University of Massachusetts Medical School

eScholarship@UMMS

GSBS Dissertations and Theses

Graduate School of Biomedical Sciences

2018-02-27

Identification of Novel Genetic Variations for Amyotrophic Lateral Sclerosis (ALS)

Guang Xu
University of Massachusetts Medical School

Repository Citation

Xu G. (2018). Identification of Novel Genetic Variations for Amyotrophic Lateral Sclerosis (ALS). GSBS Dissertations and Theses. https://doi.org/10.13028/M2VQ38. Retrieved from https://escholarship.umassmed.edu/gsbs_diss/958

IDENTIFICATION OF NOVEL GENETIC VARIATIONS FOR AMYOTROPHIC LATERAL SCLEROSIS (ALS)

A Masters Thesis Presented

By

GUANG XU

Submitted to the Faculty of the

University of Massachusetts Graduate School of Biomedical Sciences, Worcester
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

February 2th, 2018

NEUROLOGY & BIOINFORMATICS

IDENTIFICATION OF NOVEL GENETIC VARIATIONS FOR AMYOTROPHIC LATERAL SCLEROSIS (ALS)

A Masters Thesis Presented
By
Guang Xu
The signatures of the Master's Thesis Committee signify completion and approval as to style and content of the Thesis

Lawrence Hayward, Chair of Committee

Robert Brown, Member of Committee

Jeff Bailey, Member of Committee

Janice Dominov, Member of Committee

Mary Ellen Lane, Member of Committee

The signature of the Dean of the Graduate School of Biomedical Sciences signifies that the student has met all master's degree graduation requirements of the school.

Anthony Carruthers, Ph.D.,

Dean of the Graduate School of Biomedical Sciences Program
Bioinformatics & Computational Biology Program
Month, Day and Year
February 2th, 2018

ACKNOWLEDGEMENT

I would first like to thank my thesis advisor Dr. Robert Brown and Dr. Jeff Bailey for being supportive and inspiring mentors. I’m thankful for the computational skill training in my first three years with Dr. Bailey, which has huge impact on my future career. I’d also like to express my thanks for Dr. Brown giving me unparalleled support for my ALS genetics research by providing sufficient funding, giving useful guidance and offering me huge freedom to test my ideas.

I would also like to thank Peter Sapp, Nicholas Wightman and Ru-Ju Chian for years of technical support at Brown lab. I’d also like to thank Jake Metterville and Alexandra Weiss for their hard work for DPP6 and UNC13A mice projects. The whole Bailey lab also has provided me support for my computational work, which I feel extremely thankful. In addition, I'd like to express my thanks to Dr. Lingtao Peng for his years of advice on RNA biology, Dr. Anastasia Grigorenko for her guidance on NGS library preparation, Dr. Claudia Fallini for her advice on microscopy usage, Dr. Yvonne Chan for her help on circular dichroism, and also Covarrubias lab for their electrophysiology work for DPP6 projects. More, I’d also like to thank all other lab members at Brown lab for always providing support, discussing science and exchanging ideas, and I give my best wishes to everyone for their future career.

I would also like to acknowledge my committee: Dr. Lawrence Hayward, Dr. Janice Dominov, Dr. Zuoshang Xu, Dr. Johns Landers, Dr. Elinor Karlsson, Dr. Konstantin Zeldovich and Dr. Mary Ellen Lane. Their insightful and helpful suggestions contributed a lot to the progress of my research.

Finally, I owe my very sincere and special gratitude to my parents and my wife for their unconditional support and encouragement for whatever choice I’ve made for my career. This accomplishment would not have been possible without them.

ABSTRACT

A list of genes have been identified to carry mutations causing familial ALS such as SOD1, TARDBP, C9orf72. But for sporadic ALS, which is 90% of all ALS cases, the underlying genetic variants are still largely unknown. There are multiple genome-wide association study (GWAS) for sporadic ALS, but usually a large number nominated SNP can hardly be replicated in larger cohort analysis. Also majority of GWAS SNP lie within noncoding region of genome, imposing a huge challenge to study their biological role in ALS pathology. With the rapid development of next-generation sequencing technology, we are able to sequence exome and whole-genome of a large number of ALS patients to search for novel genetic variants and their potential biological function. Here by analyzing exam data, we discovered two novel or extremely rare missense mutations of DPP6 from a Mestizo Mexican ALS family. We showed the two mutations could exert loss-of-function effect by affecting electrophysiological properties of Potassium channels as well as the membrane localization of DPP6. To our knowledge this is the first report of DPP6 nonsynonymous mutations in familial ALS patients. In addition, by analyzing whole-genome data, we discovered strong linkage disequilibrium between SNP rs12608932, a repeatedly significant ALS GWAS signal, and one polymorphic TGGA tetra-nucleotide tandem repeat, which is further flanked by large TGGA repetitive sequences. We also demonstrated rs12608932 risk allele is associated with reduced UNC13A expression level in human cerebellum and

UNC13A knockout could lead to shorter survival in SOD1-G93A ALS mice. Thus the TGGA repeat might be the real underlying genetic variation that confer risk to sporadic ALS.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS, ABBREVIATIONS OR NOMENCLATURE

ALS: Amyotrophic Lateral Sclerosis

FALS: Familial Amyotrophic Lateral Sclerosis

SALS: Sporadic Amyotrophic Lateral Sclerosis

FTD: Frontotemporal Dementia

C9ORF72: Chromosome 9 Open Reading Frame 72

TDP-43: TAR DNA Binding Protein-43

GWAS: Genome-wide Association Study

DPP6: Dipeptidyl peptidase-like protein 6

NGS: Next-generation Sequencing

eQTL: Expression quantitative trait loci

CD: Circular Dichroism

PREFACE

The thesis is submitted for Master of Science degree at University of Massachusetts Medical School. The research described above has been conducted under the supervision of Dr. Robert Brown in Department of Neurology and Dr. Jeff Bailey in the Department of Bioinformatics and Computational Biology. All figures and data in this thesis were directly generated by myself except that the UNC13A mice data (Figure 7) were totally from the help of Alex Weiss from Brown lab.

The work in this thesis is to my best knowledge original, except where acknowledgement and reference were made. Neither this, nor anything substantially similar has been or is being submitted for any other degree, diploma at any other universities or institutions.

Chapter 1. Introduction

Chapter 1.1 Introduction of ALS

Amyotrophic lateral sclerosis (ALS) is a progressive, devastating neurodegenerative disorder caused by motor neuron death of upper motor neurons in cortex and lower motor neurons in spinal cord and brainstem. This will lead to paralysis and eventual respiratory failure with an average survival of 3 years from symptom onset. The mean onset age is 55-60 years old and world-wide incidence is approximately 2 per 100000 individuals (1). Around 10% of ALS cases are familial while the remaining sporadic cases are generally considered multifactorial with both genetic factors and environmental risks conferring susceptibility (2). ALS was traditionally regarded as a pure motor neuron disease, but recent findings about the sensory and spinocerebellar pathways in ALS, as well as the pleiotropy of ALS-associated genes in other syndromes (3), have implied that ALS is a multisystem disorder in which motor neurons tend to be affected most severely.

The pathology underlying ALS still remains largely elusive. Genetics studies indicate an extremely complicated etiology which may involve multiple pathways, such as oxidative stress, mitochondrial dysfunction (1), protein aggregation (4), excitotoxicity, axonal transport impairment and dysregulated RNA processing (5). There's also a growing evidence that the disrupted communication with surrounding glial cells may contribute to motor neuron injury (6). Besides genetics, environmental factors such as smoking, diet and toxic exposure may also put individuals at a higher risk for ALS (7). The identification of an RNA-binding protein TDP-43 as a major constituent of ubiquitinated protein inclusions in ALS has become the hallmark for the study of ALS pathology (8). Normally TDP-43 are predominantly localized within the nucleus, but will be ubiquitinated and shuffled into cytoplasm for most cases of FALS and SALS. The mutations in TARDBP, the TDP-43 coding gene, were discovered in several FALS pedigrees, further consolidating the role TDP-43 may play in ALS (9). However, for most cases, we don't know the genetic variants which lead to TDP-43 translocation. The knowledge for biological role of TDP-43 is still evolving, with recent research indicating it may involve in self-regulation or binding to other ALS-related proteins like FUS (10).

Chapter 1.2 FALS genetics

The genetic cause study of ALS has proved quite difficult mostly due to late-onset, short survival time and incomplete penetrance. Mendelian patterns, mostly dominant inheritance, have been recognized by linkage study in a few large familial ALS pedigrees.

SOD1 is the first gene identified to be associated with ALS. The gene encodes 153 evolutionarily conserved amino acids and catalyzes the reduction of superoxide to protect cell from harmful free radicals. 166 ALS-associated SOD1 mutations have been found, 147 of which are missense type (1). However, around 20% of individuals carrying SOD1 mutations won't show ALS symptoms even at very old age, indicating penetrance of autosomal dominant mutations in ALS can be incomplete.

The cytoplasmic inclusion of TDP-43 led to the discovery of mutations of TARDBP which encodes this protein. The mutations result in redistribution of TDP-43 from nucleus to cytoplasm. Mutations have also been identified in FUS gene (12), whose function resembles TDP-43. The recessive mutations in ALS2 gene, which produces Alsin protein, can cause juvenile-onset ALS (1). Ataxin 2 (ATXN2), a polyglutamine (polyQ) protein mutated in spinocerebellar ataxia type 2 emerges as a potential risk factor (13). The intermediate- length polyQ expansions (27–33 glutamines) in ATXN2 are reported to significantly associated with ALS. Moreover, mutations in UBQLN2, which encodes a ubiquitin-like protein, have been found to cause dominant X-linked ALS (14).

For a long time, linkage study has pointed to 9p21 as a potential locus for SALS. Very recently, it was identified that causal variant is hexanucleotide expansion, (GGGGCC)n, between the first noncoding exons of unknown gene C9ORF72 (15,16). And this expansion can account for a large number of cases of FALS, SALS and FTD, replacing SOD1 as the most common genetic abnormality of ALS patients (17,18).

Chapter 1.3 SALS genetics

As for sporadic ALS, although almost all FALS mutations can be found in SALS, the majority of SALS cannot be explained. However, a number of observations suggest genetic factor role in SALS. Twin studies give an estimate of SALS heritability of 0.6 by comparing monozygotic and dizygotic twins (19). And some analysis showed first-degree relatives of SALS patients have larger risk for developing ALS. Genome-wide association studies (GWAS) have been conducted for ALS samples.

FGGY (FGGY carbohydrate kinase domain containing) is one of the very first putative genes implicated by GWAS using 386 white SALS patients and 542 neurologically normal white controls followed by two independent replications (20). Around the same time, another group from the Netherland reported that ITPR2 (inositol 1,4,5-trisphosphate receptor type 2) may be associated with ALS in three European populations (21). However, when the same Dutch team extended their analysis to include more samples, they found DPP6 (dipeptidyl peptidase like 6) rather than ITPR2 was strongly associated with ALS for European populations (22). Facing the conflicting results, one Irish group tried to conduct GWAS on a more homogeneous population which exhibits extended linkage disequilibrium and lower allelic heterogeneity. They used 221 cases and 211 controls all from Ireland, and found the strongest signal also came from variant in DPP6 (23). However, all the previous identified genes FGGY, ITPR2 and DPP6 cannot be replicated in other later studies (24). Other candidate genes from GWAS include UNC13A (25) which encodes presynaptic proteins found in neuromuscular synapses and KIFAP3 (26), which encodes a kinesin-associated protein.

In contrast to the conflicts and uncertainty above, chromosome 9p21 has been identified in several independent large GWAS of both ALS and FTD (25,27,28, 29), implicating the genetic defect at chromosome 9 in SALS. And it was recently unveiled that the defect is noncoding hexanucleotide repeat expansion in the gene C9ORF72. And in a large-scale population study involving 386 apparently sporadic cases, 19 (5%) cases of apparently sporadic ALS had the C9orf72 repeat expansion (18).

Chapter 1.4 Complex disease and missing heritability of GWAS

ALS is very complex disease related to multiple types of factors. The classical model for complex disease is “threshold liability model”, in which, multiple genetic variants, combined with environmental risks all contribute to the liability of disease. Such liability is normally distributed in the population and disease will only occur for those whose burden is above a particular threshold.

GWAS has been extensively used to discover variants which may confer disease susceptibility and elucidate the architecture of complex traits. Initially GWAS was based on the simple common disease–common variant hypothesis, which has been refuted due to “missing heritability problem”: Only a very small proportion of heritability of complex traits can be explained by variants from GWAS (30).

There's a heated debate about where the missing heritability can be found. The potential sources of missing heritability can be:

1.The rich indels and large structural variants in human genome. The discovery and genotyping of such variants are far lagged behind the SNP study (30).

2.Rare variants may play an important role in disease etiology (31,32), while current methodologies are underpowered for the detection of rare variants due to low allele frequency and allelic heterogeneity (33).

3.Gene-environment interactions (34). For example, people carry genetic factors that confer susceptibility or resistance to a certain disorder only in a particular environment.

4.The epigenetic effects, such as parent-of-origin genetic information and DNA methylation patterns (35), and gene-gene interaction or epistasis (36).

Chapter 1.5 Structural variation

The structural variants (SV) of human genome include deletion, insertion, duplication, inversion, copy-number variation, short tandem repeats, and chromosomal translocation. SV play an important role in human complex disease (37, 38, 39). Copy number variation (CNV) is one type of SV (37). Specifically, recent studies have established that rare and de novo SV/CNV contribute to the genetic risk of a wide range of neurological and neuropsychiatric diseases including autism, schizophrenia and bipolar disorder (40, 41, 42, 43, 44, 45, 46). In addition, short tandem repeat expansion is common for neurological disorder, such as Huntington's disease. And C9orf72 is the most exciting discovery of structural variants for ALS. Hundreds or even thousands of GGGGCC hexanucleotide repeats were found in ALS patients, though it is not clear exactly how these hexanucleotide repeats cause the disease (15).

Genome-wide CNV study has also been applied to ALS samples. One study carried out SNP array for 406 patients with sporadic ALS and 404 controls, and found no loci statistically significant after Bonferroni correction in the association test (47). Similarly, another study around the same time which focused on 408 Irish individuals and 868 Dutch individuals (48), detected 26 copy number gains and 58 copy number losses that showed nominal association with ALS at p value < 0.05, but all of them failed to reach the significance by Bonferroni correction. Later in a genome-wide screen of 1875 cases and 8731 controls, no evidence was found for the difference in global CNV burden between cases and controls. And in the gene-based association study, two genes DPP6 and NIPA1 were highlighted (49).

Chapter 1.6 Next-generation sequencing

Genetic variants study by GWAS heavily relies on linkage with disease-causing variants and barely reports the exact length and breakpoints of structural variants. The availability of next-generation sequencing (NGS) technology are poised to fundamentally change the variant mapping landscape by providing full sequence information. Many computational algorithms have been developed to identify variants using NGS data (50). For example, the two most popular SNP genotype tools are Samtools and the Genome Analysis Toolkit Unified Genotyper (GATK) (51).

The dramatic cost reduction of NGS has enabled whole-genome sequencing of a couple of human genome. However, it still remains unaffordable to sequence the whole genome of a large number of individuals even at a low coverage. Thus exome sequencing becomes an effective alternative approach to capture functionally important exons at a reasonable cost. At present the main application of exome sequencing is to determine SNP and indels, and has enabled the discovery of causal variants of several Mendelian diseases (52, 53, 54), including finding a new gene (valosin-containing protein) from an Italian family with FALS (54). Also, recent trio-based studies using exome sequencing have demonstrated highly disruptive de novo exonic mutations may contribute substantially to the etiology of autism spectrum disorders (55, 56, 57). In addition, algorithms and softwares have been developed to identify SV/CNV based on exome (58, 59, 60) and whole-genome (61, 62) sequencing data.

Chapter 1.7 DPP6 introduction

Dipeptidyl peptidase-like protein 6 (DPP6) is one of the putative ALS genes implicated by SALS GWAS. SNP rs10260404 in DPP6 shows strong association with susceptibility to ALS in several independent studies (22) but fails the replication in large joint analysis (24). DPP6 is an auxiliary subunit of Kv4 family of voltage-gated potassium channels, which underlies the transient subthreshold-activating A-type current in neurons (63, 64). DPP6 knockdown in heterologous expression system shows that DPP6 enhances Kv4 surface expression and accelerates channel activation and inactivation (65). Recent reports also reveal DPP6 has important impact on formation and stability of dendritic filopodia during early neuronal development (66).

Chapter 1.8 UNC13A introduction

UNC13A participates in vesicle maturation during exocytosis as a target of the diacylglycerol second messenger pathway. UNC13A plays a crucial role in neurotransmitter release at synapse by priming synaptic vesicles to fuse with plasma membrane (67). Thus biologically UNC13A is also an attractive candidate for ALS. rs12608932, an intronic SNP within UNC13A is one of the very few risk loci supported by multiple ALS GWAS (24). It's also identified as the shared risk locus for ALS and FTD-TDP in one meta-analysis. Also multiple ALS GWAS for European population have all demonstrated rs12608932 risk allele is associated with shorter survival of ALS, indicating a potential genetic modifier role of UNC13A in ALS (68, 69).

UNC13A protein is composed of one C1 domain, one MUN domain and three C2 domain including RIM-binding C2A domain and calcium-binding C2B domain (70). In addition, UNC13A belongs to UNC13 family where UNC13B, UNC13C and UNC13D which all play certain roles in endocytosis, exocytosis and protein secretion. Also, UNC13A-deficient mice show morphological defects in spinal cord motor neurons, muscle and neuromuscular synapses (71). For transgenic C. elegans expressing mutant TDP-43, UNC13A is required for inducing innate immunity, and deletion of UNC13A could suppress motor neuron degeneration (72).

Chapter 2. Material and methods

Chapter 2.1 Sequencing samples

For DPP6 project, blood samples are collected from a Mexican Mestizo family, where two patients of aunt and niece relationship were identified (Supplementary Figure 1). Interestingly, the mother of the niece is a obligate carrier but didn’t develop ALS. For UNC13A project, we used genomic DNA prepared from blood samples of familial and sporadic ALS patients in 96-well plate, as well as brain DNA of Alzheimer's Disease Research Center (ADRC) Brain Research Program.

Chapter 2.2 NGS library preparation

Around 5 ug of genomic DNA was first diluted in EB buffer and sent for Covaris shearing. DNA fragments were blunted by DNA repair kit (# ER0720 Epicentre), followed by “A tailing” of fragments using Klenow Exo-minus (#KL0810250 Epicentre). Adapters were then added (NEXTflexTM ligation mix and barcodes). The ligation mix was then amplified by PCR for 9 cycles. The PCR product was run on 2% gels and cut for desired size around 350~400bp. The cut gel was then purified to obtain DNA library. We analyzed the library on Agilent Bioanalyzer.

Chapter 2.3 Bioinformatics pipeline for SNP calling

1. SNP calling

We first aligned the 100bp short reads using BWA (Burrows-Wheeler Aligner), generating bam files for each sequencing lane. Each lane-level bam file was processed by indel realignment and base quality recalibrator under GATK package. Then lane-level bam files were merged for both 10282 and 7800 library. We then removed PCR duplicates by Picard's MarkDuplicates. SNP and indels were then called by GATK UnifiedGenotyper. The results were then refined using GATK variants quality recalibrator.

2. Deleterious mutation prioritization

The SNP and indels were first filtered for novel or rare variants with minor allele frequency (MAF) <= 0.1% according to both NHLBI Exome (6500 version) (73) and 1000 Genome Project (2015Aug version) (74) databases. Then we picked up those variants only shared by the two patients. Then by Annovar (75) we tried to annotate the variants and looked for those that are either nonsynonymous or affecting splicing sites.

Chapter 2.4 PCR sequencing

We first pulled out exon or gene sequences according to RefSeq annotation in UCSC Genome Browser. We then designed primers for all sequence fragments of interest by using Primer 3.0. Around 10~15ng patient genomic DNA were used for PCR for sequencing one mutation. AmpliTaq Gold 360 Master mix were used for general PCR; for GC-rich region, we applied Advantage GC Genomic LA PCR kit. Single SNP or indel were identified using novoSNP based on ab1 files. For UNC13A project, we manually checked all sequencing data to get the genotypes for repeat polymorphism.

Chapter 2.5 Fluorescence microscopy

Inserts of DPP6 mRNA sequences were first prepared by PCR using plasmids used in electrophysiology study as template, then sub-cloned into the XhoI and Pst1 sites of pAcGFP1-N1 Vector. The plasmids were confirmed by Sanger sequencing. HEK-293 cells were seeded on glass bottom dish and transfected with the DPP6-GFP constructs and Mem-mCherry marker (76). After 24 hours, live image were captured by a Nikon fluorescence microscopy. The pictures were processed by ImageJ.

Chapter 2.6 Short tandem repeat analysis of online NGS data

SNP information was directly retrieved from VCF files for both 1000 Genome Project and Simon Genome Diversity Project. For tandem repeat polymorphism calling, raw bam files were downloaded for both datasets, and lobSTR (77) with default parameters were used to call short tandem repeat (STR) polymorphism.

Then r^2 score was calculated based on SNP and STR calling using PLINK (78). Also because of low-coverage of 1000 Genome data, the r^2 score accuracy was then improved by setting threshold for quality score of lobSTR callings to 0.5.

Chapter 2.7 G-quadruplex identification *in vitro*

IDT oligo, 200bp Ultramar oligo were ordered for both alleles containing 5 copies of TGGA (5-copy) and containing 9 copies of TGGA (9-copy), centering on the TGGA tetra-repeats sequences (See Supplementary Table 1). The samples were annealed by heating at 95 degree for 10min and slowly cooled overnight to room temperature in the presence or absence of KCl. KCL could allow the tandem repeats of guanines to fold into the G-quadruplex (79). The samples were then tested using circular dichroism (CD) with default parameters and CD spectral features indicative of G-quadruplex were analyzed.

Chapter 2.8 Genome-wide TGGA enrichment study

In order to find all repeats that resemble UNC13A repeats across the whole genome, we applied Bedtools (80) to intersect/cluster all TGGA or TCCA tandem repeats in RepeatMasker less than 150bp away from each other, but with total length greater than 500bp. By such standard, we identified 640 such TGGA/ TCCA repeat cluster genome-wide, and 350 of them are within 297 genes. Then we tried to search for gene enrichment for these repeat clusters using software GREAT (81).

Chapter 2.9 eQTL study for cerebellums

We sequenced ~800 ADRC brain DNA and selected 30 samples homozygous for rs12608932 non-risk allele (AA) and 20 cerebellum samples homozygous for risk allele (CC) after controlling for age, gender, diagnosis and tissue specificity. We then extracted total RNA from the cerebellums, checked for RNA quality, and prepared cDNA. Then we applied TaqMan qPCR assay (assay ID: Hs01000584_m1) to measure expression level of UNC13A using GAPDH as control.

Chapter 3. Results

Chapter 3.1 Bioinformatics analysis of NGS data

We carried out whole-genome Illumina sequencing for two patients (RB_10282 and RB_7800) of aunt-niece relationship from a Mexican ALS family. These two samples have been tested and shown negative for all major known ALS mutations including SOD1, FUS, TARDBP and C9orf72. High-quality sequences were achieved by Illumina HiSeq 2000 for the two patients with average whole-genome coverage 23.7 and 11 (See Methods).

We then conducted bioinformatics analysis for the two whole-genome sequencing data. The pipeline is shown in Figure 1. We first aligned the short reads to hg19 human reference genome using BWA, then called SNPs and indels using UnifiedGenotyper of GATK (51). We filtered out common variations with minor allele frequency (MAF) > 0.001 according to both NHLBI Exome and 1000 Genome Project databases (73, 74), then picked up those shared by the two patients, followed by functional annotation to prioritize for deleterious variants which are either nonsynonymous or affecting splicing sites (75) (See Methods). Finally we manually checked the list to remove obvious artifacts (for example caused by low read-depth). This led to discovery of 72 possibly deleterious mutations shared by the two patients (Supplementary Table 2). PCR sequencing verified all of them. We also calculated evolutionary conservation score as well as functional effects score to predict deleteriousness of mutations using multiple programs (data not shown). It's interesting that among 72 genes we see several channel-related genes such as CACNA2D1, TRPM2, DPP6, which are all related to channel activity.

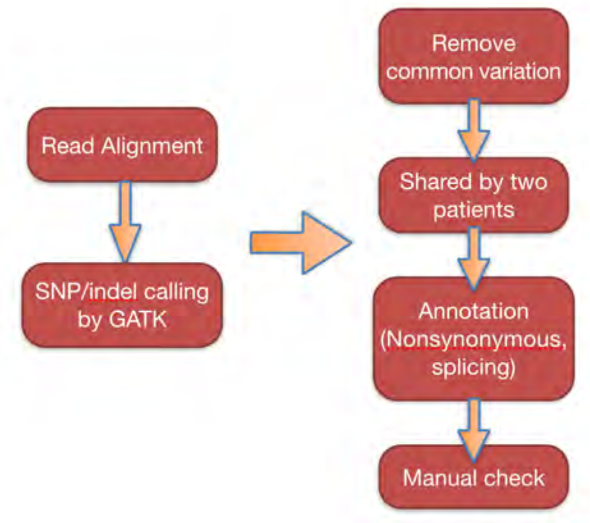


Figure 1. Bioinformatic pipeline for mutation identification. Left panel is about generating high-quality SNP/indel callings while the right panel shows procedures to prioritize for potential deleterious or disease-causing mutations.

Chapter 3.2 Two DPP6 mutations identified for the two Mexican patients.

Among the 76 rare or novel verified mutations, two are within DPP6 gene (V343E and A716V, see Figure 2). DPP6 has been shown associated with SALS in several GWAS and acts as a transmembrane protein with a large extracellular C-terminal domain. Functionally DPP6 is mostly studied as a part of A-type Potassium channel complex consisting of pore-forming Kv4 channel, Kv channel-interacting protein (KChIP) and DPP6. The mutated amino acids are both located on the large extracellular domain. The mutation V343E is predicted as very deleterious by Polyphen2 (82) while A716V is predicted as possibly-damaging. Also, V343E is only one amino acid downstream a N-glycosylation locus and Valine to Glutamate change is very likely to repress N-glycosylation efficiency (83,84). Sequencing of other family members confirmed these two mutations are on the same haplotype. Sequencing of 90 Mexican controls found neither of the two mutations (Table 1). We also sequenced 75 familial and 190 sporadic ALS samples without any known ALS mutations, but we didn't find these two mutations.

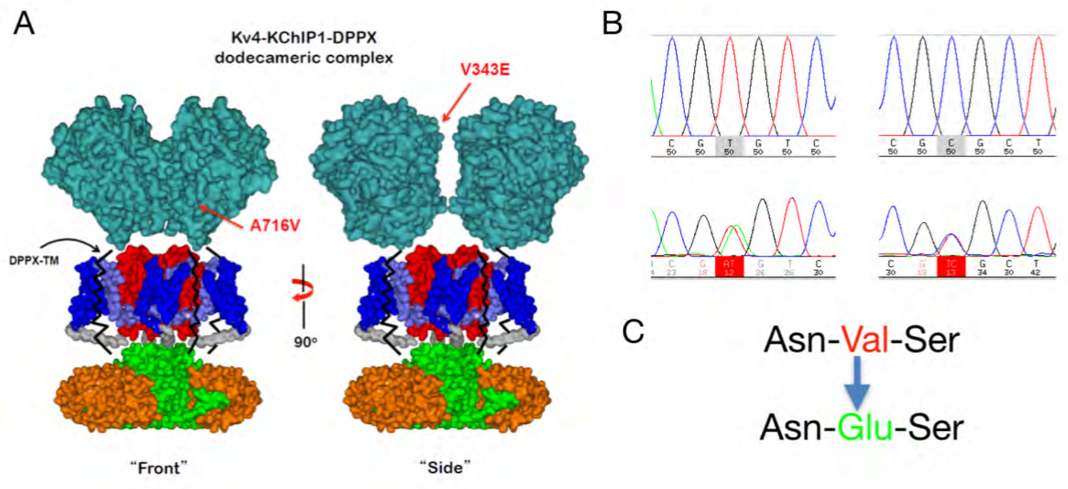
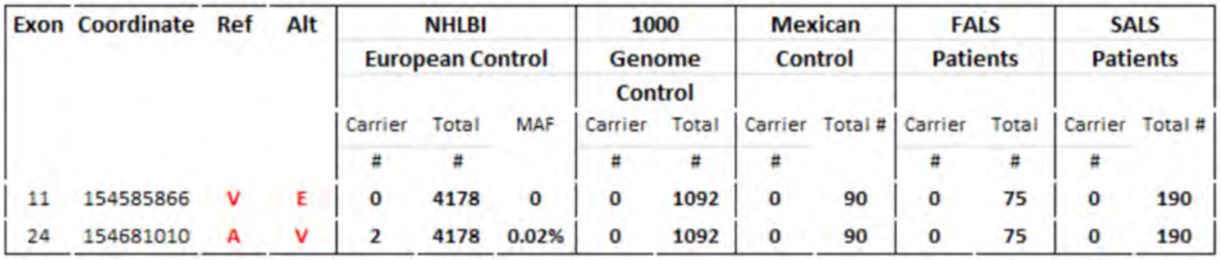


Figure 2. Two DPP6 missense mutations. (A) The structures (65) of Kv4-KChIP-DPP6 complex and the position of two mutated amino acids on the huge

extracellular domain of DPP6 (Panel A picture was created by Dr. Robert Brown). (B) Sanger sequencing conformation of the two missense mutations. Left panel: V343E; right panel: A716V. (C) V343E might change N-linked glycosylation efficiency. N-linked glycosylation generally occurs at the sequon Asn-X-Ser/Thr, where oligosaccharide is attached to the nitrogen atom of Asparagine. V343E is next to a Asparagine and likely to repress the glycosylation efficiency.



Exon	Coordinate	Ref	Alt	NHLBI European Control			1000 Genome Control		Mexican Control		FALS Patients		SALS Patients	
				Carrier #	Total #	MAF	Carrier #	Total #	Carrier #	Total #	Carrier #	Total #	Carrier #	Total #
11	154585866	V	E	0	4178	0	0	1092	0	90	0	75	0	190
24	154681010	A	V	2	4178	0.02%	0	1092	0	90	0	75	0	190

Table 1. The Minor allele frequency (MAF) of two missions mutations. The coordinate is based on hg19 version of human genome. The mutation V343E is not seen in any of the control populations including NHLBI (73), 1000 Genome Project (74) and our in-house Mexican controls, while A716V is seen at a frequency of 0.02% in NHLBI project. Neither mutation is seen in our further screen of ALS patients without known ALS mutations.

In order to search for more DPP6 nonsynonymous mutations, we designed primers and sequenced all exons of DPP6 (including different isoforms) for 75 familial Caucasian ALS patients without any known ALS mutations to look for novel DPP6 mutations. We didn’t find any more DPP6 mutations that are nonsynonymous or changing splicing sites. To our knowledge, no other DPP6 nonsynonymous mutations have been found for ALS patients except for one mutation 883G>A found for one sporadic patient (85).

Chapter 3.3 V343E disrupts DPP6 localization.

One possibility of DPP6 poor expression is its membrane localization is disrupted. To investigate if the two mutations affect DPP6 localization, We first tried to sub-clone the rat DPP6 into pEGFP-N1 vectors and generated three mutants: V343E-DPP6-GFP, A716V-DPP6-GFP and V343E-A716V-DPP6-GFP. The constructs were then co-transfected to HEK-WT cells with membrane marker Mem-mCherry, and live images were captured by fluorescence microscope. We found that V343E obviously disrupts DPP6 membrane localization resulting in a diffused localization pattern in cytoplasm, while A716V shows similar results as in WT-DPP6 (Figure 3). And V343E-A716V double mutant, not surprisingly, disrupted DPP6 membrane localization, but also demonstrated punctate. Further experiments are needed to verify the punctate and the possible additive effects by double mutants. In summary, the above data are quite consistent with electrophysiology study that DPP6 is not expressing well, indicating loss-of-function effect of V343E.

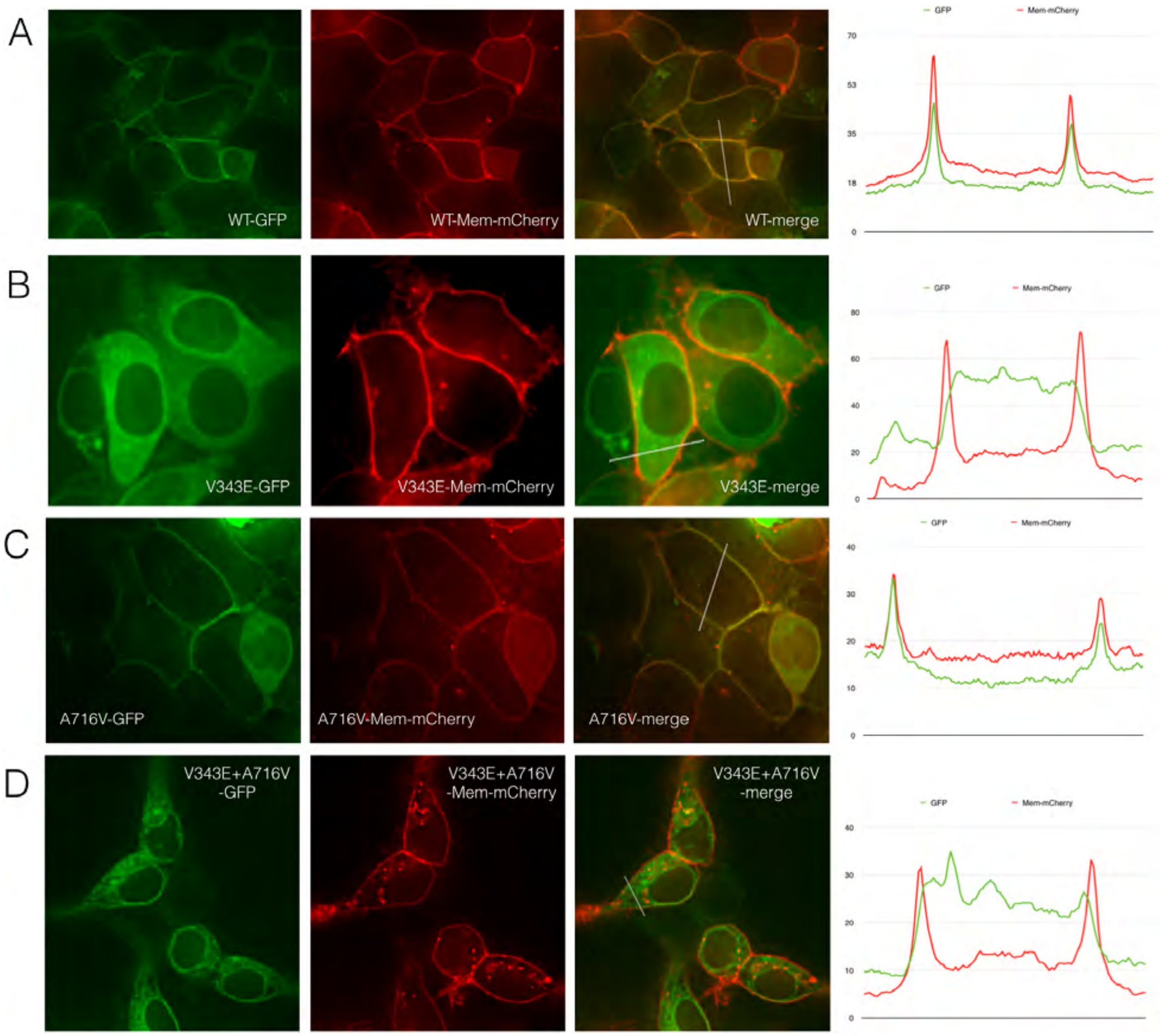


Figure 3. DPP6 membrane localization is affected by the patient mutations. Membrane marker Mem-mCherry was co-expressed with DPP6-GFP fusion protein in HEK-293 cells. An intensity plot along the white straight line is also shown. (A) WT-DPP6-GFP fusion protein is exclusively expressed on the plasma membrane, co-localizing with Mem-mCherry marker in around 50% HEK-293 cells. (B) V343E-DPP6-GFP fusion protein is diffused in cytoplasm in almost

100% HEK-293 cells, indicating mutation V343E disrupts DPP6 membrane localization. (C) A716V-DPP6-GFP fusion protein shows similar localization pattern to WT-DPP6-GFP, that is, exclusively expressed on the plasma membrane in around 50% HEK-293 cells. (D) V343E-A716V-DPP6-GFP double mutant is expressed in cytoplasm in almost 100% HEK-293 cells, and interestingly we could also see some punctate.

Chapter 3.4 Discovery of TGGA tandem repeats of UNC13A

For regions surrounding rs12608932, we first tried to search for any functional annotation including transcription level, histone modification, DNaseI hypersensitivity clusters as well as transcription factor binding from ENCODE project database. Also rs12608932 falls into intron-19 that lacks of functional annotation. However, the whole intron-19 is highly conserved in primates and to some extent conserved in other distant species, suggesting possible biological function of this intron.

We then set out to look at the genomic sequences surrounding rs12608932. We discovered there’s a possible perfect linkage disequilibrium between rs12608932 and one TGGA tetranucleotide tandem repeat around 200bp downstream the SNP by manually investigating 20 available whole-genome sequences (Supplementary Figure 2). The sequencing depth is sufficient enough for calling indels. We inferred from the sequencing data that rs12608932 non-risk allele is linked with 5- or 7-copy TGGA repeat; while risk allele is linked with 9-copy TGGA repeat or beyond (Figure 4). We then confirmed this linkage by sequencing rs12608932 and microsatellite analysis of TGGA repeat copy number in additional 380 Caucasian ALS DNA samples as well as 550 Caucasian control DNA samples (Supplementary Table 3), as well as online NGS data from 1000 Genome Project and Simon Genome Diversity Project (Supplementary Table 4).

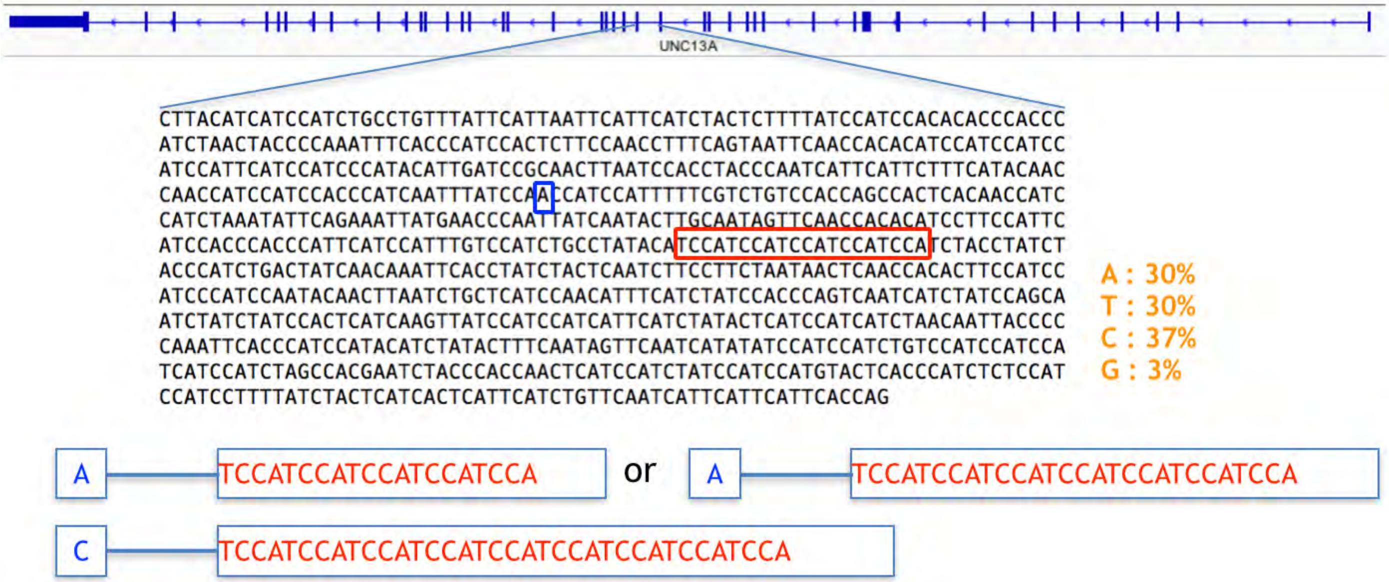

UNC13A
CTTACATCATCCATCTGCCTGTTTATTCATTAATTCATTCATCTACTCTTTTATCCATCCACACACCCACCC
ATCTAACTACCCCAAATTTCACCCATCCACTCTTCCAACCTTTCAGTAATTCAACCACACATCCATCCATCC
ATCCATTCATCCATCCCATACATTGATCCGCAACTTAATCCACCTACCCAATCATTCATTCTTTCATACAAC
CAACCATCCATCCACCCATCAATTTATCCAACCATCCATTTTTCGTCTGTCCACCAGCCACTCACAACCATC
CATCTAAATATTCAGAAATTATGAACCCAATTATCAATACTTGCAATAGTTCAACCACACATCCTTCCATTC
ATCCACCCACCCATTCATCCATTTGTCCATCTGCCTATACATCCATCCATCCATCCATCCATCTACCTATCT
ACCCATCTGACTATCAACAAATTCACCTATCTACTCAATCTTCCTTCTAATAACTCAACCACACTTCCATCC
ATCCCATCCAATACAACTTAATCTGCTCATCCAACATTTCATCTATCCACCCAGTCAATCATCTATCCAGCA
ATCTATCTATCCACTCATCAAGTTATCCATCCATCATTCATCTATACTCATCCATCATCTAACAATTACCCC
CAAATTCACCCATCCATACATCTATACTTTCAATAGTTCAATCATATATCCATCCATCTGTCCATCCATCCA
TCATCCATCTAGCCACGAATCTACCCACCAACTCATCCATCTATCCATCCATGTACTCACCCATCTCTCCAT
CCATCCTTTTATCTACTCATCACTCATTCATCTGTTCAATCATTCATTCATTCACCAG
A : 30%
T : 30%
C : 37%
G : 3%
A
TCCATCCATCCATCCATCCA
or
A
TCCATCCATCCATCCATCCATCCATCCA
C
TCCATCCATCCATCCATCCATCCATCCATCCATCCA

Figure 4. There's strong linkage disequilibrium between rs12608932 and TGGA/ TCCA tandem repeats. The figure here shows the TCCA repeats (opposite strand of TGGA repeats) in consistent with human genome reference. rs12608932 is located in the middle of gene, and in perfect association with only two other SNPs (Supplementary Table 5). TCCA tandem repeats (red rectangle) and

rs12608932 (blue rectangle) are further flanked by TCCA simple repeats. And allele A is linked with either 5 or 7 copies of TCCA, while risk allele C is always linked with 9 copies of TCCA.

Linkage disequilibrium block analysis (data not shown) based on 1000 Genome Project data by Haploview showed region surrounding rs12608932 lacks strong linkage disequilibrium. Only two neighboring SNPs achieved an r^2 score >= 0.8 with rs12608932 and they are all within middle of intron without obvious function (Supplementary Table 5). We sequence several neighboring SNPs in control and ALS samples and none of them achieved a higher odds ratio than rs12608932. This strengthens the likelihood of tetranucleotide repeats as causal variants.

Most tandem repeats studied in literature are flanked by unique sequences. However, in our case we found the TGGA tetranucleotide tandem repeats and rs12608932 are further flanked by larger TGGA simple sequence cluster, which are annotated as aggregation of closely spaced smaller TGGA simple sequence region by RepeatMasker (Figure 4). The simple sequence cluster in intron-19 is around 1.4 kb long, with only 3% sequences as guanine. Interestingly there are three introns within UNC13A containing such TGGA simple sequence cluster (Supplementary Figure 3). We then tried to search such TGGA/TCCA simple sequence clusters genome-wide (See methods). We identified 640 such cluster, with 350 of them located within 297 genes. Gene enrichment analysis by GREAT (81) showed these 297 genes are enriched for channel and membrane genes (Supplementary Table 6). This may indicate the specific role of such TGGA simple sequence cluster in neuronal genes and functions.

Chapter 3.5 Potential biophysical properties of TGGA repeats.

Then what’s the biological function of such repeats? The intron containing TGGA repeats are highly conserved among primates, and conserved to some extent among other mammals according to UCSC Genome Browser. Surprisingly, mouse UNC13A gene also contains similar repetitive sequences in the corresponding intron (defined by two conserved adjacent exons).

We then hypothesized that such repeats many have distinct functions due to its unique repetitive nature, for example, TGGA repeats serve as binding motif of certain protein or protein complex. We searched all possible online ChIP-seq or RIP-seq database such as ENCODE Project database (86), however, we didn’t find any potential transcription factor or splicing factor that specifically bind such sequences. Then we switched to Epigenome Roadmap Project database (87), and interestingly discovered two ChIP-seq datasets where TGGA simple sequence clusters are specifically bound. One is CBP (Creb-binding protein) in K562 line (Supplementary Figure 4) and the other is H3K56ac and H3K23me2 in H1 line. Both H3K56ac and H3K23me2 are strongly associated with DNA replication, damage and repair processes.

Also, it's reasonable to hypothesize the possible G-quadruplex structures for TGGA repeats due to the two consecutive guanines in the repeats. This can be easily predicted by G-quadruplex predicting software such as QGRS (79). We then tested both 5-copy and 9-copy IDT oligos (~200bp) in circular dichroism and indeed observed curve indicative of G-quadruplex and addition of Potassium could further induce G-quadruplex structures (79). Also, 9-copy sequence might have a stronger G-quadruplex structure than 5-copy (Figure 5).

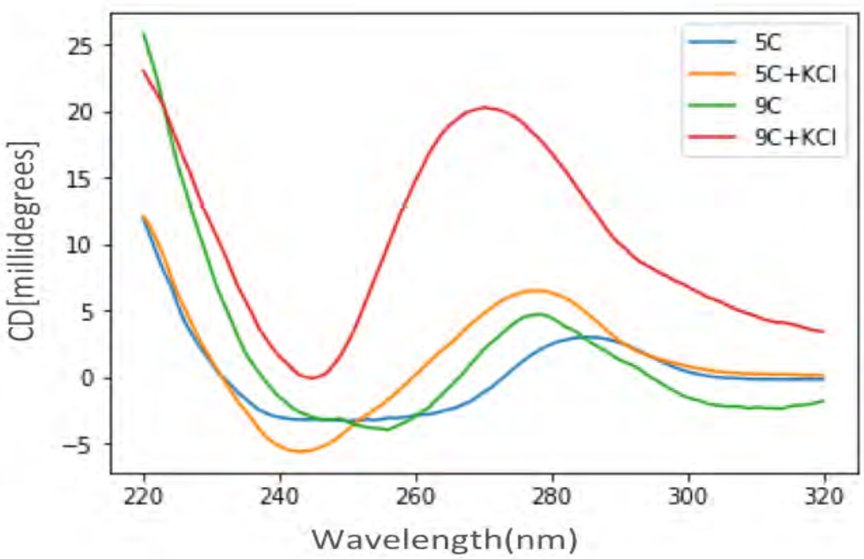


Figure 5. Circular dichroism on 200bp oligos verified the G-quadruplex structures in vitro. Addition of KCl cause a slight leftward shift of peaks for both 5-copy and 9-copy oligos, but also cause a big increase in CD value of 9-copy oligos, suggesting the presence of a strong G-quadruplex structure.
(5C : 5-copy oligo, 9C : 9-copy oligo)

Chapter 3.6 Influence of rs12608932 on UNC13A gene expression

We are interested in if rs12608932 haplotype (including TGGA tetra nucleotide polymorphism) may affect UNC13A gene expression. We first constructed two clones where we inserted each rs12608932 haplotype (the whole intron-19 plus partial flanking exon) into the dual luciferase system. We then transfected the plasmid into Hela cell, and later measured mRNA expression. We observed the splicing efficiency in risk-allele plasmid is decreased significantly (data not shown). However, we later found cloning is very challenging and cannot guarantee the unstable repeats sequences are always correct in our construct thus we don't trust the results of this luciferase assay.

We then started to test if rs12608932 affects UNC13A expression in neuronal tissues. We genotyped and cut 44 human cerebellums homozygous for rs12608932 non-risk allele (AA) and 24 cerebellums homozygous for risk allele (CC), prepared RNA and measured UNC13A expression level by qPCR using

Taqman probe. We found that UNC13A expression is slightly yet significantly reduced in risk-allele samples (Figure 6). We pulled out cerebellum RNA expression data from Braineac project and found their data shared similar trend though not significant. Interestingly, Braineac data (88) identified a significant association between rs12608932 risk allele and expression of KCNN1, a voltage-independent calcium-activated potassium channel gene. We also used the same sets of samples to compare splicing efficiency between AA and CC genotypes, but found no significant difference (data not shown). Then in order to test the loss-of-function hypothesis of UNC13A in ALS, we acquired UNC13A knockout mice from Professor Nils Brose, crossed with SOD1-G93A mice and discovered that UNC13A knockout could lead to a slight yet significant shorter survival of ALS mice (Figure 7).

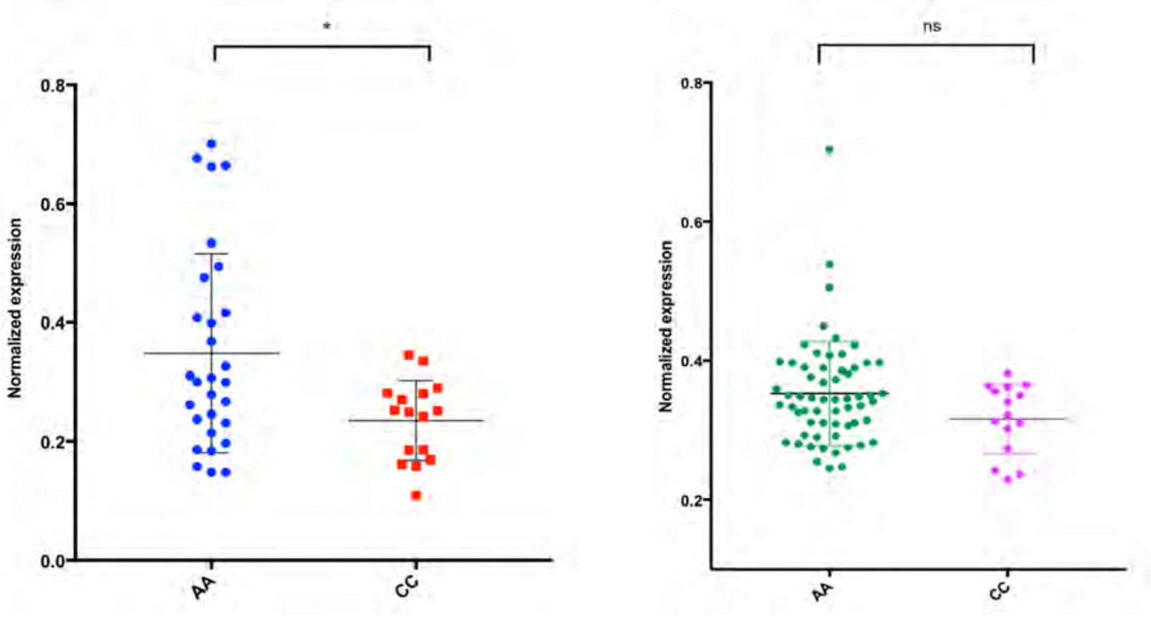


Figure 6. Left panel: Homozygous risk allele(CC) of rs12608932 is significantly yet slightly associated with reduced overall expression of UNC13A in cerebellums (T-test p-value = 0.013). Here UNC13A expression level is normalized by GAPDH. Right panel: Cerebellum data pulled from Braineac Project show the same trend but not significant (T-test P-value = 0.08). The RNA expression data from Braineac is based on exon array platform and already normalized.

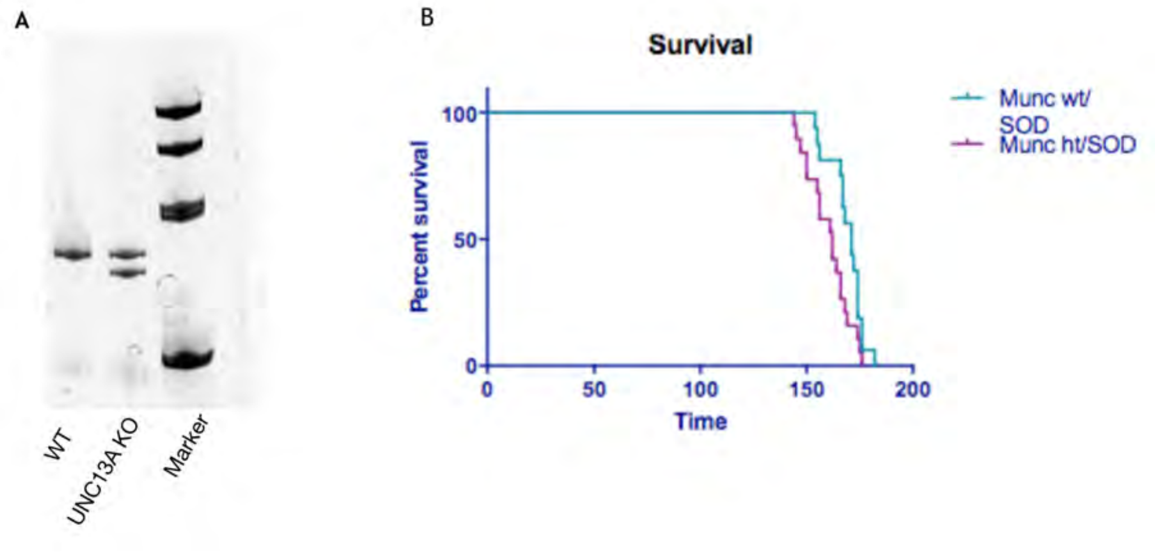


Figure 7. (A) Genotyping of UNC13A knockout mice. Double band indicates one copy of the gene has been knocked out. (B) UNC13A knockout SOD1 ALS mice show a slightly yet significantly shorter survival compared to the wild-type. (Log rank test p-value = 0.023)

Chapter 4. DISCUSSIONS

Genome-wide association study (GWAS) has identified DPP6 as a candidate ALS gene for sporadic patients. So other groups have already done sequencing for DPP6 but barely find any mutation except for one missense mutation from a sporadic patient. Here we carried out whole-genome sequencing and discovered two missense mutations from a Mexican ALS family. V343E is totally novel while A716V has an extremely low all frequency at 0.02%. Both mutations are predicted deleterious bioinformatically. V343E is upon an N-linked glysosylation sequon and possibly affect glycosylation efficiency. Given the mutation database we used for filter do not contain information for Mestizo Mexicans, we screened Mestizo controls and didn't find either mutation. We then set out to study the biological effect of these two mutations. For mutation V343E, the electrophysiological data is quite consistent with membrane localization disruption, both suggesting poor expression of DPP6 caused by the loss-of-function effect of V343E. As for A716V, the membrane location is not affected, indicating other mechanism or function contribute to the depolarizing shift in electrophysiological study. According to our knowledge, this is the first report of discovering DPP6 nonsynonymous mutations from familial ALS family and showing biological effects of the mutations.

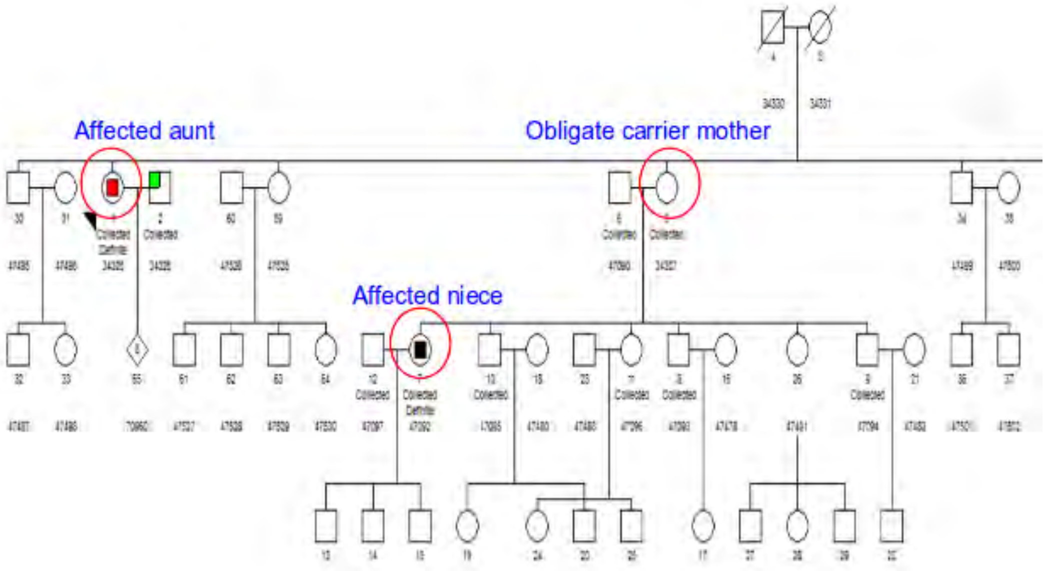
But obviously we cannot conclude such DPP6 mutations directly cause ALS in this family pedigree because first, still multiple nonsynonymous mutations were discovered from whole-genome sequencing and we cannot rule out the possibility that other mutations cause or contribute to the disease; second, this Mexican family pedigree has one obligate carrier mother, indicating the penetrance of disease-causing mutations is not complete. However, when we set out to look for more DPP6 mutations in familial ALS patients, we couldn’t find any more just like other groups. This may indicate such ALS-related DPP6 mutations are extremely rare, or play a minor genetics role contributing to the disease. There’s another explanation that DPP6 mutation serves as a genetics factor specific for Mexican population. We may need to sequence more patient samples in order to look for more evidence for DPP6 in ALS.

UNC13A is an extremely attractive candidate for sporadic ALS genetics, because it’s one of the very few GWAS signals that could be replicated in joint studies and also associated with patient survival in addition to susceptibility. The discovery of TGGA tandem repeats is very encouraging because: 1. the repeat is located within a LD-lacking genomic region, but in almost perfect linkage equilibrium with GWAS SNP interestingly only in Caucasian population; 2. we’ve already learned a huge lesson from C9ORF72 story that repetitive sequences could be the real cause to explain ALS GWAS signal. Then the huge question is what’s the biological function of such repeats?

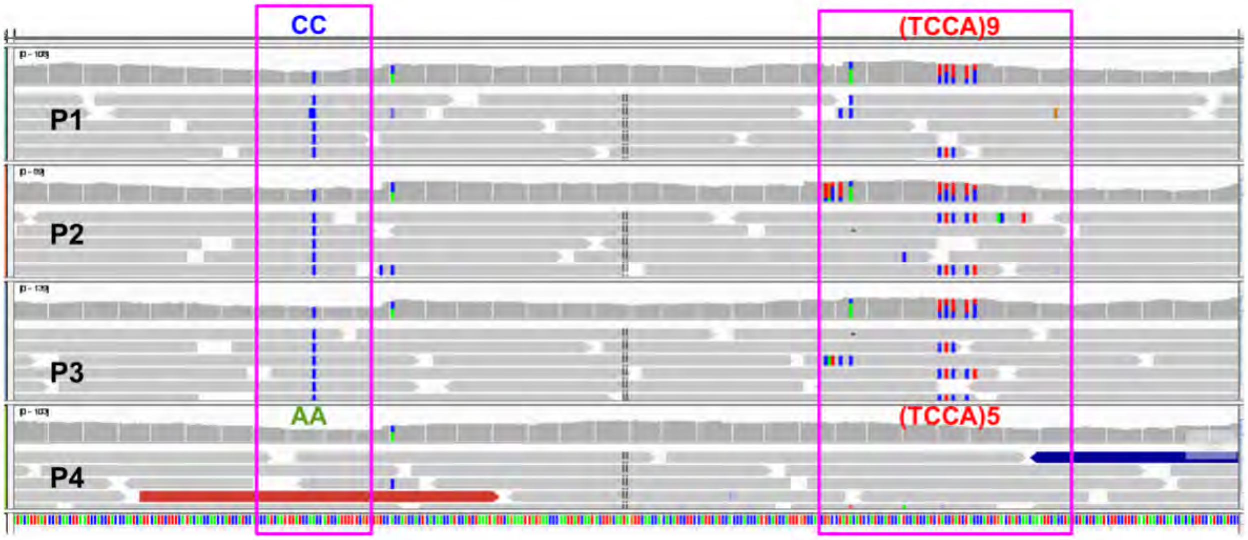
We've accumulated the following direct or indirect evidences or observations about the TGGA repeats: First, such sequences are well conserved and show up three times in UNC13A introns and carried by lots of channel-related genes, suggesting functionality of the repeats possibly in neuroscience and neurology. Second, the potential G-quadruplex with sequences resembling telomere sequences, the super unstable sequence nature, the possible binding to chromatin protein and H3K56ac (DNA damage histone marker) all suggest a possible role of TGGA large repeats in epigenetic level regulation, especially DNA damage-related processes. However, here we only verified potential G-quadruplex in vitro, and speculated super unstable nature of TGGA repeats from our failure of molecular cloning, and still need to verify the reliability of histone ChIP-seq data. So lots of work should be done further to study the potential biological function of TGGA repeats here.

Another perspective of function study is, regardless of what TGGA is doing, we could simply measure if UNC13A gene expression is different given two genotypes (eQTL study). We first tried to insert the sequence into luciferase for a reporter system, but molecular cloning of this whole-length TGGA repeats turned out extremely challenging. But what's encouraging is we see a slight yet significant reduction in UNC13A expression in cerebellums homozygous for risk alleles. Most importantly, this trend is consistent with our later mice work that UNC13A knockout mice have shorter survival compared to control. These data suggest that UNC13A could play a loss-of-function role in affecting sporadic ALS patient survival. However, all the above need to be further verified, and the followup work is pathology study for our mice, for instance, to compare the difference for ventral horn neuron count and innervation of neuromuscular junction between UNC13A knockout and control mice.

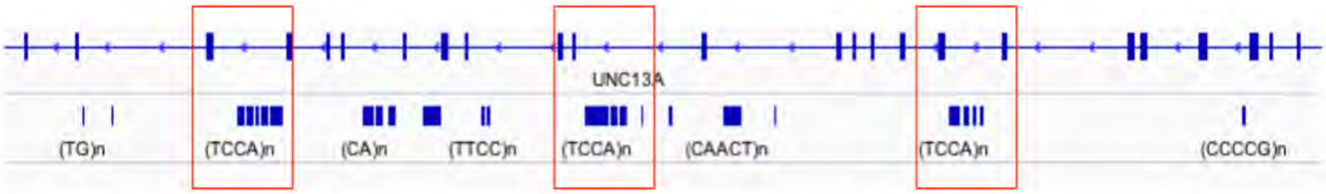
Appendices



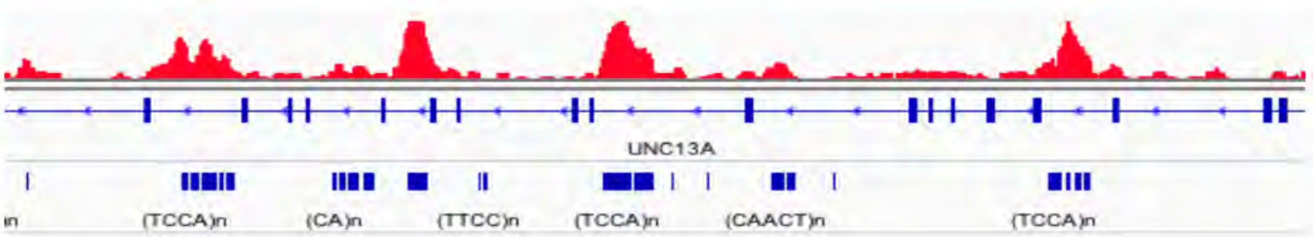
Supplementary Figure 1. Pedigree for the Mexican Mestizo ALS family, including patients of aunt and niece relationship and a obligate carrier mother.



Supplementary Figure 2. Example of haplotype inference from whole-genome sequencing. The pile-up of 100bp NGS sequences of four ALS patients (P1, P2, P3 and P4) were aligned to human reference genome and shown in IGV (Integrative Genomic Viewer). The colors for CC individual at TCCA sites are caused by misalignment of 9-copy TCCA reads onto 5-copy reference genome.



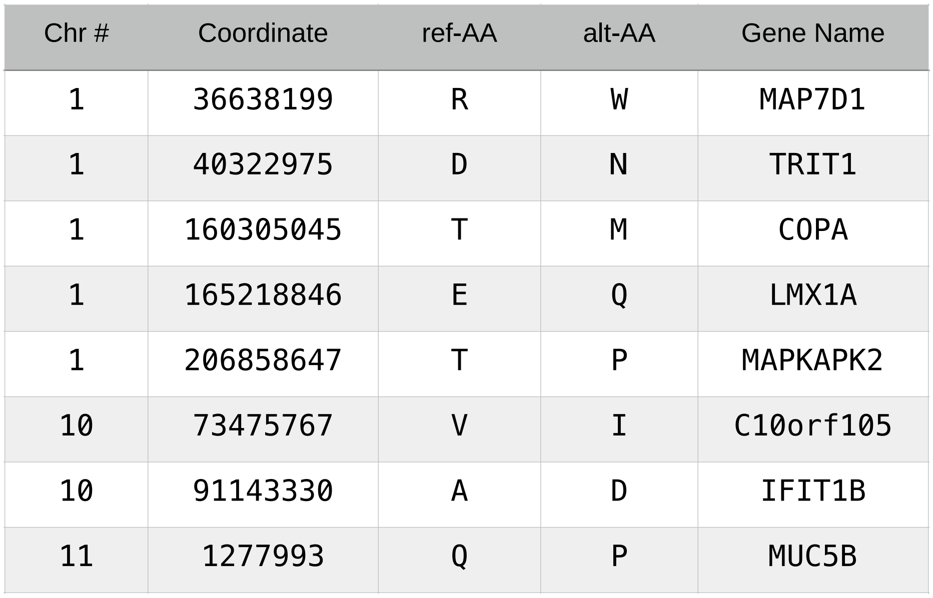
Supplementary Figure 3. There are two additional similar TGGA repeats. The upper part is gene structure of UNC13A including multiple exons and introns, and the lower part is the repeats annotation by RepeatMasker in the corresponding genomic region.



Supplementary Figure 4. ChIP-seq of Creb-binding protein in K562 lines. TGGA/ TCCA repeats are specifically bound by this chromatin regulator. For the two ChIP-seq datasets, reads are uniquely mapped for the repeats despite the repetitive nature of the sequence.

UNC13A-ssODN-5copy	ATGGGATGGATGGAAGTGTGGTTGAGTTATTAGAAGGAAG ATTGAGTAGATAGGTGAATTTGTTGATAGTCAGATGGGTAG ATAGGTAGATGGATGGATGGATGGATGGATGTATAGGCAGA TGGACAAATGGATGAATGGGTGGGTGGATGAATGGAAGGA TGTGTGGTTGAACT
UNC13A-ssODN-9copy	ATGGGATGGATGGAAGTGTGGTTGAGTTATTAGAAGGAAG ATTGAGTAGATAGGTGAATTTGTTGATAGTCAGATGGGTAG ATAGGTAGATGGATGGATGGATGGATGGATGGATGGATGG ATGGATGTATAGGCAGATGGACAAATGGATGAATGGGTGG GTGGATGAATGGAAGGATGTGTGGTTGAACT

Supplementary Table 1. Sequence of the two Ultramer oligos used in CD experiment.



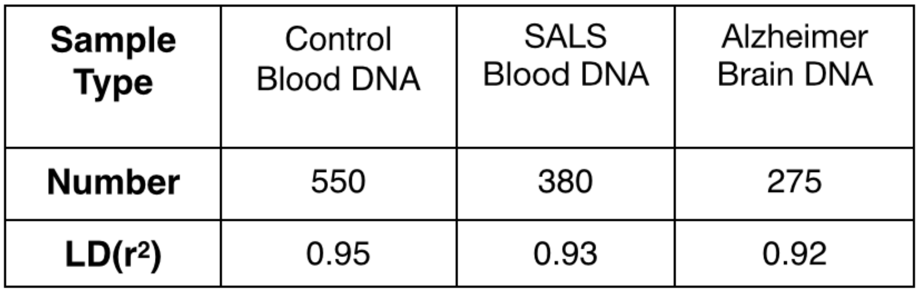
Chr #	Coordinate	ref-AA	alt-AA	Gene Name
1	36638199	R	W	MAP7D1
1	40322975	D	N	TRIT1
1	160305045	T	M	COPA
1	165218846	E	Q	LMX1A
1	206858647	T	P	MAPKAPK2
10	73475767	V	I	C10orf105
10	91143330	A	D	IFIT1B
11	1277993	Q	P	MUC5B

Chr #	Coordinate	ref-AA	alt-AA	Gene Name
11	3242950	L	S	C11orf36
11	21581854	H	Y	NELL1
11	62292219	L	M	AHNAK
11	66114821	A	T	B3GNT1
11	66473307	G	D	SPTBN2
11	124180278	P	S	OR8D1
12	55794446	M	T	OR6C65
12	56642623	D	N	ANKRD52
12	97073483	I	T	C12orf63
12	131456080	Y	D	GPR133
13	102047650	M	V	NALCN
14	67664955	P	L	FAM71D
14	73640432	R	K	PSEN1
16	10783873	E	K	TEKT5
16	30980680	P	L	SETD1A
16	31150508	P	L	PRSS36
16	58075631	G	S	MMP15
17	37099080	V	A	FBXO47
17	43318948	P	R	FMNL1
17	73620469	L	R	MYO15B
17	77705134	C	S	ENPP7
17	81006592	D	N	B3GNTL1
18	18975500	D	E	GREB1L

Chr #	Coordinate	ref-AA	alt-AA	Gene Name
19	1754783	E	D	ONECUT3
19	2226285	K	N	DOT1L
19	6707282	G	S	C3
19	18700492	T	M	C19orf60
19	36530245	R	C	THAP8
19	51607669	V	A	CTU1
19	51815108	P	A	IGLON5
19	52090222	G	V	ZNF175
19	52272549	P	L	FPR2
19	58234590	A	V	ZNF671
2	131520942	P	A	FAM123C
2	152470809	A	V	NEB
2	183095749	R	H	PDE1A
2	183291314	P	L	PDE1A
2	202412312	E	D	ALS2CR11
2	228144563	G	E	COL4A3
2	237276914	R	H	IQCA1
2	242674703	G	R	D2HGDH
21	45845642	R	W	TRPM2
21	45953710	R	C	TSPEAR
21	47666562	V	A	MCM3AP
22	36902393	S	L	FOXRED2
3	130159607	I	T	COL6A5

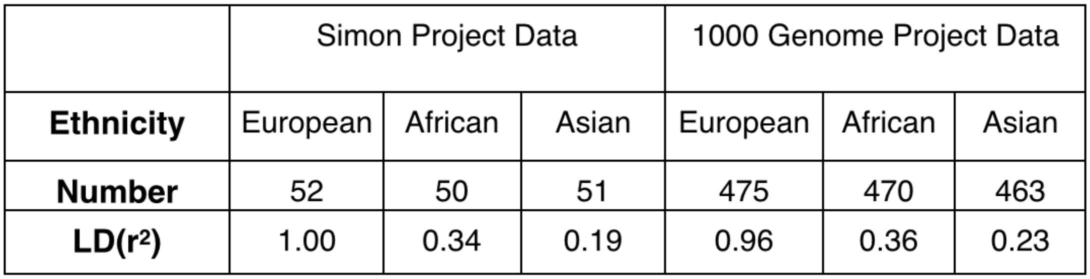
Chr #	Coordinate	ref-AA	alt-AA	Gene Name
5	80409656	E	G	RASGRF2
6	26056620	P	S	HIST1H1C
6	41774685	A	P	USP49
6	117246727	T	K	RFX6
7	81714123	V	G	CACNA2D1
7	126173250	R	Q	GRM8
7	140221738	R	H	DENND2A
7	154585866	V	E	DPP6
7	154681010	A	V	DPP6
8	21768204	R	W	DOK2
8	42693170	V	I	THAP1
8	144943082	A	V	EPPK1
8	145608403	V	L	ADCK5
9	13121859	V	L	MPDZ
9	78796352	A	V	PCSK5
9	88937978	D	G	ZCCHC6
X	2407163	M	T	ZBED1
X	16965094	C	Y	REPS2

Supplementary Table 2. List of mutations identified from DPP6 NGS bioinformatics analysis.



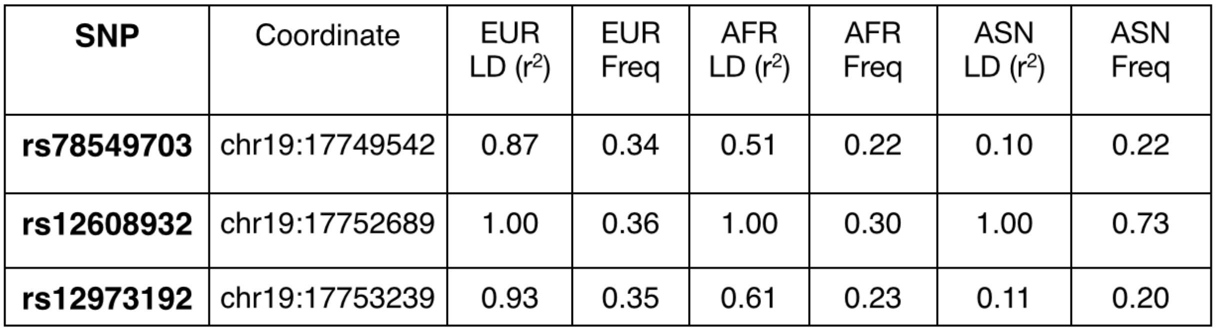
Sample Type	Control Blood DNA	SALS Blood DNA	Alzheimer Brain DNA
Number	550	380	275
LD(r^2)	0.95	0.93	0.92

Supplementary Table 3. Information of samples used for Sanger sequencing to verify LD. Linkage disequilibrium score (r^2) between rs12608932 and TGGA tandem repeats based on Sanger sequencing data for different samples. All samples are from Caucasian population. We could see rs1260892 is strongly linked with TGGA repeats.



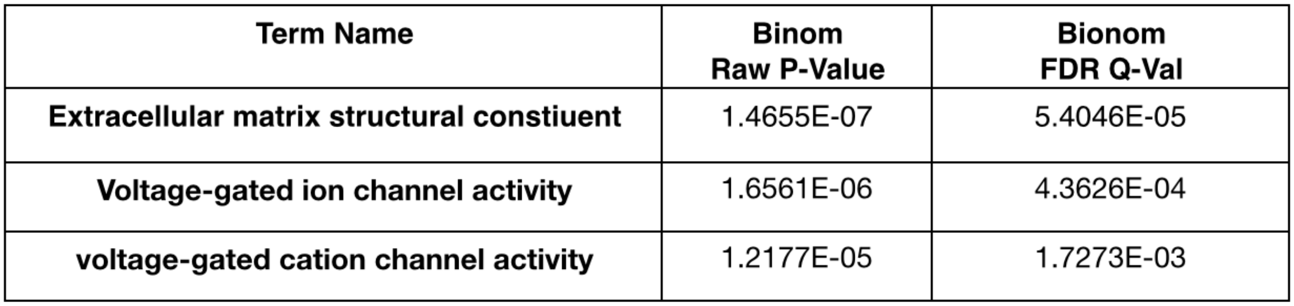
	Simon Project Data			1000 Genome Project Data		
Ethnicity	European	African	Asian	European	African	Asian
Number	52	50	51	475	470	463
LD(r^2)	1.00	0.34	0.19	0.96	0.36	0.23

Supplementary Table 4. LD score (r^2) for three ethnicities based on next-generation sequencing data from both Simon Project (high-depth) and 1000 Genome project (low-depth). We could see such strong LD only exists for Caucasian population.



SNP	Coordinate	EUR LD (r^2)	EUR Freq	AFR LD (r^2)	AFR Freq	ASN LD (r^2)	ASN Freq
rs78549703	chr19:17749542	0.87	0.34	0.51	0.22	0.10	0.22
rs12608932	chr19:17752689	1.00	0.36	1.00	0.30	1.00	0.73
rs12973192	chr19:17753239	0.93	0.35	0.61	0.23	0.11	0.20

Supplementary Table 5. Only two SNP rs78549703 and rs12973192 are in strong LD (r^2 >= 0.8) with rs12608932. According to 1000 Genome Project data, LD score and minor allele frequency of these three SNPs are shown here (EUR: European, ASN : Asian, AFR : African).



Term Name	Binom Raw P-Value	Bionom FDR Q-Val
Extracellular matrix structural constiuent	1.4655E-07	5.4046E-05
Voltage-gated ion channel activity	1.6561E-06	4.3626E-04
voltage-gated cation channel activity	1.2177E-05	1.7273E-03

Supplementary Table 6. Gene enrichment analysis by GREAT (82), which first annotates noncoding genomic region and then calculates statistical enrichments for association between the genomic region and annotation.

BIBLIOGRAPHY

1. Ferraiuolo, L., Kirby, J., Grierson, A.J., Sendtner, M. & Shaw, P.J. Molecular pathways of motor neuron injury in amyotrophic lateral sclerosis. Nat Rev Neurol. 2011 Nov;7(11):616-30.

2. Morris, H.R., Waite, A.J., Williams, N.M., Neal, J.W. & Blake, D.J. Recent Advances in the Genetics of the ALS-FTLD Complex. Curr Neurol Neurosci Rep. 2012 Jun;12(3):243-50.

3. Orr, H.T. FTD and ALS: genetic ties that bind. Neuron. 2011 Oct 20;72(2):189-90.

4. King, O.D., Gitler, A.D. & Shorter, J. The tip of the iceberg : RNA-binding proteins with prion-like domains in neurodegenerative disease. Brain Res. 2012 Jun 26;1462:61-80.

5. Polymenidou M, Lagier-Tourenne C, Hutt KR, Bennett CF, Cleveland DW, Yeo GW. Misregulated RNA processing in amyotrophic lateral sclerosis. Brain Res. 2012 Jun 26;1462:3-15.

6. Haidet-Phillips AM, Hester ME, Miranda CJ, Meyer K, Braun L, Frakes A, Song S, Likhite S, Murtha MJ, Foust KD, Rao M, Eagle A, Kammesheidt A, Christensen A, Mendell JR, Burghes AH, Kaspar BK. Astrocytes from familial and sporadic ALS patients are toxic to motor neurons. Nat Biotechnol. 2011 Aug 10;29(9):824-8.

7. Veldink, J.H., Van den Berg, L.H. & Wokk e, J.H.J. The future of motor neuron disease: the challenge is in the genes. J Neurol. 2004 Apr;251(4):491-500.

8.Neumann M, Sampathu DM, Kwong LK, Truax AC, Micsenyi MC, Chou TT, Bruce J, Schuck T, Grossman M, Clark CM, McCluskey LF, Miller BL, Masliah E, Mackenzie IR, Feldman H, Feiden W, Kretzschmar HA, Trojanowski JQ, Lee VM. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science. 2006 Oct 6;314(5796):130-3.

9. Sreedharan J, Blair IP, Tripathi VB, Hu X, Vance C, Rogelj B, Ackerley S, Durnall JC, Williams KL, Buratti E, Baralle F, de Belleroche J, Mitchell JD, Leigh PN, Al-Chalabi A, Miller CC, Nicholson G, Shaw CE. TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. Science. 2008 Mar 21;319(5870): 1668-72.

10. Tollervey JR, Curk T, Rogelj B, Briese M, Cereda M, Kayikci M, König J, Hortobágyi T, Nishimura AL, Zupunski V, Patani R, Chandran S, Rot G, Zupan B, Shaw CE, Ule J. Characterizing the RNA targets and position- dependent splicing regulation by TDP-43. Nat Neurosci. 2011 Apr;14(4):452-8.

11. Mougeot, JL, Richardson-milazi, S. & Brooks, B.R. Whole-genome association studies of sporadic amyotrophic lateral sclerosis : are retroelements involved ? Trends Mol Med. 2009 Apr;15(4):148-58.

12. Vance C, Rogelj B, Hortobágyi T, De Vos KJ, Nishimura AL, Sreedharan J, Hu X, Smith B, Ruddy D, Wright P, Ganesalingam J, Williams KL, Tripathi V, Al-Saraj S, Al-Chalabi A, Leigh PN, Blair IP, Nicholson G, de Belleroche J, Gallo JM, Miller CC, Shaw CE. Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. Science. 2009 Feb 27;323(5918): 1208-1211.

13. Elden AC, Kim HJ, Hart MP, Chen-Plotkin AS, Johnson BS, Fang X, Armakola M, Geser F, Greene R, Lu MM, Padmanabhan A, Clay-Falcone D, McCluskey L, Elman L, Juhr D, Gruber PJ, Rüb U, Auburger G, Trojanowski JQ, Lee VM, Van Deerlin VM, Bonini NM, Gitler AD. Ataxin-2 intermediate-length polyglutamine

expansions are associated with increased risk for ALS. Nature. 2010 Aug 26;466(7310):1069-75.

14. Deng HX, Chen W, Hong ST, Boycott KM, Gorrie GH, Siddique N, Yang Y, Fecto F, Shi Y, Zhai H, Jiang H, Hirano M, Rampersaud E, Jansen GH, Donkervoort S, Bigio EH, Brooks BR, Ajroud K, Sufit RL, Haines JL, Mugnaini E, Pericak-Vance MA, Siddique T. Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia. Nature. 2011 Aug 21;477(7363): 211-5.

15. DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, Nicholson AM, Finch NA, Flynn H, Adamson J, Kouri N, Wojtas A, Sengdy P, Hsiung GY, Karydas A, Seeley WW, Josephs KA, Coppola G, Geschwind DH, Wszolek ZK, Feldman H, Knopman DS, Petersen RC, Miller BL, Dickson DW, Boylan KB, Graff-Radford NR, Rademakers R. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron. 2011 Oct 20;72(2):245-56.

16. Renton AE1, Majounie E, Waite A, Simón-Sánchez J, Rollinson S, Gibbs JR, Schymick JC, Laaksovirta H, van Swieten JC, Myllykangas L, Kalimo H, Paetau A, Abramzon Y, Remes AM, Kaganovich A, Scholz SW, Duckworth J, Ding J, Harmer DW, Hernandez DG, Johnson JO, Mok K, Ryten M, Trabzuni D, Guerreiro RJ, Orrell RW, Neal J, Murray A, Pearson J, Jansen IE, Sondervan D, Seelaar H, Blake D, Young K, Halliwell N, Callister JB, Toulson G, Richardson A, Gerhard A, Snowden J, Mann D, Neary D, Nalls MA, Peuralinna T, Jansson L, Isoviita VM, Kaivorinne AL, Hölttä-Vuori M, Ikonen E, Sulkava R, Benatar M, Wuu J, Chiò A, Restagno G, Borghero G, Sabatelli M; ITALSGEN Consortium, Heckerman D, Rogaeva E, Zinman L, Rothstein JD, Sendtner M, Drepper C, Eichler EE, Alkan C, Abdullaev Z, Pack SD, Dutra A, Pak E, Hardy J, Singleton A, Williams NM, Heutink P, Pickering-Brown S, Morris HR, Tienari PJ, Traynor BJ. A

Hexanucleotide Repeat Expansion in C9ORF72 Is the Cause of Chromosome 9p21-Linked ALS-FTD. Neuron. 2011 Oct 20;72(2):257-68.

17. Gijselinck I, Van Langenhove T, van der Zee J, Sleegers K, Philtjens S, Kleinberger G, Janssens J, Bettens K, Van Cauwenberghe C, Pereson S, Engelborghs S, Sieben A, De Jonghe P, Vandenberghe R, Santens P, De Bleecker J, Maes G, Bäumer V, Dillen L, Joris G, Cuijt I, Corsmit E, Elinck E, Van Dongen J, Vermeulen S, Van den Broeck M, Vaerenberg C, Mattheijssens M, Peeters K, Robberecht W, Cras P, Martin JJ, De Deyn PP, Cruts M, Van Broeckhoven C. A C9orf72 promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum: a gene identification study. Lancet Neurol. 2012 Jan;11(1):54-65.

18. Byrne S, Elamin M, Bede P, Shatunov A, Walsh C, Corr B, Heverin M, Jordan N, Kenna K, Lynch C, McLaughlin RL, Iyer PM, O'Brien C, Phukan J, Wynne B, Bokde AL, Bradley DG, Pender N, Al-Chalabi A, Hardiman O. Cognitive and clinical characteristics of patients with amyotrophic lateral sclerosis carrying a C9orf72 repeat expansion: a population-based cohort study. Lancet Neurol. 2012 Mar; 11(3):232-40.

19. Al-Chalabi A, Fang F, Hanby MF, Leigh PN, Shaw CE, Ye W, Rijsdijk F. An estimate of amyotrophic lateral sclerosis heritability using twin data. J Neurol Neurosurg Psychiatry. 2010 Dec;81(12):1324-6.

20. Dunckley T, Huentelman MJ, Craig DW, Pearson JV, Szelinger S, Joshipura K, Halperin RF, Stamper C, Jensen KR, Letizia D, Hesterlee SE, Pestronk A, Levine T, Bertorini T, Graves MC, Mozaffar T, Jackson CE, Bosch P, McVey A, Dick A, Barohn R, Lomen-Hoerth C, Rosenfeld J, O'connor DT, Zhang K, Crook R, Ryberg H, Hutton M, Katz J, Simpson EP, Mitsumoto H, Bowser R, Miller RG, Appel SH, Stephan DA. Whole-Genome Analysis of Sporadic Amyotrophic Lateral Sclerosis. N Engl J Med. 2007 Aug 23;357(8):775-88.

21. van Es MA, Van Vught PW, Blauw HM, Franke L, Saris CG, Andersen PM, Van Den Bosch L, de Jong SW, van 't Slot R, Birve A, Lemmens R, de Jong V, Baas F, Schelhaas HJ, Sleegers K, Van Broeckhoven C, Wokke JH, Wijmenga C, Robberecht W, Veldink JH, Ophoff RA, van den Berg LH. ITPR2 as a susceptibility gene in sporadic amyotrophic lateral sclerosis : a genome-wide association study. Lancet Neurol. 2007 Oct;6(10):869-77.

22. van Es MA, van Vught PW, Blauw HM, Franke L, Saris CG, Van den Bosch L, de Jong SW, de Jong V, Baas F, van't Slot R, Lemmens R, Schelhaas HJ, Birve A, Sleegers K, Van Broeckhoven C, Schymick JC, Traynor BJ, Wokke JH, Wijmenga C, Robberecht W, Andersen PM, Veldink JH, Ophoff RA, van den Berg LH. Genetic variation in DPP6 is associated with susceptibility to amyotrophic lateral sclerosis. Nat Genet. 2008 Jan;40(1):29-31.

23. Cronin S, Berger S, Ding J, Schymick JC, Washecka N, Hernandez DG, Greenway MJ, Bradley DG, Traynor BJ, Hardiman O. A genome-wide association study of sporadic ALS in a homogenous Irish population. Hum Mol Genet. 2008 Mar 1;17(5):768-74

24. Chiò A, Schymick JC, Restagno G, Scholz SW, Lombardo F, Lai SL, Mora G, Fung HC, Britton A, Arepalli S, Gibbs JR, Nalls M, Berger S, Kwee LC, Oddone EZ, Ding J, Crews C, Rafferty I, Washecka N, Hernandez D, Ferrucci L, Bandinelli S, Guralnik J, Macciardi F, Torri F, Lupoli S, Chanock SJ, Thomas G, Hunter DJ, Gieger C, Wichmann HE, Calvo A, Mutani R, Battistini S, Giannini F, Caponnetto C, Mancardi GL, La Bella V, Valentino F, Monsurrò MR, Tedeschi G, Marinou K, Sabatelli M, Conte A, Mandrioli J, Sola P, Salvi F, Bartolomei I, Siciliano G, Carlesi C, Orrell RW, Talbot K, Simmons Z, Connor J, Pioro EP, Dunkley T, Stephan DA, Kasperaviciute D, Fisher EM, Jabonka S, Sendtner M, Beck M, Bruijn L, Rothstein J, Schmidt S, Singleton A, Hardy J, Traynor BJ. A two-stage

genome-wide association study of sporadic amyotrophic lateral sclerosis. Hum Mol Genet. 2009 Apr 15;18(8):1524-32.

25. van Es MA, Veldink JH, Saris CG, Blauw HM, van Vught PW, Birve A, Lemmens R, Schelhaas HJ, Groen EJ, Huisman MH, van der Kooi AJ, de Visser M, Dahlberg C, Estrada K, Rivadeneira F, Hofman A, Zwarts MJ, van Doormaal PT, Rujescu D, Strengman E, Giegling I, Muglia P, Tomik B, Slowik A, Uitterlinden AG, Hendrich C, Waibel S, Meyer T, Ludolph AC, Glass JD, Purcell S, Cichon S, Nöthen MM, Wichmann HE, Schreiber S, Vermeulen SH, Kiemeney LA, Wokke JH, Cronin S, McLaughlin RL, Hardiman O, Fumoto K, Pasterkamp RJ, Meininger V, Melki J, Leigh PN, Shaw CE, Landers JE, Al-Chalabi A, Brown RH Jr, Robberecht W, Andersen PM, Ophoff RA, van den Berg LH. Genome-wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. Nat Genet. 2009 Oct;41(10):1083-7.

26. Landers JE, Melki J, Meininger V, Glass JD, van den Berg LH, van Es MA, Sapp PC, van Vught PW, McKenna-Yasek DM, Blauw HM, Cho TJ, Polak M, Shi L, Wills AM, Broom WJ, Ticozzi N, Silani V, Ozoguz A, Rodriguez-Leyva I, Veldink JH, Ivinson AJ, Saris CG, Hosler BA, Barnes-Nessa A, Couture N, Wokke JH, Kwiatkowski TJ Jr, Ophoff RA, Cronin S, Hardiman O, Diekstra FP, Leigh PN, Shaw CE, Simpson CL, Hansen VK, Powell JF, Corcia P, Salachas F, Heath S, Galan P, Georges F, Horvitz HR, Lathrop M, Purcell S, Al-Chalabi A, Brown RH Jr. Reduced expression of the Kinesin-Associated Protein 3 (KIFAP3) gene increases survival in sporadic amyotrophic lateral sclerosis. Proc Natl Acad Sci U S A. 2009 Jun 2;106(22):9004-9.

27. Laaksovirta H, Peuralinna T, Schymick JC, Scholz SW, Lai SL, Myllykangas L, Sulkava R, Jansson L, Hernandez DG, Gibbs JR, Nalls MA, Heckerman D, Tienari PJ, Traynor BJ. Chromosome 9p21 in amyotrophic lateral sclerosis in Finland : a genome-wide association study. Lancet Neurol. 2010 Oct;9(10):978-85.

28. Shatunov A, Mok K, Newhouse S, Weale ME, Smith B, Vance C, Johnson L, Veldink JH, van Es MA, van den Berg LH, Robberecht W, Van Damme P, Hardiman O, Farmer AE, Lewis CM, Butler AW, Abel O, Andersen PM, Fogh I, Silani V, Chiò A, Traynor BJ, Melki J, Meininger V, Landers JE, McGuffin P, Glass JD, Pall H, Leigh PN, Hardy J, Brown RH Jr, Powell JF, Orrell RW, Morrison KE, Shaw PJ, Shaw CE, Al-Chalabi A. Chromosome 9p21 in sporadic amyotrophic lateral sclerosis in the UK and seven other countries : a genome-wide association study. Lancet Neurol. 2010 Oct;9(10):986-94

29. Renton AE, Chiò A, Traynor BJ. State of play in amyotrophic lateral sclerosis genetics. Nat Neurosci. 2014 Jan;17(1):17-23.

30. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH. Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet. 2010 Jun;11(6):446-50.

31. Cirulli, E.T. & Goldstein, D.B. Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet. 2010 Jun;11(6):415-25.

32. Gibson, G. Rare and common variants : twenty arguments. Nat Rev Genet. 2012 Jan 18;13(2):135-45

33. McClellan, J. & King, MC. Genetic heterogeneity in human disease. Cell. 2010 Apr 16;141(2):210-7.

34. Thomas, D. Gene – environment-wide association studies : emerging approaches. Nat Rev Genet. 2010 Apr;11(4):259-72.

35. Slatkin, M. Epigenetic Inheritance and the Missing Heritability Problem. Genetics. 2009 Jul;182(3):845-50.

36. Cordell, H.J. Detecting gene – gene interactions that underlie human diseases. Nat Rev Genet. 2009 Jun;10(6):392-404.

37. Girirajan, S., Campbell, C.D. & Eichler, E.E. Human Copy Number Variation and Complex Genetic Disease. Annu Rev Genet. 2011;45:203-26

38. Stankiewicz, P. & Lupski, J.R. Structural variation in the human genome and its role in disease. Annu Rev Med. 2010;61:437-55

39. Alkan, C., Coe, B.P. & Eichler, E.E. Genome structural variation discovery and genotyping. Nat Rev Genet. 2011 May;12(5):363-76

40. Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, Nord AS, Kusenda M, Malhotra D, Bhandari A, Stray SM, Rippey CF, Roccanova P, Makarov V, Lakshmi B, Findling RL, Sikich L, Stromberg T, Merriman B, Gogtay N, Butler P, Eckstrand K, Noory L, Gochman P, Long R, Chen Z, Davis S, Baker C, Eichler EE, Meltzer PS, Nelson SF, Singleton AB, Lee MK, Rapoport JL, King MC, Sebat J. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science. 2008 Apr 25;320(5875):539-43.

41. Sebat, J., Levy, D.L. & Mccarthy, S.E. Rare structural variants in schizophrenia : one disorder, multiple mutations; one mutation, multiple disorders. Trends Genet. 2009 Dec;25(12):528-35.

42. Mccarroll, S.A. Extending genome-wide association studies to copy-number variation. Hum Mol Genet. 2008 Oct 15;17(R2):R135-42.

43. Rucker JJ, Breen G, Pinto D, Pedroso I, Lewis CM, Cohen-Woods S, Uher R, Schosser A, Rivera M, Aitchison KJ, Craddock N, Owen MJ, Jones L, Jones I, Korszun A, Muglia P, Barnes MR, Preisig M, Mors O, Gill M, Maier W, Rice J, Rietschel M, Holsboer F, Farmer AE, Craig IW, Scherer SW, McGuffin P. Genome-

wide association analysis of copy number variation in recurrent depressive disorder. Mol Psychiatry. 2013 Feb;18(2):183-9.

44. Elia J, Glessner JT, Wang K, Takahashi N, Shtir CJ, Hadley D, Sleiman PM, Zhang H, Kim CE, Robison R, Lyon GJ, Flory JH, Bradfield JP, Imielinski M, Hou C, Frackelton EC, Chiavacci RM, Sakurai T, Rabin C, Middleton FA, Thomas KA, Garris M, Mentch F, Freitag CM, Steinhausen HC, Todorov AA, Reif A, Rothenberger A, Franke B, Mick EO, Roeyers H, Buitelaar J, Lesch KP, Banaschewski T, Ebstein RP, Mulas F, Oades RD, Sergeant J, Sonuga-Barke E, Renner TJ, Romanos M, Romanos J, Warnke A, Walitza S, Meyer J, Pálmason H, Seitz C, Loo SK, Smalley SL, Biederman J, Kent L, Asherson P, Anney RJ, Gaynor JW, Shaw P, Devoto M, White PS, Grant SF, Buxbaum JD, Rapoport JL, Williams NM, Nelson SF, Faraone SV, Hakonarson H. Genome-wide copy number variation study associates metabotropic glutamate receptor gene networks with attention deficit hyperactivity disorder. Nat Genet. 2011 Dec 4;44(1):78-84

45. Dauber A, Yu Y, Turchin MC, Chiang CW, Meng YA, Demerath EW, Patel SR, Rich SS, Rotter JI, Schreiner PJ, Wilson JG, Shen Y, Wu BL, Hirschhorn JN. Genome-wide Association of Copy-Number Variation Reveals an Association between Short Stature and the Presence of Low-Frequency Genomic Deletions. Am J Hum Genet. 2011 Dec 9;89(6):751-9

46. Malhotra, D. & Sebat, J. Review CNVs : Harbingers of a Rare Variant Revolution in Psychiatric Genetics. Cell. 2012 Mar 16;148(6):1223-41.

47. Blauw HM, Veldink JH, van Es MA, van Vught PW, Saris CG, van der Zwaag B, Franke L, Burbach JP, Wokke JH, Ophoff RA, van den Berg LH. Copy-number variation in sporadic amyotrophic lateral sclerosis: a genome-wide screen. Lancet Neurol. 2008 Apr;7(4):319-26.

48. Cronin S, Blauw HM, Veldink JH, van Es MA, Ophoff RA, Bradley DG, van den Berg LH, Hardiman O. Analysis of genome-wide copy number variation in Irish and Dutch ALS populations. Hum Mol Genet. 2008 Nov 1;17(21):3392-8.

49. Blauw HM, Al-Chalabi A, Andersen PM, van Vught PW, Diekstra FP, van Es MA, Saris CG, Groen EJ, van Rheenen W, Koppers M, Van't Slot R, Strengman E, Estrada K, Rivadeneira F, Hofman A, Uitterlinden AG, Kiemeney LA, Vermeulen SH, Birve A, Waibel S, Meyer T, Cronin S, McLaughlin RL, Hardiman O, Sapp PC, Tobin MD, Wain LV, Tomik B, Slowik A, Lemmens R, Rujescu D, Schulte C, Gasser T, Brown RH Jr, Landers JE, Robberecht W, Ludolph AC, Ophoff RA, Veldink JH, van den Berg LH. A large genome scan for rare CNVs in amyotrophic lateral sclerosis. Hum Mol Genet. 2010 Oct 15;19(20):4091-9.

50. Nielsen, R., Paul, J.S., Albrechtsen, A. & Song, Y.S. Genotype and SNP calling from next-generation sequencing data. Nat Rev Genet. 2011 Jun;12(6): 443-51.

51. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next- generation DNA sequencing data. Nat Genet. 2011 May;43(5):491-8

52. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. articles Exome sequencing identifies the cause of a mendelian disorder. Nat Genet. 2010 Jan; 42(1):30-5.

53. Shi Y, Li Y, Zhang D, Zhang H, Li Y, Lu F, Liu X, He F, Gong B, Cai L, Li R, Liao S, Ma S, Lin H, Cheng J, Zheng H, Shan Y, Chen B, Hu J, Jin X, Zhao P, Chen Y, Zhang Y, Lin Y, Li X, Fan Y, Yang H, Wang J, Yang Z. Exome Sequenc-

ing Identifies ZNF644 Mutations in High Myopia. PLoS Genet. 2011 Jun; 7(6):e1002084.

54. Johnson JO, Mandrioli J, Benatar M, Abramzon Y, Van Deerlin VM, Trojanowski JQ, Gibbs JR, Brunetti M, Gronka S, Wuu J, Ding J, McCluskey L, Martinez-Lage M, Falcone D, Hernandez DG, Arepalli S, Chong S, Schymick JC, Rothstein J, Landi F, Wang YD, Calvo A, Mora G, Sabatelli M, Monsurrò MR, Battistini S, Salvi F, Spataro R, Sola P, Borghero G; ITALSGEN Consortium, Galassi G, Scholz SW, Taylor JP, Restagno G, Chiò A, Traynor BJ. Exome Sequencing Reveals VCP Mutations as a Cause of Familial ALS. Neuron. 2010 Dec 9;68(5): 857-64

55. O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, Karakoc E, Mackenzie AP, Ng SB, Baker C, Rieder MJ, Nickerson DA, Bernier R, Fisher SE, Shendure J, Eichler EE. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat Genet. 2011 Jun;43(6):585-9.

56. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, Levy R, Ko A, Lee C, Smith JD, Turner EH, Stanaway IB, Vernot B, Malig M, Baker C, Reilly B, Akey JM, Borenstein E, Rieder MJ, Nickerson DA, Bernier R, Shendure J, Eichler EE. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature. 2012 Apr 4;485(7397):246-50.

57. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, Ercan-Sencicek AG, DiLullo NM, Parikshak NN, Stein JL, Walker MF, Ober GT, Teran NA, Song Y, El-Fishawy P, Murtha RC, Choi M, Overton JD, Bjornson RD, Carriero NJ, Meyer KA, Bilguvar K, Mane SM, Sestan N, Lifton RP, Günel M, Roeder K, Geschwind DH, Devlin B, State MW. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature. 2012 Apr 4;485(7397):237-41.

58. Sathirapongsasuti JF, Lee H, Horst BA, Brunner G, Cochran AJ, Binder S, Quackenbush J, Nelson SF. Exome Sequencing-Based Copy-Number Variation and Loss of Heterozygosity Detection : ExomeCNV. Bioinformatics. 2011 Oct 1;27(19):2648-54.

59. Krumm N, Sudmant PH, Ko A, O'Roak BJ, Malig M, Coe BP; NHLBI Exome Sequencing Project, Quinlan AR, Nickerson DA, Eichler EE. Copy number variation detection and genotyping from exome sequence data. Genome Res. 2012 Aug;22(8):1525-32.

60. Karakoc E, Alkan C, O'Roak BJ, Dennis MY, Vives L, Mark K, Rieder MJ, Nickerson DA, Eichler EE. Detection of structural variants and indels within exome data. Nat Methods. 2011 Dec 18;9(2):176-8.

61. Korbel JO, Abyzov A, Mu XJ, Carriero N, Cayting P, Zhang Z, Snyder M, Gerstein MB. PEMer: a computational framework with simulation-based error models for inferring genomic structural variants from massive paired-end sequencing data. Genome Biol. 2009 Feb 23;10(2):R23.

62. Abyzov A, Urban AE, Snyder M, Gerstein M. CNVnator: An approach to discover, genotype and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. 2011 Jun;21(6):974-84.

63. Nadin BM, Pfaffinger PJ. DPP6 is Required for Normal Electrophysiological Properties of Cerebellar Granule Cells. J Neurosci. 2010 Jun 23;30(25):8551-65.

64. Jerng HH, Pfaffinger PJ. Modulatory mechanisms and multiple functions of somatodendritic A-type K+ channel auxiliary subunits. Front Cell Neurosci. 2014 Mar 27;8:82.

65. Kaulin YA, De Santiago-Castillo JA, Rocha CA, Nadal MS, Rudy B, Covarrubias M. The dipeptidyl-peptidase-like protein DPP6 determines the

unitary conductance of neuronal Kv4.2 channels. J Neurosci. 2009 Mar 11;29(10):3242-51.

66. Lin L, Sun W, Throesch B, Kung F, Decoster JT, Berner CJ, Cheney RE, Rudy B, Hoffman DA. DPP6 regulation of dendritic morphogenesis impacts hippocampal synaptic development. Nat Commun. 2013;4:2270.

67. Südhof TC. A molecular machine for neurotransmitter release: synaptotagmin and beyond. Nat Med. 2013 Oct;19(10):1227-31

68. Chiò A, Mora G, Restagno G, Brunetti M, Ossola I, Barberis M, Ferrucci L, Canosa A, Manera U, Moglia C, Fuda G, Traynor BJ, Calvo A. UNC13A influences survival in Italian amyotrophic lateral sclerosis patients: a population-based study. Neurobiol Aging. 2013 Jan;34(1):357.e1-5.

69. Diekstra FP, van Vught PW, van Rheenen W, Koppers M, Pasterkamp RJ, van Es MA, Schelhaas HJ, de Visser M, Robberecht W, Van Damme P, Andersen PM, van den Berg LH, Veldink JH. UNC13A is a modifier of survival in amyotrophic lateral sclerosis. Neurobiol Aging. 2012 Mar;33(3):630.e3-8.

70. Liu X, Seven AB, Camacho M, Esser V, Xu J, Trimbuch T, Quade B, Su L, Ma C, Rosenmund C, Rizo J. Functional synergy between the Munc13 C-terminal C1 and C2 domains. Elife. 2016 May 23;5. pii: e13696

71. Varoqueaux F, Sons MS, Plomp JJ, Brose N. Aberrant Morphology and Residual Transmitter Release at the Munc13-Deficient Mouse Neuromuscular Synapse. Mol Cell Biol. 2005 Jul;25(14):5973-84.

72. Vérièpe J, Fossouo L, Parker JA. Neurodegeneration in C. elegans models of ALS requires TIR-1/Sarm1 immune pathway activation in neurons. Nat Commun. 2015 Jun 10;6:7319.

73. Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA (URL: http://evs.gs.washington.edu/EVS/) [date (month, yr) accessed].

74. 1000 Genomes Project Consortium. et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012 Nov 1;491(7422):56-65

75. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from next-generation sequencing data. Nucleic Acids Res. 2010 Sep;38(16):e164

76. Lin L, Long LK, Hatch MM, Hoffman DA. DPP6 domains responsible for its localization and function. J Biol Chem. 2014 Nov 14;289(46):32153-65.

77. Melissa Gymrek, David Golan, Saharon Rosset and Yaniv Erlich. lobSTR: A short tandem repeat profiler for personal genomes. Genome Res. 2012 Jun; 22(6): 1154–1162.

78. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ & Sham PC (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics, 81.

79. Kikin O, D'Antonio L, Bagga PS. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. Nucleic Acids Res. 2006 Jul 1;34(Web Server issue):W676-82.

80. Aaron R. Quinlan, Ira M. Hall. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010 Mar 15; 26(6): 841–842.

81. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010 May;28(5):495-501.

82. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. Nat Methods. 2010 Apr;7(4):248-9.

83. Shakin-Eshleman SH, Spitalnik SL, Kasturi L. The Amino Acid at the X Position of an Asn-X-Ser Sequon Is an important Determinant of N-linked Core-glycosylation Efficiency. J Biol Chem. 1996 Mar 15;271(11):6363-6.

84. Strop P, Bankovich AJ, Hansen KC, Garcia KC, Brunger AT. Structure of a human A-type potassium channel interacting protein DPPX, a member of the dipeptidyl aminopeptidase family. J Mol Biol. 2004 Oct 29;343(4):1055-65.

85. Kenna KP, McLaughlin RL, Byrne S, Elamin M, Heverin M, Kenny EM, Cormican P, Morris DW, Donaghy CG, Bradley DG, Hardiman O. Delineating the genetic heterogeneity of ALS using targeted high-throughput sequencing. J Med Genet. 2013 Nov;50(11):776-83.

86. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012 Sep 6;489(7414):57-74.

87. Roadmap Epigenomics Consortium. et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015 Feb 19;518(7539):317-30.

88. Ramasamy A, Trabzuni D, Guelfi S, Varghese V, Smith C, Walker R, De T; UK Brain Expression Consortium; North American Brain Expression Consortium, Coin L, de Silva R, Cookson MR, Singleton AB, Hardy J, Ryten M, Weale ME. Genetic variability in the regulation of gene expression in ten regions of the human brain.Nat Neurosci. 2014 Oct;17(10):1418-1428.