

RESEARCH

Open Access



# Long-read sequencing identifies *ATXN3* repeat expansions, and transcriptomics reveals disease progression biomarkers and druggable targets for spinocerebellar ataxia type 3

Chang Liu<sup>1†</sup>, Xin Wang<sup>2,3†</sup>, Chao Xu<sup>2,4†</sup>, Xiaoxiang Liu<sup>5</sup>, Liyan Ke<sup>2,6</sup>, Ying Li<sup>2,7</sup>, Hang Zhang<sup>2,6</sup>, Jianqiang Tan<sup>8</sup>, Senwei Tan<sup>2,3</sup>, Zitong Zhang<sup>2,6</sup>, Liang Cheng<sup>2,3\*</sup>, Yaqiong Ren<sup>9,10\*</sup> and Lei Shi<sup>2,6\*</sup>

## Abstract

**Background** Hereditary ataxias (HAs) are neurodegenerative disorders characterized by progressive cerebellar degeneration, with autosomal dominant spinocerebellar ataxias (SCAs) representing the most prevalent subtype. SCA3, the most common form worldwide, is caused by CAG repeat expansions in *ATXN3*, resulting in pathogenic ataxin-3 aggregation. However, the underlying molecular mechanisms driving disease progression remain incompletely understood.

**Methods** We utilized an integrated multi-omics strategy to investigate a five-generation Chinese HA pedigree. Genetic analyses included targeted ataxia panel sequencing (TS), whole-exome sequencing (WES), and long-read whole-genome sequencing (LR-WGS) of blood-derived DNA to identify causal variants and confirm diagnosis. Transcriptomic profiling revealed disease-associated gene expression signatures, followed by functional annotation and cross-species validation. To ensure analytical rigor, we further validated our bioinformatic pipeline using an independent ulcerative colitis (UC) dataset.

**Results** Genetic analysis identified pathogenic *ATXN3*-CAG repeat expansions that co-segregated with clinical symptoms in affected family members. Transcriptomic profiling showed significant enrichment in ECM-receptor interaction and focal adhesion pathways, along with immune dysregulation and RNA splicing defects associated with disease progression. Cross-species analysis discovered conserved blood biomarkers (*C3/ALS2/SLC35A2*↓ and

<sup>†</sup>Chang Liu, Xin Wang and Chao Xu contributed equally to this work.

\*Correspondence:

Liang Cheng  
liangcheng@hrbmu.edu.cn

Yaqiong Ren  
ryq327@163.com

Lei Shi  
shilei\_0328@hrbmu.edu.cn

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

*THBS1/CAMTA1*↑), strongly correlated with clinical progression. Protein-protein interaction network emphasized AKT1 as a central regulator, along with other key hubs (e.g., TGFB1, MAPK3, CALM3, APP), while brain-specific analyses highlighted Mobp, Mal, Gja1 and Klk6 as potential therapeutic targets.

**Conclusions** This study genetically confirms SCA3 in a Chinese pedigree using LR-WGS, overcoming the diagnostic limitations of short-read sequencing. Comprehensive analyses revealed conserved SCA3 progression signatures with potential biomarkers for future non-invasive monitoring. Mechanistically, this study identified dysregulation in ECM-receptor interaction/focal adhesion, immune response, and RNA splicing as key pathogenic contributors. These findings provide both actionable therapeutic targets and demonstrate the clinical utility of integrated multi-omics approaches for SCA3 diagnosis and patient stratification, with broader implications for repeat expansion disorders.

**Trial registration** Not Applicable.

**Keywords** Spinocerebellar Ataxia Type 3, Machado-Joseph Disease, SCA3, MJD, ATXN3, Repeat Expansion, Long-read Whole-genome Sequencing, Biomarker

## Introduction

Hereditary ataxia (HA) represents an extensively wide group of clinically and genetically heterogeneous neurodegenerative and movement disorders characterized by progressive dysfunction of the cerebellum and degeneration of spinocerebellar tracts and the spinal cord, with spinocerebellar ataxia (SCA) being the most common autosomal dominant subtype [1, 2]. Globally, the prevalence of dominant HA/SCA ranges from 0.0 to 5.6 per 100,000 individuals ( $0.0\text{--}5.6/10^5$ ), with an average estimate of  $2.7/10^5$  ( $1.5\text{--}4.0/10^5$ ) and higher rates observed in isolated populations [3, 4]. SCA's epidemiology varies greatly across regions, depending on factors such as historical ancestries, population migration, founder effects, and regional diagnostic capabilities [5–8]. SCA3, or Machado-Joseph disease (MJD), stands out as the most prevalent SCA worldwide, with a striking geographical distribution driven by founder effects [5–8]. SCA3 is very frequent in Portugal, particularly in the Azores Islands/Flores Island [4, 9], and also high frequent in Brazil [7, 10, 11], China [12, 13], Germany, and the Netherlands [9]. By contrast, SCA3 is less frequent in Mexico, Peru, Australia, India, South Africa, and Italy [5, 7, 9, 11, 13]. The regional clustering of population prevalence underscores the impact of historical migration and genetic drift on SCA3 dissemination. The uneven global distribution emphasizes the importance of targeted genetic screening in high-risk areas, such as Portugal and East Asia, to enable early diagnosis and intervention.

The diagnosis of SCA integrates regional epidemiology, clinical evaluation (family history), laboratory assessment, and molecular genetic testing. Clinicians assess characteristic ataxia symptoms such as gait disturbances, poor coordination, and dysarthria, and autosomal dominant inheritance patterns [14], with definitive confirmation through the detection of pathological variants [14, 15]. According to the OMIM database (Phenotypic Series - PS164400), there are up to 40 genes and 45 loci associated with 48 SCAs, most of them belonging to nucleotide

repeat expansion disorders (REDs). Next-generation sequencing (NGS) has improved diagnosis of inherited ataxia, but its short-read lengths usually fail to accurately detect large repeat expansions, leading to underestimation and diagnostic uncertainty [16]. Long-read whole-genome sequencing (LR-WGS) could overcome these limitations by precisely measuring repeat expansion length and identifying complex structural variants (SVs) and copy number variants (CNVs) undetectable by conventional NGS, significantly enhancing diagnostic accuracy for SCAs [17–19].

The pathophysiology of SCA3 stems from abnormal CAG repeat expansion in the *ATXN3* gene, leading to production of ataxin-3 protein with an extended polyglutamine (polyQ) tract that misfolds and then forms neuronal aggregates [5, 20]. These toxic inclusions predominantly accumulate in the cerebellum, brainstem, and spinal cord, disrupting neuronal function and causing widespread neurodegeneration [5, 20]. The nucleotide repeat disorders like SCA3 exhibit genetic anticipation, with longer tandem repeats correlating with earlier onset and increased severity across generations [5]. Normally, ataxin-3 functions as a deubiquitinating enzyme that maintains protein homeostasis by cleaving ubiquitin chains and aiding misfolded proteins for proteasomal degradation. However, the expanded polyQ of ataxin-3 impairs this function, resulting in accumulation of toxic protein aggregates [21]. These aggregates disrupt multiple cellular processes like protein degradation, mitochondrial function, and neuronal pathways, ultimately leading to neuronal death and progressive neurodegeneration [5, 20, 22].

Transcriptomic profiling through RNA sequencing (RNA-seq) have advanced our understanding of the molecular mechanisms of disease and identified potential molecular biomarkers for monitoring SCA3 disease progression. Studies have revealed dysregulated genes in the blood of SCA3 patients, with their expression correlating to disease severity [23–25]. These findings highlight the

utility of less-invasive and transcriptional blood biomarkers for tracking brain disease progression and evaluating therapeutic efficacy, offering valuable tools for basic research and clinical management.

In this study, we conducted comprehensive clinical and molecular genetic investigations on a Chinese pedigree presenting with suspected hereditary cerebellar ataxia. By employing multi-omic approaches including advanced sequencing technologies and transcriptome profiling, we intended to: (1) elucidate the molecular etiology of this family, (2) discover novel blood-based biomarkers for disease progression monitoring, and (3) explore key hub genes and potential therapeutic targets. Our integrated approach aims to improve diagnostic accuracy, facilitate early intervention, and contribute to the development of targeted therapies for this ataxia through cutting-edge genomic and transcriptomic analyses.

## Materials and methods

### Hereditary ataxia pedigree

We recruited a five-generation family with suspected hereditary ataxia from the Neurology Department of the First Traditional Chinese Medicine Hospital in Changde City, Hunan Province, China, comprising fourteen family members who underwent genetic testing (Fig. 1A). Symptomatic individuals received a standardized neurological workup like clinical examination, routine laboratory tests, brain magnetic resonance imaging (MRI), and electronystagmography (ENG). Additionally, three neuropsychological assessments — Zung's Self-Rating Anxiety Scale (SAS) [26], Self-Rating Depression Scale (SDS) [27], and the Symptom Checklist-90-Revised (SCL-90-R) [28] were administered. All participants provided informed consent for the use of clinical data, peripheral blood samples, and genetic analysis; and this study was approved by the ethics committees of the First Traditional Chinese Medicine Hospital and Harbin Medical University.

### Magnetic resonance imaging

MRI examinations were conducted on a 1.5 T Siemens MAGNETOM Avanto scanner using a dedicated head coil, following patient safety screening and standardized supine positioning with eyes closed. The imaging process commenced with a three-plane localizer (TR/TE = 20/5 ms, 20 s) for spatial planning, followed by diagnostic sequences: T1-weighted imaging (TR/TE = 500/15 ms, 2–5 min) for anatomical reference, T2-weighted (TR/TE = 4000/100 ms, 3–6 min) for pathological evaluation, fluid attenuated inversion recovery (FLAIR, TR/TI = 9000/2500 ms, 4–7 min) with cerebrospinal fluid suppression to enhance lesion conspicuity, diffusion-weighted imaging (DWI,  $b = 1000 \text{ s/mm}^2$ , 1–2 min) for acute ischemic detection, and susceptibility-weighted

imaging (SWI, Three-dimensional Gradient Recalled Echo, 3–5 min) for microhemorrhage and calcification assessment, with BLADE/PROPELLER motion correction applied throughout. Post-processing methods included multiplanar reconstruction (MPR) for anatomical correlation and maximum intensity projection (MIP) for vascular visualization, ensuring comprehensive diagnostic interpretation while maintaining optimal image quality.

### High-throughput sequencing

#### *Targeted, whole-exome, and long-read whole-genome sequencing*

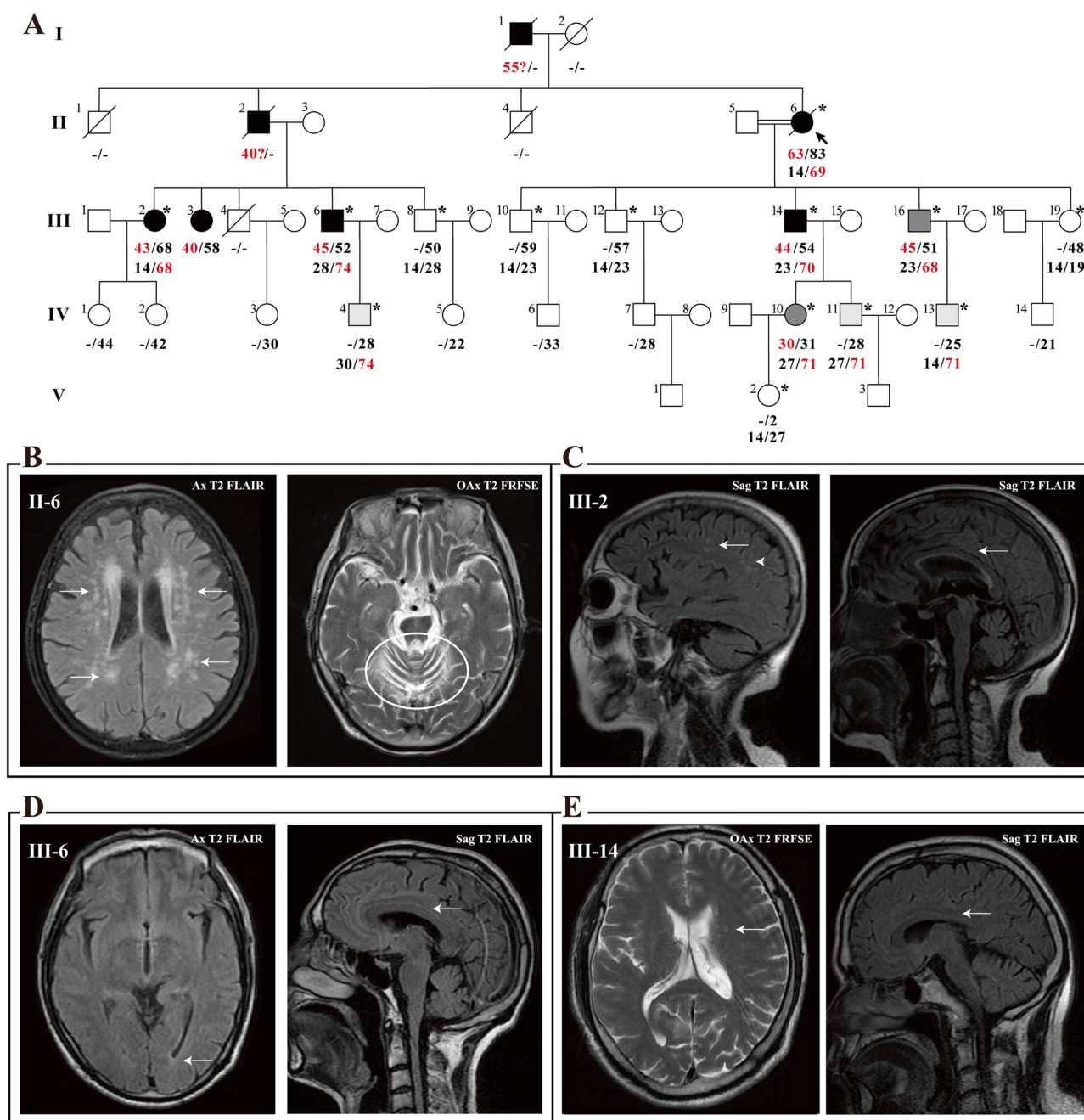
Genomic DNA was extracted from peripheral blood using the FlexiGene DNA Kit (QIAGEN, 51206). To identify causal variants, short-read ataxia targeted sequencing (TS, 1954 neurological disease-related genes) were performed on the proband II-6 and whole-exome sequencing (WES) were conducted on ten family members (II-6, III-2, III-6, III-8, III-10, III-12, III-14, III-16, III-19, and IV-10) using the Illumina NovaSeq 6000 System (Novogene, Beijing). Raw data were processed with Illumina bcl2fastq, aligned to the GRCh37/hs37d5 reference genome using Burrows-Wheeler Aligner (v0.7.8-r455) [29], and deduplicated using Sambamba (v0.6.6) [30]. The proband II-6 also underwent Oxford Nanopore LR-WGS (Biomarker, Beijing); raw signals were base-called with Guppy, quality-filtered, and aligned using Minimap2 [31, 32]. STR analysis was performed using RepeatHMM [33, 34], which employs a hidden Markov model and Gaussian mixture model-based peak detection for repeat count inference.

#### *Transcriptome sequencing*

Total RNA was extracted from peripheral blood samples, with integrity and quantity verified using an Agilent 2100 Bioanalyzer to ensure sequencing suitability. Sequencing libraries were prepared with the RNA Library Preparation Kit (NEB, E7770L) and subjected to paired-end sequencing ( $2 \times 150 \text{ bp}$ ) on the Illumina NovaSeq 6000 System (Novogene, Beijing). Raw FASTQ were quality-filtered to remove low-quality reads, and all subsequent analyses were performed using the high-quality clean data.

#### *Length of CAG repeat expansion*

To precisely quantify *ATXN3*-CAG repeat expansions, we utilized a multi-method approach combining direct PCR-Sanger sequencing, PCR-capillary electrophoresis, and T-A cloning followed by sequencing. PCR was performed using a FAM-labeled forward (5'-CCAGTGAC TACTTTGATTCGTGA-3') and reverse (5'-GAATGG TGAGCAGGCCTTAC-3') primers. Direct PCR-Sanger sequencing determined the short CAG repeats; while



**Fig. 1** The pedigree and MRI findings of the SCA3 cohort. **(A)** The pedigree identifies 14 tested members (asterisks) with color-coding: black (severe symptoms), dark gray (mild symptoms), and light gray (asymptomatic carriers). Age data (onset in red/current in black) and CAG repeat numbers (normal in black and expanded in red, formatted as Allele1/Allele2) are displayed. Fourteen members underwent molecular testing: IV-4, IV-11, IV-13, V-2 completed PCR-capillary electrophoresis and T-A cloning sequencing; ten others additionally had WES; the proband II-6 also received TS and LR-WGS. All 14 individuals contributed to RNA-seq. **(B-E)** MRI scans of severe individuals show: left panels — WM hyperintensities in the periventricular (**B,C,E**; arrows), subcortical (**C**; arrowhead), and left ventricular posterior horn (**D**) regions; right panels — atrophy in the cerebellar vermis/peduncles (**B**, circles) and posterior corpus callosum (**C-E**, arrows).

long repeat quantification was performed by capillary electrophoresis data, with expansion sizes calculated as: [flanking sequences + (CAG count  $\times$  3)]. Using the proband II-6's established flanking sequences (182 bp for expanded, 174 bp for normal alleles), we determined

repeat lengths for all affected individuals, enabling consistent measurement of both normal and pathogenic expansions throughout the pedigree. T-A cloning followed by sequencing validated the previous results.

### Data acquisition

To address the scarcity of blood transcriptome data from SCA3 patients across different disease stages, we incorporated mouse models for cross-species validation, comparing blood differentially expressed genes (DEGs), expression patterns, functional clusters, and protein-protein interaction (PPI) hub nodes associated with disease severity both in SCA3 family members and mouse models. This comparative approach significantly strengthened our findings, as human and mouse blood transcriptomes showed remarkable consistency in gene expression patterns. Mouse model selection and dataset utilization were guided by four criteria: (A) Pathological severity alignment based on age, genotype, and CAG repeat length; (B) Cross-species pathological comparability; (C) Availability of high-throughput sequencing (HTS) data for both blood and relevant brain regions in SCA3 and control mice; and (D) Inclusion of  $\geq$  five matched SCA3-control pairs in HTS analyses. Mouse datasets were publicly available at the NCBI GEO (<https://www.ncbi.nlm.nih.gov/geo/>; GSE145613, GSE107958, GSE108069; Table S2–Sheet1) [35, 36], while human blood RNA-seq data are provided in the appendix. All analyses were conducted using R (v4.3.1).

### Differentially expressed genes analysis

Family individuals were stratified into three severity groups: Severe [II-6, III-2, III-6, and III-14; Mean ( $\pm$  SD) current age was  $64.25 \pm 14.38$ ], Mild-Asymptomatic [III-16, IV-10, IV-4, IV-11, and IV-13; Mean ( $\pm$  SD) current age was  $32.60 \pm 10.50$ ], and Normal [III-8, III-10, III-12, and III-19; Mean ( $\pm$  SD) current age was  $53.50 \pm 5.32$ ]. Mouse severity classifications followed the original studies [35, 36]. Differential expression analysis was performed using DESeq2 (v1.40.2) [37] with significance thresholds set at  $p < 0.05$  and  $|\log_2FC| \geq 0.5$  for both human and mouse datasets.

### Cluster and enrichment analysis

Raw data were normalized as fragments per kilobase per million mapped reads (FPKM) for differential expression analysis. Following DEG identification, we performed mRNA expression trend clustering using ClusterGVis (v0.1.1) (<https://github.com/junjunlab/ClusterGVis>) and then conducted functional enrichment analyses (GO/KEGG) for each cluster. All analyses employed clusterProfiler (v4.8.2) [38] for enrichment and Gene Set Enrichment Analysis (GSEA), with gene ID conversion performed using org.Hs.eg.db (v3.17.0) [39] and org.Mm.eg.db (v3.17.0) [40] for human and mouse respectively, along with biomaRt (v2.56.1) [41] for cross-species gene mapping. The results were visualized using GseaVis (v0.0.9) (<https://github.com/junjunlab/GseaVis>).

### Molecular hub and druggability analysis

PPI networks were constructed using the STRING (confidence threshold: 0.4) (<https://cn.string-db.org/>) and visualized in Cytoscape (v3.10.2). Hub nodes were identified via cytoHubba's Degree algorithm, with the top ten genes classified as molecular hubs. Potential therapeutic targets were further screened using the Open Targets Platform (<https://platform.opentargets.org/>) and Therapeutic Target Database (<https://db.idrblab.net/ttd/>), prioritizing genes with existing clinical applications or trial evidence.

## Results

### Neurological, imaging, and neuropsychiatric abnormalities in the HA pedigree

The proband II-6 presented with a 7-year history of urinary urgency/frequency (recently worsened), progressive lower limb weakness, gait ataxia, dysarthria, and functional dependence, with a background of grade 3 hypertension (systolic blood pressure  $> 180$  mmHg, long-term use of amlodipine besylate) and posterior circulatory ischemia but no history of trauma, stroke, or psychiatric signs. This pedigree revealed consanguinity (II-6 and husband II-5 were first cousins) and multi-generational involvement (I-1; II-2, II-6; III-2, III-3, III-6, III-14, III-16; IV-10) with similar motor symptoms. Evaluation of six affected individuals (II-6, III-2, III-6, III-14, III-16, IV-10; Table S1–Sheet1) demonstrated cerebellar dysfunction, oculomotor abnormalities, and extrapyramidal signs, with the II-6, III-2, III-6, and III-14 exhibiting more severe manifestations than III-16 and IV-10. Earlier onset in successive generations (particularly on the right branch) suggested genetic anticipation (Fig. 1A). Routine lab tests (hematology, hepatic and renal function, and metabolic panels; Table S1–Sheet2) were unremarkable. These findings roughly correspond to SCA presentation.

Cranial MRI in the proband indicated cerebral small vessel disease (CSVD) and demyelinating lesions, characterized by punctate long T1/T2 signals and T2-FLAIR hyperintensities in periventricular white matter (WM), accompanied by mild posterior corpus callosum atrophy and severe cerebellar vermis/peduncle atrophy (Fig. 1B). These neuroimaging abnormalities were consistently observed across all five symptomatic patients (III-2, III-6, III-14, III-16, IV-10), with myelinopathy and varying degrees of cerebellar/callosal atrophy present in all cases, while CSVD features were absent only in III-6 (Fig. 1C–E, S1A; Table S1–Sheet1). Complementary ENG test revealed vestibular dysfunction and spontaneous nystagmus in four patients (III-2, III-6, III-14, III-16; Table S1–Sheet1). These findings further support central neurological involvement, despite the relative rarity of corpus callosum atrophy and CSVD in SCA spectrum disorders.

Five symptomatic patients subsequently underwent comprehensive neuropsychiatric evaluation using three

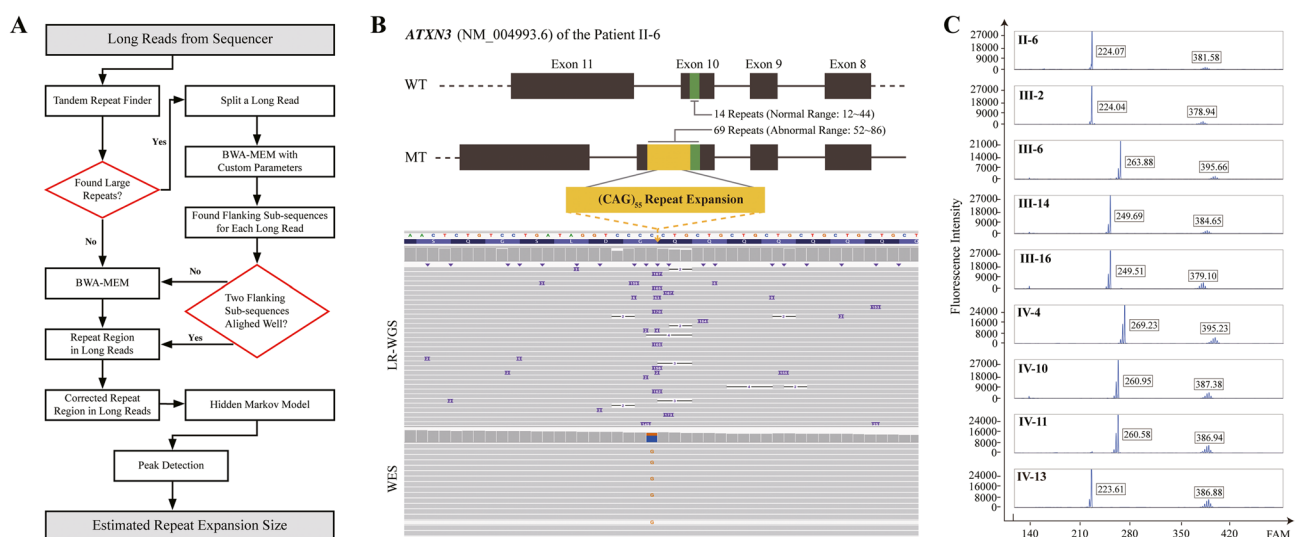
standardized scales. The Zung's SAS and SDS scales identified mild anxiety and depression in III-2 and mild depression in III-6, while the SCL-90-R scale revealed suboptimal physical and mental health states in III-2, III-6, and III-16 (Table S1–Sheet1). In addition, the clinical mental state score indicated moderate cognitive impairment in III-2, while mild cognitive impairment in III-6 and IV-10 (data not shown).

### Identification and validation of the abnormal *ATXN3* CAG repeat expansion

We implemented a sequential genetic testing strategy, beginning with targeted HA panel (having 1954 neurological disease-related genes and all known SCA genes through 2020) sequencing, which failed to discover pathogenic variants. Given the presence of unusual clinical features (e.g., corpus callosum atrophy and CSVD), incomplete genetic anticipation, and the absence of pathogenic findings on TS, we hypothesized a novel genetic basis for this familial ataxia. However, WES and linkage analysis across ten family members did not discover disease-causing variants or suggestive loci. LR-WGS ultimately identified a pathogenic 69-CAG repeat expansion in *ATXN3* exon of the proband II-6. This variant was missed by WES and TS probably due to limitations in exome capture and short-read variant calling (Fig. 2A–B). PCR-capillary electrophoresis confirmed *ATXN3*-CAG repeat expansions (68–74 repeats) cosegregating with ataxia phenotypes across affected members (Fig. 2C, S1B). Notably, asymptomatic carriers (IV-4, IV-11, IV-13) harbored 71–74 CAG repeats but had not reached disease onset age, while unaffected individuals

(III-8, III-10, III-12, III-19, V-2) showed normal repeat ranges (14–28 repeats; Table S1–Sheet3). T-A cloning and subsequent sequencing validated the expansions in nine patients (II-6, III-2, III-6, III-14, III-16, IV-10, IV-4, IV-11, IV-13; Fig. S1C). SCA3/MJD exhibits two distinct ancestral backgrounds: the Machado lineage, originally identified in the Azores Islands (Portugal) and predominantly found in Portuguese descendants; and the Joseph lineage, potentially of Asian origin with global distribution [42]. To determine the ancestral origin of *ATXN3*-CAG repeat expansion in this family, we integrated LR-WGS for comprehensive genome-wide coverage with WES for precise exon-adjacent region analysis. Haplotype phasing of ten family members was performed using a panel of 32 SNPs (6 core, 14 extended core, and 12 supplementary) [42–44]. The proband II-6 demonstrated 96.88% (31/32) genetic similarity to the Joseph lineage when considering all 32 SNPs, while analysis of the 6 core SNPs indicated that the five additional affected members also exhibited high genetic concordance with the Joseph lineage (Table S1–Sheet4). These findings strongly implicate a Joseph lineage origin for the pathogenic expansion in this family.

Using the model proposed by Tezenas du Montcel [45], we estimated the age of onset for SCA3 patients (Table S1–Sheet3) based on the published parameters ( $\alpha_0 = 7.4908$ ,  $\alpha_c = -0.0564$ ,  $\gamma_G = 0.2167$ ). The model predicted an earlier onset age for all patients except IV-10, whose prediction was later than the actual onset age, and IV-11 and IV-13 had not yet occurred (Fig. S1D). This may be due to differences in genetic background between races that may have an impact on prediction of disease onset



**Fig. 2** Identification and validation of *ATXN3* CAG repeat expansion. **(A)** The workflow outlines the LR-WGS pipeline. **(B)** LR-WGS and WES visualization reveal *ATXN3* CAG repeat expansion and rs12895357 heterozygous variant, with a schematic illustrating the regional exon structure of *ATXN3*. Color coding: green (normal repeats) and yellow (expanded repeats). **(C)** PCR-capillary electrophoresis results validate the repeat expansions across affected individuals, confirming the genetic cosegregation of the pathogenic variant.

age and the interactions between environment and genes. In addition, there may be certain errors in collecting data on patient age of onset like recall bias. Despite the above differences, this prediction is still of great value in early diagnosis of SCA3. This predictive model can provide clinicians with a reference for early risk assessment and help carry out disease monitoring and intervention in advance.

Upon re-examination of WES BAM files, we identified a heterozygous variant rs12895357 (Hg38, chr14:92071010, C>G; NM\_004993.6: c.916G>C, p.Gly306Arg; Fig. 2B) in the proband II-6, with a minor allele frequency of 10.62% for the G allele (reads ratio C40/G16). This variant was observed in all affected individuals, appearing in homozygous state in III-6 and IV-10, and heterozygous state in III-2, III-14, and III-16, as well as in one unaffected individual (III-8). However, the variant's location adjacent to the CAG repeat region of *ATXN3* raises technical concerns regarding the reliability of short-reads in this genomic context. Comprehensive evaluation of rs12895357 through multiple approaches failed to establish its pathogenic significance. Population frequency data (G allele = 0.1062) and absence of clinical associations in the ClinVar database suggest it represents a benign polymorphism. eQTLs analysis revealed no significant regulatory effects on *ATXN3* expression in the GTEx database. *In-silico* pathogenicity predictions yielded conclusive results of neutral variant, and 3D structural remodeling indicated the p.Gly306Arg substitution would not substantially perturb protein function. Analysis of other reported SCA3 modifier variants similarly failed to demonstrate significant cosegregation with disease status (disease onset and severity) or consistent phenotypic effects within this pedigree (Table S1–Sheet5). The potential cumulative impact of these variants remains uncertain given current sample size limitations and the complex genetic architecture of phenotypic modulation in SCA3. These findings suggest that while rs12895357 and other candidate variants were detected, their contribution to SCA3 pathogenesis or clinical variability appears minimal based on available evidence. Further investigation employing variant validation methods and expanded cohort studies would be required to definitively assess the role of genetic modifiers in SCA3.

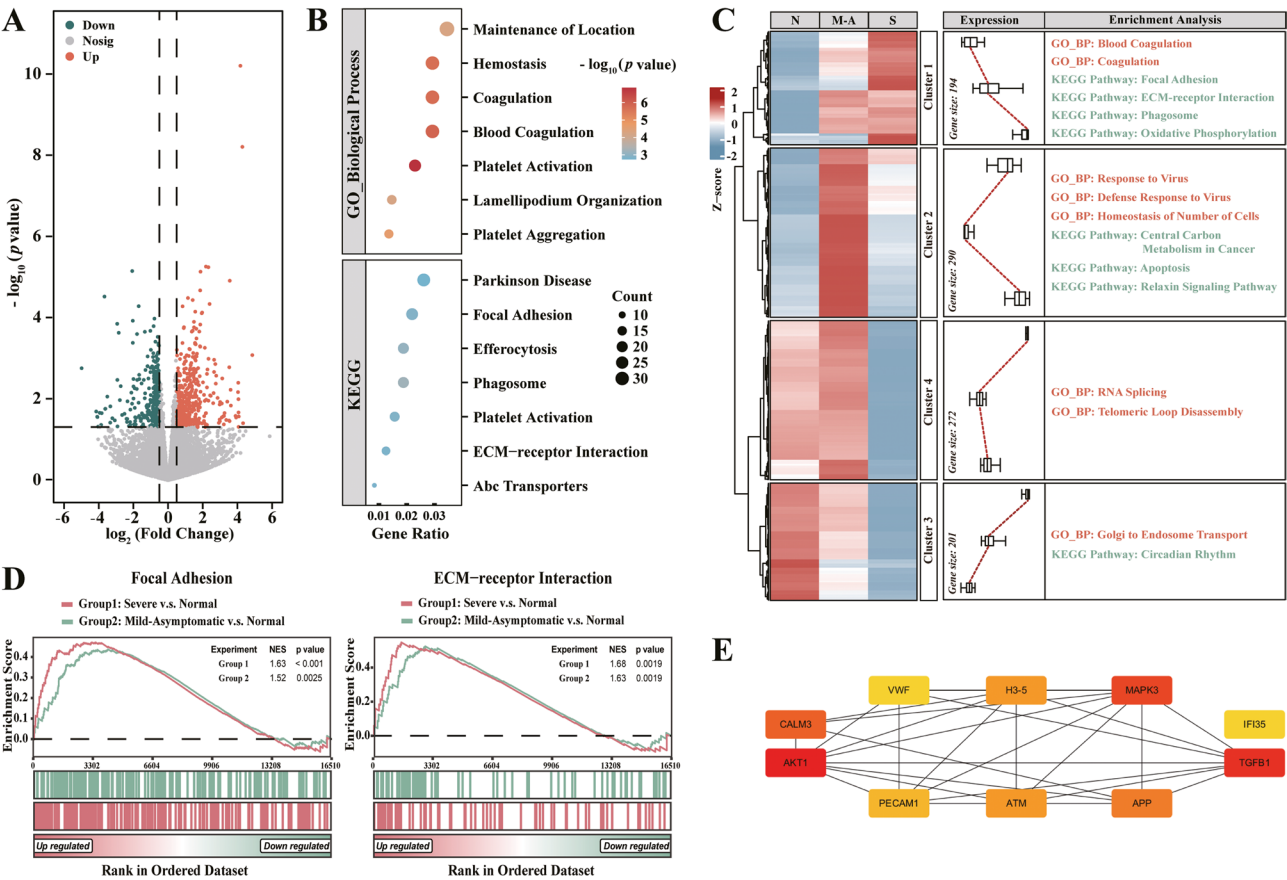
#### Human blood: genes and pathways related to SCA3 progression

Differential gene expression analysis of human peripheral blood showed a total of 795 DEGs between the severe and normal groups, comprising 444 up-regulated and 351 down-regulated genes (Fig. 3A; Table S2–Sheet2). Functional GO/KEGG enrichment analysis of these DEGs highlighted significant associations ( $p_{\text{adjust}} < 0.05$ ) with

the coagulation/anticoagulation processes, viral response pathways, immune cell migration/activation, cytoskeletal organization, and key pathways such as 'Focal Adhesion' and 'ECM-Receptor Interaction' (Fig. S2A; Table S2–Sheet3).

To identify gene transcripts correlated with SCA3 progression, we analyzed two DEG datasets: Dataset A [243 genes; intersection of significant DEGs in (severe vs normal,  $p < 0.05$ ) and (mild-asymptomatic vs normal,  $p < 0.05$ ) comparisons] and Dataset B [838 genes; significant only in (severe vs normal,  $p < 0.05$ )]. From Dataset A, the 119 DEGs showing progressive expression changes ( $|\log_2FC_{\text{severe vs normal}}| > |\log_2FC_{\text{mild-asymptomatic vs normal}}|$ ) were selected as potential biomarkers (Sub-Dataset A; Table S2–Sheet4). Combined analysis of the 957 DEGs (Sub-Dataset A and Dataset B) revealed significant enrichment ( $p_{\text{adjust}} < 0.05$ ) in terms related to the coagulation/anticoagulation processes, cytoskeletal organization, and key pathways including 'Focal Adhesion', 'ECM-Receptor Interaction', and 'Phagosome' (Fig. 3B; Table S2–Sheet5). Trend analysis clustered these DEGs into four patterns, with cluster 1, 3, and 4 showing disease-stage-dependent expression: the cluster 1 enriched for coagulation and pathways like 'Focal Adhesion', 'ECM-Receptor Interaction', 'Phagosome', and 'Oxidative Phosphorylation'; the cluster 3 for 'RNA Splicing', and cluster 4 for 'Golgi to Endosome Transport' (Fig. 3C;  $p < 0.01$ , Table S2–Sheet6). To capture genes with subtle mRNA changes and maximize biological insights, we performed GSEA analysis comparing 'severe vs normal' and 'mild-asymptomatic vs normal' groups. Pathways were prioritized based on the normalized enrichment scores (NES), with only 'Focal Adhesion' and 'ECM-Receptor Interaction' pathways meeting stringent thresholds ( $|NES_{\text{severe vs normal}}| > |NES_{\text{mild-asymptomatic vs normal}}|$ ,  $|NES| > 1$ ,  $p_{\text{adjust}} < 0.25$ ; Fig. 3D; Table S2–Sheet7).

PPI network analysis (constructed via STRING and analyzed in Cytoscape) of the 957 human blood DEGs identified ten hub molecules potentially regulating SCA3 progression, and then ranked them by interaction degree: AKT1 (103), TGFB1 (57), MAPK3 (53), CALM3 (52), APP (50), H3-5 (42), ATM (42), PECAM1 (41), IFI35 (40), and VWF (40) (Fig. 3E; Table S2–Sheet8). We subsequently validated these genes by assessing their tissue-specific expression patterns using GTEx database profiles. With the exception of H3-5, all genes exhibited high expression levels in brain tissue (Fig. S2E). Notably, *CALM3* showed significant expression in both neuronal and peripheral tissues (brain, heart, and skeletal muscle). As a key modulator of intracellular calcium homeostasis, its elevated neural expression implies potential involvement in neuronal excitability and synaptic plasticity — mechanisms that could underlie the neuronal dysfunction characteristic of SCA3 pathogenesis.



**Fig. 3** The bioinformatic analysis of peripheral blood transcriptomes from grouped family members. **(A)** Differential gene expression analysis identifies 795 DEGs associated with SCA3 occurrence. **(B–E)** Progression-related analyses include: **(B)** GO/KEGG enrichment analysis of 957 DEGs; **(C)** expression clustering and the followed functional enrichment via heatmap (N, Normal; M-A, Mild-Asymptomatic; S, Severe); **(D)** GSEA for pathway-level insights; and **(E)** PPI network construction with hub node identification, highlighting key regulatory molecules. Color scale: red (most interactors) → yellow (fewest interactors).

### Mouse blood: genes and pathways related to SCA3 progression

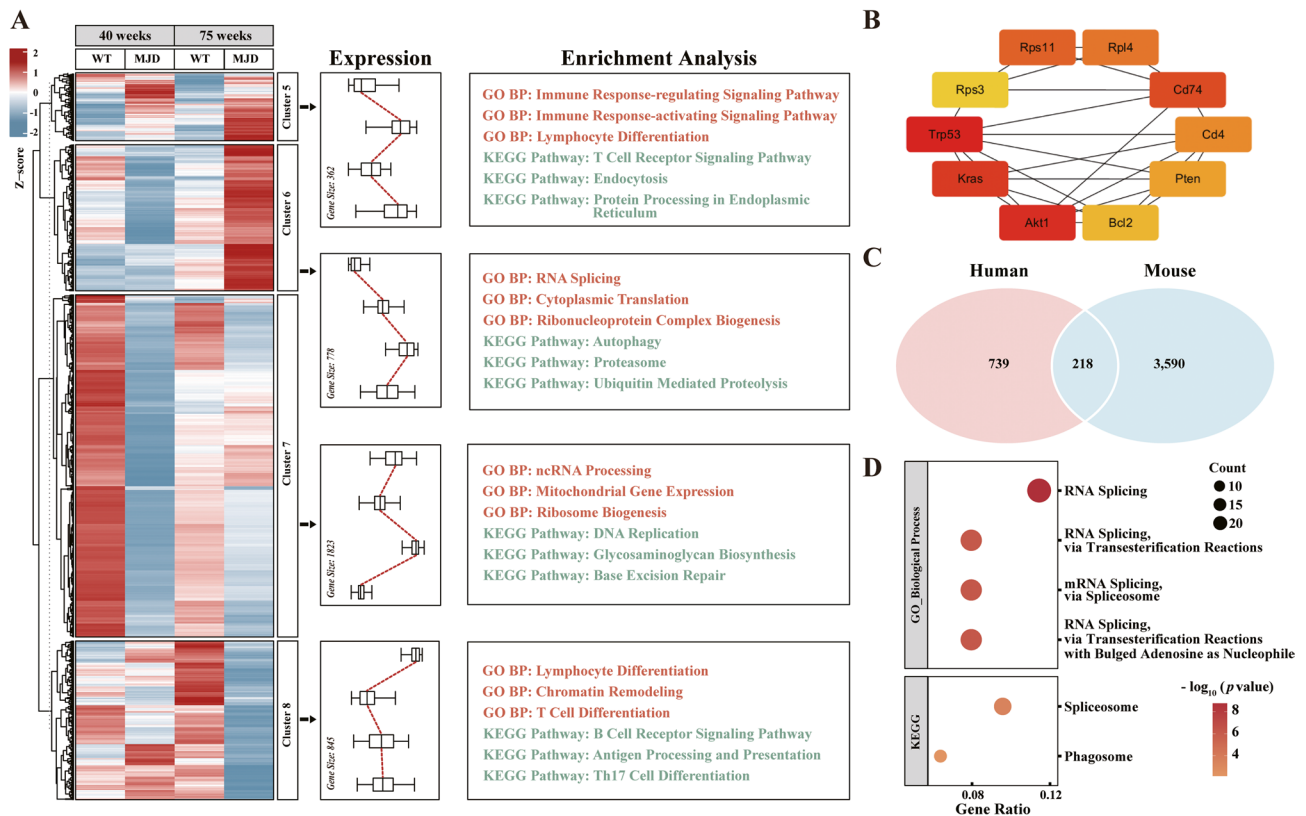
To validate human findings, we further analyzed mouse blood RNA-seq dataset (GSE108069), defining two datasets: Dataset C [109 genes; intersection of DEGs from (MJD84.2 mice, 70w-84Q vs. -WT,  $p < 0.05$ ) and (40w-84Q vs. -WT,  $p < 0.05$ )] and Dataset D [3,729 genes; significant only in (70w-84Q vs. -WT,  $p < 0.05$ )]. From Dataset C, 79 DEGs showing disease progressive changes were stored into Sub-Dataset C. Hierarchical clustering of the combined 3,808 genes (Sub-Dataset C and Dataset D; Table S2–Sheet9) revealed two major categories: clusters 5/6 and 7/8, with the clusters 6 and 8 showing age- and disease-dependent patterns in MJD84.2 mice. Functional enrichment identified the cluster 6 genes associated with RNA splicing, autophagy, and proteasome, while cluster 8 genes were linked to immune response (Fig. 4A;  $p < 0.005$ , Table S2–Sheet10). PPI network from the 3,808 mouse blood DEGs highlighted Trp53 (490 hits) and Akt1 (402 hits) as top hubs, with

AKT1 conservation between species reinforcing its potential regulatory role (Fig. 4B; Table S2–Sheet11).

Intersection analysis of human and mouse blood biomarkers identified 218 conserved DEGs as potential cross-species SCA3 progression biomarkers (Fig. 4C; Table S2–Sheet12), with functional enrichment highlighting mRNA splicing and spliceosome pathways (Fig. 4D;  $p_{\text{adjust}} < 0.05$ , Table S2–Sheet13). Further filtering for consistent expression trends [ $(\log_2FC_{\text{severe vs. normal or 70w-84Q vs. -WT}} \times \log_2FC_{\text{mild-asymptomatic vs. normal or 40w-84Q vs. -WT}} > 0)$ ,  $|\Delta FC| > 20\%$ ] yielded ten high-confidence candidates: seven positively (*CAMTA1*, *IFIT2*, *IFIT3*, *IFITM1*, *PGRMC1*, *THBS1*, *TSC22D3*) and three negatively (*ALS2*, *C3*, *SLC35A2*) associated with SCA3 severity, all of them with established neurological disease relevance (Table S2–Sheet14).

### Mouse brain: genes and pathways related to SCA3 progression

To identify SCA3 progression biomarkers in the brain, we analyzed mouse RNA-seq from cerebellum,



**Fig. 4** Cross-species blood transcriptomic analyses of SCA3 progression. **(A)** Mouse blood DEGs are clustered and functionally annotated, revealing disease-stage-associated patterns. **(B)** PPI network analysis identifies ten hub genes in mouse blood. Color scale: red (most interactors) → yellow (fewest interactors). **(C)** Intersection analysis highlights conserved DEGs between human and mouse blood. **(D)** Functional enrichment of the cross-species 218 DEGs elucidates evolutionarily conserved pathways in SCA3 pathogenesis.

brainstem, and striatum in two datasets (GSE107958 and GSE145613). GSE107958 revealed 522, 2,346, and 2,046 DEGs in the above brain areas respectively, with 94 shared DEGs linked to SCA3 occurrence (Table S2–Sheet15~16). GSE145613 yielded a total of 3,398 cerebellar DEGs associated with SCA3 progression (Table S2–Sheet17). Intersection analysis identified 41 conserved brain DEGs (Fig. S2B; Table S2–Sheet18), functionally enriched in neurodevelopmental and metabolic pathways (Fig. S2C;  $p.adjust < 0.05$ , Table S2–Sheet19). PPI network analysis of these DEGs highlighted ten hub nodes, including four with therapeutic potential: Mobp and Klk6 (multiple sclerosis targets) [46, 47], Mal (B-cell leukemia target) [48], and Gja1 (clinical trials for diabetic foot ulcers/radiation syndrome) [49, 50] (Fig. S2D; Table S2–Sheet20~21).

#### Robustness assessment of an SCA3 bioinformatic pipeline

To validate the robustness of our SCA3 bioinformatic pipeline, we sought independent neurological or movement disorder datasets (PD/HD/other SCAs) with severity-stratified RNA-seq data, but none were available. We therefore applied our analytical framework to a non-neurological condition, ulcerative colitis (UC; GSE128682),

comprising mucosal RNA-seq profiles from 16 healthy control patients, 14 remission, and 14 active-phase UC patients. Given the extensive UC-DEGs identified, we implemented more stringent statistical thresholds to enhance specificity.

Comparative transcriptomic analysis between active UC patients and healthy controls identified 666 DEGs ( $p.adjust < 0.05$ ,  $|\log_2 FC| \geq 2$ ), consisting of 456 upregulated and 210 downregulated genes (Table S2–Sheet22). Employing our established analytical framework, we subsequently stratified the data into two distinct datasets: Dataset A, representing the intersection of DEGs significant in both active versus control and remission versus control (both  $p.adjust < 0.001$ ) comparisons; and Dataset B, comprising DEGs significant in active versus control ( $p.adjust < 0.001$ ) but not in remission versus control ( $p.adjust > 0.05$ ) analyses (Table S2–Sheets22~23). This stratification yielded 2,015 DEGs in Dataset A and 2,927 DEGs in Dataset B. From Dataset A, we further refined the analysis by selecting a total of 1,144 DEGs exhibiting progressively increased expression during active disease (defined as  $|\log_2 FC_{Active UC vs. Controls}| > |\log_2 FC_{Remission UC vs. Controls}|$ ) as candidate UC biomarkers. The combined 4,071 UC progression DEGs from Sub-Dataset

A and Dataset B underwent comprehensive functional enrichment analysis, revealing significant associations with lymphocyte differentiation and immune regulation pathways (Table S2–Sheets24~25), and subsequently aligning with established mechanisms of immune pathway mediation in IBD and its subtypes [51].

PPI network analysis of these progression-associated DEGs identified ten central hub genes ranked by connectivity: TNF (565 interactions), IL6 (483), EGFR (462), IL1B (451), ACTB (447), IFNG (430), STAT3 (401), CD8A (396), PTPRC (358), and STAT1 (313) (Table S2–Sheet26). Prior studies have demonstrated elevated expression of both TNF and IL6 in active UC, with IL6 serving as a validated marker of disease activity [52]. Furthermore, P2RY13 activation of the IL6/STAT3 pathway contributes to intestinal barrier dysfunction and UC disease progression, corroborating the pathological relevance of our identified hub genes and signaling pathways [53]. The consistency between our identified UC progression-associated DEGs/central hub genes and previously validated UC biomarkers confirms the reliability and robustness of our SCA3 analytical pipeline.

## Discussion

This study genetically confirmed a suspected hereditary cerebellar ataxia as SCA3 in a Chinese family through comprehensive genetic analysis. Blood transcriptomic profiling identified disease-associated DEGs involved in coagulation, immune regulation, and cytoskeletal dynamics, particularly highlighting dysregulation in FA and ECM-Receptor Interaction pathways with AKT1, TGFB1, MAPK3, CALM3, APP as central molecular hubs. Cross-species validation reinforced these findings, revealing conserved biomarkers (*C3/ALS2/SLC35A2*↓ and *THBS1/CAMTA1*↑) and demonstrating significant disturbances in immune and RNA splicing function. Brain-specific mRNA profiles in SCA3 mice uncovered critical neurodevelopmental and metabolic alterations, pinpointing promising therapeutic targets (Mobp, Mal, Gja1, Klk6). This work provides mechanistic insights into SCA3 pathogenesis while offering valuable biomarkers for disease monitoring and novel therapeutic avenues.

### SCA3 phenotypic diversity and pathology

In this study, the family members with severe SCA3 (III-2, III-6, III-14) exhibited a high frequency of extrapyramidal signs, consistent with a previous report [54]. These manifestations reflect nigrostriatal dopaminergic pathway dysfunction, as evidenced by the reduced 99mTc-TRODAT-1 binding on brain single photon emission computed tomography (SPECT) [55]. Cranial MRI findings revealed variable atrophy in the corpus callosum, cerebellar vermis, peduncles, and suggested concurrent myelinopathy and CSVD, in this pedigree. While

widespread WM damage and atrophy begin pre-ataxia [56, 57], with most prominent changes occurring in cerebellar vermis and peduncles [56–59], corpus callosum atrophy remains exceptionally rare in SCA3 patients [60], having been reported only in an Indonesian family [61] despite being common in *TCF4*-related intellectual disability [62–64]. Human SCA3 is associated with cerebellar demyelination, but the extent of cerebral cortical pathology remains uncertain [65, 66]. In contrast, various SCA3 mouse and cell models demonstrate consistent widespread demyelination and oligodendrocyte impairment [36, 65–68]. Blood-brain barrier impairment in SCA3, attributed to tight junction protein dysregulation [69], appears distinct from CSVD, which lacks clear documentation in SCA3. Additionally, vestibular eye movement abnormalities (e.g., central nystagmus, saccadic pursuit, and square-wave jerks) indicate multisystem neurological involvement. Neuropsychiatric symptoms — including cognitive decline, anxiety, and depression — represent prevalent features of SCA3 [70] and show progressive worsening with disease advancement [71, 72]. The underlying pathophysiology may involve widespread neuronal damage [73], cerebellar dopaminergic/metabolite dysfunction [73, 74], and cerebellar/brainstem atrophy [73, 75]. However, our pedigree exhibited no clear correlation between ataxia severity and neuropsychiatric/neuroimaging findings. The concurrent presentation of atypical features (corpus callosum atrophy, CSVD) with neuropsychiatric symptoms may reflect population-specific genetic influences, environmental factors, or study limitations inherent to small cohort sizes.

### Genetic anticipation in SCA3 age trends

In STR-associated diseases, genetic anticipation commonly occurs, resulting in a decline in the age of onset over generations. In this SCA3 pedigree, the first and second generations exhibited relatively later onset, while the third generation typically presented symptoms at 40–45 years, and the fourth generation showed symptomatic individuals as early as 30 years, consistent with genetic anticipation. This trend is particularly evident in the right branch of this pedigree (II-6: 63 years, III-14: 44 years, IV-10: 30 years), correlating with progressive CAG expansion in *ATXN3* (69 to 70 to 71 repeats). In the other branch, anticipation is less clear due to incomplete age-of-onset data; for instance, I-1 (55? years) and II-2 (40? years) relied on family recall, which may lack precision. SCA3 onset is also influenced by genetic and epigenetic factors, complicating the pattern. Notably, phenotypic heterogeneity exists even among male patients of similar ages — severe in III-6 (28/74 repeats) and III-14 (23/70 repeats) but mild in III-16 (23/68 repeats) — suggesting modifiers beyond CAG repeat length.

Genetic and epigenetic modifiers of SCA3 are diverse. Although *ATXN3* CAG repeat length remains the primary determinant of clinical decline [76], other factors like CAG tract lengths in *ATXN2* [77] and other polyQ-related genes [78–80], as well as common variants in *ATXN3* [81, 82] and *ATXN2* [83]. SCA3 onset and phenotype are also linked to genetic variants that affecting lipid metabolism [84], mitochondrial function [85], autophagy [86], DNA methylation [87], and neurotransmitters [88]. Recent genome-wide association [89] and WES [90] studies have identified dozens of modifiers. Rare exonic variants, such as *PRKN*-V380L (rs1801582, impairing mitophagy) [91], *PIAS1*-S510G (rs755539001, delaying onset via SUMOylation regulation) [92, 93], and a 9-bp duplication in *ATXN2* (exacerbating phenotype) [94], further modulate SCA3. Mathematical algorithms integrating CAG length and polyQ-related gene data now aid in predicting onset age [78–80, 95, 96].

#### Precision medicine for SCA genetic diagnosis

In this study, HTS analysis was performed on proband II-6 or nine other members, including short-read TS, WES with genome linkage analysis, and LR-WGS, to determine causal variants in the SCA pedigree. Short-read TS and WES failed to reliably detect pathological variants, likely due to the high GC content in STR expansion regions of *ATXN3*, which complicates genetic variant capture, mapping, and calling. In contrast, LR-WGS provided a comprehensive genomic review, encompassing coding and non-coding regions, including STRs and telomeres, while improving sequence capture, mapping, and phasing. LR-WGS is particularly valuable for understanding the genetic architecture and consequences of SVs and STRs [97–100], and also excels in identifying insertions, transposable elements, complex deletions, and disease-related STRs, overcoming limitations of short-read WGS [101]. With optimized bioinformatic processing pipelines [98, 102], LR-WGS outperforms short-read methods in detecting pathogenic SVs and STRs, especially in REDs [97, 99, 103]. Although it has improved diagnostic yield [104, 105] and accuracy [106, 107] in known STR-associated diseases, these preliminary findings require validation in larger cohorts to confirm their clinical significance. Additionally, LR-WGS provides critical insights into STR configuration/haplotype phasing, intergenerational variation, phenotypic heterogeneity, and co-occurrence in SCA individuals [108–110].

The diagnosis of SCA3, caused by CAG repeat expansions in the *ATXN3*, requires a strategic approach balancing clinical assessment with appropriate molecular genetic testing. Initial evaluation should focus on characteristic symptoms like progressive ataxia, ophthalmoplegia, pyramidal impairment, dysarthria, and distinctive features like dystonia or bulging eyes, along

with a detailed family history [111]. For the individuals with classic SCA3 presentation, targeted *ATXN3* CAG repeat analysis using triplet repeat-primed PCR followed by capillary electrophoresis remains the gold-standard for its high accuracy, rapid turnaround, and cost-effectiveness [112–116]. Triplet repeat-primed PCR usually serves as an essential supplementary validation tool for pathogenic repeat expansion detection in novel STR disorders [117–120]. However, when clinical features overlap with atypical manifestations, other HAs or SCAs, and movement disorders, the combination of HA repeat expansion panel or triplet repeat-primed PCR and WES provides a practical intermediate method [121–124]. An alternative approach involves integrating WES raw data with specialized bioinformatic tools (e.g., ExpansionHunter, exSTRa, and STRetch) for repeat expansion analysis. While these methods enable rapid detection of small repeat expansions with high sensitivity and specificity, they remain limited in identifying large or novel repeat expansions and high-GC repeats [115, 125]. In diagnostically challenging individuals where initial molecular testing is negative, WGS offers comprehensive detection of repeat expansions and simultaneously evaluating for the other types of genetic etiologies. Thus, WGS has been endorsed by multiple authoritative institutions as the first-line approach for clinical genetic testing, including rare disease and RED analysis [116, 126–128]. For REDs specifically, WGS demonstrates near-perfect sensitivity and specificity (approaching 100%) relative to conventional triplet repeat-primed PCR assays [116]. Furthermore, WGS surpasses WES by providing comprehensive genome-wide coverage and enabling detection of all variant types across coding and non-coding regions [126]. A tiered diagnostic strategy — beginning with targeted triplet repeat-primed PCR for high-probability cases, progressing to targeted repeat expansion panels and WES for broader evaluation, and reserving WGS for complex cases — optimizes both diagnostic yield and resource utilization (Table S1–Sheet6). This strategy recognizes WGS's growing importance in broad analysis while preserving PCR's proven utility for routine SCA3 diagnostics.

In this study, the proband II-6 underwent comprehensive clinical, laboratory, and imaging evaluations, revealing presentations atypical for SCA, including corpus callosum atrophy, CSVD, and incomplete genetic anticipation. Since conventional SCA diagnostic frameworks rely on prior clinical confirmation — challenging in our pedigree — a multi-omics genomic analysis (TS/WES/LR-WGS) was performed. Although more time-consuming and costly than triplet repeat-primed PCR assays (Table S1–Sheet6), such an approach is justified in complex cases, as robust cost-effectiveness analysis often derive from large-scale cohort studies rather than

single-family analyses. These findings underscore the need for diagnostic algorithms that integrate the specificity and sensitivity of genetic testing, cost-effectiveness, and re-evaluation of prior clinical and molecular data, advancing the paradigm for HAs/SCAs/REDs diagnosis.

#### **ECM-receptor interaction and focal adhesion pathways in SCA3 pathogenesis**

Transcriptomic profiling of SCA3 patient blood samples revealed conserved dysregulation in ‘ECM-Receptor Interaction’ and ‘Focal Adhesion’ pathways, mirroring findings in other neurodegenerative and neuropsychiatric disorders, including Parkinson’s disease (PD) [129, 130], Alzheimer’s disease (AD) [131, 132], and schizophrenia [133]. The extracellular matrix (ECM), a network of structural proteins (e.g., elastin, laminins, collagens) and glycans, interacts with cell-surface receptors (integrins, CD44, TLR2/4) to regulate neurodevelopmental processes such as neuronal migration, axonal guidance, and synaptic plasticity [134–136]. Focal adhesion (FA) facilitates ECM-cytoskeletal interactions, cellular experiments reveal that cytoskeletal disruption and integrin signaling defects arise from either ataxin-3 deficiency or its pathogenic expansion [137, 138]. Pre-clinical SCA3 models further demonstrate early cytoskeletal disruption, impaired axonal transport, and metabolic dysfunction [139, 140], indicating that ECM-FA-cytoskeleton disruption as a key driver of SCA3 pathology. The presymptomatic appearance of these defects positions ECM/FA pathways as attractive targets for early therapeutic intervention.

PolyQ-ataxin-3 disrupts cytoskeletal integrity, impairing axonal transport, neuronal morphology, and metabolic function, ultimately leading to neuronal cell death [137–140]. Defective integrin signaling via FA components destabilizes synaptic connections [137, 141], while ECM degradation and its aberrant receptor signaling disrupt neurodevelopmental processes and synaptic plasticity [134, 135, 141, 142]. These results identify ECM-FA-cytoskeleton dysregulation as a core pathogenic mechanism in SCA3, with peripheral tissue detection enabling potential early biomarker development.

#### **Immune dysregulation and RNA splicing defects in SCA3**

Immune dysregulation plays a significant role in SCA3 pathogenesis, as evidenced by altered immune cell migration/activation pathways in transcriptomic analyses. *Atn3*-knockout mice exhibit lymphocyte depletion and impaired immune defense [143], while ataxin-3 deficiency suppresses innate immune responses via NOD2/TLR2 pathways [144]. PolyQ-ataxin-3 further disrupts myeloid cell function in patients and knock-in mice [145]. Neuroinflammation is supported by microglial/astrocyte activation and inflammatory gene upregulation

in human SCA3 pons [146, 147] and mouse models [143, 148]. RNA splicing defects also contribute to SCA3 progression. PolyQ-ataxin-3 disrupts Cajal body function in snRNP biogenesis [149, 150], forms RNA foci that sequester MBNL1 splicing factors [151], and causes early spliceosome protein dysregulation [139]. These findings underscore immune dysfunction and RNA processing defects as key mechanisms in SCA3, offering additional therapeutic targets.

#### **Conserved biomarkers and novel pathways in SCA3 progression**

Ten conserved blood-derived DEGs were identified across human and mouse SCA3 models, with five key genes — *C3*↓, *ALS2*↓, *SLC35A2*↓, *THBS1*↑, *CAMTA1*↑ — showing strong neurological associations. *C3* and *THBS1*, linked to ‘ECM-Receptor Interaction’, ‘Focal Adhesion’, and ‘Phagosome’ pathways, exhibit context-dependent effects: *C3* reduction improves synaptic integrity in multiple sclerosis [152] and attenuates astrogliosis in epilepsy [153] but impairs neurodevelopment [154]; *THBS1* elevation exacerbates epileptogenesis [155] yet protects against Aβ toxicity in AD [156]. Clinical findings show inverse correlations between *C3* levels and neurological outcomes [157, 158], while *THBS1* elevation correlates positively [159, 160], suggesting their biomarker potential.

*ALS2* and *CAMTA1* genes further demonstrate critical neurological roles. *ALS2* regulates endosomal trafficking, mitochondrial localization, and axonal growth [161–163], with its deficiency causing motor deficits in mice [164] and humans [165]. *CAMTA1* variants can lead to human ataxia [166, 167], and its knockout mice exhibit ataxia and Purkinje cell degeneration [168]. *SLC35A2*, a glycosylation regulator, is implicated in SCA3 due to its association with cerebellar atrophy and WM defects [169, 170]. These findings highlight novel biomarkers and therapeutic targets, bridging hematological transcriptomics with central nervous system (CNS) pathology.

Our findings align with and extend previous research on transcriptional dysregulation in SCA3/MJD, particularly regarding blood-based biomarkers identified across multiple cohort studies. Building on the work of Mafalda Raposo, who identified *FCGR3B*, *P2RY13*, and *SELPLG* as consistently upregulated genes from patient blood samples [23], our findings further support their relevance in the immune cell migration/activation pathway. We notice that *P2RY13*’s significance as a disease-onset transcriptional biomarker has been independently validated in blood and brain samples [25], enhancing the emerging connection between purinergic signaling (mediated by *P2RY13* receptor) [171] and SCA3 pathogenesis. Further support comes from studies highlighting *DDIT4* and *TRIM13* as high-precision discriminators of

CAG-dependent expression patterns [25], aligning with our observed downregulation of *C3*, *ALS2*, and *SLC35A2* and reinforcing AKT1's central role in SCA3 progression-related PPI network. These findings highlight convergent dysregulation in stress response and ubiquitin-proteasome pathways. Another research by Mafalda Raposo [24] has further elucidated relevant pathways interacting with FA and immune responses. This proposes *SAFB2*, *SFSWAP*, and *LTBP4* as potential pre-ataxic biomarkers — findings that demonstrate strong conceptual alignment with our results showing ECM receptor/FA-mediated ECM remodeling processes. These studies validate blood-based transcriptional profiling as a meaningful proxy for CNS pathology in SCA3. Our current data integrates these prior observations by identifying potential biomarkers (*THBS1*, *CAMTA1*) and drug targets (*AKT1*, *KLK6*), while providing pathway-level insights into SCA3 progression mechanisms. This growing body of evidence solidifies blood transcriptomics as a robust platform for both patient stratification and therapeutic development in SCA3 research.

In summary, this study systematically investigated SCA3 DEGs, progression-related biomarkers/hub nodes, and therapeutic targets using a genetically homogeneous pedigree, minimizing variation through analysis of direct blood relatives while controlling for age-related confounders [172]. Our multi-species transcriptomic approach revealed stage-specific molecular changes, though several limitations warrant consideration: single-family analysis may not capture population heterogeneity; imperfect age/gender matching in patients and controls; small sample sizes potentially limiting generalizability; and absence of independent clinical and experimental validation. Future directions should include expanded cohorts, diagnostic efficacy studies, and validation in diverse populations/biological models to translate these findings into clinically applicable biomarkers. These findings nevertheless establish a robust foundation for SCA3 stratification and therapeutic development, with particular value for understanding disease progression in genetically defined subgroups.

#### Abbreviations

|         |                                     |
|---------|-------------------------------------|
| CNS     | Central Nervous System              |
| CNV     | Copy Number Variant                 |
| CSVD    | Cerebral Small Vessel Disease       |
| DEG     | Differentially Expressed Gene       |
| DWI     | Diffusion-weighted Imaging          |
| ENG     | Electronystagmography               |
| FLAIR   | Fluid-attenuated Inversion Recovery |
| HA      | Hereditary Ataxia                   |
| LR-WGS  | Long-read Whole-genome Sequencing   |
| MJD     | Machado-Joseph Disease              |
| MRI     | Magnetic Resonance Imaging          |
| NGS     | Next-generation Sequencing          |
| polyQ   | Polyglutamine                       |
| PPI     | Protein-protein Interaction         |
| RNA-seq | RNA Sequencing                      |

|          |                                 |
|----------|---------------------------------|
| SAS      | Self-rating Anxiety Scale       |
| SCA      | Spinocerebellar Ataxia          |
| SCL-90-R | Symptom Checklist-90-Revised    |
| SDS      | Self-rating Depression Scale    |
| SV       | Structural Variant              |
| SWI      | Susceptibility-weighted Imaging |
| TS       | Targeted Sequencing             |
| WES      | Whole-exome Sequencing          |

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12883-025-04378-z>.

Supplementary Material 1

#### Acknowledgements

We extend our sincere gratitude to our mentor, Professor Xue Zhang of Harbin Medical University and the Chinese Academy of Medical Sciences & Peking Union Medical College, for his invaluable guidance, unwavering support, and dedicated mentorship throughout this study.

#### Author contributions

Conceptualization, LS, YQ-R, and LC; Methodology, LS, YQ-R, CL, XW, and CX; Software, XW, CX, LC, and YQ-R; Validation, CX, YQ-R, XW, and CL; Formal Analysis, XW, CX, and YQ-R; Investigation, XW, LY-K, YL, HZ, CL, XX-L, and JQ-T; Resources, CL, XW, XX-L, SW-T, and ZT-Z; Data Curation, XW, LY-K, YL, HZ, CL, and XX-L; Writing — original draft preparation, LS, CL, CX, XW, and YQ-R; Writing — review and editing, LS, XW, LC, CL, and YQ-R; Visualization, XW, YQ-R, and YL; Supervision, LS, LC, and CL; Project Administration, XW, CL, YQ-R, and CX; Funding Acquisition, LS and LC. All authors have read and agreed to the published version of the manuscript.

#### Funding

This research was funded by the National Key R&D Program of China (2020YFA0804000) and the National Natural Science Foundation of China (62222104, 62172130).

#### Data availability

The mouse tissue RNA-seq data in this study are publicly available in the NCBI GEO database (<https://www.ncbi.nlm.nih.gov/geo/>) under accession numbers GSE145613, GSE107958, and GSE108069. Human WES, LR-WGS, and RNA-seq data have been deposited in the Genome Sequence Archive for Human (GSA-Human, <https://ngdc.cncb.ac.cn/gsa-human/>) with accession numbers HRA010775 (<https://ngdc.cncb.ac.cn/search/specific?db=hra&q=HRA010775>), HRA010777 (<https://ngdc.cncb.ac.cn/search/specific?db=hra&q=HRA010777>), and HRA010785 (<https://ngdc.cncb.ac.cn/search/specific?db=hra&q=HRA010785>), respectively. Due to Chinese data protection regulations, these datasets are also available upon request from the corresponding author.

#### Declarations

##### Ethics approval

The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of the First Traditional Chinese Medicine Hospital in Changde City (Approval Number: 2020–020528) and Harbin Medical University (Approval Number: HMUIRB2022006).

##### Informed consent

Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the patient(s) to publish this paper.

##### Competing interests

The authors declare no competing interests.

##### Author details

<sup>1</sup>Department of Neurology, The First Traditional Chinese Medicine Hospital in Changde City, Changde 415000, China

<sup>2</sup>National Health Commission Key Laboratory of Molecular Probes and Targeted Diagnosis and Therapy, Harbin Medical University, Harbin 150081, China

<sup>3</sup>College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, China

<sup>4</sup>Department of Pediatrics, The Second Hospital of Harbin Medical University, Harbin 150001, China

<sup>5</sup>Department of Radiology, The First People's Hospital of Changde City, Changde 415000, China

<sup>6</sup>Department of Medical Genetics, College of Basic Medical Sciences, Harbin Medical University, Harbin 150081, China

<sup>7</sup>Department of Child and Adolescent Health, School of Public Health, Harbin Medical University, Harbin 150081, China

<sup>8</sup>Department of Genetics and Endocrinology, Liuzhou Maternity and Child Health Care Hospital, Liuzhou 545001, China

<sup>9</sup>McKusick-Zhang Center for Genetic Medicine, State Key Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Science and Peking Union Medical College, Beijing 100005, China

<sup>10</sup>Laboratory of Pediatric Research, Children's Hospital of Wujiang District, Suzhou 215234, China

Received: 26 February 2025 / Accepted: 30 July 2025

Published online: 01 September 2025

## References

- Jayadev S, Bird TD. Hereditary ataxias: overview. *Genet Med*. 2013;15(9):673–83.
- Hersheson J, Haworth A, Houlden H. The inherited ataxias: genetic heterogeneity, mutation databases, and future directions in research and clinical diagnostics. *Hum Mutat*. 2012;33(9):1324–32.
- Ruano L, Melo C, Silva MC, Coutinho P. The global epidemiology of hereditary ataxia and spastic paraplegia: a systematic review of prevalence studies. *Neuroepidemiology*. 2014;42(3):174–83.
- Coutinho P, Ruano L, Loureiro JL, Cruz VT, Barros J, Tuna A, Barbot C, Guimarães J, Alonso I, Silveira I, et al. Hereditary ataxia and spastic paraplegia in Portugal. *JAMA Neurol*. 2013;70(6):746–55.
- Klockgether T, Mariotti C, Paulson HL. Spinocerebellar ataxia. *Nat Rev Dis Primers*. 2019;5(1):24.
- Scott SSO, Pedrosa JL, Barsottini OGP, França-Junior MC, Braga-Neto P. Natural history and epidemiology of the spinocerebellar ataxias: insights from the first description to nowadays. *J Neurol Sci*. 2020;417: 117082.
- Teive HAG, Meira AT, Camargo CHF, Munhoz RP. The geographic diversity of spinocerebellar ataxias (SCAs) in the Americas: a systematic review. *Mov Disord Clin Pract*. 2019;6(7):531–40.
- Rodríguez-Labrada R, Martins AC, Magaña JJ, Vazquez-Mojena Y, Medrano-Montero J, Fernandez-Ruiz J, Cisneros B, Teive H, McFarland KN, Saraiva-Pereira ML, et al. Founder effects of spinocerebellar ataxias in the American continents and the Caribbean. *Cerebellum*. 2020;19(3):446–58.
- De Mattei F, Ferrandes F, Gallone S, Canosa A, Calvo A, Chiò A, Vasta R. Epidemiology of spinocerebellar ataxias in Europe. *Cerebellum*. 2023;23(3):1176–83.
- Alvarenga MP, Siciliani LC, Carvalho RS, Ganimi MC, Penna PS. Spinocerebellar ataxia in a cohort of patients from Rio de Janeiro. *Neurol Sci*. 2022;43(8):4997–5005.
- Galecio-Castillo M, Gutierrez-Arratia J, Abad-Murillo A, Sarapura-Castro E, Araujo-Allaga I, Saldarriaga-Mayo A, Illanes-Manrique M, Cornejo-Olivas M. Epidemiology of autosomal dominant spinocerebellar ataxias in Latin America: a systematic review and meta-analysis. *Cerebellum*. 2025;24(3):75.
- Chen Z, Wang P, Wang C, Peng Y, Hou X, Zhou X, Li T, Peng H, Qiu R, Xia K, et al. Updated frequency analysis of spinocerebellar ataxia in China. *Brain*. 2018;141(4):e22–22.
- van Prooije T, Ibrahim NM, Azmin S, van de Warrenburg B. Spinocerebellar ataxias in Asia: prevalence, phenotypes and management. *Parkinsonism Relat Disord*. 2021;92:112–8.
- Witek N, Hawkins J, Hall D. Genetic ataxias: update on classification and diagnostic approaches. *Curr Neurol Neurosci Rep*. 2021;21(3):13.
- Coarelli G, Wirth T, Tranchant C, Koenig M, Durr A, Anheim M. The inherited cerebellar ataxias: an update. *J Neurol*. 2022;270(1):208–22.
- Tenorio RB, Camargo CHF, Donis KC, Almeida CCB, Teive HAG. Diagnostic yield of NGS tests for hereditary ataxia: a systematic review. *Cerebellum*. 2023;23(4):1552–65.
- Pellerin D, Iruzubieta P, Xu IRL, Danzi MC, Cortese A, Synofzik M, Houlden H, Zuchner S, Brais B. Recent advances in the genetics of ataxias: an update on novel autosomal dominant repeat expansions. *Curr Neurol Neurosci Rep*. 2025;25(1):16.
- Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. *Nat Rev Genet*. 2020;21(10):597–614.
- Porubsky D, Eichler EE. A 25-year odyssey of genomic technology advances and structural variant discovery. *Cell*. 2024;187(5):1024–37.
- Paulson HL, Shakkottai VG, Clark HB, Orr HT. Polyglutamine spinocerebellar ataxias — from genes to potential treatments. *Nat Rev Neurosci*. 2017;18(10):613–26.
- Zhou Z-D, Jankovic J, Ashizawa T, Tan E-K. Neurodegenerative diseases associated with non-coding CGG tandem repeat expansions. *Nat Reviews Neurol*. 2022;18(3):145–57.
- Ashizawa T, Öz G, Paulson HL. Spinocerebellar ataxias: prospects and challenges for therapy development. *Nat Rev Neurol*. 2018;14(10):590–605.
- Raposo M, Bettencourt C, Maciel P, Gao F, Ramos A, Kazachkova N, Vasconcelos J, Kay T, Rodrigues AJ, Bettencourt B, et al. Novel candidate blood-based transcriptional biomarkers of machado-joseph disease. *Mov Disord*. 2015;30(7):968–75.
- Raposo M, Hübener-Schmid J, Ferreira AF, Vieira Melo AR, Vasconcelos J, Pires P, Kay T, Garcia-Moreno H, Giunti P, Santana MM, et al. Blood transcriptome sequencing identifies biomarkers able to track disease stages in spinocerebellar ataxia type 3. *Brain*. 2023;146(10):4132–43.
- Ferreira AF, Raposo M, Shaw ED, Liu L, Vasconcelos J, Kay T, Bettencourt C, Saraiva-Pereira ML, Jardim LB et al. Costa MdC : Blood DDT4 and TRIM13 Transcript Levels Mark the Early Stages of Machado–Joseph Disease. *Annals of Neurology* 2025, 98(1):107–119.
- Zung WWK. A rating instrument for anxiety disorders. *Psychosomatics*. 1971;12(6):371–9.
- Zung WWK. A self-rating depression scale. *Arch Gen Psychiatry*. 1965;12(1):63–70.
- Derogatis LR. Symptom Checklist-90-R: administration, scoring & procedure manual for the revised version of the SCL-90. MN: National Computer Systems, Minneapolis; 1994.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009;25(14):1754–60.
- Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P, Sambamba: fast processing of NGS alignment formats. *Bioinformatics*. 2015;31(12):2032–4.
- Li H, Birol I. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34(18):3094–100.
- Li H, Alkan C. New strategies to improve minimap2 alignment accuracy. *Bioinformatics*. 2021;37(23):4572–4.
- Liu Q, Zhang P, Wang D, Gu W, Wang K. Interrogating the unsequenceable genomic trinucleotide repeat disorders by long-read sequencing. *Genome Med*. 2017;9(1):65.
- Liu Q, Tong Y, Wang K. Genome-wide detection of short tandem repeat expansions by long-read sequencing. *BMC Bioinformatics*. 2020;21(S21):542.
- Toonen LJA, Overzier M, Evers MM, Leon LG, van der Zeeuw SAJ, Mei H, Kielbasa SM, Goeman JJ, Hettne KM, Magnusson OT, et al. Transcriptional profiling and biomarker identification reveal tissue specific effects of expanded ataxin-3 in a spinocerebellar ataxia type 3 mouse model. *Mol Neurodegener*. 2018;13(1):31.
- Haas E, Incebacak RD, Hentrich T, Huridou C, Schmidt T, Casadei N, Maringer Y, Bahl C, Zimmermann F, Mills JD, et al. A novel SCA3 knock-in mouse model mimics the human SCA3 disease phenotype including neuropathological, behavioral, and transcriptional abnormalities especially in oligodendrocytes. *Mol Neurobiol*. 2021;59(1):495–522.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
- Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, et al. ClusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation (Camb)*. 2021;2(3): 100141.
- Carlson M. org.Hs.eg.db: genome wide annotation for human. In.; 2019.
- Carlson M. org.Mm.eg.db: genome wide annotation for mouse. In.; 2019.
- Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the r/bioconductor package biomart. *Nat Protoc*. 2009;4(8):1184–91.

42. Martins S, Calafell F, Gaspar C, Wong VCN, Silveira I, Nicholson GA, Brunt ER, Tranebjaerg L, Stevanin G, Hsieh M, et al. Asian origin for the worldwide-spread mutational event in Machado-Joseph disease. *Arch Neurol*. 2007;64(10):1502–8.
43. Martins S, Soong B-W, Wong VCN, Giunti P, Stevanin G, Ranum LPW, Sasaki H, Riess O, Tsuji S, Coutinho P, et al. Mutational origin of Machado-Joseph disease in the Australian aboriginal communities of Groote Eylandt and Yirrkala. *Arch Neurol*. 2012;69(6):746–51.
44. Ogun SA, Martins S, Adebayo PB, Dawodu CO, Sequeiros J, Finkel MF. Machado-Joseph disease in a Nigerian family: mutational origin and review of the literature. *Eur J Hum Genet*. 2014;23(2):271–3.
45. Tezenas du Montcel S, Durr A, Rakowicz M, Nanetti L, Charles P, Sulek A, Mariotti C, Rola R, Schols L, Bauer P, et al. Prediction of the age at onset in spinocerebellar ataxia type 1, 2, 3 and 6. *J Med Genet*. 2014;51(7):479–86.
46. Kaye JF, Kerlero de Rosbo N, Mendel I, Flechter S, Hoffman M, Yust I, Ben-Nun A. The central nervous system-specific Myelin oligodendrocytic basic protein (MOBP) is encephalitogenic and a potential target antigen in multiple sclerosis (MS). *J Neuroimmunol*. 2000;102(2):189–98.
47. Liang G, Chen X, Aldous S, Pu S-F, Mehdi S, Powers E, Giovanni A, Kongsamut S, Xia T, Zhang Y, et al. Virtual screening and X-ray crystallography for human Kallikrein 6 inhibitors with an Amidinothiophene P1 group. *ACS Med Chem Lett*. 2012;3(2):159–64.
48. Antosz H, Sajewicz J, Marzec-Kotarska B, Dmoszyńska A, Baszak J, Jargiello-Baszak M. Aberrant TIRAP and MyD88 expression in B-cell chronic lymphocytic leukemia. *Blood Cells Mol Dis*. 2013;51(1):48–55.
49. Marsan E, Velmeshev D, Ramsey A, Patel RK, Zhang J, Koontz M, Andrews MG, de Mayo M, Mora C, Blumenfeld J, et al. Astroglial toxicity promotes synaptic degeneration in the thalamocortical circuit in frontotemporal dementia with GRN mutations. *J Clin Invest*. 2023. <https://doi.org/10.1172/JCI164919>.
50. Takase EO, Yamasaki R, Nagata S, Watanabe M, Masaki K, Yamaguchi H, Kira J-i, Takeuchi H, Isobe N. Astroglial connexin 43 is a novel therapeutic target for chronic multiple sclerosis model. *Sci Rep*. 2024;14(1):10877.
51. de Souza HSP, Fiocchi C. Immunopathogenesis of IBD: current state of the art. *Nat Rev Gastroenterol Hepatol*. 2015;11(1):13–27.
52. Nakase H, Sato N, Mizuno N, Ikawa Y. The influence of cytokines on the complex pathology of ulcerative colitis. *Autoimmun Rev*. 2022;21(3): 103017.
53. Wu X, Wei S, Chen M, Li J, Wei Y, Zhang J, Dong W. P2RY13 exacerbates intestinal inflammation by damaging the intestinal mucosal barrier via activating IL-6/STAT3 pathway. *Int J Biol Sci*. 2022;18(13):5056–69.
54. Tezenas du Montcel S, Petit E, Olubajo T, Faber J, Lallemand-Dudek P, Bushara K, Perlman S, Subramony SH, Morgan D, Jackman B, et al. Baseline clinical and blood biomarkers in patients with preataxic and Early-Stage disease spinocerebellar ataxia 1 and 3. *Neurology*. 2023;100(17):e1836–48.
55. Yen TC, Lu CS, Tzen KY, Wey SP, Chou YH, Weng YH, Kao PF, Ting G. Decreased dopamine transporter binding in Machado-Joseph disease. *J Nucl Med*. 2000;41(6):994–8.
56. Rezende TJR, de Paiva JLR, Martinez ARM, Lopes-Cendes I, Pedrosa JL, Barsotti OGP, Cendes F, França MC. Structural signature of SCA3: from presymptomatic to late disease stages. *Ann Neurol*. 2018;84(3):401–8.
57. Qiu H, Wu C, Liang J, Hu M, Chen Y, Huang Z, Yang Z, Zhao J, Chu J. Structural alterations of spinocerebellar ataxias type 3: from pre-symptomatic to symptomatic stage. *Eur Radiol*. 2022;33(4):2881–94.
58. Piccinin CC, Rezende TJR, de Paiva JLR, Moysés PC, Martinez ARM, Cendes F, França MC. A 5-year longitudinal clinical and magnetic resonance imaging study in spinocerebellar ataxia type 3. *Mov Disord*. 2020;35(9):1679–84.
59. Liu H, Lin J, Shang H. Voxel-based meta-analysis of gray matter and white matter changes in patients with spinocerebellar ataxia type 3. *Front Neurol*. 2023;14:1197822.
60. de Oliveira MS, D'Abreu A, França MC Jr, Lopes-Cendes I, Cendes F, Castellano G. MRI-Texture analysis of corpus callosum, thalamus, putamen, and caudate in Machado-Joseph disease. *J Neuroimaging*. 2010;22(1):46–52.
61. Sobana SA, Huda F, Hermawan R, Sribudiani Y, Koan TS, Dian S, Ong PA, Dahlan NL, Utami N, Pusparini I, et al. Brain MRI volumetry analysis in an Indonesian family of SCA 3 patients: a case-based study. *Front Neurol*. 2022;13:912592.
62. Heide S, Keren B, Billette de Villemeur T, Chantot-Bastaraud S, Depienne C, Nava C, Mignot C, Jacqueline A, Fonteneau E, Lejeune E, et al. Copy number variations found in patients with a corpus callosum abnormality and intellectual disability. *J Pediatr*. 2017;185:160–e166161.
63. Mesman S, Bakker R, Smidt MP. Tcf4 is required for correct brain development during embryogenesis. *Mol Cell Neurosci*. 2020;106: 103502.
64. Miyamoto S, Kato M, Hiraide T, Shiohama T, Goto T, Hojo A, Ebata A, Suzuki M, Kobayashi K, Chong PF, et al. Comprehensive genetic analysis confers high diagnostic yield in 16 Japanese patients with corpus callosum anomalies. *J Hum Genet*. 2021;66(11):1061–8.
65. Costa MC, Radzwion M, McLoughlin HS, Ashraf NS, Fischer S, Shakkottai VG, Maciel P, Paulson HL, Öz G. In vivo molecular signatures of cerebellar pathology in spinocerebellar ataxia type 3. *Mov Disord*. 2020;35(10):1774–86.
66. Schuster KH, DiFranco DM, Putka AF, Mato JP, Jarrah SI, Stec NR, Sundararajan VO, McLoughlin HS. Disease-associated oligodendrocyte signatures are spatiotemporally dysregulated in spinocerebellar ataxia type 3. *Front Neurosci*. 2023;17: 1118429.
67. Schuster KH, Zalon AJ, Zhang H, DiFranco DM, Stec NR, Haque Z, Blumenstein KG, Pierce AM, Guan Y, Paulson HL, et al. Impaired oligodendrocyte maturation is an early feature in SCA3 disease pathogenesis. *J Neurosci*. 2022;42(8):1604–17.
68. Schuster KH, Putka AF, McLoughlin HS. Pathogenetic mechanisms underlying spinocerebellar ataxia type 3 are altered in primary oligodendrocyte culture. *Cells*. 2022. <https://doi.org/10.3390/cells11162615>.
69. Duarte Lobo D, Nobre RJ, Oliveira Miranda C, Pereira D, Castelhan J, Sereno J, Koeppen A, Castelo-Branco M, Pereira de Almeida L. The blood-brain barrier is disrupted in Machado-Joseph disease/spinocerebellar ataxia type 3: evidence from transgenic mice and human post-mortem samples. *Acta Neuropathol Commun*. 2020;8(1):152.
70. Pedrosa JL, França MC, Braga-Neto P, D'Abreu A, Saraiva-Pereira ML, Saute JA, Teive HA, Caramelli P, Jardim LB, Lopes-Cendes I, et al. Nonmotor and extracerebellar features in Machado-Joseph disease: A review. *Mov Disord*. 2013;28(9):1200–8.
71. Roeske S, Filla I, Heim S, Amunts K, Helmstaedter C, Wüllner U, Wagner M, Klockgether T, Minnerop M. Progressive cognitive dysfunction in spinocerebellar ataxia type 3. *Mov Disord*. 2013;28(10):1435–8.
72. Hengel H, Martus P, Faber J, Giunti P, Garcia-Moreno H, Solanki N, Klockgether T, Reetz K, van de Warrenburg BP, Santana MM, et al. The frequency of non-motor symptoms in SCA3 and their association with disease severity and lifestyle factors. *J Neurol*. 2022;270(2):944–52.
73. Lopes TM, D'Abreu A, Junior MCF, Yasuda CL, Betting LE, Samara AB, Castellano G, Somazzi JC, Balthazar MLF, Lopes-Cendes I, et al. Widespread neuronal damage and cognitive dysfunction in spinocerebellar ataxia type 3. *J Neurol*. 2013;260(9):2370–9.
74. Braga-Neto P, Felicio AC, Hoexter MQ, Pedrosa JL, Dutra LA, Alessi H, Minetti T, Santos-Galduroz RF, da Rocha AJ, Garcia LAL, et al. Cognitive and olfactory deficits in Machado-Joseph disease: a dopamine transporter study. *Parkinsonism Relat Disord*. 2012;18(7):854–8.
75. Ye Z-X, Bi J, Qiu L-L, Chen X-Y, Li M-C, Chen X-Y, Qiu Y-S, Yuan R-Y, Yu X-T, Huang C-Y, et al. Cognitive impairment associated with cerebellar volume loss in spinocerebellar ataxia type 3. *J Neurol*. 2023;271(2):918–28.
76. Leotti VB, de Vries JJ, Oliveira CM, de Mattos EP, Te Meerman GJ, Brunt ER, Kampinga HH, Jardim LB, Verbeek DS. CAG repeat size influences the progression rate of spinocerebellar ataxia type 3. *Ann Neurol*. 2020;89(1):66–73.
77. Jardim L, Silveira I, Pereira ML, Do Céu Moreira M, Mendonça P, Sequeiros J, Giugliani R. Searching for modulating effects of SCA2, SCA6 and DRPLA CAG tracts on the Machado-Joseph disease (SCA3) phenotype. *Acta Neurol Scand*. 2003;107(3):211–4.
78. Tezenas du Montcel S, Durr A, Bauer P, Figueroa KP, Ichikawa Y, Brunsino A, Forlani S, Rakowicz M, Schöls L, Mariotti C, et al. Modulation of the age at onset in spinocerebellar ataxia by CAG tracts in various genes. *Brain*. 2014;137(9):2444–55.
79. Chen Z, Zheng C, Long Z, Cao L, Li X, Shang H, Yin X, Zhang B, Liu J, Ding D, et al. (CAG) n loci as genetic modifiers of age-at-onset in patients with Machado-Joseph disease from Mainland China. *Brain*. 2016;139(8):e41–41.
80. Chen Z, Wang C, Zheng C, Long Z, Cao L, Li X, Shang H, Yin X, Zhang B, Liu J, et al. Ubiquitin-related network underlain by (CAG) n loci modulate age at onset in Machado-Joseph disease. *Brain*. 2017;140(4):e25–25.
81. Le W, Long Z, Chen Z, Wang C, Huang F, Peng H, Peng H, Hou X, Ding D, Ye W, Wang J, et al. Two novel SNPs in ATXN3 3' UTR May decrease age at onset of SCA3/ MJD in Chinese patients. *PLoS ONE*. 2015;10(2):e0117488.
82. Melo ARV, Raposo M, Ventura M, Martins S, Pavão S, Alonso I, Bettencourt C, Lima M. Genetic variation in ATXN3 (Ataxin-3) 3'UTR: insights into the downstream regulatory elements of the causative gene of Machado-Joseph disease/spinocerebellar ataxia type 3. *Cerebellum*. 2022;22(1):37–45.
83. Ding D, Li K, Wang C, Chen Z, Long Z, Peng Y, Zhou X, Peng H, Qiu R, Xia K, et al. ATXN2 polymorphism modulates age at onset in Machado-Joseph disease. *Brain*. 2016;139(10):e59.

84. Peng H, Wang C, Chen Z, Sun Z, Jiao B, Li K, Huang F, Hou X, Wang J, Shen L, et al. The APOE  $\epsilon 2$  allele May decrease the age at onset in patients with spinocerebellar ataxia type 3 or Machado-Joseph disease from the Chinese Han population. *Neurobiol Aging*. 2014;35(9):e21792115–8.
85. Chen S, Gan SR, Cai PP, Ni W, Zhou Q, Dong Y, Wang N, Wu ZY. Mitochondrial NADH dehydrogenase subunit 3 polymorphism associated with an earlier age at onset in male Machado-Joseph disease patients. *CNS Neurosci Ther*. 2015;22(1):38–42.
86. Kazachkova N, Raposo M, Ramos A, Montiel R, Lima M. Promoter variant alters expression of the autophagic BECN1 gene: implications for clinical manifestations of Machado-Joseph disease. *Cerebellum*. 2017;16(5–6):957–63.
87. Ding D, Wang C, Chen Z, Peng H, Li K, Zhou X, Peng Y, Wang P, Hou X, Li T, et al. Polymorphisms in DNA methylation-related genes are linked to the phenotype of Machado-Joseph disease. *Neurobiol Aging*. 2019;75: e225221–8.
88. Ding D, Chen Z, Wang C, Tang X, Zhang L, Fang Q, Qiu R, Jiang H. A variant in genes of the NPY system as modifier factor of Machado-Joseph disease in the Chinese population. *Front Aging Neurosci*. 2022;14:822657.
89. Akçimen F, Martins S, Liao C, Bourassa CV, Catoire H, Nicholson GA, Riess O, Raposo M, França MC, Vasconcelos J, et al. Genome-wide association study identifies genetic factors that modify age at onset in Machado-Joseph disease. *Aging*. 2020;12(6):4742–56.
90. Raposo M, Bettencourt C, Melo ARV, Ferreira AF, Alonso I, Silva P, Vasconcelos J, Kay T, Saraiva-Pereira ML, Costa MD, et al. Novel Machado-Joseph disease-modifying genes and pathways identified by whole-exome sequencing. *Neurobiol Dis*. 2022;162: 105578.
91. Weber JJ, Cizisch L, Pereira Sena P, Fath F, Huridou C, Schwarz N, Incebacak Eltemur RD, Würth A, Weishäupl D, Döcker M, et al. The parkin V380L variant is a genetic modifier of Machado-Joseph disease with impact on mitophagy. *Acta Neuropathol*. 2024;148(1):14.
92. Lee YH, Tsai YS, Chang CC, Ho CC, Shih HM, Chen HM, Lai HL, Lee CW, Lee YC, Liao YC, et al. A PIAS1 protective variant S510G delays PolyQ disease onset by modifying protein homeostasis. *Mov Disord*. 2021;37(4):767–77.
93. Chang Y-C, Tsai Y-C, Chang E-C, Hsu Y-C, Huang Y-R, Lee Y-H, Tsai Y-S, Chen Y-Q, Lee Y-C, Liao Y-C, et al. PIAS1 S510G variant acts as a genetic modifier of spinocerebellar ataxia type 3 by selectively impairing mutant ataxin-3 proteostasis. *Int J Biochem Cell Biol*. 2024;176: 106662.
94. Laffita-Mesa JM, Nennesmo I, Paucar M, Svenningsson P. A novel duplication in ATXN2 as modifier for spinocerebellar ataxia 3 (SCA3) and C9ORF72-ALS. *Mov Disord*. 2020;36(2):508–14.
95. Peng L, Chen Z, Chen T, Lei L, Long Z, Liu M, Deng Q, Yuan H, Zou G, Wan L, et al. Prediction of the age at onset of spinocerebellar ataxia type 3 with machine learning. *Mov Disord*. 2020;36(1):216–24.
96. Peng L, Chen Z, Long Z, Liu M, Lei L, Wang C, Peng H, Shi Y, Peng Y, Deng Q, et al. New model for estimation of the age at onset in spinocerebellar ataxia type 3. *Neurology*. 2021;96(23):e2885–95.
97. Chintalaphani SR, Pineda SS, Deveson IW, Kumar KR. An update on the neurological short tandem repeat expansion disorders and the emergence of long-read sequencing diagnostics. *Acta Neuropathol Commun*. 2021;9(1):98.
98. Warburton PE, Sebra RP. Long-read DNA sequencing: recent advances and remaining challenges. *Annu Rev Genom Hum Genet*. 2023;24(1):109–32.
99. Kernohan KD, Boycott KM. The expanding diagnostic toolbox for rare genetic diseases. *Nat Rev Genet*. 2024;25(6):401–15.
100. Olivucci G, Iovino E, Innella G, Turchetti D, Pippucci T, Magini P. Long read sequencing on its way to the routine diagnostics of genetic diseases. *Front Genet*. 2024;15: 1374860.
101. Zhao X, Collins RL, Lee W-P, Weber AM, Jun Y, Zhu Q, Weisburd B, Huang Y, Audano PA, Wang H, et al. Expectations and blind spots for structural variation detection from long-read assemblies and short-read genome sequencing technologies. *Am J Hum Genet*. 2021;108(5):919–28.
102. Ahsan MU, Liu Q, Perdomo JE, Fang L, Wang K. A survey of algorithms for the detection of genomic structural variants from long-read sequencing data. *Nat Methods*. 2023;20(8):1143–58.
103. Su Y, Fan L, Shi C, Wang T, Zheng H, Luo H, Zhang S, Hu Z, Fan Y, Dong Y, et al. Deciphering neurodegenerative diseases using long-read sequencing. *Neurology*. 2021;97(9):423–33.
104. Mizuguchi T, Toyota T, Miyatake S, Mitsuhashi S, Doi H, Kudo Y, Kishida H, Hayashi N, Tsuburaya RS, Kinoshita M, et al. Complete sequencing of expandedSAMD12repeats by long-read sequencing and Cas9-mediated enrichment. *Brain*. 2021;144(4):1103–17.
105. Hiatt SM, Lawlor JMJ, Handley LH, Latner DR, Bonnstetter ZT, Finnila CR, Thompson ML, Boston LB, Williams M, Rodríguez Nunez I, et al. Long-read genome sequencing and variant reanalysis increase diagnostic yield in neurodevelopmental disorders. *Genome Res*. 2024;34(11):1747–62.
106. Ebbert MTW, Farrugia SL, Sens JP, Jansen-West K, Gendron TF, Prudencio M, McLaughlin IJ, Bowman B, Seetin M, DeJesus-Hernandez 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.
107. Stevanovski I, Chintalaphani SR, Gamaarachchi H, Ferguson JM, Pineda SS, Scriba CK, Tchan M, Fung V, Ng K, Cortese A, et al. Comprehensive genetic diagnosis of tandem repeat expansion disorders with programmable targeted nanopore sequencing. *Sci Adv*. 2022;8(9): eabm5386.
108. Rosenbohm A, Pott H, Thomsen M, Rafahi H, Kaya S, Szymczak S, Volk AE, Mueller K, Silveira I, Weishaupt JH, et al. Familial cerebellar ataxia and amyotrophic lateral sclerosis/frontotemporal dementia with DAB1 and C9ORF72 repeat expansions: an 18-year study. *Mov Disord*. 2022;37(12):2427–39.
109. Zou J, Wang F, Gong Z, Wang R, Chen S, Zhang H, Sun R, Gao C, Li W, Shang J, et al. A Chinese SCA36 pedigree analysis of NOP56 expansion region based on long-read sequencing. *Front Genet*. 2023;14: 1110307.
110. Sanchez-Flores M, Corral-Juan M, Gasch-Navalón E, Cirillo D, Sanchez I, Matilla-Dueñas A. Novel genotype-phenotype correlations, differential cerebellar allele-specific methylation, and a common origin of the (ATTC)n insertion in spinocerebellar ataxia type 37. *Hum Genet*. 2024;143(3):211–32.
111. Du Y-C, Dong Y, Cheng H-L, Li Q-F, Yang L, Shao Y-R, Ma Y, Ni W, Gan S-R, Wu Z-Y. Genotype-phenotype correlation in 667 Chinese families with spinocerebellar ataxia type 3. *Parkinsonism Relat Disord*. 2020;78:116–21.
112. Sharafi S, Rezvani Z. Investigation of spinocerebellar ataxia (SCA) disease in Iranian patients and accurate trinucleotide repeat detection in the SCA3 by TP-PCR method. *Mol Neurobiol*. 2024;62(3):2756–63.
113. Lian M, Zhao M, Phang G-P, Soong Y-T, Yoon C-S, Lee CG, Law H-Y, Chong SS. Rapid molecular screen of spinocerebellar ataxia types 1, 2, and 3 by triplet-primed PCR and melting curve analysis. *J Mol Diagn*. 2021;23(5):565–76.
114. Melo ARV, Ramos A, Kazachkova N, Raposo M, Bettencourt BF, Rendeiro AR, Kay T, Vasconcelos J, Bruges-Armas J, Lima M. Triplet repeat primed PCR (TP-PCR) in molecular diagnostic testing for spinocerebellar ataxia type 3 (SCA3). *Mol Diagn Ther*. 2016;20(6):617–22.
115. Méreaux J-L, Davoine C-S, Coutelier M, Guillot-Noël L, Castrioto A, Charles P, Coarelli G, Ewencyk C, Klebe S, Heinzmann A, et al. Fast and reliable detection of repeat expansions in spinocerebellar ataxia using exomes. *J Med Genet*. 2023;60(7):717–21.
116. Ibañez K, Polke J, Hagelstrom RT, Dolzhenko E, Pasko D, Thomas ERA, Daugherty LC, Kasperaviciute D, Smith KR, Deans ZC, et al. Whole genome sequencing for the diagnosis of neurological repeat expansion disorders in the UK: a retrospective diagnostic accuracy and prospective clinical validation study. *Lancet Neurol*. 2022;21(3):234–45.
117. Xi J, Wang X, Yue D, Dou T, Wu Q, Lu J, Liu Y, Yu W, Qiao K, Lin J, et al. 5' UTR CGG repeat expansion in GIPC1 is associated with oculopharyngodistal myopathy. *Brain*. 2021;144(2):601–14.
118. Sun Q-Y, Xu Q, Tian Y, Hu Z-M, Qin L-X, Yang J-X, Huang W, Xue J, Li J-C, Zeng S, et al. Expansion of GGC repeat in the human-specific NOTCH2NL gene is associated with essential tremor. *Brain*. 2020;143(1):222–33.
119. Tian Y, Wang J-L, Huang W, Zeng S, Jiao B, Liu Z, Chen Z, Li Y, Wang Y, Min H-X, et al. Expansion of human-specific GGC repeat in neuronal intranuclear inclusion disease-related disorders. *Am J Hum Genet*. 2019;105(1):166–76.
120. Zeng S, Zhang M-y, Wang X-j, Hu Z-m, Li J-c, Li N, Wang J-l, Liang F, Yang Q, Liu Q, et al. Long-read sequencing identified intronic repeat expansions inSAMD12from Chinese pedigrees affected with familial cortical myoclonic tremor with epilepsy. *J Med Genet*. 2019;56(4):265–70.
121. Baviera-Muñoz R, Carretero-Vilarroig L, Vázquez-Costa JF, Morata-Martínez C, Campins-Romeu M, Muelas N, Sastre-Bataller I, Martínez-Torres I, Pérez-García J, Sivera R, et al. Diagnostic efficacy of genetic studies in a series of hereditary cerebellar ataxias in eastern Spain. *Neurol Genet*. 2022;8(6): e200038.
122. Wan N, Chen Z, Wan L, Yuan H, Tang Z, Liu M, Peng Y, Peng L, Lei L, Xie Y, et al. Genetic etiology of a Chinese ataxia cohort: expanding the mutational spectrum of hereditary ataxias. *Parkinsonism Relat Disord*. 2021;89:120–7.
123. Kim M, Kim AR, Kim JS, Park J, Youn J, Ahn JH, Mun JK, Lee C, Kim N-S, Kim NKD, et al. Clarification of undiagnosed ataxia using whole-exome sequencing with clinical implications. *Parkinsonism Relat Disord*. 2020;80:58–64.
124. Yu AC-S, Yim AK-Y, Chan AY-Y, Yuen LYP, Au WC, Cheng THT, Lin X, Li J-W, Chan LWL, Mok VCT, et al. A targeted gene panel that covers coding, Non-coding and short tandem repeat regions improves the diagnosis of patients with neurodegenerative diseases. *Front Neurosci*. 2019;13:1324.

125. Tankard RM, Bennett MF, Degorski P, Delatycki MB, Lockhart PJ, Bahlo M. Detecting expansions of tandem repeats in cohorts sequenced with short-read sequencing data. *Am J Hum Genet.* 2018;103(6):858–73.
126. Guo F, Liu R, Pan Y, Collins C, Bean L, Ma Z, Mathur A, Da Silva C, Nallamilli B, Guraju N, et al. Evidence from 2100 index cases supports genome sequencing as a first-tier genetic test. *Genet Med.* 2024;26(1): 100995.
127. Souche E, Beltran S, Brosens E, Belmont JW, Fossum M, Riess O, Gilissen C, Ardesheidavani A, Houge G, van Gijn M, et al. Recommendations for whole genome sequencing in diagnostics for rare diseases. *Eur J Hum Genet.* 2022;30(9):1017–21.
128. Rajan-Babu I-S, Peng JJ, Chiu R, Birch P, Couse M, Guimond C, Lehman A, Mwenifumbo J, van Karnebeek C, Friedman J, et al. Genome-wide sequencing as a first-tier screening test for short tandem repeat expansions. *Genome Med.* 2021;13(1):126.
129. Hu Y, Deng L, Zhang J, Fang X, Mei P, Cao X, Lin J, Wei Y, Zhang X, Xu R. A pooling genome-wide association study combining a pathway analysis for typical sporadic parkinson's disease in the Han population of Chinese Mainland. *Mol Neurobiol.* 2015;53(7):4302–18.
130. Rike WA, Stern S. Proteins and transcriptional dysregulation of the brain extracellular matrix in Parkinson's disease: a systematic review. *Int J Mol Sci.* 2023. <https://doi.org/10.3390/ijms24087435>.
131. Dourlen P, Kilinc D, Malmanche N, Chapuis J, Lambert J-C. The new genetic landscape of alzheimer's disease: from amyloid cascade to genetically driven synaptic failure hypothesis? *Acta Neuropathol.* 2019;138(2):221–36.
132. Santiago JA, Bottero V, Potashkin JA. Transcriptomic and network analysis identifies shared and unique pathways across dementia spectrum disorders. *Int J Mol Sci.* 2020. <https://doi.org/10.3390/ijms21062050>.
133. McCarthy NS, Melton PE, Ward SV, Allan SM, Dragovic M, Clark ML, Morar B, Rubio JP, Blangero J, Badcock JC, et al. Exome array analysis suggests an increased variant burden in families with schizophrenia. *Schizophr Res.* 2017;185:9–16.
134. Dityatev A, Schachner M, Sonderegger P. The dual role of the extracellular matrix in synaptic plasticity and homeostasis. *Nat Rev Neurosci.* 2010;11(11):735–46.
135. Cope EC, Gould E. Adult neurogenesis, glia, and the extracellular matrix. *Cell Stem Cell.* 2019;24(5):690–705.
136. Fawcett JW, Ohashi T, Pizzorusso T. The roles of perineuronal nets and the perinodal extracellular matrix in neuronal function. *Nat Rev Neurosci.* 2019;20(8):451–65.
137. Rodrigues A-J, do Carmo Costa M, Silva T-L, Ferreira D, Bajanica F, Logarinho E, Maciel P. Absence of ataxin-3 leads to cytoskeletal disorganization and increased cell death. *Biochim Et Biophys Acta (BBA) - Mol Cell Res.* 2010;1803(10):1154–63.
138. Neves-Carvalho A, Logarinho E, Freitas A, Duarte-Silva S, Costa MdC, Silva-Fernandes A, Martins M, Serra SC, Lopes AT, Paulson HL, et al. Dominant negative effect of polyglutamine expansion perturbs normal function of ataxin-3 in neuronal cells. *Hum Mol Genet.* 2015;24(1):100–17.
139. Wiatr K, Piasecki P, Marczak Ł, Wojciechowski P, Kurkowiak M, Płoski R, Rydzanicz M, Handschuh L, Jungverdorben J, Brüstle O, et al. Altered levels of proteins and phosphoproteins, in the absence of early causative transcriptional changes, shape the molecular pathogenesis in the brain of young pre-symptomatic K91 SCA3/MJD mouse. *Mol Neurobiol.* 2019;56(12):8168–202.
140. Wiatr K, Marczak Ł, Pérot J-B, Brouillet E, Flament J, Figiel M. Broad influence of mutant Ataxin-3 on the proteome of the adult brain, young neurons, and axons reveals central molecular processes and biomarkers in SCA3/MJD using knock-in mouse model. *Front Mol Neurosci.* 2021;14:658339.
141. Eva R, Fawcett J. Integrin signalling and traffic during axon growth and regeneration. *Curr Opin Neurobiol.* 2014;27:179–85.
142. Seetharaman S, Etienne-Manneville S. Cytoskeletal crosstalk in cell migration. *Trends Cell Biol.* 2020;30(9):720–35.
143. Hübener J, Casadei N, Teismann P, Seeliger MW, Björkqvist M, von Hörsten S, Riess O, Nguyen HP. Automated behavioral phenotyping reveals presymptomatic alterations in a SCA3 genetrapped mouse model. *J Genet Genomics.* 2012;39(6):287–99.
144. Chapman TP, Corridoni D, Shiraishi S, Pandey S, Aulicino A, Wigfield S, do Carmo Costa M, Thézénas M-L, Paulson H, Fischer R, et al. Ataxin-3 links NOD2 and TLR2 mediated innate immune sensing and metabolism in myeloid cells. *Front Immunol.* 2019;10:1495.
145. Sowa AS, Haas E, Hübener-Schmid J, Lorentz A. Ataxin-3, the spinocerebellar ataxia type 3 neurodegenerative disorder protein, affects mast cell functions. *Front Immunol.* 2022;13: 870966.
146. Evert BO, Vogt IR, Kindermann C, Ozimek L, de Vos RAI, Brunt ERP, Schmitt I, Klockgether T, Wüllner U. Inflammatory genes are upregulated in expanded Ataxin-3-expressing cell lines and spinocerebellar ataxia type 3 brains. *J Neurosci.* 2001;21(15):5389–96.
147. Evert BO, Vogt IR, Vieira-Saecker AM, Ozimek L, de Vos RAI, Brunt ERP, Klockgether T, Wüllner U. Gene expression profiling in Ataxin-3 expressing cell lines reveals distinct effects of normal and mutant Ataxin-3. *J Neuropathol Exp Neurol.* 2003;62(10):1006–18.
148. Switonski PM, Słachcic WJ, Krzyzosiak WJ, Figiel M. A new humanized ataxin-3 knock-in mouse model combines the genetic features, pathogenesis of neurons and glia and late disease onset of SCA3/MJD. *Neurobiol Dis.* 2015;73:174–88.
149. Yamada M, Sato T, Shimohata T, Hayashi S, Igarashi S, Tsuji S, Takahashi H. Interaction between neuronal intranuclear inclusions and promyelocytic leukemia protein nuclear and coiled bodies in CAG repeat diseases. *Am J Pathol.* 2001;159(5):1785–95.
150. Sun J, Xu H, Negi S, Subramony SH, Hebert MD. Differential effects of polyglutamine proteins on nuclear organization and artificial reporter splicing. *J Neurosci Res.* 2007;85(11):2306–17.
151. Mykowska A, Sobczak K, Wojciechowska M, Kozłowski P, Krzyzosiak WJ. CAG repeats mimic CUG repeats in the misregulation of alternative splicing. *Nucleic Acids Res.* 2011;39(20):8938–51.
152. Bourel J, Planche V, Dubourdiou N, Oliveira A, Séré A, Ducourneau E-G, Tible M, Maitre M, Lesté-Lasserre T, Nadjar A, et al. Complement C3 mediates early hippocampal neurodegeneration and memory impairment in experimental multiple sclerosis. *Neurobiol Dis.* 2021;160: 105533.
153. Schartz ND, Aroor A, Li Y, Pinzón-Hoyos N, Brewster AL. Mice deficient in complement C3 are protected against recognition memory deficits and astrogliosis induced by status epilepticus. *Front Mol Neurosci.* 2023;16:1265944.
154. Gorelik A, Sapir T, Ben-Reuven L, Reiner O. Complement C3 affects Rac1 activity in the developing brain. *Front Mol Neurosci.* 2018;11:150.
155. Sun H, Ma L, Zhang Y, Pan X, Wang C, Zhang J, Zhang X, Sun H, Wang Q, Zhu W. A purinergic P2 receptor family-mediated increase in Thrombospondin-1 bolsters synaptic density and epileptic seizure activity in the amygdala-kindling rat model. *Front Cell Neurosci.* 2018;12:302.
156. Kim DH, Lim H, Lee D, Choi SJ, Oh W, Yang YS, Oh JS, Hwang HH, Jeon HB. Thrombospondin-1 secreted by human umbilical cord blood-derived mesenchymal stem cells rescues neurons from synaptic dysfunction in Alzheimer's disease model. *Sci Rep.* 2018;8(1):354.
157. Flierl MA, Gaudiani JL, Sabel AL, Long CS, Stahel PF, Mehler PS. Complement C3 serum levels in anorexia nervosa: a potential biomarker for the severity of disease? *Ann Gen Psychiatry.* 2011;10(1):16.
158. Li H, Zhang Q, Li N, Wang F, Xiang H, Zhang Z, Su Y, Huang Y, Zhang S, Zhao G, et al. Plasma levels of Th17-related cytokines and complement C3 correlated with aggressive behavior in patients with schizophrenia. *Psychiatry Res.* 2016;246:700–6.
159. Dong X-Q, Yu W-H, Zhu Q, Cheng Z-Y, Chen Y-H, Lin X-F, Ten X-L, Tang X-B, Chen J. Changes in plasma thrombospondin-1 concentrations following acute intracerebral hemorrhage. *Clin Chim Acta.* 2015;450:349–55.
160. Wang JL, Jin GL, Yuan ZG, Yu XB, Li JQ, Qiu TL, Dai RX. Plasma thrombospondin-1 and clinical outcomes in traumatic brain injury. *Acta Neurol Scand.* 2016;134(3):189–96.
161. Otomo A. ALS2, a novel guanine nucleotide exchange factor for the small GTPase Rab5, is implicated in endosomal dynamics. *Hum Mol Genet.* 2003;12(14):1671–87.
162. Lai C, Xie C, McCormack SG, Chiang H-C, Michalak MK, Lin X, Chandran J, Shim H, Shimoji M, Cookson MR, et al. Amyotrophic lateral sclerosis 2-deficiency leads to neuronal degeneration in amyotrophic lateral sclerosis through altered AMPA receptor trafficking. *J Neurosci.* 2006;26(45):11798–806.
163. Jacquier A, Buhler E, Schäfer MKE, Bohl D, Blanchard S, Beclin C, Haase G. Alsln/Rac1 signaling controls survival and growth of spinal motoneurons. *Ann Neurol.* 2006;60(1):105–17.
164. Cai H, Lin X, Xie C, Laird FM, Lai C, Wen H, Chiang H-C, Shim H, Farah MH, Hoke A, et al. Loss of ALS2 function is insufficient to trigger motor neuron degeneration in knock-out mice but predisposes neurons to oxidative stress. *J Neurosci.* 2005;25(33):7567–74.
165. Hadano S, Hand CK, Osuga H, Yanagisawa Y, Otomo A, Devon RS, Miyamoto N, Showguchi-Miyata J, Okada Y, Singaraja R, et al. A gene encoding a putative GTPase regulator is mutated in familial amyotrophic lateral sclerosis 2. *Nat Genet.* 2001;29(2):166–73.

166. Thevenon J, Lopez E, Keren B, Heron D, Mignot C, Altuzarra C, Béri-Dexheimer M, Bonnet C, Magnin E, Burglen L, et al. Intragenic *CAMTA1* rearrangements cause non-progressive congenital ataxia with or without intellectual disability. *J Med Genet*. 2012;49(6):400–8.
167. Wijnen IGM, Veenstra-Knol HE, Vansenne F, Gerkes EH, de Koning T, Vos YJ, Tijssen MAJ, Sival D, Darin N, Vanhoutte EK, et al. De novo variants in *CAMTA1* cause a syndrome variably associated with spasticity, ataxia, and intellectual disability. *Eur J Hum Genet*. 2020;28(6):763–9.
168. Long C, Grueter CE, Song K, Qin S, Qi X, Kong YM, Shelton JM, Richardson JA, Zhang C-L, Bassel-Duby R, et al. Ataxia and Purkinje cell degeneration in mice lacking the *CAMTA1* transcription factor. *Proc Natl Acad Sci U S A*. 2014;111(31):11521–6.
169. Ng BG, Sosicka P, Agadi S, Almannai M, Bacino CA, Barone R, Botto LD, Burton JE, Carlston C, Chung BHY, et al. *SLC35A2*-CDG: functional characterization, expanded molecular, clinical, and biochemical phenotypes of 30 unreported individuals. *Hum Mutat*. 2019;40(7):908–25.
170. Bonduelle T, Hartlieb T, Baldassari S, Sim NS, Kim SH, Kang H-C, Kobow K, Coras R, Chipaux M, Dorfmueller G, et al. Frequent *SLC35A2* brain mosaicism in mild malformation of cortical development with oligodendroglial hyperplasia in epilepsy (MOGHE). *Acta Neuropathol Commun*. 2021;9(1): 3.
171. Perez-Sen R, Gomez-Villafuertes R, Ortega F, Gualix J, Delicado EG, Miras-Portugal MT. An update on P2Y(13) receptor signalling and function. *Adv Exp Med Biol*. 2017;1051:139–68.
172. Lima M, Raposo M, Ferreira A, Melo ARV, Pavão S, Medeiros F, Teves L, Gonzalez C, Lemos J, Pires P, et al. The homogeneous Azorean Machado-Joseph disease cohort: characterization and contributions to advances in research. *Biomedicines*. 2023;11(2): 247.

## Publisher's note

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