Neurobiology of Aging 122 (2023) 76–87

Contents lists available at ScienceDirect

Neurobiology of Aging

journal homepage: www.elsevier.com/locate/neuaging.org

Pathological Aging

Whole genome sequencing analysis reveals post-zygotic mutation variability in monozygotic twins discordant for amyotrophic lateral sclerosis

Gijs H.P. Tazelaar[a,*], Paul J. Hop[a], Meinie Seelen[a], Joke J.F.A. van Vugt[a], Wouter van Rheenen[a], Lindy Kool[a], Kristel R. van Eijk[a], Marleen Gijzen[b], Dennis Dooijes[b], Matthieu Moisse[c,d], Andrea Calvo[e,f,g], Cristina Moglia[e,f,g], Maura Brunetti[e,f,g], Antonio Canosa[e,f,g], Angelica Nordin[h], Jesus S. Mora Pardina[i], John Ravits[j], Ammar Al-Chalabi[k,l], Adriano Chio[e,f,g], Russell L. McLaughlin[m], Orla Hardiman[n,o], Philip Van Damme[c,d], Mamede de Carvalho[p,q], Christoph Neuwirth[r], Markus Weber[r], Peter M Andersen[h], Leonard H. van den Berg[a], Jan H. Veldink[a], Michael A. van Es[a,*]

[a] Department of Neurology and Neurosurgery, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht, the Netherlands
[b] Department of Genetics, University Medical Center Utrecht, Utrecht, the Netherlands
[c] Neurology Department University Hospitals Leuven, Department of Neurosciences and Leuven Brain Institute (LBI) KU Leuven—University of Leuven, Leuven, Belgium
[d] VIB, Center for Brain & Disease Research, Leuven, Belgium
[e] ALS Centre, "Rita Levi Montalcini" Department of Neuroscience, University of Turin, Turin, Italy
[f] Azienda Ospedaliero-Universitaria Città della Salute e della Scienza di Torino, SC Neurologia 1U, Turin, Italy
[g] Neuroscience Institute of Turin (NIT), Turin, Italy.
[h] Department of Clinical Science, Neurosciences, Umeå University Umeå, Sweden
[i] ALS Unit, Hospital San Rafael, Madrid, Spain.
[j] Department of Neurosciences, University of California at San Diego, La Jolla, CA, USA.
[k] Department of Basic and Clinical Neuroscience, Maurice Wohl Clinical Neuroscience Institute and United Kingdom Dementia Research Institute, King's College London, London, UK
[l] Department of Neurology, King's College Hospital, London, UK
[m] Population Genetics Laboratory, Smurfit Institute of Genetics, Trinity College Dublin, Republic of Ireland
[n] Academic Unit of Neurology, Trinity College Dublin, Trinity Biomedical Sciences Institute, Dublin, Republic of Ireland
[o] Department of Neurology, Beaumont Hospital, Dublin, Republic of Ireland
[p] Department of Neurosciences, Hospital de Santa Maria-CHLN, Lisbon, Portugal
[q] Institute of Physiology, Institute of Molecular Medicine, Faculty of Medicine, University of Lisbon, Lisbon, Portugal
[r] Neuromuscular Diseases Unit / ALS Clinic, Kantonsspital St.Gallen, St.Gallen, Switzerland

ARTICLE INFO

Article history:
Received 11 May 2022
Revised 23 October 2022
Accepted 8 November 2022
Available online 17 November 2022

Keywords:
Amyotrophic Lateral Sclerosis
Post-zygotic mutations
Genetic modifiers
Repeat expansions

ABSTRACT

Amyotrophic lateral sclerosis is a heterogeneous, fatal neurodegenerative disease, characterized by motor neuron loss and in 50% of cases also by cognitive and/or behavioral changes. Mendelian forms of ALS comprise approximately 10-15% of cases. The majority is however considered sporadic, but also with a high contribution of genetic risk factors. To explore the contribution of somatic mutations and/or epigenetic changes to disease risk, we performed whole genome sequencing and methylation analyses using samples from multiple tissues on a cohort of 26 monozygotic twins discordant for ALS, followed by in-depth validation and replication experiments. The results of these analyses implicate several mechanisms in ALS pathophysiology, which include a role for de novo mutations, defects in DNA damage repair and accelerated aging.

© 2022 The Author(s). Published by Elsevier Inc.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

Abbreviations: ALS, Amyotrophic lateral sclerosis; CNV, copy number variant; DMPs, differentially methylated CpG positions; FTD, frontotemporal dementia; GWAS, Genome-wide association studies; pSNM, post-zygotic single nucleotide mutations; SNVs, single nucleotide variants; WES, whole exome sequencing; WGS, whole genome sequencing.

* Corresponding authors at: Department of Neurology and Neurosurgery, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht, the Netherlands. Tel.: 00 31 88 7557939; Fax: 00 31 30 2542100.

0197-4580/$ – see front matter © 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)
https://doi.org/10.1016/j.neurobiolaging.2022.11.010

1. Introduction

Amyotrophic lateral sclerosis (ALS) is a lethal neurodegenerative disorder characterized by the loss of motor neurons, resulting in progressive muscle weakness, spasticity, and ultimately respiratory failure (Brown and Al-Chalabi, 2017; van Es et al., 2017). Approximately, 50% of patients also exhibit cognitive and/or behavioral changes within the spectrum of frontotemporal dementia (FTD).

In 5%–15% of patients there is a positive family history. Over the past 15 years many familial ALS genes have been identified and currently about 70% of all familial cases can be explained mutations in >20 genes (Andersen and Al-Chalabi, 2011). Thus, the genetics of familial ALS are now relatively well understood. The majority of patients are however sporadic (with a negative family history), which is thought to be a multifactorial disease caused by both environmental and genetic risk factors. Twin and family heritability studies have estimated the genetic contribution to the risk of sporadic ALS to be between 40 and 60% (Al-Chalabi et al., 2010; Ryan et al., 2019; Wingo et al., 2011). Mutations in familial ALS genes are found in ±10% of apparently sporadic patients, but these are likely mislabeled familial cases due to factors such as non-penetrance, incomplete family histories, small family sizes and phenotypic heterogeneity. Genome-wide association studies (GWAS) have also successfully identified multiple common genetic risk factors, but which unfortunately only partially explain its heritability.

So despite the many discoveries and the relatively large heritability of sporadic ALS its genetic underpinnings remain incompletely understood. Recent germline mutations might form an explanation for the "missing heritability" in GWAS and somatic mosaicism resulting from post-zygotic mutations could explain the sporadic nature of the disease (Acuna-Hidalgo et al., 2015).

Indeed, de novo mutations have been identified in other neurological and psychiatric disorders with a substantial genetic component, especially in young-onset diseases such as autism, intellectual disability, epilepsy and schizophrenia (Gilissen et al., 2014; Wang et al., 2019), and to lesser extent in late-onset diseases such as Parkinson's disease (Kun-Rodrigues et al., 2015). In ALS, there are reports of de novo mutations in known ALS genes in sporadic cases (Chio et al., 2011; DeJesus-Hernandez et al., 2010; Kim et al., 2015; Laffita-Mesa et al., 2013) and similarly copy number variant (CNV), whole exome (WES) and genome sequencing (WGS) ALS trio studies also found de novo mutations (Chesi et al., 2013; Pamphlett and Morahan, 2011; Steinberg et al., 2015; van Doormaal et al., 2017).

In order to further assess the contribution of de novo mutations and epigenetic changes to ALS, we set up a large international cohort of 21 monozygotic twins discordant for ALS and compared their WGS and DNA methylation data and replicated the findings in an additional set of 5 twin pairs.

2. Materials and methods

2.1. Monozygotic twins

Monozygotic twins were selected from an ongoing prospective, population-based study on ALS in The Netherlands as well as from ALS databases throughout Europe and the United States of America. Monozygosity was later on confirmed using obtained whole genome sequencing data. Discordancy was based on evaluation by a neurologist of the unaffected twin confirming the absence of motor neuron disease after a period of at least 5 years after disease onset in the affected sibling. Sporadic patients had no first or second-degree relatives with motor neuron disease. Genomic DNA was extracted from whole blood, saliva and or fibroblast biopsy by means of standard procedures. All material was obtained with the ethical approval of the relevant institutional review boards. All subjects provided written informed consent.

2.2. Whole genome sequencing, variant detection and validation

DNA samples were sequenced using PCR free library preparation and paired-end (100bp) sequencing on the HiSeq 2000 platform (Illumina, San Diego, Illumina). Reads were aligned to the hg19 human genome build using BWA alignment software. Three different variant callers were used to obtain variants; Isaac's, Strelka2, and GATK. Illumina Isaac's Variant Calling and Illumina Strelka2 Somatic Variant Calling (v2.0.14) were performed by Illumina and filtered on standard Illumina criteria. For variant calling using the Genome Analysis Toolkit (https://www.broadinstitute.org/gatk) aligned files were first preprocessed and then called using the HaplotypeCaller (v3.4) method followed by Variant Recalibration according to best practice recommendations and parameters with Truth Sensitivity Tranche at VSQ 99.5 for SNVs and 99.0 for Indels and coverage threshold >10. Detected SNVs were compared per twin pair and SNVs with discordant genotypes were selected after variant filtering according to best practice guidelines, per pair comparison for 21 twins resulted in a list of over 8 million discordant single nucleotide variants (Supplementary Figure 1a). We excluded multi-allelic SNVs, homozygous variant versus homozygous reference and homozygous variant versus heterozygous variant from further analysis, given the uncertainty as to which one would be de novo and high probability of genotyping errors. To minimalize failed variant detection in the reference call we additionally filtered for reference genotype quality using genome variant files (Phred Scaled Qualityscore of 30 or higher) and if the same SNV was detected in one of the other twins of the opposite phenotype it was excluded, since it could not only be considered highly unlikely, but also be considered as not contributing to the phenotypic discordance. This resulted in a list of almost 150,000 SNVs.

We performed 2 separate validation steps using 2 different methods of next generation sequencing. First, the obtained list of SNVs was validated on a custom designed 650K Affymetrix Axiom myDesign Genotyping Array after exclusion of SNVs in ambiguous regions. Array results were analyzed using standard quality control and genotyping parameters with AxiomGTv1 algorithm. These results provided us with a possible true variant dataset of 1,124 SNVs. Give the high performance status of the Strelka2 somatic variant calling, but the limited Quality control data available from the VCF files for further filtering, we subsequently re-analyzed the entire data set using another somatic calling method GATK - The Mutect2 (v3.4). This provided us with additional quality control data and potential new candidates not picked up by Strelka2.

With a "candidate de novo" list available, we re-evaluated the discordant WGS data (now including Mutect2) and applied additional filters based on 4 "confidence" criteria: (1) Multiple caller criteria: Somatic variant calling method + 1 other method; (2) Additional quality criteria: Mapping quality score (MAPQ) > 50 & Genotype quality score (GQ) > 80; (3) Strandbias criteria: Strand Odds Ratio (SOR) < 2 & FisherStrand (FS) < 10; and (4) Allelic fraction (AF) criteria: allelic fraction of variant in variant call above 0.1 & allelic fraction of variant in reference call below 0.1. For Indels only, we additionally excluded Indels positioned in low complexity regions (LCR) (obtained from UCSC Genome Browser, https://genome.ucsc.edu/). We additionally selected SNVs and indels using the following 4 "impact criteria": (1) Previously vali-

E-mail address: m.a.vanes@umcutrecht.nl (M.A. van Es).

dated in the genotyping array; (2) Consequence prediction based on the Ensembl Variant Effect Predictor (VEP) annotation (transcript ablation, frameshift, splice acceptor/donor, missense, stop gain, stop/start lost, in frame insertion/deletion, protein altering or incomplete terminal codon variants); (3) High allelic fraction of variant (AF > 0.3) (meaning possible early post-zygotic). Last, we designed a Agilent SureSelect XT custom bait library for target sequencing of regions in which SNVs were located and subsequently performed a high-depth targeted sequencing using paired-end (150bp) sequencing on the Illumina HiSeq 4000 platform and performed somatic variant calling using best practice guidelines of GATK-Mutect2 and Strelka2.

2.3. Regulatory function analysis

Variants were mapped to a gene using gene regulatory networks that were derived from Hi-C and ATAC sequencing data from central nervous system cell-types from different sources and combined with Genehancer (Fishilevich et al., 2017). Transcription factor (TF) binding sites were identified using FIMO (*p*-value threshold 10^{-4}) with TF motifs from the JASPAR database.

2.4. Mutation signature analysis

Mutation signature analysis and graphs were performed using and according to the publicly available methods of the R package MutationPattern (Blokzijl et al., 2018).

2.5. WGS de novo mutation prediction

In order to determine all possible de novo calls in the second twin WGS dataset, we used a random forest classifier using the caret model and randomForest R libraries. First, since all of possible de novo SNVs were detected by a somatic variant caller and 95% (948/995) were detected by both, we selected all discordant variants that were picked up by both Mutect2 and Strelka2 (n=9413). This provided us with the following features (and a minimum of missing data (n = 5)): QSS = Quality score for any somatic SNV, that is, for the ALT allele to be present at a significantly different frequency between twins (Mutect2 & Strelka2); QSS_NT = Quality score reflecting the joint probability of a somatic variant and genotype of the normal sample (Strelka2); QSS_REF = Quality score in reference call (Mutect2); NLOD = Normal LOD score (Mutect2); TLOD = Tumor LOD score (Mutect2); AF_ALT = Allelic fraction of variant in variant call (Mutect2 & Strelka 2); AF_REF = Allelic fraction of variant in reference call (Mutect2); AF_RATIO = AF_ALT of Mutect2 divided by AF_ALT of Strelka2; AD_REF = Allelic depths for ref alleles (in both variant & ref call)(Mutect2); AD_ALT = Allelic depths for alt alleles (in both variant & ref call)(Mutect2). We then performed recursive feature elimination to extract a set of most informative features. The entire set of 9408 validated SNVs were used for training and 10 iteration rounds were run using a 5-fold cross-validation. Through this procedure, we identified the following set of 6 optimal features: TLOD, AF_ALT, AD_REF (in both variant and reference sample), AF_RATIO and NLOD. Next we randomly split (70/30) the dataset into a training (n = 6562) and evaluation (n = 2846) set. Using the training set, we trained a random forest classifier, again using 10 iterations with 5-fold cross-validation, resulting in average model with a classification accuracy of 92.4% (kappa = 0.64). When we applied the prediction model to our evaluation set we found a similar accuracy, namely 92.6%. Lastly we applied the prediction model on the discordant SNVs identified by both Strelka2 and Mutect2 from the WGS data obtained from the replication cohort.

2.6. DNA methylation, QC & Normalization

The following metrics were used to exclude samples: Median methylated or unmethylated intensity <1500; Median red/green intensity ratios <0.5 or >2 as calculated in type I probes; Discordance between reported sex and predicted sex based on the getSex function in the minfi R package (Aryee et al., 2014). The OP (non-polymorphic controls) and Hyb (hybridization controls) metrics as implemented in the MethylAid R package (van Iterson et al., 2014); Incomplete bisulfite conversion rate (<80%) based on the bscon metric as implemented in the wateRmelon R package (Pidsley et al., 2013); >5% of probes with detection *p*-value > 1×10^{-16} and/or >5% of probes measured by <3 beads.

After removing samples that failed on any of the steps listed above, we performed PCA on the control probes present on the array. Samples that had values larger than 3 standard deviations from the mean on the first two PCs were excluded. After removing samples that failed on any of the steps listed above, we performed PCA on the normalized β-values (described in next section), and excluded samples that had values larger than 3 standard deviations from the mean of the first two PCs. The omicsPrint R package was used to confirm the relation between twin pairs (van Iterson et al., 2018). We performed identity-by-state (IBS) using the allele-sharing function on 263 probes that reliably measured underlying common SNPs. Pairs were considered twins when the intra-pair IBS was greater than 1.9.

After quality control, signal intensities were normalized using the dasen function as implemented in the wateRmelon R package (Pidsley et al., 2013). After normalization, we set all the measurements with detection *p*-value $>1 \times 10^{-16}$ or measured by <3 beads to missing. We then removed probes with >5% missing data.

Probes were further filtered based on the following criteria: (1) a ≥14bp 3′-subsequence inexact match to the C9 repeat expansion (Hop et al., 2020); (2) a ≥30bp 3′-subsequence inexact off-target match to the reference genome; and (3) low mapping quality based as determined by Zhou et al.(Zhou et al., 2017).

2.7. Power analyses

Given the complexity of calculating power for the Wilcoxson signed-rank test, we performed power analyses for a paired *t*-test instead, as has been previously suggested (Souren et al., 2019). Power was calculated using the pwr.t.test function implemented in the pwr R package with a sample size of 20 (number of twin pairs that passed QC), and a genome-wide significance level of 2.4×10^{-7} (Saffari et al., 2018).

2.8. Differential methylation analysis

To account for the confounding effects of white blood cell composition (WBC), we regressed the β-values of each site on the estimated WBC fraction and used the residuals from this regression for subsequent analyses. At each site we tested for an association using a 2-sided Wilcoxson-signed rank test (Souren et al., 2019). Sites with $p < 2.4 \times 10^{-7}$ were considered genome-wide significant (Saffari et al., 2018).

2.9. Polymethylation scores

We imputed white blood cell fractions (CD8T cells, CD4T cells, Monocytes, Granulocytes, B-cells, and NK-cells) using the EpiDish package, where we used the 'RPC' (Robust Partial Correlations) algorithm (Teschendorff et al., 2017). Since the WBC fractions always add up to one, we dropped 1 cell-type (B-cells) in the analyses

to prevent multicollinearity among the WBC covariates. Methylation age was predicted using the multi-tissue Horvath clock and the blood-based clock developed by Zhang et al. (Horvath, 2013; Zhang et al., 2019). We calculated a smoking score as previously described in Elliot et al. and implemented in the EpiSmoker package (Bollepalli et al., 2019; Elliott et al., 2014).

3. Results

3.1. Study population

We obtained DNA from 21 monozygotic twins, of which only one was affected by ALS, with at least a 5-year period between onset of the disease and inclusion in this study. Twin characteristics are listed in Supplementary Table 1a. Notably, one of the subjects that was not affected by ALS had a medical history of chronic lymphatic leukemia (CLL). This twin was intentionally not excluded, as whole blood DNA from a CLL patient could serve as a positive control for increased post-zygotic mutations. Monozygosity was confirmed in all twins using the WGS data. None of the reported unaffected twin siblings developed ALS during the additional 5-year course of this study. All twins had a negative family history for motor neuron disease. One twin pair reported a distant family member with unspecified muscle weakness and another twin pair had a first degree relative with FTD. We screened all twins variants in known ALS genes. In the twin pair, with a positive family history for FTD, both twins carried an intronic hexanucleotide repeat expansion in *C9orf72*. Unfortunately, attempts to determine possible repeat expansion size differences with Southern blot analysis failed, probably due to insufficient amount of high quality DNA. In another pair, both twins carried a missense mutation in *TARDBP* (1:11082610:G>A / p.Ala382Thr), previously reported in several familial ALS patients (Chio et al, 2010). Three other twin pairs had rare amino-acid changing variants of unknown clinical significance in known ALS genes (*NEK1, TBK1* and *UNC13A*) (Supplementary Table 1).

3.2. Detection and validation of early post-zygotic mutations

WGS has many advantages over WES when it comes to variant detection, such as its breadth of coverage (even in exonic regions) and its potential to additionally detect non-exonic and various structural variants. However, de novo mutation identification from WGS twin or trio data relies on complex computational prediction and filtering, mostly due to limited validation capacity (Francioli et al., 2015). We analyzed our ~30x coverage Illumina HiSeq2000 WGS twin datasets using both the Genome Analysis Toolkit Haplotype Caller (GATK-HC) as well as Isaac Variant Caller and added a somatic variant caller, Strelka2, considering the ALS affected twin a somatic variant of the healthy one and vice versa (allowing for potentially protective mutations as well). The discordant single nucleotide variants (SNVs) that were identified in the WGS data were subsequently validated using a different technique (custom Axiom arrays) which resulted in 1,124 single nucleotide de novo mutations. A total 95% (n = 1,067) of these validated SNVs were detected using a somatic variant calling method. whereas application of the other variant calling methods resulted in the additional identification of 57 potential de novo mutations.

Next, we added a second validation method using targeted enrichment and sequencing, increasing coverage to at least ~70–80x for regions of interest, in order to confirm the identified variants and evaluate variants that failed custom genotyping array design, such as small base pair insertions and deletions (Indels).

This 2-step validation resulted in the identification of 998 definite post-zygotic single nucleotide mutations (pSNM) (validated by both methods) and 181 possible pSNM (validated with a single method). Using targeted sequencing only, we validated 91 small deletions and 39 small insertions. There were another 175 SNVs that had a discordant genotype using targeted sequencing, but not on the genotyping array, possibly due to sequencing artifacts or a low allelic fraction not picked up using an array due to late post-zygotic events; these were excluded from further analysis. Thus, through extensive 2-tier validation, we uncovered a high number of post-zygotic mutations for further analysis.

3.3. A skewed distribution of post-zygotic mutations between twins

As expected, the majority of the 1179 uncovered definite or possible pSNMs were detected in the control twin that suffered from chronic leukemia (1-CON: 868 mutations). After its exclusion, we observed a non-significant difference in the average of 11.6 (SD: 26.8) post-zygotic mutations in ALS twins compared to 3.3 (SD: 3.97) in non-affected twins. However, the number of mutations was significantly unevenly distributed (K-S test: $p < 0.001$) with 2 ALS affected outliers (2-ALS: 111 mutations; 8-ALS: 66 mutations) (Fig. 1A). Importantly, their unaffected twin sibling did not share the same vulnerability (2-CON: 0 mutations; 8-CON: 2 mutations). This same pattern was identified in the distribution of Indels (Fig. 1B).

3.4. Post-zygotic mutations can occur at early stage of development

To identify early pSNMs we additionally analyzed WGS data obtained from fibroblast (available in Twin 1, 2, 6, 7, 8, 10, and 11) and saliva (available in Twin 6, 8, 10, and 11). For Twin 2, none of the identified mutations could be additionally identified in DNA obtained from fibroblasts, which could indicate either late pSNMs. In contrast, 95% of the possible and definite pSNMs in Twin 8 were identified in DNA obtained from saliva (63/66), 6 of which were additionally detected in fibroblast. In all twins combined, there were ten pSNMs detected in all tissues (ALS: 5, CON: 5). Overall, the allelic fraction (AF) was not significantly different between saliva and whole blood, which could be explained by the presence of leukocytes in saliva (Fig. 2A and B). We did observe significantly higher AFs in the pSNMs picked up in all tissues compared to blood and saliva only (0.48 vs. 0.22, $p = 0.004$), close to the germline AF of 0.5 (Fig. 2C) confirming that these detected mutations occurred at an early stage during embryonic development and potentially affect the nervous system.

3.5. Mapping of identified post-zygotic mutations

After exclusion of the twin with CLL, we identified only one confident pSNM with possible amino-acid changing consequences; a stop-gain mutation in an ALS twin (8) in the exonic region of *LMX1A* (p.R208*). This variant does not occur in public exome/WGS databases such as gnomAD (http://gnomad.broadinstitute.org) or the online databrowser of Project MinE (http://databrowser.projectmine.com), containing WGS data of 4366 ALS cases and 1832 controls (van der Spek et al, 2019). A gene-based burden test on the 11 coding variants identified in *LMX1A* in the Project MinE database does not show a significant association with ALS ($p = 0.797$; Firth logistic regression) and given a low AF of 0.145 in whole blood, its' clinical relevance remains unknown. Second, we evaluated the genomic distribution of post-zygotic mutations identified in ALS or control twins, which did not show a clear clustering, except for a small locus on chromosome 8 (S1 Fig); this clustering however, appears to be more twin specific rather than phenotype specific as all three mutations can be

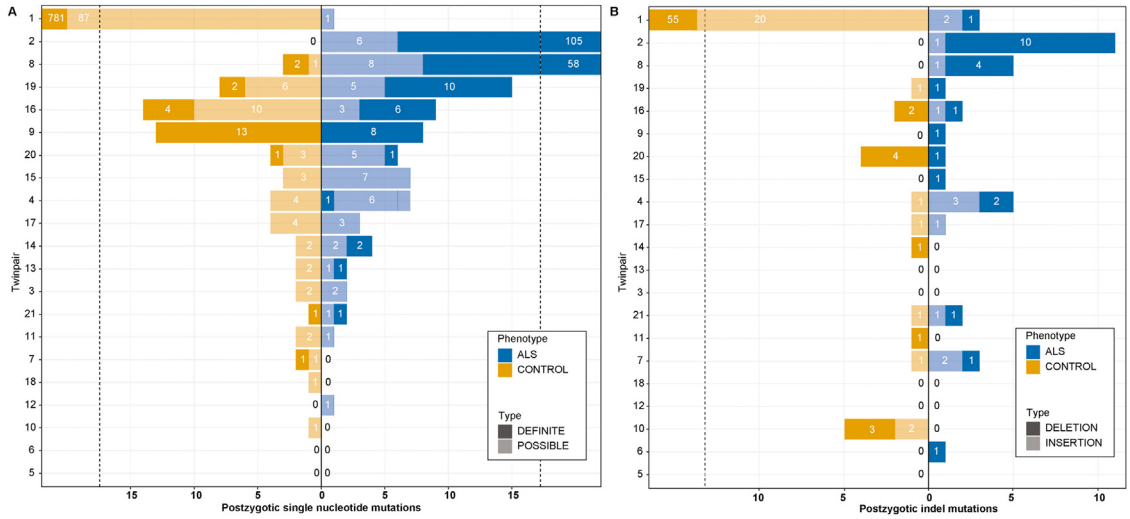

Fig. 1. Distribution of post-zygotic mutations amongst ALS discordant monozygotic twin pairs. (A) Number of identified definite (dark shade) and possible (light shade) post-zygotic single nucleotide mutations per twin per phenotype (blue: ALS affected twin; orange: unaffected control twin) with three high number outliers: 1 leukemia affected control twin (Twin 1) and 2 ALS affected twins (Twin 2 and 8). (B) A similar distribution can be identified in the possible post-zygotic small base pair deletions (dark shade) and insertions (light shade).

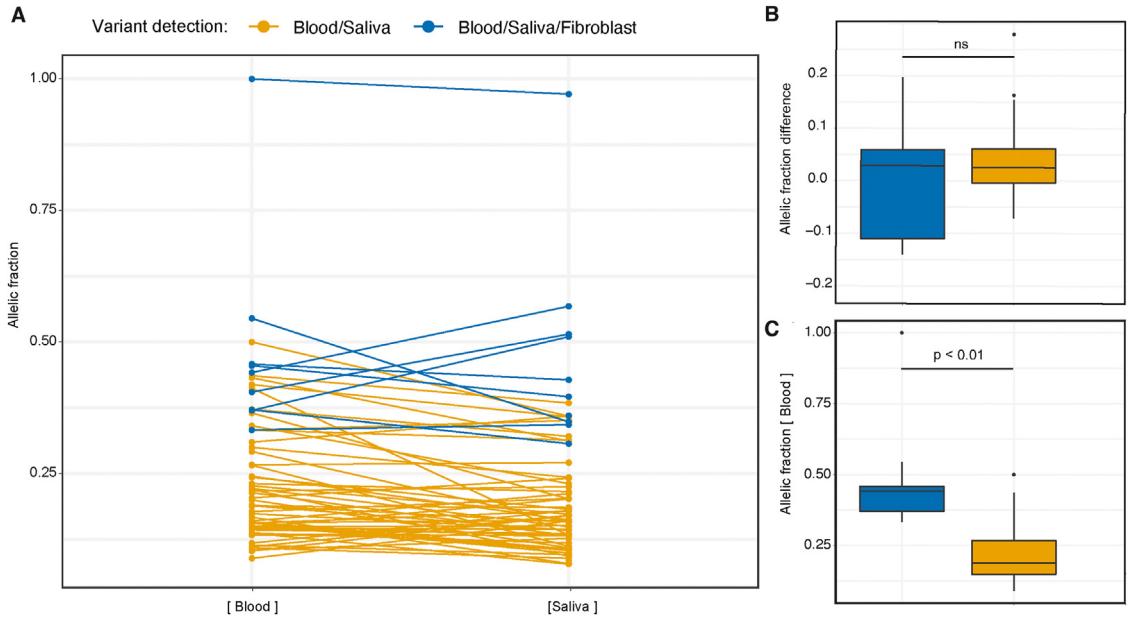

Fig. 2. Detection of post-zygotic mutation in DNA obtained from multiple tissue samples. (A) Per mutation comparison of detected allelic fraction of post-zygotic mutations that were identified in WGS data obtained from whole-blood and saliva (orange). Variants that were additionally detected in DNA obtained from fibroblast are shown in blue. (B) Boxplot of the difference in allelic fraction for the same variant between blood and saliva (defined as AF blood – AF saliva). c Boxplot of the allelic fraction of the variant allele in whole blood WGS data shows higher allelic fractions in variants discovered in all tissues compared to two-tissue only. One variant has an allelic fraction of 1 given a post-zygotic mutation in a male on the X-chromosome outside the pseudo-autosomal region.

attributed to a single twin. Lastly, we analyzed whether gene promotor or enhancer regions could be affected using the GeneHancer database (Fishilevich et al., 2017) (Supplementary Table 2). Interestingly, one ALS twin we identified a mutation within a predicted regulatory element for *UGT8*, of which the promotor region has been previously linked to ALS in a CNV trio-study (Pamphlett and Morahan, 2011). This post-zygotic mutation was picked up in both blood as well as saliva. Another post-zygotic mutation in an ALS twin with a high mutation rate is located in the regulatory region of *TDP2*, a DNA repair gene implicated in neurodegenerative diseases such as spinocerebellar ataxia (Errichiello et al., 2020; Zagnoli-Vieira et al., 2018).

3.6. Mutational signature analysis implies accelerated aging

To evaluate if the increased number of post-zygotic single nucleotide mutations in the ALS samples could be caused by an underlying problem in DNA-repair, we performed a mutational signature analysis. We compared the characteristics of the post-zygotic mutations identified in the 2 ALS hypermutation outliers

with both the mutations identified in the twin affected by leukemia, which served as a positive control, and those identified in all the other twins combined, and additionally with a list of validated de novo mutations publicly available from the Genome of the Netherlands (GoNL) project (Francioli et al., 2015; Genome of the Netherlands, 2014). We identified 3 different signatures distinguishing the de novo mutations identified in leukemia and those in GoNL, but found no signature distinguishing the ALS outlier and all other mutations (S2 Fig). Next, we calculated the contribution of known COSMIC mutational signatures (S3 Fig) (Tate et al., 2019). As expected, in the mutation profile of the twin with chronic lymphatic leukemia there was a relatively high contribution (35%) of COSMIC signature 9, which is derived from a dataset with chronic lymphocytic leukemias and malignant B-cell lymphomas(Alexandrov and Stratton, 2014). In the GoNL de novo set the largest contribution is signature 5 (31%), whereas in both the ALS outliers and all other twin identified mutations this was signature 1 (resp. 29% and 34%). Signature 1 has been linked to age at cancer diagnosis and is most likely the result of somatic mutational processes, whereas the GoNL combination with signature 5, fits the pattern of de novo mutations found in the human germline (Alexandrov et al., 2015). Thus, the increased number of mutations in ALS seems to be the result of mutational processes that occur during aging, despite the fact that these twins were the same age at sampling as their twin siblings and were not significantly older than the other monozygotic twins. In support, we found a significant transcription bias of C>A / G>T transversions in the ALS twins with an increased mutation rate (p = 0.016, 2-sided Poisson test; S4 Fig), indicative of DNA damage caused by oxidative stress and linked to accelerated aging (Shibutani et al., 1991; Zhang et al., 2017).

3.7. Increased mutation rate replicates in second ALS-discordant monozygotic twin set

During the study, an additional 5 ALS-discordant monozygotic twins that met the inclusion criteria were collected, resulting in a smaller replication cohort . Samples were analyzed using WGS, but were not taken into account for the custom genotyping array and targeted sequencing design. To determine likely de novo SNVs from the WGS data, we first build a prediction model by applying a random forest machine-learning algorithm (trained on the validated 1052 true positives and 8186 false positive post-zygotic single nucleotide mutations with complete data by both somatic variant callers) with an estimated accuracy of 93.1% and a positive predictive value of 72.4%. We then analyzed the WGS data of the replication twins with Strelka2 and Mutect2, and applied the prediction model to 1870 discordant SNVs identified with both calling algorithms. This resulted in a total of 300 possible post-zygotic mutations (Fig. 3). Again, we found an ALS-affected twin to have a high number of mutations in comparison to the other samples. WGS analysis of DNA obtained from saliva from the same twin pair confirmed 202 out of 233 post-zygotic mutations (ALS: 186, Control: 16). Functional annotation of identified mutations revealed three missense mutations (Supplementary Table 3). Two pSNMs were identified in ALS twins in the *OR52J3* and *OR4N2* genes respectively and are predicted to be deleterious by SIFT, although one is predicted to be benign by Polyphen. Interestingly, both genes are involved in olfactory transduction and this pathway showed a significant association in a KEGG analysis (p = 0.006). One pSNM was identified in an unaffected twin in the *PBX4* gene and was predicted to be benign by both SIFT and Polyphen. However, due to the unique status and that there is no clear increased gene-burden in the Project MinE ALS databrowser, the clinical relevance of these variants remains uncertain (van der Spek et al., 2019).

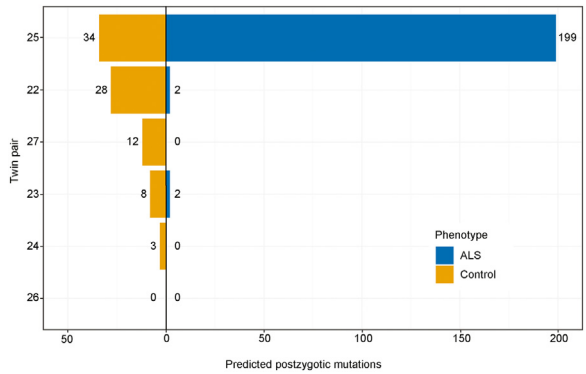

Fig. 3. Distribution of predicted post-zygotic mutations in a replication set of ALS discordant monozygotic twin pairs. Number of predicted post-zygotic single nucleotide mutations from WGS data per twin pair and per phenotype (blue: ALS affected twin; orange: unaffected control twin) in a second set of disease discordant monozygotic twins.

3.8. Analysis of CAG trinucleotide repeats shows overall repeat stability

DNA repeat expansions are an important part of the genetic architecture of ALS.; Intronic hexanucleotide repeats in *C9orf72* are is the most frequent genetic cause of ALS. Additionally, CAA trinucleotide expansions in *NIPA1* and CAG expansions in *ATXN1* and *ATXN2* have been associated with the disease (Blauw et al., 2012; DeJesus-Hernandez et al., 2011; Elden et al., 2010; Tazelaar et al., 2020; Tazelaar et al., 2019). Repetitive elements comprise a large part of the human genome, but despite recent advances, accurate repeat-length detection from WGS-data remains challenging (de Koning et al., 2011; Dolzhenko et al., 2020; Dolzhenko et al., 2019). Therefore, in addition to the hexanucleotide expansion in *C9orf72,* we used PCR to screen for a group of seven CAG trinucleotide repeats, given their role in neurodegenerative disease with pyramidal symptoms: *ATXN1, ATXN2, ATXN3, CACNA1A, ATXN7, PPP2R2B and TBP* (Holmes et al., 2001; Paulson et al., 2017). Overall, we found similar repeat sizes for all determined CAG repeats in 24 out of 26 monozygotic twins (Supplementary Table 4). In one twin pair however, we found five out of seven repeats to be discordant on one or both alleles. Although CAG repeat length instability is common in expanded repeats, this observed high variability of normal length alleles is rare and could indicate defects in DNA repair.

3.9. Differential DNA methylation analysis

Since phenotypic discordance in monozygotic twins has recently been (partially) attributed to epigenetic differences in neurological diseases such as multiple sclerosis and ALS (Souren et al., 2019; Tarr et al., 2019; Young et al., 2017), we performed a DNA methylation analysis using the Infinium HumanMethylation450 DNA methylation array (450K) on whole blood DNA obtained from 22 ALS-discordant (20 from the original set, excluding the twin with CLL, plus two from replication set). After quality control, probe filtering and adjustment for predicted white blood cell proportions, we performed a pair-wise analysis using a Wilcoxon singed-rank test on the mean within-pair β-value differences ($\Delta\beta$) to detect differentially methylated CpG positions (DMPs). We applied a stringent p-value threshold of 2.4×10^{-7} and a suggestive threshold of 5.0×10^{-6} (Saffari et al., 2018; Souren et al., 2019; Tsai and Bell, 2015). We found no DMP reaching genome wide significance while three DMPs reached a suggestive p-value: *LBX1* (cg13112154, mean $\Delta\beta$-value = -0.019), *SND1* (cg11857142,

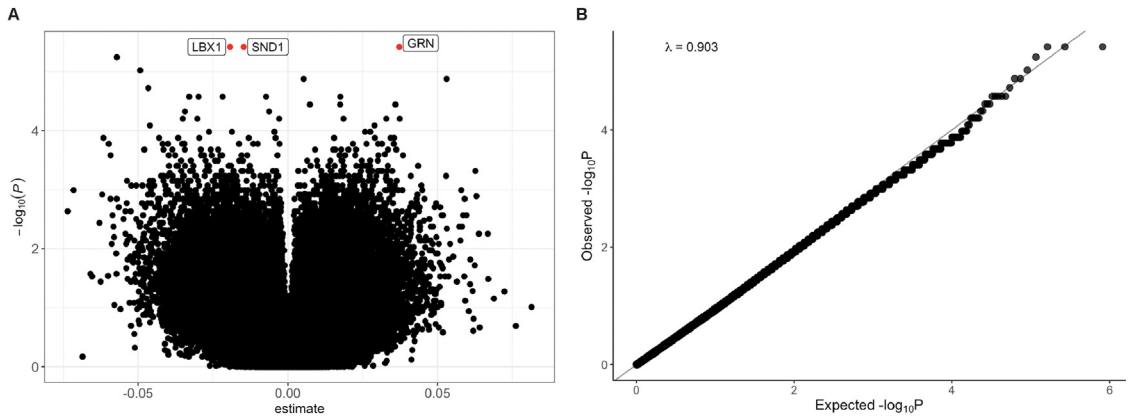

Fig. 4. Differential DNA methylation analysis on 22 ALS-discordant monozygotic twin pairs. (A) Volcano plot of the p-values resulting from the nonparametric two-tailed Wilcoxon signed-rank test against the mean within-pair β-value difference after adjustment for cell-type composition for each CpG. Red dots indicate CpGs with a *p*-value $<5 \times 10^{-6}$. (B) Q-Q plot of the p-values resulting from the nonparametric two-tailed Wilcoxon signed-rank test shown in (A).

mean $\Delta\beta$-value = -0.015) and *GRN* (cg16500497, mean $\Delta\beta$-value = 0.037) (Fig. 4 and S5 Fig). We found no significant differences in methylation in known ALS genes (S6 Fig).

DNA methylation patterns can also be used to predict certain biomarkers such as smoking, body mass index (BMI) as well as aging. The difference between the epigenetic age and true chronological age serves as an indicator for age acceleration. Previous monozygotic twin studies in ALS have shown increased predicted age for ALS-affected twins using the Horvath age prediction model (Tarr et al., 2019; Young et al., 2017). We used the recently published predictor from Zhang et al. and compared this to other models, including Horvath and Hannum predictors (Hannum et al., 2013; Horvath, 2013; Zhang et al., 2019). Since precision can decrease with increased chronological age, we only used the twin pairs which were sampled at the same age. Overall, we found the highest correlation using the Zhang model (r = 0.95) (Fig. 5A) and found no evidence of age acceleration for ALS-affected twins in comparison with their non-affected siblings (6 out of 14 pairs, Fig. 5B). We did find more ALS dependent acceleration using Hannum and Horvath prediction models (with resp. 11 and 9 out of 14 pairs, S7 Fig). However, all three models show inconsistency when it comes to which twin shows acceleration, which indicates a lack of precision to predict subtle differences in a relatively small dataset. Using other prediction models, we also did not find ALS-dependent differences in other predicted biomarkers: BMI, smoking, HDL-cholesterol, CRP and alcohol use (S8 Fig).

4. Discussion

Multiple studies have demonstrated a linear relationship between the log incidence and log age of onset of ALS, which is consistent with a multistep model of disease (Al-Chalabi et al., 2014). The model assumes there is a genetic predisposition and that subsequent downstream events (environmental exposures, (epi)genetic changes) trigger the disease. Data from these studies suggest that 6 steps are required to trigger ALS (Al-Chalabi et al., 2014; Vucic et al., 2020). Interestingly, it has been demonstrated that in familial ALS, multiple steps are also required (albeit fewer), suggesting that the presumed pathogenic mutation itself is not solely responsible for the disease (Chio et al., 2018). This is a possible explanation for the frequent observation of non-penetrance and high phenotypic variability in familial ALS pedigrees as well as the identification of pathogenic mutations in apparently sporadic cases and is also compatible with the findings of this study, in which we identified (potentially) pathogenic mutations (*C9orf72* repeat expansions, *TARDBP, NEK1, TBK1*, and *unc13a*) within twin pairs of which only one developed disease. In line with the multistep model, this implies that additional event(s) have taken place in one, but not the other twin. Discordant monozygotic twins both carrying *C9orf72* repeat expansions have been reported previously (PMID: 25209579, PMID: 31164693) and even a triplet carrying a p.I114T *SOD1* mutation, of which only was affected (Tarr et al., 2019). Similarly, there are multiple reports of familial ALS pedigrees with pathogenic mutations in ≥2 ALS genes and sporadic cases carrying pathogenic mutations in combination with ALS risk factors (e.g. both C9orf72 and ATXN2 repeat expansions) (McCann et al., 2020; van Blitterswijk et al., 2012). Moreover, GWAS results also show that the genetic risk for ALS is driven by variants with a low frequency (1%–10%)(van Rheenen et al., 2021). Therefore, ALS is considered to be an oligogenic disease.

Given the oligogenic and multistep model of disease, the aim of this study was to identify novel (epi)genetic risk factors that might increase risk for ALS by identifying variants present in one twin but not the other. By performing in-depth sequencing and extensive validation we identified post-zygotic mutations of high confidence leading to 3 exonic candidates in the *LMX1A, OR52J3* and *OR4N2I* genes. Burden testing in the larger Project MinE cohort did not provide further evidence for association with ALS. It must be noted however, that the variant in *LMX1A* is a stop-gain mutation whereas burden testing was performed using non-synonymous variants. Loss-of-function mutations were not identified in *LMX1A* in the Project MinE database and gnomAD contains only 1 LoF mutation.

We also considered variants outside of coding genome, which yielded 8 post-zygotic mutations in predicted transcriptional binding sites of gene enhancers. We identified a mutation in blood as well as saliva within a predicted regulatory element for *UGT8*. An unbiased genome-wide screen for de novo DNA mutations in ALS trios previously identified rare and potentially pathogenic CNVs in the promotor region of *UGT8* (Pamphlett and Morahan, 2011). UGT8 plays a role in the biosynthesis of galactocerebroside, which is a sphingolipid that forms myelin membranes in both the central and peripheral nervous system. Transgenic mice that lack the *UGT8* orthologue have a motor phenotype with disrupted nerve conduction and degeneration of myelin (Bosio et al., 1996; Coetzee et al., 1996). UGT8 is also a molecular marker for breast cancer and lung metastases (Dziegiel et al., 2010). One of the ALS twins with a high mutation rate carried a post-zygotic mutation in the regulatory region of *TDP2*, which is a DNA repair gene. We cannot comment on whether the high mutation rate is a direct consequence

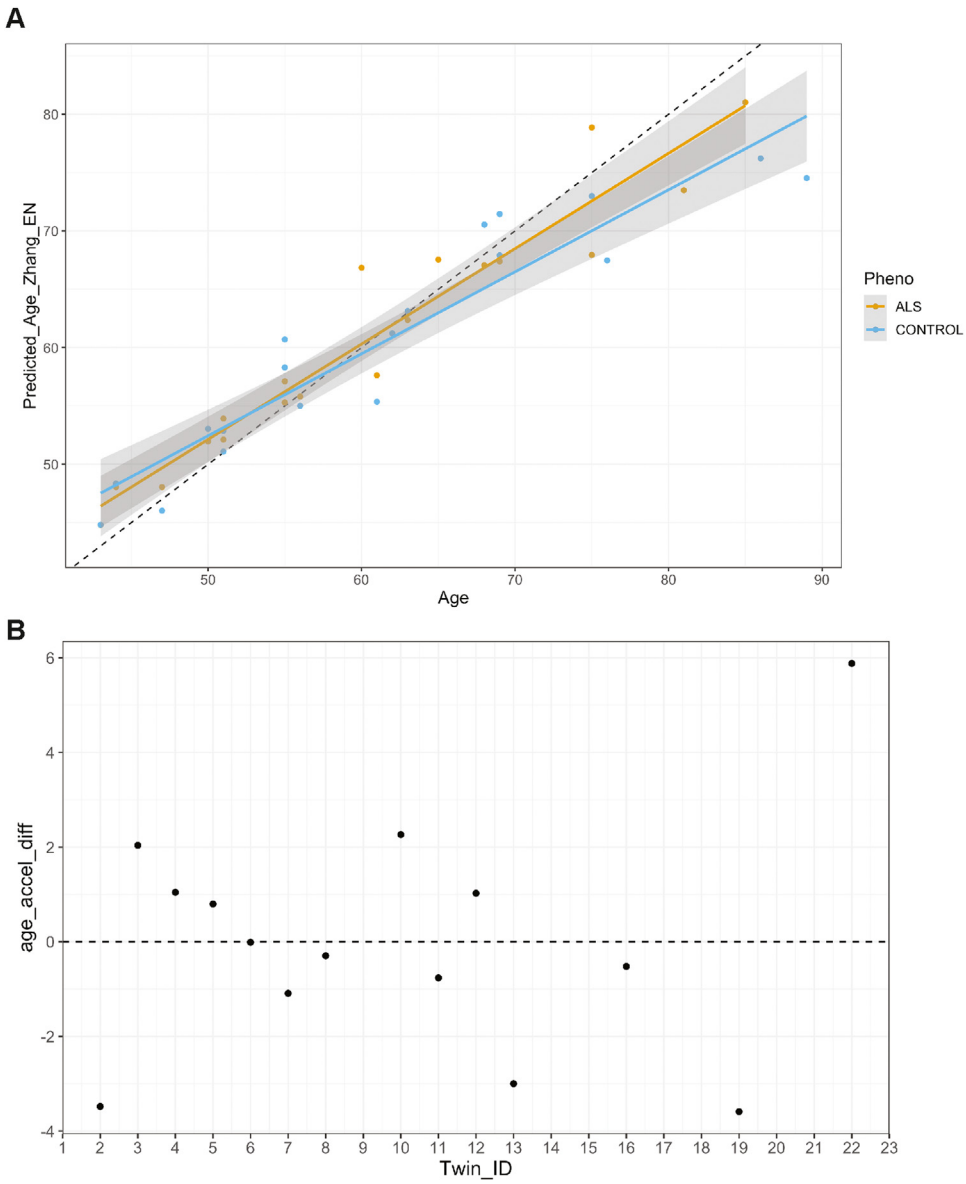

Fig. 5. Individual age prediction using methylation data. (A) Correlation between biological age and epigenetic age for each subject using Zhang prediction model shows high correlation ($r = 0.95$) for both phenotypes (orange: ALS, blue: control). (B) Plot of methylome age differences for each twin pair based on the difference in epigenetic age of ALS sibling minus that of the unaffected sibling (only for twins which were same age at sampling).

of impaired TDP2 function. Interestingly, homozygous mutations in TDP2 cause autosomal recessive spinocerebellar ataxia type 2, has been functionally implicated in other neurodegenerative diseases such as Parkinson's disease and is also an oncogenic factor (Li et al., 2011; Zoghi et al., 2021).

Our methylation analysis identified three candidate genes (*SND1, LBX1,* and *GRN*), all of which have been well-established to play a role in multiple forms of cancer. Two of these can be directly linked to ALS: *SND1* was enriched for de novo mutations in a previous ALS trio-study (Steinberg et al., 2015), whereas *GRN* (Granulin Precursor) has been extensively linked to FTD and *GRN* mutations have been identified in ALS-FTD patients (Cannon et al., 2013).

Another important finding is the skewed distribution of the mutation rate amongst twins with a very high number in a subset of the affected siblings. This finding replicated in the second independent twin cohort and was present in multiple tissues, ruling out sample contamination or environmental mutagenic exposure. In support, prediction of environmental risk factors, such as smoking, from methylation profiles did not indicate clear epigenetic differences amongst siblings. A possible explanation for the observed relative hypermutation in ALS could be accelerated aging, which is backed-up by previous studies investigating epigenetic differences in ALS discordant monozygotic twins (Tarr et al., 2019; Young et al., 2017). However, we were unable to replicate these findings as age prediction between twins from methylation data seems to be method-dependent and probably lacks precision for detecting small epigenetic age differences in a pair-wise analysis, especially at old(er) age. We did however, find an ALS-hypermutation dependent transcriptional bias in C>A / G>T transversions, which can be attributed to DNA damage caused by oxidative stress. Interestingly, high levels of DNA oxidation have been identified in mice lacking Sod1, the first identified ALS gene, which also suffer from an accelerated aging phenotype (Zhang et al., 2017). Differences in aging processes might additionally explain the observed CAG repeat length differences in one twin pair, as aging and oxidative dam-

age are known to influence the somatic stability of CAG repeats (Kovtun et al., 2007; Sanchez-Contreras and Cardozo-Pelaez, 2017).

This study has several limitations. Firstly, the sample size is small, which means that the statistical power to detect associations was low. The underlying assumption for this study was however that the effect of discordant variants would be large. Although this indeed may be the case, the ultra-rare nature of the identified variants makes replication challenging and thus complicates interpreting their relevance.

The other major limitation of this study was that we did not have access to tissue from the nervous system. We attempted to (partially) overcome this by additionally analyzing DNA isolated from dermal fibroblast and/or saliva in some twin pairs. Taken into consideration that the embryonic origin of dermal fibroblasts depends on site of biopsy, but often taken from ventral / limb which derives from lateral plate mesoderm (Driskell and Watt, 2015), and saliva consists of a mixed population of leukocytes of mesodermal origin and epithelial cells of ectodermal origin (Theda et al., 2018), we were able to confirm that several mutations were present in multiple tissues and indeed had higher allelic fractions, indicative of very early post-zygotic events. This is an important finding, as it demonstrates that, at least in some twins, genetic differences already occur at an early stage and are therefore more likely to affect the central nervous system. Vulnerability to somatic mutations might also explain why were unable to identify obvious pathogenic variants in whole blood DNA, as somatic mutations could also have occurred in the nervous system itself and were therefore not picked up in this study.

If indeed ultra-rare and randomly induced mutations are one of the factors that trigger ALS, this will pose a huge challenge to the field. Even more so, if these triggering events only take place in certain tissues or cells. Studies using readily accessible tissues (such as blood, fibroblasts or muscle) would not be able to detect these changes and they could even be missed in autopsy material from brain or spinal cord due to sampling. However, a somatic mutation in the nervous system that triggers disease would fit well the frequently very focal onset of the disease.

In conclusion, our findings support the multi-step hypothesis for ALS, which is not limited to predisposing genetic factors and subsequent environmental factors, but could also be the result of ongoing (epi) genetic changes. Intriguingly, our analyses identified several genes that have been implicated in ALS and neurodegeneration previously, including *GRN, SND1, TDP2* and *UGT8*. Strikingly, these genes all play a role in multiple forms of cancer, which also the result of a multi-step process. The skewed distribution of the mutation rate amongst twins with a very high number in a subset of affected siblings as well as CAG repeat the length differences in 1 twin pair are in line with these findings.

5. Conclusions

In summary, this study implicates several mechanisms in ALS pathophysiology, which include a role for de novo mutations, defects in DNA damage repair and accelerated aging. Although it is difficult to pinpoint a single post-zygotic downstream events responsible for disease discordancy in monozygotic twins, the variety in discordant factors support a multistep model of disease that trigger the onset of ALS.

Verification

The data contained in the manuscript have not been previously published and have not been and will not be submitted elsewhere while under consideration at Neurobiology of Aging.

Disclosure statement

We have the following potential conflicts of interest to disclose: L.H. van den Berg serves on scientific advisory boards for the Prinses Beatrix Spierfonds, Thierry Latran Foundation, Biogen and Cytokinetics; and serves on the editorial board of Amyotrophic Lateral Sclerosis And Frontotemporal Degeneration and The Journal of Neurology, Neurosurgery, and Psychiatry. O. Hardiman has received speaking honoraria from Novarits, Biogen Idec, Sanofi Aventis and Merck-Serono, has been a member of advisory panels for Biogen Idec, Allergen, Ono Pharmaceuticals, Novartis, Cytokinetics and Sanofi Aventis and serves as Editor-in-Chief of Amyotrophic Lateral Sclerosis and Frontotemporal Dementia. A. Al-Chalabi has consulted for OrionPharma, Biogen Idec, Cytokinetics Inc, Treeway Inc, and Chronos Therapeutics. J.H. Veldink reports that his institute received consultancy fees from Vertex Pharmaceuticals outside the submitted work. M.A. van Es has received travel grants from Baxalta and serves on the biomedical research advisory panel of the motor neurone disease association (MNDA). Other authors declare no actual or potential conflicts of interest.

Acknowledgements

This study was funded by the Thierry Latran Foundation and the Dutch ALS foundation. Project MinE Belgium was supported by a grant from IWT (n° 140935), the ALS Liga België, the National Lottery of Belgium and the KU Leuven Opening the Future Fund. M.A.v.E. is additionally supported by the Rudolf Magnus Brain Center Talent Fellowship. P.V.D. holds a senior clinical investigatorship of FWO-Vlaanderen and is supported by the E. von Behring Chair for Neuromuscular and Neurodegenerative Disorders, the ALS Liga België and the KU Leuven funds “Een Hart voor ALS,” “Laeversfonds voor ALS Onderzoek” and the “Valéry Perrier Race against ALS Fund.” Several authors of this publication are member of the European Reference Network for Rare Neuromuscular Diseases (ERN-NMD).

A.A.-C. receives salary support from the National Institute for Health Research (NIHR) Dementia Biomedical Research Unit and Biomedical Research Centre in Mental Health at South London and Maudsley NHS Foundation Trust and King’s College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. O.H. is funded by the Health Research Board Clinician Scientist Programme and Science Foundation Ireland. R.L.M is also supported by the Thierry Latran Foundation (ALSIBD) and the ALS Association (2284). The Swedish Brain Foundation (grants nr. 2012-0262, 2012-0305, 2013-0279, 2016-0303), the Swedish Science Council (grants nr 2012-3167, 2017-03100), the Knut and Alice Wallenberg Foundation (grants nr. 2012.0091, 2014.0305), the Bertil Hållsten Foundation, the Ulla-Carin Lindquist Foundation, the Neuroförbundet Association, the Torsten and Ragnar Söderberg Foundation, Umeå University Insamlingsstiftelsen (223-2808-12, 223-1881-13, 2.1.12-1605-14), Västerbotten County Council, Swedish Brain Power, King Gustaf V:s and Queen Victoria’s Freemason’s Foundation.

CRediT authorship contribution statement

Gijs H.P. Tazelaar: Methodology, Validation, Investigation, Writing – original draft, Writing – review & editing. **Paul J. Hop:** Methodology, Validation, Investigation, Data curation, Writing – original draft. **Meinie Seelen:** Methodology, Validation, Investigation. **Joke J.F.A. van Vugt:** Investigation, Data curation, Writing – original draft. **Wouter van Rheenen:** Investigation, Data curation, Writing – original draft. **Lindy Kool:** Investigation, Data curation.

Kristel R. van Eijk: Investigation, Data curation, Writing – original draft. **Marleen Gijzen:** Investigation, Validation. **Dennis Dooijes:** Investigation, Validation. **Matthieu Moisse:** Methodology, Investigation, Data curation. **Andrea Calvo:** Resources, Writing – review & editing. **Cristina Moglia:** Resources, Writing – review & editing. **Maura Brunetti:** Resources, Writing – review & editing. **Antonio Canosa:** Resources, Writing – review & editing. **Angelica Nordin:** Resources, Writing – review & editing. **Jesus S. Mora Pardina:** Resources, Writing – review & editing. **John Ravits:** Resources, Writing – review & editing. **Ammar Al-Chalabi:** Resources, Writing – review & editing. **Adriano Chio:** Resources, Writing – review & editing. **Russell L. McLaughlin:** Resources, Writing – review & editing. **Orla Hardiman:** Resources, Writing – review & editing. **Philip Van Damme:** Resources, Writing – review & editing. **Mamede de Carvalho:** Resources, Writing – review & editing. **Christoph Neuwirth:** Resources, Writing – review & editing. **Markus Weber:** Resources, Writing – review & editing. **Peter M Andersen:** Resources, Writing – review & editing. **Leonard H. van den Berg:** Resources, Writing – review & editing, Supervision, Funding acquisition. **Jan H. Veldink:** Conceptualization, Methodology, Data curation, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition. **Michael A. van Es:** Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neurobiolaging.2022.11.010.

References

Acuna-Hidalgo, R., Bo, T., Kwint, M.P., van de Vorst, M., Pinelli, M., Veltman, J.A., Hoischen, A., Vissers, L.E., Gilissen, C., 2015. Post-zygotic point mutations are an underrecognized source of de novo genomic variation. Am J Hum Genet 97 (1), 67–74.

Al-Chalabi, A., Calvo, A., Chio, A., Colville, S., Ellis, C.M., Hardiman, O., Heverin, M., Howard, R.S., Huisman, M.H.B., Keren, N., Leigh, P.N., Mazzini, L., Mora, G., Orrell, R.W., Rooney, J., Scott, K.M., Scotton, W.J., Seelen, M., Shaw, C.E., Sidle, K.S., Swingler, R., Tsuda, M., Veldink, J.H., Visser, A.E., van den Berg, L.H., Pearce, N., 2014. Analysis of amyotrophic lateral sclerosis as a multistep process: a population-based modelling study. Lancet Neurol 13 (11), 1108–1113.

Al-Chalabi, A., Fang, F., Hanby, M.F., Leigh, P.N., Shaw, C.E., Ye, W., Rijsdijk, F., 2010. An estimate of amyotrophic lateral sclerosis heritability using twin data. J Neurol Neurosurg Psychiatry 81 (12), 1324–1326.

Alexandrov, L.B., Jones, P.H., Wedge, D.C., Sale, J.E., Campbell, P.J., Nik-Zainal, S., Stratton, M.R., 2015. Clock-like mutational processes in human somatic cells. Nat Genet 47 (12), 1402–1407.

Alexandrov, L.B., Stratton, M.R., 2014. Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. Curr Opin Genet Dev 24, 52–60.

Andersen, P.M., Al-Chalabi, A., 2011. Clinical genetics of amyotrophic lateral sclerosis: what do we really know? Nat Rev Neurol 7 (11), 603–615.

Aryee, M.J., Jaffe, A.E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A.P., Hansen, K.D., Irizarry, R.A., 2014. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 30 (10), 1363–1369.

Blauw, H.M., van Rheenen, W., Koppers, M., Van Damme, P., Waibel, S., Lemmens, R., van Vught, P.W., Meyer, T., Schulte, C., Gasser, T., Cuppen, E., Pasterkamp, R.J., Robberecht, W., Ludolph, A.C., Veldink, J.H., van den Berg, L.H., 2012. NIPA1 polyalanine repeat expansions are associated with amyotrophic lateral sclerosis. Hum Mol Genet 21 (11), 2497–2502.

Blokzijl, F., Janssen, R., van Boxtel, R., Cuppen, E., 2018. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. Genome Med 10 (1), 33.

Bollepalli, S., Korhonen, T., Kaprio, J., Anders, S., Ollikainen, M., 2019. EpiSmokEr: a robust classifier to determine smoking status from DNA methylation data. Epigenomics 11 (13), 1469–1486.

Bosio, A., Binczek, E., Stoffel, W., 1996. Functional breakdown of the lipid bilayer of the myelin membrane in central and peripheral nervous system by disrupted galactocerebroside synthesis. Proc Natl Acad Sci U S A 93 (23), 13280–13285.

Brown, R.H., Al-Chalabi, A., 2017. Amyotrophic Lateral Sclerosis. N Engl J Med 377 (2), 162–172.

Cannon, A., Fujioka, S., Rutherford, N.J., Ferman, T.J., Broderick, D.F., Boylan, K.B., Graff-Radford, N.R., Uitti, R.J., Rademakers, R., Wszolek, Z.K., Dickson, D.W., 2013. Clinicopathologic variability of the GRN A9D mutation, including amyotrophic lateral sclerosis. Neurology 80 (19), 1771–1777.

Chesi, A., Staahl, B.T., Jovicic, A., Couthouis, J., Fasolino, M., Raphael, A.R., Yamazaki, T., Elias, L., Polak, M., Kelly, C., Williams, K.L., Fifita, J.A., Maragakis, N.J., Nicholson, G.A., King, O.D., Reed, R., Crabtree, G.R., Blair, I.P., Glass, J.D., Gitler, A.D., 2013. Exome sequencing to identify de novo mutations in sporadic ALS trios. Nat Neurosci 16 (7), 851–855.

Chio, A., Calvo, A., Moglia, C., Ossola, I., Brunetti, M., Sbaiz, L., Lai, S.L., Abramzon, Y., Traynor, B.J., Restagno, G., 2011. A de novo missense mutation of the FUS gene in a "true" sporadic ALS case. Neurobiol Aging 32 (3), 553 e523-556.

Chio, A., Mazzini, L., D'Alfonso, S., Corrado, L., Canosa, A., Moglia, C., Manera, U., Bersano, E., Brunetti, M., Barberis, M., Veldink, J.H., van den Berg, L.H., Pearce, N., Sproviero, W., McLaughlin, R., Vajda, A., Hardiman, O., Rooney, J., Mora, G., Calvo, A., Al-Chalabi, A., 2018. The multistep hypothesis of ALS revisited: the role of genetic mutations. Neurology 91 (7), e635–e642.

Coetzee, T., Fujita, N., Dupree, J., Shi, R., Blight, A., Suzuki, K., Suzuki, K., Popko, B., 1996. Myelination in the absence of galactocerebroside and sulfatide: normal structure with abnormal function and regional instability. Cell 86 (2), 209–219.

de Koning, A.P., Gu, W., Castoe, T.A., Batzer, M.A., Pollock, D.D., 2011. Repetitive elements may comprise over two-thirds of the human genome. PLoS Genet 7 (12), e1002384.

DeJesus-Hernandez, M., Kocerha, J., Finch, N., Crook, R., Baker, M., Desaro, P., Johnston, A., Rutherford, N., Wojtas, A., Kennelly, K., Wszolek, Z.K., Graff-Radford, N., Boylan, K., Rademakers, R., 2010. De novo truncating FUS gene mutation as a cause of sporadic amyotrophic lateral sclerosis. Hum Mutat 31 (5), E1377–E1389.

DeJesus-Hernandez, M., Mackenzie, I.R., Boeve, B.F., Boxer, A.L., Baker, M., Rutherford, N.J., Nicholson, A.M., Finch, N.A., Flynn, H., Adamson, J., Kouri, N., Wojtas, A., Sengdy, P., Hsiung, G.Y., Karydas, A., Seeley, W.W., Josephs, K.A., Coppola, G., Geschwind, D.H., Wszolek, Z.K., Feldman, H., Knopman, D.S., Petersen, R.C., Miller, B.L., Dickson, D.W., Boylan, K.B., Graff-Radford, N.R., Rademakers, R., 2011. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron 72 (2), 245–256.

Dolzhenko, E., Bennett, M.F., Richmond, P.A., Trost, B., Chen, S., van Vugt, J., Nguyen, C., Narzisi, G., Gainullin, V.G., Gross, A.M., Lajoie, B.R., Taft, R.J., Wasserman, W.W., Scherer, S.W., Veldink, J.H., Bentley, D.R., Yuen, R.K.C., Bahlo, M., Eberle, M.A., 2020. ExpansionHunter Denovo: a computational method for locating known and novel repeat expansions in short-read sequencing data. Genome Biol 21 (1), 102.

Dolzhenko, E., Deshpande, V., Schlesinger, F., Krusche, P., Petrovski, R., Chen, S., Emig-Agius, D., Gross, A., Narzisi, G., Bowman, B., Scheffler, K., van Vugt, J., French, C., Sanchis-Juan, A., Ibanez, K., Tucci, A., Lajoie, B.R., Veldink, J.H., Raymond, F.L., Taft, R.J., Bentley, D.R., Eberle, M.A., 2019. ExpansionHunter: a sequence-graph-based tool to analyze variation in short tandem repeat regions. Bioinformatics 35 (22), 4754–4756.

Driskell, R.R., Watt, F.M., 2015. Understanding fibroblast heterogeneity in the skin. Trends Cell Biol 25 (2), 92–99.

Dziegiel, P., Owczarek, T., Plazuk, E., Gomulkiewicz, A., Majchrzak, M., Podhorska-Okolow, M., Driouch, K., Lidereau, R., Ugorski, M., 2010. Ceramide galactosyltransferase (UGT8) is a molecular marker of breast cancer malignancy and lung metastases. Br J Cancer 103 (4), 524–531.

Elden, A.C., Kim, H.J., Hart, M.P., Chen-Plotkin, A.S., Johnson, B.S., Fang, X., Armakola, M., Geser, F., Greene, R., Lu, M.M., Padmanabhan, A., Clay-Falcone, D., McCluskey, L., Elman, L., Juhr, D., Gruber, P.J., Rub, U., Auburger, G., Trojanowski, J.Q., Lee, V.M., Van Deerlin, V.M., Bonini, N.M., Gitler, A.D., 2010. Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. Nature 466 (7310), 1069–1075.

Elliott, H.R., Tillin, T., McArdle, W.L., Ho, K., Duggirala, A., Frayling, T.M., Davey Smith, G., Hughes, A.D., Chaturvedi, N., Relton, C.L., 2014. Differences in smoking associated DNA methylation patterns in South Asians and Europeans. Clin Epigenetics 6 (1), 4.

Errichiello, E., Zagnoli-Vieira, G., Rizzi, R., Garavelli, L., Caldecott, K.W., Zuffardi, O., 2020. Characterization of a novel loss-of-function variant in TDP2 in two adult patients with spinocerebellar ataxia autosomal recessive 23 (SCAR23). J Hum Genet 65 (12), 1135–1141.

Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T., Rosen, N., Kohn, A., Twik, M., Safran, M., Lancet, D., Cohen, D., 2017. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. Database (Oxford) 2017.

Francioli, L.C., Polak, P.P., Koren, A., Menelaou, A., Chun, S., Renkens, I., Genome of the Netherlands, C., van Duijn, C.M., Swertz, M., Wijmenga, C., van Ommen, G., Slagboom, P.E., Boomsma, D.I., Ye, K., Guryev, V., Arndt, P.F., Kloosterman, W.P., de Bakker, P.I.W., Sunyaev, S.R., 2015. Genome-wide patterns and properties of de novo mutations in humans. Nat Genet 47 (7), 822–826.

Genome of the Netherlands, C., 2014. Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nat Genet 46 (8), 818–825.

Gilissen, C., Hehir-Kwa, J.Y., Thung, D.T., van de Vorst, M., van Bon, B.W., Willemsen, M.H., Kwint, M., Janssen, I.M., Hoischen, A., Schenck, A., Leach, R., Klein, R., Tearle, R., Bo, T., Pfundt, R., Yntema, H.G., de Vries, B.B., Kleefstra, T., Brunner, H.G., Vissers, L.E., Veltman, J.A., 2014. Genome sequencing identifies major causes of severe intellectual disability. Nature 511 (7509), 344–347.

Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., Klotzle, B., Bibikova, M., Fan, J.B., Gao, Y., Deconde, R., Chen, M., Rajapakse, I., Friend, S.,

Ideker, T., Zhang, K., 2013. Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol Cell 49 (2), 359–367.
Holmes, S.E., Hearn, E.O., Ross, C.A., Margolis, R.L., 2001. SCA12: an unusual mutation leads to an unusual spinocerebellar ataxia. Brain Res Bull 56 (3-4), 397–403.
Hop, P.J., Zwamborn, R.A.J., Hannon, E.J., Dekker, A.M., van Eijk, K.R., Walker, E.M., Iacoangeli, A., Jones, A.R., Shatunov, A., Khleifat, A.A., Opie-Martin, S., Shaw, C.E., Morrison, K.E., Shaw, P.J., McLaughlin, R.L., Hardiman, O., Al-Chalabi, A., Van Den Berg, L.H., Mill, J., Veldink, J.H., 2020. Cross-reactive probes on Illumina DNA methylation arrays: a large study on ALS shows that a cautionary approach is warranted in interpreting epigenome-wide association studies. NAR Genom Bioinform 2 (4), lqaa105.
Horvath, S., 2013. DNA methylation age of human tissues and cell types. Genome Biol 14 (10), R115.
Kim, Y.E., Oh, K.W., Kwon, M.J., Choi, W.J., Oh, S.I., Ki, C.S., Kim, S.H., 2015. De novo FUS mutations in 2 Korean patients with sporadic amyotrophic lateral sclerosis. Neurobiol Aging 36 (3), 1604 e1617-1609.
Kovtun, I.V., Liu, Y., Bjoras, M., Klungland, A., Wilson, S.H., McMurray, C.T., 2007. OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells. Nature 447 (7143), 447–452.
Kun-Rodrigues, C., Ganos, C., Guerreiro, R., Schneider, S.A., Schulte, C., Lesage, S., Darwent, L., Holmans, P., Singleton, A., International Parkinson's Disease Genomics, C., Bhatia, K., Bras, J., 2015. A systematic screening to identify de novo mutations causing sporadic early-onset Parkinson's disease. Hum Mol Genet 24 (23), 6711–6720.
Laffita-Mesa, J.M., Rodriguez Pupo, J.M., Moreno Sera, R., Vazquez Mojena, Y., Kouri, V., Laguna-Salvia, L., Martinez-Godales, M., Valdevila Figueira, J.A., Bauer, P.O., Rodriguez-Labrada, R., Gonzalez Zaldivar, Y., Paucar, M., Svenningsson, P., Velazquez Perez, L., 2013. De novo mutations in ataxin-2 gene and ALS risk. PLoS One 8 (8), e70560.
Li, C., Sun, S.Y., Khuri, F.R., Li, R., 2011. Pleiotropic functions of EAPII/TTRAP/TDP2: cancer development, chemoresistance and beyond. Cell Cycle 10 (19), 3274–3283.
McCann, E.P., Henden, L., Fifita, J.A., Zhang, K.Y., Grima, N., Bauer, D.C., Chan Moi Fat, S., Twine, N.A., Pamphlett, R., Kiernan, M.C., Rowe, D.B., Williams, K.L., Blair, I.P., 2020. Evidence for polygenic and oligogenic basis of Australian sporadic amyotrophic lateral sclerosis. J Med Genet 58 (2), 87–95.
Pamphlett, R., Morahan, J.M., 2011. Copy number imbalances in blood and hair in monozygotic twins discordant for amyotrophic lateral sclerosis. J Clin Neurosci 18 (9), 1231–1234.
Paulson, H.L., Shakkottai, V.G., Clark, H.B., Orr, H.T., 2017. Polyglutamine spinocerebellar ataxias - from genes to potential treatments. Nat Rev Neurosci 18 (10), 613–626.
Pidsley, R., CC, Y.W., Volta, M., Lunnon, K., Mill, J., Schalkwyk, L.C., 2013. A data–driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics 14, 293.
Ryan, M., Heverin, M., McLaughlin, R.L., Hardiman, O., 2019. Lifetime Risk and Heritability of Amyotrophic Lateral Sclerosis. JAMA Neurol 76 (11), 1367–1374.
Saffari, A., Silver, M.J., Zavattari, P., Moi, L., Columbano, A., Meaburn, E.L., Dudbridge, F., 2018. Estimation of a significance threshold for epigenome-wide association studies. Genet Epidemiol 42 (1), 20–33.
Sanchez-Contreras, M., Cardozo-Pelaez, F., 2017. Age-related length variability of polymorphic CAG repeats. DNA Repair (Amst) 49, 26–32.
Shibutani, S., Takeshita, M., Grollman, A.P., 1991. Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. Nature 349 (6308), 431–434.
Souren, N.Y., Gerdes, L.A., Lutsik, P., Gasparoni, G., Beltran, E., Salhab, A., Kumpfel, T., Weichenhan, D., Plass, C., Hohlfeld, R., Walter, J., 2019. DNA methylation signatures of monozygotic twins clinically discordant for multiple sclerosis. Nat Commun 10 (1), 2094.
Steinberg, K.M., Yu, B., Koboldt, D.C., Mardis, E.R., Pamphlett, R., 2015. Exome sequencing of case-unaffected-parents trios reveals recessive and de novo genetic variants in sporadic. ALS. Sci Rep 5, 9124.
Tarr, I.S., McCann, E.P., Benyamin, B., Peters, T.J., Twine, N.A., Zhang, K.Y., Zhao, Q., Zhang, Z.H., Rowe, D.B., Nicholson, G.A., Bauer, D., Clark, S.J., Blair, I.P., Williams, K.L., 2019. Monozygotic twins and triplets discordant for amyotrophic lateral sclerosis display differential methylation and gene expression. Sci Rep 9 (1), 8254.
Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E., Fish, P., Harsha, B., Hathaway, C., Jupe, S.C., Kok, C.Y., Noble, K., Ponting, L., Ramshaw, C.C., Rye, C.E., Speedy, H.E., Stefancsik, R., Thompson, S.L., Wang, S., Ward, S., Campbell, P.J., Forbes, S.A., 2019. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res 47 (D1), D941–D947.
Tazelaar, G.H.P., Boeynaems, S., De Decker, M., van Vugt, J., Kool, L., Goedee, H.S., McLaughlin, R.L., Sproviero, W., Iacoangeli, A., Moisse, M., Jacquemyn, M., Daelemans, D., Dekker, A.M., van der Spek, R.A., Westeneng, H.J., Kenna, K.P., Assialioui, A., Da Silva, N., Project Min, E.A.L.S.S.C., Povedano, M., Pardina, J.S.M., Hardiman, O., Salachas, F., Millecamps, S., Vourc'h, P., Corcia, P., Couratier, P., Morrison, K.E., Shaw, P.J., Shaw, C.E., Pasterkamp, R.J., Landers, J.E., Van Den Bosch, L., Robberecht, W., Al-Chalabi, A., van den Berg, L.H., Van Damme, P., Veldink, J.H, van Es, M.A., 2020. ATXN1 repeat expansions confer risk for amyotrophic lateral sclerosis and contribute to TDP-43 mislocalization. Brain Commun 2 (2), fcaa064.
Tazelaar, G.H.P., Dekker, A.M., van Vugt, J., van der Spek, R.A., Westeneng, H.J., Kool, L., Kenna, K.P., van Rheenen, W., Pulit, S.L., McLaughlin, R.L., Sproviero, W., Iacoangeli, A., Hubers, A., Brenner, D., Morrison, K.E., Shaw, P.J., Shaw, C.E., Panades, M.P., Mora Pardina, J.S., Glass, J.D., Hardiman, O., Al-Chalabi, A., van Damme, P., Robberecht, W., Landers, J.E., Ludolph, A.C., Weishaupt, J.H., van den Berg, L.H., Veldink, J.H., van Es, M.A., Project Min, E.A.L.S.S.C., 2019. Association of NIPA1 repeat expansions with amyotrophic lateral sclerosis in a large international cohort. Neurobiol Aging 74, 234 e239-234 e215.
Teschendorff, A.E., Breeze, C.E., Zheng, S.C., Beck, S., 2017. A comparison of reference-based algorithms for correcting cell-type heterogeneity in Epigenome-Wide Association Studies. BMC Bioinformatics 18 (1), 105.
Theda, C., Hwang, S.H., Czajko, A., Loke, Y.J., Leong, P., Craig, J.M., 2018. Quantitation of the cellular content of saliva and buccal swab samples. Sci Rep 8 (1), 6944.
Tsai, P.C., Bell, J.T., 2015. Power and sample size estimation for epigenome-wide association scans to detect differential DNA methylation. Int J Epidemiol 44 (4), 1429–1441.
van Blitterswijk, M., van Es, M.A., Hennekam, E.A., Dooijes, D., van Rheenen, W., Medic, J., Bourque, P.R., Schelhaas, H.J., van der Kooi, A.J., de Visser, M., de Bakker, P.I., Veldink, J.H., van den Berg, L.H., 2012. Evidence for an oligogenic basis of amyotrophic lateral sclerosis. Hum Mol Genet 21 (17), 3776–3784.
van der Spek, R.A.A., van Rheenen, W., Pulit, S.L., Kenna, K.P., van den Berg, L.H., Veldink, J.H., Project Min, E.A.L.S.S.C.p.s., 2019. The project MinE databrowser: bringing large-scale whole-genome sequencing in ALS to researchers and the public. Amyotroph Lateral Scler Frontotemporal Degener 20 (5-6), 432–440.
van Doormaal, P.T.C., Ticozzi, N., Weishaupt, J.H., Kenna, K., Diekstra, F.P., Verde, F., Andersen, P.M., Dekker, A.M., Tiloca, C., Marroquin, N., Overste, D.J., Pensato, V., Nurnberg, P., Pulit, S.L., Schellevis, R.D., Calini, D., Altmuller, J., Francioli, L.C., Muller, B., Castellotti, B., Motameny, S., Ratti, A., Wolf, J., Gellera, C., Ludolph, A.C., van den Berg, L.H., Kubisch, C., Landers, J.E., Veldink, J.H., Silani, V., Volk, A.E., 2017. The role of de novo mutations in the development of amyotrophic lateral sclerosis. Hum Mutat 38 (11), 1534–1541.
van Es, M.A., Hardiman, O., Chio, A., Al-Chalabi, A., Pasterkamp, R.J., Veldink, J.H., van den Berg, L.H., 2017. Amyotrophic lateral sclerosis. Lancet 390 (10107), 2084–2098.
van Iterson, M., Cats, D., Hop, P., Consortium, B., Heijmans, B.T., 2018. omicsPrint: detection of data linkage errors in multiple omics studies. Bioinformatics 34 (12), 2142–2143.
van Iterson, M., Tobi, E.W., Slieker, R.C., den Hollander, W., Luijk, R., Slagboom, P.E., Heijmans, B.T., 2014. MethylAid: visual and interactive quality control of large Illumina 450k datasets. Bioinformatics 30 (23), 3435–3437.
van Rheenen, W., van der Spek, R.A.A., Bakker, M.K., van Vugt, J., Hop, P.J., Zwamborn, R.A.J., de Klein, N., Westra, H.J., Bakker, O.B., Deelen, P., Shireby, G., Hannon, E., Moisse, M., Baird, D., Restuadi, R., Dolzhenko, E., Dekker, A.M., Gawor, K., Westeneng, H.J., Tazelaar, G.H.P., van Eijk, K.R., Kooyman, M, Byrne, R.P., Doherty, M., Heverin, M., Al Khleifat, A., Iacoangeli, A., Shatunov, A., Ticozzi, N., Cooper-Knock, J., Smith, B.N., Gromicho, M., Chandran, S., Pal, S., Morrison, K.E., Shaw, P.J., Hardy, J., Orrell, R.W., Sendtner, M., Meyer, T., Basak, N., van der Kooi, A.J., Ratti, A., Fogh, I., Gellera, C., Lauria, G., Corti, S., Cereda, C., Sproviero, D., D'Alfonso, S., Soraru, G., Siciliano, G., Filosto, M., Padovani, A., Chio, A., Calvo, A., Moglia, C., Brunetti, M., Canosa, A., Grassano, M., Beghi, E., Pupillo, E., Logroscino, G., Nefussy, B., Osmanovic, A., Nordin, A., Lerner, Y., Zabari, M., Gotkine, M., Baloh, R.H., Bell, S., Vourc'h, P., Corcia, P., Couratier, P., Millecamps, S., Meininger, V., Salachas, F., Mora Pardina, J.S., Assialioui, A., Rojas-Garcia, R., Dion, P.A., Ross, J.P., Ludolph, A.C., Weishaupt, J.H., Brenner, D., Freischmidt, A., Bensimon, G., Brice, A., Durr, A., Payan, C.A.M., Saker-Delye, S., Wood, N.W., Topp, S., Rademakers, R., Tittmann, L., Lieb, W., Franke, A., Ripke, S., Braun, A., Kraft, J., Whiteman, D.C., Olsen, C.M., Uitterlinden, A.G., Hofman, A., Rietschel, M., Cichon, S., Nothen, M.M., Amouyel, P., Consortium, S., Consortium, P., Consortium, S., Consortium, S., Traynor, B.J., Singleton, A.B., Mitne Neto, M., Cauchi, R.J., Ophoff, R.A., Wiedau-Pazos, M., Lomen-Hoerth, C., van Deerlin, V.M., Grosskreutz, J., Roediger, A., Gaur, N., Jork, A., Barthel, T., Theele, E., Ilse, B., Stubendorff, B., Witte, O.W., Steinbach, R., Hubner, C.A., Graff, C., Brylev, L., Fominykh, V., Demeshonok, V., Ataulina, A., Rogelj, B., Koritnik, B., Zidar, J., Ravnik-Glavac, M., Glavac, D., Stevic, Z., Drory, V., Povedano, M., Blair, I.P., Kiernan, M.C., Benyamin, B., Henderson, R.D., Furlong, S., Mathers, S., McCombe, P.A., Needham, M., Ngo, S.T., Nicholson, G.A., Pamphlett, R., Rowe, D.B., Steyn, F.J., Williams, K.L., Mather, K.A., Sachdev, P.S., Henders, A.K., Wallace, L., de Carvalho, M., Pinto, S., Petri, S., Weber, M., Rouleau, G.A., Silani, V., Curtis, C.J., Breen, G., Glass, J.D., Brown Jr., R.H., Landers, J.E., Shaw, C.E., Andersen, P.M., Groen, E.J.N., van Es, M.A., Pasterkamp, R.J., Fan, D., Garton, F.C., McRae, A.F., Davey Smith, G., Gaunt, T.R., Eberle, M.A., Mill, J., McLaughlin, R.L., Hardiman, O., Kenna, K.P., Wray, N.R., Tsai, E., Runz, H., Franke, L., Al-Chalabi, A., Van Damme, P., van den Berg, L.H., Veldink, J.H., 2021. Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. Nat Genet 53 (12), 1636–1648.
Vucic, S., Higashihara, M., Sobue, G., Atsuta, N., Doi, Y., Kuwabara, S., Kim, S.H., Kim, I., Oh, K.W., Park, J., Kim, E.M., Talman, P., Menon, P., Kiernan, M.C., Consortium, P., 2020. ALS is a multistep process in South Korean, Japanese, and Australian patients. Neurology 94 (15), e1657–e1663.
Wang, W., Corominas, R., Lin, G.N., 2019. De novo Mutations From Whole Exome Sequencing in Neurodevelopmental and Psychiatric Disorders: From Discovery to Application. Front Genet 10, 258.

Wingo, T.S., Cutler, D.J., Yarab, N., Kelly, C.M., Glass, J.D., 2011. The heritability of amyotrophic lateral sclerosis in a clinically ascertained United States research registry. PLoS One 6 (11), e27985.

Young, P.E., Kum Jew, S., Buckland, M.E., Pamphlett, R., Suter, C.M., 2017. Epigenetic differences between monozygotic twins discordant for amyotrophic lateral sclerosis (ALS) provide clues to disease pathogenesis. PLoS One 12 (8), e0182638.

Zagnoli-Vieira, G., Bruni, F., Thompson, K., He, L., Walker, S., de Brouwer, A.P.M., Taylor, R.W., Niyazov, D., Caldecott, K.W., 2018. Confirming TDP2 mutation in spinocerebellar ataxia autosomal recessive 23 (SCAR23). Neurol Genet 4 (4), e262.

Zhang, Q., Vallerga, C.L., Walker, R.M., Lin, T., Henders, A.K., Montgomery, G.W., He, J., Fan, D., Fowdar, J., Kennedy, M., Pitcher, T., Pearson, J., Halliday, G., Kwok, J.B., Hickie, I., Lewis, S., Anderson, T., Silburn, P.A., Mellick, G.D., Harris, S.E., Redmond, P., Murray, A.D., Porteous, D.J., Haley, C.S., Evans, K.L., McIntosh, A.M., Yang, J., Gratten, J., Marioni, R.E., Wray, N.R., Deary, I.J., McRae, A.F., Visscher, P.M., 2019. Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. Genome Med 11 (1), 54.

Zhang, Y., Unnikrishnan, A., Deepa, S.S., Liu, Y., Li, Y., Ikeno, Y., Sosnowska, D., Van Remmen, H., Richardson, A., 2017. A new role for oxidative stress in aging: The accelerated aging phenotype in Sod1(-/)(-) mice is correlated to increased cellular senescence. Redox Biol 11, 30–37.

Zhou, W., Laird, P.W., Shen, H., 2017. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. Nucleic Acids Res 45 (4), e22.

Zoghi, S., Khamirani, H.J., Hassanipour, H., Bostanian, P., Masoudian, R., Dastgheib, S.A., 2021. A novel non-sense mutation in TDP2 causes spinocerebellar ataxia autosomal recessive 23 accompanied by bilateral upward gaze; report of a case and review of the literature. Eur J Med Genet 64 (12), 104348.