The First Insight into the Hereditary Fusion Gene Landscape of Amyotrophic Lateral Sclerosis

Jinfeng Yang¹, Fenghua Yuan², Anna Palovcak², Ling Fei³, Noah Zhuo⁴, NYGC ALS Consortium*, Yanbin Zhang², Degen Zhuo⁴

¹Department of Anesthesiology, Hunan Cancer Hospital, The Affiliated Cancer Hospital of Xiangya School of Medicine, Central South University, Changsha, Hunan, China; ²the Department of Biochemistry and Molecular Biology, University of Miami Miller School of Medicine, Miami, Florida 33136, ³Dept of Cardiology, The Affiliated Hospital of North Sichuan, Chengdu Xinhua Hospital, Cheng Du, 610055, China; ⁴SplicinagCodes, BioTailor Inc, 7328 SW 82nd Street, #B114, Miami, FL 33143 USA

Correspondence should be addressed to: SplicinagCodes, BioTailor Inc, 7328 SW 82nd Street, #B114, Miami, FL 33143 USA, degen.zhuo@gmail.com

Abstract

Amyotrophic lateral sclerosis (ALS) is a progressive nervous system disease that causes loss of muscle control. Over 30 mutated genes are associated with ASL. However, 90-95% of ASL cases have been found without a family history. Here, we have analyzed RNA-Seq data of NYGC ALS Consortium and identified fusion transcripts from ASL patients and non-neurologic controls (NNC). In this study, we combined previously-curated 1180 monozygotic (MZ) hereditary fusion genes (HFGs), and 204 HFGs discovered from NNC to analyze ASL fusion transcripts and identified 348 HFGs. Comparative analysis between ASL and GTEx shows that 139 HFGs are associated with ASL and ranged from 10.4% to 98.7% of 77 ASL patients. The most recurrent HFG is ZNF528-ZNF880, detected in 98.7% of 77 ASL patients and 4.5% of 133 GTEx brain cortices. Alignments of HFG transcripts from ASL with fusion transcripts from mesial temporal lobe epilepsy (MTLE) and Alzheimer’s disease (AD) showed that 43.9% and 11.6% of the ASL HFGs were present in MTLE and AD, respectively. The most recurrent and common HFG among ASL, MTLE, and AD was ADAMTSL3-SH3GL3, which behaves like ubiquitously-expressed SH3GL3-ADAMTSL3 epigenetic fusion gene (EFG) and shows that ADAMTSL3-SH3GL3 is a potential dormant or differentially-expressed HFG (dHFG), suggesting that they have common pathophysiological mechanisms. These HFGs associated with ASL have shown that HFGs are the missing genetic heritability and provide novel therapeutic targets for more efficient therapeutic drugs and methods to treat and cure many neurological diseases.
Keywords: Hereditary fusion gene, Amyotrophic Lateral Sclerosis, epilepsy, Alzheimer’s disease, RNA-Seq, Epigenetic fusion gene

Introduction

Amyotrophic lateral sclerosis (ALS) is a progressive, fatal neurodegenerative disease characterized by progressive degenerative changes in upper and lower motor neurons that overlap with frontotemporal lobar degeneration clinically, morphologically and genetically\textsuperscript{1}. ALS clinical symptoms include weakness of the bulbar and limb muscles, hyperreflexia, spasticity of the arms and legs, and respiratory failure\textsuperscript{2, 3}. These symptoms start with onset typically in the fifth or sixth decade of life, and death
commonly occurs within 2-4 years of diagnosis\(^2\). ALS is closely related to frontotemporal dementia (FTD), a progressive neurodegenerative disease characterized by degeneration of the frontal and temporal lobes of the brain, resulting in behavior, personality, and language disturbances. Up to 15% of ALS patients are also diagnosed with frontotemporal dementia (FTD-ALS patients)\(^4\). Many studies have shown clinical, pathological, and genetic commonalities between them. Therefore, ASL and FTD have now considered two manifestations of one disease continuum of the FTD-ALS spectrums\(^5\).

ALS is associated with over 30 mutated genes, including \(SOD1\), \(TARDBP\), \(FUS\), \(OPTN\), \(VCP\), \(UBQLN2\), \(C9ORF72\), and \(PFN1\)\(^6-8\). However, they are typically present in fewer than 5-10% of all ALS patients, and most ASL cases remain unexplained\(^9\)\(^10\). Genome-wide associations (GWAS) have genotyped numerous single nucleotide polymorphisms (SNPs) present in at least 1% of the human population\(^11\) and identified some common and rare SNPs associated with ASL across the genome\(^12-14\). More recently, the discovery of genes associated with ASL has been shifted to next-generation sequencing, such as whole-genome sequencing (WGS) and RNA-Seq. Many WGS studies have confirmed previous findings and identified many common and rare variants associated with ASL\(^15, 16\). Brown et al. have performed RNA-Seq analysis on the ASL postmortem cortex samples of NYGC ALS Consortium and found that TDP-43 loss resulted from retrotransposon and ALS-risk SNPs drive mis-splicing and depletion of UNC13A\(^17, 18\). Many studies have speculated that structural variants may be a source of missing ASL heritability\(^10, 19\). More recently, Raghav et al. analyzed RNA-Seq data from Target ALS and the NYGC ALS Consortium and identified 607 unique fusion genes, some significantly higher in ASL samples than those in control\(^20\), suggesting that fusion genes generated by structural variants may be responsible for ASL.
As discussed above, advances in RNA-Seq have made it possible to discover fusion transcripts generated by structural variants. However, misconceptions and poor algorithms for repetitive sequences have made fusion transcripts discovered incomplete, irreproducible, and unreliable. In order to overcome these problems, we have used the splicingcode theory to develop SCIF (SplicingCodes Identify Fusion Genes) to discover more reproducible and accurate fusion transcripts and make it possible to compare and analyze data among different experiments. Validation of highly-recurrent fusion genes has cast doubts that fusion genes are cancerous and generated from somatic genomic abnormalities. \textit{KANSARL} (\textit{KANSL1-ARL17A}) is validated as the first predisposition (hereditary) fusion genes specific to 29\% of populations of European ancestry. TPM4-KLF2, detected in 92.2\% of 727 multiple myeloma patients and initially thought to be a multiple myeloma-specific somatic fusion gene, has been detected in all five healthy bone marrows. Based on their genomic structures, genetics, and evolution, we have classified the fusion genes into somatic, epigenetic, and hereditary. Hereditary fusion genes (HFGs) are the fusion genes that offspring inherit from their parents, excluding readthrough transcripts of two identical-stand neighboring genes, named epigenetic fusion genes (EFGs). Furthermore, we have used 1180 HFGs identified from monozygotic (MZ) twins to analyze fusion transcripts from multiple myeloma and acute myeloid leukemia (AML) and discovered that HFGs are dominant inherited factors and mutated genes have played minor roles in cancer inheritance. EFGs have reflected developmentally interactive consequences between genetic and environmental factors.

In this study, we analyzed RNA-Seq data of NYGC ALS Consortium and identified large numbers of fusion transcripts from RNA-Seq data of ASL patients. Surprisingly, we have observed large numbers of highly-recurrent fusion genes from their non-neurological controls, 204 of which are determined to be HFGs. We have used these newly-identified HFGs and the previously-curated 1180 HFGs to analyze
the fusion transcripts. From 358 HFGs, we have identified 139 HFGs associated with ASL ranging from 10.4% to 98.7% of ASL patients. We have discovered that many highly recurrent ASL HFGs have also been detected in mesial temporal lobe epilepsy (MTLE) and Alzheimer's disease (AD), suggesting common pathophysiological mechanisms. \textit{ADAMTSL3-SH3GL3} HFG behaves like \textit{SH3GL3-ADAMTSL3} EFG, suggesting it has a much higher recurrent frequency than its GTEx brain cortex counterpart. This study has shown that HFGs are the missing heritability of complex diseases and traits and provides potential therapeutic targets for more efficient novel therapeutic drugs and methods for ASL and other neurological diseases.

\textbf{Materials and Methods}

\textbf{Materials}

\textbf{Amyotrophic lateral sclerosis RNA-Seq data:} Amyotrophic lateral sclerosis (ALS) RNA-Seq data (GEO: GSE124439) was downloaded from NCBI (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA512012). The ASL dataset contained 176 RNA-Seq data and included 145 samples of ASL spectrum from 77 unique ASL patients, 14 samples of other neurological disorders (named as OND to avoid confusion) from seven unique individuals, and 17 non-neurological controls (NNC) from 12 unique individuals. The RNA-Seq was performed on postmortem patients’ frontal cortex, medial motor cortex, and lateral motor cortex.

\textbf{Mesial temporal lobe epilepsy RNA-Seq data:} The mesial temporal lobe epilepsy (MTLE) RNA-Seq data (GEO: GSE134697) was downloaded from NCBI (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA556159). The MTLE RNA-Seq dataset contained the hippocampus and neocortex of 17 MTLE patients and two healthy controls. This study used only RNA-Seq data from the neocortex of 17 MTLE patients.

The First Insight into the Hereditary Fusion Gene Landscape of ASL
Alzheimer's disease RNA-Seq data: The Alzheimer's disease (AD) RNA-Seq data (GEO: GSE53697) was downloaded from NCBI (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA232669). The AD data had 17 brain samples, including nine patients and eight controls.

GTEx brain cortex RNA-Seq data: the Genotype-Tissue Expression (GTEx) RNA-Seq data (dbGaP: phs000424) was downloaded from NCBI (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA75899). We selected and performed RNA-Seq data from 133 brain cortexes of 133 unique individuals as the healthy controls.

Methods

Identification of fusion transcripts from RNA-Seq data: We used SCIF (SplicingCodes Identify Fusion Genes) to perform RNA-Seq analysis at the default conditions. Since SCIF used raw RNA-Seq reads without any assembly to discover the fusion junctions and followed by extended sequence alignments, each RNA-Seq read was either a fusion transcript or not a fusion transcript. Human was not involved in determining whether a sample with a fusion junction or a sample with three fusion junctions was hereditary fusion gene-positive. However, one of the worst defects in discovering the hereditary fusion genes from RNA-Seq data was that levels of gene expression, RNA-Seq qualities, and the number of RNA-Seq reads would affect the analysis. Consequently, we must take caution to handle and analyzing data.

Statistical analysis: To compare two different populations, we have used the two-tailed Z score analyses to determine whether the two populations differ significantly in their genetic characteristics. We set the null hypothesis to be that there is no difference between the two population proportions. Z scores are calculated based on the following formula:

\[
Z = \frac{(\bar{p}_1 - \bar{p}_2) - 0}{\sqrt{\frac{\bar{p}(1 - \bar{p})}{n_1} + \frac{\bar{p}(1 - \bar{p})}{n_2}}}
\]

The First Insight into the Hereditary Fusion Gene Landscape of ASL
\[ \bar{p}_1 \text{ and } \bar{p}_2 \text{ were the population standard errors of } p_1 \text{ and } p_2. \]

\[ \bar{p}_1 = x_1 / n_1; \quad \bar{p}_2 = x_2 / n_2; \quad \bar{p} = (x_1 + x_2) / (n_1 + n_2). \]

To perform HFG comparisons between ASL and GTEx, if the recurrent frequency was zero in GTEx, we used one to replace zero of the raw count to avoid the infinite.

Results

Discovering novel HFGs from RNA-Seq data of the non-neurologic controls.

We downloaded the RNA-Seq data of NYGC ALS Consortium (GEO: GSE124439), which contained a total of 176 frontal cortex/motor cortex, including 145 samples of ASL spectrum from 77 unique ASL patients, 14 samples of other neurological disorders (named as OND to avoid confusions) from seven unique individuals, and 17 non-neurological controls (NNC) from 12 unique individuals (Fig.1a). We used SCIF to analyze RNA-Seq data of this dataset and identified 64,000 fusion transcripts from samples of ASL and OND and 8,190 fusion transcripts from 12 non-neurological controls. Manual inspections showed that fusion genes had been discovered at such high recurrent frequencies in non-neurological samples that traditional logic and biological models cannot explain. For example, \textit{ADAMTSL3-SH3GL3}, an inversion of \textit{SH3GL3-ADAMTSL3}, had been observed in all 12 non-neurological controls. If structural variants generating a fusion gene per individual had a rate of \(3.6 \times 10^{-29}\), the chance of all 12 NNC individuals having the identical \textit{ADAMTSL3-SH3GL3} fusion gene generated by random somatic genomic alterations was \(4.7 \times 10^{-18}\) and a mathematically impossible event, suggesting it had to be inherited from their parents\(^{24}\). Fig.1b shows the procedure to discover hereditary fusion genes (HFGs) from NNC. If we set the cutoff of recurrent frequencies at two or more in 12 NNC individuals, the minimum recurrent frequencies of 16.7% were at least 11.9 fold of 1.4% of the random chance. We initially identified 257 putative HFG transcripts whose recurrent frequencies...
ranged from 16.7% to 100%. After performing Blast analyses in Aceview and NCBI, we removed 24 EFG transcripts and potential alternative-spliced transcripts and obtained 233 HFG transcripts encoded by 204 HFGs (Table 1). Out of 233 HFG transcripts, we discovered that 33 (16.2%) HFGs coding for 38 transcripts were from tandem gene duplications (Table 2), suggesting that tandem gene duplications were essential for genotype and phenotype diversities.

**Identifications of HFGs from RNA-Seq data of amyotrophic lateral sclerosis (ALS) patients**

Next, we used these 204 HFGs discovered from NNC plus the previously curated 1180 MZ HFGs (Fig.1c) to analyze the total fusion transcripts of 154 ASL and 14 OND samples. We identified 358 HFGs encoding 870 HFG transcripts, ranging from 0.6% to 98.1% of 159 ASL and OND samples, the average of which was 2.4 per HFG, suggesting that many HFGs had multiple isoforms. Fig.1d showed that 42 out of 358 HFGs were present in both NNC and MZ HFG datasets and counted for 21.4% of NNC 196 HFGs, supporting our rationales of discovering novel HFGs from NNC. Further analysis showed that only an additional 17 of 196 HFGs were present in GTEx blood samples (data not shown) and suggested that most 196 HFGs were potentially differentially expressed and not expressed in blood and other samples. This differential HFG expression potentially masked the accuracy of the HFG frequencies detected in blood samples. Since these HFG transcripts were from the frontal cortex, medial motor cortex, and lateral motor cortex of ASL and OND samples, we must study whether these HFGs were differentially expressed among the frontal cortex and medial and lateral motor cortices and affected the results.

To understand whether HFGs were differentially expressed and affected HFG results analyzed afterward, we performed comparative analyses of HFGs among the frontal cortices, medial motor cortex, and lateral motor cortex. After inspecting the HFG distributions of the frontal cortices, medial...
motor cortex, and lateral motor cortices, we found no HFGs with high recurrent frequencies were differentially expressed among these three brain tissues. STable 3 showed that 179 HFGs were encoding 268 HFG transcripts in all frontal cortex, medial motor cortex, and lateral cortex. Majorities of differences in the HFG transcripts among the three groups were small and statistically insignificant. We identified 22 ASL patients with data from all frontal cortex, medial motor cortex, and lateral motor cortex, similar to 22 three-duplication experiments. We performed a comparative data analysis of these 22 ASL patients. STable 4 showed that we identified 29 HFGs encoding 38 HFG transcripts expressed in all frontal cortex, medial motor cortex, and lateral motor cortex and suggested that these HFGs discovered by SCIF were highly reproducible. For example, ZNF528-ZNF880 and SNORD114_-MEG8_a were observed in all three tissues of 21 of 22 ASL patients and showed 100% reproducibility. Here, SNORD114_ was an abbreviated SNORD114-1-SNORD114-2-SNORD114-3^{25} while MEG8_ was MEG8-SNORD112-SNORD113-3^{25}. Careful inspections showed that some HFGs were alternatively spliced. For example, the frequency difference of the main SNORD114_-MEG8_ isoform a was within statistical errors among the frontal cortex, medial motor cortex, and lateral cortex. However, the recurrent frequencies of the SNORD114_-MEG8_ isoform c in the medial motor cortex and lateral motor cortices were 16.4% and 13.7% higher than their counterpart in the frontal cortex. They reflected alternatively splicing in the frontal, medial, and lateral motor cortices. Since we generally used the recurrent frequency of the main HFG isoform as the HFG recurrent frequency for genotype analysis, we could combine HFG data from the frontal cortex and medial and lateral motor cortices to perform further analysis.

Identification and characterization of HFGs associated with ASL

We used the combined HFG dataset described above to analyze the total fusion transcripts from 77 ASL patients. We identified 346 HFGs encoding 819 HFG transcripts whose recurrent frequencies
ranged from 1.3% to 98.7%. 345 (42.1%) HFG transcripts were detected in only one ASL patient, suggesting they were either causal HFGs or lowly-expressed HFG transcripts. For a more convenient analysis, we set the cutoff of the recurrent frequency at 10%. STable 5 showed that we observed 147 HFGs coding for 190 HFG transcripts ranging from 10.4% to 98.7%. Table 1 showed that the top sixteen HFGs encoding twenty-one HFG transcripts with recurrent frequencies of ≥50% ranged from 51.3% to 98.7%. Fig.2a showed that the most recurrent ZNF528-ZNF880, an inversion of ZNF880-ZNF528 on the plus strand of chromosome 19q13.41, was detected in 76 (98.7%) of 77 ASL patients and encoded a truncated zinc finger protein 880 without 35 aa N-terminus sequences (S Fig.1). Fig.2b showed that SNORD114_-MEG8_ was one of the most highly-expressed HFGs and coded for 20 isoforms, three of which were 92.1%, 85.5%, and 53.9% (Table 1). When we performed a comparative analysis between ASL and NCC, many HFG differences were insignificant. For example, ERMP-KIAA2026, reported in stomach cancer31, was detected in 75% of 75 ASL patients and 12 NNC samples. As described previously, if structural variants generating a fusion gene per individual had a rate of 3.6x10⁻² 29, 30, the chance of having nine NNC individuals who possessed ERMP-KIAA202631 by random somatic genomic abnormalities was 1.0X10⁻¹³ and a mathematically-impossible event. It was most likely that NNC was extraordinarily distorted for some unexplained reasons and unsuitable to use as a standard control.

To overcome the NNC distortion, we selected RNA-Seq data from 133 GTEx brain cortexes as healthy controls. We used 346 HFG from 77 ASL patients to analyze 26,000 fusion transcripts discovered from the GTEx brain cortexes and identified 165 HFGs coding for 188 HFG transcripts from GTEx brain cortexes, suggesting that 47.7% of 346 ASL HFGs were present in the GTEx brain cortexes. The most recurrent GTEx HFGs were KANSARL (KANSL1-ARL17A)22 and TRNAN35-FAM91A3P (Fig.2c) HFGs were detected in 34.6% and 31.6% of 133 GTEx brain cortexes. In comparison, 56.6% of the ASL
HFGs with recurrent frequencies of ≥10% overlapped with HFGs of GTEx brain cortexes, suggesting that the more recurrent ASL HFGs were, the more likely they were present in GTEx brain cortexes. Next, we performed a comparative analysis of the HFG differences between ASL and GTEx. We discovered that 139 out of 147 ASL HFGs ranged from 10.4% to 98.7% and were 1.2-244 folds higher than their GTEx brain cortex counterparts, which were statistically significant (STable 5). Table 1 showed that sixteen HFGs were encoding twenty-one HFG transcripts, whose recurrent frequencies ranged from 51.3% to 98.7% and were 1.8 to 120 folds of the GTEx brain cortex counterparts, which were statistically significant by the Z-test. Table 1 showed that ZNF528-ZNF880 (Fig.2a) was the most recurrent HFG detected in 98.7% of 77 ASL patients. In comparison, it was detected in 4.5% of 133 GTEx brain cortexes. The former was 21 folds of the latter, suggesting that ZNF528-ZNF880 HFG was necessary for ASL initiation and prognosis. The most significant difference of HFG recurrent frequencies between ASL and GTEx was SNORD114- MEG8 (Fig.2b), a local inversion of SNORD114-1-SNORD114-2-SNORD114-3 EFG25 and MEG8-SNORD112-SNORD113-3 EFG25, resulted in potentially more diverse expression of snoRNAs. TRNAN35-FAM91A3P was the HFG with the least frequency difference between ASL and GTEx. It was an inversion of TRNAN35, encoding transfer RNA asparagine 35 and FAM91A3P, coding for the family with sequence similarity 91 member A3 pseudogene (Fig.2c), detected in 31.6% of GTEx brain cortexes. Table 1 showed that TRNAN35-FAM91A3P, TRNAN35-SRGAP2P2, and TRNAN35-SEC22B were three TRNAN35-fused HFGs coded for putative 124 a.a, 177 a.a, 171 a.a, and 67 a.a polypeptides, which were potentially regulated by TRNAN35 via asparaginyl-tRNA synthetase and formed a natural network. Four HFGs, including SDHAP2-SDHAP2, YY1AP1-YY1AP1 (S Fig.2), FMNL2-FMNL2 (S Fig.3), and SPECC1L-SPECC1L, were shown to be generated via gene tandem duplications, suggesting that tandem gene duplications were more widespread than what was reported previously. FMNL2 encoded an FMNL2 protein with two exon insertions. FMNL2 has been shown to
regulate gliovascular interactions and is associated with vascular risk factors and cerebrovascular pathology in Alzheimer’s disease. YY1AP1-YY1AP1 encoded a truncated YY1-associated protein. A genome-wide sequence showed that a frameshift mutation and a large deletion in YY1AP1 resulted in a girl with a panvascular artery disease. SPECC1L-SPECC1L seemed to encode intact SPECC1L protein. It had been shown that mutated SPECC1L caused congenital diaphragmatic hernia -- a prominent feature of a SPECC1L-related syndrome.

Identification of common HFGs among ASL, Alzheimer’s disease (AD), and mesial temporal lobe epilepsy (MTLE)

To investigate whether the OND had different HFGs, fusion transcripts identified from 14 OND samples of seven patients were aligned with ASL HFG transcripts. We identified 206 HFGs encoding 318 HFG transcripts whose recurrent frequencies ranged from 14.3% to 100%, suggesting that 59.5% of 346 ASL HFGs were also found in seven OND patients, many of which had much higher recurrent frequencies in OND than in ASL. SNORD114-MEG8, FMNL2-FMNL2, and ADAMTSL3-SH3GL3 were the three top HFGs detected in all seven (100%) OND, while they were observed in 90.1%, 68.8%, and 87% of 77 ASL patients. Many other HFGs associated with ASL were also observed in OND at similar or even higher recurrent frequencies than their ASL counterparts. 59.5% of ASL HFGs observed in OND raised interesting questions about whether other neurological diseases, such as temporal lobe epilepsy (MTLE) and Alzheimer’s disease (AD), had common HFGs.

We used 819 HFG transcripts to analyze fusion transcripts of nine advanced AD patients and 17 neocortex samples of seventeen MTLE patients produced for a 2019 failed NIH grant application. We discovered that 152 HFGs coding for 214 HFG transcripts overlapped between ASL and MTLE and counted for 43.9% of 346 HFGs discovered in ASL, suggesting that ASL and MTLE had significantly inherited genetic similarity. In comparison, when 817 HFG fusion transcripts were aligned with fusion The First Insight into the Hereditary Fusion Gene Landscape of ASL
transcripts from nine AD patients, we identified only 40 HFGs encoding 74 HFG transcripts, accounting for only 11.6% of 346 HFGs observed in ASL and suggested that AD was much more divergent from ASL than MTLE.

Table 2 showed comparisons of highly-recurrent sixteen HFGs between ASL, MTLE, and AD, while the GTEx brain cortex counterparts were used as controls. All 21 HFG transcripts of sixteen ASL HFGs were present in seventeen MTLE patients. Surprisingly, The TRNAN35-FAM91A3P, TRNAN35-SRGAP2P2, and TRNAN35-SEC22B recurrent frequencies in ASL were 2.3 to 3.8 folds of the MTLE counterparts. The TRNAN35-FAM91A3P recurrent frequency of MTLE patients was 23.5% and lower than 31.6% of GTEx brain cortexes (Table 2).

Table 2 also showed that only half of the top sixteen highly recurrent ASL HFGs were detected in AD patients, while another half was absent in AD patients. The most notable HFGs absent in AD patients were YY1AP1-YY1AP1, FMNL2-FMNL2, and SPECC1L-SPECC1L generated by tandem duplications.

The most critical HFGs shared between ASL and AD were ZNF528-ZNF880 and ADAMTSL3-SH3GL3, detected in 98.7% and 87% of 77 ASL patients, and 88.9%, and 100% of nine AD patients, respectively. These highly recurrent HFGs were present in ASL, MTLE, and AD and supported that many neurological diseases might have common pathophysiological mechanisms.

Identification of ADAMTSL3-SH3GL3 as a potentially dormant or differentially-expressed HFG

ADAMTSL3-SH3GL3, reported recently, was an inversion of SH3GL3-ADAMTSL3 on 15q25, encoding thirteen isoforms and coded for 42 alternatively-spliced isoforms (Fig.3). Its ADAMTSL3-SH3GL3 main isoform a was expressed only at 19% of the KANSARL (KANASL1-ARL17A) main isoform level and encoded a conceptual 202 aa ADAMTSL3-SH3GL3 fusion protein over two-thirds of which was from SH3GL3. Since SH3GL3 was most expressed only in the brain and testis while ADAMTSL3 was expressed at a minimum level in the brain but at high levels in other tissues,
ADAMTSL3-SH3GL3 may alter SH3GL3 expression patterns and regulations. Table 3 showed that ADAMTSL3-SH3GL3 was also detected in ASL, OND, AD, and MTLE and ranged from 82.4% to 100%. Surprisingly, ADAMTSL3-SH3GL3 was also detected in 100% of 12 NNC samples and 87.5% of eight AD-healthy controls, significantly higher than 1.5% of 133 GTEx brain cortexes (Table 3). To address these enormous gaps among these controls, we used SH3GL3-ADAMTSL3, recently reported, as an example to compare both HFG expression patterns and behaviors. Table 3 showed that SH3GL3-ADAMTSL3, coding for an intact ADAMTS-like 3 protein, was ubiquitously expressed in all datasets analyzed except for GTEx brain cortexes, in 7.5% of which it was detected. Since the SH3GL3-ADAMTSL3 locus was present in almost all human individuals, its expression might have been induced due to interactions between environmental and genetic lesions during aging. Hence, we speculated that the ADAMTSL3-SH3GL3 frequency in human populations was higher than 1.5%. Most ADAMTSL3-SH3GL3 was dormant or differentially expressed at younger ages and activated by interactions between environmental and genetic lesions during aging. We named this dormant or differentially expressed fusion gene inherited from their parents the dormant or differentially-expressed HFG (dHFG). Hence, ADAMTSL3-SH3GL3 was the first potential dHFG. Table 3 showed that SH3GL3-ADAMTSL3 and ADAMTSL3-SH3GL3 were early-stage aging and neurological disease biomarkers.

Discussion
We analyzed the RNA-Seq data of the NYGC ALS Consortium and identified fusion transcripts from ASL, OND, and NNC. Many NNC fusion transcripts were highly recurrent. For example, ADAMTSL3-SH3GL3 was detected in 100% of 12 NNC samples and is consistent with our previous finding that ADAMTSL3-SH3GL3 was detected in 87.5% of eight AD-healthy controls. If structural variants generating a fusion gene per individual had a rate of 3.6x10^-2, generating seven and twelve
ADAMTSL3-SH3GL3 simultaneously in a dataset was 7.8\times10^{-11} and 4.7\times10^{-18}. Such low chances from two different sources\textsuperscript{18, 36} made it mathematically impossible for ADAMTSL3-SH3GL3 to be generated via random somatic genomic abnormalities. These observations supported that fusion transcripts generated via genomic alterations in healthy controls were potential HFGs. Hence, we developed a novel approach using fusion transcripts of NNC to discover HFGs. We used a mathematical model and blast against AceView and NCBI and identified 204 HFGs encoding 233 HFG transcripts. Bioinformatic analysis showed that 33 (16.2\%) HFGs were from tandem gene duplications. As we described previously, SCIF was not intended to discover fusion transcripts generated via tandem duplications, and their discovery was due to gene duplications and pseudogenes\textsuperscript{24}. Hence, HFGs produced via genomic tandem duplications were essential biomarkers for diseases and complex traits and might be significantly underestimated.

We combined 204 HFGs discovered from NNC and the previously-curated 1180 MZ HFGs into the HFG dataset to analyze the fusion transcripts of 154 RNA-Seq data from 77 ASL patients. We discovered 346 HFGs encoding 819 HFG transcripts from ASL patients, whose recurrent frequencies ranged from 1 (1.3\%) to 76 (98.7\%) of 77 ASL patients. Out of 819 HFG transcripts, 345 (42.1\%) HFG transcripts were detected in only one of 77 ASL patients and were shown to be causal HFGs and lowly-expressed HFG transcripts. When a cutoff of recurrent frequency was set at 10\%, we identified 147 HFGs coding for 188 HFG transcripts ranging from 10.4\% to 98.7\%. As described above, HFGs discovered from 12 NNC samples were extremely recurrent, some were detected in all 12 NNC samples, and the NNC genotypes were distorted. So, we have yet to discover any reasons which can explain these HFG genotype distortions. Hence, we selected 133 GTEx brain cortexes as healthy controls. Table 1 showed that 139 HFGs were 1.2-121 folds higher than their GTEx brain cortex counterparts and were shown by statistical analysis to be associated with ASL. Out of 139 HFGs...
associated with ASL, Table 2 showed sixteen HFGs with recurrent frequencies of ≥50%, some of which were also highly recurrent in OND, AD, and MTLE. 139 HFGs associated with ASL were four folds of more than 30 mutated genes associated with ASL. The HFG numbers and their recurrent frequencies suggested that HFGs were much more critical "inherited" genetic factors than the mutated genes. Hence, we can conclude that highly recurrent HFGs are the missing ASL heritability. More importantly, many neurological diseases such as ASL, MTLE, and AD shared highly recurrent HFGs and supported that these neurological diseases had common pathophysiological mechanisms. They potentially allowed us to develop more effective novel therapeutic drugs and methods to treat and cure neurological diseases.

One of the most highly-recurrent HFGs was ADAMTSL3-SH3GL3, an inversion of SH3GL3-ADAMTSL3 EFG (Fig.3). ADAMTSL3-SH3GL3 HFG was detected in at least 82.4% of ASL, OND, AD, and MTLE samples. More surprisingly, it was detected in 100% and 87.5% of 12 NNC samples and nine AD controls. Even though OND, AD, and TMLE data sizes are relatively small, the divergent sources of these data compensated for shortcomings. These data suggested that ADAMTSL3-SH3GL3 would be more recurrent among general populations than 1.5% of GTEx brain cortexes. To address this apparent disparity, we could take a lesson from SH3GL3-ADAMTSL3 EFG, the ancestors of ADAMTSL3-SH3GL3. Table 3 showed that SH3GL3-ADAMTSL3 was almost ubiquitously expressed in ASL, OND, AD, MTLE, and two controls. In comparison, it was detected in 7.5% of 133 GTEx brain cortexes and suggested that SH3GL3-ADAMTSL3 and ADAMTSL3-SH3GL3 behaved similarly. Since SH3GL3-ADAMTSL3 EFG can be treated as a permanent HFG with a recurrent frequency of 100%, we speculated that ADAMTSL3-SH3GL3 had a much higher recurrent frequency than 1.5% of GTEx brain cortexes and was activated to be expressed by interactions between genetic and environmental lesions during aging. We named this type of HFG the dormant or differentially-

The First Insight into the Hereditary Fusion Gene Landscape of ASL
expressed HFGs. We needed genotyping general populations to validate if ADAMTSL3-SH3GL3 was a dHFG or if other factors caused the disparity between neurological disease-related data and GTEx brain cortexes.

To study whether hereditary fusion genes were associated with ASL, AD, MTLE, and other neurological diseases, one of the most critical issues was the accuracy and reproducibility of fusion transcripts discovered from RNA-Seq data. Alternatively, the software systems to discover fusion transcripts from RNA-Seq data were essential to generate accurate and reproducible fusion transcript data. Since more than 20 software systems were available, scientific communities needed more standards to evaluate the qualities of these systems. In this study, we showed 22 ASL patients had frontal cortex, medial motor cortex, and lateral cortex and were more like three data duplications. It is important to note that we should avoid such a strategy to study epigenetic fusion genes. As shown in STable 2, we identified 29 HFGs encoding 38 HFG transcripts expressed in all three brain tissues. For example, ZNF528-ZNF880 and SNORD114-MEG8 were observed in the frontal cortex, medial motor cortex, and lateral cortex of 21 of 22 ASL patients and showed 100% reproducibility of fusion transcripts generated by SCIF. Hence, we recommended that scientists and bioinformaticians use these data as reference standards to select a suitable software system for one's research projects. Since our fusion transcripts could be traced to each RNA-Seq read, one could use fusion junction sequences provided in this study to search the original datasets to validate HFG fusion transcripts, most of which were reproducible except for fusion transcripts generated by DNA duplications.
Acknowledges

We appreciate the support from The Target ALS Human Postmortem Tissue Core, the New York Genome Center for Genomics of Neurodegenerative Disease, the Amyotrophic Lateral Sclerosis Association, and the Tow Foundation. We thank the support from the science and technology innovation Program of Hunan Province (Program #: 2021SK4014).

References


The First Insight into the Hereditary Fusion Gene Landscape of ASL


The First Insight into the Hereditary Fusion Gene Landscape of ASL


The First Insight into the Hereditary Fusion Gene Landscape of ASL
# Tables

## Table 1 Comparisons of the top sixteen HFGs between ASL and GTEx

<table>
<thead>
<tr>
<th>HFGs</th>
<th>5' Junction Sequence</th>
<th>3' Junction Sequence</th>
<th>ASL (77)</th>
<th>GTEx (133)</th>
<th>Folds</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZNF528-ZNFB80</td>
<td>CATCAAAGGTGTGAAACACAG</td>
<td>GACACCTGGCATTCCAGGGA</td>
<td>76</td>
<td>98.7</td>
<td>6</td>
<td>4.5</td>
</tr>
<tr>
<td>SNORD144-MEG8a</td>
<td>AGCCTGGACACTAAAGAA</td>
<td>GATAAGTCTAATTTCCTCC</td>
<td>70</td>
<td>90.9</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>TRNAV35-FAM91A1P</td>
<td>TAAAGACTGCTGTTTGGCAG</td>
<td>GCTCTATGTCATTCTCC</td>
<td>69</td>
<td>89.6</td>
<td>42</td>
<td>31.6</td>
</tr>
<tr>
<td>ADAMTS13-SH3GL3a</td>
<td>CTTATCGCTATGATGACAG</td>
<td>GTAGAACAGTCGCCAGTTTT</td>
<td>67</td>
<td>87.0</td>
<td>2</td>
<td>1.5</td>
</tr>
<tr>
<td>SNORD144-MEG8b</td>
<td>TGTGTACAGGAAAGACTCCTT</td>
<td>GATAAGTCTAATTTCCTCC</td>
<td>65</td>
<td>84.4</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>SDHAP2-SOHAP2a</td>
<td>AATATCTGGACTTGGAAAT</td>
<td>TGCAAGCAGGAGAACACA</td>
<td>60</td>
<td>77.9</td>
<td>7</td>
<td>4.5</td>
</tr>
<tr>
<td>YY1AP1-YY1AP1a</td>
<td>TACCCAAGGCTAGAACA</td>
<td>TTCCAGGATGAGTGAGATT</td>
<td>58</td>
<td>75.3</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>SDHAP2-SOHAP2b</td>
<td>TAGACTGGAATGCAAGACTG</td>
<td>TGCAAGCAGGAGAACACA</td>
<td>57</td>
<td>74.0</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>ERMP1-KIAA2026</td>
<td>GTTACCTGAGATGTTGAAC</td>
<td>TGAGCTTCTTCTTAAGTG</td>
<td>57</td>
<td>74.0</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>TRNAV35-SRGAP2P2</td>
<td>TGGAGACTGCTGTTTGGCAG</td>
<td>GCAAGTGAACGGAGAACAA</td>
<td>56</td>
<td>72.7</td>
<td>6</td>
<td>4.5</td>
</tr>
<tr>
<td>ADAMTS13-SH3GL3b</td>
<td>TGCACTCTCCCTCCGAGCAG</td>
<td>GTAGAACAGTCGCCAGTTTT</td>
<td>54</td>
<td>70.1</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>FMNL2-FMNL2</td>
<td>GGAAAGACATTGTGACCTGG</td>
<td>GAACGATTTCCAGTGAA</td>
<td>53</td>
<td>68.8</td>
<td>1</td>
<td>2.5</td>
</tr>
<tr>
<td>SMOX-RAD5L2</td>
<td>CAAAGCTGCTGAATTAACAG</td>
<td>CACCCCTGCAAGTGACAGT</td>
<td>50</td>
<td>64.9</td>
<td>4</td>
<td>3.0</td>
</tr>
<tr>
<td>PTPRS-PHACTR3</td>
<td>CCGAGTTCCTGATGACTACAG</td>
<td>ATGAGATGACAAAAAGCCC</td>
<td>47</td>
<td>61.0</td>
<td>16</td>
<td>12.0</td>
</tr>
<tr>
<td>FAM166B-NUCB2</td>
<td>CAGCTGCGCGCGCAGGGCG</td>
<td>TAGAGACAGGTTCACAGT</td>
<td>43</td>
<td>55.8</td>
<td>24</td>
<td>18.0</td>
</tr>
<tr>
<td>TMEFF1-RWDD2B</td>
<td>AAGAGACACTGCTGACTCTG</td>
<td>GCTGGAAGCTGCGCGCGA</td>
<td>42</td>
<td>54.5</td>
<td>3</td>
<td>2.3</td>
</tr>
<tr>
<td>TRNAV35-SEC22B</td>
<td>TGGAGACTGCTGTTTGGCAG</td>
<td>CGACCTCAAGGACAGTAAAA</td>
<td>41</td>
<td>53.2</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>SNORD144-MEG8c</td>
<td>GAGCTGGACACTAAAGAA</td>
<td>AATGGAGGCCACAGAGTGAA</td>
<td>41</td>
<td>53.2</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>YY1AP1-YY1AP1b</td>
<td>GAACTGGACAGACAGCATG</td>
<td>TGGTTCTGCACACCACTTAG</td>
<td>39</td>
<td>50.6</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>PTTPNM3-KIAA0753</td>
<td>GAAACTGGACAGGACATCAAG</td>
<td>TGGTTCTGCACACCACTTAG</td>
<td>39</td>
<td>50.6</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>SPECC1L-SPECC1L</td>
<td>CAAAATCCAGACTTACAGCAG</td>
<td>ATGGCATGTAATGGCACTCA</td>
<td>39</td>
<td>50.6</td>
<td>1</td>
<td>0.8</td>
</tr>
</tbody>
</table>

The First Insight into the Hereditary Fusion Gene Landscape of ASL
Table 2 Distributions of the top sixteen HFGs among OND, MTLE, and AD

<table>
<thead>
<tr>
<th>HFGs</th>
<th>ASL (77)</th>
<th>OND (7)</th>
<th>MTLE (17)</th>
<th>AD (9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZNF528-ZNF880</td>
<td>76</td>
<td>98.7</td>
<td>6</td>
<td>85.7</td>
</tr>
<tr>
<td>SNORD114_-MEG8_ a</td>
<td>70</td>
<td>90.9</td>
<td>5</td>
<td>71.4</td>
</tr>
<tr>
<td>TRNAN35-FAM91A3P</td>
<td>69</td>
<td>89.6</td>
<td>5</td>
<td>71.4</td>
</tr>
<tr>
<td>ADAMTSL3-SH3GL3 a</td>
<td>67</td>
<td>87.0</td>
<td>7</td>
<td>100.0</td>
</tr>
<tr>
<td>SNORD114_-MEG8_ b</td>
<td>65</td>
<td>84.4</td>
<td>7</td>
<td>100.0</td>
</tr>
<tr>
<td>SDHAP2-SDHAP2 a</td>
<td>60</td>
<td>77.9</td>
<td>4</td>
<td>57.1</td>
</tr>
<tr>
<td>YY1AP1-YY1AP1 a</td>
<td>58</td>
<td>75.3</td>
<td>6</td>
<td>85.7</td>
</tr>
<tr>
<td>SDHAP2-SDHAP2 b</td>
<td>57</td>
<td>74.0</td>
<td>4</td>
<td>57.1</td>
</tr>
<tr>
<td>ERMP1-KIAA2026</td>
<td>57</td>
<td>74.0</td>
<td>5</td>
<td>71.4</td>
</tr>
<tr>
<td>TRNAN35-SRGAP2P2</td>
<td>56</td>
<td>72.7</td>
<td>4</td>
<td>57.1</td>
</tr>
<tr>
<td>ADAMTSL3-SH3GL3 b</td>
<td>54</td>
<td>70.1</td>
<td>6</td>
<td>85.7</td>
</tr>
<tr>
<td>FMNL2-FMNL2</td>
<td>53</td>
<td>68.8</td>
<td>7</td>
<td>100.0</td>
</tr>
<tr>
<td>SMOX-RAD54L2</td>
<td>50</td>
<td>64.9</td>
<td>3</td>
<td>42.9</td>
</tr>
<tr>
<td>PTRRS-PHACTR3</td>
<td>47</td>
<td>61.0</td>
<td>5</td>
<td>71.4</td>
</tr>
<tr>
<td>FAM168B-NUCH2</td>
<td>43</td>
<td>55.8</td>
<td>3</td>
<td>42.9</td>
</tr>
<tr>
<td>TMEEF1-RWD22B</td>
<td>42</td>
<td>54.5</td>
<td>6</td>
<td>85.7</td>
</tr>
<tr>
<td>TRNAN35-SEC22B</td>
<td>41</td>
<td>53.2</td>
<td>5</td>
<td>71.4</td>
</tr>
<tr>
<td>SNORD114_-MEG8_ c</td>
<td>41</td>
<td>53.2</td>
<td>5</td>
<td>71.4</td>
</tr>
<tr>
<td>YY1AP1-YY1AP1 b</td>
<td>39</td>
<td>50.6</td>
<td>4</td>
<td>57.1</td>
</tr>
<tr>
<td>PITPNM3-KIAA0753</td>
<td>39</td>
<td>50.6</td>
<td>3</td>
<td>42.9</td>
</tr>
<tr>
<td>SPECCL1-SPECCL1</td>
<td>39</td>
<td>50.6</td>
<td>6</td>
<td>85.7</td>
</tr>
</tbody>
</table>

The First Insight into the Hereditary Fusion Gene Landscape of ASL
Table 3: *ADAMTSL3-SH3GL3* behaved like *SH3GL3-ADAMTSL3* and was a potential dormant HFG.

<table>
<thead>
<tr>
<th>Data</th>
<th><em>SH3GL3-ADAMTSL3</em> (%)</th>
<th><em>ADAMTSL3-SH3GL3</em> (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amyotrophic lateral sclerosis (77)</td>
<td>98.7</td>
<td>87.0</td>
</tr>
<tr>
<td>Non-neurological controls (12)</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Other neurological disorders (7)</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Mesial temporal lobe epilepsy (17)</td>
<td>94.1</td>
<td>82.4</td>
</tr>
<tr>
<td>Alzheimer's disease (9)</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Alzheimer's disease control (8)</td>
<td>100</td>
<td>87.5</td>
</tr>
<tr>
<td>GTEx brain cortex (133)</td>
<td>7.5</td>
<td>1.5</td>
</tr>
</tbody>
</table>
The First Insight into the Hereditary Fusion Gene Landscape of ASL

Figure Legends

**Fig. 1** Identification and characterization of hereditary fusion genes (HFGs) from RNA-Seq data of the non-neurological controls (NCC) and amyotrophic lateral sclerosis (ASL).

a). characteristics of the amyotrophic lateral sclerosis RNA-Seq dataset (GEO: GSE124439), including 77 amyotrophic lateral sclerosis (ASL) patients, seven patients with other neurological disorders (OND), and 12 non-neurological controls (NNC); b). The procedure of identifying HFGs from 12 NCC individuals; c). The total HFGs used to analyze HFG transcripts of amyotrophic lateral sclerosis, including previously-curated 1180 MZ HFGs and 204 Ctr HFGs; d). The components of HFGs to identify ASL HFGs.

**Fig. 2** Potential genomic mechanisms to generate ZNF528-ZNF880, TRNAN35-FAM91A3P, and SNORD114-MEG8. a). A potential local inversion to form ZNF528-ZNF880; b) A putative local translocation to produce TRNAN35-FAM91A3P; and c). A potential local inversion to generate SNORD114-MEG8. Black and red horizontal arrows indicated 5’ and 3’ genes of HFGs. Gray horizontal arrows were genes surrounding 5’ and 3’ genes. The vertical arrows showed potential genomic rearrangements.

**Fig. 3** A potential local inversion of SH3GL3-ADAMTSL3 to produce ADAMTSL3-SH3GL3. Two identical-stand neighboring SH3GL3 and ADAMTSL3 genes (a) produced an SH3GL3-ADAMTSL3 readthrough pre-mRNA (b), which was spliced to form three of thirteen SH3GL3-ADAMTSL3 EFG transcripts(c). SH3GL3-ADAMTSL3 was locally inverted or locally amplified to ADAMTSL3-SH3GL3 hereditary fusion gene (HFG) (d), transcribed into ADAMTSL3-SH3GL3 pre-mRNAs €, which was spliced into three examples of 42 ADAMTSL3-SH3GL3 HFG.
transcripts. Black and red horizontal solid arrows showed 5’ and 3’ genes of HFGs. Black and red rectangles were exons. Black triangles were introns. The double dashed lines were putative breakpoints. Horizontal dashed lines indicated omitted exons and introns. An open arch arrow indicated a local translocation or genomic amplification.
Supplemental Materials

NYGC ALS Consortium


Center for Genomics of Neurodegenerative Disease (CGND), New York Genome Center, New York, NY, USA. Department of Neurology, Lewis Katz School of Medicine, Temple University, Philadelphia, PA, USA. Cedars-Sinai Department of Biomedical Sciences, Board of Governors Regenerative Medicine Institute and Brain Program, Cedars-Sinai Medical Center, University of California, Los Angeles, CA, USA. Department of Medicine, University of California, Los Angeles, CA, USA. Department of Biochemistry and Molecular Biology, Penn State Institute for The First Insight into the Hereditary Fusion Gene Landscape of ASL.
Personalized Medicine, The Pennsylvania State University, Hershey, PA, USA. 10Department of Neurology, The Pennsylvania State University, Hershey, PA, USA. 11Department of Neurology, Henry Ford Health System, Detroit, MI, USA. 12Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. 13Department of Neurology, Center for Motor Neuron Biology and Disease, Institute for Genomic Medicine, Columbia University, New York, NY, USA. 1Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. 14Department of Neurology, Johns Hopkins School of Medicine, Baltimore, MD, USA. 15Department of Neurogenetics, Academic Medical Centre, Amsterdam, The Netherlands. 16Leiden University Medical Center, Leiden, The Netherlands. 17Department of Medicine, Lung Biology Center, University of California, San Francisco, CA, USA. 18ALS Multidisciplinary Clinic, Neuromuscular Division, Department of Neurology, Harvard Medical School, Boston, MA, USA. 19Neurological Clinical Research Institute, Massachusetts General Hospital, Boston, MA, USA. 19Centre for Neuroscience and Trauma, Blizard Institute, Barts, Queen Mary University of London, London, UK. 20The London School of Medicine and Dentistry, Queen Mary University of London, London, UK. 21Department of Neurology, Basildon University Hospital, Basildon, UK. 22Institute of Neurology, National Hospital for Neurology and Neurosurgery, University College London, London, UK. 23The Jackson Laboratory, Bar Harbor, ME, USA. 24Department of Psychiatry and Human Behavior, Department of Biological Chemistry, School of Medicine, University of California, Irvine, CA, USA. 25Department of Neurobiology and Behavior, School of Biological Sciences, University of California, Irvine, CA, USA. 26Taube/Koret Center for Neurodegenerative Disease Research, Roddenberry Center for Stem Cell Biology and Medicine, Gladstone Institute, San Francisco, CA, USA. 27Department of Neurology and Sensory Organs, University of Thessaly, Thessaly, Greece. 28Department of Neurology, The First Insight into the Hereditary Fusion Gene Landscape of ASL.
The First Insight into the Hereditary Fusion Gene Landscape of ASL
The First Insight into the Hereditary Fusion Gene Landscape of ASL Alzheimer's Disease, Icahn School of Medicine at Mount Sinai, New York, NY, USA. 46 Department of Neurology, Harvard Medical School, Neurological Clinical Research Institute, Massachusetts General Hospital, Boston, MA, USA. 47 Department of Neurology, Hospital for Special Surgery, New York, NY, USA. 48 Weill Cornell Medical Center, New York, NY, USA. 49 Medical Genetics, Atlantic Health System, Morristown Medical Center, Morristown, NJ, USA. 50 Overlook Medical Center, Summit, NJ, USA. 51 Center of Clinical Research, Experimental Surgery and Translational Research, Biomedical Research Foundation of the Academy of Athens (BRFAA), Athens, Greece. 52 1st Department of Neurology, Eginition Hospital, Medical School, National and Kapodistrian University of Athens, Athens, Greece. 53 Neuromuscular/EMG service and ALS/Motor Neuron Disease Clinic, Hebrew University-Hadassah Medical Center, Jerusalem, Israel. 54 Board of Governors Regenerative Medicine Institute, Los Angeles, CA, USA. 55 Department of Neurology, Cedars-Sinai Medical Center, Los Angeles, CA, USA. 56 Departments of Neuroscience, Genetics and Genomic Sciences, Ronald M. Loeb Center for Alzheimer's disease, Icahn School of Medicine at Mount Sinai, New York, NY, USA. 57 Harvard Medical School, Department of Physical Medicine and Rehabilitation, Spaulding Rehabilitation Hospital, Boston, MA, USA. 58 Center for Cellular and Molecular Therapeutics, Children's Hospital of Philadelphia, Philadelphia, PA, USA. 59 Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. 60 Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK. 61 Euan MacDonald Centre for Motor Neurone Disease Research, University of Edinburgh, Edinburgh, UK. 62 Department of Genetics and Genomic Sciences, Icahn Institute of Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. 63 Department of Neuropathology, Georgetown Brain Bank, Georgetown Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington, DC, USA. 64 Neuroradiology Section, Department of Radiology
Supplemental Table 1: Lists of hereditary fusion genes (HFGs) discovered from the non-neurological controls (NNC)

Supplemental Table 2: Lists of NNC HFGs formed via tandem gene duplications

The First Insight into the Hereditary Fusion Gene Landscape of ASL
Supplemental Table 3: Lists of HFGs that were discovered in amyotrophic lateral sclerosis and present in the frontal cortex, medial motor cortex, and lateral motor cortex.

Supplemental Table 4: Lists and recurrent frequencies of HFGs present in the frontal cortex, medial motor cortex, and lateral motor cortex of 22 ASL patients.

Supplemental Table 5: Comparative analysis of recurrent frequencies of HFGs between ASL patients and GTEx brain cortexes.

Supplemental Fig.1 Sequence alignment of conceptual ZNF528-ZNF880 fusion protein and ZNF880 protein. Dashed line indicated the sequence gap.

Supplemental Fig.2 Sequence alignment of conceptual FMNL2-FMNL2 fusion protein and formin like 2 protein. Dashed line indicated the sequence gap.

Supplemental Fig.3 Sequence alignment of conceptual YY1AP1-YY1AP1 fusion protein and YY1 associated protein 1. Dashed line indicated the sequence gap.
Fig. 1

a) Bar chart showing the distribution of NNC, OND, and ASL.

b) Workflow diagram explaining the process from Total Fusion Transcripts (TFs) to HFG Transcripts, detailing the steps of FTs From Genomic Alterations, FTs w/ ≥16% BLAST, and resulting in HFG Transcripts.

c) Bar chart comparing MZ HFGs and Ctr HFGs.

d) Venn diagram illustrating the comparison between MZ and NNC, with 205, 42, and 196 as the numbers in each section.
Fig. 3