newton, M.R., Nair, U., Gazina, E. V, Milligan, C.J., et al. (2016). A targeted resequencing gene panel for focal epilepsy. Neurology *86*, 1605-1612..
13. Bamshad, M.J., Ng, S.B., Bigham, A.W., Tabor, H.K., Emond, M.J., Nickerson, D. A., and Shendure, J. (2011). Exome sequencing as a tool for Mendelian disease gene discovery. Nat. Rev. Genet. *12*, 745–755.
14. Kiezun, A., Garimella, K., Do, R., Stitziel, N.O., Neale, B.M., McLaren, P.J., Gupta, N., Sklar, P., Sullivan, P.F., Moran, J.L., et al. (2012). Exome sequencing and the genetic basis of complex traits. Nat. Genet. *44*, 623–630.

15. Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., and Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. *29*, 308–311.

16. Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., and McVean, G.A. (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061–1073.

17. NHLBI Exome Sequencing Project Exome Variant Server. http://evs.gs.washington.edu/EVS/

18. Exome Aggregation Consortium (ExAC) ExAC. http://exac.broadinstitute.org

19. Hakenberg, J., Cheng, W.-Y., Thomas, P., Wang, Y.-C., Uzilov, A. V., and Chen, R. (2016). Integrating 400 million variants from 80,000 human samples with extensive annotations: towards a knowledge base to analyze disease cohorts. BMC Bioinformatics *17*, 24.

20. Itan, Y., Shang, L., Boisson, B., Patin, E., Bolze, A., Moncada-Vélez, M., Scott, E., Ciancanelli, M.J., Lafaille, F.G., Markle, J.G., et al. (2015). The human gene damage index as a gene-level approach to prioritizing exome variants. Proc. Natl. Acad. Sci. U. S. A. *112*, 13615–13620.

21. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat. Methods *7*, 248–249.

22. Kumar, P., Henikoff, S., and Ng, P.C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. *4*, 1073–1081.

23. Reva, B., Antipin, Y., and Sander, C. (2011). Predicting the functional impact of protein mutations: Application to cancer genomics. Nucleic Acids Res. *39*, 37–43.

24. Jiang, R., Yang, H., Zhou, L., Kuo, C.-C.J., Sun, F., and Chen, T. (2007). Sequence-based prioritization of nonsynonymous single-nucleotide polymorphisms for the study of disease mutations. Am. J. Hum. Genet. *81*, 346–360.

25. Siepel, A., Pollard, K.S., and Haussler, D. (2006). Research in Computational Molecular Biology. Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) *3909*, 190–205.

26. Cooper, G.M., Stone, E.A., Asimenos, G., Green, E.D., Batzoglou, S., and Sidow, A. (2005). Distribution and intensity of constraint in mammalian genomic sequence. Genome Res. *15*, 901–913.

27. Garber, M., Guttman, M., Clamp, M., Zody, M.C., Friedman, N., and Xie, X. (2009). Identifying novel constrained elements by exploiting biased substitution patterns. Bioinformatics *25*, 54–62.

28. Lehmann, K. Van, and Chen, T. (2013). Exploring functional variant discovery in non-coding regions with SInBaD. Nucleic Acids Res. *41*, 1–10.

29. Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K., and Liu, X. (2015). Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum. Mol. Genet. *24*, 2125–2137.

30. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. *46*, 310–315.

31. Quang, D., Chen, Y., and Xie, X. (2015). DANN: A deep learning approach for annotating the pathogenicity of genetic variants. Bioinformatics *31*, 761–763.

32. Wu, M., Wu, J., Chen, T., and Jiang, R. (2015). Prioritization of nonsynonymous single nucleotide variants for exome sequencing studies via integrative learning on multiple genomic data. Sci. Rep. *5*, 14955.

33. Smedley, D., and Robinson, P.N. (2015). Phenotype-driven strategies for exome prioritization of human Mendelian disease genes. Genome Med. *7*, 81.

34. Robinson, P.N., Köhler, S., Oellrich, A., Genetics, S.M., Wang, K., Mungall, C.J., Lewis, S.E., Washington, N., Bauer, S., Seelow, D., et al. (2014). Improved exome prioritization of disease genes through cross-species phenotype comparison. Genome Res. *24*, 340–348.

35. James, R.A., Campbell, I.M., Chen, E.S., Boone, P.M., Rao, M.A., Bainbridge, M.N., Lupski, J.R., Yang, Y., Eng, C.M., Posey, J.E., et al. (2016). A visual and curatorial approach to clinical variant prioritization and disease gene discovery in genome-wide diagnostics. Genome Med. *8*, 13.

36. MacArthur, D.G., Manolio, T. a, Dimmock, D.P., Rehm, H.L., Shendure, J., Abecasis, G.R., Adams, D.R., Altman, R.B., Antonarakis, S.E., Ashley, E.A., et al. (2014). Guidelines for investigating causality of sequence variants in human disease. Nature *508*, 469–476.

37. Paudel, R., Hardy, J., Revesz, T., Holton, J.L., and Houlden, H. (2012). Review: genetics and neuropathology of primary pure dystonia. Neuropathol. Appl. Neurobiol. *38*, 520–534.

38. Argyelan, M., Carbon, M., Niethammer, M., Ulug, A.M., Voss, H.U., Bressman, S.B., Dhawan, V., and Eidelberg, D. (2009). Cerebellothalamocortical connectivity regulates penetrance in dystonia. J. Neurosci. *29*, 9740–9747.

39. Lohmann, K., Schmidt, A., Schillert, A., Winkler, S., Albanese, A., Baas, F., Bentivoglio, A.R., Borngräber, F., Brüggemann, N., Defazio, G., et al. (2013). Genome-wide association study in musician's dystonia: A risk variant at the arylsulfatase G locus? Mov. Disord. *29*, 921–927.

40. Sander, J.D., and Joung, J.K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. Nat. Biotechnol. *32*, 347–355.

41. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

42. Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell *157*, 1262–1278.

43. Peng, R., Lin, G., and Li, J. (2015). Potential pitfalls of CRISPR/Cas9-mediated genome editing. FEBS J. 1–14.

44. Duarri, A., Lin, M.A., Fokkens, M.R., Meijer, M., Smeets, C.J.L.M., Nibbeling, E.A.R., Boddeke, E., Sinke, R.J., Kampinga, H.H., Papazian, D.M., et al. (2015). Spinocerebellar ataxia type 19 / 22 mutations alter heterocomplex Kv4.3 channel function and gating in a dominant manner. Cell. Mol. Life Sci. *72*, 3387–3399.

45. Martins, S., Pearson, C.E., Coutinho, P., Provost, S., Amorim, A., Dubé, M.-P., Sequeiros, J., and Rouleau, G.A. (2014). Modifiers of (CAG)n instability in Machado-Joseph disease (MJD/SCA3) transmissions: an association study with DNA replication, repair and recombination genes. Hum. Genet. *133*, 1311–1318.

46. Kawaguchi, Y., Okamoto, T., Taniwaki, M., Aizawa, M., Inoue, M., Katayama, S., Kawakami, H., Nakamura, S., Nishimura, M., Akiguchi, I., et al. (1994). CAG expansions in a novel gene for Machado-Joseph Disease at chromosome 14q32.1. Nat. Genet. *8*, 221–228.

47. Van De Warrenburg, B.P.C., Hendriks, H., Dürr, A., Van Zuijlen, M.C.A., Stevanin, G., Camuzat, A., Sinke, R.J., Brice, A., and Kremer, B.P.H. (2005). Age at onset variance analysis in spinocerebellar ataxias: A study in a Dutch-French cohort. Ann. Neurol. *57*, 505–512.

48. Tezenas du Montcel, S., Durr, A., Bauer, P., Figueroa, K.P., Ichikawa, Y., Brussino, A., Forlani, S., Rakowicz, M., Schöls, L., Mariotti, C., et al. (2014). Modulation of the age at onset in spinocerebellar ataxia by CAG tracts in various genes. Brain *137*, 2444–2455.

49. Arning, L., and Epplen, J.T. (2013). Genetic modifiers in Huntington's disease: Fiction or fact? Neurogenetics *14*, 171–172.

50. Hallett, M., Albanese, A., Dressler, D., Segal, K.R., Simpson, D.M., Truong, D., and Jankovic, J. (2013). Evidence-based review and assessment of botulinum neurotoxin for the treatment of movement disorders. Toxicon *67*, 94–114.

51. Simpson, D.M., and Miyasaki, J.M. (2008). Assessment : Botulinum neurotoxin for the treatment of spasticity (an evidence-based review) Report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. Neurology *70*, 1691–1698.

52. Morgante, F., and Klein, C. (2013). Dystonia. Continuum (N. Y). *19*, 1225–1241.

53. Vidailhet, M., Jutras, M.-F., Grabli, D., and Roze, E. (2013). Deep brain stimulation for dystonia. J. Neurol. Neurosurg. Psychiatry *84*, 1029–1042.

54. Keiser, M.S., Boudreau, R.L., and Davidson, B.L. (2014). Broad therapeutic benefit after RNAi expression vector delivery to deep cerebellar nuclei: implications for spinocerebellar ataxia type 1 therapy. Mol. Ther. *22*, 588–595.

55. Evers, M.M., Tran, H.-D., Zalachoras, I., Pepers, B.A., Meijer, O.C., den Dunnen, J.T., van Ommen, G.-J.B., Aartsma-Rus, A., and van Roon-Mom, W.M.C. (2013). Ataxin-3 protein modification as a treatment strategy for spinocerebellar ataxia type 3: removal of the CAG containing exon. Neurobiol. Dis. *58*, 49–56.

56. van Putten, M., and Aartsma-Rus, A. (2011). Opportunities and challenges for the development of antisense treatment in neuromuscular disorders. Expert Opin. Biol. Ther. *11*, 1025–1037.

57. Siva, K., Covello, G., and Denti, M.A. (2014). Exon-skipping antisense oligonucleotides to correct missplicing in neurogenetic diseases. Nucleic Acid Ther. *24*, 69–86.

58. Maguire, C.A., Ramirez, S.H., Merkel, S.F., Sena-Esteves, M., and Breakefield, X.O. (2014). Gene therapy for the nervous system: challenges and new strategies. Neurotherapeutics *11*, 817–839.

59. Dell'Orco, J.M., Wasserman, A.H., Chopra, R., Ingram, M.A. C., Hu, Y.S., Singh, V., Wulff, H., Opal, P., Orr, H.T., and Shakkottai, V.G. (2015). Neuronal atrophy early in degenerative ataxia is a compensatory mechanism to regulate membrane excitability. J. Neurosci. *35*, 11292–11307.

60. Muguruma, K., Nishiyama, A., Kawakami, H., Hashimoto, K., and Sasai, Y. (2015). Self-organization of polarized cerebellar tissue in 3D culture of human pluripotent stem cells. Cell Rep. *10*, 537–550.

61. Morino, H., Matsuda, Y., Muguruma, K., Miyamoto, R., Ohsawa, R., Ohtake, T., Otobe, R., Watanabe, M., Maruyama, H., Hashimoto, K., et al. (2015). A mutation in the low voltage-gated calcium channel CACNA1G alters the physiological properties of the channel, causing spinocerebellar ataxia. Mol. Brain *8*, 89.

62. Cendelin, J. (2016). Transplantation and stem cell therapy for cerebellar degenerations. The Cerebellum *15*, 48–50.