Published in final edited form as:

Expert Rev Neurother. 2013 February ; 13(2): . doi:10.1586/ern.12.165.

Next generation sequencing in understanding complex neurological disease

Adam E. Handel, BA, BMBCh^{1,2} [clinical fellow], Giulio Disanto, MD^{1,2} [doctoral student], and Sreeram V. Ramagopalan, MA, DPhil^{1,2,**} [program leader]

¹Department of Physiology, Anatomy and Genetics and Medical Research Council Functional Genomics Unit, University of Oxford, South Parks Road, Oxford, OX1 3PT, United Kingdom

²Blizard Institute, Queen Mary University of London, Barts and The London School of Medicine and Dentistry, Newark Street, London, E1 2AT, United Kingdom

Abstract

Next generation sequencing techniques have made vast quantities of data on human genomes and transcriptomes available to researchers. Huge progress has been made towards understanding the basis of many Mendelian neurological conditions, but progress has been considerably slower in complex neurological diseases (multiple sclerosis, migraine, Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis *etc.*). We review current next generation sequencing methodologies and present selected studies illustrating how these have been used to cast light on the genetic *aetiology* of complex neurological diseases, with specific focus on multiple sclerosis. We highlight particular pitfalls in next generation sequencing experiments and speculate on both clinical and research applications of these sequencing platforms for complex neurological disorders in the future.

Keywords

Next generation sequencing; transcriptomics; functional genomics; complex disease; neurological disease

Genetics of complex neurological disease

Our understanding of the human genome in health and disease has changed drastically over the last few decades. In 2003 the first complete read on the human genome was obtained, and following this there has been an explosion in the number of individuals whose genome has been fully or partially sequenced [1–3]. This has led to breakthroughs in the genetic *aetiology* of many diseases, especially those with Mendelian inheritance patterns, as reviewed in [4]. Only now, though, are we beginning to bring the power of sequencing to bear on the far more difficult problem of complex diseases: those in which multiple genetic risk factors are likely to act in different individuals and in which genetic risk factors are modified by interactions with the environment. We review the role next generation sequencing methodologies have in the field of neurology for understanding complex disease, with particular reference to multiple sclerosis (MS).

^{**}Correspondence can be addressed to: Dr. Sreeram Ramagopalan, Blizard Institute, Queen Mary University of London, Barts and The London School of Medicine and Dentistry, 4 Newark Street, London, E1 2AT, United Kingdom Tel: 020 7377 7472, Fax: 020 7377 7033, s.ramagopalan@qmul.ac.uk.

Conflicts of interest The authors have no conflicts of interest to declare.

Whole genome or whole exome sequencing has now become routine, however the techniques employed to identify genetic risk factors in disease prior to this identified many important risk factors both in Mendelian and complex disease. Due to the inherent study design limitations, these were most frequently common variants (minor allele frequency (MAF) $\geq 5\%$) rather than rare variants (MAF $<3\%$).

Linkage analysis using single nucleotide polymorphisms (SNPs) or other common variants was effective at localising regions conferring risk in disease pedigrees, although were generally limited by poor resolution, implicating genomic regions containing tens or hundreds of possible genes [5]. Genome-wide association studies (GWAS) using hundreds of thousands of common SNPs to tag regions across the genome were able to compare the frequency of these SNPs between thousands of cases and thousands of controls. This study design has been extremely effective at identifying genetic risk *loci* (table 1 and table S1), but only some of these have a direct detectable effect on gene expression, although many more have other regulatory effects e.g. influencing transcription factor binding [6,7]. For the sample of SNPs included in table S1, only 7.3% are shown to alter gene expression, but 47.6% are predicted to alter transcription factor binding [8]. As some disease associated SNPs do not appear to have an as yet described function, this has often been interpreted to imply that the true causal variant is likely to be a nearby rare variant. This can either be because the causative variant is in linkage disequilibrium (LD) with the tagging SNP or by synthetic association, whereby one or possibly many rare variants are stochastically more likely to be present in a population if a particular variant of the SNP is also present (figure 1) [9]. Frequently targeted sequencing of the region immediately surrounding the implicated SNP was undertaken to search for a rare variant, and this strategy has been successful in a number of conditions [4,10]. Estimates of synthetic association suggest that rare variants can be located megabases away from the detected SNP and overly restrictive sequencing would fail to find such causative variants [9]. However, much of the synthetic association theory is based on genomic modelling and there is currently little direct evidence that this would account for many of the associations detected by GWAS [11].

Next generation sequencing techniques promise to circumvent a large number of these limitations. However, these sequencing methods generate problems of their own. Next generation sequencing methods have identified rare variants responsible for conditions inherited in a Mendelian pattern but the major complex neurological diseases (multiple sclerosis, migraine, Parkinson's disease, Alzheimer's disease and amyotrophic lateral sclerosis) have proven more difficult to unravel, as reviewed in [12].

Whole genome and whole exome sequencing

Next generation sequencing relies upon multiple, short, overlapping reads of fragmented DNA which can be aligned against a reference genome or assembled "de novo" in the case that no information on the reference genome is available (figure 2) [12]. It is far faster than the previously available Sanger-sequencing method but, due to the need for overlapping reads to allow fragments to be aligned, the depth of sequencing (number of reads per nucleotide position) required is increased. It is also less reliable in genomic regions with extensive nucleotide repeats.

Whole exome sequencing uses precisely the same method except that complementary strands to known exons (protein-coding regions of DNA) are used to extract fragments covering exonic regions prior to sequencing. This has the advantage of requiring less sequencing (and is therefore less costly and the analysis less computationally demanding) than whole genome sequencing [12]. However, it will, by definition, only capture variants within exons and so will miss any potential causative variants elsewhere in the genome. This

may be a very important limitation as current estimates suggest that the majority of variants altering gene expression are located outside of protein-coding regions [13]. PolyQ diseases (expansions of glutamine codons) and those associated with copy number variants, such as Alzheimer's and Parkinson's disease, are also difficult to study with next generation sequencing methods [14,15].

Sequencing experiments have produced a number of rare variants important in neurological conditions (table 2). In MS, such approaches have identified three rare variants of potential importance to disease susceptibility, two generated by whole exome sequencing and one by targeted resequencing [16–18]. These are variants in *SIAE*, *CYP27B1* and *TYK2*. SNP genotyping arrays such as the ImmunoChip offer a cost-effective way of replicating variants identified by whole genome or whole exome sequencing, as reviewed in [19].

The study that identified rare variants in *SIAE* used targeted re-sequencing of the *SIAE* gene based on work in mice which had implicated this gene as important for B-cell tolerance [16]. The authors found that functional rare variants across *SIAE* were significantly associated with MS (odds ratio (OR) 10.45 (95% confidence intervals (95% CI) 1.77 to 61.73), $p=0.02$). However, further direct genotyping of these rare variants did not support a role for *SIAE* in MS susceptibility [20].

Our group performed whole exome sequencing in a cohort of MS patients and identified functional rare variants within *CYP27B1* [17]. These affected 1- α -hydroxylation of vitamin D, deficiency of which is associated with MS susceptibility [21]. A rare variant in *TYK2* was also identified in an extended pedigree with an unusually high frequency of MS by whole exome sequencing [18]. This gene is a tyrosine kinase involved in modulating the function of multiple immune-related genes. Further study of *CYP27B1* and *TYK2* is warranted however, as ultimately appropriate power to confirm or refute rare variant associations will require tens of thousands of cases and controls.

A recent study undertook whole genome sequencing and RNA-seq (see below) in pairs of monozygotic twins discordant for MS [22]. The authors were unable to identify any differences in DNA or mRNA sequence between twins that would explain the development of MS in one twin but not the other. However, it is possible that the sequencing methods used and analysis of the data were sub-optimal, thus meaning that important risk variants may have been missed [23]. For example, whole genome sequencing studies in monozygotic twins discordant for schizophrenia at a greater depth did identify sequence differences between twins [24].

Whole exome sequencing experiments are not always able to identify rare variants common to all cases. One example of this was a recent study in idiopathic generalised epilepsy [25]. The authors identified multiple rare variants present hetero- or homozygously in cases but not controls. However, none of these variants individually reached statistical significance. This suggests that for some conditions underlying genetic heterogeneity may render finding unifying causal rare variants very difficult without enormous sample sizes.

A further difficulty in dealing with next generation sequencing data is deciding how to analyse multiple functional rare variants arising within a single gene. In Parkinson's disease, pooling all rare variants within a candidate gene (*GBA*) generated a large odds ratio, but future studies will likely use more sophisticated methods based on weighting by functional effect and collapsing rare variants over different regions of interest [26,27].

These examples show that next generation sequencing has the potential to identify causative variants in many complex neurological diseases. However, the chance for false positives is also high, given the vast quantities of sequencing data generated. For example, whole exome

sequencing in our cohort of MS patients identified more than 58,000 variants in each individual. Clearly not all of these will be causal, and integrating information from GWAS can help to pick plausible candidates. Certainly current techniques of *in silico* prediction of altered function from rare variants are not always reliable methods of selecting likely causal variants. Several tools exist that attempt to predict damaging variants (PolyPhen [28], SIFT [29], Condel [30] and MutationTaster [31]). The sensitivity and specificity of these for distinguishing causative rare variants from known benign variants are 80-90% and 75-80% respectively [30,32]. A recent review of this area did point out the example of *LRRK2* in Parkinson's disease as reason to interpret the output of these tools with caution: the causative variant was labelled as "benign" by PolyPhen and "tolerated" by SIFT [12,33,34]. *In vitro* studies can supplement bioinformatics risk prediction, but these methods are assay-dependent and may miss some functionally important variants. Potentially important variants that affect transcription factor binding may also be missed (and these will often be missed by whole exome sequencing in any case). Recent data from ENCODE show that many disease-associated SNPs occur in regions of transcription factor binding, thus suggesting that at least some of these signals may be from genetic variants disrupting normal patterns of transcription factor binding, as was suggested from a study of vitamin D receptor binding and autoimmune disease [35,36].

Replication is vital in confirming the role of potentially causative variants, especially as the accuracy of power calculation for next generation sequencing methodologies is less clear than for genotype-based studies [37]. Replication through genotyping is certainly possible and is not predicted to differ hugely from sequencing-based replication in large samples [38]. However, the frequency of rare variants and their importance to disease *aetiology* is likely to be very different in ethnically diverse cohorts (and, indeed, this would be expected to be the case if synthetic association is important in the genetic *aetiology* of complex disease [9]). As happened with GWAS, as large numbers of genomes are sequenced so the genetic architecture of complex neurological disease will become clearer.

Transcriptomics

Next generation sequencing methods are not restricted to sequencing genomic DNA. mRNA can also be sequenced using probes for polyadenylated mRNA and reverse transcriptase to generate cDNA which can then be sequenced using conventional next generation techniques (figure 2) [39]. Similar techniques can be used to analyse different types of RNA, for example non-coding RNA or whole RNA. Sequencing the transcriptome has the potential to identify splice variants that may be of importance in complex neurological disease, as well as providing a quantitative estimate of gene expression. Clearly, however, it is vital that the correct tissue is used since the transcriptome will differ markedly by cell type. This is particularly important when using whole genome or whole exome sequencing in parallel with transcriptomics to identify functional variants, since there is evidence that genetic effects on expression may be cell type-specific [40,41].

RNA-seq carried out on cerebellar samples from normal and autistic individuals have shown differences in adenosine-to-inosine post-transcriptional editing [42]. Thus transcriptomics offers a powerful tool to identify functional differences associated with complex neurological disease.

RNA-seq has been used to study gene expression in CD4+ cells from discordant monozygotic twins as mentioned earlier without much success [22]. No significant differences were found when compared by disease status but there were some interesting findings of allelic imbalance in gene expression. There are significant differences in mRNA

by MS phenotype [43]. The importance of the transcriptome for susceptibility remains to be uncovered as the timing of sampling and choice of cell type is vital in such studies [23].

Such studies will be made far easier in the future by the creation of transcriptomic maps of normal human brain tissue. These resources are being generated by applying RNA-seq to post-mortem samples from different developmental stages in humans [44].

Transcriptomic approaches to complex neurological disease have shown some evidence of success (table 3). The main difficulties lie in obtaining tissue, which often means autopsy samples must be studied, a generally poor surrogate for living tissue especially since RNA is extremely prone to degradation *ex vivo* [45]. There are also potentially serious issues in analysing RNA-seq data with the possibility of sequencing and alignment errors [46]. Another problem is that of distinguishing cause and effect, since transcriptomics in case-control studies is undertaken after the onset of disease.

Chromatin immunoprecipitation and next generation sequencing (ChIP-seq)

ChIP-seq is a method whereby a protein of interest bound to fragments of DNA is precipitated out using antibodies. The resultant DNA fragments are then sequenced and can be mapped back to the reference genome to provide a map of where that protein is binding (figure 2) [47]. This provides a way of using next generation sequencing techniques to provide data on functional genomics. One major limitation of this method is that, because polymerase chain reaction (PCR) amplification of DNA can only be undertaken after immunoprecipitation, a large number of cells are required to give adequate coverage in order to align ChIP-seq peaks back to the genome, although methods are being trialled to improve this.

MS is particularly amenable to study using ChIP-seq. This is because it is an autoimmune disorder and so peripheral immune cells are likely to cast light on disease mechanisms. Our group recently demonstrated that vitamin D receptor (VDR) binding in lymphoblastoid cell lines overlapped significantly with genomic regions implicated by GWAS (2.2-fold enrichment, $p=0.0003$) [36]. Vitamin D deficiency is a candidate risk factor in MS and so the observation that VDR binding overlaps with previously implicated genetic risk *loci* suggests the possibility of gene-environment interactions. Given the significant proportion of GWAS associated SNPs lying in transcription factor binding sites, ChIP-seq approaches for a number of different transcription factors in relevant tissues will aid in understanding the role that associated variants play in disease development.

As with transcriptomics, ChIP-seq requires samples of cells relevant to the underlying disease process, which generates problems when studying complex neurological conditions limited to involvement of the brain. Unlike other uses of next generation sequencing methods, it also requires a strong *a priori* hypothesis that determines the target of the antibody used in the immunoprecipitation step. Several neurological diseases with peripheral manifestations (e.g. stroke, migraine *etc.*) may well be suitable for study using this approach in the future [48,49] and the use of neuronal cell lines may generate useful genomic binding data as a surrogate for brain tissue, as has been done with α -synuclein binding in Parkinson's disease [50].

Next generation sequencing approaches to epigenetics

Epigenetics is the study of reversible modifications to the structure of DNA associated with gene expression differences. The most common modifications are histone acetylation or

methylation and DNA methylation. Epigenetic alterations have been suggested as the cause of several complex diseases, as reviewed in [51].

Next generation sequencing techniques provide powerful ways of interrogating the epigenome, as reviewed in [52]. ChIP-seq can be used to identify regions of the genome enriched for certain modified forms of chromatin.

Similarly, by using antibodies directed against 5-methylcytosine, methyl-DNA immunoprecipitation (Me-DIP) can be used to identify genomic regions with high levels of DNA methylation. DNA methylation can also be assessed by using methyl-DNA sensitive enzymes to produce DNA enriched for 5-methylcytosine (restricted representation bisulphite sequencing, RRBS) as well as whole genome bisulphite sequencing. In both of these methods, unmethylated cytosine residues are chemically modified to uracil by bisulphite treatment prior to sequencing on next generation sequencing platforms.

RRBS has been applied to the discordant monozygotic MS twins mentioned earlier [22]. No significant differences were found in DNA methylation, which was unexpected given differences seen in monozygotic twins at birth [53]. It is possible that the analytical methods used were not sufficiently powerful to show potentially interesting differences in methylation [23].

Epigenetic differences are likely to be cell- and time-dependent and so, just as with transcriptomics, the selection of tissue and the timing of sampling are vital. Similar concerns also exist over the issue of cause *versus* effect when these methods are applied to case-control studies.

Clinical relevance of next generation sequencing

The most direct relevance of next generation sequencing to clinical medicine is in molecular diagnostics. Considerable headway has already been made in this area and has identified the causative mutation in previously undiagnosed or clinically complex conditions, including neuroacanthocytosis, GM₁-gangliosidosis and Charcot-Marie-Tooth syndrome [54–56]. In some cases, there is tentative evidence that discovering the underlying biochemical pathway of a rare form of motor neuron disease may enable a subset of patients to be treated [57]. Next generation sequencing methodologies are highly likely to have a role in the future diagnosis of complex phenotypes [58].

A better understanding of the genetic architecture of human disease may enable more accurate risk prediction and genetic counselling. The public desire for this service has already been shown by the popularity of “self-test” DNA chips to generate risk profiles for common diseases from SNP genotyping [59]. However, the ethical issues inherent to this are far from trivial. It may be that ethical issues become more straightforward as whole exome and whole genome sequencing become more routine in clinical practice, as has happened with HIV testing [60,61]. However, it is important to bear in mind that HIV is a treatable condition whereas determinants of genetic risk have far broader implications to the individual and other family members.

Integrating whole exome and whole genome sequencing data from large numbers of patients with functional genomic approaches presents a potentially powerful approach to identifying important biochemical pathways in disease *aetiology*. This in turn generates both potential treatment targets and improved diagnostic tests. Although personalised medicine has not yet lived up to its full promise, next generation sequencing methodologies may allow treatments to be tailored to individual patients in a way that previous genetic techniques could not [62]. Complex neurological diseases, especially Alzheimer’s disease and amyotrophic lateral

sclerosis, are particularly important focuses for genomic approaches to identify treatment options, as existing treatments are woefully inadequate.

Conclusions

Next generation sequencing platforms offer powerful new tools to dissect out genetic contributions to disease *aetiology*. In addition to whole exome and whole genome sequencing, which have already been applied to complex neurological conditions with varying success, functional genomic methods, such as RNA-seq and ChIP-seq, present new ways of identifying disease-associated biochemical pathways. In tandem with these improved sequencing methods, new computational approaches are required to select likely causative variants and deal with the immense amounts of data generated by next generation sequencing techniques. These studies will yield important insights as to disease mechanisms with potential improvements in treatment and provide novel clinical applications for genetic diagnostics.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding statement This work was supported by the Medical Research Council.

References

1. Schmutz J, Wheeler J, Grimwood J, et al. Quality assessment of the human genome sequence. *Nature*. 2004; 429(6990):365–368. [PubMed: 15164052]
2. The human genome at ten. *Nature*. 2010; 464(7289):649–650. [PubMed: 20360688]
3. Mardis ER. A decade's perspective on DNA sequencing technology. *Nature*. 2011; 470(7333):198–203. [PubMed: 21307932]
4. Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat. Rev. Genet.* 2010; 11(6):415–425. [PubMed: 20479773]
5. Bailey-Wilson JE, Wilson AF. Linkage analysis in the next-generation sequencing era. *Hum. Hered.* 2011; 72(4):228–236. [PubMed: 22189465]
6. Hindorff LA, Sethupathy P, Junkins HA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U.S.A.* 2009; 106(23):9362–9367. [PubMed: 19474294]
7. Handel AE, Handunnetthi L, Berlanga AJ, Watson CT, Morahan JM, Ramagopalan SV. The effect of single nucleotide polymorphisms from genome wide association studies in multiple sclerosis on gene expression. *PLoS ONE*. 2010; 5(4):e10142. [PubMed: 20405052]
8. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 2012; 22(9):1790–1797. [PubMed: 22955989]
9. Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. Rare variants create synthetic genome-wide associations. *PLoS Biol.* 2010; 8(1):e1000294. [PubMed: 20126254]
10. Nejentsev S, Walker N, Riches D, Egholm M, Todd JA. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science*. 2009; 324(5925):387–389. [PubMed: 19264985]
11. Anderson CA, Soranzo N, Zeggini E, Barrett JC. Synthetic associations are unlikely to account for many common disease genome-wide association signals. *PLoS Biol.* 2011; 9(1):e1000580. [PubMed: 21267062]
12. Foo J-N, Liu J-J, Tan E-K. Whole-genome and whole-exome sequencing in neurological diseases. *Nat Rev Neurol.* 2012; 8(9):508–517. [PubMed: 22847385]

13. Vernot B, Stergachis AB, Maurano MT, et al. Personal and population genomics of human regulatory variation. *Genome Res.* 2012; 22(9):1689–1697. [PubMed: 22955981]
14. Hooli BV, Mohapatra G, Mattheisen M, et al. Role of common and rare APP DNA sequence variants in Alzheimer disease. *Neurology.* 2012; 78(16):1250–1257. [PubMed: 22491860]
15. Pankratz N, Dumitriu A, Hetrick KN, et al. Copy number variation in familial Parkinson disease. *PLoS ONE.* 2011; 6(8):e20988. [PubMed: 21829596]
16. Surolia I, Pirnie SP, Chellappa V, et al. Functionally defective germline variants of sialic acid acetyltransferase in autoimmunity. *Nature.* 2010; 466(7303):243–247. [PubMed: 20555325]
17. Ramagopalan SV, Dymment DA, Cader MZ, et al. Rare variants in the CYP27B1 gene are associated with multiple sclerosis. *Ann. Neurol.* 2011; 70(6):881–886. [PubMed: 22190362]
18. Dymment DA, Cader MZ, Chao MJ, et al. Exome sequencing identifies a novel multiple sclerosis susceptibility variant in the TYK2 gene. *Neurology.* 2012; 79(5):406–411. [PubMed: 22744673]
19. Cortes A, Brown MA. Promise and pitfalls of the Immunochip. *Arthritis Research & Therapy.* 2011; 13(1):101. [PubMed: 21345260]
20. Hunt KA, Smyth DJ, Balschun T, et al. Rare and functional SIAE variants are not associated with autoimmune disease risk in up to 66,924 individuals of European ancestry. *Nat. Genet.* 2012; 44(1):3–5. [PubMed: 22200769]
21. Munger KL, Levin LI, Hollis BW, Howard NS, Ascherio A. Serum 25-hydroxyvitamin D levels and risk of multiple sclerosis. *JAMA.* 2006; 296(23):2832–2838. [PubMed: 17179460]
22. Baranzini SE, Mudge J, Van Velkinburgh JC, et al. Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. *Nature.* 2010; 464(7293):1351–1356. [PubMed: 20428171]
23. Handunnetthi L, Handel AE, Ramagopalan SV. Contribution of genetic, epigenetic and transcriptomic differences to twin discordance in multiple sclerosis. *Expert Rev Neurother.* 2010; 10(9):1379–1381. [PubMed: 20819009]
24. Reumers J, De Rijk P, Zhao H, et al. Optimized filtering reduces the error rate in detecting genomic variants by short-read sequencing. *Nat. Biotechnol.* 2012; 30(1):61–68. [PubMed: 22178994]
25. Heinzen EL, Depondt C, Cavalleri GL, et al. Exome sequencing followed by large-scale genotyping fails to identify single rare variants of large effect in idiopathic generalized epilepsy. *Am. J. Hum. Genet.* 2012; 91(2):293–302. [PubMed: 22863189]
26. Sidransky E, Nalls MA, Aasly JO, et al. Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. *N. Engl. J. Med.* 2009; 361(17):1651–1661. [PubMed: 19846850]
27. Dering C, Hemmelmann C, Pugh E, Ziegler A. Statistical Analysis of Rare Sequence Variants: An Overview of Collapsing Methods. *Genet Epidemiol.* 2011; 35(Suppl 1):S12–S17. [PubMed: 22128052]
28. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat. Methods.* 2010; 7(4):248–249. [PubMed: 20354512]
29. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 2003; 31(13):3812–3814. [PubMed: 12824425]
30. González-Pérez A, López-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am. J. Hum. Genet.* 2011; 88(4):440–449. [PubMed: 21457909]
31. Schwarz JM, Rödelberger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods.* 2010; 7(8):575–576. [PubMed: 20676075]
32. Tchernitchko D, Goossens M, Wajcman H. In silico prediction of the deleterious effect of a mutation: proceed with caution in clinical genetics. *Clin. Chem.* 2004; 50(11):1974–1978. [PubMed: 15502081]
33. Healy DG, Falchi M, O'Sullivan SS, et al. Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case-control study. *Lancet Neurol.* 2008; 7(7):583–590. [PubMed: 18539534]
34. Kumari U, Tan EK. LRRK2 in Parkinson's disease: genetic and clinical studies from patients. *FEBS J.* 2009; 276(22):6455–6463. [PubMed: 19804413]

35. Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations with regulatory information in the human genome. *Genome Res.* 2012; 22(9):1748–1759. [PubMed: 22955986]
36. Ramagopalan SV, Heger A, Berlanga AJ, et al. A ChIP-seq defined genome-wide map of vitamin D receptor binding: associations with disease and evolution. *Genome Res.* 2010; 20(10):1352–1360. [PubMed: 20736230]
37. Sugaya Y, Akazawa Y, Saito A, Kamitsuji S. NDesign: software for study design for the detection of rare variants from next-generation sequencing data. *J. Hum. Genet.* 2012; 57(10):676–678. [PubMed: 22786579]
38. Liu DJ, Leal SM. Replication Strategies for Rare Variant Complex Trait Association Studies via Next-Generation Sequencing. *The American Journal of Human Genetics.* 2010; 87(6):790–801.
39. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 2009; 10(1):57–63. [PubMed: 19015660]
40. Fairfax BP, Makino S, Radhakrishnan J, et al. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat. Genet.* 2012; 44(5):502–510. [PubMed: 22446964]
41. Grundberg E, Small KS, Hedman ÅK, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat. Genet.* 2012; 44(10):1084–1089. [PubMed: 22941192]
42. Eran A, Li JB, Vatalaro K, et al. Comparative RNA editing in autistic and neurotypical cerebella. *Mol. Psychiatry.* 2012
43. Gandhi KS, McKay FC, Cox M, et al. The multiple sclerosis whole blood mRNA transcriptome and genetic associations indicate dysregulation of specific T cell pathways in pathogenesis. *Hum. Mol. Genet.* 2010; 19(11):2134–2143. [PubMed: 20190274]
44. Kang HJ, Kawasawa YI, Cheng F, et al. Spatio-temporal transcriptome of the human brain. *Nature.* 2011; 478(7370):483–489. [PubMed: 22031440]
45. Koppelkamm A, Vennemann B, Lutz-Bonengel S, Fracasso T, Vennemann M. RNA integrity in post-mortem samples: influencing parameters and implications on RT-qPCR assays. *Int. J. Legal Med.* 2011; 125(4):573–580. [PubMed: 21584656]
46. Pickrell JK, Gilad Y, Pritchard JK. Comment on “Widespread RNA and DNA sequence differences in the human transcriptome”. *Science.* 2012; 335(6074):1302. author reply 1302. [PubMed: 22422963]
47. Barski A, Zhao K. Genomic location analysis by ChIP-Seq. *J. Cell. Biochem.* 2009; 107(1):11–18. [PubMed: 19173299]
48. Denes A, Thornton P, Rothwell NJ, Allan SM. Inflammation and brain injury: acute cerebral ischaemia, peripheral and central inflammation. *Brain Behav. Immun.* 2010; 24(5):708–723. [PubMed: 19770034]
49. Hamed SA. The vascular risk associations with migraine: relation to migraine susceptibility and progression. *Atherosclerosis.* 2009; 205(1):15–22. [PubMed: 19054516]
50. Martins M, Rosa A, Guedes LC, et al. Convergence of miRNA expression profiling, -synuclein interacton and GWAS in Parkinson's disease. *PLoS ONE.* 2011; 6(10):e25443. [PubMed: 22003392]
51. Handel AE, Ebers GC, Ramagopalan SV. Epigenetics: molecular mechanisms and implications for disease. *Trends Mol Med.* 2010; 16(1):7–16. [PubMed: 20022812]
52. Ku CS, Naidoo N, Wu M, Soong R. Studying the epigenome using next generation sequencing. *J. Med. Genet.* 2011; 48(11):721–730. [PubMed: 21825079]
53. Gordon L, Joo JE, Powell JE, et al. Neonatal DNA methylation profile in human twins is specified by a complex interplay between intrauterine environmental and genetic factors, subject to tissue-specific influence. *Genome Res.* 2012; 22(8):1395–1406. [PubMed: 22800725]
54. Walker RH, Schulz VP, Tikhonova IR, et al. Genetic diagnosis of neuroacanthocytosis disorders using exome sequencing. *Mov. Disord.* 2012; 27(4):539–543. [PubMed: 22038564]
55. Pierson TM, Adams DA, Markello T, et al. Exome sequencing as a diagnostic tool in a case of undiagnosed juvenile-onset GM1-gangliosidosis. *Neurology.* 2012; 79(2):123–126. [PubMed: 22675082]

56. Montenegro G, Powell E, Huang J, et al. Exome sequencing allows for rapid gene identification in a Charcot-Marie-Tooth family. *Ann. Neurol.* 2011; 69(3):464–470. [PubMed: 21254193]
57. Johnson JO, Gibbs JR, Megarbane A, et al. Exome sequencing reveals riboflavin transporter mutations as a cause of motor neuron disease. *Brain.* 2012; 135(Pt 9):2875–2882. [PubMed: 22740598]
58. Need AC, Shashi V, Hitomi Y, et al. Clinical application of exome sequencing in undiagnosed genetic conditions. *J. Med. Genet.* 2012; 49(6):353–361. [PubMed: 22581936]
59. Samuel GN, Jordens CFC, Kerridge I. Direct-to-consumer personal genome testing: ethical and regulatory issues that arise from wanting to “know” your DNA. *Intern Med J.* 2010; 40(3):220–224. [PubMed: 20446967]
60. Maman S, King E. Changes in HIV testing policies and the implications for women. *J Midwifery Womens Health.* 2008; 53(3):195–201. [PubMed: 18455093]
61. Burke RC, Sepkowitz KA, Bernstein KT, et al. Why don't physicians test for HIV? A review of the US literature. *AIDS.* 2007; 21(12):1617–1624. [PubMed: 17630557]
62. Altman RB, Kroemer HK, McCarty CA, Ratain MJ, Roden D. Pharmacogenomics: will the promise be fulfilled? *Nat. Rev. Genet.* 2011; 12(1):69–73. [PubMed: 21116304]
63. Seshadri S, Fitzpatrick AL, Ikram MA, et al. Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA.* 2010; 303(18):1832–1840. [PubMed: 20460622]
64. Hollingworth P, Harold D, Sims R, et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat. Genet.* 2011; 43(5):429–435. [PubMed: 21460840]
65. Lambert J-C, Grenier-Boley B, Harold D, et al. Genome-wide haplotype association study identifies the FRMD4A gene as a risk locus for Alzheimer's disease. *Mol. Psychiatry.* 2012
66. Harold D, Abraham R, Hollingworth P, et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat. Genet.* 2009; 41(10):1088–1093. [PubMed: 19734902]
67. Kamboh MI, Barmada MM, Demirci FY, et al. Genome-wide association analysis of age at-onset in Alzheimer's disease. *Mol. Psychiatry.* 2012; 17(12):1340–1346. [PubMed: 22005931]
68. Naj AC, Beecham GW, Martin ER, et al. Dementia revealed: novel chromosome 6 locus for late-onset Alzheimer disease provides genetic evidence for folate-pathway abnormalities. *PLoS Genet.* 2010; 6(9)
69. Naj AC, Jun G, Beecham GW, et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat. Genet.* 2011; 43(5):436–441. [PubMed: 21460841]
70. Reiman EM, Webster JA, Myers AJ, et al. GAB2 alleles modify Alzheimer's risk in APOE epsilon4 carriers. *Neuron.* 2007; 54(5):713–720. [PubMed: 17553421]
71. Van Es MA, Veldink JH, Saris CGJ, et al. Genome-wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. *Nat. Genet.* 2009; 41(10):1083–1087. [PubMed: 19734901]
72. Shatunov A, Mok K, Newhouse S, et al. Chromosome 9p21 in sporadic amyotrophic lateral sclerosis in the UK and seven other countries: a genome-wide association study. *Lancet Neurol.* 2010; 9(10):986–994. [PubMed: 20801717]
73. Laaksovirta H, Peuralinna T, Schymick JC, et al. Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: a genome-wide association study. *Lancet Neurol.* 2010; 9(10):978–985. [PubMed: 20801718]
74. Landers JE, Melki J, Meininger V, et al. Reduced expression of the Kinesin-Associated Protein 3 (KIFAP3) gene increases survival in sporadic amyotrophic lateral sclerosis. *Proc. Natl. Acad. Sci. U.S.A.* 2009; 106(22):9004–9009. [PubMed: 19451621]
75. Van Es MA, Van Vught PWJ, Blauw HM, et al. Genetic variation in DPP6 is associated with susceptibility to amyotrophic lateral sclerosis. *Nat. Genet.* 2008; 40(1):29–31. [PubMed: 18084291]
76. Wang K, Zhang H, Ma D, et al. Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature.* 2009; 459(7246):528–533. [PubMed: 19404256]

77. Anney R, Klei L, Pinto D, et al. A genome-wide scan for common alleles affecting risk for autism. *Hum. Mol. Genet.* 2010; 19(20):4072–4082. [PubMed: 20663923]
78. Mead S, Poulter M, Uphill J, et al. Genetic risk factors for variant Creutzfeldt-Jakob disease: a genome-wide association study. *Lancet Neurol.* 2009; 8(1):57–66. [PubMed: 19081515]
79. Sanchez-Juan P, Bishop MT, Aulchenko YS, et al. Genome-wide study links MTMR7 gene to variant Creutzfeldt-Jakob risk. *Neurobiol. Aging.* 2012; 33(7):1487.e21–28. [PubMed: 22137330]
80. Guo Y, Baum LW, Sham PC, et al. Two-stage genome-wide association study identifies variants in CAMSAP1L1 as susceptibility loci for epilepsy in Chinese. *Hum. Mol. Genet.* 2012; 21(5):1184–1189. [PubMed: 22116939]
81. Stefansson H, Steinberg S, Petursson H, et al. Variant in the sequence of the LINGO1 gene confers risk of essential tremor. *Nat. Genet.* 2009; 41(3):277–279. [PubMed: 19182806]
82. Thier S, Lorenz D, Nothnagel M, et al. Polymorphisms in the glial glutamate transporter SLC1A2 are associated with essential tremor. *Neurology.* 2012; 79(3):243–248. [PubMed: 22764253]
83. Sanson M, Hosking FJ, Shete S, et al. Chromosome 7p11.2 (EGFR) variation influences glioma risk. *Hum. Mol. Genet.* 2011; 20(14):2897–2904. [PubMed: 21531791]
84. Shete S, Hosking FJ, Robertson LB, et al. Genome-wide association study identifies five susceptibility loci for glioma. *Nat. Genet.* 2009; 41(8):899–904. [PubMed: 19578367]
85. Wrensch M, Jenkins RB, Chang JS, et al. Variants in the CDKN2B and RTEL1 regions are associated with high-grade glioma susceptibility. *Nat. Genet.* 2009; 41(8):905–908. [PubMed: 19578366]
86. Dobbins SE, Broderick P, Melin B, et al. Common variation at 10p12.31 near MLLT10 influences meningioma risk. *Nat. Genet.* 2011; 43(9):825–827. [PubMed: 21804547]
87. Freilinger T, Anttila V, De Vries B, et al. Genome-wide association analysis identifies susceptibility loci for migraine without aura. *Nat. Genet.* 2012; 44(7):777–782. [PubMed: 22683712]
88. Anttila V, Stefansson H, Kallela M, et al. Genome-wide association study of migraine implicates a common susceptibility variant on 8q22.1. *Nat. Genet.* 2010; 42(10):869–873. [PubMed: 20802479]
89. Chasman DI, Schürks M, Anttila V, et al. Genome-wide association study reveals three susceptibility loci for common migraine in the general population. *Nat. Genet.* 2011; 43(7):695–698. [PubMed: 21666692]
90. De Jager PL, Jia X, Wang J, et al. Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. *Nat. Genet.* 2009; 41(7):776–782. [PubMed: 19525953]
91. Hafler DA, Compston A, Sawcer S, et al. Risk alleles for multiple sclerosis identified by a genomewide study. *N. Engl. J. Med.* 2007; 357(9):851–862. [PubMed: 17660530]
92. Sawcer S, Hellenthal G, Pirinen M, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature.* 2011; 476(7359):214–219. [PubMed: 21833088]
93. Nischwitz S, Cepok S, Kroner A, et al. Evidence for VAV2 and ZNF433 as susceptibility genes for multiple sclerosis. *J. Neuroimmunol.* 2010; 227(1-2):162–166. [PubMed: 20598377]
94. Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. *Nat. Genet.* 2009; 41(7):824–828. [PubMed: 19525955]
95. Sanna S, Pitzalis M, Zoledziewska M, et al. Variants within the immunoregulatory CBLB gene are associated with multiple sclerosis. *Nat. Genet.* 2010; 42(6):495–497. [PubMed: 20453840]
96. Aulchenko YS, Hoppenbrouwers IA, Ramagopalan SV, et al. Genetic variation in the KIF1B locus influences susceptibility to multiple sclerosis. *Nat. Genet.* 2008; 40(12):1402–1403. [PubMed: 18997785]
97. Jakkula E, Leppä V, Sulonen A-M, et al. Genome-wide association study in a high-risk isolate for multiple sclerosis reveals associated variants in STAT3 gene. *Am. J. Hum. Genet.* 2010; 86(2):285–291. [PubMed: 20159113]
98. Wang JH, Pappas D, De Jager PL, et al. Modeling the cumulative genetic risk for multiple sclerosis from genome-wide association data. *Genome Med.* 2011; 3(1):3. [PubMed: 21244703]

99. Matesanz F, González-Pérez A, Lucas M, et al. Genome-wide association study of multiple sclerosis confirms a novel locus at 5p13.1. *PLoS ONE*. 2012; 7(5):e36140. [PubMed: 22570697]
100. Patsopoulos NA, Esposito F, Reischl J, et al. Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. *Ann. Neurol*. 2011; 70(6):897–912. [PubMed: 22190364]
101. Hallmayer J, Faraco J, Lin L, et al. Narcolepsy is strongly associated with the T-cell receptor alpha locus. *Nat. Genet*. 2009; 41(6):708–711. [PubMed: 19412176]
102. Hor H, Kutalik Z, Dauvilliers Y, et al. Genome-wide association study identifies new HLA class II haplotypes strongly protective against narcolepsy. *Nat. Genet*. 2010; 42(9):786–789. [PubMed: 20711174]
103. Miyagawa T, Kawashima M, Nishida N, et al. Variant between CPT1B and CHKB associated with susceptibility to narcolepsy. *Nat. Genet*. 2008; 40(11):1324–1328. [PubMed: 18820697]
104. Wang K, Diskin SJ, Zhang H, et al. Integrative genomics identifies LMO1 as a neuroblastoma oncogene. *Nature*. 2011; 469(7329):216–220. [PubMed: 21124317]
105. Capasso M, Devoto M, Hou C, et al. Common variations in BARD1 influence susceptibility to high-risk neuroblastoma. *Nat. Genet*. 2009; 41(6):718–723. [PubMed: 19412175]
106. Lill CM, Roehr JT, McQueen MB, et al. Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: The PDGene database. *PLoS Genet*. 2012; 8(3):e1002548. [PubMed: 22438815]
107. Do CB, Tung JY, Dorfman E, et al. Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease. *PLoS Genet*. 2011; 7(6):e1002141. [PubMed: 21738487]
108. Pankratz N, Beecham GW, DeStefano AL, et al. Meta-analysis of Parkinson's disease: identification of a novel locus, RIT2. *Ann. Neurol*. 2012; 71(3):370–384. [PubMed: 22451204]
109. Nalls MA, Plagnol V, Hernandez DG, et al. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet*. 2011; 377(9766):641–649. [PubMed: 21292315]
110. Simón-Sánchez J, Schulte C, Bras JM, et al. Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat. Genet*. 2009; 41(12):1308–1312. [PubMed: 19915575]
111. Satake W, Nakabayashi Y, Mizuta I, et al. Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nat. Genet*. 2009; 41(12):1303–1307. [PubMed: 19915576]
112. Hamza TH, Zabetian CP, Tenesa A, et al. Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. *Nat. Genet*. 2010; 42(9):781–785. [PubMed: 20711177]
113. Edwards TL, Scott WK, Almonte C, et al. Genome-wide association study confirms SNPs in SNCA and the MAPT region as common risk factors for Parkinson disease. *Ann. Hum. Genet*. 2010; 74(2):97–109. [PubMed: 20070850]
114. Höglinger GU, Melhem NM, Dickson DW, et al. Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. *Nat. Genet*. 2011; 43(7):699–705. [PubMed: 21685912]
115. Winkelmann J, Czamara D, Schormair B, et al. Genome-wide association study identifies novel restless legs syndrome susceptibility loci on 2p14 and 16q12.1. *PLoS Genet*. 2011; 7(7):e1002171. [PubMed: 21779176]
116. Schormair B, Kemlink D, Roeske D, et al. PTPRD (protein tyrosine phosphatase receptor type delta) is associated with restless legs syndrome. *Nat. Genet*. 2008; 40(8):946–948. [PubMed: 18660810]
117. Bellenguez C, Bevan S, Gschwendtner A, et al. Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. *Nat. Genet*. 2012; 44(3):328–333. [PubMed: 22306652]
118. Ikram MA, Seshadri S, Bis JC, et al. Genomewide association studies of stroke. *N. Engl. J. Med*. 2009; 360(17):1718–1728. [PubMed: 19369658]
119. Gretarsdottir S, Thorleifsson G, Manolescu A, et al. Risk variants for atrial fibrillation on chromosome 4q25 associate with ischemic stroke. *Ann. Neurol*. 2008; 64(4):402–409. [PubMed: 18991354]

120. Blumkin L, Kivity S, Lev D, et al. A compound heterozygous missense mutation and a large deletion in the KCTD7 gene presenting as an opsoclonus-myoclonus ataxia-like syndrome. *J. Neurol.* 2012; 259(12):2590–2598. [PubMed: 22638565]
121. Xiao J, Uitti RJ, Zhao Y, et al. Mutations in CIZ1 cause adult onset primary cervical dystonia. *Ann. Neurol.* 2012; 71(4):458–469. [PubMed: 22447717]
122. Rosewich H, Thiele H, Ohlenbusch A, et al. Heterozygous de-novo mutations in ATP1A3 in patients with alternating hemiplegia of childhood: a whole-exome sequencing gene-identification study. *Lancet Neurol.* 2012; 11(9):764–773. [PubMed: 22850527]
123. Heinzen EL, Swoboda KJ, Hitomi Y, et al. De novo mutations in ATP1A3 cause alternating hemiplegia of childhood. *Nat. Genet.* 2012; 44(9):1030–1034. [PubMed: 22842232]
124. Iossifov I, Ronemus M, Levy D, et al. De novo gene disruptions in children on the autistic spectrum. *Neuron.* 2012; 74(2):285–299. [PubMed: 22542183]
125. Neale BM, Kou Y, Liu L, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature.* 2012; 485(7397):242–245. [PubMed: 22495311]
126. Chahrour MH, Yu TW, Lim ET, et al. Whole-exome sequencing and homozygosity analysis implicate depolarization-regulated neuronal genes in autism. *PLoS Genet.* 2012; 8(4):e1002635. [PubMed: 22511880]
127. Bi C, Wu J, Jiang T, et al. Mutations of ANK3 identified by exome sequencing are associated with Autism susceptibility. *Hum. Mutat.* 2012; 33(12):1635–1638. [PubMed: 22865819]
128. Sanders SJ, Murtha MT, Gupta AR, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature.* 2012; 485(7397):237–241. [PubMed: 22495306]
129. Lee H, Graham JM Jr, Rimoin DL, et al. Exome sequencing identifies PDE4D mutations in acrodyostosis. *Am. J. Hum. Genet.* 2012; 90(4):746–751. [PubMed: 22464252]
130. Winkelmann J, Lin L, Schormair B, et al. Mutations in DNMT1 cause autosomal dominant cerebellar ataxia, deafness and narcolepsy. *Hum. Mol. Genet.* 2012; 21(10):2205–2210. [PubMed: 22328086]
131. Pottier C, Hannequin D, Coutant S, et al. High frequency of potentially pathogenic SORL1 mutations in autosomal dominant early-onset Alzheimer disease. *Mol. Psychiatry.* 2012; 17(9):875–879. [PubMed: 22472873]
132. Ishiura H, Sako W, Yoshida M, et al. The TRK-Fused Gene Is Mutated in Hereditary Motor and Sensory Neuropathy with Proximal Dominant Involvement. *Am. J. Hum. Genet.* 2012; 91(2):320–329. [PubMed: 22883144]
133. Harms MB, Sommerville RB, Allred P, et al. Exome sequencing reveals DNAJB6 mutations in dominantly-inherited myopathy. *Ann. Neurol.* 2012; 71(3):407–416. [PubMed: 22334415]
134. Staropoli JF, Karaa A, Lim ET, et al. A homozygous mutation in KCTD7 links neuronal ceroid lipofuscinosis to the ubiquitin-proteasome system. *Am. J. Hum. Genet.* 2012; 91(1):202–208. [PubMed: 22748208]
135. Sailer A, Scholz SW, Gibbs JR, et al. Exome sequencing in an SCA14 family demonstrates its utility in diagnosing heterogeneous diseases. *Neurology.* 2012; 79(2):127–131. [PubMed: 22675081]
136. Li M, Pang S, Song Y, Kung M, Ho S-L, Sham P-C. Whole exome sequencing identifies a novel mutation in the transglutaminase 6 gene for spinocerebellar ataxia in a Chinese family. *Clin. Genet.* 2012
137. Delmaghani S, Aghaie A, Michalski N, Bonnet C, Weil D, Petit C. Defect in the gene encoding the EAR/EPTP domain-containing protein TSPEAR causes DFNB98 profound deafness. *Hum. Mol. Genet.* 2012; 21(17):3835–3844. [PubMed: 22678063]
138. Jimenez-Escrig A, Gobernado I, Garcia-Villanueva M, Sanchez-Herranz A. Autosomal recessive Emery-Dreifuss muscular dystrophy caused by a novel mutation (R225Q) in the lamin A/C gene identified by exome sequencing. *Muscle Nerve.* 2012; 45(4):605–610. [PubMed: 22431096]
139. Dündar H, Özgül RK, Yalınzı lu D, et al. Identification of a novel Twinkle mutation in a family with infantile onset spinocerebellar ataxia by whole exome sequencing. *Pediatr. Neurol.* 2012; 46(3):172–177. [PubMed: 22353293]

140. Schossig A, Wolf NI, Fischer C, et al. Mutations in ROGDI Cause Kohlschütter-Tönz Syndrome. *Am. J. Hum. Genet.* 2012; 90(4):701–707. [PubMed: 22424600]
141. Raja Rayan DL, Haworth A, Sud R, et al. A new explanation for recessive myotonia congenita: exon deletions and duplications in CLCN1. *Neurology.* 2012; 78(24):1953–1958. [PubMed: 22649220]
142. Wan J, Yourshaw M, Mamsa H, et al. Mutations in the RNA exosome component gene EXOSC3 cause pontocerebellar hypoplasia and spinal motor neuron degeneration. *Nat. Genet.* 2012; 44(6):704–708. [PubMed: 22544365]
143. Zhou J, Tawk M, Tiziano FD, et al. Spinal Muscular Atrophy Associated with Progressive Myoclonic Epilepsy Is Caused by Mutations in ASAH1. *Am. J. Hum. Genet.* 2012; 91(1):5–14. [PubMed: 22703880]
144. Choi B-O, Koo SK, Park M-H, et al. Exome sequencing is an efficient tool for genetic screening of Charcot-Marie-Tooth disease. *Hum. Mutat.* 2012; 33(11):1610–1615. [PubMed: 22730194]
145. Timal S, Hoischen A, Lehle L, et al. Gene identification in the congenital disorders of glycosylation type I by whole-exome sequencing. *Hum. Mol. Genet.* 2012; 21(19):4151–4161. [PubMed: 22492991]
146. Beetz C, Pieber TR, Hertel N, et al. Exome sequencing identifies a REEP1 mutation involved in distal hereditary motor neuropathy type V. *Am. J. Hum. Genet.* 2012; 91(1):139–145. [PubMed: 22703882]
147. Merner ND, Girard SL, Catoire H, et al. Exome sequencing identifies FUS mutations as a cause of essential tremor. *Am. J. Hum. Genet.* 2012; 91(2):313–319. [PubMed: 22863194]
148. Wu C-H, Fallini C, Ticozzi N, et al. Mutations in the profilin 1 gene cause familial amyotrophic lateral sclerosis. *Nature.* 2012; 488(7412):499–503. [PubMed: 22801503]
149. Herdewyn S, Zhao H, Moisse M, et al. Whole-genome sequencing reveals a coding non pathogenic variant tagging a non-coding pathogenic hexanucleotide repeat expansion in C9orf72 as cause of amyotrophic lateral sclerosis. *Hum. Mol. Genet.* 2012; 21(11):2412–2419. [PubMed: 22343411]
150. Williams KL, Warraich ST, Yang S, et al. UBQLN2/ubiquilin 2 mutation and pathology in familial amyotrophic lateral sclerosis. *Neurobiol. Aging.* 2012; 33(10):2527.e3–10. [PubMed: 22717235]
151. Wu J, Shen E, Shi D, Sun Z, Cai T. Identification of a novel Cys146X mutation of SOD1 in familial amyotrophic lateral sclerosis by whole-exome sequencing. *Genet. Med.* 2012
152. Elo JM, Yadavalli SS, Euro L, et al. Mitochondrial phenylalanyl-tRNA synthetase mutations underlie fatal infantile Alpers encephalopathy. *Hum. Mol. Genet.* 2012; 21(20):4521–4529. [PubMed: 22833457]
153. Lee JH, Huynh M, Silhavy JL, et al. De novo somatic mutations in components of the PI3K-AKT3-mTOR pathway cause hemimegalencephaly. *Nat. Genet.* 2012; 44(8):941–945. [PubMed: 22729223]
154. Edvardson S, Cinnamon Y, J alas C, et al. Hereditary sensory autonomic neuropathy caused by a mutation in dystonin. *Ann. Neurol.* 2012; 71(4):569–572. [PubMed: 22522446]
155. Veeramah KR, O'Brien JE, Meisler MH, et al. De novo pathogenic SCN8A mutation identified by whole-genome sequencing of a family quartet affected by infantile epileptic encephalopathy and SUDEP. *Am. J. Hum. Genet.* 2012; 90(3):502–510. [PubMed: 22365152]
156. Berger I, Dor T, Halvardson J, et al. Intractable epilepsy of infancy due to homozygous mutation in the EFHC1 gene. *Epilepsia.* 2012; 53(8):1436–1440. [PubMed: 22690745]
157. Cole JW, Stine OC, Liu X, et al. Rare variants in ischemic stroke: an exome pilot study. *PLoS ONE.* 2012; 7(4):e35591. [PubMed: 22536414]
158. Edvardson S, Cinnamon Y, Ta-Shma A, et al. A deleterious mutation in DNAJC6 encoding the neuronal-specific clathrin-uncoating co-chaperone auxilin, is associated with juvenile parkinsonism. *PLoS ONE.* 2012; 7(5):e36458. [PubMed: 22563501]
159. Steenweg ME, Ghezzi D, Haack T, et al. Leukoencephalopathy with thalamus and brainstem involvement and high lactate “LTBL” caused by EARS2 mutations. *Brain.* 2012; 135(Pt 5):1387–1394. [PubMed: 22492562]

160. Horvath R, Holinski-Feder E, Neeve VCM, et al. A new phenotype of brain iron accumulation with dystonia, optic atrophy, and peripheral neuropathy. *Mov. Disord.* 2012; 27(6):789–793. [PubMed: 22508347]
161. Bras J, Verloes A, Schneider SA, Mole SE, Guerreiro RJ. Mutation of the parkinsonism gene ATP13A2 causes neuronal ceroid-lipofuscinosis. *Hum. Mol. Genet.* 2012; 21(12):2646–2650. [PubMed: 22388936]
162. Saito H, Kato M, Osaka H, et al. CASK aberrations in male patients with Ohtahara syndrome and cerebellar hypoplasia. *Epilepsia.* 2012; 53(8):1441–1449. [PubMed: 22709267]
163. Andrade DM, Paton T, Turnbull J, Marshall CR, Scherer SW, Minassian BA. Mutation of the CLN6 Gene in Teenage-Onset Progressive Myoclonus Epilepsy. *Pediatr. Neurol.* 2012; 47(3): 205–208. [PubMed: 22883287]
164. Pyle A, Griffin H, Yu-Wai-Man P, et al. Prominent Sensorimotor Neuropathy Due to SACS Mutations Revealed by Whole-Exome Sequencing. *Sensorimotor Neuropathy Due to SACS Mutations. Archives of neurology.* 2012; 1–4.
165. Twine NA, Janitz K, Wilkins MR, Janitz M. Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease. *PLoS ONE.* 2011; 6(1):e16266. [PubMed: 21283692]
166. Hazelett DJ, Chang J-C, Lakeland DL, Morton DB. Comparison of Parallel High-Throughput RNA Sequencing Between Knockout of TDP-43 and Its Overexpression Reveals Primarily Nonreciprocal and Nonoverlapping Gene Expression Changes in the Central Nervous System of *Drosophila*. *G3 (Bethesda).* 2012; 2(7):789–802. [PubMed: 22870402]
167. Fogel BL, Wexler E, Wahnich A, et al. RBFOX1 regulates both splicing and transcriptional networks in human neuronal development. *Hum. Mol. Genet.* 2012; 21(19):4171–4186. [PubMed: 22730494]
168. Forman OP, De Risio L, Stewart J, Mellersh CS, Beltran E. Genome-wide mRNA sequencing of a single canine cerebellar cortical degeneration case leads to the identification of a disease associated SPTBN2 mutation. *BMC Genet.* 2012; 13:55. [PubMed: 22781464]
169. Kriesel JD, Hobbs MR, Jones BB, Milash B, Nagra RM, Fischer KF. Deep sequencing for the detection of virus-like sequences in the brains of patients with multiple sclerosis: detection of GBV-C in human brain. *PLoS ONE.* 2012; 7(3):e31886. [PubMed: 22412845]
170. Heap GA, Yang JHM, Downes K, et al. Genome-wide analysis of allelic expression imbalance in human primary cells by high-throughput transcriptome resequencing. *Hum. Mol. Genet.* 2010; 19(1):122–134. [PubMed: 19825846]
171. Lin M, Pedrosa E, Shah A, et al. RNA-Seq of human neurons derived from iPS cells reveals candidate long non-coding RNAs involved in neurogenesis and neuropsychiatric disorders. *PLoS ONE.* 2011; 6(9):e23356. [PubMed: 21915259]

Expert commentary

Genome-wide association studies have made great inroads into understanding the genetic architecture of complex neurological diseases. However, a large proportion of the genetic *aetiology* of these disorders remains unexplained. Next generation sequencing methods have made it possible to generate vast amounts of genetic sequence data in cases and control individuals. Already whole genome and whole exome sequencing have identified a multitude of rare variants important in Mendelian conditions, although to provide similar insights into complex neurological diseases is likely to require the sequencing of many more individual genomes. This outpouring of novel genetic data requires leaps forward in statistical analysis of such data, a movement which has started but still has much work to do. Next generation sequencing methods are being combined with functional genomics: mRNA expression, epigenetics and nuclear factor binding. This combinatorial approach to unravelling the genetic basis of complex neurological disease is likely to lead to huge advances in our understanding, with applications for diagnostics and potentially novel treatments.

5-year view

Next generation sequencing methods are likely to advance to encompass single strand DNA sequencing, requiring far less input DNA and thus allowing single cell sequencing. Functional genomics will benefit in a similar way and it is likely that the integration of sequence, expression and epigenomic data in a very small number of well characterised cells will provide unprecedentedly detailed views of disease *aetiology* and pathophysiology. Statistical genomics will improve to allow quicker and more powerful analysis of next generation sequencing data. It is very likely that our understanding of functional genomics will change radically with the application of these methods, with similar advances in the understanding of cellular biology and genetics in health and neurological disease. These advances are likely to have applications initially in diagnostics and later in therapeutics.

Key issues

- Next generation sequencing allows rapid generation of whole genome and whole exome data from healthy and diseased individuals.
- These methods have distinct drawbacks, with particularly reduced accuracy in regions of repeat DNA.
- Statistical approaches to next generation sequencing data are still being developed but are likely to advance rapidly over the forthcoming years.
- Combining next generation sequencing methods with functional genomics enables detailed views of mRNA expression, epigenomics and nuclear receptor binding.
- Extending next generation sequencing methods to complex neurological disease will require the sequencing of many more individuals than was required to understand Mendelian diseases.
- Integration of multiple next generation sequencing methods is likely to provide advances in understanding the cellular biology of complex neurological diseases, diagnosis and potentially treatment.

Figure 1. Synthetic association

A rare, disease-causing variant (orange) is located downstream of a tagging single nucleotide polymorphism (SNP) present in two alleles (red and green). Stochastically, the rare variant is located in individuals with the second (green) allele more frequently than the first (red). Therefore, genome-wide association studies will pick up the signal from the rare variant as an association with the second allele (green).

Figure 2. Next generation sequencing methodologies

This is a schematic illustrating next generation sequencing tools. RNA-seq uses poly-A tags to extract mRNA from total RNA. This is then converted into cDNA and sequenced. Whole genome sequencing fragments amplified genomic DNA and sequences the resulting DNA. Whole exome sequencing is similar to whole genome sequencing but uses tags to purify out all known exons prior to sequencing. Chromatin immunoprecipitation with massively parallel sequencing (ChIP-seq) uses antibodies directed against a protein of interest to immunoprecipitate fragments of genomic DNA bound by that particular protein. These fragments of interest are then sequenced.

Table 1
Single nucleotide polymorphisms associated with complex neurological diseases

This list is taken from the GWAS Catalogue and shows a summary of genes containing SNPs associated with complex neurological disease where $p < 10^{-7}$ (genome-wide significance) [6]. We have included only diseases and not studies focused on traits.[63–119]

Disease	Genes
Alzheimer's	ABCA7, APOC1, APOE, BIN1, CD2AP, CD33, CLU, CRI1, EPHA1, FRMD4A, GAB2, MS4A4A, MS4A6A, MS4A4E, MTHFD1L, PICALM, PVRL2, TOMM40
Amyotrophic lateral sclerosis	C9orf72, DPP6, IFNK, KIFAP3, MOBKL2B, SOD1, UNC13A, ZNF746
Autism	CDH10, CDH9, MACROD2
Creutzfeldt-Jakob disease	PRNP, STMN2
Creutzfeldt-Jakob disease (variant)	MTMR7, NPAS2, PRNP
Epilepsy	CAMSAP1L1
Essential tremor	LINGO1, SLC1A2
Glioma	CCDC26, CDKN2A, CDKN2B, EGFR, PHLDDB1, RTEL1, TERT
Meningioma	MLLT10
Migraine	ASTN2, FHL5, LRPI, MEF2D, MTDH, PGCP, PHACTR1, PRDM16, TGFBR2, TRPM8
Multiple sclerosis	AH11, BACH2, BATF, CBLB, CD40, CD58, CD6, CD86, CLECL1, CXCR4, CYP24A1, CYP27B1, DKKL1, EOMES, EXTL2, GPR65, HHEX, HLA-B, HLA-DQA1, HLA-DRA, HLA-DRB1, IL12B, IL22RA, IL2RA, IRF8, KIF1B, KPNB1, MALT1, MANBA, MAPK1, MERTK, METTL1, MLANA, MPV17L2, MYC, NA, NFKBIZ, NFKA7, SLC30A7, SOX8, SPI140, STAT3, TAGAP, TBKBP1, TBX21, THADA, TNFRSF1A, TNFSF14, ZBTB46, ZFP36L1, ZNF767
Narcolepsy	CPT1B, HLA-DQA2, TRA-alpha, TRAJ10
Neuroblastoma	BARD1, LMO1, FLJ22536, FLJ44180
Parkinson's disease	ACMSD, BST1, C10orf32, CCDC62, CNRM2, CYP17A1, DGKQ, GAK, GBA, HIP1R, HLA-DRA, HLA-DRB5, IMP5, LAMP3, LOC642072, LRRK2, MAPT, MCCC1, MMRN1, NSF, NUCKS1, PARK16, PLEKHM1, PM20D1, RAB7L1, RAI1, RIT2, SCARB2, SFXN2, SLC41A1, SLC45A3, SNCA, SREBF1, STH, STK39, SYT11, WNT3
Progressive supranuclear palsy	CD8B, EIF2AK3, IL2, IL21, IRF4, MAPT, MOBP, SLC6A1A2, STX6
Restless leg syndrome	BC034767, BTBD9, MAP2K5, MEIS1, PTPRD, SKOR1, TOX3
Stroke	HDAC9, NINJ2, NR

Table 2
Rare variants associated with neurological disease identified by whole genome or whole exome sequencing approaches

This table shows a selection of studies published within the last year that use whole genome or whole exome study designs to identify rare variants. WES = whole exome sequencing, WGS = whole genome sequencing. [18,25,55,57,120–164]

Condition	Design	Gene	Reference
Acute onset myoclonus and opsoclonus-like eye movements	WES	KCTD7	Blumkin et al. 2012 J Neurol
Adult onset primary cervical dystonia	WES	CIZ1	Xiao et al. 2012 Ann Neurol
Alternating hemiplegia of childhood	WES	ATP1A3	Rosewich et al. 2012 Lancet Neurol, Heinzen et al. 2012 Nat Genet
Autistic spectrum disorder	WES	Multiple gene disruptions, CHD8, KATNAL2	Iossifov et al. 2012 Neuron, Neale et al. 2012 Nature
Autism	WES	UBE3B, CLTCL1, NCKAP5L, ZNF18, ANK3, SCN2A	Chahrour et al. 2012 PLoS Genet, Bi et al. 2012 Hum Mutat, Sanders et al. 2012 Nature
Autosomal dominant acrodysostosis	WES	PDE4D	Lee et al. 2012 Am J Hum Genet
Autosomal dominant cerebellar ataxia, deafness and narcolepsy	WES	DNMT1	Winkelmann et al. 2012 Hum Mol Genet
Autosomal dominant early-onset Alzheimer disease	WES	SORL1	Pottier et al. 2012 Mol Psychiatry
Autosomal dominant hereditary motor and sensory neuropathy with proximal dominant involvement	WES	TFG	Ishiura et al. 2012 Am J Hum Genet
Autosomal dominant limb girdle muscular dystrophy	WES	DNAJB6	Harms et al. 2012 Ann Neurol
Autosomal dominant neuronal ceroid lipofuscinosis	WES	KCTD7	Staropoli et al. 2012 Am J Hum Genet
Autosomal dominant spinocerebellar ataxia	WES	PRKCG, TGM6	Sailer et al. 2012 Neurology, Li et al. 2012 Clin Genet
Autosomal recessive congenital profound sensorineural deafness	WES	TSPEAR	Delmaghani et al. 2012 Hum Mol Genet
Autosomal recessive Emery-Dreifuss muscular dystrophy	WES	LMNA	Jimenez-Escrig et al. 2012 Muscle Nerve
Autosomal recessive infantile onset spinocerebellar ataxia	WES	C10orf2	Dündar et al. 2012 Pediatr Neurol
Autosomal recessive Kohlschütter-Tönnz syndrome	WES	ROGDI	Schossig et al. 2012 Am J Hum Genet
Autosomal recessive myotonia congenita	WES	CLCN1	Raja Rayan et al. 2012 Neurology
Autosomal recessive pontocerebellar hypoplasia and spinal motor neuron degeneration	WES	EXOSC3	Wan et al. 2012 Nat Genet
Autosomal recessive spinal muscular atrophy associated with progressive myoclonic epilepsy	WES	ASAHI	Zhou et al. 2012 Am J Hum Genet
Brown-Vialetto-Van Laere syndrome (early-onset ALS)	WES	SLC52A3	Johnson et al. 2012 Brain
Charcot-Marie-Tooth disease	WES	Multiple heterozygous variants	Choi et al. 2012 Hum Mutat
Congenital disorders of glycosylation type I	WES	ALG13, DPAGT1, PGM1	Timal et al. 2012 Hum Mol Genet
Distal hereditary motor neuropathy type V	WES	REEP1	Beetz et al. 2012 Am J Hum Genet
Essential tremor	WES	FUS	Merner et al. Am J Hum Genet
Familial amyotrophic lateral sclerosis	WES	PFN1, UBQLN2, SOD1	Wu et al. 2012 Nature, Williams et al. 2012 Neurobiol Aging, Wu et al. 2012 Genet Med
Familial amyotrophic lateral sclerosis	WGS	C9orf72	Herdewyn et al. 2012 Hum Mol Genet

Condition	Design	Gene	Reference
Fatal infantile Alpers encephalopathy	WES	mtPheRS	Elo et al. 2012 Hum Mol Genet
GM1-gangliosidosis	WES	GLB1	Pierson et al. 2012 Neurology
Hemimegalencephaly	WES	PIK3CA, AKT3 and MTOR	Lee et al. 2012 Nat Genet
Hereditary sensory autonomic neuropathy	WES	DST	Edvardson et al. 2012 Ann Neurol
Idiopathic generalised epilepsy	WES	None identified that reached statistical significance	Heinzen et al. 2012 Am J Hum Genet
Infantile epileptic encephalopathy and SUDEP	WGS	SCN8A	Veeramah et al. 2012 Am J Hum Genet
Intractable epilepsy of infancy	WES	EFHC1	Berger et al. 2012 Epilepsia
Ischaemic stroke	WES	CSN3	Cole et al. 2012 PLoS ONE
Juvenile parkinsonism	WES	DNAJC6	Edvardson et al. 2012 PLoS ONE
Leukoencephalopathy with thalamus and brainstem involvement and high lactate	WES	EARS2	Steenweg et al. 2012 Brain
Multiple sclerosis	WES	TYK2	Dyment et al. 2012 Neurology
Neurodegeneration with brain iron accumulation	WES	C19orf12	Horvath et al. 2012 Mov Disord
Neuronal ceroid-lipofuscinosis	WES	ATP13A2	Bras et al. 2012 Hum Mol Genet
Ohtahara syndrome	WES	CASK	Saitu et al. 2012 Epilepsia
Teenage-onset progressive myoclonus epilepsy	WES	CLN6	Andrade et al. 2012 Pediatr Neurol
Undiagnosed sensorimotor neuropathy, ataxia, and spasticity	WES	SACS	Pyle et al. 2012 Arch Neurol

Table 3
Transcriptomic approaches to complex neurological disease

Some examples of studies using transcriptomic next generation sequencing methods to investigate complex neurological disease.[42,165–171]

Condition	Design	Gene	Reference
Alzheimer's disease	RNA-seq of human brain tissue	APOE	Twine et al. 2011 PLoS ONE
Amyotrophic lateral sclerosis	RNA-seq in Drosophila	TBPH	Hazelett et al. 2012 G3 (Bethesda)
Autism	RNA-seq of human brain tissue	Adenosine-to-inosine editing	Eran et al. 2012 Mol Psychiatry
Autism	RNA-seq in human primary neural stem cells	RBFOX1 splicing network	Fogel et al. 2012 Hum Mol Genet
Canine neonatal cerebellar cortical degeneration	RNA-seq in dogs	SPTBN2	Forman et al. 2012 BMC Genet
Multiple sclerosis	RNA-seq of viral RNA	Viral RNA	Kriesel et al. 2012 PLoS ONE
Multiple sclerosis	RNA-seq in human primary CD4+ cells	CD6	Heap et al. 2010 Hum Mol Genet
Schizophrenia, bipolar disorder, autism	RNA-seq of cultured neurons	Non-coding RNA	Lin et al. 2011 PLoS ONE