

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; ² the 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 ALS. However, 90-95% of ALS cases have been found without a family history. Here, we have analyzed RNA-Seq data of NYGC ALS Consortium and identified fusion transcripts from ALS 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 ALS fusion transcripts and identified 348 HFGs. Comparative analysis between ALS and GTEx shows that 139 HFGs are associated with ALS and ranged from 10.4% to 98.7% of 77 ALS patients. The most recurrent HFG is *ZNF528-ZNF880*, detected in 98.7% of 77 ALS patients and 4.5% of 133 GTEx brain cortexes. Alignments of HFG transcripts from ALS with fusion transcripts from mesial temporal lobe epilepsy (MTLE) and Alzheimer's disease (AD) showed that 43.9% and 11.6% of the ALS HFGs were present in MTLE and AD, respectively. The most recurrent and common HFG among ALS, 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 ALS 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.

The First Insight into the Hereditary Fusion Gene Landscape of ALS

1
2
3 Keywords: Hereditary fusion gene, Amyotrophic Lateral Sclerosis, epilepsy, Alzheimer's disease, RNA-
4 Seq, Epigenetic fusion gene

5
6
7
8
9
10
11
12
13
14
15
16
17
18

19 **Introduction**

20 Amyotrophic lateral sclerosis (ALS) is a progressive, fatal neurodegenerative disease characterized by
21 progressive degenerative changes in upper and lower motor neurons that overlap with frontotemporal
22 lobar degeneration clinically, morphologically and genetically¹. ALS clinical symptoms include
23 weakness of the bulbar and limb muscles, hyperreflexia, spasticity of the arms and legs, and respiratory
24 failure^{2, 3}. These symptoms start with onset typically in the fifth or sixth decade of life, and death

The First Insight into the Hereditary Fusion Gene Landscape of ALS

1 commonly occurs within 2-4 years of diagnosis². ALS is closely related to frontotemporal dementia
2 (FTD), a progressive neurodegenerative disease characterized by degeneration of the frontal and
3 temporal lobes of the brain, resulting in behavior, personality, and language disturbances. Up to 15%
4 of ALS patients are also diagnosed with frontotemporal dementia (FTD-ALS patients)⁴. Many studies
5 have shown clinical, pathological, and genetic commonalities between them. Therefore, ALS and FTD
6 have now considered two manifestations of one disease continuum of the FTD-ALS spectrums⁵.

7
8 ALS is associated with over 30 mutated genes, including *SOD1*, *TARDBP*, *FUS*, *OPTN*, *VCP*,
9 *UBQLN2*, *C9ORF72*, and *PFN1*⁶⁻⁸. However, they are typically present in fewer than 5-10% of all ALS
10 patients, and most ALS cases remain unexplained^{9, 10}. Genome-wide associations (GWAS) have
11 genotyped numerous single nucleotide polymorphisms (SNPs) present in at least 1% of the human
12 population¹¹ and identified some common and rare SNPs associated with ALS across the genome¹²⁻¹⁴.
13 More recently, the discovery of genes associated with ALS has been shifted to next-generation
14 sequencing, such as whole-genome sequencing (WGS) and RNA-Seq. Many WGS studies have
15 confirmed previous findings and identified many common and rare variants associated with ALS^{15, 16}.
16 Brown et al. have performed RNA-Seq analysis on the ALS postmortem cortex samples of NYGC ALS
17 Consortium and found that TDP-43 loss resulted from retrotransposon and ALS-risk SNPs drive mis-
18 splicing and depletion of *UNC13A*^{17, 18}. Many studies have speculated that structural variants may be a
19 source of missing ALS heritability^{10, 19}. More recently, Raghav et al. analyzed RNA-Seq data from
20 Target ALS and the NYGC ALS Consortium and identified 607 unique fusion genes, some significantly
21 higher in ALS samples than those in control²⁰, suggesting that fusion genes generated by structural
22 variants may be responsible for ALS.

23

The First Insight into the Hereditary Fusion Gene Landscape of ALS

1 As discussed above, advances in RNA-Seq have made it possible to discover fusion transcripts
2 generated by structural variants. However, misconceptions and poor algorithms for repetitive
3 sequences have made fusion transcripts discovered incomplete, irreproducible, and unreliable²¹. In
4 order to overcome these problems, we have used the splicingcode theory to develop SCIF
5 (SplicingCodes Identify Fusion Genes) to discover more reproducible and accurate fusion transcripts
6 and make it possible to compare and analyze data among different experiments²². Validation of highly-
7 recurrent fusion genes has cast doubts that fusion genes are cancerous and generated from somatic
8 genomic abnormalities. *KANSARL* (*KANSL1-ARL17A*) is validated as the first predisposition
9 (hereditary) fusion genes specific to 29% of populations of European ancestry²². *TPM4-KLF2*,
10 detected in 92.2% of 727 multiple myeloma patients and initially thought to be a multiple myeloma-
11 specific somatic fusion gene, has been detected in all five healthy bone marrows²³. Based on their
12 genomic structures, genetics, and evolution, we have classified the fusion genes into somatic,
13 epigenetic, and hereditary²⁴. Hereditary fusion genes (HFGs) are the fusion genes that offspring inherit
14 from their parents, excluding readthrough transcripts of two identical-stand neighboring genes²⁵, named
15 epigenetic fusion genes (EFGs)²⁴. Furthermore, we have used 1180 HFGs identified from
16 monozygotic (MZ) twins to analyze fusion transcripts from multiple myeloma and acute myeloid
17 leukemia (AML) and discovered that HFGs are dominant inherited factors and mutated genes have
18 played minor roles in cancer inheritance^{26, 27}. EFGs have reflected developmentally interactive
19 consequences between genetic and environmental factors^{23, 27}.

20
21 In this study, we analyzed RNA-Seq data of NYGC ALS Consortium and identified large numbers of
22 fusion transcripts from RNA-Seq data of ALS patients. Surprisingly, we have observed large numbers
23 of highly-recurrent fusion genes from their non-neurological controls, 204 of which are determined to be
24 HFGs. We have used these newly-identified HFGs and the previously-curated 1180 HFGs to analyze

The First Insight into the Hereditary Fusion Gene Landscape of ALS

1 the fusion transcripts. From 358 HFGs, we have identified 139 HFGs associated with ASL ranging from
2 10.4% to 98.7% of ASL patients. We have discovered that many highly recurrent ASL HFGs have also
3 been detected in mesial temporal lobe epilepsy (MTLE) and Alzheimer's disease (AD), suggesting
4 common pathophysiological mechanisms²⁸. *ADAMTSL3-SH3GL3* HFG behaves like *SH3GL3-*
5 *ADAMTSL3* EFG, suggesting it has a much higher recurrent frequency than its GTEx brain cortex
6 counterpart. This study has shown that HFGs are the missing heritability of complex diseases and
7 traits and provides potential therapeutic targets for more efficient novel therapeutic drugs and methods
8 for ASL and other neurological diseases.

9

10 **Materials and Methods**

11 **Materials**

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

19 **Mesial temporal lobe epilepsy RNA-Seq data:** The mesial temporal lobe epilepsy (MTLE) RNA-Seq
20 data (GEO: GSE134697) was downloaded from NCBI
21 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA556159>). The MTLE RNA-Seq dataset contained the
22 hippocampus and neocortex of 17 MTLE patients and two healthy controls. This study used only RNA-
23 Seq data from the neocortex of 17 MTLE patients.

24

The First Insight into the Hereditary Fusion Gene Landscape of ASL

1 **Alzheimer's disease RNA-Seq data:** The Alzheimer's disease (AD) RNA-Seq data (GEO: GSE53697)
2 was downloaded from NCBI (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA232669>). The AD data
3 had 17 brain samples, including nine patients and eight controls.

4
5 **GTEX brain cortex RNA-Seq data:** the Genotype-Tissue Expression (GTEx) RNA-Seq data (dbGaP:
6 phs000424) was downloaded from NCBI (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA75899>). We
7 selected and performed RNA-Seq data from 133 brain cortexes of 133 unique individuals as the healthy
8 controls.

10 **Methods**

11 **Identification of fusion transcripts from RNA-Seq data:** We used SCIF (SplicingCodes Identify
12 Fusion Genes) to perform RNA-Seq analysis at the default conditions²². Since SCIF used raw RNA-
13 Seq reads without any assembly to discover the fusion junctions and followed by extended sequence
14 alignments, each RNA-Seq read was either a fusion transcript or not a fusion transcript. Human was
15 not involved in determining whether a sample with a fusion junction or a sample with three fusion
16 junctions was hereditary fusion gene-positive. However, one of the worst defects in discovering the
17 hereditary fusion genes from RNA-Seq data was that levels of gene expression, RNA-Seq qualities,
18 and the number of RNA-Seq reads would affect the analysis. Consequently, we must take caution to
19 handle and analyzing data.

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

$$Z = \frac{(\bar{p}_1 - \bar{p}_2) - 0}{\sqrt{\bar{p}(1 - \bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

24

The First Insight into the Hereditary Fusion Gene Landscape of ASL

1 \bar{p}_1 and \bar{p}_2 were the population standard errors of p_1 and p_2 .

2 $\bar{p}_1 = x_1 / n_1$; $\bar{p}_2 = x_2 / n_2$; $\bar{p} = (x_1 + x_2) / (n_1 + n_2)$.

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

5

6

7 **Results**

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

9 We downloaded the RNA-Seq data of NYGC ALS Consortium (GEO: GSE124439), which contained a
10 total of 176 frontal cortex/motor cortex, including 145 samples of ASL spectrum from 77 unique ASL
11 patients, 14 samples of other neurological disorders (named as OND to avoid confusions) from seven
12 unique individuals, and 17 non-neurological controls (NNC) from 12 unique individuals (Fig.1a). We
13 used SCIF to analyze RNA-Seq data of this dataset and identified 64,000 fusion transcripts from
14 samples of ASL and OND and 8,190 fusion transcripts from 12 non-neurological controls. Manual
15 inspections showed that fusion genes had been discovered at such high recurrent frequencies in non-
16 neurological samples that traditional logic and biological models cannot explain. For example,
17 *ADAMTSL3-SH3GL3*, an inversion of *SH3GL3-ADAMTSL3*, had been observed in all 12 non-
18 neurological controls. If structural variants generating a fusion gene per individual had a rate of $3.6 \times 10^{-29, 30}$,
19 the chance of all 12 NNC individuals having the identical *ADAMTSL3-SH3GL3* fusion gene
20 generated by random somatic genomic alterations was 4.7×10^{-18} and a mathematically impossible
21 event, suggesting it had to be inherited from their parents²⁴. Fig.1b shows the procedure to discover
22 hereditary fusion genes (HFGs) from NNC. If we set the cutoff of recurrent frequencies at two or more
23 in 12 NNC individuals, the minimum recurrent frequencies of 16.7% were at least 11.9 fold of 1.4% of
24 the random chance. We initially identified 257 putative HFG transcripts whose recurrent frequencies

The First Insight into the Hereditary Fusion Gene Landscape of ASL

1 ranged from 16.7% to 100%. After performing Blast analyses in Aceview and NCBI, we removed 24
2 EFG transcripts and potential alternative-spliced transcripts and obtained 233 HFG transcripts encoded
3 by 204 HFGs (STable 1). Out of 233 HFG transcripts, we discovered that 33 (16.2%) HFGs coding for
4 38 transcripts were from tandem gene duplications (Stable 2), suggesting that tandem gene
5 duplications were essential for genotype and phenotype diversities²⁴.

6

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

8 Next, we used these 204 HFGs discovered from NNC plus the previously curated 1180 MZ HFGs²⁴
9 (Fig.1c) to analyze the total fusion transcripts of 154 ASL and 14 OND samples. We identified 358
10 HFGs encoding 870 HFG transcripts, ranging from 0.6% to 98.1% of 159 ASL and OND samples, the
11 average of which was 2.4 per HFG, suggesting that many HFGs had multiple isoforms. Fig.1d showed
12 that 42 out of 358 HFGs were present in both NNC and MZ HFG datasets and counted for 21.4% of
13 NNC 196 HFGs, supporting our rationales of discovering novel HFGs from NNC. Further analysis
14 showed that only an additional 17 of 196 HFGs were present in GTEx blood samples (data not shown)
15 and suggested that most 196 HFGs were potentially differentially expressed and not expressed in blood
16 and other samples. This differential HFG expression potentially masked the accuracy of the HFG
17 frequencies detected in blood samples. Since these HFG transcripts were from the frontal cortex,
18 medial motor cortex, and lateral motor cortex of ASL and OND samples, we must study whether these
19 HFGs were differentially expressed among the frontal cortex and medial and lateral motor cortexes and
20 affected the results.

21

22 To understand whether HFGs were differentially expressed and affected HFG results analyzed
23 afterward, we performed comparative analyses of HFGs among the frontal cortexes, medial motor
24 cortex, and lateral motor cortex. After inspecting the HFG distributions of the frontal cortexes, medial

The First Insight into the Hereditary Fusion Gene Landscape of ASL

1 motor cortex, and lateral motor cortexes, we found no HFGs with high recurrent frequencies were
2 differentially expressed among these three brain tissues. STable 3 showed that 179 HFGs were
3 encoding 268 HFG transcripts in all frontal cortex, medial motor cortex, and lateral cortex. Majorities of
4 differences in the HFG transcripts among the three groups were small and statistically insignificant. We
5 identified 22 ASL patients with data from all frontal cortex, medial motor cortex, and lateral motor
6 cortex, similar to 22 three-duplication experiments. We performed a comparative data analysis of these
7 22 ASL patients. STable 4 showed that we identified 29 HFGs encoding 38 HFG transcripts
8 expressed in all frontal cortex, medial motor cortex, and lateral motor cortex and suggested that these
9 HFGs discovered by SCIF were highly reproducible. For example, *ZNF528-ZNF880* and *SNORD114_-*
10 *MEG8_* a were observed in all three tissues of 21 of 22 ASL patients and showed 100% reproducibility.
11 Here, *SNORD114_* was an abbreviated *SNORD114-1-SNORD114-2-SNORD114-3²⁵* while *MEG8_*
12 was *MEG8-SNORD112-SNORD113-3²⁵*. Careful inspections showed that some HFGs were
13 alternatively spliced. For example, the frequency difference of the main *SNORD114_-MEG8_* isoform
14 a was within statistical errors among the frontal cortex, medial motor cortex, and lateral cortex.
15 However, the recurrent frequencies of the *SNORD114_-MEG8_* isoform c in the medial motor cortex
16 and lateral motor cortexes were 16.4% and 13.7% higher than their counterpart in the frontal cortex.
17 They reflected alternatively splicing in the frontal, medial, and lateral motor cortexes. Since we
18 generally used the recurrent frequency of the main HFG isoform as the HFG recurrent frequency for
19 genotype analysis, we could combine HFG data from the frontal cortex and medial and lateral motor
20 cortexes to perform further analysis.

21

22 **Identification and characterization of HFGs associated with ASL**

23 We used the combined HFG dataset described above to analyze the total fusion transcripts from 77
24 ASL patients. We identified 346 HFGs encoding 819 HFG transcripts whose recurrent frequencies

The First Insight into the Hereditary Fusion Gene Landscape of ASL

1 ranged from 1.3% to 98.7%. 345 (42.1%) HFG transcripts were detected in only one ASL patient,
2 suggesting they were either causal HFGs or lowly-expressed HFG transcripts. For a more convenient
3 analysis, we set the cutoff of the recurrent frequency at 10%. Table 5 showed that we observed 147
4 HFGs coding for 190 HFG transcripts ranging from 10.4% to 98.7%. Table 1 showed that the top
5 sixteen HFGs encoding twenty-one HFG transcripts with recurrent frequencies of $\geq 50\%$ ranged from
6 51.3% to 98.7%. Fig.2a showed that the most recurrent *ZNF528-ZNF880*, an inversion of *ZNF880-*
7 *ZNF528* on the plus strand of chromosome 19q13.41, was detected in 76 (98.7%) of 77 ASL patients
8 and encoded a truncated zinc finger protein 880 without 35 aa N-terminus sequences (S Fig.1). Fig.2b
9 showed that *SNORD114_-MEG8_* was one of the most highly-expressed HFGs and coded for 20
10 isoforms, three of which were 92.1%, 85.5%, and 53.9% (Table 1). When we performed a comparative
11 analysis between ASL and NCC, many HFG differences were insignificant. For example, *ERMP-*
12 *KIAA2026*, reported in stomach cancer³¹, was detected in 75% of 75 ASL patients and 12 NNC
13 samples. As described previously, if structural variants generating a fusion gene per individual had a
14 rate of 3.6×10^{-2} ^{29, 30}, the chance of having nine NNC individuals who possessed *ERMP-KIAA2026*³¹
15 by random somatic genomic abnormalities was 1.0×10^{-13} and a mathematically-impossible event. It
16 was most likely that NNC was extraordinarily distorted for some unexplained reasons and unsuitable to
17 use as a standard control.

18 To overcome the NNC distortion, we selected RNA-Seq data from 133 GTEx brain cortexes as healthy
19 controls. We used 346 HFG from 77 ASL patients to analyze 26,000 fusion transcripts discovered from
20 the GTEx brain cortexes and identified 165 HFGs coding for 188 HFG transcripts from GTEx brain
21 cortexes, suggesting that 47.7% of 346 ASL HFGs were present in the GTEx brain cortexes. The most
22 recurrent GTEx HFGs were *KANSARL (KANSL1-ARL17A)*²² and *TRNAN35-FAM91A3P* (Fig.2c) HFGs
23 were detected in 34.6% and 31.6% of 133 GTEx brain cortexes. In comparison, 56.6% of the ASL

1 HFGs with recurrent frequencies of $\geq 10\%$ overlapped with HFGs of GTEx brain cortexes, suggesting
2 that the more recurrent ASL HFGs were, the more likely they were present in GTEx brain cortexes.
3 Next, we performed a comparative analysis of the HFG differences between ASL and GTEx. We
4 discovered that 139 out of 147 ASL HFGs ranged from 10.4% to 98.7% and were 1.2-244 folds higher
5 than their GTEx brain cortex counterparts, which were statistically significant (STable 5). Table 1
6 showed that sixteen HFGs were encoding twenty-one HFG transcripts, whose recurrent frequencies
7 ranged from 51.3% to 98.7% and were 1.8 to 120 folds of the GTEx brain cortex counterparts, which
8 were statistically significant by the *Z*-test. Table 1 showed that *ZNF528-ZNF880* (Fig.2a) was the most
9 recurrent HFG detected in 98.7% of 77 ASL patients. In comparison, it was detected in 4.5% of 133
10 GTEx brain cortexes. The former was 21 folds of the latter, suggesting that *ZNF528-ZNF880* HFG was
11 necessary for ASL initiation and prognosis. The most significant difference of HFG recurrent
12 frequencies between ASL and GTEx was *SNORD114_-MEG8_* (Fig.2b), a local inversion of
13 *SNORD114-1-SNORD114-2-SNORD114-3* EFG²⁵ and *MEG8-SNORD112-SNORD113-3* EFG²⁵,
14 resulted in potentially more diverse expression of snoRNAs. *TRNAN35-FAM91A3P* was the HFG with
15 the least frequency difference between ASL and GTEx. It was an inversion of *TRNAN35*, encoding
16 transfer RNA asparagine 35 and *FAM91A3P*, coding for the family with sequence similarity 91 member
17 A3 pseudogene (Fig.2c), detected in 31.6% of GTEx brain cortexes.
18 Table 1 showed that *TRNAN35-FAM91A3P*, *TRNAN35-SRGAP2P2*, and *TRNAN35-SEC22B* were
19 three *TRNAN35*-fused HFGs coded for putative 124 a.a, 177 a.a, 171 a.a, and 67 a.a polypeptides,
20 which were potentially regulated by *TRNAN35* via asparaginyl-tRNA synthetase and formed a natural
21 network²⁴. Four HFGs, including *SDHAP2-SDHAP2*, *YY1AP1-YY1AP1* (S Fig.2), *FMNL2-FMNL2*
22 (S Fig.3), and *SPECC1L-SPECC1L*, were shown to be generated via gene tandem duplications,
23 suggesting that tandem gene duplications were more widespread than what was reported previously³².
24 *FMNL2-FMNL2* encoded an FMNL2 protein with two exon insertions. *FMNL2* has been shown to

1 regulate gliovascular interactions and is associated with vascular risk factors and cerebrovascular
2 pathology in Alzheimer's disease³³. *YY1AP1-YY1AP1* encoded a truncated YY1-associated protein 1.
3 A genome-wide sequence showed that a frameshift mutation and a large deletion in *YY1AP1* resulted
4 in a girl with a panvascular artery disease³⁴. *SPECC1L-SPECC1L* seemed to encode intact SPECC1L
5 protein. It had been shown that mutated *SPECC1L* caused congenital diaphragmatic hernia -- a
6 prominent feature of a *SPECC1L*-related syndrome³⁵.

7
8 **Identification of common HFGs among ASL, Alzheimer's disease (AD), and mesial temporal lobe**
9 **epilepsy (MTLE)**

10 To investigate whether the OND had different HFGs, fusion transcripts identified from 14 OND samples
11 of seven patients were aligned with ASL HFG transcripts. We identified 206 HFGs encoding 318 HFG
12 transcripts whose recurrent frequencies ranged from 14.3% to 100%, suggesting that 59.5% of 346
13 ASL HFGs were also found in seven OND patients, many of which had much higher recurrent
14 frequencies in OND than in ASL. *SNORD114_-MEG8_*, *FMNL2-FMNL2*, and *ADAMTSL3-SH3GL3*
15 were the three top HFGs detected in all seven (100%) OND, while they were observed in 90.1%,
16 68.8%, and 87% of 77 ASL patients. Many other HFGs associated with ASL were also observed in
17 OND at similar or even higher recurrent frequencies than their ASL counterparts. 59.5% of ASL HFGs
18 observed in OND raised interesting questions about whether other neurological diseases, such as
19 temporal lobe epilepsy (MTLE) and Alzheimer's disease (AD), had common HFGs.

20 We used 819 HFG transcripts to analyze fusion transcripts of nine advanced AD patients³⁶ and 17
21 neocortex samples of seventeen MTLE patients produced for a 2019 failed NIH grant application. We
22 discovered that 152 HFGs coding for 214 HFG transcripts overlapped between ASL and MTLE and
23 counted for 43.9% of 346 HFGs discovered in ASL, suggesting that ASL and MTLE had significantly
24 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

1 transcripts from nine AD patients, we identified only 40 HFGs encoding 74 HFG transcripts, accounting
2 for only 11.6% of 346 HFGs observed in ASL and suggested that AD was much more divergent from
3 ASL than MTLE.

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

10 Table 2 also showed that only half of the top sixteen highly recurrent ASL HFGs were detected in AD
11 patients, while another half was absent in AD patients. The most notable HFGs absent in AD patients
12 were *YY1AP1-YY1AP1*, *FMNL2-FMNL2*, and *SPECC1L-SPECC1L* generated by tandem duplications.
13 The most critical HFGs shared between ASL and AD were *ZNF528-ZNF880* and *ADAMTSL3-SH3GL3*,
14 detected in 98.7% and 87% of 77 ASL patients, and 88.9%, and 100% of nine AD patients,
15 respectively. These highly recurrent HFGs were present in ASL, MTLE, and AD and supported that
16 many neurological diseases might have common pathophysiological mechanisms²⁸.

17

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

19 *ADAMTSL3-SH3GL3*, reported recently^{20, 37}, was an inversion of *SH3GL3-ADAMTSL3*³⁸ on 15q25,
20 encoding thirteen isoforms and coded for 42 alternatively-spliced isoforms (Fig.3). Its *ADAMTSL3-*
21 *SH3GL3* main isoform a was expressed only at 19% of the *KANSARL (KANASL1-ARL17A)* main
22 isoform level and encoded a conceptual 202 aa *ADAMTSL3-SH3GL3* fusion protein over two-thirds of
23 which was from *SH3GL3*. Since *SH3GL3* was most expressed only in the brain and testis while
24 *ADAMTSL3* was expressed at a minimum level in the brain but at high levels in other tissues,

The First Insight into the Hereditary Fusion Gene Landscape of ASL

1 *ADAMTSL3-SH3GL3* may alter *SH3GL3* expression patterns and regulations. Table 3 showed that
2 *ADAMTSL3-SH3GL3* was also detected in ASL, OND, AD, and MTLE and ranged from 82.4% to 100%.
3 Surprisingly, *ADAMTSL3-SH3GL3* was also detected in 100% of 12 NNC samples and 87.5% of eight
4 AD-healthy controls, significantly higher than 1.5% of 133 GTEx brain cortexes (Table 3). To address
5 these enormous gaps among these controls, we used *SH3GL3-ADAMTSL3*, recently reported³⁸, as an
6 example to compare both HFG expression patterns and behaviors. Table 3 showed that *SH3GL3-*
7 *ADAMTSL3*, coding for an intact ADAMTS-like 3 protein, was ubiquitously expressed in all datasets
8 analyzed except for GTEx brain cortexes, in 7.5% of which it was detected. Since the *SH3GL3-*
9 *ADAMTSL3* locus was present in almost all human individuals, its expression might have been induced
10 due to interactions between environmental and genetic lesions during aging. Hence, we speculated
11 that the *ADAMTSL3-SH3GL3* frequency in human populations was higher than 1.5%. Most
12 *ADAMTSL3-SH3GL3* was dormant or differentially expressed at younger ages and activated by
13 interactions between environmental and genetic lesions during aging. We named this dormant or
14 differentially expressed fusion gene inherited from their parents the dormant or differentially-expressed
15 HFG (dHFG). Hence, *ADAMTSL3-SH3GL3* was the first potential dHFG. Table 3 showed that
16 *SH3GL3-ADAMTSL3* and *ADAMTSL3-SH3GL3* were early-stage aging and neurological disease
17 biomarkers.

18

19 **Discussion**

20 We analyzed the RNA-Seq data of the NYGC ALS Consortium and identified fusion transcripts from
21 ASL, OND, and NNC. Many NNC fusion transcripts were highly recurrent. For example, *ADAMTSL3-*
22 *SH3GL3* was detected in 100% of 12 NNC samples and is consistent with our previous finding that
23 *ADAMTSL3-SH3GL3* was detected in 87.5% of eight AD-healthy controls³⁶. If structural variants
24 generating a fusion gene per individual had a rate of 3.6×10^{-2} ^{29, 30}, generating seven and twelve

The First Insight into the Hereditary Fusion Gene Landscape of ASL

1 *ADAMTSL3-SH3GL3* simultaneously in a dataset was $7.8\%10^{-11}$ and $4.7\%10^{-18}$. Such low chances
2 from two different sources^{18, 36} made it mathematically impossible for *ADAMTSL3-SH3GL3* to be
3 generated via random somatic genomic abnormalities. These observations supported that fusion
4 transcripts generated via genomic alterations in healthy controls were potential HFGs. Hence, we
5 developed a novel approach using fusion transcripts of NNC to discover HFGs. We used a
6 mathematical model and blast against AceView and NCBI and identified 204 HFGs encoding 233 HFG
7 transcripts. Bioinformatic analysis showed that 33 (16.2%) HFGs were from tandem gene duplications.
8 As we described previously, SCIF was not intended to discover fusion transcripts generated via tandem
9 duplications, and their discovery was due to gene duplications and pseudogenes²⁴. Hence, HFGs
10 produced via genomic tandem duplications were essential biomarkers for diseases and complex traits
11 and might be significantly underestimated.

12
13 We combined 204 HFGs discovered from NNC and the previously-curated 1180 MZ HFGs into the
14 HFG dataset to analyze the fusion transcripts of 154 RNA-Seq data from 77 ASL patients. We
15 discovered 346 HFGs encoding 819 HFG transcripts from ASL patients, whose recurrent frequencies
16 ranged from 1 (1.3%) to 76 (98.7) of 77 ASL patients. Out of 819 HFG transcripts, 345 (42.1%) HFG
17 transcripts were detected in only one of 77 ASL patients and were shown to be causal HFGs and lowly-
18 expressed HFG transcripts. When a cutoff of recurrent frequency was set at 10%, we identified 147
19 HFGs coding for 188 HFG transcripts ranging from 10.4% to 98.7%. As described above, HFGs
20 discovered from 12 NNC samples were extremely recurrent, some were detected in all 12 NNC
21 samples, and the NNC genotypes were distorted. So, we have yet to discover any reasons which can
22 explain these HFG genotype distortions. Hence, we selected 133 GTEx brain cortexes as healthy
23 controls. Table 1 showed that 139 HFGs were 1.2-121 folds higher than their GTEx brain cortex
24 counterparts and were shown by statistical analysis to be associated with ASL. Out of 139 HFGs

The First Insight into the Hereditary Fusion Gene Landscape of ASL

1 associated with ASL, Table 2 showed sixteen HFGs with recurrent frequencies of $\geq 50\%$, some of which
2 were also highly recurrent in OND, AD, and MTLE. 139 HFGs associated with ASL were four folds of
3 more than 30 mutated genes associated with ASL. The HFG numbers and their recurrent frequencies
4 suggested that HFGs were much more critical "inherited" genetic factors than the mutated genes⁶⁻⁸.
5 Hence, we can conclude that highly recurrent HFGs are the missing ASL heritability. More importantly,
6 many neurological diseases such as ASL, MTLE, and AD shared highly recurrent HFGs and supported
7 that these neurological diseases had common pathophysiological mechanisms²⁸. They potentially
8 allowed us to develop more effective novel therapeutic drugs and methods to treat and cure
9 neurological diseases.

10
11 One of the most highly-recurrent HFGs was *ADAMTSL3-SH3GL3*, an inversion of *SH3GL3-ADAMTSL3*
12 EFG (Fig.3). *ADAMTSL3-SH3GL3* HFG was detected in at least 82.4% of ASL, OND, AD, and MTLE
13 samples. More surprisingly, it was detected in 100% and 87.5% of 12 NNC samples and nine AD
14 controls. Even though OND, AD, and TMLE data sizes are relatively small, the divergent sources of
15 these data compensated for shortcomings. These data suggested that *ADAMTSL3-SH3GL3* would be
16 more recurrent among general populations than 1.5% of GTEx brain cortexes. To address this
17 apparent disparity, we could take a lesson from *SH3GL3-ADAMTSL3* EFG, the ancestors of
18 *ADAMTSL3-SH3GL3*. Table 3 showed that *SH3GL3-ADAMTSL3* was almost ubiquitously expressed in
19 ASL, OND, AD, MTLE, and two controls. In comparison, it was detected in 7.5% of 133 GTEx brain
20 cortexes and suggested that *SH3GL3-ADAMTSL3*³⁸ and *ADAMTSL3-SH3GL3*^{20, 37} behaved similarly.
21 Since *SH3GL3-ADAMTSL3* EFG can be treated as a permanent HFG with a recurrent frequency of
22 100%, we speculated that *ADAMTSL3-SH3GL3* had a much higher recurrent frequency than 1.5% of
23 GTEx brain cortexes and was activated to be expressed by interactions between genetic and
24 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

1 expressed HFGs. We needed genotyping general populations to validate if *ADAMTSL3-SH3GL3* was
2 a dHFG or if other factors caused the disparity between neurological disease-related data and GTEx
3 brain cortexes.

4
5
6 To study whether hereditary fusion genes were associated with ASL, AD, MTLE, and other neurological
7 diseases, one of the most critical issues was the accuracy and reproducibility of fusion transcripts
8 discovered from RNA-Seq data. Alternatively, the software systems to discover fusion transcripts from
9 RNA-Seq data were essential to generate accurate and reproducible fusion transcript data. Since more
10 than 20 software systems were available, scientific communities needed more standards to evaluate
11 the qualities of these systems ³⁹. In this study, we showed 22 ASL patients had frontal cortex, medial
12 motor cortex, and lateral cortex and were more like three data duplications. It is important to note that
13 we should avoid such a strategy to study epigenetic fusion genes. As shown in STable 2, we identified
14 29 HFGs encoding 38 HFG transcripts expressed in all three brain tissues. For example, *ZNF528-*
15 *ZNF880* and *SNORD114_-MEG8_* were observed in the frontal cortex, medial motor cortex, and lateral
16 cortex of 21 of 22 ASL patients and showed 100% reproducibility of fusion transcripts generated by
17 SCIF. Hence, we recommended that scientists and bioinformaticians use these data as reference
18 standards to select a suitable software system for one's research projects. Since our fusion transcripts
19 could be traced to each RNA-Seq read, one could use fusion junction sequences provided in this study
20 to search the original datasets to validate HFG fusion transcripts, most of which were reproducible
21 except for fusion transcripts generated by DNA duplications.

22
23
24

The First Insight into the Hereditary Fusion Gene Landscape of ASL

1

2

3

4

5

6

7

8

9 **Acknowledges**

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

14

15

16

17

18

19

20 **References**

21 1. Siddique N, Siddique T. Amyotrophic Lateral Sclerosis Overview. In: Adam MP, Everman DB,
22 Mirzaa GM, et al., eds. GeneReviews((R)). Seattle (WA)1993.

23 2. Chio A, Logroscino G, Hardiman O, et al. Prognostic factors in ALS: A critical review.
24 Amyotroph Lateral Scler 2009;10:310-323.

The First Insight into the Hereditary Fusion Gene Landscape of ASL

- 1 3. Masrori P, Van Damme P. Amyotrophic lateral sclerosis: a clinical review. *Eur J Neurol*
2 2020;27:1918-1929.
- 3 4. Nguyen HP, Van Broeckhoven C, van der Zee J. ALS Genes in the Genomic Era and their
4 Implications for FTD. *Trends Genet* 2018;34:404-423.
- 5 5. Parobkova E, Matej R. Amyotrophic Lateral Sclerosis and Frontotemporal Lobar Degenerations:
6 Similarities in Genetic Background. *Diagnostics* 2021;11.
- 7 6. Renton AE, Chio A, Traynor BJ. State of play in amyotrophic lateral sclerosis genetics. *Nature*
8 *neuroscience* 2014;17:17-23.
- 9 7. Alsultan AA, Waller R, Heath PR, Kirby J. The genetics of amyotrophic lateral sclerosis:
10 current insights. *Degenerative neurological and neuromuscular disease* 2016;6:49-64.
- 11 8. Suzuki N, Nishiyama A, Warita H, Aoki M. Genetics of amyotrophic lateral sclerosis: seeking
12 therapeutic targets in the era of gene therapy. *J Hum Genet* 2023;68:131-152.
- 13 9. Chia R, Chio A, Traynor BJ. Novel genes associated with amyotrophic lateral sclerosis:
14 diagnostic and clinical implications. *The Lancet Neurology* 2018;17:94-102.
- 15 10. Mejzini R, Flynn LL, Pitout IL, Fletcher S, Wilton SD, Akkari PA. ALS Genetics, Mechanisms,
16 and Therapeutics: Where Are We Now? *Front Neurosci* 2019;13:1310.
- 17 11. International HapMap C, Frazer KA, Ballinger DG, et al. A second generation human haplotype
18 map of over 3.1 million SNPs. *Nature* 2007;449:851-861.
- 19 12. Benyamin B, He J, Zhao Q, et al. Cross-ethnic meta-analysis identifies association of the GPX3-
20 TNIP1 locus with amyotrophic lateral sclerosis. *Nature communications* 2017;8:611.

- 1 13. van Es MA, Veldink JH, Saris CG, et al. Genome-wide association study identifies 19p13.3
2 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. *Nat Genet*
3 2009;41:1083-1087.
- 4 14. Laaksovirta H, Peuralinna T, Schymick JC, et al. Chromosome 9p21 in amyotrophic lateral
5 sclerosis in Finland: a genome-wide association study. *The Lancet Neurology* 2010;9:978-985.
- 6 15. Udine E, Jain A, van Blitterswijk M. Advances in sequencing technologies for amyotrophic
7 lateral sclerosis research. *Molecular neurodegeneration* 2023;18:4.
- 8 16. van Rheenen W, van der Spek RAA, Bakker MK, et al. Common and rare variant association
9 analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and
10 neuron-specific biology. *Nat Genet* 2021;53:1636-1648.
- 11 17. Tam OH, Rozhkov NV, Shaw R, et al. Postmortem Cortex Samples Identify Distinct Molecular
12 Subtypes of ALS: Retrotransposon Activation, Oxidative Stress, and Activated Glia. *Cell reports*
13 2019;29:1164-1177 e1165.
- 14 18. Brown AL, Wilkins OG, Keuss MJ, et al. TDP-43 loss and ALS-risk SNPs drive mis-splicing
15 and depletion of UNC13A. *Nature* 2022;603:131-137.
- 16 19. Theunissen F, Flynn LL, Anderton RS, et al. Structural Variants May Be a Source of Missing
17 Heritability in sALS. *Front Neurosci* 2020;14:47.
- 18 20. Raghav Y, Dilliot AA, Petrozziello T, et al. Identification of gene fusions associated with
19 amyotrophic lateral sclerosis. *medRxiv* 2022:2022.2006.2004.22275962.
- 20 21. Conesa A, Madrigal P, Tarazona S, et al. A survey of best practices for RNA-seq data analysis.
21 *Genome Biol* 2016;17:13.

- 1 22. Zhou JX, Yang X, Ning S, et al. Identification of KANSARL as the first cancer predisposition
2 fusion gene specific to the population of European ancestry origin. *Oncotarget* 2017;8:50594-50607.
- 3 23. Fei L, Zhuo N, Zhuo D. The Hereditary and Epigenetic Fusion Gene Signatures of Multiple
4 Myeloma. *bioRxiv* 2022:2022.2012.2005.519213.
- 5 24. Zhuo D. The First Glimpse of *Homo sapiens* Hereditary Fusion Genes. *Journal of*
6 *Pharmacogenomics & Pharmacoproteomics* 2022;13:1-7.
- 7 25. Thierry-Mieg D, Thierry-Mieg J. AceView: a comprehensive cDNA-supported gene and
8 transcripts annotation. *Genome Biol* 2006;7 Suppl 1:S12 11-14.
- 9 26. Ling F, Zhuo D. Hereditary Fusion Genes Are Associated with the Inheritance of Acute Myeloid
10 Leukemia. *bioRxiv* 2022:2022.2010.2031.514615.
- 11 27. Ling F, Zhuo N, Zhuo D. The First Insight into the Epigenetic Fusion Gene Landscape of Acute
12 Myeloid Leukemia. *bioRxiv* 2022:2022.2012.2006.519396.
- 13 28. Boison D, Aronica E. Comorbidities in Neurology: Is adenosine the common link?
14 *Neuropharmacology* 2015;97:18-34.
- 15 29. Conrad DF, Pinto D, Redon R, et al. Origins and functional impact of copy number variation in
16 the human genome. *Nature* 2010;464:704-712.
- 17 30. Campbell CD, Eichler EE. Properties and rates of germline mutations in humans. *Trends Genet*
18 2013;29:575-584.
- 19 31. Group PTC, Calabrese C, Davidson NR, et al. Genomic basis for RNA alterations in cancer.
20 *Nature* 2020;578:129-136.
- 21 32. Iafrate AJ, Feuk L, Rivera MN, et al. Detection of large-scale variation in the human genome.
22 *Nat Genet* 2004;36:949-951.

The First Insight into the Hereditary Fusion Gene Landscape of ASL

- 1 33. Lee AJ, Raghavan NS, Bhattarai P, et al. FMNL2 regulates gliovascular interactions and is
2 associated with vascular risk factors and cerebrovascular pathology in Alzheimer's disease. *Acta*
3 *Neuropathol* 2022;144:59-79.
- 4 34. Raggio V, Dell'Oca N, Simoes C, et al. Whole genome sequencing reveals a frameshift mutation
5 and a large deletion in YY1AP1 in a girl with a panvascular artery disease. *Hum Genomics* 2021;15:28.
- 6 35. Wild KT, Gordon T, Bhoj EJ, et al. Congenital diaphragmatic hernia as a prominent feature of a
7 SPECC1L-related syndrome. *Am J Med Genet A* 2020;182:2919-2925.
- 8 36. Scheckel C, Drapeau E, Frias MA, et al. Regulatory consequences of neuronal ELAV-like
9 protein binding to coding and non-coding RNAs in human brain. *Elife* 2016;5.
- 10 37. Azatyan A, Zaphiropoulos PG. Circular and Fusion RNAs in Medulloblastoma Development.
11 *Cancers* 2022;14.
- 12 38. Singh S, Qin F, Kumar S, et al. The landscape of chimeric RNAs in non-diseased tissues and
13 cells. *Nucleic Acids Res* 2020;48:1764-1778.
- 14 39. Simoneau J, Dumontier S, Gosselin R, Scott MS. Current RNA-seq methodology reporting limits
15 reproducibility. *Briefings in bioinformatics* 2021;22:140-145.
- 16
- 17

1 **Tables**

2

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

HFGs	5' Junction Sequence	3' Junction Sequence	ASL (77)		GTEx (133)		Folds	p
			#	%	#	%		
<i>ZNF528-ZNF880</i>	CATCAAAGGTGTGAACACAG	GGACACTTGGCATTGAGGGA	76	98.7	6	4.5	20.9	***
<i>SNORD114_-MEG8_a</i>	GAGCCTGGACACTAATAGAA	GATAAGTCTACTTTCTTCC	70	90.9	1	0.8	119.9	***
<i>TRNAN35-FAM91A3P</i>	TGAGGACTGCCTGTTTGACAG	GCTCTGATTGACTTTCTCC	69	89.6	42	31.6	1.8	***
<i>ADAMTSL3-SH3GL3_a</i>	CTTATCGCTATGATGACCAG	GTAGAACAAGTCAGCCAGTT	67	87.0	2	1.5	56.9	***
<i>SNORD114_-MEG8_b</i>	TGTGTACAGGAAGACTCTT	GATAAGTCTACTTTCTTCC	65	84.4	1	0.8	111.3	***
<i>SDHAP2-SDHAP2_a</i>	AATATCTGGACTTTGAAATT	TGCACAGCCAGAGAACAACA	60	77.9	1	0.8	102.6	***
<i>YY1AP1-YY1AP1_a</i>	TTCACCAAGGCTGAGGACAA	TTCCAAGATGAGATGGGATT	58	75.3	1	0.8	99.2	***
<i>SDHAP2-SDHAP2_b</i>	TAGACTGGATGCAGAGCATG	TGCACAGCCAGAGAACAACA	57	74.0	1	0.8	97.5	***
<i>ERMP1-KIAA2026</i>	GTTACGTGGATGTTTGTAAAC	TGAGCTTTCTCTTAAGTGTG	57	74.0	1	0.8	97.5	***
<i>TRNAN35-SRGAP2P2</i>	TGAGGACTGCCTGTTTGACAG	GCAAGTGAACCGGAGCAAAC	56	72.7	6	4.5	15.1	***
<i>ADAMTSL3-SH3GL3_b</i>	TGCACTCTCCCTCCCGCAG	GTAGAACAAGTCAGCCAGTT	54	70.1	1	0.8	92.3	***
<i>FMNL2-FMNL2</i>	GGAAGACATTCTGCACTGCG	GAACGATTCCAGGTGAAGAA	53	68.8	2	1.5	44.8	***
<i>SMOX-RAD54L2</i>	CCAAAGTGCTGGAATTACAG	CACCCCTGCAGTGGACCATG	50	64.9	4	3.0	20.6	***
<i>PTPRS-PHACTR3</i>	CCGAGTGGCTGGGACTACAG	ATGAGATGGACCAAACGCC	47	61.0	16	12.0	4.1	***
<i>FAM168B-NUCB2</i>	CAGGTCGCCGCCGCGGCGCG	TAGAGACGAGGTTTCACCGT	43	55.8	24	18.0	2.1	***
<i>TMEFF1-RWDD2B</i>	AAGAGGACCATGCTACTCTG	GCTGGAGTGCAGTGGCGCGA	42	54.5	3	2.3	23.2	***
<i>TRNAN35-SEC22B</i>	TGAGGACTGCCTGTTTGACAG	CTACATTATTGAGCAGGGGG	41	53.2	1	0.8	69.8	***
<i>SNORD114_-MEG8_c</i>	GAGCCTGGACACTAATAGAA	AATGGAGGCTCCAGAGTGAA	41	53.2	9	6.8	6.9	***
<i>YY1AP1-YY1AP1_b</i>	GGCTCCATTCTGGTTAAAG	CGAATGAATTTCCCTGTTTG	39	50.6	1	0.8	66.4	***
<i>PITPNM3-KIAA0753</i>	GAAACTGGACGAGCATCAAG	TGTTTCTGCCACCCACTTAG	39	50.6	1	0.8	66.4	***
<i>SPECC1L-SPECC1L</i>	CAAATCCAAGATTAGCACAG	ATTTGCTGTAAATGCATCA	39	50.6	1	0.8	66.4	***

4

5

6

7

8

9

10

11

12

13

14

1

2 **Table 2 Distributions of the top sixteen HFGs among OND, MTLE, and AD**

HFGs	ASL (77)		OND (7)		MTLE (17)		AD (9)	
	#	%	#	%	#	%	#	%
<i>ZNF528-ZNF880</i>	76	98.7	6	85.7	16	94.1	8	88.9
<i>SNORD114_-MEG8_a</i>	70	90.9	5	71.4	14	82.4	5	55.6
<i>TRNAN35-FAM91A3P</i>	69	89.6	5	71.4	4	23.5	4	44.4
<i>ADAMTSL3-SH3GL3 a</i>	67	87.0	7	100.0	14	82.4	9	100.0
<i>SNORD114_-MEG8_b</i>	65	84.4	7	100.0	13	76.5	5	55.6
<i>SDHAP2-SDHAP2 a</i>	60	77.9	4	57.1	14	82.4	6	66.7
<i>YY1AP1-YY1AP1 a</i>	58	75.3	6	85.7	12	70.6	0	0.0
<i>SDHAP2-SDHAP2 b</i>	57	74.0	4	57.1	9	52.9	3	33.3
<i>ERMP1-KIAA2026</i>	57	74.0	5	71.4	4	23.5	4	44.4
<i>TRNAN35-SRGAP2P2</i>	56	72.7	4	57.1	5	29.4	4	44.4
<i>ADAMTSL3-SH3GL3 b</i>	54	70.1	6	85.7	9	52.9	4	44.4
<i>FMNL2-FMNL2</i>	53	68.8	7	100.0	9	52.9	0	0.0
<i>SMOX-RAD54L2</i>	50	64.9	3	42.9	7	41.2	0	0.0
<i>PTPRS-PHACTR3</i>	47	61.0	5	71.4	1	5.9	0	0.0
<i>FAM168B-NUCB2</i>	43	55.8	3	42.9	4	23.5	0	0.0
<i>TMEFF1-RWDD2B</i>	42	54.5	6	85.7	5	29.4	0	0.0
<i>TRNAN35-SEC22B</i>	41	53.2	5	71.4	4	23.5	0	0.0
<i>SNORD114_-MEG8_c</i>	41	53.2	5	71.4	5	29.4	0	0.0
<i>YY1AP1-YY1AP1 b</i>	39	50.6	4	57.1	4	23.5	0	0.0
<i>PITPNM3-KIAA0753</i>	39	50.6	3	42.9	4	23.5	2	22.2
<i>SPECC1L-SPECC1L</i>	39	50.6	6	85.7	8	47.1	0	0.0

3

4

5

6

7

8

9

10

11

12

13

1
2
3
4
5
6
7

Table 3 *ADAMTSL3-SH3GL3* behaved like *SH3GL3-ADAMTSL3* and was a potential dormant HFG.

Data	<i>SH3GL3-ADAMTSL3</i> (%)	<i>ADAMTSL3-SH3GL3</i> (%)
Amyotrophic lateral sclerosis (77)	98.7	87.0
Non-neurological controls (12)	100	100
Other neurological disorders (7)	100	100
mesial temporal lobe epilepsy (17)	94.1	82.4
Alzheimer's disease (9)	100	100
Alzheimer's disease control (8)	100	87.5
GTEX brain cortex (133)	7.5	1.5

8
9

1 **Figure Legends**

2

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

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

11

12 **Fig. 2 Potential genomic mechanisms to generate *ZNF528-ZNF880*, *TRNAN35-FAM91A3P*,**

13 **and *SNORD114_-MEG8_*.** a). A potential local inversion to form *ZNF528-ZNF880*; b) A

14 putative local translocation to produce *TRNAN35-FAM91A3P*; and c). A potential local

15 inversion to generate *SNORD114_-MEG8_*, Black and red horizontal arrows indicated 5' and 3'

16 genes of HFGs. Gray horizontal arrows were genes surrounding 5' and 3' genes. The vertical

17 arrows showed potential genomic rearrangements.

18

19 **Fig. 3 A potential local inversion of *SH3GL3-ADAMTSL3* to produce *ADAMTSL3-SH3GL3*.**

20 Two identical-stand neighboring *SH3GL3* and *ADAMTSL3* genes (a) produced an *SH3GL3-*

21 *ADAMTSL3* readthrough pre-mRNA (b), which was spliced to form three of thirteen *SH3GL3-*

22 *ADAMTSL3* EFG transcripts(c). *SH3GL3-ADAMTSL3* was locally inverted or locally amplified to

23 *ADAMTSL3-SH3GL3* hereditary fusion gene (HFG) (d), transcribed into *ADAMTSL3-SH3GL3*

24 pre-mRNAs €, which was spliced into three examples of 42 *ADAMTSL3-SH3GL3* HFG

1 transcripts. Black and red horizontal solid arrows showed 5' and 3' genes of HFGs. Black and
2 red rectangles were exons. Black triangles were introns. The double dashed lines were
3 putative breakpoints. Horizontal dashed lines indicated omitted exons and introns. An open
4 arch arrow indicated a local translocation or genomic amplification.
5

1 Supplemental Materials

2 NYGC ALS Consortium

3 Hemali Phatnani⁵, Justin Kwan⁶, Dhruv Sareen^{7,8}, James R. Broach⁹, Zachary Simmons¹⁰,
4 Ximena Arcila-Londono¹¹, Edward B. Lee¹², Vivianna M. Van Deerlin¹², Neil A. Shneider¹³,
5 Ernest Fraenkel¹, Lyle W. Ostrow¹⁴, Frank Baas^{15,16}, Noah Zaitlen¹⁷, James D. Berry^{18,19}, Andrea
6 Malaspina^{19,20,21}, Pietro Fratta²², Gregory A. Cox²³, Leslie M. Thompson^{24,25}, Steve Finkbeiner²⁶,
7 Efthimios Dardiotis²⁷, Timothy M. Miller²⁸, Siddharthan Chandran²⁹, Suvankar Pal²⁹, Eran
8 Hornstein³⁰, Daniel J. MacGowan³¹, Terry Heiman-Patterson³², Molly G. Hammell³³, Nikolaos. A.
9 Patsopoulos^{34,35}, Oleg Butovsky³⁶, Joshua Dubnau³⁷, Avindra Nath³⁸, Robert Bowser^{39,40},
10 Matthew Harms⁴¹, Eleonora Aronica⁴², Mary Poss⁴³, Jennifer Phillips-Cremins⁴⁴, John Crary⁴⁵,
11 Nazem Atassi⁴⁶, Dale J. Lange^{47,48}, Darius J. Adams^{49,50}, Leonidas Stefanis^{51,52}, Marc Gotkine⁵³,
12 Robert H. Baloh^{54,55}, Suma Babu¹⁹, Towfique Raj⁵⁶, Sabrina Paganoni⁵⁷, Ophir Shalem^{58,59},
13 Colin Smith^{60,61}, Bin Zhang⁶², Brent Harris⁶³, Iris Broce⁶⁴, Vivian Drory⁶⁵, John Ravits⁶⁶, Corey
14 McMillan⁶⁷, Vilas Menon⁶⁸, Lani Wu⁶⁹, Steven Altschuler⁶⁹, Yossef Lerner⁷⁰, Rita Sattler⁷¹,
15 Kendall Van Keuren-Jensen⁷², Orit Rozenblatt-Rosen⁷³, Kerstin Lindblad-Toh⁷³, Katharine
16 Nicholson⁷⁴, Peter Gregersen⁷⁵, Jeong-Ho Lee⁷⁶, Sulev Kokos⁷⁷, Stephen Muljo⁷⁸, & Bryan J.
17 Traynor⁷⁹.

18

19 ⁵Center for Genomics of Neurodegenerative Disease (CGND), New York Genome Center, New
20 York, NY, USA. ⁶Department of Neurology, Lewis Katz School of Medicine, Temple University,
21 Philadelphia, PA, USA. ⁷Cedars-Sinai Department of Biomedical Sciences, Board of Governors
22 Regenerative Medicine Institute and Brain Program, Cedars-Sinai Medical Center, University of
23 California, Los Angeles, CA, USA. ⁸Department of Medicine, University of California, Los
24 Angeles, CA, USA. ⁹Department of Biochemistry and Molecular Biology, Penn State Institute for

The First Insight into the Hereditary Fusion Gene Landscape of ASL

1 Personalized Medicine, The Pennsylvania State University, Hershey, PA, USA. ¹⁰Department of
2 Neurology, The Pennsylvania State University, Hershey, PA, USA. ¹¹Department of Neurology,
3 Henry Ford Health System, Detroit, MI, USA. ¹²Department of Pathology and Laboratory
4 Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.
5 ¹³Department of Neurology, Center for Motor Neuron Biology and Disease, Institute for Genomic
6 Medicine, Columbia University, New York, NY, USA. ¹Department of Biological Engineering,
7 Massachusetts Institute of Technology, Cambridge, MA, USA. ¹⁴Department of Neurology,
8 Johns Hopkins School of Medicine, Baltimore, MD, USA. ¹⁵Department of Neurogenetics,
9 Academic Medical Centre, Amsterdam, The Netherlands. ¹⁶Leiden University Medical Center,
10 Leiden, The Netherlands. ¹⁷Department of Medicine, Lung Biology Center, University of
11 California, San Francisco, CA, USA. ¹⁸ALS Multidisciplinary Clinic, Neuromuscular Division,
12 Department of Neurology, Harvard Medical School, Boston, MA, USA. ¹⁹Neurological Clinical
13 Research Institute, Massachusetts General Hospital, Boston, MA, USA. ¹⁹Centre for
14 Neuroscience and Trauma, Blizard Institute, Barts, Queen Mary University of London, London,
15 UK. ²⁰The London School of Medicine and Dentistry, Queen Mary University of London,
16 London, UK. ²¹Department of Neurology, Basildon University Hospital, Basildon, UK. ²²
17 Institute of Neurology, National Hospital for Neurology and Neurosurgery, University College
18 London, London, UK. ²³The Jackson Laboratory, Bar Harbor, ME, USA. ²⁴Department of
19 Psychiatry and Human Behavior, Department of Biological Chemistry, School of Medicine,
20 University of California, Irvine, CA, USA. ²⁵Department of Neurobiology and Behavior, School of
21 Biological Sciences, University of California, Irvine, CA, USA. ²⁶Taube/Koret Center for
22 Neurodegenerative Disease Research, Roddenberry Center for Stem Cell Biology and
23 Medicine, Gladstone Institute, San Francisco, CA, USA. ²⁷Department of Neurology and
24 Sensory Organs, University of Thessaly, Thessaly, Greece. ²⁸Department of Neurology,

The First Insight into the Hereditary Fusion Gene Landscape of ASL

1 Washington University in St Louis, St Louis, MO, USA. ²⁹Centre for Clinical Brain Sciences,
2 Anne Rowling Regenerative Neurology Clinic, Euan MacDonald Centre for Motor Neurone
3 Disease Research, University of Edinburgh, Edinburgh, UK. ³⁰Department of Molecular
4 Genetics, Weizmann Institute of Science, Rehovot, Israel. ³¹Department of Neurology, Icahn
5 School of Medicine at Mount Sinai, New York, NY, USA. ³²Center for Neurodegenerative
6 Disorders, Department of Neurology, the Lewis Katz School of Medicine, Temple University,
7 Philadelphia, PA, USA. ³³Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA. ³⁴
8 Computer Science and Systems Biology Program, Ann Romney Center for Neurological
9 Diseases, Department of Neurology and Division of Genetics in Department of Medicine,
10 Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. ³⁵Program in
11 Medical and Population Genetics, Broad Institute, Cambridge, MA, USA. ³⁶Ann Romney Center
12 for Neurologic Diseases, Brigham and Women's Hospital, Harvard Medical School, Boston, MA,
13 USA. ³⁷ Department of Anesthesiology, Stony Brook University, Stony Brook, NY, USA.
14 ³⁸Section of Infections of the Nervous System, National Institute of Neurological Disorders and
15 Stroke, NIH, Bethesda, MD, USA. ³⁹ Department of Neurology, Barrow Neurological Institute,
16 St Joseph's Hospital, Phoenix, AZ, USA. ⁴⁰Medical Center, Department of Neurobiology,
17 Barrow Neurological Institute, St Joseph's Hospital and Medical Center, Phoenix, AZ, USA.
18 ⁴¹Department of Neurology, Division of Neuromuscular Medicine, Columbia University, New
19 York, NY, USA. ⁴²Department of Neuropathology, Academic Medical Center, University of
20 Amsterdam, Amsterdam, The Netherlands. ⁴³Department of Biology and Veterinary and
21 Biomedical Sciences, The Pennsylvania State University, University Park, PA, USA. ⁴⁴ New
22 York Stem Cell Foundation, Department of Bioengineering, School of Engineering and Applied
23 Sciences, University of Pennsylvania, Philadelphia, PA, USA. ⁴⁵Department of Pathology,
24 Fishberg Department of Neuroscience, Friedman Brain Institute, Ronald M. Loeb Center for

The First Insight into the Hereditary Fusion Gene Landscape of ASL

1 Alzheimer's Disease, Icahn School of Medicine at Mount Sinai, New York, NY, USA. 46
2 Department of Neurology, Harvard Medical School, Neurological Clinical Research Institute,
3 Massachusetts General Hospital, Boston, MA, USA. 47 Department of Neurology, Hospital for
4 Special Surgery, New York, NY, USA. ⁴⁸Weill Cornell Medical Center, New York, NY, USA. 49
5 Medical Genetics, Atlantic Health System, Morristown Medical Center, Morristown, NJ, USA. 50
6 Overlook Medical Center, Summit, NJ, USA. ⁵¹Center of Clinical Research, Experimental
7 Surgery and Translational Research, Biomedical Research Foundation of the Academy of
8 Athens (BRFAA), Athens, Greece. ⁵²^{1st} Department of Neurology, Eginition Hospital, Medical
9 School, National and Kapodistrian University of Athens, Athens, Greece. ⁵³Neuromuscular/EMG
10 service and ALS/Motor Neuron Disease Clinic, Hebrew University-Hadassah Medical Center,
11 Jerusalem, Israel. ⁵⁴ Board of Governors Regenerative Medicine Institute, Los Angeles, CA,
12 USA. ⁵⁵Department of Neurology, Cedars-Sinai Medical Center, Los Angeles, CA, USA. 56
13 Departments of Neuroscience, Genetics and Genomic Sciences, Ronald M. Loeb Center for
14 Alzheimer's disease, Icahn School of Medicine at Mount Sinai, New York, NY, USA. 57 Harvard
15 Medical School, Department of Physical Medicine and Rehabilitation, Spaulding Rehabilitation
16 Hospital, Boston, MA, USA. 58 Center for Cellular and Molecular Therapeutics, Children's
17 Hospital of Philadelphia, Philadelphia, PA, USA. ⁵⁹Department of Genetics, Perelman School of
18 Medicine, University of Pennsylvania, Philadelphia, PA, USA. ⁶⁰Centre for Clinical Brain
19 Sciences, University of Edinburgh, Edinburgh, UK. ⁶¹Euan MacDonald Centre for Motor
20 Neurone Disease Research, University of Edinburgh, Edinburgh, UK. ⁶²Department of Genetics
21 and Genomic Sciences, Icahn Institute of Data Science and Genomic Technology, Icahn School
22 of Medicine at Mount Sinai, New York, NY, USA. ⁶³Department of Neuropathology, Georgetown
23 Brain Bank, Georgetown Lombardi Comprehensive Cancer Center, Georgetown University
24 Medical Center, Washington, DC, USA. ⁶⁴Neuroradiology Section, Department of Radiology

The First Insight into the Hereditary Fusion Gene Landscape of ASL

1 and Biomedical Imaging, University of California, San Francisco, San Francisco, CA, USA.
2 ⁶⁵Neuromuscular Diseases Unit, Department of Neurology, Tel Aviv Sourasky Medical Center,
3 Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv, Israel. ⁶⁶Department of
4 Neuroscience, University of California San Diego, La Jolla, CA, USA. ⁶⁷Department of
5 Neurology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA.
6 ⁶⁸Department of Neurology, Columbia University Medical Center, New York, NY, USA.
7 ⁶⁹Department of Pharmaceutical Chemistry, University of California San Francisco, San
8 Francisco, CA, USA. ⁷⁰Hadassah Hebrew University, Jerusalem, Israel. ⁷¹Department of
9 Translational Neuroscience, Barrow Neurological Institute, Phoenix, AZ, USA. ⁷²The
10 Translational Genomics Research Institute (TGen), Phoenix, AZ, USA. ⁷³Broad Institute,
11 Cambridge, MA, USA. ⁷⁴Massachusetts General Hospital, Boston, MA, USA. ⁷⁵Institute of
12 Molecular Medicine, Feinstein Institutes for Medical Research, Northwell Health, Manhasset,
13 NY, USA. ⁷⁶Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South
14 Korea. ⁷⁷Perron Institute for Neurological and Translational Science, Nedlands, Western
15 Australia, Australia. ⁷⁸Integrative Immunobiology Section, National Institute of Allergy and
16 Infectious Disease, NIH, Bethesda, MD, USA. ⁷⁹Neuromuscular Disease Research Section,
17 National Institute of Aging, NIH, Bethesda, MD, USA

18

19 **Supplemental Table 1: Lists of hereditary fusion genes (HFGs) discovered from the non-**
20 **neurological controls (NNC)**

21

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

23

1 **Supplemental Table 3: Lists of HFGs that were discovered in amyotrophic lateral**
2 **sclerosis and present in the frontal cortex, medial motor cortex, and lateral motor cortex**

3

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

6

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

9

10

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

13

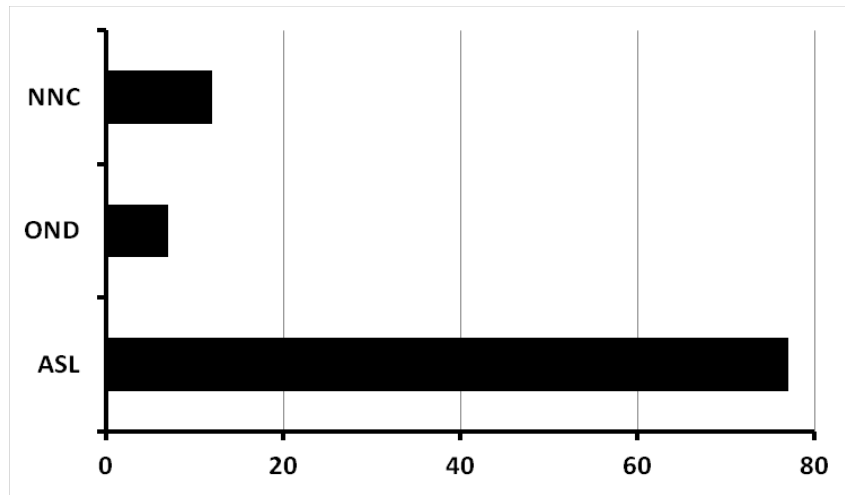
14 **Supplemental Fig.2 Sequence alignment of conceptual FMNL2-FMNL2 fusion**
15 **protein and formin like 2 protein. Dashed line indicated the sequence gap.**

16

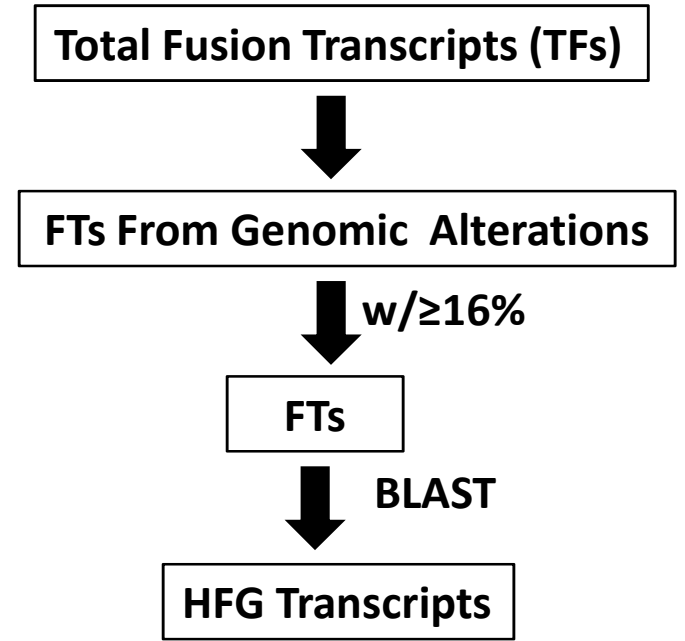
17 **Supplemental Fig.3 Sequence alignment of conceptual YY1AP1-YY1AP1 fusion**
18 **protein and YY1 associated protein 1. Dashed line indicated the sequence gap.**

Fig.1

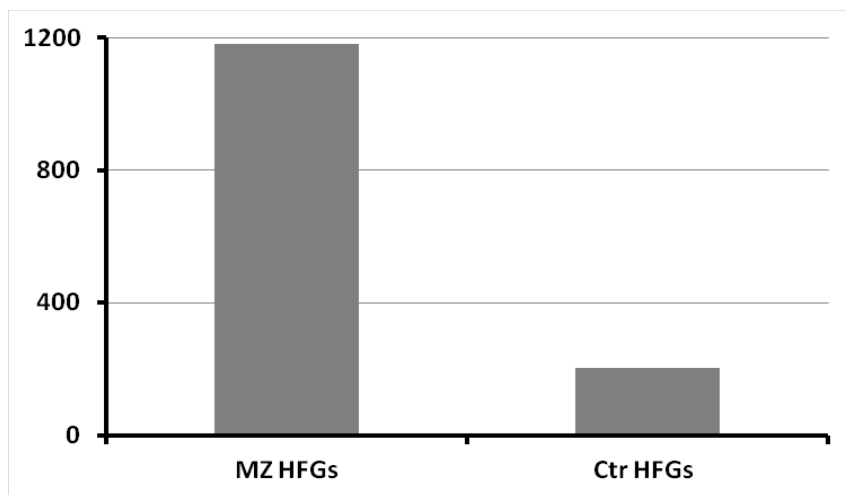
a



b



c



d

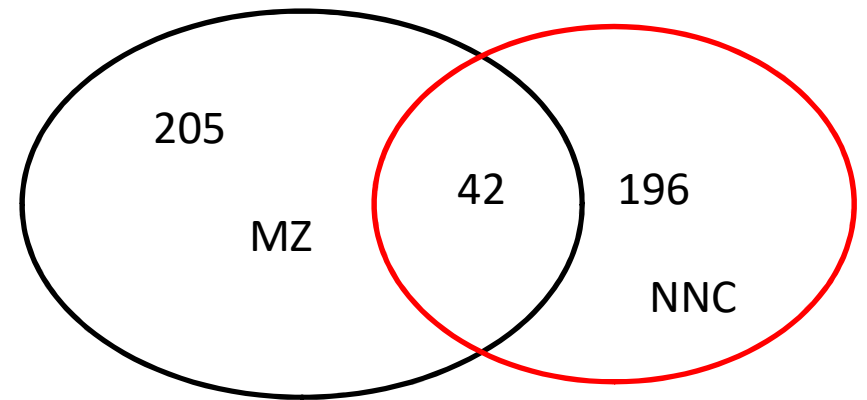
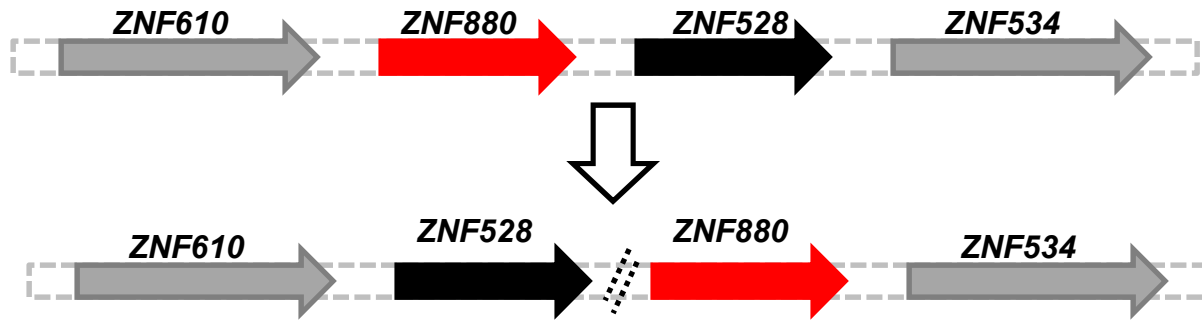
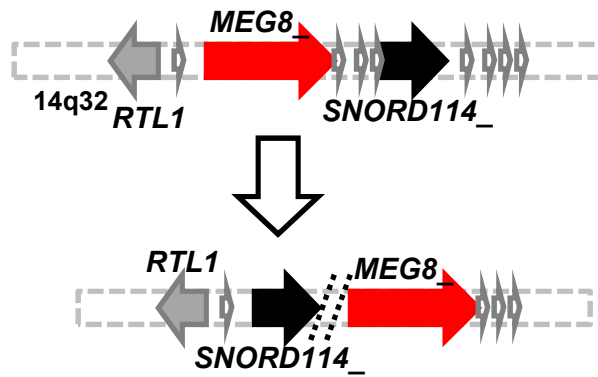


Fig. 4

a



b



c

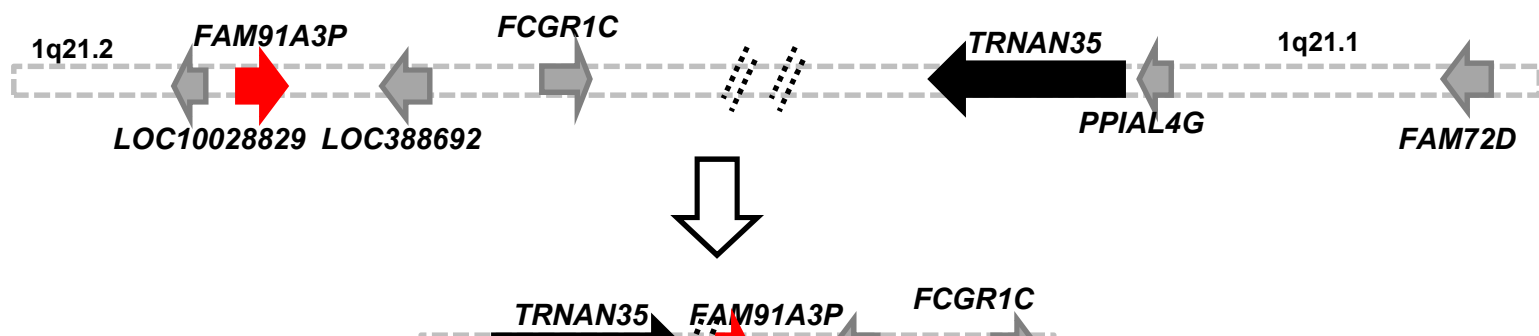


Fig.3

