

REVIEW

Open Access



Advances in sequencing technologies for amyotrophic lateral sclerosis research

Evan Udine^{1,2}, Angita Jain^{1,2,3} and Marka van Blitterswijk^{1*}

Abstract

Amyotrophic lateral sclerosis (ALS) is caused by upper and lower motor neuron loss and has a fairly rapid disease progression, leading to fatality in an average of 2-5 years after symptom onset. Numerous genes have been implicated in this disease; however, many cases remain unexplained. Several technologies are being used to identify regions of interest and investigate candidate genes. Initial approaches to detect ALS genes include, among others, linkage analysis, Sanger sequencing, and genome-wide association studies. More recently, next-generation sequencing methods, such as whole-exome and whole-genome sequencing, have been introduced. While those methods have been particularly useful in discovering new ALS-linked genes, methodological advances are becoming increasingly important, especially given the complex genetics of ALS. Novel sequencing technologies, like long-read sequencing, are beginning to be used to uncover the contribution of repeat expansions and other types of structural variation, which may help explain missing heritability in ALS. In this review, we discuss how popular and/or upcoming methods are being used to discover ALS genes, highlighting emerging long-read sequencing platforms and their role in aiding our understanding of this challenging disease.

Keywords Amyotrophic lateral sclerosis, DNA sequencing, Long-read sequencing, SMRT sequencing, Nanopore sequencing, Multi-omics

Background

Amyotrophic lateral sclerosis (ALS) is a fatal neuromuscular disease caused by degeneration of both upper and lower motor neurons in the brain, brainstem, and spinal cord, typically displaying accumulation of cytoplasmic TDP-43 [1, 2, 3]. The most common clinical presentations are asymmetric limb weakness, which is seen in about 75% of ALS cases, and bulbar segment onset in about 25% of cases [4]. In addition to the motor

symptoms, approximately 60% of patients diagnosed with ALS will experience cognitive and/or behavioral changes, while up to 15% of cases may also receive a diagnosis of frontotemporal dementia (FTD) [5, 6]. Considerable clinical heterogeneity exists in terms of age of disease onset, ranging from 20 to 70 years old [7, 8], and survival after diagnosis, which is generally 2-5 years after onset, with approximately 10% of the patients living for 10 years or more [9, 10]. Diagnosing ALS often proves to be challenging, with the median time of definitive diagnosis between 1 and 4 years to distinguish ALS from other motor neuron diseases (MNDs) [11–13]. Clinical history and physical examination remain the gold standard for diagnosing ALS, even with the advancement of genetic testing [14].

In terms of ALS genetics, approximately 10% of all cases can be classified as familial (fALS), and the remaining 90% of cases are considered sporadic (sALS) [15, 16]. While most fALS cases are caused by mutations in a single gene

*Correspondence:

Marka van Blitterswijk

VanBlitterswijk.Marka@mayo.edu

¹ Department of Neuroscience, Mayo Clinic, 4500 San Pablo Road S, Jacksonville, FL 32224, USA

² Mayo Clinic Graduate School of Biomedical Sciences, 4500 San Pablo Road S, Jacksonville, FL 32224, USA

³ Center for Clinical and Translational Sciences, Mayo Clinic, 4500 San Pablo Road S, Jacksonville, FL 32224, USA



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(monogenic), a subset can be attributed to mutations in several genes (oligogenic) [17, 18]. With a heritability around 60% [19], sALS cases are thought to arise from a combination of variants in many genes (polygenic), probably in addition to environmental factors [17, 18, 20]. It should be noted that dividing ALS into fALS and sALS, although convenient, may not be straightforward. In fact, fALS cases are greatly underreported and can be misclassified as sporadic due to a short disease duration, small pedigrees, genetic heterogeneity, phenotypic variability, and incomplete penetrance [21–24].

Since the discovery of the first ALS gene in 1993, *SOD1* [25], additional genes have been implicated, ranging from causative genes to potential risk factors and disease modifiers (Table 1). Many types of genetic variants may contribute to ALS, such as single nucleotide variants (SNVs) and structural variants. SNVs in coding sequences can be pathogenic missense mutations that lead to the production of proteins with incorrect amino acid sequences, while SNVs in non-coding regions can confer disease risk by affecting the expression or splicing of nearby or distal genes. The other major class of variants, structural variants, encompass large genomic alterations in the form of insertions, deletions, inversions, translocations, repeat expansions, and copy number variations. Structural variants can also occur in non-coding regions of the genome, which often do not change the composition of the mature protein [26], and have been implicated in ALS and FTD (e.g. repeat expansions in *C9orf72*) [27, 28]. Pathogenic mutations in the genes *SOD1*, *C9orf72*, *FUS*, and *TARDBP* are the most frequently observed genetic causes of ALS [25, 27, 29, 30] and comprise both SNVs (i.e. *SOD1*) and structural variants (i.e. *C9orf72*).

Early genetic studies of ALS relied on mapping a chromosomal location in ALS pedigrees to nominate disease genes using a method called DNA linkage analysis. One of the most notable examples is the identification of chromosomal region 9p21 [86–90] and its subsequent refinement [91–94], which eventually led to the groundbreaking discovery of a repeat expansion in the gene *C9orf72* [27, 28]. Another popular method is a genome-wide association study (GWAS). The goal of a GWAS usually is to quantify differences in allele frequencies across the genome between cases and controls. It is an unbiased approach to identify disease-associated common genetic variants. This is often performed by genotyping numerous single nucleotide polymorphisms (SNPs), which are SNVs that are present in at least 1% of the population [95], using various methods (e.g. microarrays). A GWAS can reveal significant associations; please note, however, that association does not equal causation. In the ALS field, GWAS has resulted in several discoveries, including that of *UNC13A* and *KIF5A*

[61, 91]. Thus far, GWAS has only explained a small proportion of genetic susceptibility to ALS, suggesting rare and structural variants may account for a substantial proportion of missing heritability [96].

Sequencing methods

Sanger sequencing

The methods described above were used to discover disease-associated genomic loci and genes, but, on their own, are unable to provide sequence information about the genes themselves. For that, sequencing methods could be used. Sanger sequencing was one of the earliest sequencing methods developed to determine the DNA sequence. Nowadays, a modified version is used, where genomic DNA is amplified using primers that target a region of interest. Subsequently, the amplicon is sequenced by capillary electrophoresis. Sanger sequencing is a reliable method with up to 99.9% accuracy. Currently, in the ALS field, it is most commonly used to screen samples for mutations in well-known ALS genes [97] and to confirm the presence of mutations identified through other methods, such as "next-generation" sequencing [98].

Next-generation sequencing

More recently, sequencing efforts have shifted toward next-generation sequencing techniques, such as whole-exome sequencing (WES) and whole-genome sequencing (WGS). This enabled researchers to continue their search for ALS-linked genes, even in cases without a multigenerational family history and in families with limited DNA sample availability. These technologies leverage high-throughput, large-scale parallel DNA sequencing of all coding sequences (WES) or the entire genome (WGS) and can be very powerful in addressing monogenic disorders [99–101]. Briefly, to perform next-generation sequencing, DNA must be fragmented, regularly by shearing, sonication, or enzyme digestion. Then, linkers or specialized adaptors are often added at the ends of the fragmented molecules to create template libraries. The resulting clusters of DNA fragments are typically amplified on a chip, producing millions of copies of double-stranded DNA. Frequently, a signal for each base is detected using fluorescence during the sequencing procedure. Using this method, it is possible to produce either single (one direction of sequencing) or paired (both directions of sequencing) end reads that will need to be analyzed by a process known as base calling. Different software programs are used to sort and align DNA sequences to the reference genome and analyze the data efficiently [102]. As the cost of sequencing is declining, bigger cohorts are being sequenced. This enables the identification of coding and non-coding variants

Table 1 Discovery methods of ALS-associated genes (Adapted from Goutmann et al- Emerging insights into the complex genetics and pathophysiology of amyotrophic lateral sclerosis [31])

Gene by Discovery Method	Chromosomal Position	Inheritance Pattern	ALSoD Category	Pathogenic Mechanism	fALS	sALS	Year of Discovery	References
Linkage Analysis								
<i>ALS2</i>	2q33.1	AR	Tenuous	Trafficking and degradation of proteins	< 1%	< 1%	2001	[32]
<i>C9orf72</i>	9p21.2	AD	Definitive ALS gene	Repeat expansion, Trafficking, degradation of proteins, RNA foci, DPRs	40%	7%	2011	[27, 28]
<i>DAO</i>	12q24.11	AD	Moderate evidence	Oxidative stress	< 1%	< 1%	2010	[33]
<i>DCTN1</i>	2p13.1	AD, AR	Tenuous	Trafficking and degradation of proteins	< 1%	< 1%	2003	[34]
<i>ERBB4</i>	2q34	AD	Moderate evidence	Neuronal cell migration and development	< 1%	< 1%	2013	[35]
<i>FUS</i>	16p11.2	AD	Definitive ALS gene	RNA processing/splicing	4%	1%	2009	[29, 36]
<i>GLT8D1</i>	3p21.1	AD	Tenuous	Unknown	< 1%	< 1%	2019	[37]
<i>hnRNPA1</i>	12q13.13	AD	Definitive ALS gene	RNA processing/splicing	< 1%	< 1%	2013	[38]
<i>hnRNPA2B1</i>	7p15.2	AD	Tenuous	RNA processing/splicing	< 1%	< 1%	2013	[38]
<i>MATR3</i>	5q31.2	AD	Tenuous	RNA processing/splicing	< 1%	< 1%	2014	[39]
<i>SETX</i>	9q34.13	AD	Tenuous	RNA processing/splicing	< 1%	< 1%	1998	[40]
<i>SOD1</i>	21q22.11	AD, AR	Definitive ALS gene	Gain of toxic protein	12%	1-2%	1993	[25]
<i>SPG11</i>	15q21.11	AR	Tenuous	DNA Repair damage	< 1%	< 1%	2010	[41, 42]
<i>TARDBP</i>	1p36.22	AD	Definitive ALS gene	RNA processing/splicing	4%	1%	2008	[30, 43, 44]
<i>UBQLN2</i>	Xp11.21	XL	Definitive ALS gene	Trafficking and degradation of proteins	< 1%	< 1%	2011	[45]
<i>VAPB</i>	20q13.32	AD	Definitive ALS gene	Trafficking and degradation of proteins	< 1%	< 1%	2004	[46]
Candidate Gene Analysis								
<i>ANG</i>	14q11.2	AD	Moderate evidence	RNA processing/splicing	< 1%	< 1%	2006	[47]
<i>ATXN2</i>	12q24.12	AD	Clinical modifier	RNA processing/splicing	< 1%	< 1%	2010	[48]
<i>CHMP2B</i>	3p11.2	AD	Moderate evidence	Trafficking and degradation of proteins	< 1%	< 1%	2006	[49]
<i>CHRNA3</i>	15q24	N/A	N/A	Synaptic dysfunction	< 1%	< 1%	2009	[50]
<i>EWSR1</i>	22q12.2	N/A	Tenuous	RNA processing/splicing	< 1%	< 1%	2012	[51]
<i>FIG4</i>	6q21	AD	Moderate evidence	Trafficking and degradation of proteins	< 1%	< 1%	2009	[52]
<i>GLE1</i>	9q34.11	N/A	Moderate evidence	RNA processing/splicing	N/A	N/A	2016	[53]
<i>NEFH</i>	22q12.2	AD, AR	Tenuous	Trafficking and degradation of proteins	< 1%	< 1%	1999	[54]
<i>PON1-3</i>	7q21.3	N/A	Tenuous	Lipid metabolism	< 1%	< 1%	2009	[55]
<i>PRPH</i>	12q13.12	AD, AR	Tenuous	Trafficking and degradation of proteins	< 1%	< 1%	2004	[56]
<i>SIGMAR1</i>	9p13.3	AR	Tenuous	Trafficking and degradation of proteins	< 1%	< 1%	2011	[49]

Table 1 (continued)

Gene by Discovery Method	Chromosomal Position	Inheritance Pattern	ALSoD Category	Pathogenic Mechanism	fALS	sALS	Year of Discovery	References
<i>SQSTM1</i>	5q35.3	AD	Tenuous	Trafficking and degradation of proteins	~1%	<1%	2011	[57]
<i>TAF15</i>	17q12	N/A	Tenuous	RNA processing/splicing	N/A	N/A	2011	[58]
GWAS								
<i>CFAP410 (C21orf2)</i>	21q22.3	N/A	Strong evidence	Cytoskeletal defects	<1%	<1%	2016	[59]
<i>CAMTA1</i>	1p36.31-p36.23	AD	Clinical modifier	Trafficking and degradation of proteins	N/A	N/A	2016	[60]
<i>CCNF</i>	16p13.3	AD	Strong evidence		~1-3.3%	<1%	2016	[59]
<i>KIF5A</i>	12q13.3	AD	Definitive ALS gene	Axonal pathology	N/A	N/A	2018	[61]
<i>NIPA1</i>	15q11.2	AD	Strong evidence	Repeat expansion	N/A	N/A	2019	[62]
<i>SARM1</i>	17q11.2	N/A	Moderate evidence	Axonal pathology	N/A	N/A	2021	[63]
<i>TIA1</i>	2p13.3	AD	Tenuous	Oxidative stress	~2.2%	<1%	2017	[64]
<i>UNC13A</i>	19p13.11	N/A	Definitive ALS gene	Synaptic dysfunction	N/A	N/A	2012	[65, 66]
WES/WGS								
<i>ANXA11</i>	10q22.3	AD	Definitive ALS gene	Trafficking and degradation of proteins	~1%	~1.7%	2017	[67]
<i>CHCHD10</i>	22q11.23	AD	Definitive ALS gene	Trafficking and degradation of proteins (Mitochondria)	<1%	<1%	2014	[68, 69]
<i>DNAJC7</i>	17q21.2	N/A	Moderate evidence	Trafficking and degradation of proteins	<1%	<1%	2019	[70]
<i>NEK1</i>	4q33	AD	Definitive ALS gene	Trafficking and degradation of proteins	~1-2%	<1%	2015	[71–74]
<i>PFN1</i>	17p13.2	AD	Definitive ALS gene	Trafficking and degradation of proteins	<1%	<1%	2012	[75]
<i>TBK1</i>	12q14.2	AD	Definitive ALS gene	Trafficking and degradation of proteins	~3%	<1%	2015	[76–78]
<i>TUBA4A</i>	2q35	AD	Strong evidence	Trafficking and degradation of proteins	<1%	<1%	2014	[79, 80]
<i>VCP</i>	9p13.3	AD	Definitive ALS gene	Trafficking and degradation of proteins	1%	1%	2010	[81]
<i>ERLIN2</i> Homozygosity mapping	10q24.31	AR	Tenuous	Lipid transport	N/A	N/A	2012	[82]
<i>EPHA4</i> Functional study	2q36.1	N/A	Definitive ALS gene	Axonal pathology	N/A	N/A	2012	[83]
<i>OPTN</i> Homozygosity mapping	10p13	AD, AR	Definitive ALS gene	Trafficking and degradation of proteins	1%	<1%	2010	[84]
<i>VEGFA</i> Functional testing	6p21.1	N/A	Tenuous	Angiogenesis	N/A	N/A	2009	[85]

AD Autosomal dominant inheritance, AR Autosomal recessive inheritance, XL X-linked inheritance. ALS Online Database (ALSoD) category definitions = *Definitive ALS gene* Variants in these genes show increase in risk of ALS based on statistical test, *Clinical modifier* Variants in these genes have been linked to a difference in clinical phenotype of ALS, *Strong evidence* Variants in these genes have shown to increase risk in well-conducted recent studies but require replication or resolution of conflicting evidence, *Moderate evidence* Variants in these genes have been associated with ALS in smaller studies or there may be contradictory evidence, *Tenuous* Variants in these genes have been associated with ALS in smaller studies a while ago but have not stood up to replication. They may also be genes that have been associated with ALS-like diseases, which are no longer recognized as a clinical diagnosis of ALS. N/A Information is not available. fALS represents the reported frequency of these variants in familial ALS cases and sALS represents the reported frequency of these variants in sporadic ALS cases

associated with ALS. To prioritize variants, analyses like unsupervised learning [103, 104], linear mixed-modeling, and gene burden testing [105] have been employed. For example, an exome-wide rare variant burden analysis confirmed the significant GWAS hit in *KIF5A*, and

additionally, revealed significant associations for *TBK1* and *NEK1* [71, 76, 106]. To extend aforementioned studies, Project MinE (see section ‘Collaborative sequencing efforts’ for additional details) aims at performing WGS on 15,000 ALS patients and 7500 matched controls. In

addition to *KIF5A* and *NEK1*, this project already identified *CFAP410* [59, 61, 71, 107], as well as detected structural variants in *C9orf72*, *VCP*, and *ERBB4* [108]. Nonetheless, WES and WGS are not entirely without limitations. One major drawback of WES is its restriction to exonic regions, generally missing intronic, promoter, and enhancer variants. Both WES and WGS depend on read quality and sequencing depth, and additionally, they encounter issues calling structural variation. For example, the intronic repeat in *C9orf72* has been identified through linkage mapping and the locus has been implicated in several GWASs; however, the expansion is challenging/arduous to capture and size by WGS because of the difficulty aligning short-read data of microsatellite and minisatellite DNA sequences [109, 110]. Though bioinformatic tools have now been developed to detect structural variation in short-read data (ExpansionHunter, HipSTR, GangSTR, etc.) [102, 111, 112], other approaches that potentially provide a more in-depth characterization will be valuable in understanding the missing heritability in ALS and/or identifying disease modifiers [113].

Long-read sequencing

While the previously discussed approaches have been crucial to gene discovery in ALS, newer technologies are beginning to take precedent to address outstanding genomic questions. One approach that is continuing to gain popularity is long-read sequencing. Though there are multiple platforms available, and others in development, long-read sequencing can broadly be defined as any single-molecule sequencing approach that is capable of generating reads that are multiple kilobases in length. Platforms from Pacific Biosciences (PacBio) [114, 115] and Oxford Nanopore Technologies (ONT) [116, 117] appear to have emerged as the leading long-read sequencing technologies (Fig. 1).

PacBio has developed a technology called single-molecule real-time (SMRT) sequencing [114, 118]. Originally developed in 2009, PacBio's SMRT sequencing produces long reads by incorporating phospholinked nucleotides labelled with different colored fluorophores. SMRT sequencing is achieved at zero mode waveguides (ZMWs), which are tiny wells with a glass bottom, that can hold a single DNA molecule. PacBio's SMRT cells accommodate millions of these ZMWs for sequencing to occur. At each ZMW, an anchored DNA polymerase will incorporate a labelled nucleotide, complementary to that from the template DNA molecule. When this occurs, light is emitted and the signal, which is unique to each base, is measured in real time [114, 118, 119]. More recently, PacBio has implemented updated sequencing technologies with the Sequel II system, and in 2019 they introduced high fidelity (HiFi) sequencing, which

drastically improves the accuracy of the sequencing by utilizing circularized adapters (SMRT Bell Adapters) so that each molecule can be sequenced multiple times [120]. With this option, the user can computationally call a consensus sequence – a circular consensus sequence (ccs read) – to obtain the most accurate read possible, with reads being >99.9% accurate [120]. While HiFi sequencing may be the best option for obtaining the most accurate sequences, another method, called continuous long-read sequencing (clr) generates longer reads. Additionally, in 2022 PacBio announced a new platform called the Revio. The Revio is a sequencer that performs HiFi sequencing at a much greater scale than what was previously achievable. It is able to generate data at a 15x higher throughput than the Sequel II, containing 25 million ZMWs in a single SMRT cell [121].

The other most prominent long-read sequencing technology is from ONT. ONT has pioneered the nanopore technology for long-read sequencing, where they are able to sequence extremely long reads while also yielding many reads [116, 122]. Nanopore sequencing works by using a motor protein and tether to pull a single-stranded DNA molecule through a nanopore. The change in ionic current is measured as each nucleotide is passed through the nanopore, with unique signals for each base [116, 123]. There are three main ONT sequencing platforms, including the MinION [116, 122, 124–126], GridION [127, 128], and PromethION [129, 130], which are different sequencers with the same underlying technology, but varying strengths and weaknesses. The MinION is the smallest and most cost effective of the machines offering desktop and portable sequencing options, but has the lowest yield, lowest accuracy (initial estimates around 60%) and can only sequence one flow cell at a time [116]. The GridION allows for up to five flow cells at a time and can generate 250 Gb of sequencing data but does not offer much improvement from the MinION other than scalability [131]. Lastly, the PromethION offers up to 48 flow cells and produces the most accurate sequencing data offered by ONT, with a read accuracy of up to 99% [132]. Overall, considerations of cost, read length, read depth, and sequencing accuracy need to be considered when choosing which long-read sequencing technology to use.

Long-read sequencing applications

Long-read sequencing has been used in other fields to create reference genomes and/or transcriptomes for a diverse number of species [122, 133–136]. More recently, in humans, long-read WGS has been utilized by the telomere-to-telomere consortium to sequence the first “complete human genome” [137, 138]. This

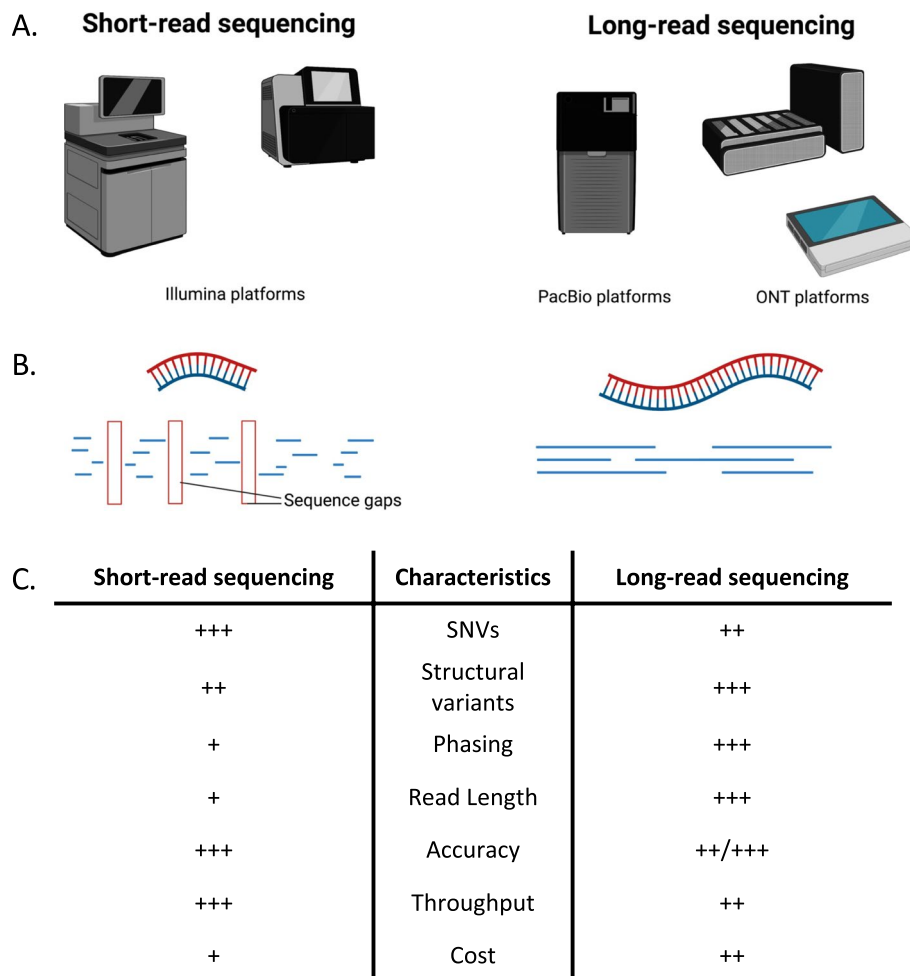


Fig. 1 Overview of short-read and long-read sequencing technologies. **A** Examples of widely used platforms for short-read and long-read sequencing technologies. **B** The primary difference between short-read and long-read sequencing technologies is the significant increase in read length. In contrast to short-read sequencing (150–300 bp), long-read sequencing has the capacity to sequence reads spanning multiple kilobases in one single read, thereby requiring fewer reads to cover the same gene. The read overlap seen with long-read data reduces the sequence gaps as observed in short-read data. **C** Semi-quantitative comparison of short-read and long-read sequencing of various features including the ability to detect single nucleotide variants (SNVs), structural variants, and complete genome phasing, as well as the overall read length, accuracy, throughput, and sequencing cost

expedition begun to build a reference genome without any gaps in humans, where researchers used both PacBio and ONT to sequence every part of the genome, including the telomeres and centromeres that were previously too difficult to capture [137, 138]. One of the main advantages of long-read WGS is its ability to cover these kinds of complex genomic regions and find structural variation [139–142]. Structural variation, including insertions, deletions, inversions, translocations, expansions, and copy number variations are difficult to capture with short-read sequencing because the length of each sequencing read is often shorter than the size of the structural variant [140, 141, 143]. As previously mentioned, (see [Background](#)), structural variation may

explain some of the missing heritability in ALS [113]. Thus far to our knowledge, only one published study has performed long-read WGS in the context of ALS, where they focused on *C9orf72* repeat expansions [109]. Using the ONT MinION, they could not detect any reads covering the *C9orf72* expansion, while with PacBio SMRT sequencing there was 8x coverage of the expansion [109]. Currently, no large-scale association studies in ALS have been reported (yet) with long-read WGS. Studies have utilized long reads, however, to identify many structural variants in a small number of subjects [144], and been used to resolve complex regions that harbor known polymorphisms [145–148] or to validate structural variants that have been determined by other methods [149, 150].

Therefore, there is great promise for this technology to be used in the future of ALS research.

Rather than performing genome-wide long-read sequencing, targeted sequencing approaches can be used to scrutinize highly complex regions of the genome where there is known genetic risk. For very long variants or repeat expansions, WGS may not have enough reads to sufficiently cover those regions [109]. Therefore, targeted methods are extremely useful for understanding repeat expansions and have been applied in many neurological diseases, such as those associated with repeat expansions in *FMRI* [151, 152], *NOTCH2NLC* [153], *DMI* [154] and *HTT* [155] to name a few. PacBio and ONT both offer targeted sequencing platforms that select a specific region of the genome using primers, probes, or CRISPR-based methods. PacBio's targeted sequencing method, No-Amp (no amplification) sequencing is a DNA sequencing approach that can be used with CRISPR-Cas9 and custom designed guide RNAs to target a specific region in the genome [156]. The main advantages of this approach over alternative methods are that it can measure the exact length and sequence of the expansion, while detecting DNA methylation (as can ONT) [156]. In the design of No-Amp studies, researchers can elect to capture the flanking regions around the expansion so that expansion length, which may act as a disease modifier in certain diseases, can be accurately sized by ensuring the entire expanded region is captured. This has been applied, for example, to sequence through the *C9orf72* repeat expansion [109, 157]. No-Amp sequencing that has been completed in this region has demonstrated that the expansion length from No-Amp is correlated with the estimated length from Southern blotting, the current gold standard for sizing the *C9orf72* expansion [157]. Targeted sequencing has been done for other ALS genes, where long-read sequencing revealed an unstable intronic repeat with variation in the sequence of the gene *WDR7*, which was missed by other sequencing technologies [158]. No-Amp sequencing can be done on a number of genes at the same time, especially with smaller expansions. This multi-gene approach has been used to look at repetitive regions that cause various spinocerebellar ataxias (SCAs) [159], such as SCA1, SCA2, SCA10, and SCA36, as well as myotonic dystrophy type 1, where it was possible to size the repeats and detect interruptions in the sequence within the each repeat [159, 160]. One of these diseases, SCA2 is caused by a repeat expansion in the gene *ATXN2* [161], which has been demonstrated to be a genetic modifier of ALS [48, 162, 163].

Here, we have highlighted the power of long-read DNA sequencing (Fig. 1), specifically demonstrating its ability to sequence through highly complex regions of the genome [141]. As mentioned previously, the two

technologies highlighted above both detect DNA methylation, however, tools for analyzing this data are in earlier stages of development [164, 165]. In addition to these two main platforms, other options from companies such as Beijing Genomics Institute, 10x genomics, and Illumina are available or in development. Alternatively, non-sequencing, optical mapping approaches from Bionano and OpGen can be used to visualize large chromosomal abnormalities. Despite the many advantages of long-read sequencing, there remain limitations. Primarily, it is generally more expensive than alternative approaches, while generating fewer reads than short-read sequencing [166]. Additionally, there is a great computational cost. Data files can be on the scale of terabytes of data per flow cell, which makes data storage and processing costly. Moreover, the quality of the material required to guarantee sequencing integrity can be a challenge when working with frozen tissue, particularly tissue from the central nervous system. Finally, though much longer reads can be obtained than with traditional sequencing methods, it is inevitable that some structural variants will exceed read length capabilities. Nevertheless, this technology is continuing to advance, with reduction of cost and rapid improvements to the number, the length, and the accuracy of reads that are generated.

Multi-omics

Thus far, we have focused on the use of single DNA sequencing techniques to identify causal variants and genes, as well as genetic modifiers and/or risk factors of ALS. While these approaches have and will continue to be widely useful, there is tremendous value in integrating multiple data types to further prioritize disease-relevant or causative genes. Functional genomic and/or multi-omic approaches rely on incorporating DNA sequencing data with other data types to look at the epigenome, transcriptome, proteome, etc. Methods for these analyses are very powerful for highly polygenic diseases, where multiple common variants may confer some disease risk if not sufficient to cause disease. Given the apparent polygenic nature of sALS [61, 107], it will be important to use these integrative approaches to nominate genes that may be impacted by identified genetic variants. Herein, this review will discuss common functional genomic/multi-omic approaches while highlighting how they have been used in ALS or related diseases.

Perhaps the most commonly used approach in multi-omic research is bulk short-read RNA sequencing (RNAseq). RNAseq is a next-generation sequencing method that can be used to quantify gene expression and splicing for many genes across the entire transcriptome. RNAseq analyses are commonly used in animal and cell

models of ALS to determine the transcriptomic effects of gene knockout or overexpression [167–171]. Standard disease-relevant RNAseq analyses in humans are used to perform case vs. control analyses to identify differential gene expression and differential splicing across the transcriptome. Other analyses, like network and pathway analyses can be done to find networks of genes whose expression is correlated and determine the dysregulated molecular pathways, rather than single genes. This has been done many times across neurodegenerative diseases including in the ALS field [172–175]. Differential expression analysis of human brain tissue in the context of ALS, for instance, has revealed that sALS and *C9orf72*-linked ALS demonstrate wide-spread splicing alterations, but have unique transcriptomic profiles [172, 174], and later showed that repetitive elements are increased in *C9orf72*-linked ALS [167, 173]. Another RNAseq study in ALS identified three major unique molecular subtypes - retrotransposon activation, oxidative damage, and glial activation - of ALS, based on unique transcriptomic profiles [175]. Additional RNAseq data was generated in human ALS tissue, which was used to show truncated transcripts of *STMN2*, a microtubule gene that has been implicated in ALS and FTD, are present specifically in tissues with TDP-43 pathology [176–178]. Further analyses of RNAseq data revealed the mechanism by which ALS-associated SNPs in the gene *UNC13A* are likely pathogenic [65, 91, 107]. Moreover, RNAseq was performed in multiple cell types and in human tissues to show that variants in *UNC13A* increase the inclusion of a cryptic exon, which is an exon that is present within a normally intronic region and is incorrectly included in the mature mRNA, possibly by preventing TDP-43 from binding to the cryptic splice site [179, 180]. These cryptic exons may be particularly relevant in ALS, as one of the roles of TDP-43 is to prevent their inclusion into mature RNAs [167]. Cryptic splicing events may continue to be observed in additional genes relevant to ALS and are proposed to be pathogenic by either introducing an early stop codon, causing a loss of expression, or by being incorporated into the mature RNA, and thus potentially leading to the production of a toxic protein. Future studies, like the ones described here, are essential for increasing our understanding about how genetic variants may confer pathogenicity [167].

Currently, single-cell and single-nuclei RNAseq approaches are being used to identify cell type changes and transcriptomic alterations within specific cell types. Many single-cell studies have been utilized in the cancer field and in other neurodegenerative diseases, like Alzheimer's disease [181]. More recently, researchers have begun to perform single-cell sequencing in the ALS field. These studies have pointed toward alterations in multiple

cell types, suggesting that genetic risk of ALS is conferred through interneurons, motor neurons/Betz cells, and oligodendrocytes [182]. This goes beyond bulk RNAseq methods, allowing researchers to find cell type alterations that are unable to be detected with current pathological measures.

Other newer approaches can be used to pinpoint genes and proteins that change in specific regions of a cell or tissue. Spatial transcriptomics has been used more widely in cancer and tumor biology, with more limited applications in neurodegeneration and ALS [183, 184]. One study performed spatial transcriptomics in mouse and human ALS tissue and found alterations in microglia and astrocyte dynamics in the spinal cord [183]. Another study found 16 transcripts that were dysregulated in the granular cell layer of ALS spinal cords [184]. This approach can be further applied to look for transcriptomic changes surrounding the various pathological features of ALS (i.e. TDP-43), as has been done in the context of amyloid pathology in Alzheimer's disease [185].

In addition to these short-read RNAseq approaches, long-read RNAseq can also be used to improve upon short-read approaches by detecting more alternative splicing events than short-read sequencing [186] and identifying novel transcript variants and genes, which may be particularly relevant to ALS given the strong implication of RNA-binding proteins in disease pathogenesis [187–192]. PacBio [159] and ONT [193] also dominate the long-read RNAseq field with RNAseq possible on all the previously described platforms. Efforts are currently ongoing to apply long-read RNA sequencing to sizeable human datasets, and thus far have primarily been used for transcriptome reference assembly.

Multi-omic approaches, however, are not just limited to expression profiling. Other approaches, such as ATAC-seq (chromatin accessibility) [194], CHIP-seq (protein-DNA/RNA binding) [195], and HI-C (genome structure/interactions) [196] can be used to look at regulatory changes across the entire genome. Relevant to ALS, the Answer ALS consortium [197] is pioneering efforts to integrate many types of multi-omic data, including genomic, transcriptomic, epigenomic, proteomic, and metabolomic data, with the end goal of developing a cure for ALS (see section 'Collaborative sequencing efforts' for additional details). Various multi-omic studies relevant to ALS have been completed in induced pluripotent stem cell (iPSC) models [198, 199], with one group integrating ALS GWAS with RNAseq, ATAC-seq, CHIP-seq, and HI-C to identify *KANK1* as an ALS risk gene [199].

A common way to integrate multi-omic data is through quantitative trait loci mapping (QTLs) [200]. QTLs can be used to determine the molecular effect of a genetic

variant, where the presence of a variant can be associated with expression (eQTL), splicing (sQTL), methylation (meQTL), or other -omic measures. QTL results are used in transcriptome-wide association studies (TWAS) to nominate genes that may be impacted by disease-associated genetic variants [201]. Two TWASs have been completed in ALS, where expression was estimated from human brain tissue and blood. These studies have been able to validate previously identified GWAS loci, and nominate 7 and 5 novel genes, respectively [202, 203]. Like loci identified through GWAS and other methods, TWAS results require validation and replication, however it is clear that TWAS itself can also be used to identify novel loci.

Collaborative sequencing efforts

In many fields, not just in the ALS community, efforts are being made to generate large datasets that include many individuals in order to increase power to detect genetic variants. Project MinE aims at generating WGS from greater than 15,000 ALS patients and an additional 7500 controls [204]. The Clinical Research in ALS and related disorders for Therapeutic development (CREATe) consortium seeks to discover ALS biomarkers, creating a data repository that contains WGS and biospecimens for >1000 subjects with ALS or other MNDs [205]. Answer ALS [197], which was also mentioned earlier, is focused on developing ALS patient-derived cell lines and generating multi-omic data from cell lines and human tissue data. NeuroLINCS, which is a major contributor to Answer ALS, is a collaborative effort to perform multi-omic profiling of iPSC-derived motor neurons [198]. Likewise, the NYGC ALS Consortium, which contributes to the sequencing effort of Answer ALS, is working to integrate WGS and RNAseq data from human ALS tissue [61]. Each of these consortia and collaborative efforts have the same end goal – to work towards developing a treatment for ALS. It should be noted, though, that these efforts are all works in progress and do not provide all answers to the many challenging questions they are attempting to address. Of course, creating these large datasets will yield many new lines of investigation. However, researchers should be mindful that like other studies, validation and replication remains crucial.

These collaborative efforts use many of the previously discussed methods and technologies and apply them to global datasets to identify disease-relevant variants and genes that we currently do not have enough power to detect. Analysis of these datasets may require machine learning or deep learning approaches with the goal of deciphering causal variants and identifying disease subtypes that may contain distinct genetic drivers [206, 207]. These datasets seek to move/ will potentially

move the field towards developing therapeutics and possibly to inform personalized medicine. In the genomics field for sporadic diseases, which accounts for 90% of ALS patients, larger GWAS studies can allow researchers to calculate polygenic risk scores (PRS). These scores can be used to determine which pathways may be driving disease risk or to calculate a genetic risk for disease in a given individual. PRS can be calculated using the summary statistics from massive association studies (i.e. GWAS) and therefore can be updated with every new association study that is released [208, 209]. In the future, it may be possible to use these scores in the same way that we use current genetic testing, but rather than looking at one or a few genes, the whole genome will be considered [210]. While some PRS scores are currently being used for melanoma [211], coronary artery disease [212] and diabetes [213], PRS is still primarily in the research phase for ALS and many other diseases. One PRS has been completed recently in ALS but did not seem likely to have clinical utility based on the small proportion of heritability that it could explain [214]. Perhaps this is because ALS is driven by variation other than common SNPs, so PRS calculated from GWAS may never be sufficient [107]. Possibly some of the long-read sequencing and multi-omic approaches can be utilized to improve the predictive power of PRS. It should be noted that PRS will require large datasets for training and is highly dependent on population structure [215]. Because most GWAS have been completed in Caucasian/European populations, there is a risk that if PRS are introduced in the clinic, they may not benefit diverse populations worsening current disparities in healthcare [215]. Therefore, efforts should continue to be made to include patients of many genetic population backgrounds in sequencing studies.

Conclusions

Classical gene discovery methods have helped to uncover important genetic variation that is causative or modulates risk of developing ALS. Linkage analyses in familial studies and Sanger sequencing will continue to remain pertinent to identify variants in known genomic regions. Newer sequencing methods have facilitated discovery of pathogenic variation in individuals, families and even in large populations not just for known genes, but across the entire genome. The emergence of long-read sequencing has shed light on more complex variation, including repeat expansions and other types of structural variation in ALS. In the upcoming years, we expect that long-read sequencing technologies will continue to be used by more researchers and clinicians, especially if the costs decrease, as it provides an unbiased approach to capture the complex genetics of ALS. Integration of multiple methods using multi-omic techniques to determine

the effect of variants will also continue to help nominate genes and pathways that contribute to disease pathogenesis. Improving our understanding of the origin and course of the disease will be useful not only in developing hypotheses for research but will be equally important clinically to help with genetic testing and disease prediction, ultimately offering therapeutic solutions for this devastating disease.

Abbreviations

ALS	Amyotrophic lateral sclerosis
ATAC-seq	Assay for transposase accessible chromatin sequencing
CCS	Circular consensus sequence
CHIP-seq	Chromatin immunoprecipitation sequencing
CLR	Continuous long read
CRISPR	Clustered regularly interspaced short palindromic repeats
eQTL	Expression quantitative trait locus
fALS	Familial ALS
FTD	Frontotemporal dementia
GWAS	Genome-wide association study
HiFi	High fidelity
iPSC	Induced pluripotent stem cell
meQTL	Methylation quantitative trait locus
MND	Motor neuron disease
No-Amp	No amplification
ONT	Oxford Nanopore Technologies
PacBio	Pacific Biosciences
pQTL	Protein quantitative trait locus
PRS	Polygenic risk score
QTL	Quantitative trait locus
RNAseq	RNA sequencing
sALS	Sporadic ALS
SCA	Spinocerebellar ataxia
SMRT sequencing	Single-molecule real-time sequencing
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
TWAS	Transcriptome-wide association study
WES	Whole-exome sequencing
WGS	Whole-genome sequencing
ZMW	Zero mode waveguide

Acknowledgements

The authors thank Drs. Phil Wong and Wilfried Rossoll for inviting us to contribute this review to the “ALS: Challenges and New Advances” review series.

Authors' contributions

EU prepared the manuscript with significant contributions from AJ. All authors conceived its content and structure, reviewed, edited, and approved the final version.

Funding

MVB is supported by the National Institute of Neurological Disorders and Stroke (NINDS; NS123052 and NS121125) and the Spastic Paraplegia Foundation, Inc., (SPF), and previously by the National Ataxia Foundation and the Muscular Dystrophy Association. EU and AJ are supported by the Mayo Clinic Graduate School of Biomedical Sciences. AJ is further supported by Clinical and Translational Science Awards Program Grant Number TL1 TR002380 from the National Center for Advancing Translational Science (NCATS). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health (NIH).

Availability of data and materials

Not applicable.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

Not applicable.

Received: 6 October 2022 Accepted: 23 December 2022

Published online: 13 January 2023

References

- Rowland LP. Amyotrophic lateral sclerosis. *Curr Opin Neurol.* 1994;7(4):310–5.
- Brooks BR. El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. Subcommittee on Motor Neuron Diseases/Amyotrophic Lateral Sclerosis of the World Federation of Neurology Research Group on Neuromuscular Diseases and the El Escorial “Clinical limits of amyotrophic lateral sclerosis” workshop contributors. *J Neurol Sci.* 1994;124(Suppl):96–107.
- Sejvar JJ, Holman RC, Bresee JS, Kochanek KD, Schonberger LB. Amyotrophic lateral sclerosis mortality in the United States, 1979–2001. *Neuroepidemiology.* 2005;25(3):144–52.
- van Es MA, Hardiman O, Chio A, Al-Chalabi A, Pasterkamp RJ, Veldink JH, et al. Amyotrophic lateral sclerosis. *Lancet.* 2017;390(10107):2084–98.
- Ringholz GM, Appel SH, Bradshaw M, Cooke NA, Mosnik DM, Schulz PE. Prevalence and patterns of cognitive impairment in sporadic ALS. *Neurology.* 2005;65(4):586–90.
- Phukan J, Pender NP, Hardiman O. Cognitive impairment in amyotrophic lateral sclerosis. *Lancet Neurol.* 2007;6(11):994–1003.
- Chio A, Logroscino G, Hardiman O, Swingler R, Mitchell D, Beghi E, et al. Prognostic factors in ALS: a critical review. *Amyotroph Lateral Scler.* 2009;10(5-6):310–23.
- Cronin S, Hardiman O, Traynor BJ. Ethnic variation in the incidence of ALS: a systematic review. *Neurology.* 2007;68(13):1002–7.
- del Aguila MA, Longstreth WT Jr, McGuire V, Koepsell TD, van Belle G. Prognosis in amyotrophic lateral sclerosis: a population-based study. *Neurology.* 2003;60(5):813–9.
- Testa D, Lovati R, Ferrarini M, Salmoiraghi F, Filippini G. Survival of 793 patients with amyotrophic lateral sclerosis diagnosed over a 28-year period. *Amyotroph Lateral Scler Other Motor Neuron Disord.* 2004;5(4):208–12.
- Palese F, Sartori A, Logroscino G, Pisa FE. Predictors of diagnostic delay in amyotrophic lateral sclerosis: a cohort study based on administrative and electronic medical records data. *Amyotroph Lateral Scler Frontotemporal Degener.* 2019;20(3-4):176–85.
- Galvin M, Ryan P, Maguire S, Heverin M, Madden C, Vajda A, et al. The path to specialist multidisciplinary care in amyotrophic lateral sclerosis: a population-based study of consultations, interventions and costs. *PLoS One.* 2017;12(6):e0179796.
- Richards D, Morren JA, Pioro EP. Time to diagnosis and factors affecting diagnostic delay in amyotrophic lateral sclerosis. *J Neurol Sci.* 2020;417:117054.
- Brooks BR, Miller RG, Swash M, Munsat TL. World Federation of Neurology Research Group on Motor Neuron D. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Other Motor Neuron Disord.* 2000;1(5):293–9.
- Alsultan AA, Waller R, Heath PR, Kirby J. The genetics of amyotrophic lateral sclerosis: current insights. *Degener Neurol Neuromuscul Dis.* 2016;6:49–64.
- Chen S, Sayana P, Zhang X, Le W. Genetics of amyotrophic lateral sclerosis: an update. *Mol Neurodegener.* 2013;8:28.

17. van Blitterswijk M, van Es MA, Hennekam EA, Dooijes D, van Rheenen W, Medic J, et al. Evidence for an oligogenic basis of amyotrophic lateral sclerosis. *Hum Mol Genet.* 2012;21(17):3776–84.
18. McCann EP, Henden L, Fifta JA, Zhang KY, Grima N, Bauer DC, et al. Evidence for polygenic and oligogenic basis of Australian sporadic amyotrophic lateral sclerosis. *J Med Genet.* 2020. <https://doi.org/10.1136/jmedgenet-2020-106866>.
19. Al-Chalabi A, Fang F, Hanby MF, Leigh PN, Shaw CE, Ye W, et al. An estimate of amyotrophic lateral sclerosis heritability using twin data. *J Neurol Neurosurg Psychiatry.* 2010;81(12):1324–6.
20. Bandres-Ciga S, Noyce AJ, Hemani G, Nicolas A, Calvo A, Mora G, et al. Shared polygenic risk and causal inferences in amyotrophic lateral sclerosis. *Ann Neurol.* 2019;85(4):470–81.
21. Ryan M, Heverin M, Doherty MA, Davis N, Corr EM, Vajda A, et al. Determining the incidence of familiarity in ALS: a study of temporal trends in Ireland from 1994 to 2016. *Neurol Genet.* 2018;4(3):e239.
22. Al-Chalabi A, Lewis CM. Modelling the effects of penetrance and family size on rates of sporadic and familial disease. *Hum Hered.* 2011;71(4):281–8.
23. Williams DB, Floate DA, Leicester J. Familial motor neuron disease: differing penetrance in large pedigrees. *J Neurol Sci.* 1988;86(2–3):215–30.
24. Jones CT, Swingler RJ, Brock DJ. Identification of a novel SOD1 mutation in an apparently sporadic amyotrophic lateral sclerosis patient and the detection of Ile113Thr in three others. *Hum Mol Genet.* 1994;3(4):649–50.
25. Rosen DR, Siddique T, Patterson D, Figlewicz DA, Sapp P, Hentati A, et al. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature.* 1993;362(6415):59–62.
26. Roses AD, Akkari PA, Chiba-Falek O, Lutz MW, Gottschalk WK, Saunders AM, et al. Structural variants can be more informative for disease diagnostics, prognostics and translation than current SNP mapping and exon sequencing. *Expert Opin Drug Metab Toxicol.* 2016;12(2):135–47.
27. DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron.* 2011;72(2):245–56.
28. Renton AE, Majounie E, Waite A, Simon-Sanchez J, Rollinson S, Gibbs JR, et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron.* 2011;72(2):257–68.
29. Vance C, Rogelj B, Hortobagyi T, De Vos KJ, Nishimura AL, Sreedharan J, et al. Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science.* 2009;323(5918):1208–11.
30. Gitcho MA, Baloh RH, Chakraverty S, Mayo K, Norton JB, Levitch D, et al. TDP-43 A315T mutation in familial motor neuron disease. *Ann Neurol.* 2008;63(4):535–8.
31. Goutman SA, Hardiman O, Al-Chalabi A, Chió A, Savelieff MG, Kiernan MC, et al. Emerging insights into the complex genetics and pathophysiology of amyotrophic lateral sclerosis. *Lancet Neurol.* 2022;21(5):465–79.
32. Hadano S, Hand CK, Osuga H, Yanagisawa Y, Otomo A, Devon RS, et al. A gene encoding a putative GTPase regulator is mutated in familial amyotrophic lateral sclerosis 2. *Nat Genet.* 2001;29(2):166–73.
33. Mitchell J, Paul P, Chen HJ, Morris A, Payling M, Falchi M, et al. Familial amyotrophic lateral sclerosis is associated with a mutation in D-amino acid oxidase. *Proc Natl Acad Sci U S A.* 2010;107(16):7556–61.
34. Puls I, Jonnakuty C, LaMonte BH, Holzbaur EL, Tokito M, Mann E, et al. Mutant dynactin in motor neuron disease. *Nat Genet.* 2003;33(4):455–6.
35. Takahashi Y, Fukuda Y, Yoshimura J, Toyoda A, Kurppa K, Moritoyo H, et al. ERBB4 mutations that disrupt the neuregulin-ErbB4 pathway cause amyotrophic lateral sclerosis type 19. *Am J Hum Genet.* 2013;93(5):900–5.
36. Kwiatkowski TJ Jr, Bosco DA, Leclerc AL, Tamrazian E, Vanderburg CR, Russ C, et al. Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science.* 2009;323(5918):1205–8.
37. Cooper-Knock J, Moll T, Ramesh T, Castelli L, Beer A, Robins H, et al. Mutations in the glycosyltransferase domain of GLT8D1 are associated with familial amyotrophic lateral sclerosis. *Cell Rep.* 2019;26(9):2298–306.e5.
38. Kim HJ, Kim NC, Wang YD, Scarborough EA, Moore J, Diaz Z, et al. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature.* 2013;495(7442):467–73.
39. Johnson JO, Piro EP, Boehringer A, Chia R, Feit H, Renton AE, et al. Mutations in the Matrin 3 gene cause familial amyotrophic lateral sclerosis. *Nat Neurosci.* 2014;17(5):664–6.
40. Chen YZ, Bennett CL, Huynh HM, Blair IP, Puls I, Irobi J, et al. DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). *Am J Hum Genet.* 2004;74(6):1128–35.
41. Daoud H, Zhou S, Noreau A, Sabbagh M, Belzil V, Dionne-Laporte A, et al. Exome sequencing reveals SPG11 mutations causing juvenile ALS. *Neurobiol Aging.* 2012;33(4):839.e5–9.
42. Orlacchio A, Babalini C, Borreca A, Patrono C, Massa R, Basaran S, et al. SPATACSIN mutations cause autosomal recessive juvenile amyotrophic lateral sclerosis. *Brain.* 2010;133(Pt 2):591–8.
43. Sreedharan J, Blair IP, Tripathi VB, Hu X, Vance C, Rogelj B, et al. TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science.* 2008;319(5870):1668–72.
44. Kabashi E, Valdmanis PN, Dion P, Spiegelman D, McConkey BJ, Vande Velde C, et al. TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. *Nat Genet.* 2008;40(5):572–4.
45. Deng HX, Chen W, Hong ST, Boycott KM, Gorrie GH, Siddique N, et al. Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia. *Nature.* 2011;477(7363):211–5.
46. Nishimura AL, Mitne-Neto M, Silva HC, Richieri-Costa A, Middleton S, Cascio D, et al. A mutation in the vesicle-trafficking protein VAPB causes late-onset spinal muscular atrophy and amyotrophic lateral sclerosis. *Am J Hum Genet.* 2004;75(5):822–31.
47. Greenway MJ, Andersen PM, Russ C, Ennis S, Cashman S, Donaghy C, et al. ANG mutations segregate with familial and 'sporadic' amyotrophic lateral sclerosis. *Nat Genet.* 2006;38(4):411–3.
48. Elden AC, Kim HJ, Hart MP, Chen-Plotkin AS, Johnson BS, Fang X, et al. Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature.* 2010;466(7310):1069–75.
49. Parkinson N, Ince PG, Smith MO, Highley R, Skibinski G, Andersen PM, et al. ALS phenotypes with mutations in CHMP2B (charged multivesicular body protein 2B). *Neurology.* 2006;67(6):1074–7.
50. Sabatelli M, Eusebi F, Al-Chalabi A, Conte A, Madia F, Luigetti M, et al. Rare missense variants of neuronal nicotinic acetylcholine receptor altering receptor function are associated with sporadic amyotrophic lateral sclerosis. *Hum Mol Genet.* 2009;18(20):3997–4006.
51. Couthouis J, Hart MP, Erion R, King OD, Diaz Z, Nakaya T, et al. Evaluating the role of the FUS/TLS-related gene EWSR1 in amyotrophic lateral sclerosis. *Hum Mol Genet.* 2012;21(13):2899–911.
52. Chow CY, Landers JE, Bergren SK, Sapp PC, Grant AE, Jones JM, et al. Deleterious variants of FIG4, a phosphoinositide phosphatase, in patients with ALS. *Am J Hum Genet.* 2009;84(1):85–8.
53. Aditi GL, Dawson TR, Wente SR. An amyotrophic lateral sclerosis-linked mutation in GLE1 alters the cellular pool of human Gle1 functional isoforms. *Adv Biol Regul.* 2016;62:25–36.
54. Al-Chalabi A, Andersen PM, Nilsson P, Chioza B, Andersson JL, Russ C, et al. Deletions of the heavy neurofilament subunit tail in amyotrophic lateral sclerosis. *Hum Mol Genet.* 1999;8(2):157–64.
55. Wills AM, Cronin S, Slowik A, Kasperaviciute D, Van Es MA, Morahan JM, et al. A large-scale international meta-analysis of paraoxonase gene polymorphisms in sporadic ALS. *Neurology.* 2009;73(1):16–24.
56. Leung CL, He CZ, Kaufmann P, Chin SS, Naini A, Liem RK, et al. A pathogenic peripherin gene mutation in a patient with amyotrophic lateral sclerosis. *Brain Pathol.* 2004;14(3):290–6.
57. Fecto F, Yan J, Vemula SP, Liu E, Yang Y, Chen W, et al. SQSTM1 mutations in familial and sporadic amyotrophic lateral sclerosis. *Arch Neurol.* 2011;68(11):1440–6.
58. Couthouis J, Hart MP, Shorter J, DeJesus-Hernandez M, Erion R, Oristano R, et al. A yeast functional screen predicts new candidate ALS disease genes. *Proc Natl Acad Sci U S A.* 2011;108(52):20881–90.
59. van Rheenen W, Shatunov A, Dekker AM, McLaughlin RL, Diekstra FP, Pullit SL, et al. Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. *Nat Genet.* 2016;48(9):1043–8.

60. Fogh I, Lin K, Tiloca C, Rooney J, Gellera C, Diekstra FP, et al. Association of a Locus in the CAMTA1 gene with survival in patients with sporadic amyotrophic lateral sclerosis. *JAMA Neurol.* 2016;73(7):812–20.
61. Nicolas A, Kenna KP, Renton AE, Ticozzi N, Faghri F, Chia R, et al. Genome-wide analyses identify KIF5A as a novel ALS gene. *Neuron.* 2018;97(6):1268–83 e6.
62. Tazelaar GHP, Dekker AM, van Vugt J, van der Spek RA, Westeneng HJ, Kool L, et al. Association of NIPA1 repeat expansions with amyotrophic lateral sclerosis in a large international cohort. *Neurobiol Aging.* 2019;74:234.e9–e15.
63. Gilley J, Jackson O, Pipis M, Estiar MA, Al-Chalabi A, Danzi MC, et al. Enrichment of SARM1 alleles encoding variants with constitutively hyperactive NADase in patients with ALS and other motor nerve disorders. *Elife.* 2021;10:e70905.
64. Mackenzie IR, Nicholson AM, Sarkar M, Messing J, Purice MD, Pottier C, et al. TIA1 mutations in amyotrophic lateral sclerosis and frontotemporal dementia promote phase separation and alter stress granule dynamics. *Neuron.* 2017;95(4):808–16 e9.
65. Diekstra FP, van Vught PW, van Rheenen W, Koppers M, Pasterkamp RJ, van Es MA, et al. UNC13A is a modifier of survival in amyotrophic lateral sclerosis. *Neurobiol Aging.* 2012;33(3):630.e3–8.
66. Chio A, Mora G, Restagno G, Brunetti M, Ossola I, Barberis M, et al. UNC13A influences survival in Italian amyotrophic lateral sclerosis patients: a population-based study. *Neurobiol Aging.* 2013;34(1):357.e1–5.
67. Smith BN, Topp SD, Fallini C, Shibata H, Chen HJ, Troakes C, et al. Mutations in the vesicular trafficking protein annexin A11 are associated with amyotrophic lateral sclerosis. *Sci Transl Med.* 2017;9(388):eaad9157.
68. Bannwarth S, Ait-El-Mkadem S, Chausseot A, Genin EC, Lacas-Gervais S, Fragaki K, et al. A mitochondrial origin for frontotemporal dementia and amyotrophic lateral sclerosis through CHCHD10 involvement. *Brain.* 2014;137(Pt 8):2329–45.
69. Chausseot A, Le Ber I, Ait-El-Mkadem S, Camuzat A, de Septenville A, Bannwarth S, et al. Screening of CHCHD10 in a French cohort confirms the involvement of this gene in frontotemporal dementia with amyotrophic lateral sclerosis patients. *Neurobiol Aging.* 2014;35(12):2884.e1–4.
70. Farhan SMK, Howrigan DP, Abbott LE, Klim JR, Topp SD, Byrnes AE, et al. Publisher correction: exome sequencing in amyotrophic lateral sclerosis implicates a novel gene, DNAJC7, encoding a heat-shock protein. *Nat Neurosci.* 2020;23(2):295.
71. Kenna KP, van Doormaal PT, Dekker AM, Ticozzi N, Kenna BJ, Diekstra FP, et al. NEK1 variants confer susceptibility to amyotrophic lateral sclerosis. *Nat Genet.* 2016;48(9):1037–42.
72. Brenner D, Muller K, Wieland T, Weydt P, Bohm S, Lule D, et al. NEK1 mutations in familial amyotrophic lateral sclerosis. *Brain.* 2016;139(Pt 5):e28.
73. Nguyen HP, Van Mossevelde S, Dillen L, De Bleecker JL, Moisse M, Van Damme P, et al. NEK1 genetic variability in a Belgian cohort of ALS and ALS-FTD patients. *Neurobiol Aging.* 2018;61:255.e1–7.
74. Shu S, Lei X, Liu F, Cui B, Liu Q, Ding Q, et al. Mutation screening of NEK1 in Chinese ALS patients. *Neurobiol Aging.* 2018;71:267.e1–4.
75. Wu CH, Fallini C, Ticozzi N, Keagle PJ, Sapp PC, Piotrowska K, et al. Mutations in the profilin 1 gene cause familial amyotrophic lateral sclerosis. *Nature.* 2012;488(7412):499–503.
76. Cirulli ET, Lasseigne BN, Petrovski S, Sapp PC, Dion PA, Leblond CS, et al. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science.* 2015;347(6229):1436–41.
77. Borghero G, Pugliatti M, Marrosu F, Marrosu MG, Murru MR, Floris G, et al. TBK1 is associated with ALS and ALS-FTD in Sardinian patients. *Neurobiol Aging.* 2016;43:180.e1–5.
78. Oakes JA, Davies MC, Collins MO. TBK1: a new player in ALS linking autophagy and neuroinflammation. *Mol Brain.* 2017;10(1):5.
79. Smith BN, Ticozzi N, Fallini C, Gkazi AS, Topp S, Kenna KP, et al. Exome-wide rare variant analysis identifies TUBA4A mutations associated with familial ALS. *Neuron.* 2014;84(2):324–31.
80. Pensato V, Tiloca C, Corrado L, Bertolin C, Sardone V, Del Bo R, et al. TUBA4A gene analysis in sporadic amyotrophic lateral sclerosis: identification of novel mutations. *J Neurol.* 2015;262(5):1376–8.
81. Johnson JO, Mandrioli J, Benatar M, Abramzon Y, Van Deerlin VM, Trojanowski JQ, et al. Exome sequencing reveals VCP mutations as a cause of familial ALS. *Neuron.* 2010;68(5):857–64.
82. Al-Saif A, Bohlega S, Al-Mohanna F. Loss of ERLIN2 function leads to juvenile primary lateral sclerosis. *Ann Neurol.* 2012;72(4):510–6.
83. Van Hoecke A, Schoonaert L, Lemmens R, Timmers M, Staats KA, Laird AS, et al. EPHA4 is a disease modifier of amyotrophic lateral sclerosis in animal models and in humans. *Nat Med.* 2012;18(9):1418–22.
84. Maruyama H, Morino H, Ito H, Izumi Y, Kato H, Watanabe Y, et al. Mutations of optineurin in amyotrophic lateral sclerosis. *Nature.* 2010;465(7295):223–6.
85. Lambrechts D, Poesen K, Fernandez-Santiago R, Al-Chalabi A, Del Bo R, Van Vught PW, et al. Meta-analysis of vascular endothelial growth factor variations in amyotrophic lateral sclerosis: increased susceptibility in male carriers of the -2578AA genotype. *J Med Genet.* 2009;46(12):840–6.
86. Vance C, Al-Chalabi A, Ruddy D, Smith BN, Hu X, Sreedharan J, et al. Familial amyotrophic lateral sclerosis with frontotemporal dementia is linked to a locus on chromosome 9p13.2–21.3. *Brain.* 2006;129(Pt 4):868–76.
87. Pearson JP, Williams NM, Majounie E, Waite A, Stott J, Newsway V, et al. Familial frontotemporal dementia with amyotrophic lateral sclerosis and a shared haplotype on chromosome 9p. *J Neurol.* 2011;258(4):647–55.
88. Morita M, Al-Chalabi A, Andersen PM, Hosler B, Sapp P, Englund E, et al. A locus on chromosome 9p confers susceptibility to ALS and frontotemporal dementia. *Neurology.* 2006;66(6):839–44.
89. Boxer AL, Mackenzie IR, Boeve BF, Baker M, Seeley WW, Crook R, et al. Clinical, neuroimaging and neuropathological features of a new chromosome 9p-linked FTD-ALS family. *J Neurol Neurosurg Psychiatry.* 2011;82(2):196–203.
90. Mok K, Traynor BJ, Schymick J, Tienari PJ, Laaksovirta H, Peuralinna T, et al. Chromosome 9 ALS and FTD locus is probably derived from a single founder. *Neurobiol Aging.* 2012;33(1):209.e3–8.
91. van Es MA, Veldink JH, Saris CG, Blauw HM, van Vught PW, Birve A, 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):1083–7.
92. Laaksovirta H, Peuralinna T, Schymick JC, Scholz SW, Lai SL, Myllykangas L, et al. Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: a genome-wide association study. *Lancet Neurol.* 2010;9(10):978–85.
93. Shatunov A, Mok K, Newhouse S, Weale ME, Smith B, Vance C, 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):986–94.
94. Van Deerlin VM, Sleiman PM, Martinez-Lage M, Chen-Plotkin A, Wang LS, Graff-Radford NR, et al. Common variants at 7p21 are associated with frontotemporal lobar degeneration with TDP-43 inclusions. *Nat Genet.* 2010;42(3):234–9.
95. International HapMap C, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature.* 2007;449(7164):851–61.
96. McLaughlin RL, Vajda A, Hardiman O. Heritability of amyotrophic lateral sclerosis: insights from disparate numbers. *JAMA Neurol.* 2015;72(8):857–8.
97. Luigetti M, Lattante S, Zollino M, Conte A, Marangi G, Del Grande A, et al. SOD1 G93D sporadic amyotrophic lateral sclerosis (SALS) patient with rapid progression and concomitant novel ANG variant. *Neurobiol Aging.* 2011;32(10):1924.e15–8.
98. Lamp M, Origone P, Geroldi A, Verdiani S, Gotta F, Caponnetto C, et al. Twenty years of molecular analyses in amyotrophic lateral sclerosis: genetic landscape of Italian patients. *Neurobiol Aging.* 2018;66:179.e5–e16.
99. Lee H, Tang H. Next-generation sequencing technologies and fragment assembly algorithms. *Methods Mol Biol.* 2012;855:155–74.
100. Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM. Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol.* 2007;8(7):R143.

101. Bentley DR. Whole-genome re-sequencing. *Curr Opin Genet Dev.* 2006;16(6):545–52.
102. Dolzhenko E, van Vugt J, Shaw RJ, Bekritsky MA, van Blitterswijk M, Narzisi G, et al. Detection of long repeat expansions from PCR-free whole-genome sequence data. *Genome Res.* 2017;27(11):1895–903.
103. Ionita-Laza I, McCallum K, Xu B, Buxbaum JD. A spectral approach integrating functional genomic annotations for coding and noncoding variants. *Nat Genet.* 2016;48(2):214–20.
104. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *Am J Hum Genet.* 2016;99(4):877–85.
105. Listgarten J, Lippert C, Heckerman D. FaST-LMM-select for addressing confounding from spatial structure and rare variants. *Nat Genet.* 2013;45(5):470–1.
106. Freischmidt A, Wieland T, Richter B, Ruf W, Schaeffer V, Muller K, et al. Haploinsufficiency of TBK1 causes familial ALS and fronto-temporal dementia. *Nat Neurosci.* 2015;18(5):631–6.
107. van Rheenen W, van der Spek RAA, Bakker MK, van Vugt J, Hop PJ, Zwamborn RAJ, et al. Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. *Nat Genet.* 2021;53(12):1636–48.
108. Al Khleifat A, Iacoangeli A, van Vugt J, Bowles H, Moisse M, Zwamborn RAJ, et al. Structural variation analysis of 6,500 whole genome sequences in amyotrophic lateral sclerosis. *NPJ Genom Med.* 2022;7(1):8.
109. Ebbert MTW, Farrugia SL, Sens JP, Jansen-West K, Gendron TF, Prudencio M, et al. Long-read sequencing across the C9orf72 'GGGGCC' repeat expansion: implications for clinical use and genetic discovery efforts in human disease. *Mol Neurodegener.* 2018;13(1):46.
110. Cameron DL, Di Stefano L, Papenfuss AT. Comprehensive evaluation and characterisation of short read general-purpose structural variant calling software. *Nat Commun.* 2019;10(1):3240.
111. Mousavi N, Shleizer-Burko S, Yanicky R, Gymrek M. Profiling the genome-wide landscape of tandem repeat expansions. *Nucleic Acids Res.* 2019;47(15):e90.
112. Willems T, Zielinski D, Yuan J, Gordon A, Gymrek M, Erlich Y. Genome-wide profiling of heritable and de novo STR variations. *Nat Methods.* 2017;14(6):590–2.
113. Theunissen F, Flynn LL, Anderton RS, Mastaglia F, Pytte J, Jiang L, et al. Structural variants may be a source of missing heritability in sALS. *Front Neurosci.* 2020;14:47.
114. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing from single polymerase molecules. *Science.* 2009;323(5910):133–8.
115. Levene MJ, Korlach J, Turner SW, Foquet M, Craighead HG, Webb WW. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science.* 2003;299(5607):682–6.
116. Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.* 2016;17(1):239.
117. Jain M, Fiddes IT, Miga KH, Olsen HE, Paten B, Akeson M. Improved data analysis for the MinION nanopore sequencer. *Nat Methods.* 2015;12(4):351–6.
118. Carneiro MO, Russ C, Ross MG, Gabriel SB, Nusbaum C, DePristo MA. Pacific biosciences sequencing technology for genotyping and variation discovery in human data. *BMC Genomics.* 2012;13:375.
119. Roberts RJ, Carneiro MO, Schatz MC. The advantages of SMRT sequencing. *Genome Biol.* 2013;14(7):405.
120. Wenger AM, Peluso P, Rowell WJ, Chang PC, Hall RJ, Concepcion GT, et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat Biotechnol.* 2019;37(10):1155–62.
121. Pacific Biosciences- Revio. Available from: <https://www.pacb.com/revio/>. Accessed 23 Nov 2022.
122. Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol.* 2018;36(4):338–45.
123. Deamer D, Akeson M, Branton D. Three decades of nanopore sequencing. *Nat Biotechnol.* 2016;34(5):518–24.
124. de Lannoy C, de Ridder D, Risse J. The long reads ahead: de novo genome assembly using the MinION. *F1000Res.* 2017;6:1083.
125. Bowden R, Davies RW, Heger A, Pagnamenta AT, de Cesare M, Oikkonen LE, et al. Sequencing of human genomes with nanopore technology. *Nat Commun.* 2019;10(1):1869.
126. Oxford Nanopore Technologies - Minlon. <https://nanoporetech.com/products/minion>. Accessed 26 Sept 2022.
127. Payne A, Holmes N, Rakyar V, Loose M. BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics.* 2019;35(13):2193–8.
128. Oxford Nanopore Technologies - GridION. Available from: <https://nanoporetech.com/products/gridion>. Cited 2022 26 September.
129. Kim HS, Jeon S, Kim C, Kim YK, Cho YS, Kim J, et al. Chromosome-scale assembly comparison of the Korean Reference Genome KOREF from PromethION and PacBio with Hi-C mapping information. *Gigascience.* 2019;8(12):giz125.
130. Oxford Nanopore Technologies – PromethION. Available from: <https://nanoporetech.com/products/promethion>. Cited 2022 26 September.
131. Rang FJ, Kloosterman WP, de Ridder J. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biol.* 2018;19(1):90.
132. Oxford Nanopore Technologies- Accuracy of reads. Available from: <https://nanoporetech.com/accuracy>. Cited 2022 26 September.
133. Wang B, Tseng E, Regulski M, Clark TA, Hon T, Jiao Y, et al. Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nat Commun.* 2016;7:11708.
134. Berlin K, Koren S, Chin CS, Drake JP, Landolin JM, Phillippy AM. Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. *Nat Biotechnol.* 2015;33(6):623–30.
135. Schmidt MH, Vogel A, Denton AK, Istace B, Wormit A, van de Geest H, et al. De novo assembly of a new *Solanum pennellii* accession using nanopore sequencing. *Plant Cell.* 2017;29(10):2336–48.
136. Shi L, Guo Y, Dong C, Huddleston J, Yang H, Han X, et al. Long-read sequencing and de novo assembly of a Chinese genome. *Nat Commun.* 2016;7:12065.
137. Miga KH, Koren S, Rhie A, Vollger MR, Gershman A, Bzikadze A, et al. Telomere-to-telomere assembly of a complete human X chromosome. *Nature.* 2020;585(7823):79–84.
138. Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, et al. The complete sequence of a human genome. *Science.* 2022;376(6588):44–53.
139. Chaisson MJ, Huddleston J, Dennis MY, Sudmant PH, Malig M, Hormozdiari F, et al. Resolving the complexity of the human genome using single-molecule sequencing. *Nature.* 2015;517(7536):608–11.
140. Huddleston J, Chaisson MJP, Steinberg KM, Warren W, Hoekzema K, Gordon D, et al. Discovery and genotyping of structural variation from long-read haploid genome sequence data. *Genome Res.* 2017;27(5):677–85.
141. Chaisson MJP, Sanders AD, Zhao X, Malhotra A, Porubsky D, Rausch T, et al. Multi-platform discovery of haplotype-resolved structural variation in human genomes. *Nat Commun.* 2019;10(1):1784.
142. Ebert P, Audano PA, Zhu Q, Rodriguez-Martin B, Porubsky D, Bonder MJ, et al. Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Science.* 2021;372(6537):eabf7117.
143. Ebbert MTW, Jensen TD, Jansen-West K, Sens JP, Reddy JS, Ridge PG, et al. Systematic analysis of dark and camouflaged genes reveals disease-relevant genes hiding in plain sight. *Genome Biol.* 2019;20(1):97.
144. Merker JD, Wenger AM, Sneddon T, Grove M, Zappala Z, Fresard L, et al. Long-read genome sequencing identifies causal structural variation in a Mendelian disease. *Genet Med.* 2018;20(1):159–63.
145. Dutta UR, Rao SN, Pidugu VK, Vineeth VS, Bhattacherjee A, Bhowmik AD, et al. Breakpoint mapping of a novel de novo translocation t(X;20)(q11.1;p13) by positional cloning and long read sequencing. *Genomics.* 2019;111(5):1108–14.
146. Ishiura H, Doi K, Mitsui J, Yoshimura J, Matsukawa MK, Fujiyama A, et al. Expansions of intronic TTTC and TTTTA repeats in benign adult familial myoclonic epilepsy. *Nat Genet.* 2018;50(4):581–90.
147. Zeng S, Zhang MY, Wang XJ, Hu ZM, Li JC, Li N, et al. Long-read sequencing identified intronic repeat expansions in SAMD12 from

- Chinese pedigrees affected with familial cortical myoclonic tremor with epilepsy. *J Med Genet.* 2019;56(4):265–70.
148. Mizuguchi T, Suzuki T, Abe C, Umemura A, Tokunaga K, Kawai Y, et al. A 12-kb structural variation in progressive myoclonic epilepsy was newly identified by long-read whole-genome sequencing. *J Hum Genet.* 2019;64(5):359–68.
 149. Vialle RA, de Paiva Lopes K, Bennett DA, Crary JF, Raj T. Integrating whole-genome sequencing with multi-omic data reveals the impact of structural variants on gene regulation in the human brain. *Nat Neurosci.* 2022;25(4):504–14.
 150. Yau WY, Vandrovicova J, Sullivan R, Chen Z, Zecchinelli A, Cilia R, et al. Low prevalence of NOTCH2NL GGC repeat expansion in white patients with movement disorders. *Mov Disord.* 2021;36(1):251–5.
 151. Loomis EW, Eid JS, Peluso P, Yin J, Hickey L, Rank D, et al. Sequencing the unsequenceable: expanded CGG-repeat alleles of the fragile X gene. *Genome Res.* 2013;23(1):121–8.
 152. Ardui S, Race V, de Ravel T, Van Esch H, Devriendt K, Matthijs G, et al. Detecting AGG interruptions in females with a FMR1 premutation by long-read single-molecule sequencing: a 1 year clinical experience. *Front Genet.* 2018;9:150.
 153. Tian Y, Wang JL, Huang W, Zeng S, Jiao B, Liu Z, et al. Expansion of human-specific GGC repeat in neuronal Intranuclear inclusion disease-related disorders. *Am J Hum Genet.* 2019;105(1):166–76.
 154. Tsai YC, de Pontual L, Heiner C, Stojkovic T, Furling D, Bassez G, et al. Identification of a CCG-enriched expanded allele in DM1 patients using amplification-free long-read sequencing. *J Mol Diagn.* 2022;24(11):114.
 155. Hoijer I, Tsai YC, Clark TA, Kotturi P, Dahl N, Stattin EL, et al. Detailed analysis of HTT repeat elements in human blood using targeted amplification-free long-read sequencing. *Hum Mutat.* 2018;39(9):1262–72.
 156. Pacific Biosciences- Targeted sequencing. Available from: <https://www.pacb.com/products-and-services/applications/targeted-sequencing/>. Cited 2022 26 September.
 157. DeJesus-Hernandez M, Aleff RA, Jackson JL, Finch NA, Baker MC, Gendron TF, et al. Long-read targeted sequencing uncovers clinicopathological associations for C9orf72-linked diseases. *Brain.* 2021;144(4):1082–8.
 158. Course MM, Gudsruk K, Smukowski SN, Winston K, Desai N, Ross JP, et al. Evolution of a human-specific tandem repeat associated with ALS. *Am J Hum Genet.* 2020;107(3):445–60.
 159. Tsai Y-C, Zafar F, McEachin ZT, McLaughlin I, Blitterswijk M, Ziegler J, et al. Multiplex CRISPR/Cas9-Guided No-Amp targeted sequencing panel for spinocerebellar ataxia repeat expansions. *Neuromethods.* 2022;182: Springer Nature.
 160. Morato Torres CA, Zafar F, Tsai YC, Vazquez JP, Gallagher MD, McLaughlin I, et al. ATTCT and ATTCC repeat expansions in the ATXN10 gene affect disease penetrance of spinocerebellar ataxia type 10. *HGG Adv.* 2022;3(4):100137.
 161. Pulst SM, Nechiporuk A, Nechiporuk T, Gispert S, Chen XN, Lopes-Cendes I, et al. Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. *Nat Genet.* 1996;14(3):269–76.
 162. Van Damme P, Veldink JH, van Blitterswijk M, Corveleyn A, van Vught PW, Thijs V, et al. Expanded ATXN2 CAG repeat size in ALS identifies genetic overlap between ALS and SCA2. *Neurology.* 2011;76(24):2066–72.
 163. van Blitterswijk M, Mullen B, Heckman MG, Baker MC, DeJesus-Hernandez M, Brown PH, et al. Ataxin-2 as potential disease modifier in C9ORF72 expansion carriers. *Neurobiol Aging.* 2014;35(10):2421.e13–7.
 164. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods.* 2010;7(6):461–5.
 165. Laszlo AH, Derrington IM, Brinkerhoff H, Langford KW, Nova IC, Samson JM, et al. Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. *Proc Natl Acad Sci U S A.* 2013;110(47):18904–9.
 166. Adewale BA. Will long-read sequencing technologies replace short-read sequencing technologies in the next 10 years? *Afr J Lab Med.* 2020;9(1):1340.
 167. Ling JP, Pletnikova O, Troncoso JC, Wong PC. TDP-43 repression of nonconserved cryptic exons is compromised in ALS-FTD. *Science.* 2015;349(6248):650–5.
 168. Chiang PM, Ling J, Jeong YH, Price DL, Aja SM, Wong PC. Deletion of TDP-43 down-regulates Tbc1d1, a gene linked to obesity, and alters body fat metabolism. *Proc Natl Acad Sci U S A.* 2010;107(37):16320–4.
 169. Polymenidou M, Lagier-Tourenne C, Hutt KR, Huelga SC, Moran J, Liang TY, et al. Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. *Nat Neurosci.* 2011;14(4):459–68.
 170. Lagier-Tourenne C, Polymenidou M, Hutt KR, Vu AQ, Baughn M, Huelga SC, et al. Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. *Nat Neurosci.* 2012;15(11):1488–97.
 171. Bandyopadhyay U, Cotney J, Nagy M, Oh S, Leng J, Mahajan M, et al. RNA-Seq profiling of spinal cord motor neurons from a presymptomatic SOD1 ALS mouse. *PLoS One.* 2013;8(1):e53575.
 172. Prudencio M, Belzil VV, Batra R, Ross CA, Gendron TF, Pregent LJ, et al. Distinct brain transcriptome profiles in C9orf72-associated and sporadic ALS. *Nat Neurosci.* 2015;18(8):1175–82.
 173. Prudencio M, Gonzales PK, Cook CN, Gendron TF, Daugherty LM, Song Y, et al. Repetitive element transcripts are elevated in the brain of C9orf72 ALS/FTLD patients. *Hum Mol Genet.* 2017;26(17):3421–31.
 174. Dickson DW, Baker MC, Jackson JL, DeJesus-Hernandez M, Finch NA, Tian S, et al. Extensive transcriptomic study emphasizes importance of vesicular transport in C9orf72 expansion carriers. *Acta Neuropathol Commun.* 2019;7(1):150.
 175. Tam OH, Rozhkov NV, Shaw R, Kim D, Hubbard I, Fennessey S, et al. Postmortem cortex samples identify distinct molecular subtypes of ALS: retrotransposon activation, oxidative stress, and activated glia. *Cell Rep.* 2019;29(5):1164–77 e5.
 176. Prudencio M, Humphrey J, Pickles S, Brown AL, Hill SE, Kachergus JM, et al. Truncated stathmin-2 is a marker of TDP-43 pathology in fronto-temporal dementia. *J Clin Invest.* 2020;130(11):6080–92.
 177. Klim JR, Williams LA, Limone F, Guerra San Juan I, Davis-Dusenbery BN, Mordes DA, et al. ALS-implicated protein TDP-43 sustains levels of STMN2, a mediator of motor neuron growth and repair. *Nat Neurosci.* 2019;22(2):167–79.
 178. Melamed Z, Lopez-Erauskin J, Baughn MW, Zhang O, Drenner K, Sun Y, et al. Premature polyadenylation-mediated loss of stathmin-2 is a hallmark of TDP-43-dependent neurodegeneration. *Nat Neurosci.* 2019;22(2):180–90.
 179. Ma XR, Prudencio M, Koike Y, Vatsavayai SC, Kim G, Harbinski F, et al. TDP-43 represses cryptic exon inclusion in the FTD-ALS gene UNC13A. *Nature.* 2022;603(7899):124–30.
 180. Brown AL, Wilkins OG, Keuss MJ, Hill SE, Zanovello M, Lee WC, et al. TDP-43 loss and ALS-risk SNPs drive mis-splicing and depletion of UNC13A. *Nature.* 2022;603(7899):131–7.
 181. Mathys H, Davila-Velderrain J, Peng Z, Gao F, Mohammadi S, Young JZ, et al. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature.* 2019;570(7761):332–7.
 182. Saez-Atienzar S, Bandres-Ciga S, Langston RG, Kim JJ, Choi SW, Reynolds RH, et al. Genetic analysis of amyotrophic lateral sclerosis identifies contributing pathways and cell types. *Sci Adv.* 2021;7(3):eabd9036.
 183. Maniatis S, Ajo T, Vickovic S, Braine C, Kang K, Mollbrink A, et al. Spatiotemporal dynamics of molecular pathology in amyotrophic lateral sclerosis. *Science.* 2019;364(6435):89–93.
 184. Gregory JM, McDade K, Livesey MR, Croy I, Marion de Proce S, Aitman T, et al. Spatial transcriptomics identifies spatially dysregulated expression of GRM3 and USP47 in amyotrophic lateral sclerosis. *Neuropathol Appl Neurobiol.* 2020;46(5):441–57.
 185. Chen WT, Lu A, Craessaerts K, Pavie B, Sala Frigerio C, Corthout N, et al. Spatial transcriptomics and in situ sequencing to study Alzheimer's disease. *Cell.* 2020;182(4):976–91 e19.
 186. Leung SK, Jeffries AR, Castanho I, Jordan BT, Moore K, Davies JP, et al. Full-length transcript sequencing of human and mouse cerebral cortex identifies widespread isoform diversity and alternative splicing. *Cell Rep.* 2021;37(7):110022.
 187. Abston ED, Coronado MJ, Bucek A, Bedja D, Shin J, Kim JB, et al. Th2 regulation of viral myocarditis in mice: different roles for TLR3 versus TRIF in progression to chronic disease. *Clin Dev Immunol.* 2012;2012:129486.

188. Workman RE, Tang AD, Tang PS, Jain M, Tyson JR, Razaghi R, et al. Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat Methods*. 2019;16(12):1297–305.
189. Sharon D, Tilgner H, Grubert F, Snyder M. A single-molecule long-read survey of the human transcriptome. *Nat Biotechnol*. 2013;31(11):1009–14.
190. Deveson IW, Brunck ME, Blackburn J, Tseng E, Hon T, Clark TA, et al. Universal alternative splicing of noncoding exons. *Cell Syst*. 2018;6(2):245–55.e5.
191. Byrne A, Beaudin AE, Olsen HE, Jain M, Cole C, Palmer T, et al. Nanopore long-read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells. *Nat Commun*. 2017;8:16027.
192. Tilgner H, Grubert F, Sharon D, Snyder MP. Defining a personal, allele-specific, and single-molecule long-read transcriptome. *Proc Natl Acad Sci U S A*. 2014;111(27):9869–74.
193. Oikonomopoulos S, Wang YC, Djambazian H, Badescu D, Ragoussis J. Benchmarking of the Oxford Nanopore MinION sequencing for quantitative and qualitative assessment of cDNA populations. *Sci Rep*. 2016;6:31602.
194. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*. 2013;10(12):1213–8.
195. Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, et al. Genome-wide location and function of DNA binding proteins. *Science*. 2000;290(5500):2306–9.
196. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009;326(5950):289–93.
197. Baxi EG, Thompson T, Li J, Kaye JA, Lim RG, Wu J, et al. Answer ALS, a large-scale resource for sporadic and familial ALS combining clinical and multi-omics data from induced pluripotent cell lines. *Nat Neurosci*. 2022;25(2):226–37.
198. Neuro LC, Li J, Lim RG, Kaye JA, Dardov V, Coyne AN, et al. An integrated multi-omic analysis of iPSC-derived motor neurons from C9ORF72 ALS patients. *iScience*. 2021;24(11):103221.
199. Zhang S, Cooper-Knock J, Weimer AK, Shi M, Moll T, Marshall JNG, et al. Genome-wide identification of the genetic basis of amyotrophic lateral sclerosis. *Neuron*. 2022;110(6):992–1008.e11.
200. Aulchenko YS, de Koning DJ, Haley C. Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. *Genetics*. 2007;177(1):577–85.
201. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet*. 2016;48(5):481–7.
202. Park S, Kim D, Song J, Joo JWJ. An integrative transcriptome-wide analysis of amyotrophic lateral sclerosis for the identification of potential genetic markers and drug candidates. *Int J Mol Sci*. 2021;22(6):3216.
203. Xiao L, Yuan Z, Jin S, Wang T, Huang S, Zeng P. Multiple-tissue integrative transcriptome-wide association studies discovered new genes associated with amyotrophic lateral sclerosis. *Front Genet*. 2020;11:587243.
204. Project Min EALS. Project MinE: study design and pilot analyses of a large-scale whole-genome sequencing study in amyotrophic lateral sclerosis. *Eur J Hum Genet*. 2018;26(10):1537–46.
205. Clinical Research in ALS and Related Disorders for Therapeutic Development (CReATe) Consortium. Available from: <https://www1.rarediseasesesnetwork.org/cms/create>. Cited 2022 26 September.
206. Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. *Nat Rev Genet*. 2015;16(6):321–32.
207. Reel PS, Reel S, Pearson E, Trucco E, Jefferson E. Using machine learning approaches for multi-omics data analysis: a review. *Biotechnol Adv*. 2021;49:107739.
208. Euesden J, Lewis CM, O'Reilly PF. PRSice: polygenic risk score software. *Bioinformatics*. 2015;31(9):1466–8.
209. Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet*. 2013;9(3):e1003348.
210. Lewis ACF, Green RC. Polygenic risk scores in the clinic: new perspectives needed on familiar ethical issues. *Genome Med*. 2021;13(1):14.
211. Mavaddat N, Pharoah PD, Michailidou K, Tyrer J, Brook MN, Bolla MK, et al. Prediction of breast cancer risk based on profiling with common genetic variants. *J Natl Cancer Inst*. 2015;107(5):djv036.
212. Inouye M, Abraham G, Nelson CP, Wood AM, Sweeting MJ, Dudbridge F, et al. Genomic risk prediction of coronary artery disease in 480,000 adults: implications for primary prevention. *J Am Coll Cardiol*. 2018;72(16):1883–93.
213. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*. 2018;50(9):1219–24.
214. Restuadi R, Garton FC, Benyamin B, Lin T, Williams KL, Vinkhuyzen A, et al. Polygenic risk score analysis for amyotrophic lateral sclerosis leveraging cognitive performance, educational attainment and schizophrenia. *Eur J Hum Genet*. 2022;30(5):532–9.
215. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet*. 2019;51(4):584–91.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

