博士論文

Investigation into the *ADARB1* Regulatory Mechanism in the Human Spinal Motor Neurons —An Implication for the Amyotrophic Lateral Sclerosis Pathogenesis

（RNA 編集酵素 ADAR2 の発現を
ヒト脊髄運動神経特異的に制御する転写因子の探索
――筋萎縮性側索硬化症との病因的関連――）

廣　瀬　直　毅

Investigation into the *ADARB1* Regulatory Mechanism in the Human Spinal Motor Neurons —An Implication for the Amyotrophic Lateral Sclerosis Pathogenesis

（RNA 編集酵素 ADAR2 の発現を ヒト脊髄運動神経特異的に制御する転写因子の探索 ――筋萎縮性側索硬化症との病因的関連――）

東京大学大学院医学系研究科

脳神経医学専攻

医学博士課程

指導教員　　岩坪　威　教授

廣 瀬　直 毅

TABLE OF ABBREVIATIONS

Abbreviation	*Description*
ADAR2	Adenosine Deaminase Acting on RNA 2
ALS	Amyotrophic Lateral Sclerosis
AMPA	Alpha-Amino-3-Hydroxy-5-Methyl-4-Isoxazolepropionic Acid
bp	Base Pair
C9ORF72	Chromosome 9 Open Reading Frame 72
CAGE	Cap Analysis of Gene Expression
CAT	CAGE-Associated Transcript
CCNF	Cyclin F
ChIP-seq	Chromatin Immunoprecipitation Followed by High-Throughput Sequencing
CLIP-seq	Crosslinking Immunoprecipitation Followed by RNA-seq
CTSS	Transcription Start Site Defined by Cap Analysis of Gene Expression
DHS	DNase-Hypersensitive Site
eQTL	Expression Quantitative Trait Locus
eRNA	Enhancer RNA
FANTOM5	Functional Annotation of the Mammalian Genome 5
FPKM	Fragments per Kilobase Mapped Exon per Million
FUS	Fused in Sarcoma
GluA2	Glutamate Receptor Ionotropic, AMPA 2
GTEx	the Genotype-Tissue Expression
H3K27ac	Histone H3 Acetylation at the Lysine 27
H3K4me1	Histone H3 Mono-Methylation at the Lysine 4
H3K4me3	Histone H3 Tri-Methylation at the Lysine 4
hnRNPA1	Heterogenous Nuclear Ribonucleoprotein A1
iMN	Motor Neuron Differentiated from Induced Pluripotent Stem Cells
kb	Kilobase
lncRNA	Long Non-Coding RNA
mCpG	Methylation of a Cytosine Residue in CpG Dinucleotide
MN	Motor Neuron
Motoneuronal TF	Transcription Factor Expressed in Motor Neurons
mRNA	Messenger RNA
NPC	Nuclear Pore Complex
nt	Nucleotides
ORTI	Open-Access Repository of Transcriptional Interactions
PROMPT	Promoter Upstream Transcript
Q/R site	Glutamine/Arginine Site
SNP	Single Nucleotide Polymorphism
SOD1	Copper/Zinc Superoxide Dismutase 1
TARDBP	TAR DNA Binding Protein
TDP-43	Transactivation Responsive Region-DNA Binding Protein of 43 Kilodalton
TF	Transcription Factor
tpm	Tag per Million
TSS	Transcription Start Site
UCSC-GB	the University of California, Santa Cruz Genome Browser
WGCNA	Weighted Gene Co-Expression Network Analysis
WT	Wild-Type

1. ABSTRACT

Amyotrophic lateral sclerosis is a so far untreatable fatal neurological disease where both upper and lower motor neurons (MNs) are selectively and progressively degenerated. Although its pathogenesis in the spinal MNs is shown to stem from abnormal down-regulation of mRNA of *ADARB1* encoding RNA editing enzyme ADAR2, what factors cause this down-regulation is yet to be determined. To infer an abnormality underlying the down-regulation of *ADARB1* in the patients' MNs, the present study determined many transcription factors (TFs) possibly activating alternative promoters of *ADARB1* in the MNs, based on cap analysis of gene expression for laser-captured human MNs, multiple *in silico* analyses using public data and *in vitro* luciferase assay for the promoters of *ADARB1*. The present results on the promoters suggested that *ADARB1* in the MNs was differentially regulated from spinal dorsal horn neurons and white matter by a set of multiple TFs forming an MN-specific complex. Because age-dependent decrease of one MN-specific activator for the *ADARB1* promoters has been previously reported in the non-ALS subjects, the present study implied that an age-dependent decrease of the *ADARB1*-regulatory TFs from an inherently low expression level due to single nucleotide polymorphisms at their expression quantitative trait loci induced the down-regulation of *ADARB1* below a threshold level in the patients' MNs after middle life. Further investigation into other TFs in the MNs will be neces-

sary to reveal the whole picture of the abnormality driving the pathogenesis of sporadic amyotrophic lateral sclerosis.

2. INTRODUCTION

2-1. Regulatory mechanism of transcription

In general, transcription of a gene is regulated largely in three levels: *cis*-regulatory regions in the genome sequence, *trans*-acting factors and epigenetic effects.

2-1-1. *Cis*-regulatory regions in the genome sequence (1): a promoter

A promoter drives transcription of messenger RNAs (mRNAs) and generally consists of two core promoters in 100 base pairs (bp) and transcription start sites (TSSs) residing at the center of each of these core promoters [**Introduction Figures 1A and 1B**][1–3]. It has been demonstrated that the core promoter has several specific features in both DNA sequences and modes of transcription initiation. The core promoter is thought to consist of specific DNA sequence motifs that are highly conserved among species (e.g., TATA-box and initiator)[4–10], although no universal motif in the core promoter has been identified. From a mode of transcription initiation, the core promoter is classified as either a focused or a broad core promoter[4–6,11]: a focused core promoter harbors a single or multiple TSSs within several nucleotides (nt), typically has a TATA-box, and is mostly observed in the spatially and temporally regulated genes; a broad core promoter has multiple weak TSSs within the length of approximately 50 ~ 100 nt, lacks the characteristic core promoter motifs and is generally associated with housekeeping genes. Although these features are basically conserved in *Drosophila* and mam-

mals[4–6,11], mammalian broad core promoters specifically reside within a CpG island[6], the region where over 200 CpG dinucleotides reside within a highly GC-enriched area[12,13]. The TATA-box rarely resides within the mammalian broad core promoter associated with a CpG island[7,14].

A paired TSSs within a promoter produces transcripts on both the plus and minus strands in bidirectional and simultaneous ways [**Introduction Figure 1B**], driven by discrete transcriptional machineries[1–3,8,9,15,16]. Such transcripts are a pair of stable mRNAs[16], or alternatively, a pair of a stable mRNA and an unstable, polyadenylated non-coding promoter upstream transcript (PROMPT[17–19]), defined by an axis of 5' splice sites and polyadenylation signals within 500 bp of the TSSs[2,8,9,15,20,21]. This pair of TSSs lies within the same DNase-hypersensitive site (DHS) in a narrow window of 100 ~ 180 bp on average that is strictly and symmetrically flanked by nucleosomes[1–3,9,15,20]. Hence, importantly, a promoter can be defined within a relatively narrow region.

2-1-2. *Cis*-regulatory regions in the genome sequence (2): an enhancer

An enhancer is defined as a conserved DNA element[8,9,22–24] that triggers transcriptional bursts at promoters within its topologically proximal region[25–27] by producing bidirectional non-coding RNAs, named enhancer RNAs (eRNAs), within a single DHS[8,9,15,22,23,28–30] [**Intro-**

duction Figures 1C and 1D]. A single enhancer can activate some target promoters at the same time[25,26,31], and conversely, multiple enhancers are connected with a single promoters in a tissue-specific manner[22,32]. By clustering within an order of magnitude wider region, multiple canonical enhancers form a super enhancer[33,34] (or stretch enhancer[35]), which reinforces expression of tissue- and cell-type-specific genes with highly concentrated eRNAs and transcriptional machineries[33,34,36], thereby determining the identity of the tissue and cell-type[31,37]. Tissue-specific canonical[22,38–42], super[34] and stretch enhancers[35] are frequently linked to many disease-associated single nucleotide polymorphisms (SNPs), highlighting the importance of tissue-specific enhancers for pathogenesis.

2-1-3. *Trans*-acting factors: transcription factors

Transcription is regulated *in trans* by both many proteins and long non-coding RNAs (lncRNAs[43–46]). Because how lncRNAs regulate gene expression is largely still unknown, in this section I am going to describe only about protein regulators, transcription factors (TFs).

So far, more than one thousand TFs that cooperatively drive transcription by RNA polymerase II are known[47–50]. By binding their specific DNA sequences called "consensus motifs" within *cis*-regulatory regions[51–55], TFs act as either an activator or a suppressor for gene expression[56]: they are activators mainly when they bind to the center region between the

bidirectional TSSs within the *cis*-regulatory region[28], whereas they tend to be suppressors when they bind closely adjacent to the TSSs[9]. Expressed in a ubiquitous or tissue-specific manner[47], TFs regulate expression of functionally-related genes[57,58] in a tissue-specific manner in concert with other TFs and transcriptional co-factors[39,48,59–61] through additive or synergistic effects[54,62,63], based on their own tissue-specific hierarchical network[59,64–66].

TFs are basically classified according to the similarities of their DNA binding domains[47,49], but they are also distinguished by two other criteria. One criterion is the role for remodeling the chromatin[50,62,67–69]: pioneer TFs bind to their target DNA sequences within a nucleosome and recruit a chromatin remodeling complex there to remove the nucleosome at that position, thereby allowing other non-pioneer TFs to bind there. Many TFs play this chromatin remodeling role in a situation-depending manner[70,71]. The other criterion is the importance in the regulatory network of TFs[33,65]: master TFs (or core TFs) bind to super enhancers of the genes that are biologically important to that cell-type, thereby defining the identity of the cell. Some master TFs also act as pioneer TFs (e.g., Oct4 and Sox2[33,65,69]), implying a functional link between these two criteria.

2-1-4. Epigenetic effects: histone modifications and DNA methylation

The third levels of transcriptional regulation are largely two epigenetic effects. One of these

is a post-translational modification of histone proteins around the *cis*-regulatory regions depending on the transcriptional activity: some modifications with methyl and/or acetyl groups on the N-terminal tail of a histone H3 protein render these *cis*-regulatory regions more accessible to additional TFs and RNA polymerase II[1,9,10,15,18,19,28,72,73] in a tissue- and cell-type-specific manner[22,74,75]. Well-known modifications on the histone H3 for the active promoters and enhancers are tri-methylation at the lysine 4 (H3K4me3), and mono-methylation at the lysine 4 (H3K4me1) or acetylation at the lysine 27 (H3K27ac), respectively; a ratio of H3K4me3 to H3K4me1 at the enhancers positively correlates with their activity[76]. The other is methylation of a cytosine residue in CpG dinucleotides (mCpG), which modulates binding affinity of TFs to their targets harboring CpGs[53] and inhibits the transcriptional initiation without affecting its elongation, thereby controlling the tissue-specific usage of the alternative promoters[12,77]. Highly expressed genes tend to have few mCpGs at their active promoter and abundant mCpGs within their gene body[77–79].

2-1-5. Species-, tissue- and cell-type-specificity of gene expression

The regulation of gene expression is highly specific to the species, the tissues and the cell-types[57,64]. Namely, genes conserved among mammals are differently regulated by species-specific ways at all the three levels described above[47,48,80–82], resulting in the different pro-

file of gene expression (i.e., transcriptome) among the mammalian species[83,84]. Similarly, different tissues of the same species have different transcriptomes[83,85–89] even in cell-type resolution (e.g., subtypes of neurons and glial cells in the brain[90]).

2-2. Amyotrophic lateral sclerosis and its etiology

2-2-1. Definition of amyotrophic lateral sclerosis

Amyotrophic lateral sclerosis (ALS or Lou Gehrig's disease) is a neurodegenerative disease affecting motor function. A pathological hallmark in a patient with this disease is progressive and selective degeneration of both upper and lower motor neurons (MNs). When so far healthy people in their middle age are affected with ALS, they can live in 3 ~ 5 years after the onset due to paralysis of muscles of respiration[91]. Namely, ALS is a fatal disease. However, an etiology of this disease is currently not fully understood, leaving this disease incurable. Therefore, ALS is one of diseases whose etiology is to be elucidated and completely curable therapeutics is to be established.

2-2-2. Epidemiology of ALS and ALS-associated genes

An incidence of ALS is about 2 per 100,000 population per year, in which there are sex dif-

ferences (ratio of men to women is about 1.4:1[92]) that weakens after menopause somehow[93]. World-wide prevalence of ALS is 1.0 ~ 11.3 per 100,000 depending on reports[94,95].

No more than 10% of patients with ALS have a positive family history for the disease, and this subpopulation of the patients is designated as familial ALS[92,94]. Genetic analyses of patients with ALS all over the world have been linking nearly 50 genes to this disease[92], including copper/zinc superoxide dismutase 1 (*SOD1*)[96], fused in sarcoma (*FUS*)[97,98], TAR DNA binding protein (*TARDBP*)[99,100], optineurin[101], valosin-containing protein[102], ubiquilin 2[103], profilin 1[104], heterogenous nuclear ribonucleoprotein A1 (*hnRNPA1*)[105], chromosome 9 open reading frame 72 (*C9ORF72*)[106,107] and cyclin F (*CCNF*)[108]. Regardless of such various mutations, more than 90% of the patients have a negative family history for ALS (i.e., sporadic ALS). Although some mutations are also found in patients with sporadic ALS to a limited extent[91] (e.g., *SOD1*[109], *FUS*[110], *TARDBP*[99,100], *hnRNPA1*[105], *C9ORF72*[106,107,111,112], and *CCNF*[108]), associated genes and their proportions in all the ALS patients markedly differ between East Asian and Caucasian[113], and none of these defined mutations has enabled to develop effective therapeutics for ALS.

2-2-3. Hypotheses of sporadic ALS pathogenesis

During the past several decades, many research groups have tried to elucidate pathogenesis of

ALS[91] with, for example, axonal transport defects[114], neurofilaments abnormalities[115], decreased clearance of neurotransmitter glutamate[116], oxidative stress[117] and D-serine-induced glutamate toxicity[118]. Among such hypotheses, adenosine deaminase acting on RNA 2 (ADAR2)-glutamate receptor ionotropic, AMPA 2 (GluA2) hypothesis[119] proposed and expanded by the Kwak' s group is the most plausible. This is because it well explains both the cause of death of MNs and pathological changes specifically found in patients' MNs, based on an aberrant molecular cascade.

2-2-4. ADAR2-GluA2 hypothesis for the pathogenesis of sporadic ALS

The ADAR2-GluA2 hypothesis proposes an abnormal molecular cascade underlying the pathogenesis of sporadic ALS, as illustrated in **Introduction Figure 2A**. In MNs of patients with sporadic ALS, the expression of mRNA of ADAR2 is down-regulated in a disease-specific manner[120], thereby allowing the expression of unedited RNAs with ADAR2-mediated adenosine-to-inosine positions, including the glutamine/arginine (Q/R) site in pre-mRNA of GluA2 protein, a member of subunits of alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors[121]. When an unedited GluA2 subunit is included in an AMPA receptor, it renders this glutamate receptor permeable to Ca^{2+} ions[122–124]. Through Ca^{2+}-permeable AMPA receptors, exaggerated intra-

cellular entrance of Ca^{2+} ions activates autophagy[125] and calpain, a Ca^{2+}-dependent cysteine protease. This activated calpain then causes Transactivation Responsive Region-DNA binding protein of 43 kilodalton (TDP-43) to aggregate abnormally in the cytosol by cleaving TDP-43 into aggregation-prone fragments[126] (known as TDP-43 pathology), and also disrupts nuclear pore complex (NPC) by cleaving its constituents nucleoporins[127]. Indeed, both of TDP-43 pathology and the disruption of NPC specifically occur in the MNs of patients with ALS[128–131]. Besides, when ADAR2 is conditionally knocked out with the Cre-loxP system in MNs of mice after birth, these mice named AR2 faithfully recapitulate the disease phenotypes of ALS described above[132] [**Introduction Figure 2B**]: loss of ADAR2 in the MNs causes sequentially failure of RNA editing at the Q/R site of GluA2, activation of calpain, which results in generation of TDP-43 pathology, disruption of NPC, and activation of autophagy[133]. These sequential events ultimately result in death of the MNs and motor dysfunction in the AR2 mice. Therefore, down-regulation of ADAR2 in the mouse MNs mimics both clinical features and pathological findings seen in the patients with sporadic ALS. Moreover, down-regulation of ADAR2 mRNA is also reported in the MNs of an ALS patient carrying an *FUS*^{P525L} mutation[110] and decreased RNA editing levels at the Q/R site of GluA2 are observed in tissues of ALS patients with an abnormal expansion of G_4C_2 repeats in *C9ORF72*[134], indicating that the ADAR2-GluA2 hypothesis may partially explain the molecular cascade in he-

reditary types of ALS, as well. In short, down-regulation of ADAR2 mRNA and protein with resultant editing insufficiency at the Q/R site of GluA2 in the MNs may play a pivotal role in the pathogenesis of both sporadic and some types of familial ALS.

2-3. ADAR2 gene *ADARB1*

When the first canonical coding exon of ADAR2 gene *ADARB1* is designated as Exon 1[135], ADAR2 protein has two nuclear localization signals and two double-stranded RNA binding domains residing in Exon 2, and an adenosine deaminase domain lying from Exon 4 to Exon 6 [**Introduction Figure 3**][136,137]. So far *ADARB1* is known to have 5 alternative exons: Exon 1a[135], Exon 0[138], 5'-extended Exon 2[139], Exon 5a[140,141] and Exon 7a[142]. In addition, *ADARB1* transcripts can be in four forms in the 3' region downstream of Exon 9: the longest form harboring retained Intron 9 (L-1), two longer ones (L-2 and L3) and the shortest (S)[135,136]. Such multiple alternative exons can affect enzymatic activity of ADAR2 protein[143].

By cap analysis of gene expression (CAGE)[144], the Functional Annotation of the Mammalian Genome 5 (FANTOM5) Consortium has analyzed the transcriptional activity genome-widely at the a nucleotide level[10,22,30], and determined that *ADARB1* had five CAGE-defined TSSs (CTSSs): p1@ADARB1, p2@ADARB1, p3@ADARB1,

p4@ADARB1 and p13@ADARB1 [see **Introduction Figure 3**], where the smaller number *X* in p*X*@*Y* defines a more intense and ubiquitous CTSS of *Y* gene[10]. For simplicity, these five CTSSs of *ADARB1* are abbreviated just to p1@, p2@, p3@, p4@ and p13@ hereafter. Because of these five CTSSs, there are indeed various putative patterns of *ADARB1* transcripts, many of which have high coding potential[46] [**Introduction Figure 4**]; however, ADAR2 is enzymatically fully active only when it is translated from *ADARB1* transcripts including Exon 2 and being in one of the longer 3' end-forms (L-2 or L-3)[136,140]. According to the previous reports[10,22,30,135,138], the p1@ flanks the first 5' untranslated exon (Exon -2) and the p2@ lies in the region 714 bp upstream from the 5' end of the p1@; both CTSSs can produce full-length transcripts. On the other hand, the p3@ and the p4@ reside at the exact 5' end of Exon 0 and the position 18 bp downstream from the p3@, respectively; the transcripts from them are devoid of Exon -2, Exon -1 and Exon 1a. The p13@ is located in the region within the gene body between Exon -2 and Exon -1, and its transcripts are depleted Exon -2. Enzymatic competence of ADAR2 proteins expressed from the latter three CTSSs is currently undetermined.

According to the FANTOM CAT v1.0.0 Extended View (v1, hg19) on the FANTOM Zenbu Browser (http://fantom.gsc.riken.jp/zenbu/)[46] [see **Introduction Figure 4**], *ADARB1* is expressed in insignificantly weak correlation with intergenic lncRNA *LINC00162*. Besides,

ADARB1 is apparently regulated by two super enhancers (~10 kilobases (kb) and ~110 kb in length) and other canonical primate-specific enhancers, because both of the two super enhancers harbor many expression quantitative trait loci (eQTL) of *ADARB1* affecting its expression, and because many of its eQTL reside within DHSs whose DNA sequence is highly conserved among primates but poorly between humans and mice. There are 18 CTSSs other than the five CTSSs of *ADARB1* within the 231.8 kb region around *ADARB1*, suggesting positions of putative enhancers around this gene [annotated with "*lnc*" and "*e*" for lncRNAs and eRNAs, respectively, in **Introduction Figure 4**].

2-4. Working hypothesis and problems that should be overcome in this study

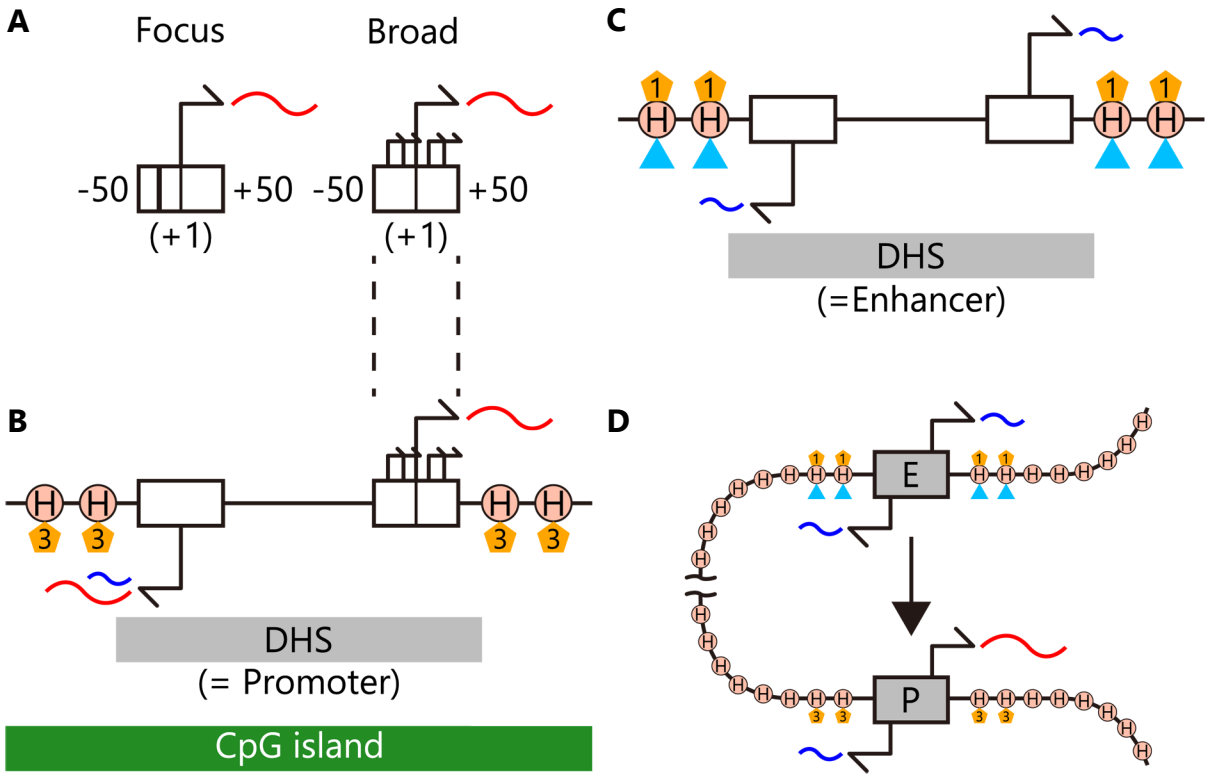
2-4-1. Working hypothesis in this study

When the ADAR2-GluA2 hypothesis holds true, why is mRNA of ADAR2 down-regulated in the MNs of patients with ALS? Expression of ADAR2 and the edited form of its substrate have been demonstrated to decrease with aging in wild-type (WT) mouse spinal MNs[145] and, corresponding phenomena are also observed in the brain of older humans compared to youngers[146,147]. Based on these observations, I set a putative long-term pathogenic cascade of sporadic ALS as a working hypothesis in the present study: abnormally decreasing in an

age-dependent manner, some *ADARB1*-regulatory TFs would fail to maintain the normal expression of *ADARB1* in the patient's MNs, thereby accelerating the age-dependent decrease of ADAR2 [**Introduction Figure 5**]. This working hypothesis fueled my interest in a regulatory mechanism of *ADARB1* expression in the human MNs, because understanding this mechanism is vital to concretely predict an abnormality underlying the down-regulation of *ADARB1* in the patients' MNs.

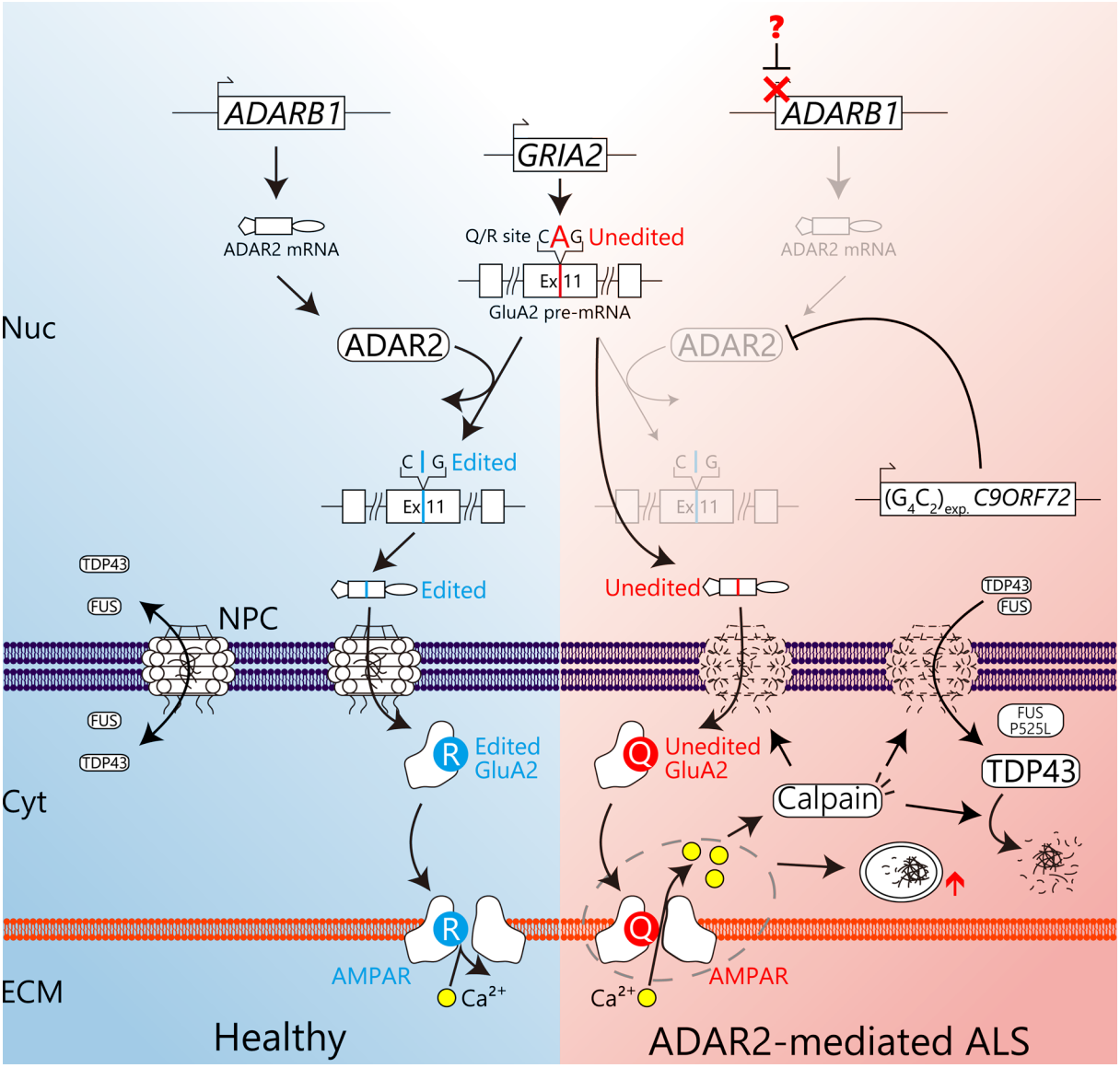
2-4-2. Problems that should be overcome in this study

To investigate the regulatory mechanism of *ADARB1* expression in the human MNs, this study should overcome the following big four problems: lack of specific profile of both expression of TFs and transcriptional activity of the *ADARB1* alternative promoters in the human MNs; lack of established methods that enable to systematically infer all the *ADARB1*-regulatory TFs in the MNs; limited information regarding the regulators for expression of the human[148,149] and murine[150–152] ADAR2 genes; lack of ontological difference in the regulatory mechanism between human and mouse ADAR2 genes. It is necessary to infer *ADARB1*-regulatory TFs *in silico*, because expression libraries of human TFs are not completely available and because it will be quite laborious to functionally screen all *ADARB1*-regulatory TFs whose library is available.



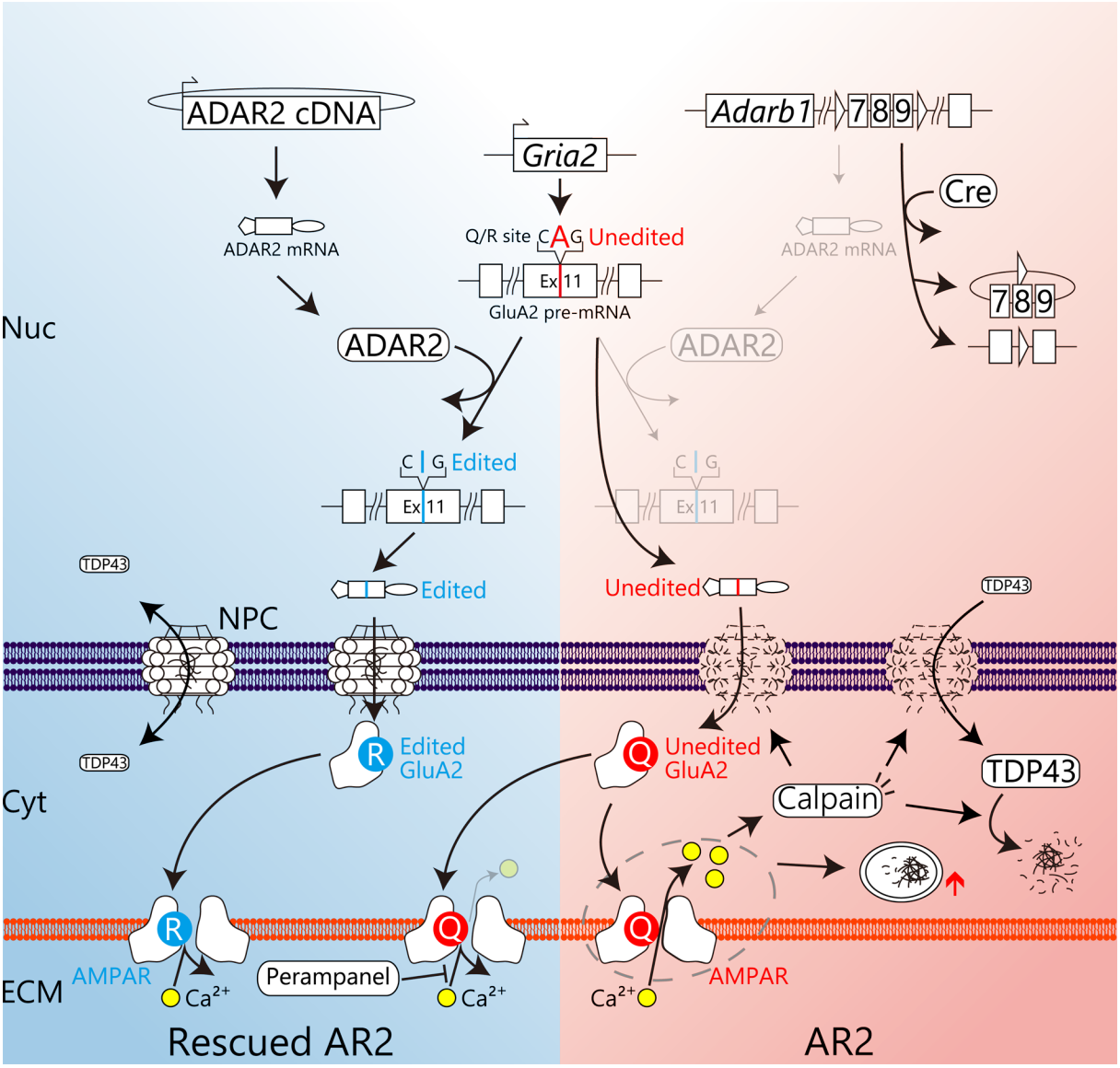
Introduction Figure 1 Definition of the promoter and enhancer, according to Andersson R. *et al*. *Mol*. *Cell* (2015)[3] and Chen Y. *et al*. *Nat*. *Genet*. (2016)[2].

A, Structure of core promoters. A TSS (+1) is indicated with an arrow, which resides at the center of a core promoter. Both focus and broad core promoters are shown with positions of the TATA-box (thicker vertical line) and/or the initiator (thinner vertical line) and mRNA (red wavy line). **B**, Structure of a promoter embraced by a CpG island (green). The promoter consists of a pair of bidirectional TSSs in 100 ~ 180 bp distance and is defined within a single DNase hypersensitive site (DHS) symmetrically flanked by nucleosomes with H3K4me3 modification (3 in an orange pentagon). Red wavy line, mRNA; blue wavy line, non-coding promoter upstream transcript; H, histone. **C**, Structure of an enhancer. The enhancer is also defined within the single DHS flanked by nucleosomes with H3K4me1 (1 in an orange pentagon) and H3K27ac (light blue triangle) modification. In the enhancer, bidirectional non-coding RNAs (enhancer RNAs) are transcribed. **D**, Schema of the enhancer activating its target promoter. Transcription levels are highly correlated between the enhancer and its target promoter.

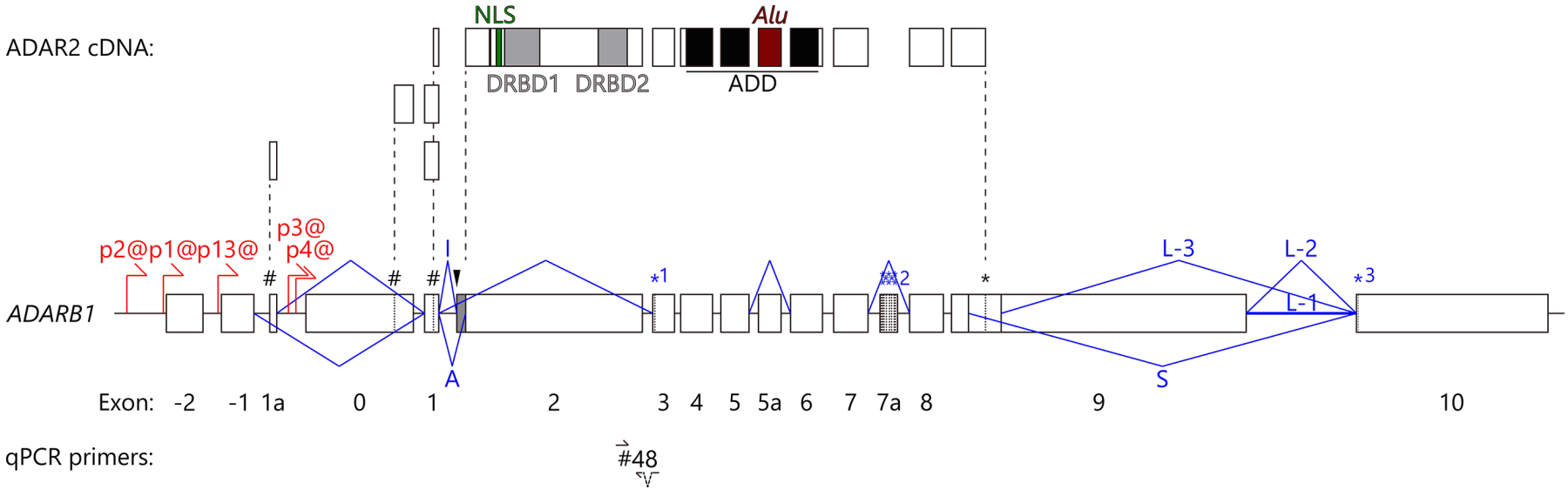


Introduction Figure 2A Scheme of ADAR2-GluA2 hypothesis of sporadic ALS pathogenesis.

Normal activity of ADAR2 in motor neurons of healthy people (left panel) is deteriorated in motor neurons of patients with ADAR2-mediated ALS, including sporadic, *FUS*^{P525L}-caused and abnormal expansion of G_4C_2 repeats in *C9ORF72*-caused ALS (right panel). As indicated with a red question mark, what factor causes the down-regulation of ADAR2 expression is still unknown. Although Ca^{2+} influx encircled by dotted line is confirmed only in sporadic ALS model mice AR2, this molecular event is sure to occur in the motor neurons of the patients as well [see **Introduction Figure 2B**]. AMPAR, AMPA receptor; Cyt, cytoplasm; ECM, extracellular matrix; NPC, nuclear pore complex; Nuc, nucleoplasm; TDP43, Transactivation Responsive Region-DNA binding Protein of 43 kilodalton; FUS, Fused in Sarcoma.

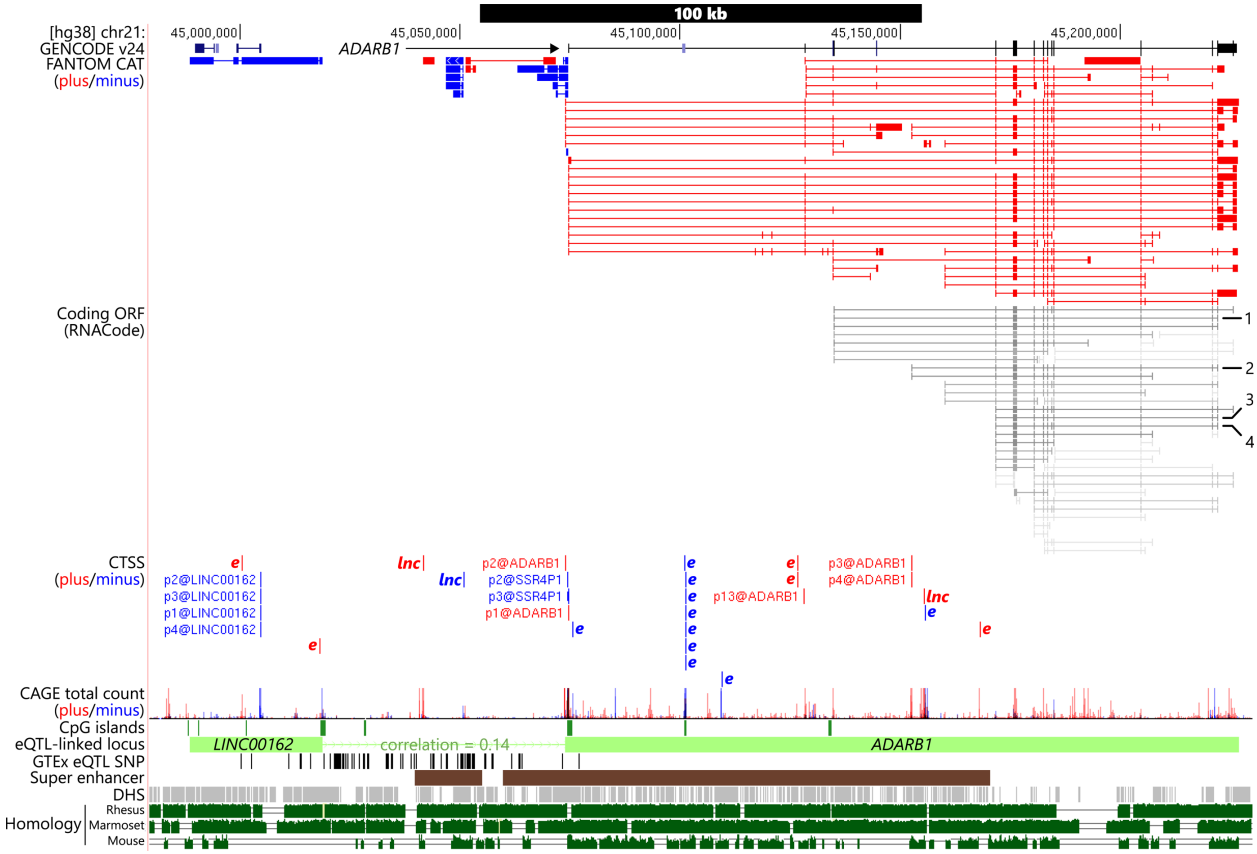


Introduction Figure 2B Molecular phenotype in the motor neurons of AR2 mice. ADAR2-conditionally knocked out mice AR2 (right panel) faithfully recapitulate the molecular phenotype observed in the motor neurons of the patients. Of note, such phenotype in AR2 mice can be rescued by intravenous injection of AAV9-ADAR2 expression vector[153] or oral administration of anti-epileptic drug perampanel[154], providing the rationale of potential therapeutics for sporadic ALS. Abbreviations are the same as in **Introduction Figure 2A**.



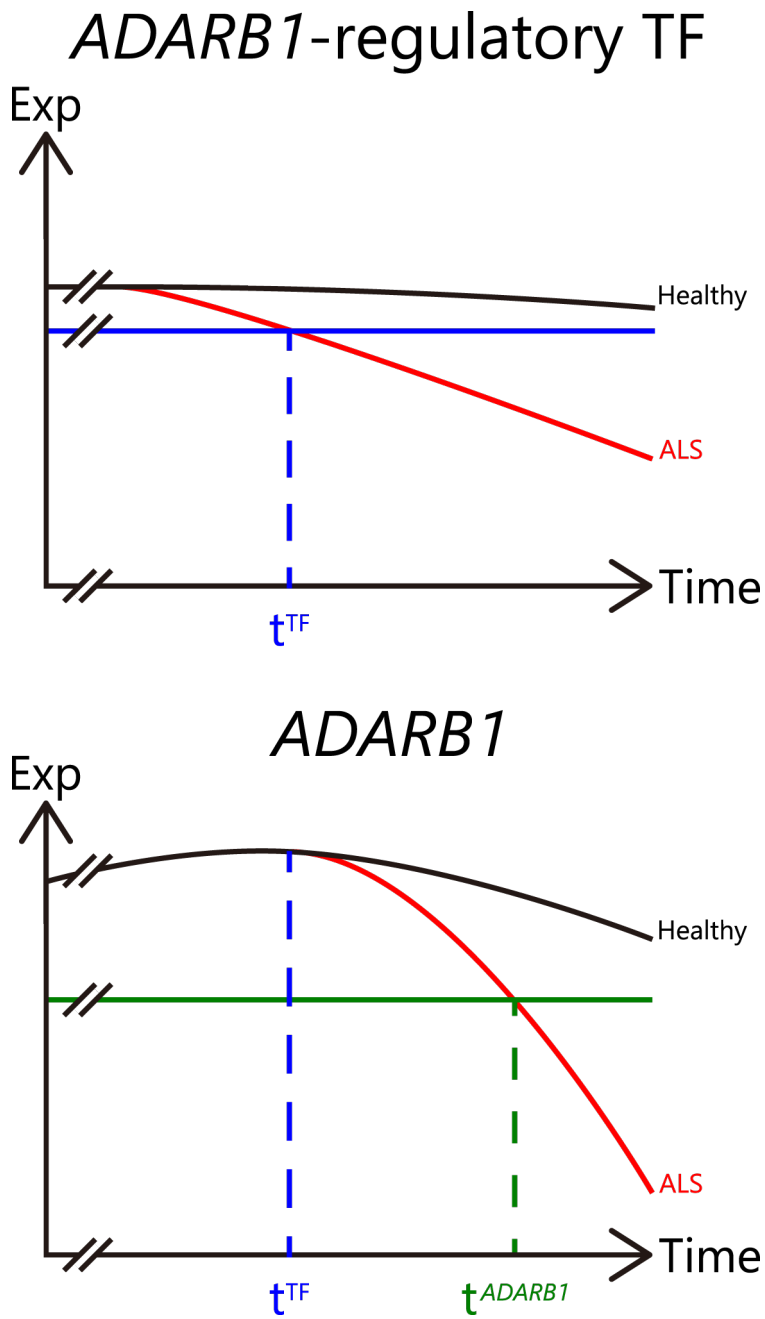
Introduction Figure 3 Structures of cDNA of ADAR2 and its gene *ADARB1*.

(Top) Only 3 representative splicing variants at the N-terminal of the cDNA are shown, and other exons are omitted for simplification. From top to bottom, the variants are started from the canonical Exon 1, Exon 0, and Exon 1a, respectively. (Bottom) Blue line, patterns of alternative splicing of *ADARB1*; red, positions and names of the CTSSs of *ADARB1*; hash mark, the start codon; black arrowhead, the self-editing site of ADAR2; I, splicing pattern in edited transcripts; A, splicing pattern in unedited transcripts; black asterisk, the canonical stop codon; *1, *2 and *3 are stop codons for the Exon 2-skipped variant, the Exon 7a-including variant, and the S variant at the 3' end of the gene, respectively. Positions of the domains and the names of splicing variants at the 3' end are according to Kawahara Y. *et al*. *Gene* (2005)[136]; the numbers of exons are according to Slavov D. and Gardiner K. *Gene* (2002)[135]. All the exons are in scale, while the introns are not. Primer positions for quantitative PCR of ADAR2 mRNAs are shown at the bottom. ADD, adenosine deaminase domain; DRBD, double-stranded RNA binding domain; NLS, nuclear localization signal.



Introduction Figure 4 Genetic environment around *ADARB1*.

ADARB1 locus, CATs (CAGE-associated transcripts), its coding potential (1, 2, 3 and 4 are all putative enzymatically active coding sequences started from Exon 1a, Exon 0, Exon 1 excluding Exon 5a, and Exon 1 including Exon 5a, respectively), CTSSs (CAGE-defined TSSs) and total count of CAGE signals (*e* and *lnc*: CTSSs of enhancer RNA and other lncRNA, respectively), CpG islands (green bar), eQTL (expression quantitative trait loci)-linked lncRNA-mRNA pair (green yellow), SNPs at eQTL of *ADARB1* (black), super enhancers (brown), DHSs (DNase-hypersensitive sites; grey), and DNA sequence conservation among mammals (green signals; lack of signals indicates human-unique sequences). All of these are publicly available data (see the Materials and methods Section).

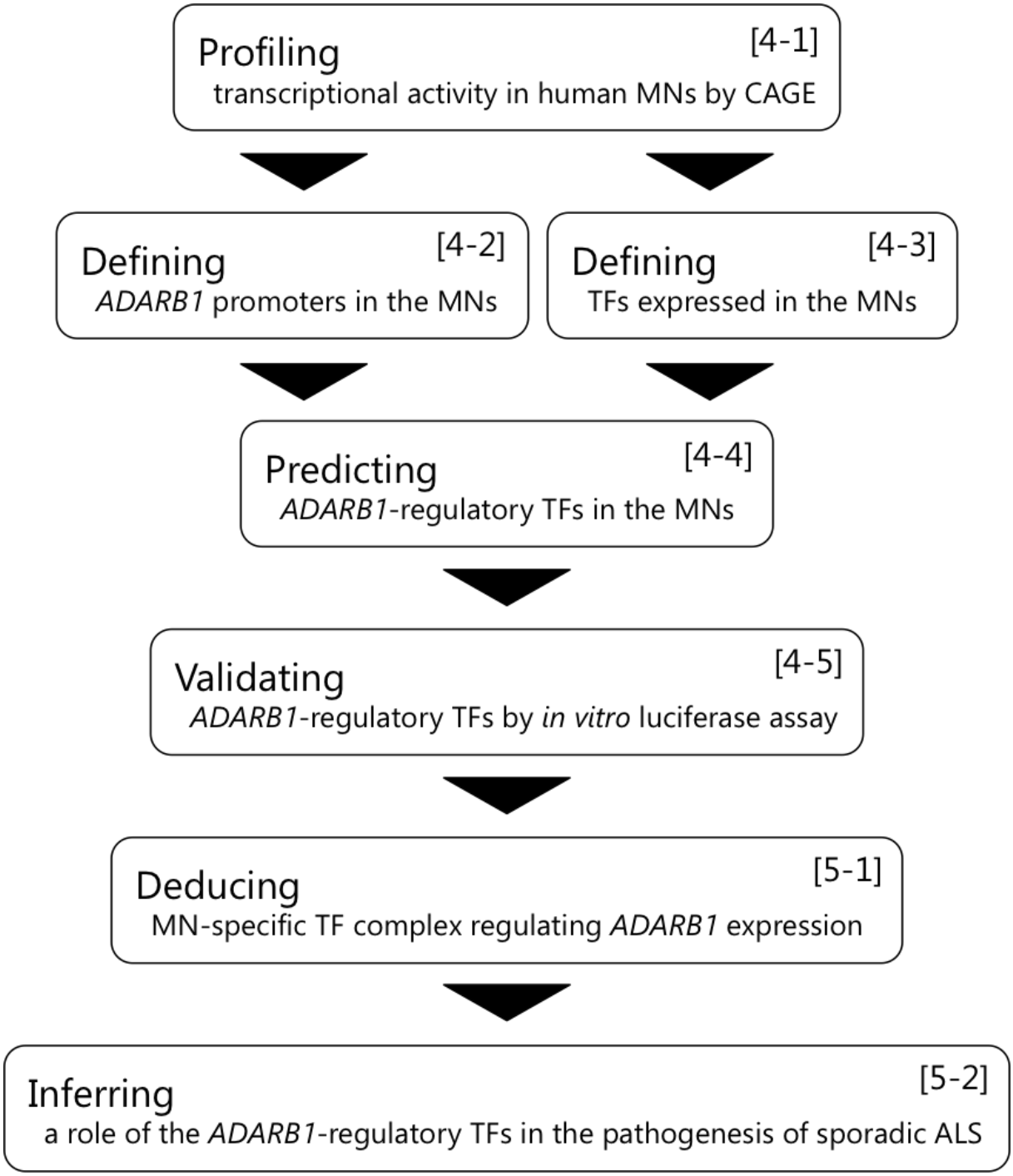


Introduction Figure 5 Working hypothesis in this study.
Abnormal age-dependent decrease of *ADARB1*-regulatory TF(s) in the motor neurons of patients with sporadic ALS (red line) compared to the healthy (black line) could cause the down-regulation of *ADARB1*. When expression of the *ADARB1*-regulatory TF decreases below the threshold for maintaining the required expression of *ADARB1* (blue horizontal line), at that time point (t^{TF}) expression of *ADARB1* might start being down-regulated. A time point T^{ADARB1} at which expression of *ADARB1* becomes lower than the threshold to sufficiently mediate RNA editing at Q/R site of GluA2 (green horizontal line) may be the entrance for the pathogenesis of sporadic ALS already proposed by ADAR2-GluA2 hypothesis.

3. MATERIALS AND METHODS

3-1. Strategy in this study

To comprehensively seek human *ADARB1*-regulatory TFs in the MNs, this study adopted a strategy assuming species-, tissue- and cell-type-specific regulation of *ADARB1* expression [**Materials and methods Figure 1**]. First, I genome-widely profiled transcriptional activity specific to laser-captured human MNs by conducting the CAGE on these neurons. Next, based on this valuable CAGE data and supportive public data, I defined active promoters of *ADARB1* in the MNs and all TFs expressed in the human MNs (motoneuronal TFs). After this, I predicted *ADARB1*-regulatory TFs in the MNs by multiple approaches that mutually complimented and supported my prediction, with the best use of various public data of TFs. Fourth, I validated many *ADARB1*-regulatory TFs from the predicted TFs by *in vitro* luciferase assay for the *ADARB1* promoters. Then, based on the results of this assay and public data of protein-protein interactions and tissue-specific expression of the TFs, I deduced an MN-specific TF complex possibly regulating *ADARB1* expression in the MNs. Finally, I inferred a role of the *ADARB1*-regulatory TFs in the pathogenesis of sporadic ALS, based on public data of age- and polymorphism-dependent change in their expression, thereby proposing an implication for the sporadic ALS pathogenesis.



Materials and methods Figure 1 Schema of the strategy in this study.
Every procedure is shown with the number of the Results and Discussion Sections in brackets.
CAGE, cap analysis of gene expression; MN, motor neuron; TF, transcription factor.

3-2. Materials and methods in this study

When a particular software or package was not denoted, all data management was performed by the free programming language R version 3.2.2[155] on RStudio version 1.0.136[156].

3-2-1. Data visualization of the genetic environment around *ADARB1*

All information was visualized on the University of California, Santa Cruz Genome Browser (UCSC-GB; https://genome.ucsc.edu/) with GRCh38/hg38 human assembly. The CTSSs around *ADARB1* and CAGE total counts were retrieved via the track data hub of FANTOM5 on the UCSC-GB. All additional information of the genetic environment around *ADARB1* was extracted from the FANTOM CAT v1.0.0 Extended View v1 hg19 on the FANTOM Zenbu Browser (http://fantom.gsc.riken.jp/zenbu/)[46], re-annotated from hg19 to hg38 using UCSC tool LiftOver (https://genome.ucsc.edu/cgi-bin/hgLiftOver), and then, visualized on the UCSC-GB; no insertion or deletion of nucleotides was observed during the re-annotation with LiftOver within the investigated region around *ADARB1*. This information included the FANTOM5 data[46] (CAGE-associated transcripts (CATs), coding potential determined by RNACode[157], lncRNA-mRNA pairs linked with eQTL and positions of super enhancers), positions of DHSs[38], SNPs at eQTL of *ADARB1* affecting an expression level of this gene[158], and positions of UCSC CpG islands.

3-2-2. Extraction of total RNAs from spinal MNs, dorsal horns and white matter

Frozen spinal cords of two autopsied non-ALS donors (donor 1 T-153 and donor 2 T-168) were respectively embedded in Tissue-Tek® O.C.T. Compound (Sakura Finetech) and sectioned by a cryostat (Leica Microsystems) at -20 °C into 20 μm. The sections were washed by methanol, dried in 5 minutes at the room temperature, stained by 0.1% toluidine blue, washed by 70% ethanol 5 times, washed by 100% ethanol, dried in 5 minutes at the room temperature, and then kept at -30 °C until following microdissection. Storage buffer of laser-captured MNs was prepared with RNeasy Micro Kit Buffer RLT (QIAGEN) and beta-mercaptoethanol (ratio = 100:1 in volume). From these sections, large anterior cells were recognized as MNs, and they were sampled one by one into the storage buffer with a laser capture microdissector (Leica Microsystems): 11,100 MNs were from donor 1 and 9,529 from donor 2 (in total, 20,629 MNs). All the samples were kept at -80 °C until RNA extraction. For the controls, the dorsal horns and white matter of the donor 2 were also macroscopically sampled with a knife. Total RNAs of all the samples were extracted using QIAGEN RNeasy Micro Kit (QIAGEN) according to the manufacturer's instruction, respectively. Sample collection and RNA extraction were conducted by Ms. Teramoto and me. The experimental procedures were approved by the Ethics Committee of the University of Tokyo (Examination Number: G1396- (32)).

3-2-3. Cap analysis of gene expression of human MNs, dorsal horns and white matter

For the 6 sets of the extracted total RNA, CAGE[144] was performed and the analyzed data files were provided by Genome Network Analysis Support Facility, RIKEN (Yokohama, Japan). The provided data files included bam files and bedGraph files mapped by BWA[159] version 0.7.10-r789 or TopHat2[160] version 2.0.12 onto the hg19 annotation. Before visualization of CAGE signals around *ADARB1* on the UCSC-GB, all bedGraph files were re-annotated from hg19 to hg38 using the UCSC LiftOver.

Using these bam files, expression levels of CAGE tags in every sample were calculated in tag per million (tpm) by a series of commands in CAGEr R package[161] as follows: *normalizeTagCount(myCAGEset, method="simpleTpm")*; *clusterCTSS(myCAGEset, threshold=1, nrPassThreshold=1, thresholdIsTpm=T, method="paraclu", maxDist=400, removeSingletons=F)*; *cumulativeCTSSdistribution(myCAGEset, clusters="tagClusters")*; *quantilePositions(myCAGEset, clusters="tagClusters", qLow=0.1, qUp=0.9)*; *exportToBed(myCAGEset, what="tagClusters", qLow=0.1, qUp=0.9, oneFile=T)*; *aggregateTagClusters(myCAGEset, tpmThreshold=5, qLow=0.1, qUp=0.9, maxDis=400)*; *getExpressionProfiles(myCAGEset, what="consensusClusters", tpmThreshold=5, nrPassThreshold=1, method="som", xDim=4, yDim=2)*; *exportToBed(myCAGEset, what="consensusClusters", colorByExpressionProfile=T)*; *extractExpression-*

Class(myCAGEset, what="consensusClusters", which="all"). Pairwise correlations among the 6 samples were calculated using *plotCorrelation(myCAGEset, samples="all", method="pearson")* function in the CAGEr R package.

There was still no consensus on a definitive list for TFs function of which was confirmed *in vivo* or *in vitro*[47]. Therefore, before I define motoneuronal TFs, two lists of TFs in previous studies[10,50] were merged beforehand to make a whole list of human putative TFs, resulting that 2,410 genes were defined as TFs in this study; this list consequently included not only *bona fide* DNA sequence-dependent TFs but also other genes broadly related to transcription (e.g., genes encoding subunits of RNA polymerases). Expression levels of TFs in the MNs were extracted by their names from data mapped by BWA and TopHat2 using BEDtools2[162] after the bed files of the consensus cluster sorted by *sortBed* were annotated onto hg19 by *bedtools intersect -s -wb -a -b* in this package.

3-2-4. Processing of RNA-seq data from human tissues and cells

In this study, two sets of RNA-seq data were analyzed: the Data Set 1 included 44 RNA-seq data from the human central nervous tissues[83,163–166] [**Materials and methods Table 1**]; the Data Set 2 consisted of RNA-seq data from 60 various tissues, primary cells and MNs differentiated from induced pluripotent stem cells (iMNs)[39,83,88,89,163–168] [**Materials and methods**

Table 2]. After sra files were converted to fastq files using *fastq-dump -I -gzip* function in sratoolkit version 2.8.1-2 (https://www.ncbi.nlm.nih.gov/sra/docs/toolkitsoft/), the first quality control of all the files was performed with *fastqc* function in FastQC version 0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and adaptors in reads were trimmed by Trimmomatic[169] version 0.36 with the following command for both a single-end and paired-end: *java -jar trimmomatic-0.36.jar SE* (or "*PE*" for the paired-end) *-phred33 ILLUMINACLIP:*[the name of the adaptor sequence file]*:2:30:10 MAXINFO:40:0.7 MINLEN:36*. After the quality control was conducted again in the same way as the first one, the filtered and trimmed reads were aligned by STAR[170] version 2.5.2b to UCSC hg38 with annotation of GENCODE[171] version 24 by *STAR --genomeDir --readFilesIn --readFilesCommand gunzip -c --outFilenamePrefix --outSAMstrandField intronMotif --outSAMtype BAM SortedByCoordinate*. Then, the aligned reads were assembled by *stringtie -o -G -A -x chrM* in StringTie-1.3.2d[172] and then merged by *stringtie --merge -G -o -i* for both data sets, respectively. With the merged transcriptome, the aligned reads were re-assembled by *stringtie -o -e -B -G -A*, and from the re-assembled transcriptome, transcript-level expression of the TFs was retrieved with their gene names, using Ballgown R package[173] and referencing a protocol article[174]. After the negligible transcripts less than 1 fragments per kilobase mapped exon per million (FPKM)[175] were discarded, gene-level ex-

pression of the TFs was gained by summing up their transcript-level expression.

3-2-5. RNA-seq from spinal MNs of WT mice and its data analysis

Single spinal MNs were collected from three WT C57BL/6 mice at 7 weeks old (Oriental Yeast Co., Ltd.) with the laser capture microdissector in the same way as described above. The numbers of the MNs collected by Ms Teramoto and me were 1,916 from WT-1, 1,933 from WT-2 and 2,042 from WT-3. After total RNA was extracted using QIAGEN RNeasy Micro Kit (QIAGEN), RNA quality was checked using Agilent RNA 6000 Pico Assay (Agilent Technologies), and then, sequencing libraries were prepared using SMARTer Stranded RNA-Seq Kit (Clontech) by Ms. Teramoto. These libraries were sequenced using Illumina HiSeq2000^{TM} by BGI Japan, and that institution provided data of the reads whose quality was controlled and adaptors were trimmed. Succeeding data processing procedures were the same as those for the RNA-seq data from the human samples described above, except for an aligned genome UCSC mm10. To visualize RNA-seq signals on the UCSC-GB, bedGraph files were produced from the aligned bam files sorted by the coordinate with *index* function in SAMtools[176] version 1.3.1, by *bamCoverage -b --normalizeUsingRPKM --outFileFormat bedgraph -o* function in deepTools[177] version 2.4.1.

3-2-6. Investigation of conservation of TFs between human and mouse

To search common TFs between human and mouse, only names and expression levels of the TFs were taken into consideration, and homology examination for neither DNA nor protein sequence was performed. Because expression of 161 TFs was only detected in the public RNA-seq data from spinal MNs, their expression levels were input as 0 tpm. On the other hand, the mean expression levels of the other 1,168 TFs (defined by BWA) and 1,046 TFs (defined by TopHat2) were calculated, respectively. After expression levels of TFs that were uniquely detected by BWA or TopHat2 were set as 0 tpm in the other data, the final mean expression values were calculated between these two data based on BWA and TopHat2. Using this processed data, quartile of expression levels of the TFs in the human MNs were examined by *quantile* function in the R. On the other hand, after the mean expression levels of the TFs among three WT mice were calculated, these mean values were used for examining the quartile. The scatter plot was drawn using *plot* function in the default R.

3-2-7. Calculation of correlation coefficients among the CAGE signals

The human CAGE summary data produced by the FANTOM5 Consortium[10,22,30,178] with annotation (hg19.cage_peak_phase1and2combined_tpm_ann.osc.txt.gz) was downloaded from its website (http://fantom.gsc.riken.jp/5/datafiles/latest/extra/CAGE_peaks/). Only human

tissue data in that file were used to investigate tissue-specific expression of the five CTSSs of *ADARB1* [**Materials and methods Table 3**]. Pearson correlation coefficients and statistical significance among CAGE signals around *ADARB1* detected by the FANTOM5 Consortium were calculated for central nervous tissue samples (n = 60) and for all of the samples of adults (n = 140), respectively, by *pairs.panels* function in psych R package[179]. Statistical analyses were performed at the same time using this *pairs.panels* function, setting the significance threshold at 0.05.

3-2-8. Investigation of CAGE signals and histone modifications around *ADARB1* in human tissues and HeLa-S3 cell line

In addition to HeLa-S3 cells, I found only seven adult human tissues from which data of concomitant analyses of both CAGE and chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) for H3K27ac, H3K4me1 and H3K4me3 were available: adipose, hippocampus, liver, pancreas, smooth muscle, spleen and substantia nigra. All the CAGE data were extracted from the aforementioned annotated summary data downloaded from the FANTOM5 website. The mean values of CAGE signals were calculated for replicated samples: the biological replicates of adipose (n = 4), hippocampus (n = 3) and substantia nigra (n = 3), and the technical replicates of HeLa-S3 cells (n= 3). For the other tissues, only

one data from a pooled sample was available. ChIP-seq data of the tissues and the HeLa-S3 cell line[39] were downloaded via the website Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) with the accession numbers summarized in **Materials and methods Table 4**. After re-annotation from hg19 to hg38 using the UCSC LiftOver, all the data were visualized on the UCSC-GB.

3-2-9. Quantification of mRNA of ADAR2 in adipose tissue and HeLa cells

Total RNA was extracted from HeLa cells using the illustra RNAspin Mini RNA Isolation Kit (GE Healthcare Bioscience) according to the manufacturer's instruction. Total RNA extracted from a human abdominal subcutaneous adipose tissue was purchased at Zenbio (RNA-T10-2; 10 μg of the RNA in 10 μL of 0.1 mM EDTA). The provided donor information was as follows: 49 year-old non-smoker female without medications; unknown ethnicity; body mass index in average was 25.8; not diabetic nor reactive to viral DNA from HIV-1, HIV-2, HTLV I, HTLV II, hepatitis B and hepatitis C. The provided Certificate of Analysis showed 7.8 in the RNA Integrity Number (RIN) of the RNA. From this total RNA, cDNAs were reversely transcribed using ReverTraAce qPCR RT Kit (TOYOBO) and quantified by real-time quantitative PCR with a set of primers for mRNAs of ADAR2 or beta-actin on LightCycler System (Roche Diagnostics). Primer sequences for ADAR2 from the 5' to the 3'

end were TTGGATCAGACGCCATCTC (19 nt; the forward) and GACAGCGTCAGCTAAAACCTG (21 nt; the reverse). Primer sequences for beta-actin in the same direction were TCCTCCCTGGAGAAGAGCTA (20 nt; the forward) and CGTG-GATGCCACAGGACT (18 nt; the reverse). Before quantification, an internal standard specific to each of the genes was amplified using the same set of primers mentioned above. The set of standard DNA and cDNA were amplified in duplex in 20 μL of reaction mixture, composed of 10 μL of 2× LightCycler 480 Probes Master Roche (Roche Diagnostics) and 10 μM of each primer and the Universal ProbeLibrary Probe (Roche Diagnostics; #48 and #27 for ADAR2 and beta-actin, respectively) by PCR: pre-incubation at 95 °C for 10 min and amplification in 45 cycles of denaturation at 95 °C for 10 sec and annealing and extension at 60 °C for 30 sec. Finally, the expression level of mRNA of ADAR2 was normalized to that of beta-actin.

3-2-10. Calculation of correlation in expression between *ADARB1* and TFs

Correlation in expression with *ADARB1* was calculated only for the 1,672 TFs defined by the FANTOM5 Consortium. For calculation, 10 tissue samples in which *ADARB1* was highly expressed and another 10 in which *ADARB1* was lowly expressed were selected from the old Phase 1 data of the FANTOM5 Consortium downloaded via the Zenbu Browser (accessed on

Sep/2014), respectively [**Materials and methods Table 5**]. The expression levels of *ADARB1* in those tissue samples was extracted from that data, whereas all expression levels of p1@TFs in the selected tissue samples were extracted from the old Phase 1 data of the FANTOM5 Consortium downloaded from its website at that time. Spearman correlation coefficients in expression levels between p1@TFs and *ADARB1* were calculated using *cor* function in the R. No statistical analysis was performed during the calculation.

3-2-11. Search on a database for TFs probably targeting *ADARB1*

ADARB1 was queried as a "target gene" on Open-access Repository of Transcriptional Interactions (ORTI)[180] (http://orti.sydney.edu.au/) only for "Homo sapiens" with the default other settings (accessed on Oct/2016). The resultant table was analyzed manually.

3-2-12. Weighted gene co-expression network analysis

As a preprocessing procedure for a weighted gene co-expression network analysis (WGCNA)[90,181–183], a batch effect residing in the raw RNA-seq data in the Data Set 1 was observed in a principal component analysis by *pca(method = "svd")* function in pcaMethods R package[184]. Therefore, the batch effect in $\log_2$ (FPKM + 1)-transformed expression data of the whole genes was corrected as to the laboratories that had produced the raw data by *ComBat* function

in sva R package[185]. Then, attenuation of the batch effect was confirmed by the principal component analysis, and from the batch-corrected data, expression data of all the TFs in the Data Set 1 were extracted. Additionally, expression data of *ADARB1* and non-TF marker genes for several types of neurons and non-neuronal cells[90] were also extracted [**Materials and methods Table 6**].

After this batch correction, the WGCNA was performed using WGCNA R package[186] on the following RNA-seq data in the Data Set 1 according to the developer's website (https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/): cerebellum (n = 5)[163], frontal cortex (n = 5)[163], hippocampus (n = 4)[164], motor cortex (n = 7)[165], laser-captured MNs from the non-ALS controls (n = 7)[166] and temporal cortex (n = 3)[83]. The RNA-seq data from laser-captured MNs of patients with sporadic ALS (n = 13) in the Data Set 1 were excluded from this WGCNA to investigate modules of TFs in the state of the human central nervous tissues free from ALS. After confirmation of no outlier in all the data by hierarchical clustering, the following parameters were set: soft-thresholding power = 6, where mean connectiviey = 6.38, median connectivity = 3.88, max connectivity = 51.28; minimum cluster size = 25; deepSplit = 4. Branches of the dendrogram were then cut by the Dynamic Tree Cut method by *cutreeDynamic* function, and the eigengene network was visualized by *plotEigengeneNetworks* function. Correlation between modules and expression

levels of *ADARB1* and the marker genes was calculated by *cor* and *corPvalueStudent* functions with the significance threshold at 0.05. Bar plots and heat maps were drawn by *plot* function in the R and *heatmap.2* function in gplots R package[187], respectively. To seek TFs significantly belonging to the Module Magenta, Bonferroni correction was adopted by multiplying the raw *p*-values by the number of the modules (n = 16) to counteract the problem of multiple comparisons.

3-2-13. Scanning ChIP-seq signals and consensus motifs of TFs

Available ChIP-seq data of 571 TFs were manually downloaded one by one as bed files via ChIP-Atlas[188] for all cell types with the significance threshold at 50 (accessed on Jun/2017). After the downloaded bed files were re-annotated from hg19 to hg38 using the UCSC LiftOver, ChIP-seq signals within the promoter regions of *ADARB1* defined in this study were scanned by the *intersect -wb -a -b* function in the BEDtools2.

Consensus motifs of 821 TFs from integrated data within rtfbsdb R package[189], including data of Catalog of Inferred Sequence Binding Preferences database[51], and from manually collected data of JASPAR 2016[190] and other articles[52,53,55] were used to predict TF binding sites. After the position frequency matrices in those data sources were converted into the position weighted matrices in probability ratio by *toPWM* function in TFBSTools R

package[191], all the position weighted matrices were scanned around *ADARB1* gene (chr21:45,010,086-45,230,085; hg38) by *tfbs.scanTFsite* function with *threshold = 6* in the rtfbsdb R package. Predicted TF binding sites within the defined promoter regions of *ADARB1* were scanned from the output bed file by the *intersect -wb -a -b* function in the BEDtools2.

3-2-14. Search for TFs differentially expressed in ALS patients' tissues

The following published supplementary data of previous studies were analyzed, in which the authors had investigated differentially expressed genes between sporadic ALS patients and non-ALS healthy controls in high-throughput ways: the RNA-seq-based data from laser-captured MNs (13 cases, 9 controls)[166,168]; microarray-based data from laser-captured MNs and the remained anterior horns (12 cases, 10 controls), respectively[192]; RNA-seq-based data from the whole spinal cords (6 cases, 5 controls)[193]; and two sets of microarray-based data from motor cortices (31 cases, 10 controls[194]; and 11 cases, 9 controls[195]). From these supplementary tables and figure, all the TFs were extracted by their names or the provided probe IDs for the TFs. Changes in *ADARB1* expression were examined directly in these supplementary data. Their original criteria and methods for the statistical analysis were kept throughout this analysis. For the microarray data from the laser-captured MNs and the re-

mained anterior horns, the significance threshold was set at 0.05.

3-2-15. Seeking for TFs interacting with FUS protein

TFs whose mRNAs were bound by FUS proteins were extracted by their names from the published supplementary data of previous studies conducting crosslinking immunoprecipitation followed by RNA-seq (CLIP-seq) for FUS in human temporal cortex[196], HeLa cells [197] and HEK293 cells[198]. Similarly, TFs interacted by FUS proteins were extracted from the published data of quantitative mass spectrometry following to immunoprecipitation of FUS in HeLa cells[199]. The name of *ADARB1* was searched directly in these supplementary data.

3-2-16. Prediction of TFs targeted by miR-141 and miR-200a

To predict effects of miR-141 and miR-200a on expression of TFs, previous experimental data where these microRNAs had been overexpressed were downloaded via the Gene Expression Omnibus with the accession numbers GSM911074 (miR-141 transfection in A498 cells), GSM911072 (miR-200a transfection in A498 cells), GSM911081 (miR-141 transfection in Caki-1 cells) and GSM911078 (miR-200a transfection in Caki-1); while the former 2 data were thought to be unpublished, the latter 2 data were previously reported[200]. From these processed data, fold-change values of all the TFs were extracted, and the mean value was

calculated for every TF detected by multiple probes. Because ZEB1 and ZEB2 are established targets of miR-141 and miR-200a[201], a threshold for down-regulation was set to the same extent for their suppressed levels (0.75 fold-change), and its reciprocal number (1.33 fold-change) was used as a threshold for up-regulation. Fold-changes in expression of *ADARB1* were also extracted and analyzed in the same way above. Additionally, TFs that might be targeted by these two microRNAs was predicted *in silico* using TargetScan Human 7.1[202] (http://www.targetscan.org/vert_71/); after querying with the microRNA names on its website, all the TFs were screened by their names from the downloaded resultant tables.

3-2-17. Plasmid construction of the selected TFs

All expression vectors of the selected 48 candidate TFs were constructed in the same way by Dr. Yamashita, Ms. Teramoto and Mr. Hosaka. The coding sequence of the TFs was amplified using KOD -Plus- Ver. 2 (TOYOBO) from the template cDNA of HEK293 cells, HeLa cells, SH-SY5Y cells or human spinal cord. The gel-purified PCR products were cloned into a pCI mammalian expression vector (Promega) using In-Fusion® HD Cloning Kit (Takara Bio). The MAX Efficiency™ DH5α™ Competent Cells (ThermoFisher Scientific) transformed by the plasmid were proliferated overnight under selection with carbenicillin in LB medium (Sigma-Aldrich) at 37 °C. After the plasmid was purified using QIAprep® Spin

Miniprep Kit (QIAGEN), the insert sequence was verified by DNA sequencing [**Materials and methods Table 7**].

3-2-18. Luciferase assay

Among the four defined *ADARB1* promoters, the p1, the p3/p4 and the p13 promoters were constructed in the same way. First, the DNA sequence of every promoter was amplified from the human genomic DNA using KOD -Plus- Ver. 2 (TOYOBO) with the following primers shown from the 5' to the 3' end: TTTTTTGGTACCCAGCCGCGGTCTCTCAGC (30 nt; forward) and TTTTTTAAGCTTGCACCCTCGCTTCTCCGC (30 nt; reverse) for amplification of the p1 promoter; TTTTTTGGTACCATTGATAGATTTTTTGTATTAGGATTTC (40 nt; forward) and TTTTTTAAGCTTCCGACGCCCCATGATGCTGAAAAAGGT (39 nt; reverse) for the p3/p4 promoter; TTTTTTGGTACCTAATGGCTTGCTGGTTGAAAACGC (36 nt; forward) and TTTTTTAAGCTTTCTCCATGAAAAGTCTTCTAAATACA (38 nt; reverse) for the p13 promoter. Conditions of PCR were as follows: pre-incubation at 94 °C for 2 min and amplification in 25 cycles of denaturation at 98 °C for 10 sec and annealing and extension at 68 °C for 1 min (for the p1 promoter); pre-incubation at 94 °C for 2 min and amplification in 35 cycles of denaturation at 98 °C for 10 sec, annealing at 65 °C for 30 sec and extension at 68 °C for 30 sec (for the p3/p4 and the p13 promoters, respectively). The

gel-purified PCR products were digested with restriction enzymes *Kpn*I-HF and *Hind*III (New England Biolabs). On the other hand, luciferase vector pGL4.15[*luc2P*/Hygro] (Promega) was also digested with the same restriction enzymes and dephosphorylated using Alkaline Phosphatase, Calf Intestinal (New England Biolabs). Then, the digested amplicons and the luciferase vector were ligated using Quick Ligation Kit™ (New England Biolabs), respectively. The succeeding purification of the luciferase plasmids was conducted in the same way as the expression vectors were purified.

Construction of the other *ADARB1* promoter defined in the present study, the p2 promoter, was performed by Dr. Yamashita and Ms. Teramoto. To circumvent a persistent mutation within the defined region, a wider region including the defined one was amplified from the human genomic DNA using KOD -Multi & Epi- (TOYOBO) with the following primers shown from the 5' to the 3' end: TGGCCTAACTGGCCGG-GAGGCCAGGGCTAGTACAA (35 nt; forward) and TCTTGA-TATCCTCGAAACAACTGCTCTTAAGACAA (35 nt; reverse). PCR was performed in the following condition: pre-incubation at 94 °C for 2 min and amplification in 35 cycles of denaturation at 98 °C for 10 sec and annealing and extension at 68 °C for 1.5 min. The gel-purified PCR products were digested with restriction enzymes *Drd*I and *Alw*NI (New England Biolabs), and the yielded fragment was blunted with Klenow Fragment (Takara Bio)

according to the manufacturer's instruction, thereby gaining the insert sequence of the defined p2 promoter. After the luciferase vector pGL4.15[*luc2P*/Hygro] were digested with *Kpn*I-HF and *Xho*I, blunted with Klenow Fragment and then dephosphorylated with Alkaline Phosphatase, Calf Intestinal, the insert sequence and the prepared luciferase vector were ligated using Quick Ligation Kit™. The MAX Efficiency™ DH5α™ Competent Cells transformed by this plasmid were proliferated in the same way above, and from these cells plasmids were purified. Because all the purified luciferase plasmids had the inverse sequence of the defined p2 promoter, these plasmids were digested with *Eco*RV and *Sfi*I (New England Biolabs) and blunted with Klenow Fragment. This digested product and the pCI vector that had been digested with *Eco*RI (New England Biolabs) and blunted with Klenow Fragment were then ligated in the same way. The purified vector harboring the inverse sequence of the p2 promoter was digested with *Kpn*I-HF and *Xho*I. After this fragment and pGL4.15[*luc2P*/Hygro] digested with *Kpn*I-HF and *Xho*I were ligated in the same way above, this product was transformed into the One Shot™ Stbl3™ Chemically Competent *E. coli* (ThermoFisher Scientific), and these cells were proliferated under selection with carbenicillin in LB medium at 30 °C for two days. From these cells, the luciferase plasmids of the p2 promoter were purified in the same way above.

All procedures of the following luciferase assay were performed by Ms. Teramoto.

After HeLa cells were seeded into 96-well plates, the cells were transfected with 0.1 μg of each plasmid of the *ADARB1* promoters and 0.1 μg of each plasmid of the TF expression vectors using Lipofectamine® 3000 Transfection Reagent (Invitrogen). The empty pCI expression vector was used as a negative control. The cells in MEM-alpha medium (WAKO) with 10% fetal bovine serum (Invitrogen), 100 U/mL penicillin and 100 μg/mL streptomycin (Invitrogen) were cultured in 5% CO_2 at 37 °C for 72 hours. After this, the luciferase activity was measured with GloMax® Navigator (Promega) using the Luciferase Assay System (Promega) according to the manufacturer's instruction.

3-2-19. Visualization of CAGE data of the human primary memory T cells

The CAGE signals of human primary $CD4^+$ $CD25^{high}$ $CD45RA^-$ regulatory and $CD4^+$ $CD25^-$ CD45RA- conventional memory T cells before and after *in vitro* expansion[203] around *ADARB1* were extracted from the aforementioned annotated summary data downloaded from the FANTOM5 Consortium website. After bedGraph files in hg38 were produced using this data and the LiftOver, the CAGE signals around *ADARB1* were visualized on the UCSC-GB.

3-2-20. Additional data analyses for the TFs examined by the luciferase assay

Data of protein-protein interactions among the examined TFs was downloaded from

STRING[204] version 10.5 and visualized using Cytoscape[205] version 3.5.1; this data was based on the experiments, databases and co-expression (> 0.15 confidence). Data of tissue-specificity in expression of the TFs and their preference to be a hub were quoted from Supplementary Table S1 of Ravasi T. *et al*. *Cell* 2010[48]. Data of differentially expressed TFs with aging was extracted from Supplementary Table S12 of Mele M. *et al*. *Science* 2015[86]. Data of SNPs at eQTL of the TFs in human tissues determined by the Genotype-Tissue Expression (GTEx) Consortium[158] were downloaded via the FANTOM CAT v1.0.0 Extended View v1 hg19, and the effect of the alternative allele at these eQTL on expression of the TFs was manually surveyed by browsing the GTEx Consortium website version V6p (https://www.gtexportal.org/home/). Data of SNPs observed in patients with sporadic ALS were retrieved from ALS Variant Server[206] (http://als.umassmed.edu/), ALS Data Browser[207] version 2 (http://alsdb.org/) and ALS Gene[208] (http://alsgene.org/), respectively; only the ALS Variant Server provided data of the numbers of alleles counted in the patients. All of these web resources mentioned here were accessed on Jul/2017.

Materials and methods Table 1. RNA-seq data in the Data Set 1.

Sample	#	Donor	Age	Data.ID	Accession	Ref
Cerebellum	1	Female	62	90_cereb	GSM1642318	
	2	Female	65	24_cereb	GSM1642314	
	3	Male	76	86_cereb	GSM1642316	
	4	Female	73	91_cereb	GSM1642320	
	5	Male	81	93_cereb	GSM1642322	1
Frontal cortex	1	Female	62	90_FCX	GSM1642319	
	2	Female	65	24_FCX	GSM1642315	
	3	Male	76	86_FCX	GSM1642317	
	4	Female	73	91_FCX	GSM1642321	
	5	Male	81	93_FCX	GSM1642323	
Hippocampus	1	Female	77	CTRL1	GSM1645000	
	2	Male	83	CTRL2	GSM1645001	2
	3	Male	> 90	CTRL3	GSM1645002	
	4	Female	85	CTRL4	GSM1645003	
Motor cortex	1	Adult	64	CTRL-12	GSM2100633	
	2	Adult	46	CTRL-3	GSM2100628	
	3	Adult	42	CTRL-4	GSM2100629	
	4	Adult	50	CTRL-8	GSM2100630	3
	5	Adult	48	CTRL-9	GSM2100631	
	6	Adult	57	CTRL-11	GSM2100632	
	7	Adult	59	CTRL-13	GSM2100634	

Materials and methods Table 1 (cont.). RNA-seq data in the Data Set 1.

Sample	#	Donor	Age	Data.ID	Accession	Ref
Spinal motor neurons	1	Male	78	10_Control	GSM1977027	
Healthy control	2	Male	77	39_Control	GSM1977030	
	3	Female	80	44_Control	GSM1977033	
	4	Male	82	65_Control	GSM1977028	
	5	Male	77	67_Control	GSM1977031	
	6	Male	68	76_Control	GSM1977032	
	7	Female	58	78_Control	GSM1977029	
Spinal motor neurons	1	Male	61	16_sALS	GSM1977044	
Patients	2	Male	84	21_sALS	GSM1977040	
(sporadic ALS)	3	Male	74	27_sALS	GSM1977045	4
	4	Female	81	34_sALS	GSM1977041	
	5	Male	67	48_sALS	GSM1977046	
	6	Female	58	60_sALS	GSM1977035	
	7	Male	52	62_sALS	GSM1977036	
	8	Male	68	63_sALS	GSM1977037	
	9	Male	55	79_sALS	GSM1977042	
	10	Male	54	82_sALS	GSM1977043	
	11	Female	77	85_sALS	GSM1977047	
	12	Female	56	84_sALS	GSM1977038	
	13	Male	36	89_sALS	GSM1977039	
Temporal cortex	1	Adult	63	63yo	GSM1901345	
	2	Adult	45	45yo	GSM1901344	5
	3	Adult	53	53yo	GSM1901347	

All the accession numbers are for Gene Expression Omnibus.
Ref: 1, Prudencio M. *et al*. *Nat*. *Neurosci*. 2015; 2, Magistri M. *et al*. *J Alzheimers Dis*. 2015; 3, Lin L. *et al*. *Hum*. *Mol*. *Genet*. 2016; 4, Batra R. *et al*. *bioRxiv* 2016; 5, Zhang Y. *et al*. *Neuron* 2016.

Materials and methods Table 2. RNA-seq data in the Data Set 2.

Tissues outside the central nervous system (n = 24)

Sample	Donor	Age	Accession	Src	Ref
Adipose tissue (subcutaneous)	Male	54	ENCFF665BAC, ENCFF745QMZ	a	1
Adrenal gland	Male	54	ENCFF855MDA, ENCFF223ZRA		
Bone marrow	5a, V230	NA	ERR315425	b	2
Duodenum	4c, V150	NA	ERR315442		
Esophagus muscularis mucosa	Male	54	ENCFF035HHL, ENCFF897WFT	a	1
Gallbladder	5c, V182	NA	ERR315360	b	2
Gastrocnemius medialis	Male	54	ENCFF139SHO, ENCFF825PCO	a	1
Heart	Male	34	ENCFF001RTS, ENCFF001RTR		
Kidney	c, V23	NA	ERR315468	b	2
Liver	Female	53	ENCFF381FTA, ENCFF151LEE	a	1
Lung	Male	54	ENCFF901BQL, ENCFF460SWJ		
Lymph node	4a, V157	NA	ERR315371	b	2
Ovary	Female	53	ENCFF609SCM, ENCFF454PGN	a	1
Pancreas	Female	53	ENCFF556VCV, ENCFF444CRZ		
Placenta	6b, V223	NA	ERR315476	b	2
Prostate gland	Male	54	ENCFF714UWF, ENCFF638FRV	a	1
Sigmoid colon	Male	54	ENCFF326CGI, ENCFF663VCC		
Spleen	Male	54	ENCFF884ZSS, ENCFF908RHV		
Stomach	Male	54	ENCFF878UMC, ENCFF438GBR		
Testis	Male	54	ENCFF961NRE, ENCFF531RII		
Thyroid gland	Male	54	ENCFF086TFZ, ENCFF351OAS		
Urinary bladder	5b, V176	NA	ERR315355	b	2
Uterus	Female	53	ENCFF208XMO, ENCFF141YVJ	a	1
Whole blood (pooled)	5 Males	NA	SRX984188 (SRP056985)	c	3

Materials and methods Table 2 (cont.). RNA-seq data in the Data Set 2.

Tissues and primary cells within the central nervous system (n = 31)

Sample	Donor	Age	Accession	Src	Ref
Astrocyte*	Adult	63	GSM1901319	d	4
Oligodendrocyte precursor cell*	Adult	63	GSM1901338		
Oligodendrocyte (replicate 1)*	Adult	63	GSM1901335		
Oligodendrocyte (replicate 2)*	Adult	63	GSM1901336		
Microglia*	Adult	63	GSM1901341		
Cerebellum**	Female	62	GSM1642318		5
Hippocampus**	Female	77	GSE67333		6
Motor cortex (Brodmann's Area 4)**	Adult	64	GSM2100633		7
Superior frontal gyrus, Grey matter	Adult	NA	SRR1035125	c	8
Superior frontal gyrus, White matter	Adult	NA	SRR1035941		
Temporal lobe (Whole)**	Adult	63	GSM1901345	d	4
Spinal motor neuron (dissected)**	Healthy (n = 7)	58-82	GSE76220		9
Spinal motor neuron (dissected)**	sALS (n = 13)	36-84			

Materials and methods Table 2 (cont.). RNA-seq data in the Data Set 2.

iPSC-derived spinal motor neurons (n = 5)

Sample	Donor	Age	Accession	Src	Ref
FUS R521G iPSC-derived MNs	Patients with *FUS* R521G mutation	NA	GSM2056804	d	10
			GSM2056805		
			GSM2056806		
Wild-type iPSC-derived MNs	Healthy	NA	GSM2056807		
			GSM2056808		

iPSC, induced pluripotent stem cell; MN, motor neuron; NA, not available; *, primary cells; **, data also in the Data Set 1.

Src: a, Encyclopedia of DNA Elements; b, European Nucleotide Archive; c, NCBI Sequence Read Archive; d, Gene Expression Omnibus

Ref: 1, ENCODE Project Consortium *Nature* 2012; 2, Fagerberg L. *et al*. *Mol*. *Cell*. *Proteomics* 2014; 3, Zhao S. *et al*. *BMC Genomics* 2015; 4, Zhang Y. *et al*. *Neuron* 2016; 5, Prudencio M. *et al*. *Nat*. *Neurosci*. 2015; 6, Magistri M. *et al*. *J*. *Alzheimers Dis*. 2015; 7, Lin L. *et al*. *Hum*. *Mol*. *Genet*. 2016; 8, Mills J. D. *et al*. *PLoS One* 2013; 9, Batra R. *et al*. *bioRxiv* 2016; 10, Kapeli K. *et al*. *Nat*. *Commun*. 2016.

Materials and methods Table 3. Adult tissue samples for calculating correlation among the CTSSs around *ADARB1*.

Tissue type	Sample			
Body fluid (2)	Blood (1)	Cerebrospinal fluid (1)		
Circulatory (9)	Aorta (1)	Artery (1)	Heart (1)	Heart, mitral valve (1)
	Heart, pulmonic valve (1)	Heart, tricuspid valve (1)	Left atrium (1)	Left ventricle (1)
	Vein (1)			
Digestive (10)	Colon (2)	Esophagus (1)	Gall bladder (1)	Liver (1)
	Pancreas (1)	Parotid gland (1)	Salivary gland (1)	Small intestine (1)
	Submaxillary gland (1)			
Endocrine (2)	Adrenal gland (1)	Thyroid (1)		
Eye (2)	Eye vitreous humor (1)	Retina (1)		
Immune (5)	Appendix (1)	Lymph node (1)	Spleen (1)	Thymus (1)
	Tonsil (1)			
Muscle (7)	Eye, inferior rectus (1)	Eye, lateral (1)	Eye, medial (1)	Eye, superior (1)
	Skeletal muscle (1)	Smooth muscle (1)	Soleus muscle (1)	
Nervous (other; 8)	Corpus callosum (1)	Optic nerve (1)	Pineal gland (3)	Pituitary gland (3)

Materials and methods Table 3 (cont.). Adult tissue samples for calculating correlation among the CTSSs around *ADARB1*.

Tissue type	Sample			
Nervous (60)	Amygdala (2)	Brain (2)	Caudate nucleus (3)	Cerebellum (3)
	Diencephalon (1)	Frontal lobe (1)	Globus pallidus (3)	Hippocampus (3)
	Insula (1)	Locus coeruleus (3)	Medial frontal gyrus (3)	Medial temporal gyrus (4)
	Medulla oblongata (3)	Nucleus accumbens (1)	Occipital cortex (2)	Occipital lobe (1)
	Occipital pole (1)	Olfactory region (1)	Paracentral gyrus (1)	Parietal cortex (1)
	Parietal lobe (3)	Pons (1)	Postcentral gyrus (1)	Putamen (4)
	Spinal cord (3)	Substantia nigra (3)	Temporal lobe (1)	Thalamus (4)
Reproductive (female; 6)	Breast (1)	Cervix (1)	Ovary (1)	Placenta (1)
	Uterus (1)	Vagina (1)		
Reproductive (male; 7)	Ductus deferens (1)	Epididymis (1)	Penis (1)	Prostate (1)
	Seminal vesicle (1)	Testis (2)		
Respiratory (4)	Lung (1)	Lung, right lower lobe (1)	Throat (1)	Trachea (1)
Urinary (3)	Bladder (1)	Kidney (1)	Urethra (1)	
Others (15)	Achilles tendon (1)	Adipose (4)	Adipose tissue (1)	Bone marrow (1)
	Cerebral_meninges (1)	Cruciate ligament (1)	Dura mater (1)	Fingernail (1)
	Skin (1)	Skin palm (1)	Tongue (1)	Tongue, epidermis (1)

The number of each sample is indicated in parentheses.

Materials and methods Table 4. GEO accession numbers of ChIP-seq data for 3 histone modifications.

Tissue/cell line	H3K27ac	H3K4me1	H3K4me3
Adipose	GSM916066	GSM669975	GSM670041
Hippocampus	GSM1112791	GSM669962	GSM670022
Liver	GSM1112809	GSM669972	GSM621675
Pancreas	GSM906397	GSM910576	GSM910581
Smooth muscle	GSM916065	GSM621642	GSM621645
Spleen	GSM906398	GSM910577	GSM910582
Substantia nigra	GSM1112778	GSM772898	GSM772901
HeLa cell	GSM733684	GSM798322	GSM733682

Materials and methods Table 5. Tissue samples for calculating correlation between *ADARB1* expression and TF.

Group	Sample	
Highest 10 in *ADARB1* expression	Aorta	Bladder
	Cerebellum (n = 3)	Dura mater
	Lung, right lower lobe	Pineal gland
	Smooth muscle	Vagina
Lowest 10 in *ADARB1* expression	Blood	Heart
	Kidney	Left ventricle
	Liver	Parotid gland
	Salivary gland	Small intestine
	Substantia nigra	Thyroid

Materials and methods Table 6. Marker genes used in the WGCNA.

Cell	Type	Module*	Top 10 non-TF genes				
Neurons	Neuron (common)	M16A	ATP6AP2	CCT2	GLS	GMFB	PAFAH1B1
			PRKAR1A	RAB5A	REEP5	SYNJ1	TPD52
	Glutamatergic	M10A	AP3B2	ATP6V0C	C9orf16	CACNB1	CLTB
			CYP46A1	DMTN	RUNDC3A	SLC17A7	SYP
	Interneuron	M17A	ASAP2	DYNLT3	EFR3A	ENTPD4	EPS15
			KCTD9	NARS	PLCB1	RCAN2	UQCRB
	Purkinje cell	M6D	CA8	CALB1	CHST2	ITPR1	LARGE
			LPL	LPPR4	PCP4	PLXDC1	PVALB
Non neural cells	Astrocyte	M15A	AGT	AHCYL1	ASCBG1	ATP1A2	GJA1
			METTL7A	NTRK2	PON2	PPAP2B	SLC1A3
	Oligodendrocyte	M9A	ENPP2	GPR37	GPRC5B	HSPA2	LAMP2
			MAL	PCSK6	PLLP	PMP22	SLC31A2
	Microglia	M4A	C1QA	C3	CD74	CSF1R	HLA-DPA1
			HLA-DRA	ITGB2	LAPTM5	RNASET2	TYROBP
	Meningeal cell	M43	CEMIP	DCN	DUSP1	FN1	FRZB
			IGFBP7	LEPR	NOV	PTGDS	VIM

TF, transcription factor; *, modules reported in Oldham M. C. *et al*. *Nat*. *Neurosci*. 2008.

Materials and methods Table 7. Sequence IDs of cloned TFs in this study.

TF	ID	TF	ID
AFF4	CCDS4164	MSX1	CCDS3378
AR	CCDS14387	MXI1	CCDS7564
ARNT	CCDS970	NFIA	CCDS44156
ATF4	CCDS13996	NR2F1	CCDS4068
CBFB	CCDS10827	NR2F2	CCDS10375
CBX3	CCDS5398	NR3C1	CCDS4278
CEBPA	CCDS54243	NR3C2	CCDS3772
CREB1	CCDS2374	RFX2	CCDS12157
EBF1	CCDS4343	RUNX1	CCDS13639
ELF2	CCDS82954	RXRG	CCDS72970
ENO1	CCDS97	SETBP1	CCDS11923
ESRRG	CCDS1517	STAT1	CCDS42793
FOXA1	CCDS9665	TCF4	CCDS11960
FOXB1	CCDS32255	TFCP2	CCDS8808
FOXD1	CCDS75259	YBX1	CCDS470
FOXP2	CCDS43635	YY1	CCDS9957
FOXP3	CCDS14323	ZEB1	CCDS44370
HIF1A	CCDS9753	ZEB2	CCDS2186
HOXB7	CCDS11532	ZFPM2	CCDS47908
ID4	CCDS4544	ZHX3	CCDS13315
KLF7	CCDS2373	ZNF25	CCDS7195
MAX	CCDS9772	ZNF672	CCDS1638
MEF2A	CCDS53978	ZNF781	CCDS12507
MLLT11	CCDS982	ZNF876P	NR_027481.1

AR has 21 CAG repeats (within the normal range).

4. RESULTS

4-1. Genome-wide profile of transcriptional activity in human MNs by the CAGE method

First of all, on assumption of species-, tissue- and cell-type-specific transcriptional activity and regulation of gene expression, in total 20,629 single human MNs were collected from autopsied spinal cords of two donors with laser capture microdissection[209] (11,100 and 9,529 MNs from each; i.e., two biological replicates) [**Results Figure 1A**]; the dorsal horns and white matter were macroscopically collected as controls, respectively. RIN of the replicated samples of the MNs, the dorsal horns and the white matter were 4.2, 4.6, 7.7, 6.1, 6.2 and 5.9, respectively [**Results Table 1A**]. Then, I genome-widely profiled transcriptional activity in these samples by the CAGE method. Since the starting amount of the total RNAs (0.45 ~ 2.75 μg) was below the requirement (5 μg), each mapped read count of the samples was in an order fewer than the average in the FANTOM5 data[178] [**Results Table 1B**]. Nevertheless, the resultant CAGE signals in the MNs were highly correlated between the biological replicates and lowly correlated with those in the dorsal horns or white matter [**Results Figure 1B**], suggesting that the gained CAGE signals in the MNs faithfully represented genome-wide transcriptional activity specific to these neurons.

4-2. Definition of promoters of *ADARB1*

As a crucial step in the present study, I defined active promoters of *ADARB1* in the human MNs based on the present CAGE data on the MNs and the previously annotated five CTSSs of *ADARB1*. This enables to survey both species-specificity in the DNA sequences and positions of the promoters of the ADAR2 genes and tissue-specificity in activity of the *ADARB1* promoters in the adult human tissues.

4-2-1. Definition of the active promoters of *ADARB1* in the human MNs

Overview of the CAGE signals around *ADARB1* in the MNs showed that the MNs consistently expressed these signals around the p1@ and the p2@, but not in the exact position of the p13@, the p3@ or the p4@ [**Results Figure 2A, bottom**], suggesting that regions harboring the former two CTSSs were active promoters of *ADARB1* in the MNs [**Results Figures 2B-D; Results Table 2**].

A single weak CAGE signal in the MNs was consistently detected at the p2@, which had no obvious PROMPT signal in the CAGE total counts [**Results Figure 2B**], probably due to technical limitation to detect unstable nascent RNAs in the CAGE method[9]. A promoter of *ADARB1* that was active in the MNs and harbored the p2@ was designated as p2 promoter and defined within a single DHS from approximately 300 bp upstream of the 5' end of the p2@ to 100 bp downstream of the 3' end of this CTSS (length: 412 bp); this definition was

according to a previous report[210] and consistent with other previous reports that an active CTSS was flanked by the closest nucleosomes with the H3K4me3 modification at -250 ~ -350 bp and about +100 bp[72,211], respectively. This p2 promoter has TATA-box (TATAAAAA) in a position unusually distant from the p2@ (more than -110 bp, compared to the canonical -33 to -28 bp, relative to the CTSS[6,11]). The DNA sequence of the p2 promoter was highly conserved in a rhesus and partially in a marmoset, but not in a mouse.

The CAGE signals in the region around the p1@ suggested that this region was also an active promoter of *ADARB1* in the MNs [see **Results Figure 2B**]. This active promoter in the MNs, designated hereafter as p1 promoter of *ADARB1*, was defined within a single DHS from 200 bp upstream of the 5' end of the p1@ to 320 bp downstream of the 3' end of this CTSS (length: 596 bp), which encompassed all the CAGE signals around the p1@ in the MNs and the previously defined two CTSSs p2@SSR4P1 and the p3@SSR4P1[10,22,30] of the putative *ADARB1* PROMPT. The DNA sequence of the p1 promoter was modestly conserved with a marmoset and partially and weakly conserved with a mouse, whereas the sequence was in no homology with a rhesus. This p1 promoter resided within a CpG island.

No CAGE signals were detected in the MNs at the p13@, which lacked an obvious PROMPT signal in the CAGE total counts [**Results Figure 2C**]. I defined a region from 300 bp upstream of the 5' end of the p13@ to 100 bp of the 3' end of this CTSS as p13 pro-

moter of *ADARB1* (length: 435 bp) whose DNA sequence was highly conserved among the primates and to a lower extent with a mouse.

Similarly, in the MNs, no CAGE signals were detected at the p3@ nor the p4@, which had no obvious PROMPT signal in the CAGE total counts [**Results Figure 2D**]. I defined a region within a single DHS from 300 bp upstream of the 5' end of the p3@ to 100 bp downstream of the 3' end of the p4@ as p3/p4 promoter of *ADARB1* (length: 455 bp), whose DNA sequence was highly conserved throughout the primate species and modestly with a mouse.

In addition, the overview of the CAGE signals around *ADARB1* in the MNs demonstrated that many MN-specific CAGE signals were detected within and upstream regions of the gene body of *ADARB1*, implying the positions of putative MN-specific enhancers [see **Results Figure 2A, bottom**]. Accordingly, although none of these signals was at the exact positions of the previously detected CTSSs[10,22,30], seven out of the 17 CAGE signals in the non-coding regions around *ADARB1* in the MNs settled within the previously defined DHSs[38].

4-2-2. Species-specificity of the promoters of the ADAR2 gene between humans and mice

To survey difference in promoters of mouse ADAR2 gene *Adarb1* from the human counter-

parts, I investigated CTSSs of *Adarb1* defined by the FANTOM5 Consortium[10,22,30] [**Results Figure 3A**], revealing that among its three annotated CTSSs, only p1@Adarb1 and p2@Adarb1 had similar features to the human p1@ (e.g., their existence within a CpG island and presence of putative PROMPT *Gm17769* in their 70 ~ 90 bp upstream), but the DNA sequence around these CTSSs was poorly conserved with humans [**Results Figure 3B**]. On the other hand, the position of the other p5@Adarb1 did not correspond to any of the human p2@, p3@, p4@ or p13@.

In conclusion, because the repertoire of the motoneuronal TFs and the promoters of the ADAR2 gene of humans were not conserved with mice, experiments on the regulatory mechanism of expression of mouse ADAR2 gene in the MNs could be of little help, and only analyses specific to the human MNs would elucidate the regulatory mechanism of *ADARB1* expression in the human MNs.

4-2-3. Tissue-specific activity of the *ADARB1* alternative promoters

To infer specificity in activity of the *ADARB1* promoters among human adult tissues, I investigated transcriptional activity at the annotated five CTSSs of *ADARB1* using the publicly available CAGE data from 140 different tissue samples of adults[10,22,30,178]. Transcription at the p1@ within the p1 promoter was ubiquitously active among the tissues; at the p2@ within

the p2 promoter and at the p3@ and p4@ within the p3/p4 promoter, transcription was less ubiquitously active, respectively; and transcription at the p13@ within the p13 promoter was specifically detected in some of the nervous tissues [**Results Figure 4A**].

To survey co-expression of the five CTSSs of *ADARB1* in the human nervous tissues, I calculated correlation in expression of all the CTSSs within the 231.8 kb region around the *ADARB1* using the CAGE data on the 60 nervous tissue samples within the aforementioned public CAGE data. It demonstrated that the p1@, the p13@ and the p2@ were co-expressed in the 60 nervous tissue samples independently from the p3@ and the p4@ [**Results Figure 4B**]. Consistently, as transcriptional activity of co-regulated promoters and that of their enhancers are highly correlated[22,28], this result also showed that the p1@, the p2@ and the p13@ may be regulated by a different set of enhancers from the p3@ and the p4@ in the nervous tissues. This independency of the p3@ and the p3@ was also observed even in calculation using the CAGE data on all the 140 tissue samples. Indeed, previous CAGE data from human primary $CD4^+$ $CD25^{high}$ $CD45RA^-$ regulatory and $CD4^+$ $CD25^-$ $CD45RA^-$ conventional memory T cells[203] showed the otherwise quiescent p3@ and p4@ were independently activated during the *in vitro* expansion[212] [**Results Figure 4C**], suggesting independent and dynamic regulation of the p3/p4 promoter depending on a biological state of the cells.

To survey tissue-specific activity of the *ADARB1* promoters from a viewpoint of the

histone modifications, I additionally analyzed the publicly available CAGE[10,22,30,178] and ChIP-seq[39] data for the active enhancer markers (H3K27ac and H3K4me1) and the active promoter marker (H3K4me3) in human seven tissues (adipose, hippocampus, liver, pancreas, smooth muscle, spleen and substantia nigra) and the HeLa-S3 cell line [**Results Figure 5**]. Among these, the p1@ was the most active CTSS in all the samples but the adipose tissue where the p2@ was more active than the p1@; the p3@ and the p4@ were weakly or scarcely expressed and the p13@ was inactive in all of these data. In accord with the lower CAGE signal at the p1@ in the adipose than that in the HeLa-S3 cell line, the commercially available total RNA from an adipose tissue indeed expressed fewer mRNAs of ADAR2 than HeLa cells (4.9×10^{-3} in adipose and 1.2×10^{-2} in HeLa cells, respectively, when the expression level of ADAR2 was quantified and then normalized to that of beta-actin by quantitative PCR). As shown in **Results Figure 5A**, ChIP-seq peaks of the H3K27ac and H3K4me1 appeared at different positions in different tissues and cell line, implying tissue- and cell-type-specific combinatorial usage of enhancers[22] for regulating *ADARB1* expression. In addition, a closer view around the p1@ and the p2@ revealed multiple peaks of the H3K4me3 around these CTSSs [**Results Figure 5B**]; the patterns of the multiple peaks were quite different between the adipose tissues actively expressing the p2@ and the other tissues dominantly expressing the p1@, demonstrating the difference in this histone modification reflecting the different ac-

tivity of these two CTSSs.

In conclusion, the results of these investigations altogether suggested the tissue-specific regulation of the *ADARB1* promoters by enhancers specific to that tissue.

4-3. Definition of the human motoneuronal TFs

To define the human motoneuronal TFs, I analyzed the present CAGE data on the human MNs and publicly available RNA-seq data on the laser-captured MNs of seven non-ALS adults[166] for complementing the relatively limited CAGE signals in the MNs. The CAGE data defined 1,173 motoneuronal TFs [**Results Figure 6**]; by adding 161 TFs uniquely detected in the public RNA-seq data, I defined in total 1,334 human motoneuronal TFs.

Next, to survey difference in the motoneuronal TFs between humans and mice, I analyzed RNA-seq data on approximately 2,000 laser-captured MNs of WT mice (n=3) [**Results Table 3**], defining 1,225 mouse motoneuronal TFs. Humans shared 74% of their motoneuronal TFs (988 out of 1,334) with mice; these shared motoneuronal TFs were in modest correlation in expression (Pearson's rho = 0.58, $p < 2.2e-16$) [**Results Figures 7A, left** and **7B**]. However, humans shared with mice only 33% (223 out of 668) of their TFs expressed greater than the median expression level of the human motoneuronal TFs [**Results Figure 7A, right**],

demonstrating species-specificity in the expression and repertoire of the motoneuronal TFs.

4-4. Prediction of motoneuronal TFs regulating *ADARB1* expression

A large number of the human motoneuronal TFs required to select a subset of them that were preferentially examined by *in vitro* luciferase assay. Thus, I predicted *ADARB1*-regulatory TFs in the MNs by multiple *in silico* analyses, based on (1) correlation in expression between each of the TFs and *ADARB1*, (2) a database for putative target genes of TFs, (3) correlation in expression between TF modules and *ADARB1*, (4) public data of ChIP-seq and consensus motifs of TFs, (5) TFs decreased in tissues of patients with sporadic ALS, and (6) possible molecular cascades of pathogenesis of mutant FUS-linked ALS.

4-4-1. Prediction from correlation in expression between each of the TFs and *ADARB1*

TFs correlatively expressed with *ADARB1* might regulate *ADARB1* expression. Therefore, I calculated the correlation between each of the TFs and *ADARB1* using the previous CAGE data from highest and lowest 10 human tissue samples expressing *ADARB1*[10,22]; this resulted in 32 TFs above the threshold (|Spearman's rho| ≥ 0.7), 24 of which were motoneuronal ones [**Results Table 4**].

4-4-2. Prediction from a database for putative target genes of TFs

Database ORTI[180] informs putative TFs targeting a particular gene based on previous low/high-throughput experiments and computational prediction. For *ADARB1*, this database showed six putative TFs as its regulator (AR, E2F4, ETS1, FOXP3, GATA2 and TP53), among which AR and E2F4 were the motoneuronal TFs.

4-4-3. Prediction from correlation in expression between TF modules and *ADARB1*

To find a subset of TFs ("module") regulating *ADARB1* in neurons in the central nervous tissues, I conducted WGCNA that enabled to dissect genes expressed in heterogeneous tissues into multiple modules that may have tissue- or cell-type-specific function[85,90,181–183]. Unfortunately, the number of available RNA-seq data from the laser-captured MNs (n= 7) of non-ALS adults[166] was below the requirement to perform this WGCNA meaningfully (n > 15). Therefore, I conducted WGCNA using expression profiles of all TFs detected at least one sample by RNA-seq on cerebellum (n = 5)[163], frontal cortex (n = 5)[163], hippocampus (n = 4)[164], motor cortex (n = 7)[165], the laser-captured MNs (n = 7)[166] and temporal cortex (n = 3)[83] in the Data Set 1 (in total 1,986 TFs). The WGCNA divided these TFs into 15 modules (colored) and one unclassified group (Module Grey) [**Results Figure 8A**]; 10 of these 15 colored modules formed the largest cluster based on their first principal component[213] ("module

eigengene") [**Results Figure 8B**], indicating inter-modular functional relationship within this module cluster. On the other hand, module eigengene of Module Magenta clearly distinguished the MNs from the other central nervous tissues [**Results Figure 8C**]; four TFs negatively contributing to this eigengene were expressed only in the MNs, one of which was a well-known MN-marker gene *MNX1* encoding Hb9 protein [**Results Figure 8D**].

Next, to find putative neuron-specific modules regulating *ADARB1* expression, I examined correlation of all the module eigengenes with expression of *ADARB1*, an MN-marker gene *CHAT* and each of 10 other marker genes for eight cell-types (neurons, glutamatergic neurons, interneurons, Purkinje cells, astrocytes, oligodendrocytes, microglial cells and meningeal cells[90]) in these central nervous data, respectively [**Results Figure 8E**]. Even though I failed to find correlation between the expression of *CHAT* and any module eigengenes due to its low expression levels, the eigengenes of the six Modules correlating various neuronal marker genes (Turquoise, Blue, Brown, Red, Green and Black) were strongly and positively correlated with the expression of *ADARB1*. All of these six modules within the largest module cluster embraced 91.6% of the TFs (339/370) whose expression was significantly correlated with the *ADARB1* expression in this WGCNA. Because 235 out of 339 were the motoneuronal TFs, they were regarded as putative *ADARB1*-regulatory TFs in the MNs.

4-4-4. Prediction from public data of ChIP-seq and consensus motifs of TFs

Possibility of TF for binding to the *ADARB1* alternative promoters could be predicted from the presence of their ChIP-seq signals and their consensus motifs within the defined regions. Among the 340 motoneuronal TFs whose ChIP-seq data were available in ChIP-Atlas[188], no more than 60 TFs similarly bound to the p1, p2 and p3/p4 promoter, respectively [**Results Table 5**], suggesting that the p3/p4 promoter is likely regulated as actively as the p1 and the p2 promoters are done in the cells. However, only MAX bound to the p13 promoter in the available data. On the other hand, among the 328 motoneuronal TFs whose consensus motifs were available in public[52,53,55,189,190], the motifs of less than 150 TFs were detected within each of the *ADARB1* promoters [see **Results Table 5**], in many cases of which these positions were overlapped by the corresponding ChIP-seq signals, validating this *in silico* motif search.

4-4-5. Extraction of TFs decreasing in the tissues of patients with sporadic ALS

Because TFs decreasing in tissues of patients with sporadic ALS might be involved in the down-regulation of *ADARB1* in those tissues, I extracted such down-regulated TFs from the published high-throughput data on tissues of the patients and the non-ALS controls: 1) RNA-seq data from laser-captured MNs (13 cases, 9 controls)[166,168]; 2) microarray data from laser-captured MNs and the remained anterior horns (12 cases, 10 controls)[192]; 3) RNA-seq

data from the whole spinal cords (6 cases, 5 controls)[193]; 4) microarray data from motor cortices (31 cases, 10 controls)[194]; and 5) another microarray data from motor cortices (11 cases, 9 controls)[195]. Conducting unsupervised hierarchical clustering based on gene expression, Aronica E. *et al*. reported that all the samples were divided into 3 groups: SALS1 group (18 cases and 1 control), SALS2 group (13 cases and 2 controls) and Control group (7 controls)[194]. As a result of this analysis, in total 619 motoneuronal TFs were regarded as candidates for the *ADARB1*-regulatory TFs [**Results Table 6**].

4-4-6. Extraction of TFs in possible molecular cascades of pathogenesis of mutant FUS-linked ALS

RNA-binding protein FUS[196,198,214] regulates gene expression via alternative splicing[168,215,216] and polyadenylation[217] (extensively reviewed elsewhere[218]); and ALS-linked mutant FUS deteriorates such biological functions[197,199,219], and also forms abnormal aggregation in the cytoplasm[97,98]. Another report suggested the role of long-term effects from aberrant FUS-miR-141/200a feed-back loop in the pathogenesis of mutant FUS-linked ALS[220]. It has been unclear, however, whether mutant FUS causes ALS phenotype through disruption of these known biological functions. A recent report demonstrated that expression of *ADARB1* is decreased in MNs of an ALS patient carrying an FUS^{P525L} mutation[110] raised the possibility

that mutant FUS may have a role in *ADARB1* regulation. To assess the role of mutant FUS in down-regulation of *ADARB1*, I investigated the possibility that mutant FUS may facilitate sequestration of either mRNAs (Putative link I) or proteins (Putative link II) of the regulatory TFs and/or *ADARB1* into cytoplasmic aggregation or (Putative link III) suppress the regulatory TFs and/or *ADARB1* through miR-141 and/or miR-200a up-regulation.

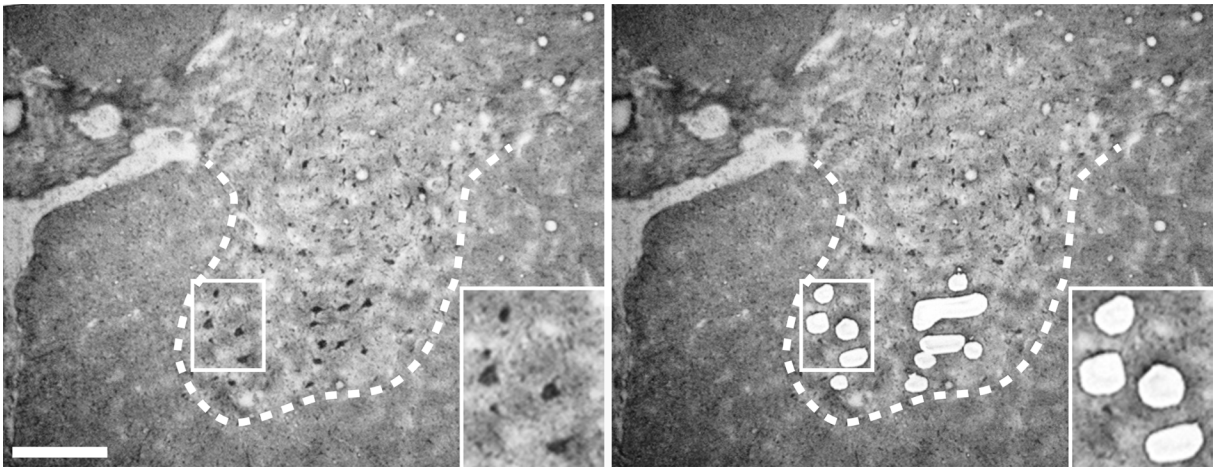
For the Putative link I, mRNAs of 834 and 579 motoneuronal TFs in total bound to WT and mutant FUS (FUS^{R521G} or FUS^{R521H}), respectively, in previous data of CLIP-seq for FUS in human temporal cortex[196], HeLa cells[197] or HEK293 cells[198] [**Results Table 7**]. These public data also showed that mRNAs of *ADARB1* bound to WT FUS in HeLa and HEK293 cells but not to mutant FUS in HEK293 cells. Another previous data[199] showed protein-level interaction of 48 motoneuronal TFs with FUS in HeLa cells (the Putative link II); ADAR2 does not interact with FUS in this cell line. According to microarray data of renal cell carcinoma cell lines with miR-141 and miR-200a overexpression[200], 99 ~ 191 motoneuronal TFs were suppressed by each of these microRNAs to the similar levels of their well-known targets ZEB1 and ZEB2, respectively (the Putative link III); TargetScan Human 7.1[202] predicted 220 ~ 255 TFs as putative targets of these two microRNAs. On the other hand, expression of *ADARB1* was not affected by miR-141 but up-regulated by miR-200a in both of the cell lines (1.4 ~ 1.5 fold-change).

4-5. *In vitro* luciferase assay of the selected TFs

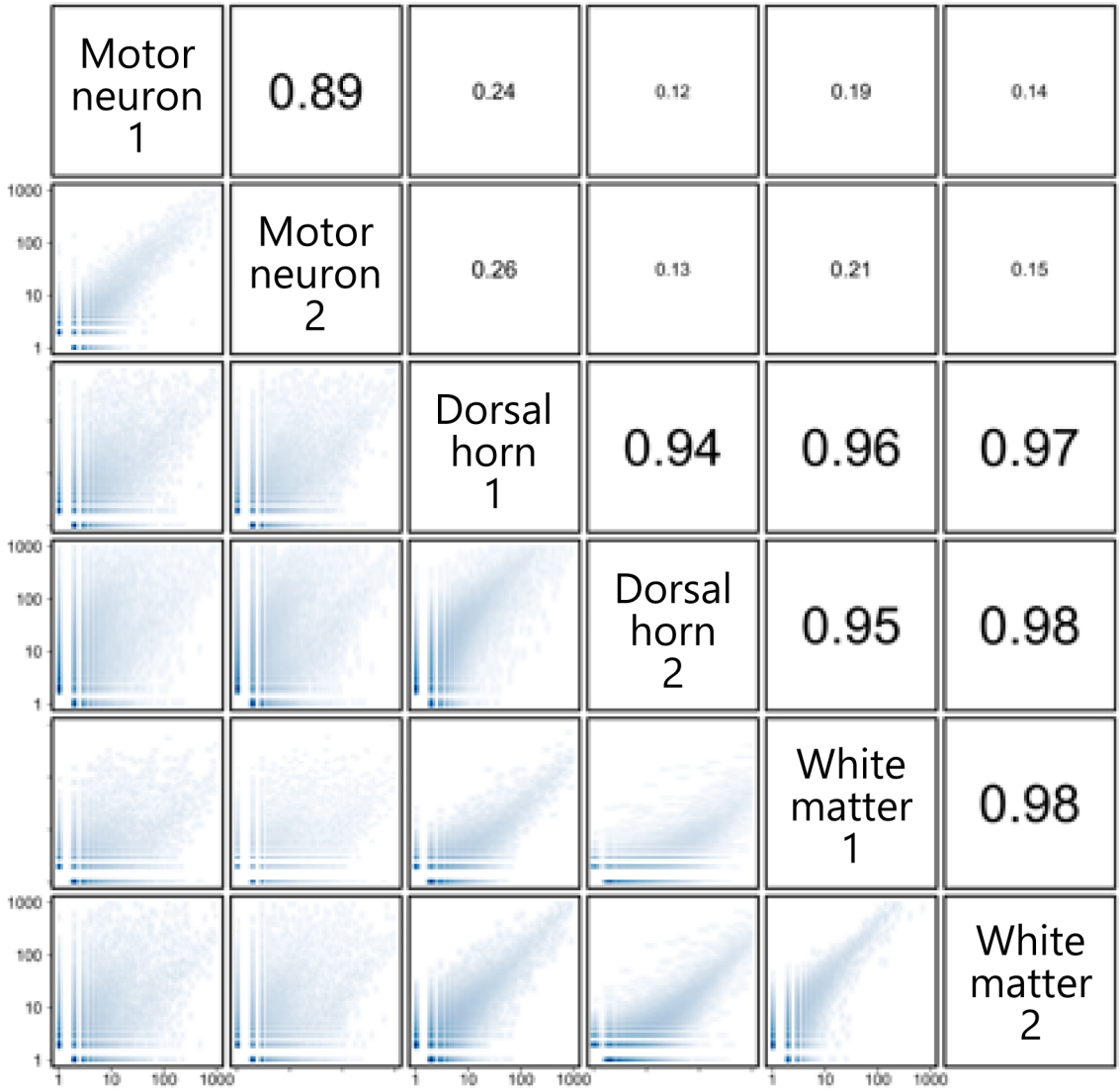
After these *in silico* predicting analyses, in total hundreds of motoneuronal TFs were listed up as candidates for *ADARB1*-regulatory TFs. Among these candidates, in total 34 motoneuronal and 14 non-motoneuronal TFs were selected to preferentially examine their regulating properties for the four *ADARB1* promoters by *in vitro* luciferase assay [**Results Figure 9**; **Results Tables 8A** and **8B**]. These candidate TFs included two motoneuronal and eight non-motoneuronal TFs that had been randomly selected before the present study, based on homology in DNA sequences around the first exon of *ADARB1* among a human, chimpanzee, mouse and rat.

Luciferase assay was conducted not only for the p2 and p1 promoters active in the MNs but also for the p13 and p3/p4 promoters to ask if these quiescent promoters in the MNs could be regulated by the same TFs as the active ones were done. Co-transfecting each of the selected TFs with each of the four promoters in the HeLa cells, respectively, revealed that all the four regions defined in the present study were competent as a promoter and each of them was regulated by a different set of the examined TFs [**Results Figure 10A**]. Among the examined 34 motoneuronal TFs, 15, 13, 21 and 23 TFs activated and 4, 8, 3 and 3 TFs suppressed transcriptional activity at the p2, the p1, the p13 and the p3/p4 promoters, respectively; 3 and 2 TFs activated and suppressed only one of the promoters, respectively; 21 and 5

TFs activated and suppressed multiple promoters, respectively (cutoff: 3/2 and 2/3 fold-change to the negative control for up- and down-regulation, respectively). Seven out of the 21 activators for multiple promoters activated all the four promoters (i.e., universal activators: ESRRG, ID4, KLF7, NFIA, NR3C1, NR3C2 and ZEB1); one out of five suppressors for multiple promoters suppressed all the promoters (i.e., a universal suppressor: YBX1). Only one motoneuronal TF, MXI1, exhibited no effect on any of the tested *ADARB1* promoters. On the other hand, among the 14 examined non-motoneuronal TFs, 5, 4, 7 and 7 TFs activated and 1, 6, 1 and 2 TFs suppressed the p2, the p1, the p13 and the p3/p4 promoters, respectively, and FOXD1 and FOXP3 activated all the four promoters [**Results Figure 10B**].

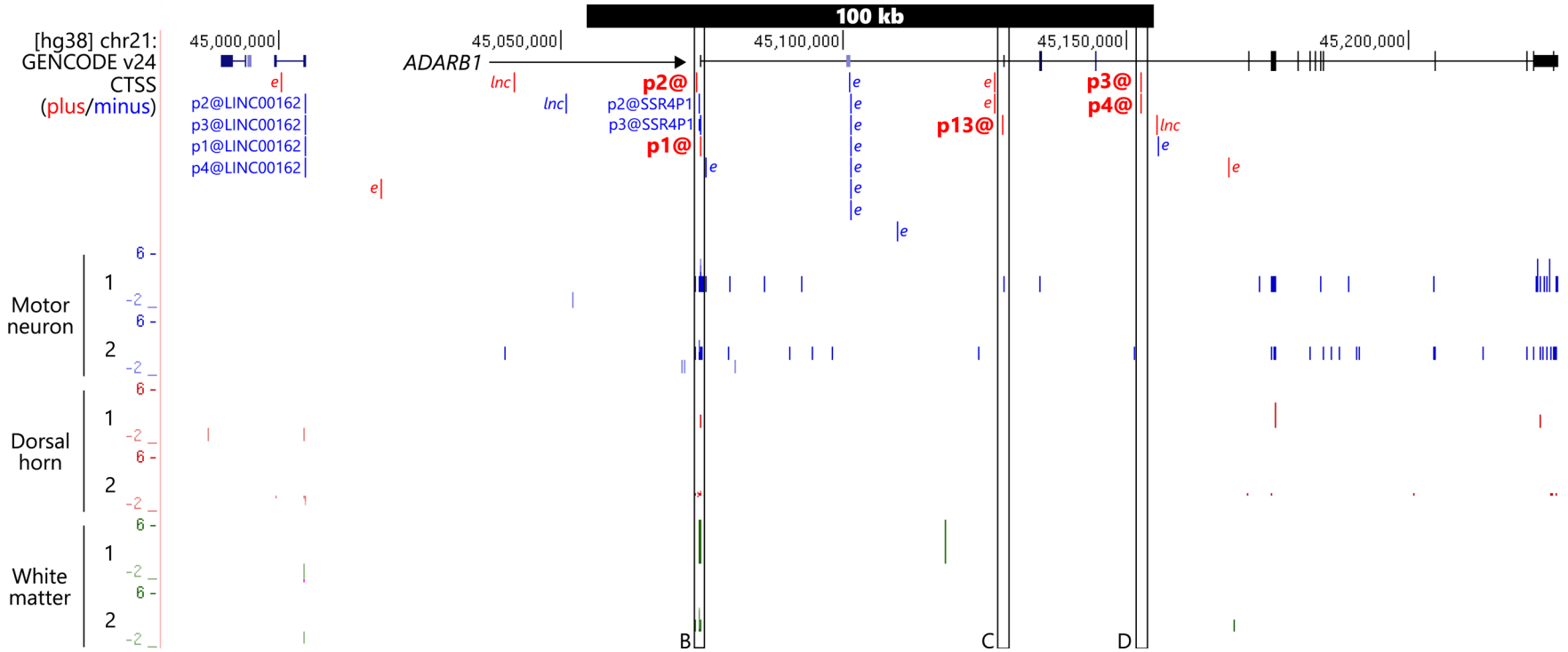


Results Figure 1A Laser capture microdissection of human spinal motor neurons (MNs). Larger anterior horn cells are regarded as MNs and selectively collected with the microdissector. Representative pictures before and after dissection are shown in left and right panels, respectively. The margin of the anterior horn is depicted with white dotted line. Inset in each panel shows a 2-fold-magnified view in a region indicated by the white rectangle. The scale (200 μm) is shown at the bottom left.



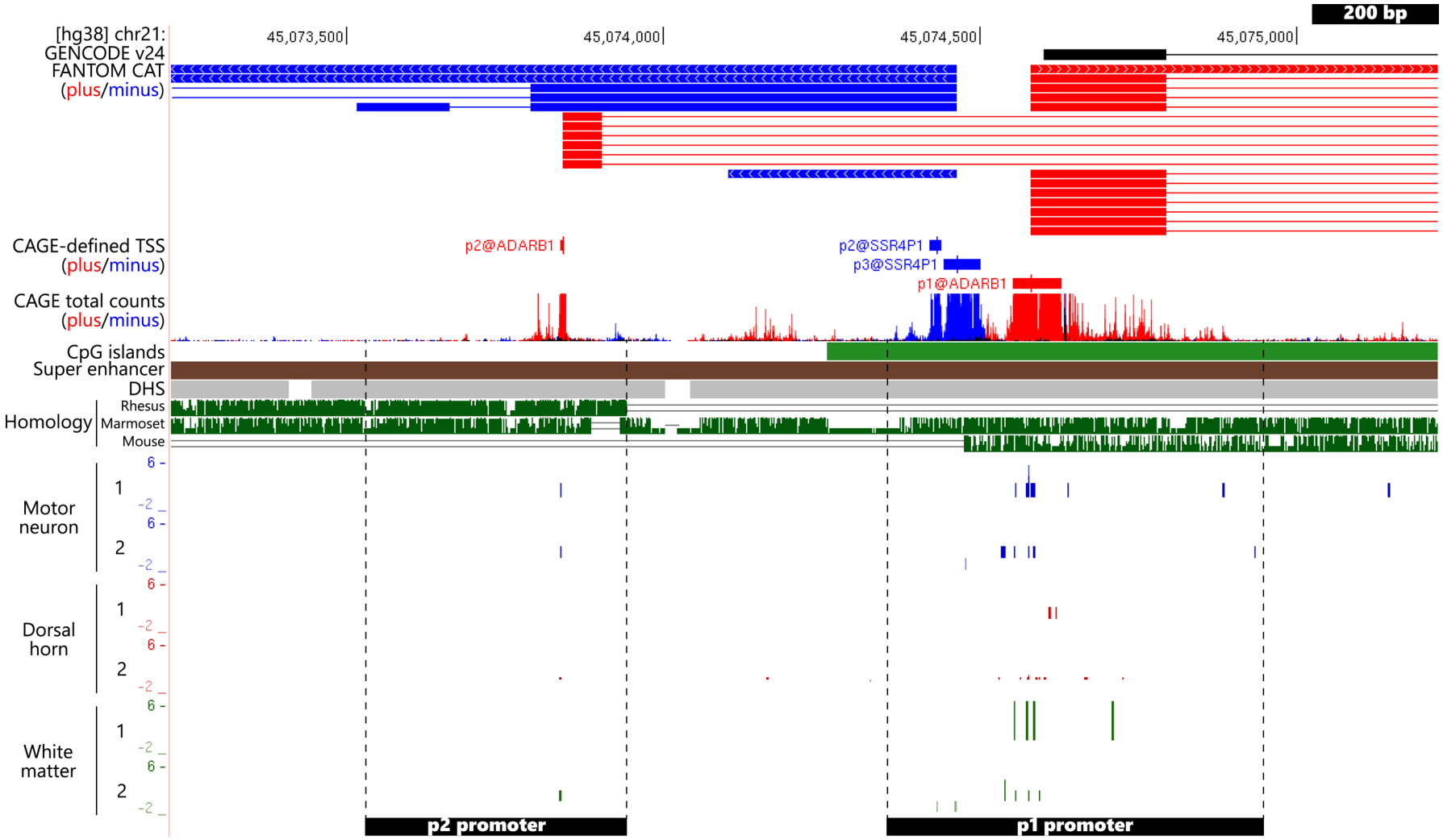
Results Figure 1B Correlation among CAGE signals detected in the laser-captured motor neurons, dorsal horns and white matter (aligned by BWA).

CAGE signals in the two biological replicates of the MNs are highly correlated with each other but just weakly with those in technical replicates of dorsal horns and white matter, demonstrating specificity of the CAGE signals in the MNs.



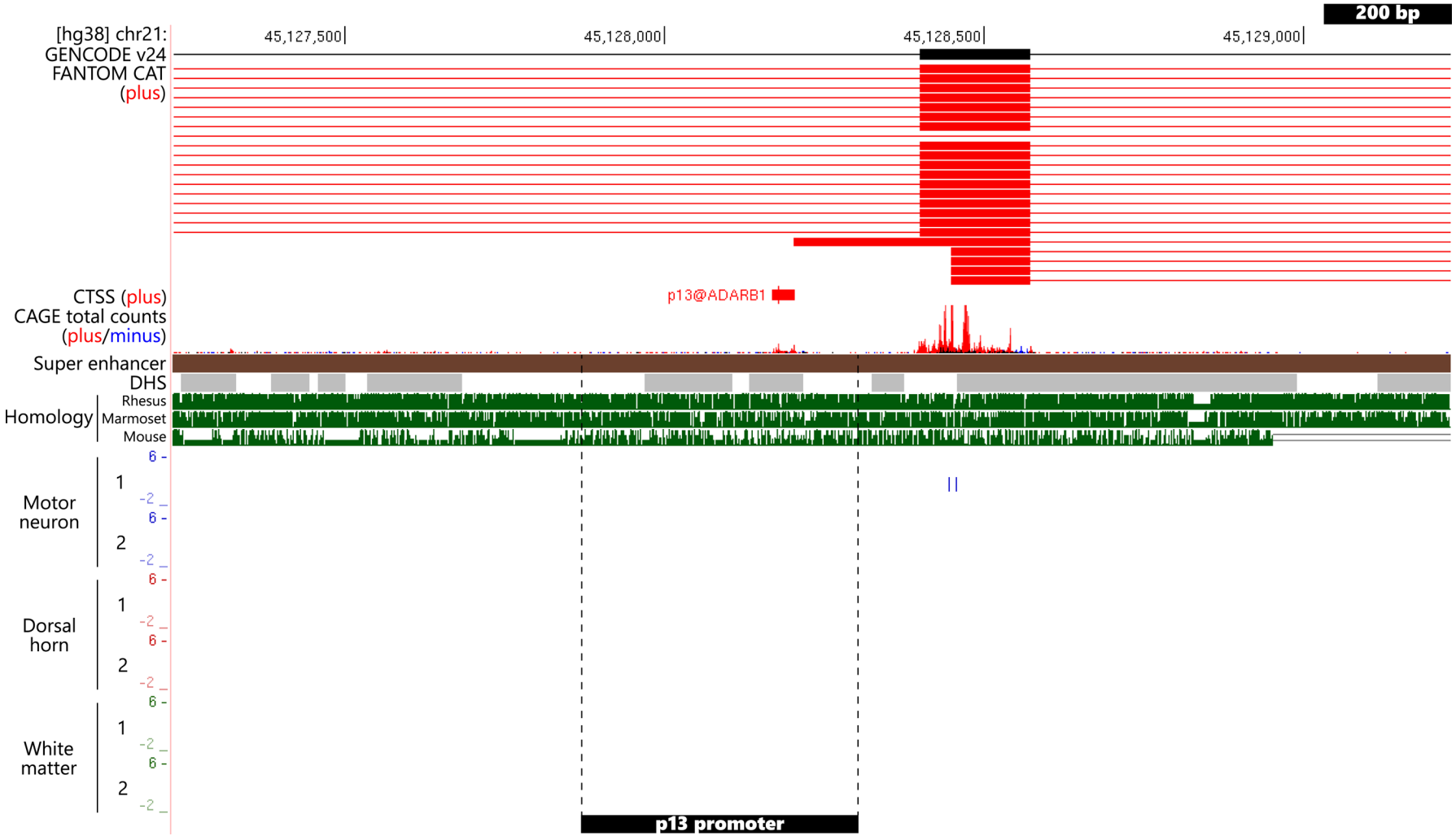
Results Figure 2A CAGE signals around *ADARB1*.

Bwa-mapped CAGE data from the human motor neurons (blue), dorsal horns (red) and white matter (green) shown at a cutoff of -2 tag per millions. Locus of *ADARB1* is indicated by the black arrow. All CTSSs (CAGE-defined TSSs) within this region are also shown (*e* and *lnc*: CTSSs of enhancer RNA and other lncRNA, respectively; also see **Introduction Figure 4**). The scale (100 kb) is shown at the top. Regions in **Results Figures 2B-D** are indicated by the rectangles, respectively.

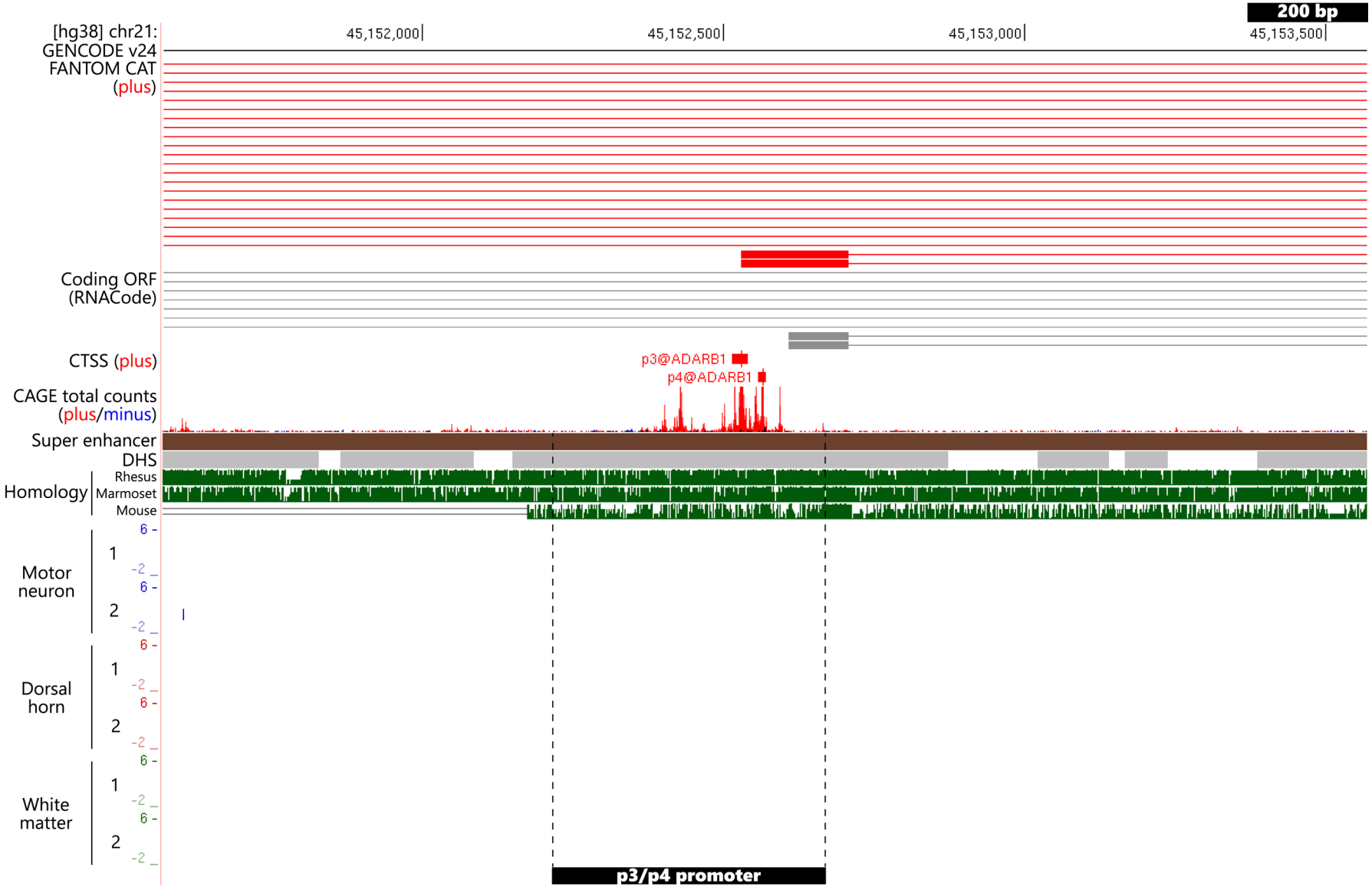


Results Figure 2B The CAGE signals and genetic environment around the p1 and the p2 promoters.

Positions of the p1 and the p2 promoters are indicated with the black bars at the bottom right and left, respectively. All data in **Results Figures 2B-D** are shown in the same way as **Results Figure 2A**. The scale is indicated at the top right.

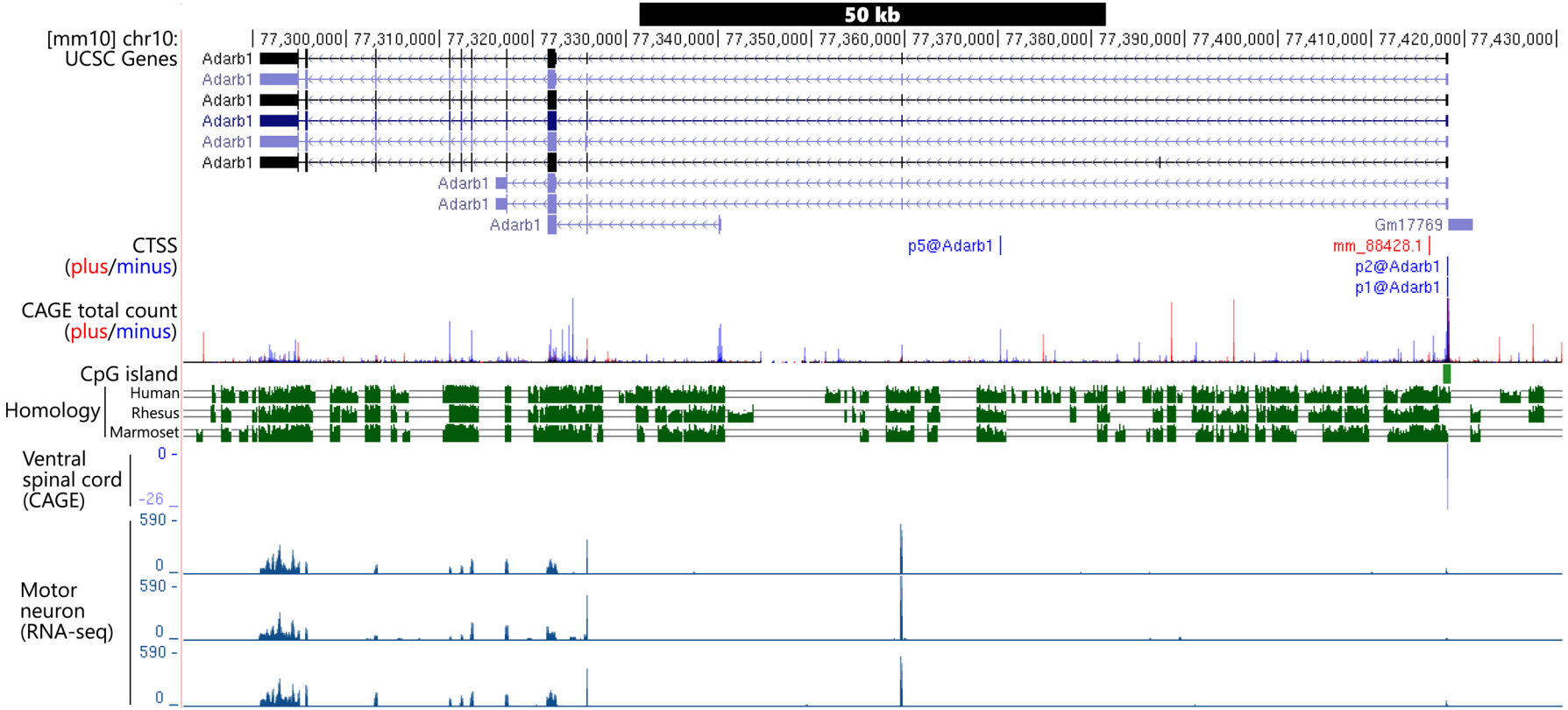


Results Figure 2C The CAGE signals and genetic environment around the p13 promoter. Position of the p13 promoter is indicated with the black bar at the bottom.

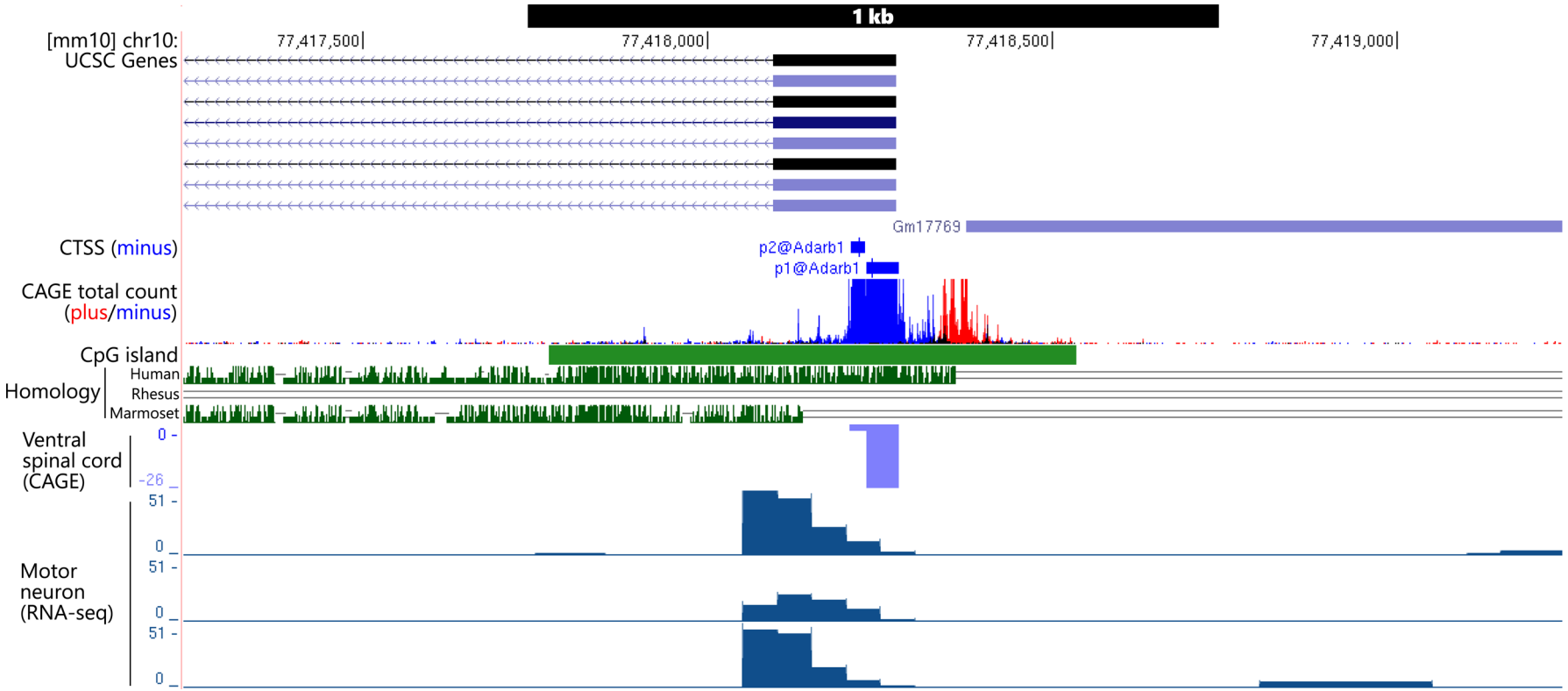


Results Figure 2D The CAGE signals and genetic environment around the p3/p4 promoter.

Position of the p3/p4 promoter is indicated with the black bar at the bottom right and left, respectively.

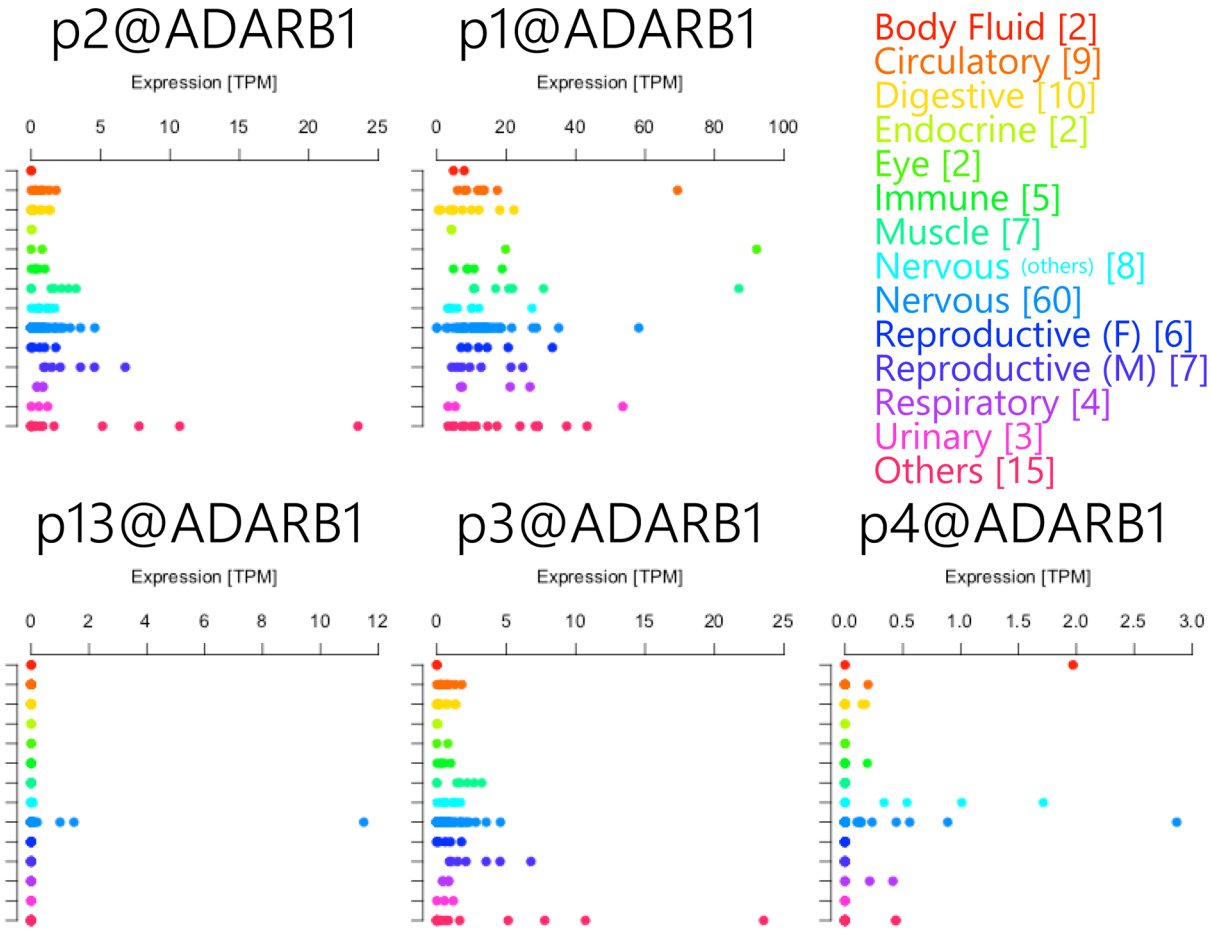


Results Figure 3A Structure and genetic environment around mouse ADAR2 gene *Adarb1* (overview).
Mice have only three CTSS (CAGE-defined TSSs) of *Adarb1* on the minus strand (blue). Data of RNA-seq from the mouse motor neurons is shown only for signals on the minus strand (the sense strand of *Adarb1*). CpG island (green bar) and DNA sequence homology (green signal) are also shown. CAGE data from the ventral spinal cord is retrieved from the FANTOM5 web site.

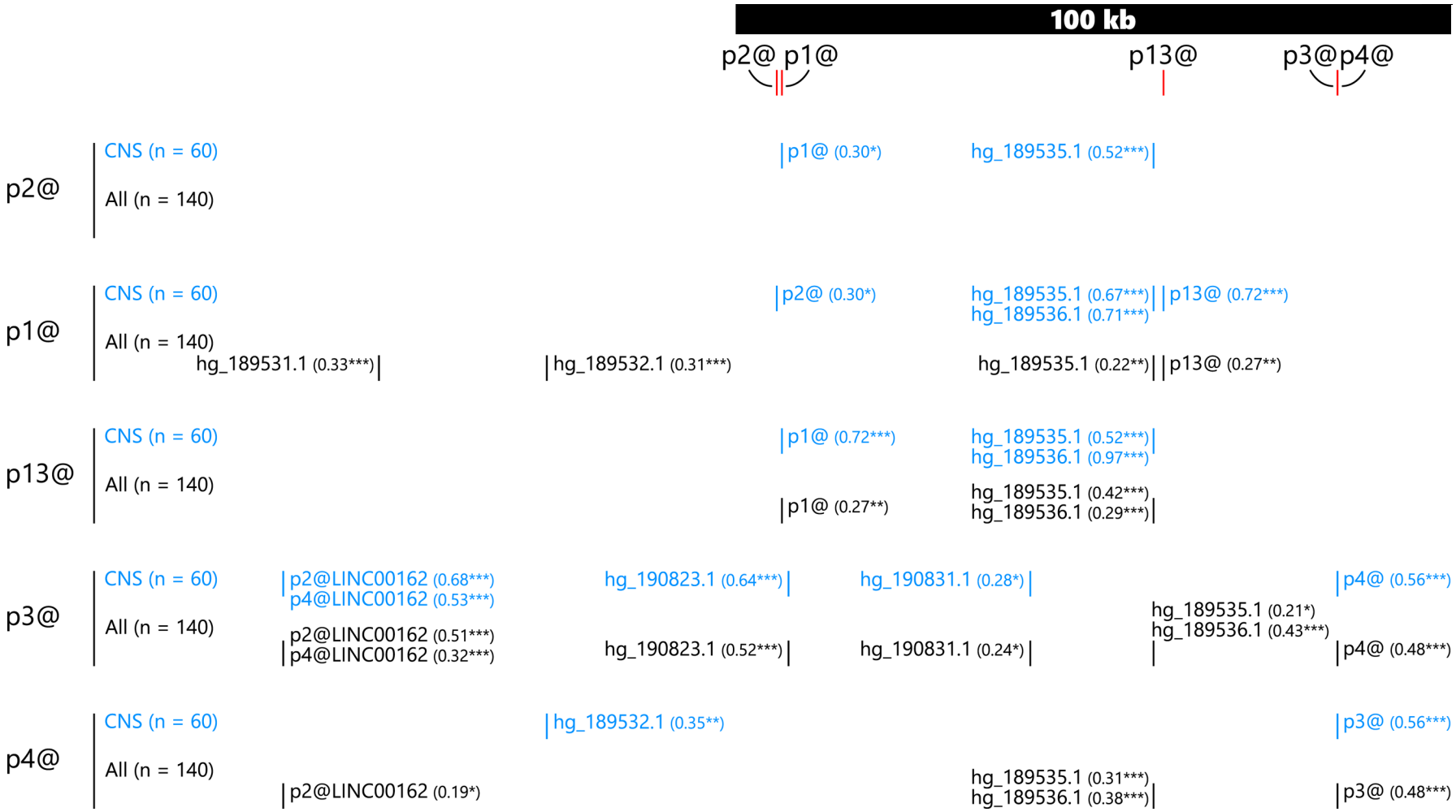


Results Figure 3B Structure and genetic environment around the first exon of *Adarb1*.

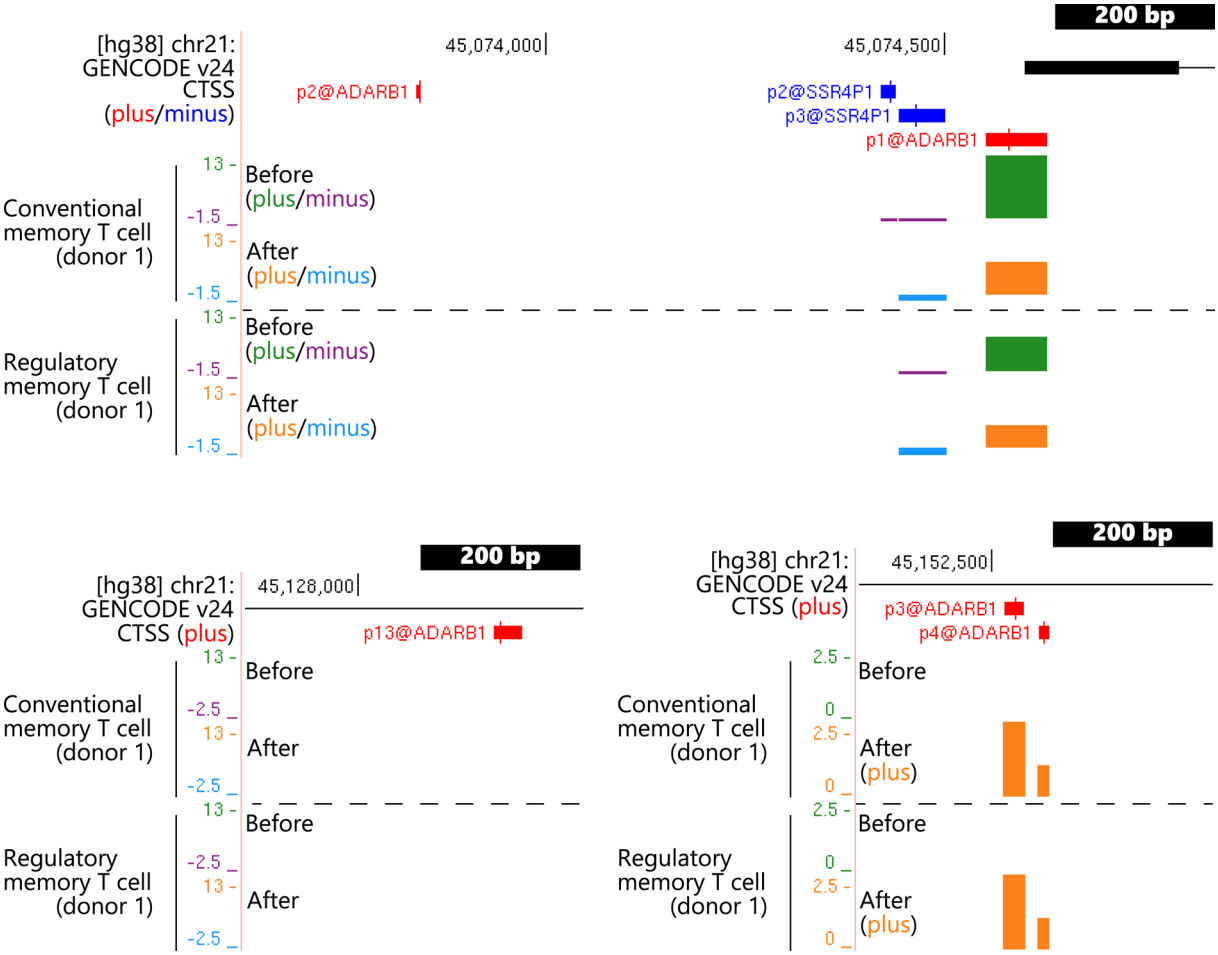
All data are shown in the same way as Results Figure 5A. Both of the p1@Adarb1 and the p2@Adarb1 are embraced with a CpG island (green bar). The DNA sequence around the p1@Adarb1 and p2@Adarb1 is poorly and partially conserved among the mammals (green signals). A strong CAGE signal of *Adarb1* in mouse ventral spinal cord is detected at the p1@Adarb1.



Results Figure 4A Tissue-specific expression of the 5 CTSSs of *ADARB1*.
Expression levels of every CTSSs in all the human tissue samples are shown in TPM (tag per million). All the human samples of healthy adults are categorized with different colors as to tissue-types. The number of samples in each category is indicated in brackets. All of the raw data are retrieved from the FANTOM5 web site; for names of these tissue samples, see **Materials and methods Table 3**.

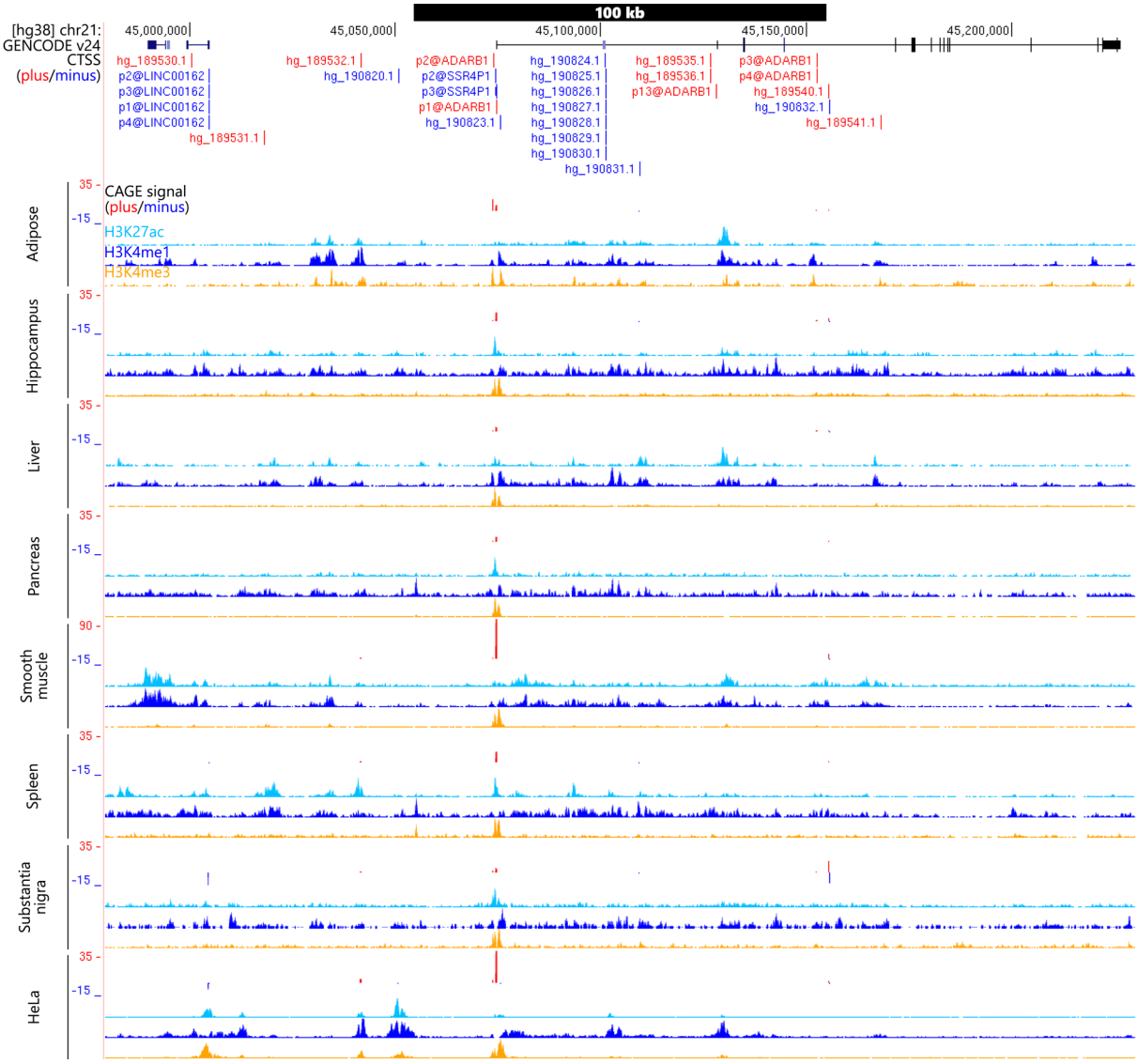


Results Figure 4B Correlation among CTSSs in CNS (blue) or all (black) of the tissues. Relative distances among the CTSSs are in scale. The scale (100 kb) is shown at the top right. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.



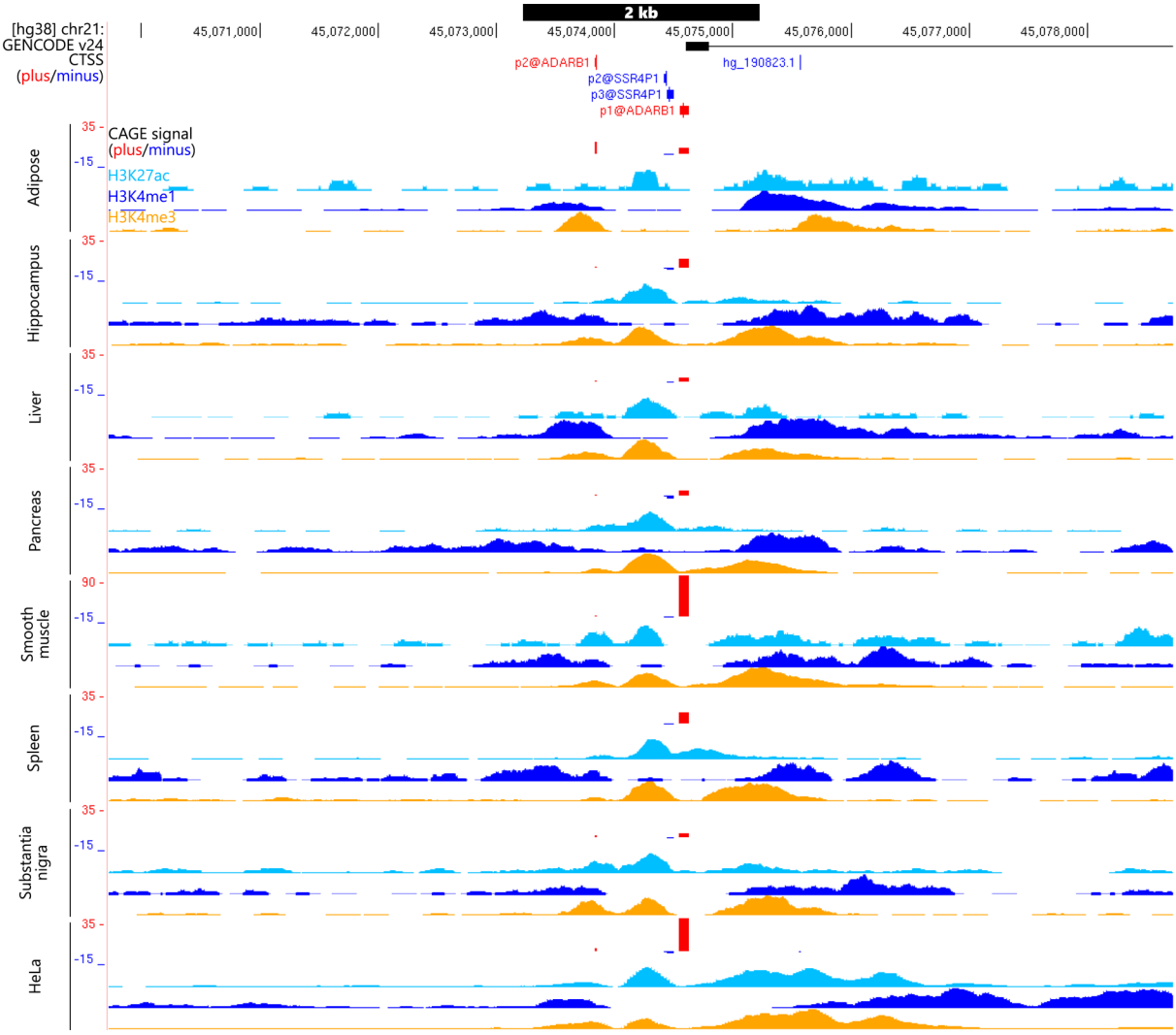
Results Figure 4C Example of independent and dynamic regulation of the p3@ and the p4@ in memory T cells.

Transcriptional activity of *ADARB1* at the p1@ and the p2@ (top), at the p3@ and the p4@ (bottom left) and at the p13@ (bottom right) before and after *in vitro* expansion of two kinds of memory T cells is shown, respectively. CAGE signals on the plus and minus strands before and after the expansion are shown in different colors, respectively, as indicated in each row. The raw data are downloaded from the FANTOM5 web site (see the Materials and methods Section). The scale (200 bp) is shown in the top right of each panel.



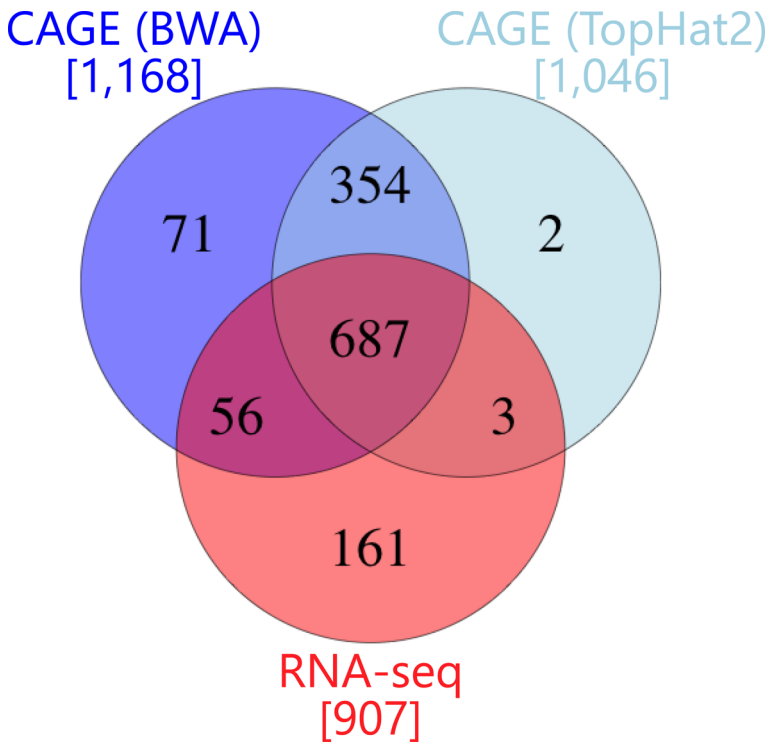
Results Figure 5A Transcriptional activity and histone modifications around *ADARB1* in 7 human tissues and HeLa cells.

CAGE signals (red, the plus strand; blue, the minus strand) and ChIP-seq signals for histone H3K27ac (light blue), H3K4me1 (dark blue) and H3K4me3 (orange) in human adipose (n = 4), hippocampus (n = 3), liver (n = 1), pancreas (n = 1), smooth muscle (n = 1), spleen (n = 1), substantia nigra (n = 3) and HeLa cells (n = 3) are visualized, respectively. For CAGE signals in adipose, hippocampus, substantia nigra and HeLa cells, the mean values of their replicates are shown, respectively. Note that the scale of CAGE signals in smooth muscle differs from the others, and that because of their auto scales, ChIP-seq signals are not comparable with each other. All of these are publicly available data (see the Materials and methods Section). The scale (100 kb) is indicated at the top.



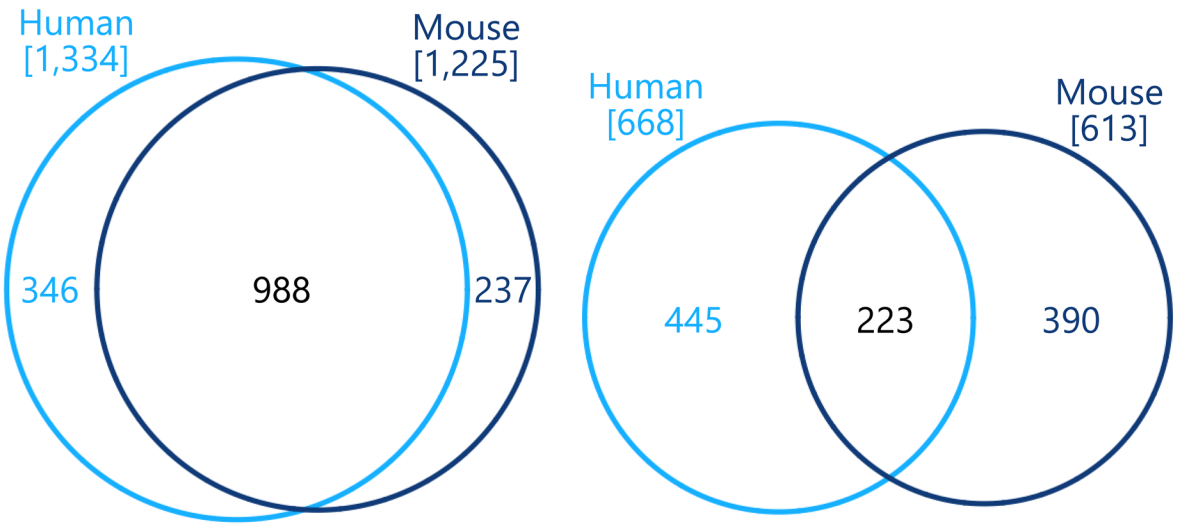
Results Figure 5B Transcriptional activity and histone modifications around *ADARB1* in 7 human tissues and HeLa cells (view around the p1@ and the p2@).

All scales, annotations and colors are the same as those in **Results Figure 5A**. All of these are publicly available data (see the Materials and methods Section). The scale (2 kb) is indicated at the top.

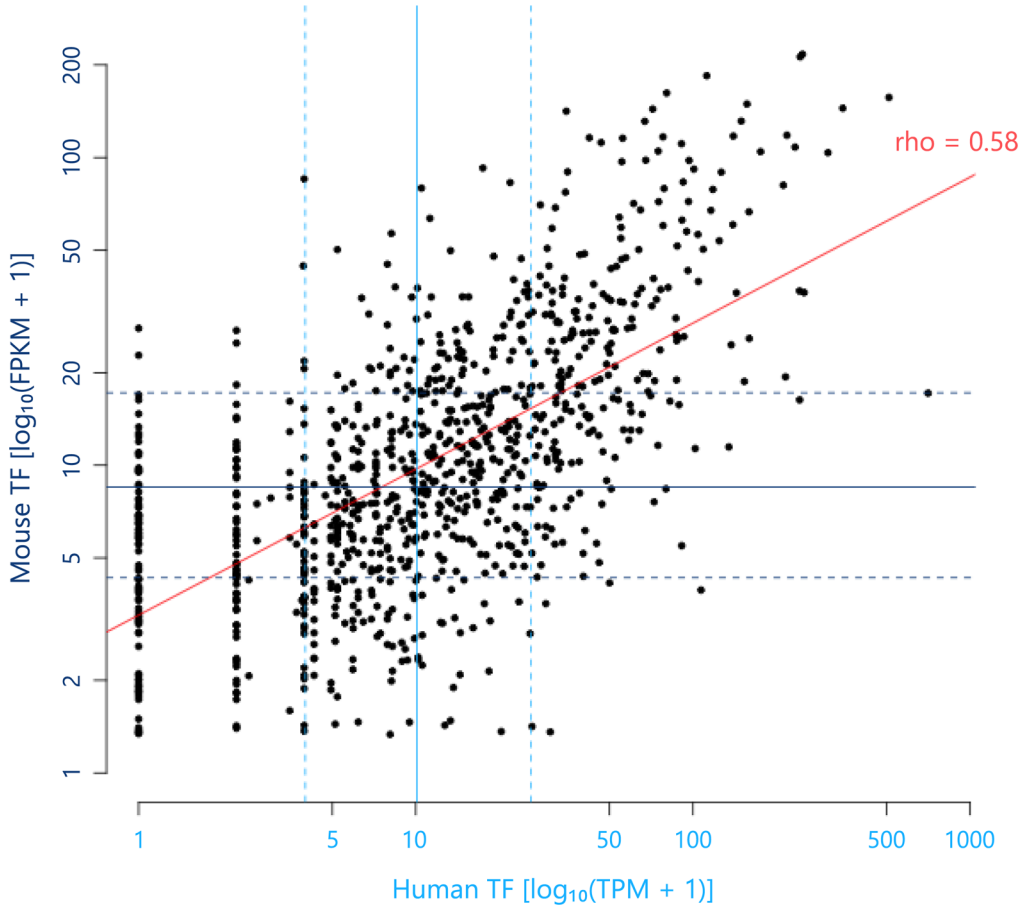


Results Figure 6 The numbers of TFs detected in the motor neurons (MNs) by the present CAGE and the previous RNA-seq.

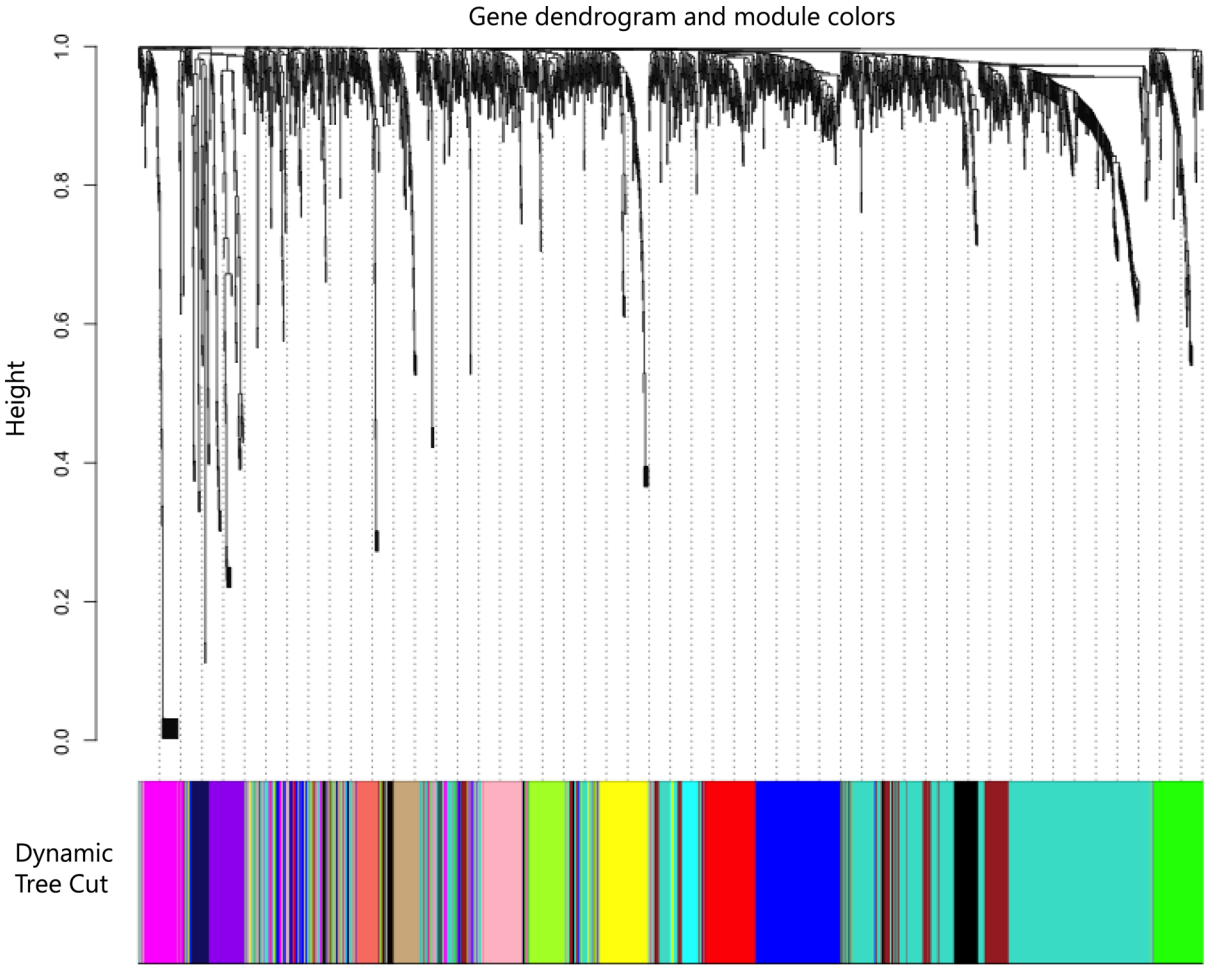
The number of TFs found in the present data of CAGE from the MNs aligned by BWA (blue) or TopHat2 (light blue) and in the publicly available previous RNA-seq data is indicated in brackets, respectively. In total, 1,334 TFs are defined as motoneuronal TFs without duplication.



Results Figure 7A The number of motoneuronal TFs expressed both in humans and mice. As a whole, 988 motoneuronal TFs are overlapped between humans and mice (left). Approximately only one-third of the motoneuronal TFs are commonly expressed between the two species at a level greater than the median expression level of total motoneuronal TFs (right). The total number of human and mouse TFs is shown in brackets, respectively.

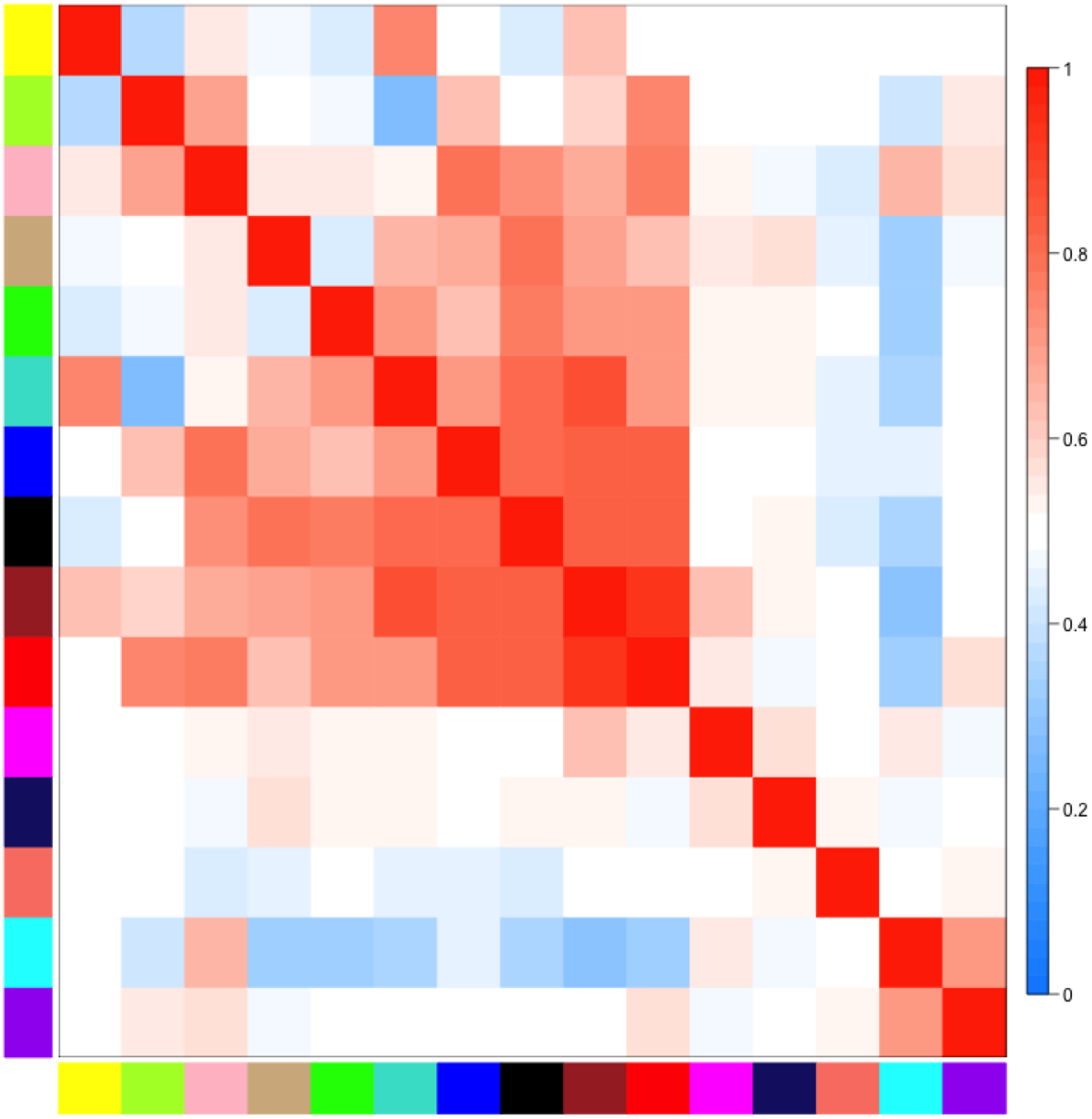


Results Figure 7B Correlation in expression between human and mouse motoneuronal TFs. Expression levels of 988 motoneuronal TFs are just modestly correlated between human (light blue) and mouse (dark blue). A red line shows a result of linear regression, and Pearson's rho is indicated also in red. Lines in light and dark blue display the median of the expression levels of the human and mouse TFs, respectively. Dotted lines in light and dark blue are the lower or upper quartile of the expression levels of the human and mouse TFs, respectively.



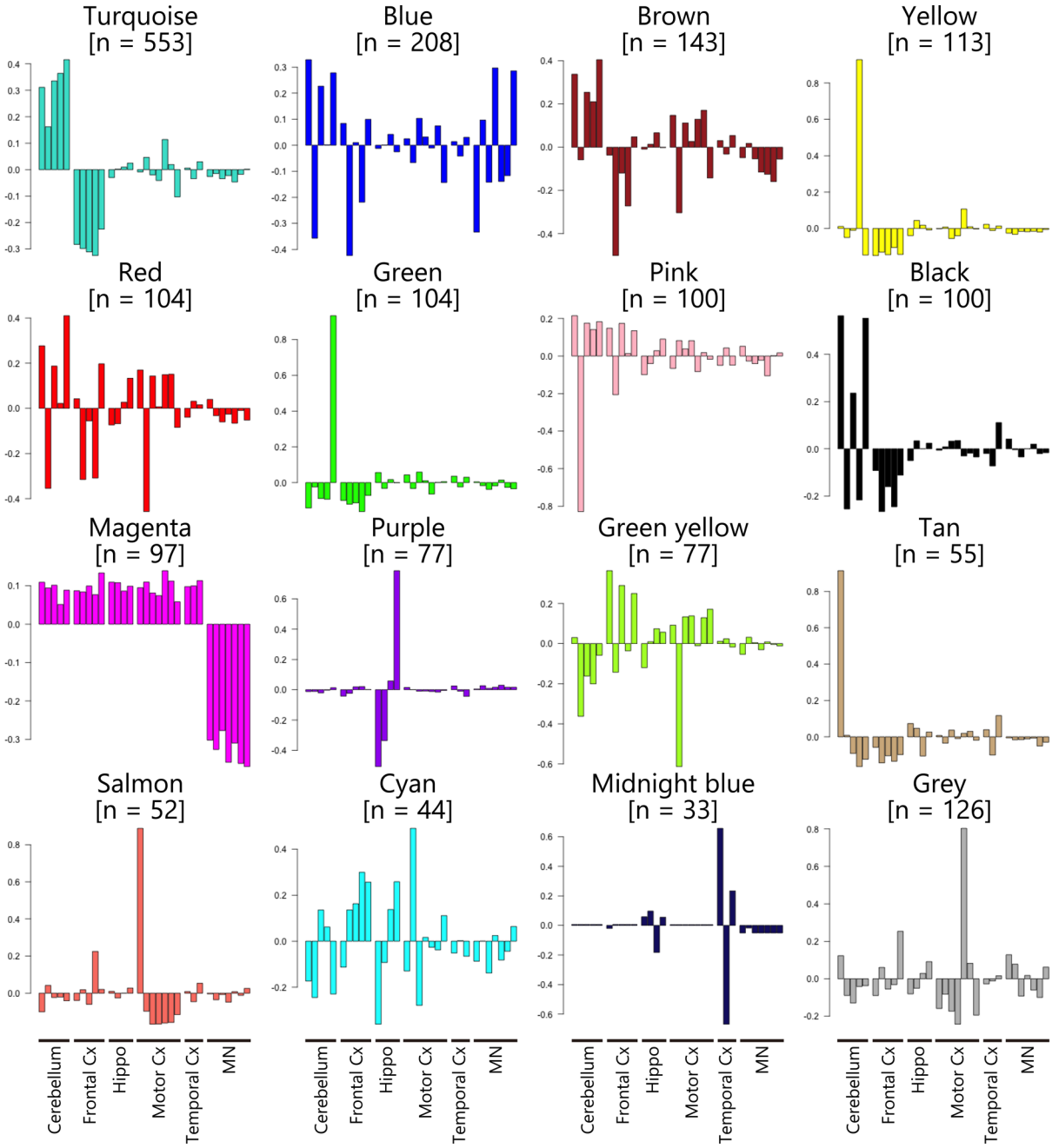
Results Figure 8A Dynamic tree cut in the WGCNA.

WGCNA (Weighted gene co-expression network analysis) with the dynamic tree cut divides TFs expressed in the human central nervous tissues into the 15 colored modules and one unclassified group (Module Grey).



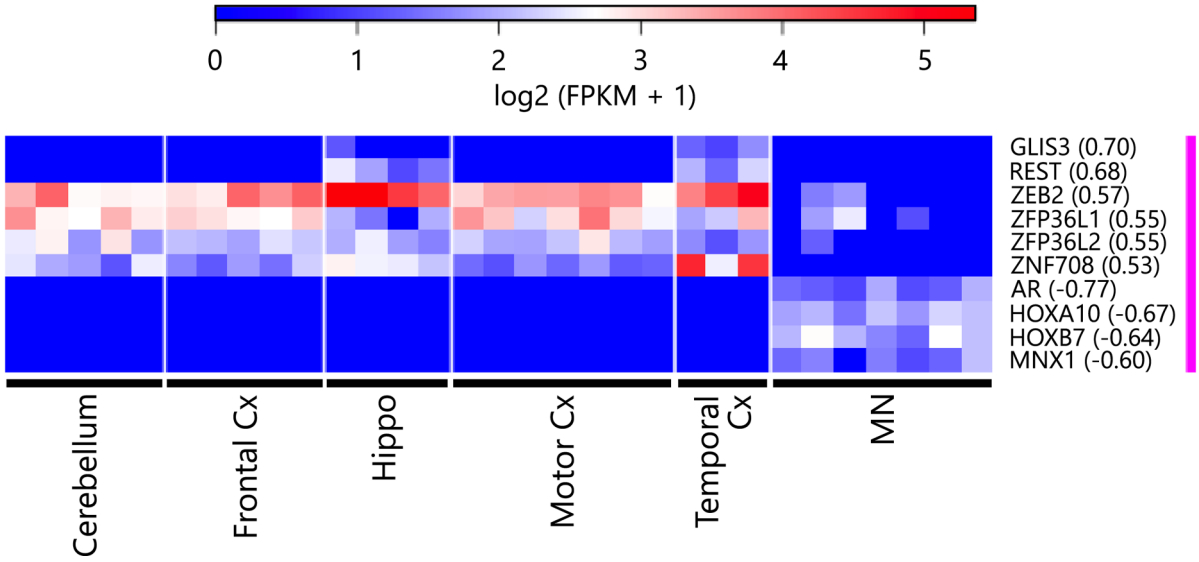
Results Figure 8B Adjacency of the eigengenes of the 15 modules.

The module eigengene (first principal component) shows the large cluster consisting of 10 modules (upper left). Module colors are depicted at the left and bottom. The modules within the large cluster may be functionally related with each other. The color key at the right side indicates the adjacency (1 + cor(Ei, Ej))/2, where Ei and Ej are the module eigengene of Module i and j, respectively, and cor(Ei, Ej) is Pearson' rho between Ei and Ej.



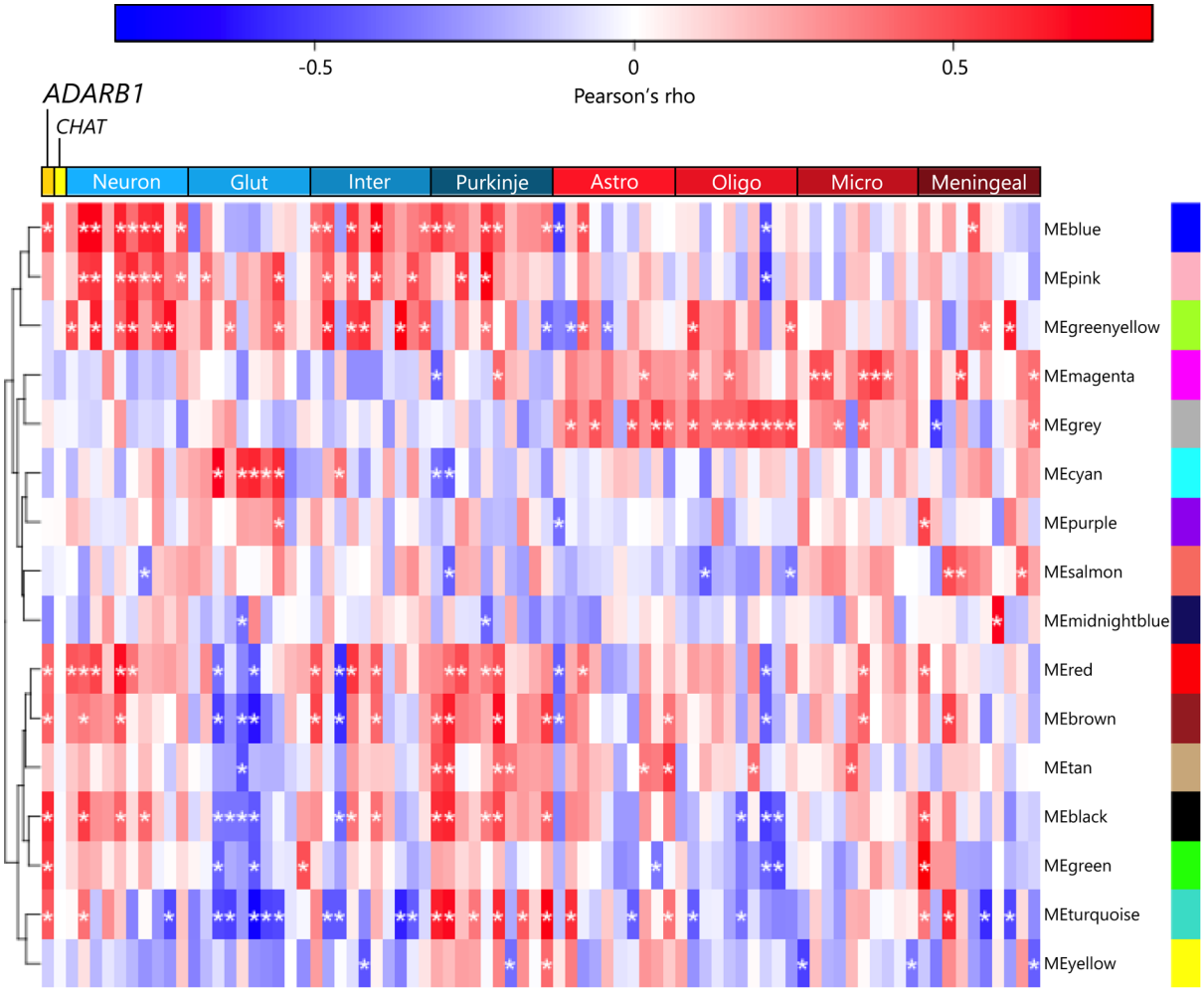
Results Figure 8C Module eigengene of all the modules.

All module eigengenes in cerebellum (n = 5), frontal cortex (Frontal Cx; n = 5), hippocampus (Hippo; n = 4), motor cortex (Motor Cx; n = 7), temporal cortex (Temporal Cx; n = 3) and motor neurons (MNs; n = 7) are shown. Eigengene of Module Magenta clearly distinguish the MNs from the other nervous tissues. Because negative module eigengene is prominent in the MNs, TFs whose expression negatively contributes to the eigengene of the Module Magenta may be specifically expressed in the MNs. The number of TFs classified in each module is indicated in brackets.



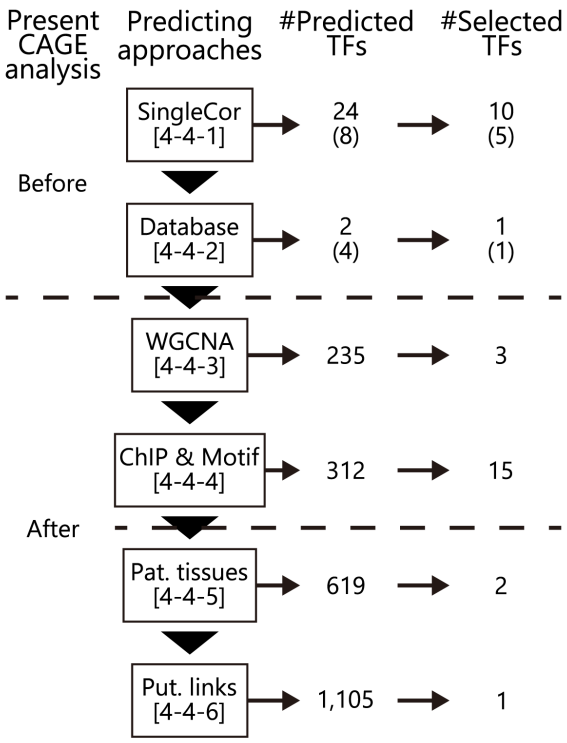
Results Figure 8D Expression levels of TFs significantly contributing to the eigengenes of the Module Magenta.

The WGCNA reveals that six TFs have significant positive contribution to the eigengene of the Module Magenta, whereas four TFs have significant negative contribution to it, as indicated in parentheses. As expected, negatively contributing four TFs are specifically expressed in the motor neurons (MN). The color key at the top directs expression levels of TFs. Cx, cortex; Hippo, hippocampus.

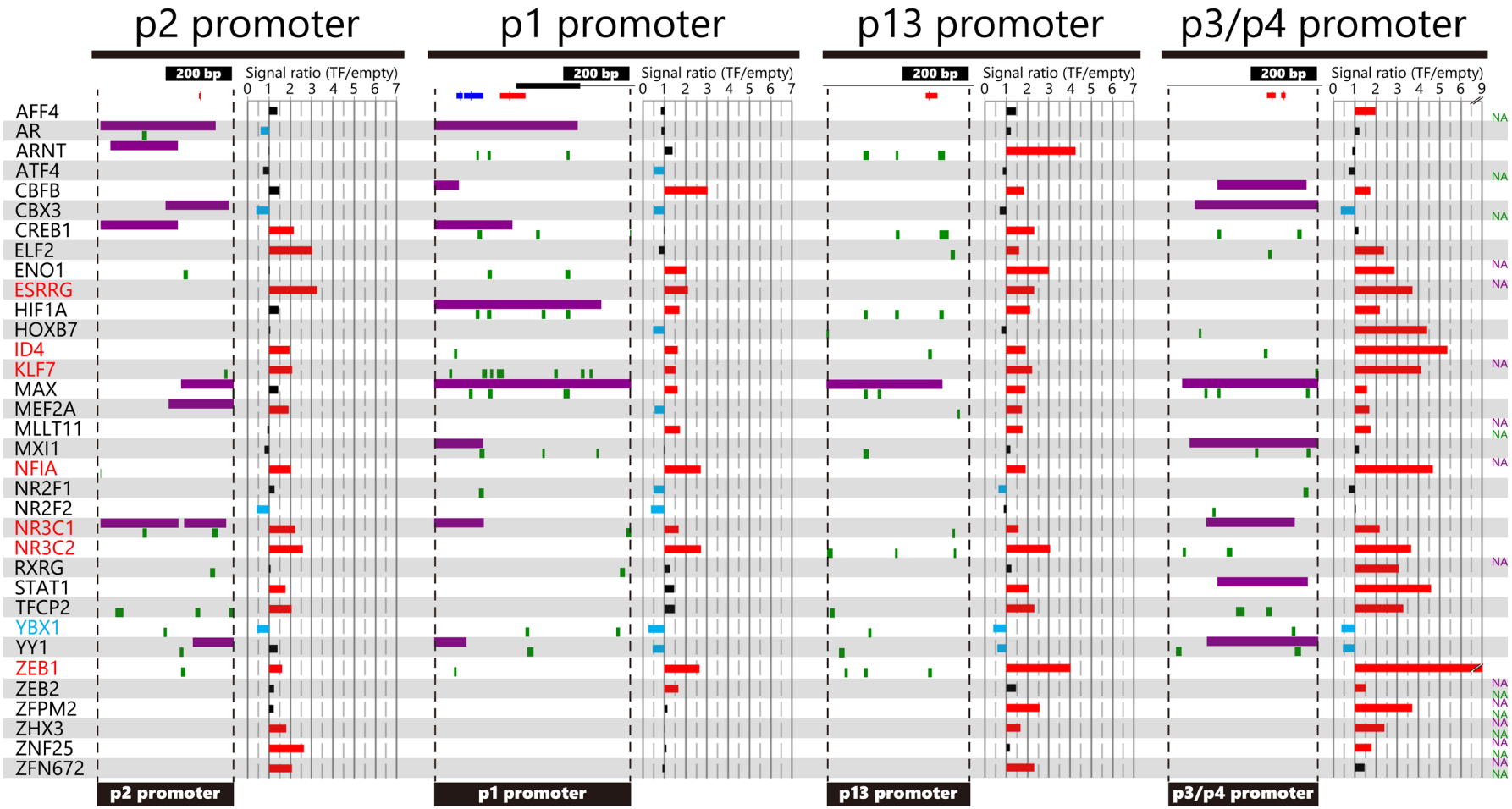


Results Figure 8E Correlation between every module eigengene and expression of *ADARB1* and other marker genes.

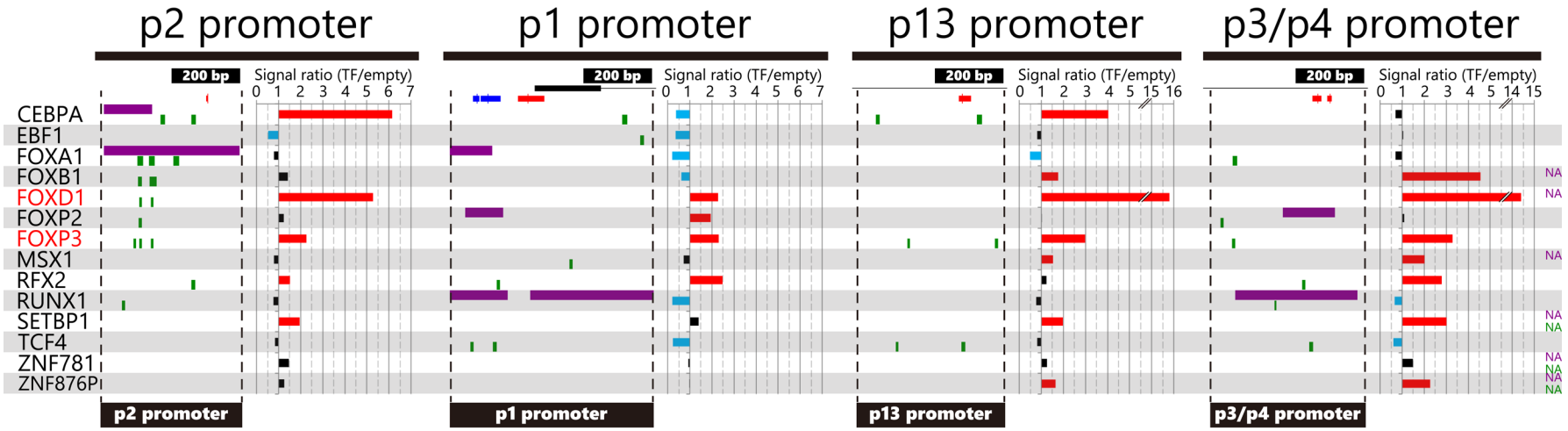
Eigengenes of some modules are significantly and uniquely correlated with expression of the selected genes, as indicated with the asterisk in cells of the heatmap. Cell-types of the marker genes are indicated at the upside of the heatmap. Labels for names of the marker genes for every cell-types are omitted but in alphabetical order [also see **Materials and methods Table 6**]. Modules are hierarchically clustered. Pearson's rho is indicated with the color key at the top. Neuron, ubiquitous neuronal marker genes; Glut, Inter, Purkinje, Astro, Oligo, Micro, Meningeal are marker genes for glutamatergic neuron, interneuron, Purkinje cell, astrocyte, oligodendrocyte, microglia and meningeal cell.



Results Figure 9 Flow chart for selecting candidate TFs for *in vitro* luciferase assay. *ADARB1*-regulatory TFs were predicted *in silico* while the present CAGE analysis on the human motor neurons (MNs) was progressed. The candidate TFs were randomly selected from "SingleCor" to "ChIP & Motif" approaches, whereas two and one TFs were intendedly selected in "Pat. tissues" and "Put. links" ones. In the "Pat. tissues" approach, two out of 619 motoneuronal TFs were additionally selected, because they were highly expressed in the present CAGE data on the laser-captured MNs and decreased in multiple tissue types of patients with sporadic ALS. In the "Put. links" approach, I selected only one TF whose molecular link with FUS-ALS pathology had been implied in previous reports[201,220]. The results of the present CAGE analysis on the laser-captured MNs revealed that some of the selected TFs in "SingleCor" and "Database" approaches were not expressed in the MNs, as indicated in the parentheses in their rows. The number of the Results section of each predicting approach is shown in brackets. By this selecting procedure, 32 motoneuronal and six non-motoneuronal TFs were selected; in addition to these, two motoneuronal and eight non-motoneuronal TFs had been already selected and prepared before the present study. SingleCor, correlation between a single TF and *ADARB1* expression; Database, the database ORTI for target genes of TFs; WGCNA, correlation between a TF in the module and *ADARB1* expression; ChIP & Motif, presence of ChIP-seq signals and consensus motifs of TFs within the promoters; Pat. tissues, significantly decreased TFs in tissues of patients with sporadic ALS; Put. links, putative molecular links between *ADARB1* regulation and FUS-linked ALS pathogenesis.



Results Figure 10A Luciferase assay of the selected motoneuronal TFs and positions of their ChIP-seq signals and consensus motifs. Fold-change in luciferase signals to the negative control is displayed as a graph at each right side of the ChIP-seq (purple) and consensus motifs (green) data shown in relative to each of the *ADARB1* promoters (black bars at the bottom). Up- (> 3/2) and down-regulated (< 2/3) signals are shown in red and blue bars in the graphs, respectively. Names of TFs that up- and down-regulated all the three promoters (i.e., universal activators and suppressors) are shown in red and blue, respectively. The margins of the promoters are indicated by vertical dotted black lines. The scale and position of the CTSSs of *ADARB1* (red boxes) are shown at the upside. NA, no data available.



Results Figure 10B Luciferase assay of the selected non-motoneuronal TFs and positions of their ChIP-seq signals and consensus motifs. Fold-change in luciferase signals to the negative control is displayed as a graph at each right side of the ChIP-seq (purple) and consensus motifs (green) data shown in relative to each of the *ADARB1* promoters (black bars at the bottom). Up- (> 3/2) and down-regulated (< 2/3) signals are shown in red and blue bars in the graphs, respectively. Names of TFs that up-regulated all the three promoters (i.e., universal activators) are shown in red. The margins of the promoters are indicated by vertical dotted black lines. The scale and position of the CTSSs of *ADARB1* (red boxes) are shown at the upside. NA, no data available.

Results Table 1A. Sample quality control and mapped read counts of the CAGE libraries.

Sample	Replicate	Sample quality control						# Mapped read	
		260/280 [1.8]	260/230 [1.8]	Concentration (µg/µL) [1]	Volume (µL)	Amount (µg) [5]	RIN [7]	bwa	tophat
Motor neuron	B 1	2.04	2.03	0.042	20	0.84	4.2	573,606	577,229
	B 2	2.18	0.09	0.029	19	0.56	4.6	685,187	690,260
Dorsal horn	T 1	2.1	2.7	0.023	24	0.55	7.7	736,433	746,413
	T 2	2.09	0.52	0.125	22	2.75	6.1	3,238,005	3,310,985
White matter	T 1	2.02	0.11	0.02	23	0.45	6.2	212,370	212,681
	T 2	2.1	0.1	0.041	21	0.86	5.9	743,609	750,320
Internal control	T 1							5,220,558	5,339,403
	T 2							5,879,554	6,006,198

B, biological; T, technical. The required value for every quality parameter is shown in brackets. Total RNA of *Mus musculus* C57BL/6J whole body, embryonic day 17.5 was used as an internal control.

Results Table 1B. Basic sequencing statistics.

Yield (Mb)	# Reads	% of ≥ Q30 bases (PF)	Mean quality score (PF)
2,644	51,837,148	95.24	37.23

Mb, megabases. % of ≥ Q30 bases (PF) = (yield of bases with Q30 or higher from clusters passing filter) / (total yield of clusters passing filter) x 100. Mean quality score (PF) = (the total sum of quality scores of clusters passing filter) / (total yield of clusters passing filter).

Results Table 2. Defined promoter regions of *ADARB1*.

Promoter	CTSS position in chr21	Promoter position in chr21	Length (bp)	Window around the CTSS (bp)	GC ratio (%)	CpG occupancy (%)
p2	45,073,837-45,073,842	45,073,531-45,073,942	412	-308, +100	46.6	3.9
p1	45,074,551-45,074,628	45,074,353-45,074,948	596	-200, +320	82.9	36.9
p13	45,128,169-45,128,205	45,127,871-45,128,305	435	-300, +100	49.0	6.4
p3/p4	45,152,515-45,152,540 45,152,558-45,152,571	45,152,217-45,152,671	455	-300, +100	56.0	9.7

Defined four *ADARB1* promoters were listed according to the chromosomal position (hg38).

Results Table 3. Sample quality control and mapped read counts of the RNA-seq of the motor neurons of wild-type mice.

Replicate	Sample quality		# Mapped read by STAR	
	Concentration (ng/μL)	RIN	Unique	Multiple
1	19.66	4.7	22,757,963	718,245
2	3.74	4.4	10,344,419	370,368
3	5.55	4.2	22,843,932	774,587

Results Table 4. TFs highly correlated or anti-correlated with the expression of *ADARB1* in the 20 tissue samples.

TF	Spearman's rho	TF	Spearman's rho
ADNP2	0.749	RLF	0.734
AFF4	**0.729**	RUNX1T1	0.852
ARNTL2	0.701	**SETBP1**	**0.788**
CRAMP1L	0.717	SFPQ	0.704
EBF1	**0.874**	**ZEB1**	**0.723**
ELF2	**0.735**	ZFP90	0.750
ELK3	0.725	**ZFPM2**	**0.833**
FOXJ3	0.719	**ZHX3**	**0.829**
FOXK1	0.774	**ZNF25**	**0.717**
HSF4	0.721	ZNF83	0.744
KLF7	**0.841**	ZNF385B	0.724
MEF2A	**0.809**	ZNF385D	0.719
MEF2D	0.777	ZNF660	0.724
NFIA	**0.704**	**ZNF672**	**-0.750**
POU2F1	0.744	**ZNF781**	**0.718**
RFX2	**0.770**	**ZNF876P**	**0.700**

TFs that were examined by *in vitro* luciferase assay are shown in bold.

Results Table 5. #ChIP-seq signals and consensus motifs of TFs within the *ADARB1* promoters.

Promoter	Data	#All TFs ChIP [571] Motif [821]	#MnTFs ChIP [340] Motif [328]	#MnTFs in total [498/1,334]
p1	ChIP	99	60	189
	Motif	347	149	
p2	ChIP	89	55	158
	Motif	319	123	
p3/p4	ChIP	91	57	177
	Motif	336	142	
p13	ChIP	2	1	113
	Motif	271	113	

The numbers of TFs whose ChIP-seq data and consensus motifs were available are shown in brackets in columns, respectively. The total numbers of motoneuronal TFs that possibly bind to the promoters are also shown. ChIP, ChIP-seq signals; MnTFs, motoneuronal TFs; Motif, consensus motifs.

Results Table 6. Differentially expressed TFs in previous studies.

Study	Region	#Case	#Cntl	Method		Change	#All TFs [2,410]	#MnTFs [1,334]	Src
1	Spinal motor neurons	13	9	Seq		Up	150	150	1-1
						Down	52	52	
2	Spinal motor neurons	12	10	Array		Up	87	87	2-1
						Down	109	109	
	Remained anterior horns	12	10	Array		Up	62	62	2-2
						Down	25	25	
3	Whole spinal cord	6	5	Seq		Up	47	6	3-1
						Down	23	18	
4	Motor cortex	31	10	Array	A1vs C	Up	224	163	4-1
						Down	93	26	
					A2vs C	Up	258	119	
						Down	560	391	
					A1vs A2	Up	685	502	
						Down	334	116	
5	Motor cortex	11	9	Array		Up	1	0	5-1
						Down	1	1	

Cntl, control; Array, microarray; Seq, RNA-seq; Up, up-regulated; Down, down-regulated; MnTF, motoneuronal TF; A1, the SALS1 group (18 cases, 1 control); A2, the SALS2 group (13 cases, 2 controls); C, the Control group (7 controls). The total numbers of all TFs and motoneuronal TFs are shown in brackets in each column, respectively.
Study: 1, Kapeli K. *et al*. 2016 *Nat. Commun.*; 2, Rabin S. J. *et al*. 2009 *Hum. Mol. Genet.*; 3, D'Erchia A. M. *et al*. 2017 *Sci. Rep.*; 4, Aronica E. *et al*. 2015 *Neurobiol. Dis.*; 5, Lederer C. W. *et al*. 2007 *BMC Genomics*.
Src: 1-1, Kapeli K. *et al*. 2016 *Nat. Commun*. Supplementary Data 8; 2-1, Rabin S. J. *et al*. 2009 *Hum. Mol. Genet*. Supplementary Table S2; 2-2, Rabin S. J. *et al*. 2009 *Hum. Mol. Genet*. Supplementary Table S3; 3-1, D'Erchia A. M. *et al*. 2017 *Sci. Rep*. Supplementary Table S2; 4-1, Aronica E. *et al*. 2015 *Neurobiol. Dis*. Supplementary Tables 2-4; 5-1, Lederer C. W. *et al*. 2007 *BMC Genomics* Figure 3.

Results Table 7. #TFs that are listed up from every putative molecular link between *ADARB1* dysregulation and FUS mutations.

Putative link	TF	Interactor	Sample	#All TFs [2,410]	#MnTFs [1,334]		Src
I	mRNA	FUS	Temporal cortex	82	66	834	1
			HeLa	184	139		2
			HEK293	1121	810		3
		mFUS	HEK293	750	579		
II	Protein	FUS	HeLa	50	48		4
III	mRNA	miR-141	A498	Up 354	Up 176		5
				Down 339	Down 168		
			Caki-1	Up 293	Up 128		6
				Down 229	Down 100		
			in silico prediction	435	255		7
		miR-200a	A498	Up 399	Up 207		8
				Down 340	Down 191		
			Caki-1	Up 226	Up 104		9
				Down 270	Down 99		
			in silico prediction	366	220		7

MnTF, motoneuronal TF; mFUS, mutant FUS (R521G/H). The total numbers of all TFs and motoneuronal TFs are shown in brackets in each column, respectively. In Row I, the total number of motoneuronal TFs bound by wild-type FUS is also indicated. Src: 1, Nakaya T. *et al*. 2013 *RNA* Supplementary Table S3; 2, Zhou Y. *et al*. 2013 *PLoS Genet*. Supplementary Table S1; 3, Hoell J. I. *et al*. 2011 *Nat. Struct. Mol. Biol*. Supplementary Dataset 2; 4, Sun S. *et al*. 2015 *Nat. Commun*. Supplementary Table S1; 5, GSM911074; 6, Yoshino H. *et al*. 2013 *J. Hum. Genet*. (GSM911081); 7, TargetScan Human 7.1; 8, GSM911072; 9, Yoshino H. *et al*. 2013 *J. Hum. Genet*. (GSM911078).

Results Table 8A. Motoneuronal TFs examined by *in vitro* luciferase assay.

Section:	4-3-1		4-3-2	4-4-2	4-4-1	4-4-3		5-1-3		5-2	5-2	4-4-5										4-4-6							
TF	Expression in MNs			In the Data base	Cor with *ADARB1*	WGCNA		Tissue Specificity		Change with aging	eQTL SNP in sALS	Differential expression in the patients' tissues										Interacted by			MiR-141			MiR-200a		
	Human		Mouse			Cor with *ADARB1*	Mod col	Score	Hub			MNs		AHs	SP	Motor cortex						FUS	mFUS		OE			OE		
																S4-1									Fold-change			Fold-change		
	Seq	CAGE	Seq									S1-1	S2-1	S2-2	S3-1	A1-C	A2-C	A1-A2	S5-1	Protein of TF	mRNA of TF	mRNA of TF	A498	Caki-1	Predict	A498	Caki-1	Predict
AFF4	✔	1/4	1/4		0.729	0.377	Blu	0.27			✔										✔	✔						✔
AR	✔	4/4	2/4	✔			Mag	1.15	Spe				D					D								0.25		✔
ARNT	✔	2/4	4/4			0.391	Red	0.21	Ubi		✔			U		U					✔	✔			✔			
ATF4	✔	1/4	1/4				Tur	0.22	Ubi		✔			D		U	D	U			✔	✔	1.62			1.60		
CBFB	✔	4/4	3/4				Pin	0.22				U	U				D	U			✔	✔		1.53			1.62	
CBX3	✔	1/4	1/4				Pin	0.21		D							D	U			✔	✔						
CREB1	✔	2/4	4/4				Tur	0.27	Ubi								U	D			✔	✔						
ELF2	✔	4/4	3/4		0.735		Blu	0.31						U							✔	✔				1.35		
ENO1	✔	1/4	1/4				Pin				✔					U		U		✔	✔	✔						
ESRRG	✔	1/4	1/4				Tan	0.97							D	D	D	D							✔	0.57		
HIF1A	✔	1/4	1/4				Red	0.39	Ubi					U			D	U			✔	✔	1.44			1.45		
HOXB7	✔	2/4	3/4				Mag	1.31		U							U	D										
ID4		2/4	3/4		0.474		Gre	1.36				U				U		U			✔	✔	1.46			0.72		✔
KLF7	✔	1/4	3/4		0.841		Blk				✔		D													1.85		
MAX	✔	1/4	2/4			0.425	Blk	0.11	Ubi		✔					U	D	U			✔				✔			
MEF2A	✔	1/4	1/4		0.809		Tur	0.43	Ubi												✔	✔						
MLLT11	✔	1/4	1/4		0.503		Yel			D	✔	D			D		D	U	D		✔	✔		0.69			0.68	
MXI1	✔	3/4	3/4				Bwn	0.30			✔	U	U				D	U			✔	✔	1.34		✔			
NFIA	✔	1/4	2/4		0.704		Tur	0.45									D	U			✔	✔			✔		0.75	
NR2F1	✔	2/4	4/4		0.520		Tur	0.63	Ubi			U						U			✔	✔						
NR2F2	✔	4/4	3/4				Mag	0.64				U	U					U			✔	✔						
NR3C1	✔	3/4	2/4				Blk	0.35	Ubi				U				D	U			✔	✔						
NR3C2	✔	2/4	1/4				Pin	0.60													✔		1.37	0.56			0.54	
RXRG	✔	1/4	3/4				Bwn	1.54	Spe			D	D	D			U	D					1.79					
STAT1	✔	1/4	2/4				Blu	0.33	Ubi								D	U			✔	✔	1.92			1.50		
TFCP2	✔	2/4	3/4		0.400		Tan	0.09			✔					U		U			✔	✔			✔			
YBX1	✔	3/4	2/4			0.445	Bwn	0.51	Ubi			U	D							✔	✔	✔						
YY1	✔	2/4	2/4				Grey	0.09	Ubi							U		U			✔	✔	0.74		✔			✔
ZEB1	✔	1/4	1/4		0.723		Tur							U							✔	✔	0.71	0.64	✔	0.65	0.64	✔
ZEB2	✔	2/4	4/4				Mag					U	U								✔	✔	0.69	0.74		0.53	0.54	✔
ZFPM2	✔	3/4	3/4		0.833	0.358	Blu	1.08				D					Am	Am			✔			1.48				
ZHX3		2/4	2/4		0.829		Yel	0.41					U			U		U			✔							
ZNF25	✔	1/4			0.717		Gre				✔						D	U			✔					1.35		✔
ZNF672		3/4			-0.750		GrY					U																

Seq, RNA-sequencing; CAGE, Cap analysis of gene expression; Cor, correlation; WGCNA, weighted gene coexpression analysis; Mod col, module color; Spe, tissue-specific hub; Ubi, ubiquitous hub; eQTL, expression quantitative trait locus; SNP, single nucleotide polymorphysim; sALS, sporadic ALS; MN, motor neuron; AH, anterior horn; SP, spinal cord; A1-C, ALS group 1 vs Control; A2-C, ALS group 2 vs Control; A1-A2, ALS group 1 vs 2; OE, overexpression; D, down; U, up; Am, ambivalently change depending on probes. 1/4, 2/4, 3/4 4/4 mean above the top quartile, between the top quartile and the median, between the median and the bottom quartile, and below the bottom quartile, respectively. TFs whose score in tissue-specificity above is regarded as tissue-specific in the previous report. S1-1 ~ S5-1 are corresponding to the source numbers in Results Table 6. For fold-changes in the column 4-5-5, those only above and below the thresholds for up- and down-regulation are shown, respectively. The number of the Results Section where the analysis was conducted is indicated at the top of the column. For details of eQTL SNP in sALS, see Discussion Table 1.
Abbreviations for module colors: Blk, Black; Blu, Blue; Bwn, Brown; Gre, Green; GrY, Greenyellow; Mag, Magenta; Mid, Midnightblue; Pin, Pink; Tur, Turquoise; Yel, Yellow.

Results Table 8B. Non-motoneuronal TFs examined by *in vitro* luciferase assay.

TF	4-4-2	4-4-1	4-4-3		5-1-3		5–2	4-4-5								4-4-6							
	In the Data base	Cor with *ADARB1*	WGCNA		Tissue Specificity		Change with aging	Differential expression in the patients' tissues								Interacted by			MiR-141			MiR-200a		
			Cor with *ADARB1*	Mod col				MNs		AHs	SP	Motor cortex				FUS	mFUS		OE			OE		
												S4-1							Fold-change		Predict	Fold-change		Predict
					Score	Hub		S1-1	S2-1	S2-2	S3-1	A1-C	A2-C	A1-A2	S5-1	Protein of TF	mRNA of TF	mRNA of TF	A498	Caki-1		A498	Caki-1	
CEBPA			-0.396	Cyan	1.4	Spe						Am	D	U			✔	✔						
EBF1		0.874	0.527	Tur													✔		0.62			0.62	0.63	
FOXA1					3							D	Am	Am					0.55	0.51	✔	0.73	0.60	✔
FOXB1													D											
FOXD1																	✔	✔				1.53	1.16	✔
FOXP2				Mid	1								U	D			✔	✔		0.72				
FOXP3	✔				1.49														0.30	0.41		0.29	0.38	
MSX1		0.427	-0.360	Tur	1.01	Spe								D					1.62	1.39	✔			
RFX2		0.770		Tur	0.45									U			✔					0.55	0.72	
RUNX1				Mag	0.81	Ubi					U		D				✔	✔			✔			
SETBP1		0.788		Tur									D	U			✔	✔	0.48					
TCF4		0.687		Bwn	0.56	Ubi							D	U			✔	✔	1.82			2.35		
ZNF781		0.718		Tur																		0.55	0.47	
ZNF876P		0.700		Yel																0.62				

Cor, correlation; WGCNA, weighted gene coexpression analysis; Mod col, module color; Ubi, ubiquitous hub; Spe, tissue-specific hub; MN, motor neuron; AH, anterior horn; SP, spinal cord; A1-C, ALS group 1 vs Control; A2-C, ALS group 2 vs Control; A1-A2, ALS group 1 vs 2; OE, overexpression; D, down; U, up; Am, ambivalently change depending on probes. TFs whose score in tissue-specificity above is regarded as tissue-specific in the previous report. S1-1 ~ S5-1 are corresponding to the source numbers in Results Table 6. For fold-changes in the column 4-5-5, those only above and below the thresholds for up- and down-regulation are shown, respectively. The number of the Results Section where the analysis was conducted is indicated at the top of the column.

Abbreviations for module colors: Bwn, Brown; Mag, Magenta; Mid, Midnightblue; Tur, Turquoise; Yel, Yellow.

5. DISCUSSION

5-1. MN-specific regulation of *ADARB1* expression

The CAGE analysis on the laser-captured MNs revealed both the active promoters of *ADARB1* and the profile of gene expression in the human MNs. Based on this CAGE data and the previous multiple data, the present *in vitro* luciferase assay demonstrated that many TFs could regulate *ADARB1* expression in the MNs.

5-1-1. CAGE may detect higher transcriptional activity specific to the human MNs

Although relatively fewer CAGE signals were obtained in the MNs, these signals were highly correlated between the two biological replicates of the MNs and independent from the signals in the control technical replicates of the dorsal horns and white matter [see **Results Figure 1B**]. Because RNAs are not seriously damaged through the procedure of the laser-capture microdissection[221], this independent correlation may reflect the difference in transcriptional activity but not the procedure for sampling between the MNs and the controls. Therefore, this result indicates that these CAGE signals confidently represent higher transcriptional activity in the MNs. In other words, this study provides valuable genome-wide high transcriptional activity at a nucleotide level specific to the human MNs for the first time. Importantly, CAGE signals around *ADARB1* in the MNs differed from those in the dorsal horns and white matter, suggesting that MN-specific enhancers regulate *ADARB1* expression in a MN-specific

way. Based on this CAGE data of the MNs and the public data[10,22,30,178], I defined the p2 and the p1 promoters as active promoters of *ADARB1* in the MNs [see **Results Figure 2B**].

5-1-2. *ADARB1* expression may be differently regulated from the mouse ADAR2 gene

Consistent with a previous report of species-specificity in the DNA sequences of promoters[80], the present study confirms poor conservation between humans and mice in the position and DNA sequences of their ADAR2 gene alternative promoters [see **Results Figures 2B-D** and **3**]. In a similar way, consistent with previous reports describing species-specificity between humans and mice in the repertories of their TFs[47] and their transcriptomes[83,84], I found poor conservation in expression and repertoire of the motoneuronal TFs between the two species [see **Results Figure 7A**]. Since signals detected by CAGE and RNA-seq in the same tissue type are highly correlated[222], the modest correlation in expression of the human and mouse motoneuronal TFs demonstrated in the present study may reflect differences between these two species rather than methodological differences between CAGE and RNA-seq [see **Results Figure 7B**]. Hence, the present study indicates that expression of human *ADARB1* is regulated at different alternative promoters by different motoneuronal TFs from the mouse counterparts and that mice might be improper for an *in vivo* experimental model to examine a role of human *ADARB1*-regulatory TFs in the human MNs.

5-1-3. *ADARB1* expression in the MNs may be regulated by an MN-specific set of multiple TFs in its alternative promoter level

So far, regulatory function of only one non-motoneuronal TF TP53 has been validated by *in vitro* luciferase assay using yeast and an osteosarcoma cell line for an *ADARB1* promoter that had been differently defined from the present study (i.e., in the region about 1.1 kb upstream of the present p2 promoter)[148]. Aside from that TP53, the present luciferase assay newly determined 33 motoneuronal and 13 non-motoneuronal TFs as *ADARB1*-regulatory TFs [see **Results Figure 10**]. The presence of possible multiple *ADARB1*-regulatory TFs is consistent with previous reports that TFs regulate gene expression in cooperative and redundant manners[39,48,54,59–63] via their tissue-specific regulatory network[59,64–66] and expression[47] while mutually buffering their effects on promoters[56].

Among the *ADARB1*-regulatory TFs determined in the present study, which ones are specifically expressed in the MNs? According to a previous report defining the tissue-specificity of TFs and their ability to be a tissue-specific hub[48] and an additional analysis in the present study on expression patterns of TFs in human tissues and cells with the publicly available RNA-seq data (the Data Set 1 and 2), expression of the 10 examined motoneuronal TFs was MN-specific (AR, ESRRG, HOXB7, ID4, KLF7, MLLT11, RXRG, ZEB1, ZEB2 and ZFPM2) [summarized in **Results Table 8A; Discussion Figures 1A and 1B**]. Alt-

hough a tissue specific score of ESRRG (0.97) was slightly below the previously defined threshold (i.e., 1.0), the RNA-seq data clearly guaranteed its tissue-specificity; tissue-specificity of MLLT11, ZEB1 and ZEB2 whose score was unavailable has been already reported[223,224]. Consistent with the results of the present WGCNA determining AR as MN-specific [see **Results Figure 8D**], AR was previously defined as a tissue-specific hub; in addition to AR, RXRG also previously defined as a tissue-specific hub.

Protein-protein interactions[204] among the examined motoneuronal TFs and other basic transcriptional machinery (e.g., subunits of RNA polymerase II) revealed that there was one large activating TF complex: this putative TF complex was composed of eight activators (ARNT, CREB1, ENO1, ESRRG, HIF1A, MAX, NR3C1 and NR3C2) and interacted with the basic transcriptional machinery via ARNT, ENO1, MAX and NR3C1 [**Discussion Figure 2**]. Consistent with a previous report that protein-protein interaction of TFs consists of both TFs ubiquitously expressed in many tissue-types and ones specifically expressed in a fewer tissue-types[48], this putative complex may exist in an MN-specific way because of MN-specific expression of ESRRG. Moreover, because the present CAGE data on the laser-captured MNs showed high expression of most of these TFs [see **Results Table 8A**], all of these TFs belonged to the TF modules in the largest module cluster that might regulate *ADARB1* expression in neurons in the present WGCNA (i.e., Module Black, Pink, Red, Tan and Tur-

quoise) [see **Results Figure 8B**], and because ChIP-seq signals and/or consensus motifs of six TFs (ARNT, CREB1, ENO1, HIF1A, MAX and NR3C1) existed within the p2 and the p1 promoters that were active in the MNs [see **Results Figure 10A**], this putative TF complex could regulate *ADARB1* expression in an MN-specific manner. This expectation is indirectly supported by a previous report showing positive correlation between MAX expression and the whole RNA editing levels in human tissues[225]. Activity of ARNT and CREB1 in the p1 promoter could be indirect through selectively activating transcription at the p13@ whose activity was significantly correlated with that at the p1@ [see **Results Figure 4B**]. Although both AR and MXI1 also interacted with the putative TF complex [see **Discussion Figure 2**], neither of these two TFs activated the p2 nor the p1 promoters in the present luciferase assay; this discrepancy between the expectation and the results of the present luciferase assay could be caused from the present experimental design that each of these TFs was singly transfected into HeLa cells together with the promoter construct or the lack of intracellular condition that requires activation of these TFs. AR might require external addition of its ligand to bind its target, and MXI1 exclusively binds to its targets by forming a heterodimer with MAX[226]. On the other hand, the other universal activators, ID4, KLF7, NFIA and ZEB1, were independent from the putative TF complex. Therefore, these universal activators possibly contribute to other TF complexes regulating *ADARB1* expression in the MNs.

The present luciferase assay demonstrates that the 10 motoneuronal TFs did not activate the p1 promoter but activated the p3/p4 promoter, whose transcriptional activity in the MNs was below the detection limit in the present CAGE analysis [see **Results Figure 10A**]. Hence, as implied in the present *in silico* analyses [see **Results Figures 4B** and **4C** and **Results Table 5**], the endogenous p3/p4 promoter in the MNs could be independently regulated as actively as the p2 and the p1 promoters, based on the selective activation of the p3/p4 promoter by these 10 TFs and possibly on epigenetic modifications around it. Indeed, motoneuronal TFs HOXB7 and MEF2A and a non-motoneuronal TF FOXB1 suppressed the p1 promoter but oppositely activated the p3/p4 promoter, implying their putative role in shifting transcription from the p1 promoter to the p3/p4 promoter, as demonstrated in the expanded memory T cells[203] [see **Results Figure 4C**].

In the present luciferase assay demonstrated that most non-motoneuronal TFs were competent to regulate transcription at one or more of the four *ADARB1* promoters. Because these non-motoneuronal TFs are expressed in adult tissues other than the MNs or fetal tissues, they may play their regulatory role of *ADARB1* expression in those tissues.

In short, in the human MNs, each of the four *ADARB1* promoters may be redundantly regulated to some extent by a different MN-specific set of multiple TFs; these TFs may keep the p2 and the p1 promoters active but the others rather quiescent, thereby complicatedly reg-

ulating the total expression levels of *ADARB1* in the MNs in its alternative promoter level.

5-2. Implication for the ALS pathogenesis

Providing knowledge of the putative regulatory mechanism of *ADARB1* expression in the human MNs, the present study enables to infer roles of the *ADARB1*-regulatory motoneuronal TFs in the pathogenesis of sporadic ALS. The aforementioned putative complicated regulation of *ADARB1* expression indicates that complicated causative abnormalities may underlie the down-regulation of *ADARB1* in the patients' MNs: abnormalities of not a single but multiple TFs may underlie the down-regulation of *ADARB1* in the patients' MNs, and such abnormalities possibly have some variation among the individual patients. Indeed, the present analysis of the public data showed that the activators of the *ADARB1* promoters MLLT11 and ZEB1 apparently decreased in the MNs of some but not all of the 13 patients with sporadic ALS[166] [see **Discussion Figure 1A**].

What kinds of abnormalities related to the *ADARB1*-regulatory TFs could cause the down-regulation of *ADARB1* in the patients' MNs? In accordance with the working hypothesis in the present study, one possible abnormality is the age-dependent decrease of the *ADARB1*-regulatory TFs, whose expression is inherently low in the patients' MNs because of

alternative alleles at eQTL of these TFs; inherently low expression of the other TFs whose expression is rather constant throughout the life might also underlie the down-regulation of *ADARB1* [**Discussion Figure 3**]. Indeed, according to a previous report that comprehensively defines all genes whose expression changes with aging in human tissues[86], 70 motoneuronal TFs are significantly decreased with aging, including MLLT11 that was demonstrated in the present study as the activator of the p1, the p13 and the p3/p4 promoters [see **Results Table 8A**]. In addition, results of exome-sequencing (ALS Variant Server[206] and ALS Data Browser[207]) and genome-wide association studies (ALS Gene[208]) on the patients with sporadic ALS report many SNPs at eQTL[158] of the 10 examined motoneuronal TFs, including MLLT11 [**Discussion Table 1**]. Although an exact effect of SNPs at eQTL in the MNs is uncertain, the patients in the investigated cohort have more homozygous minor alternative alleles at eQTL that are associated with low expression of MLLT11 than the homozygous major reference alleles of this motoneuronal TF. In other words, these homozygous minor alternative alleles at eQTL of MLLT11 could cause this TF to be inherently less expressed in the patients' MNs, resulting in low basal expression of *ADARB1*. Because MLLT11 is a neuron-specific marker[223] highly expressed in the MNs in the present CAGE and the previous RNA-seq data [see **Discussion Figure 1A** and **Results Table 8A**], MLLT11 may have its MN-specific regulatory role in gene expression. Based on this series of indi-

rect evidence, decrease of MLLT11 in the laser-captured MNs[166,168], the whole spinal cords[193], and motor cortices[194,195] of the patients with sporadic ALS implies that such decrease might be not responsive but could contribute to the down-regulation of *ADARB1* in the patients' MNs.

In addition to the age-dependent decrease of the *ADARB1*-regulatory TFs, the down-regulation of *ADARB1* could also be caused by change in their quality which, for instance, might be due to aberrant alternative splicing or post-translational modifications and by SNPs at eQTL of *ADARB1* within its promoters and enhancers [**Discussion Figure 4**]. Although the latter two are rather age-independent factors, these three abnormalities related to the *ADARB1*-regulatory TFs may have mutually not exclusive but summative effects on the down-regulation of *ADARB1*. Although several studies have already attempted but failed to explain its pathogenesis with the repeat number of CAG in AR gene[227–229] (i.e., the quality of AR protein), all of these three possible abnormalities related to the *ADARB1*-regulatory TFs should be considered when the cause of the down-regulation of *ADARB1* in the patients' MNs is further examined in the future.

Are abnormalities of the motoneuronal TFs regulating the p2 and the p1 promoters in the MNs enough to cause the down-regulation of *ADARB1* in the patients' MNs? The answer to this question might be "No," because the public high-throughput data on the patients' MNs[166,168,192] showed decrease of RXRG, the selective activator for the p3/p4 promoter [see

Results Figure 10A and **Results Table 8A**]. RXRG is a tissue-specific hub TF[48] and preferentially binds to its target DNA sequences with mCpG[53] that are enriched within actively transcribed gene bodies[77–79], suggesting its MN-specific role in regulating the endogenous mCpG-rich p3/p4 promoter within the gene body of *ADARB1* in the MNs. Hence, the decrease of RXRG observed in the patients' MNs might have contributed to its pathogenesis. If ADAR2 proteins expressed from the p3/p4 promoter are competent for RNA editing at Q/R site of GluA2, the p3/p4 promoter could be compensatively activated via its dynamic and independent regulation upon impairment of activity of the p2 and p1 promoters in the MNs. Therefore, to further infer contribution of such p3/p4 promoter-selective activators to the pathogenesis of sporadic ALS, enzymatic competence of the ADAR2 proteins expressed from the p3/p4 promoter (and might also from the p13 promoters) should be determined first.

In short, when the p2 and the p1 promoters of *ADARB1* are regulated by multiple TFs in the MNs, abnormalities of multiple TFs may cause the down-regulation of *ADARB1* in the MNs of patients with sporadic ALS.

5-3. Assessment of the approaches for predicting and validating *ADARB1*-regiulatory TFs in the present study

Here, I briefly assess all the approaches adopted in the present study to predict *ADARB1*-regulatory TFs as comprehensive as possible, based on (1) data availability, (2) efficiency of that prediction in narrowing down the candidates from all the 1,334 motoneuronal TFs, (3) consistency of that prediction with the results of the present luciferase assay of the 34 motoneuronal TFs, and (4) possible bias in that approach [**Discussion Table 2**]. Also, here I denote limitations of the present luciferase assay for validating *ADARB1*-regulatory TFs.

5-3-1. Prediction from correlation in expression between TFs and *ADARB1*

Using (1) abundantly available high-throughput public data of gene expression profiles, (2) calculating correlation can efficiently list up the candidate TFs by setting an arbitrary threshold of the correlation coefficient (as in the Results Section 4-4-1) or statistical significance (as in the Results Section 4-4-3). Surprisingly, for both of these predicting approaches, (3) the present luciferase assay showed that the candidates tended to activate the p2, the p13 and the p3/p4 promoters compared to the p1 promoter; this tendency remained even at a lower threshold of Spearman' rho ($|rho| > 0.4$ in the Results Section 4-4-1) [the parentheses in **Discussion Table 2**], suggesting preference of these approaches to predict TFs selectively activating the promoters other than the p1 promoter. Although these correlation-based approaches achieved the most profitable prediction in the present study, (4) these approaches are clearly under bias because these were based only on the total expression of *ADARB1* and dis-

missed individual activity of the *ADARB1* alternative promoters that was tested by the present luciferase assay. In other words, calculating correlation in every CTSS activity between TFs and *ADARB1* in the human tissues might enable to predict *ADARB1*-regulatory TFs more accurately in the alternative promoter level.

5-3-2. Prediction from the database for putative target genes of TFs

Based on experimental validation (e.g., ChIP assay and luciferase assay), (1) the public database ORTI[180] suggested only two motoneuronal TFs as a putative *ADARB1*-regulatory TF, enabling to (2) regard these two out of the 1,334 TFs as the candidates. The present luciferase assay showed that (3) a candidate motoneuronal TF AR and another candidate non-motoneuronal TF FOXP3 did not activate and strongly activated all the *ADARB1* promoters, respectively [see **Results Figures 10A** and **10B**]. (4) Possible bias in this approach is a different definition of the promoter regions of *ADARB1* between the present study and previous reports curated by this database.

5-3-3. Prediction from public data of ChIP-seq and consensus motifs of TFs

(1) The database ChIP-Atlas[188] and the previous reports[52,53,55,189,190] provided the ChIP-seq and consensus motif data of approximately 25% of the 1,334 motoneuronal TFs, respectively [see

Results Table 5], (2) only 18% and 45% of which were regarded as the candidates based on presence of their signals and motifs within the promoters, respectively. However, the present *in vitro* luciferase assay revealed that (3) only about a half of the examined TFs whose data were available really activated the promoter within which their ChIP-seq signals or consensus motifs existed or was inactive for the promoter within which the signals or motifs were absent. (4) The possible bias in the prediction using the ChIP-seq data is the cell-type-specific intracellular environment (e.g., difference in activity of the *ADARB1* alternative promoters and other co-expressed TFs with which the examined TF binds to the promoters in that cell); and non-functional or weak binding of the examined TF that is solely insufficient to activate or suppress the promoter[56,63]. On the other hand, (4) the possible bias in the prediction using the consensus motifs is variability of these motifs: although the consensus motifs can vary depending on binding manners of TFs and contexts of DNA sequences[53,55,61,230], not all those variants are yet to be determined.

5-3-4. Extraction of TFs decreasing in the tissues of patients with sporadic ALS

(1) High-throughput data of differentially expressed gene analyses on the tissues of patients with sporadic ALS have been accumulating in public, and such data are expected to increase further in the future. However, (2) this approach did not straightforwardly allow the putative

causative TFs to surface from all the motoneuronal TFs, because approximately a half of the 1,334 motoneuronal TFs decreased in one or more data sets of the studies on the patients' tissues[166,168,192–195] [see **Results Figure 9**]. The present luciferase assay showed that (3) nearly a half of the candidate TFs in this approach activated at least one of the *ADARB1* promoters; this also holds true when I consider only the TFs decreasing in the patients' MNs[166,168] and/or the whole spinal cords[193] [the parentheses in **Discussion Table 2**]. (4) The data of the differentially expressed gene analyses may not only be under bias in the original statistical method and their procedure for sample collection but also in indistinguishable co-existence of the causative and responsive change[195] and dismissing differential expression among the individual patients. The bias in sample collection is critical especially for the data using the laser-captured MNs, because the patients' single MN harboring the sufficiently edited GluA2 at its Q/R site contains much more ADAR2 mRNA than the one harboring the unedited GluA2[120]; therefore, when the total RNA is extracted from the pooled samples of the laser-captured MNs, the crucial down-regulation of *ADARB1* in the degenerated MNs could be masked by relatively preserved ADAR2 mRNAs in less degenerated MNs that could have been collected more often. Indeed, in the public data of the MNs used in this study, *ADARB1* in the patients' MNs was insignificantly down-regulated or expressed in near levels to that in the control subjects, depending on the individuals. In addition, given the afore-

mentioned possibility that abnormalities of the *ADARB1*-regulatory TFs might vary to some extent among the individual patients, a statistical analysis on the transcriptomes of a cohort of the patients could fail to detect such individual difference. Hence, such bias altogether suggests that statistically insignificant change in expression of the motoneuronal TFs in the patients' MNs may never exclude those TFs from the list of the putative causative TFs down-regulating *ADARB1* expression in the MNs, leaving interpreting data of the differentially expressed gene analyses quite difficult without examining their regulatory role on the *ADARB1* promoters by luciferase assay.

5-3-5. Extraction of TFs in possible molecular cascades of pathogenesis of mutant FUS-linked ALS

(1) This approach utilized many previous data available in public but (2) listed so many candidates [see **Results Figure 9**]. Although (3) about half of the motoneuronal TFs listed in this approach really activated one or more of the tested *ADARB1* promoters, (4) this approach is based on the poorest rationale, because molecular links between the pathogenesis of sporadic and mutant FUS-linked ALS yet to be revealed. The present results in this approach might be rather useful to survey the pathogenesis of mutant FUS-linked ALS in the future when most *ADARB1*-regulatory TFs causing sporadic ALS are determined.

5-3-6. Limitations of the present luciferase assay for validating *ADARB1*-regulatory TFs

The ChIP-seq and consensus motifs whose data were used in the present *in silico* prediction are mainly for the TFs highly expressed in culture cell lines, and difference in expression levels of TFs between the MNs and the culture cell lines may underlie the poor availability (approximately 25%) for these data of the motoneuronal TFs [see **Results Table 5**]. Moreover, as already mentioned, the prediction based on these data is not highly accurate, implying the possible false negative in this prediction. Nevertheless, the TFs extracted by the present *in silico* predicting approaches did demonstrate their regulatory function for the *ADARB1* promoters in the present luciferase assay, suggesting their faithful regulatory roles in *ADARB1* expression.

It is unclear to what extent the demonstrated regulatory function of the examined motoneuronal TFs for the *ADARB1* promoters in the present luciferase assay using HeLa cells exactly recapitulates their function in the MNs, because the examined TFs demonstrated their function for the *ADARB1* promoters in cooperation with other TFs that were co-expressed not in the MNs but in the HeLa cells, and because the tested promoters were independent from the original chromosomal environment and regulation by their enhancers. Nevertheless, the examined motoneuronal TFs activating the *ADARB1* promoters in the present luciferase assay

using the HeLa cells may activate the endogenous *ADARB1* promoters in the MNs as well. Although altering expression of the motoneuronal TFs in a cell line whose transcriptome mimics the MNs can reveal their true role in regulating *ADARB1* expression in the MNs, such a cell line is currently unavailable. Indeed, even iMNs have different transcriptome from the real MNs [see **Discussion Figures 1A** and **1B**], possibly because of their immaturity[231].

5-4. What should be done next, to explain why and how the down-regulation of *ADARB1* in the MNs of patients with sporadic ALS occurs?

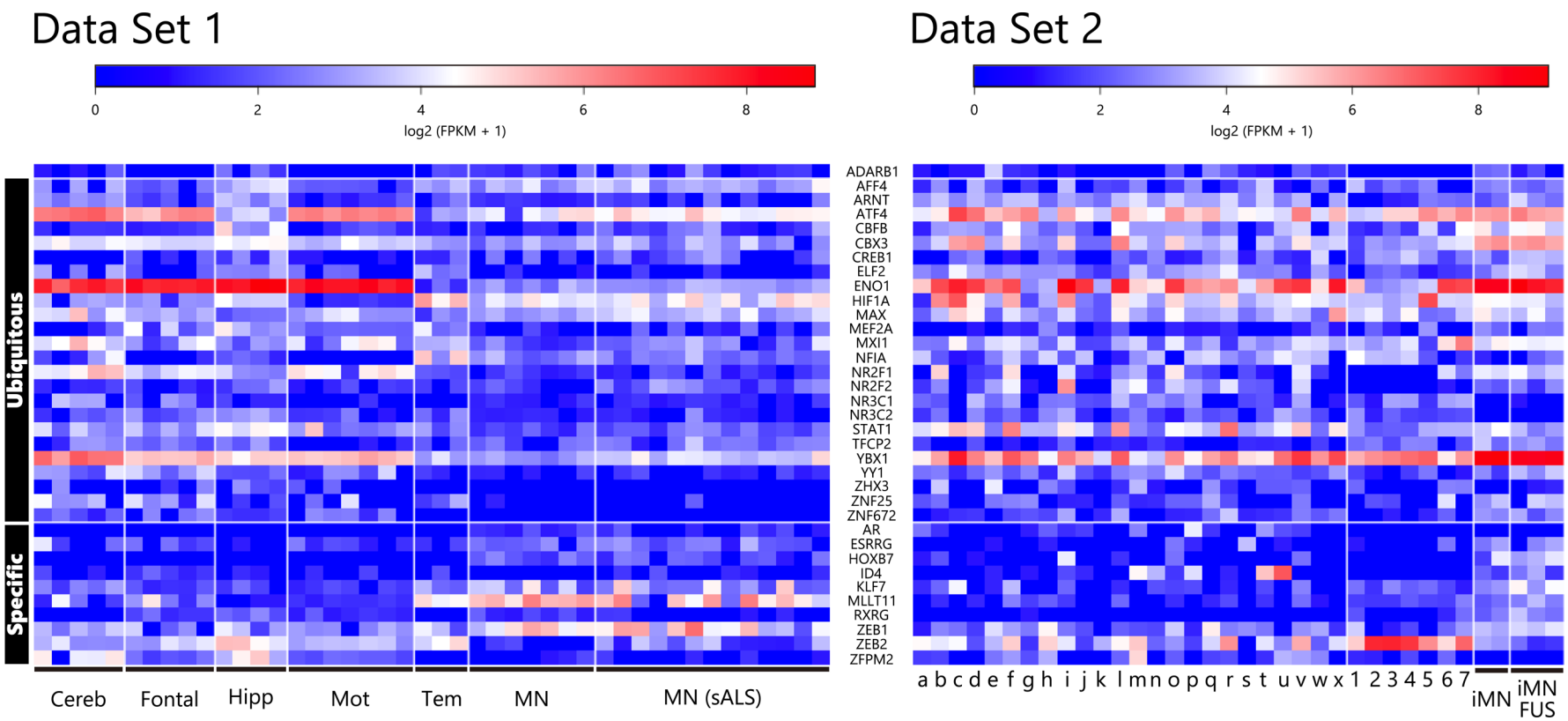
The present study implies that abnormalities of the multiple *ADARB1*-regulatory TFs may underlie the down-regulation of *ADARB1* in the MNs of patients with sporadic ALS, thereby establishing the entrance of the final goal to explain why and how such down-regulation of *ADARB1* occurs in the patients' MNs. Because of the aforementioned false negative results in the current *in silico* predicting approaches, other motoneuronal TFs that were not examined in the present *in vitro* luciferase assay can also regulate *ADARB1* expression in the MNs. Therefore, to achieve this final goal, it still requires to determine additional *ADARB1*-regulatory TFs in the MNs and then to predict a crucial subset of the *ADARB1*-regulatory TFs for maintaining *ADARB1* expression. Besides, the present study

concentrated solely on the *ADARB1* promoters to investigate the regulatory mechanism of *ADARB1* expression; because transcriptional activity of the *ADARB1* promoters in the MNs may be regulated by MN-specific enhancers [see **Results Figure 2A**], regulatory effects of such enhancers on *ADARB1* expression in the MNs should be investigated in the future.

To determine additional *ADARB1*-regulatory TFs in the MNs, it is predictable to use information of protein-protein interactions of the unexamined TFs with the activators determined in this study and positions of ChIP-seq signals and consensus motifs of the unexamined TFs within the *ADARB1* promoters, although physical interaction among TFs is not necessarily required for their cooperative function[61]. Regulatory function of newly predicted *ADARB1*-regulatory TFs on the *ADARB1* promoters should be examined by *in vitro* luciferase assay in the same way as the present study.

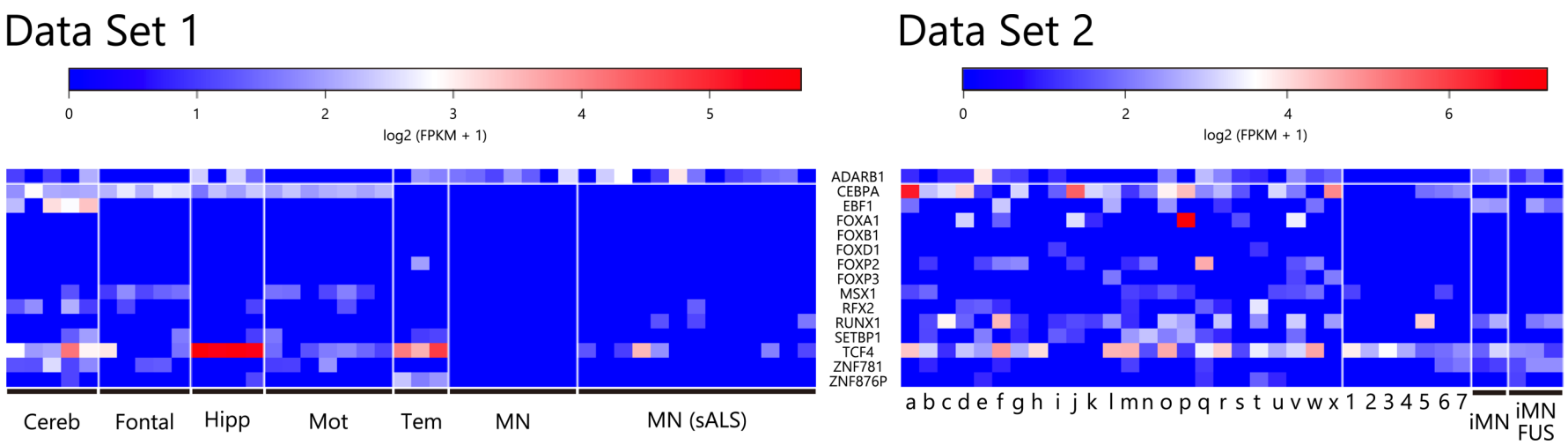
What and how many TFs may contribute to the down-regulation of *ADARB1* in the patients' MNs? Answering this difficult question demands prediction of a crucial subset of the *ADARB1*-regulatory TFs to maintain *ADARB1* expression. In this regard, because genes specifically expressed in a sub-region of the brain have been shown to significantly associate with diseases in the brain regardless of their expression levels[85], the *ADARB1*-regulatory TFs expressed not ubiquitously but specifically in the MNs should be concerned in no matter what levels they are expressed in these neurons. As described above, to ask if roles of the p3/p4

and p13 promoter-selective motoneuronal TFs can be crucial, unexamined enzymatic activity of ADAR2 expressed from transcripts with Exon 0 or Exon 1a should be determined. In addition, one executable way is to deduce the crucial subset of the TFs by seeking SNPs at eQTL[158,232] of the TFs around the CAGE signals in the MNs from the patients' genome sequences available on dbGaP website (https://www.ncbi.nlm.nih.gov/gap), where authorized users are allowed to access such data. This approach only requires additional *in silico* analyses and the authorized accession to dbGaP; notably, the GTEx Consortium updated data of eQTL during preparing this dissertation, and data of eQTL in human spinal cord tissues are now available on its website (latest version V7; https://www.gtexportal.org/home/). Other ways are to *in vitro* knockdown or to simulate *in silico* the knockdown of the multiple *ADARB1*-regulatory TFs, thereby assessing resultant change in the *ADARB1* expression; however, either way requires a currently unavailable cell line faithfully mimicking the real adult human MNs for gaining valid results and validating the simulated results, respectively. Adopting the approaches suggested here and additional ones that may be available in the future will lead to depict the whole picture of the pathogenesis of sporadic ALS.



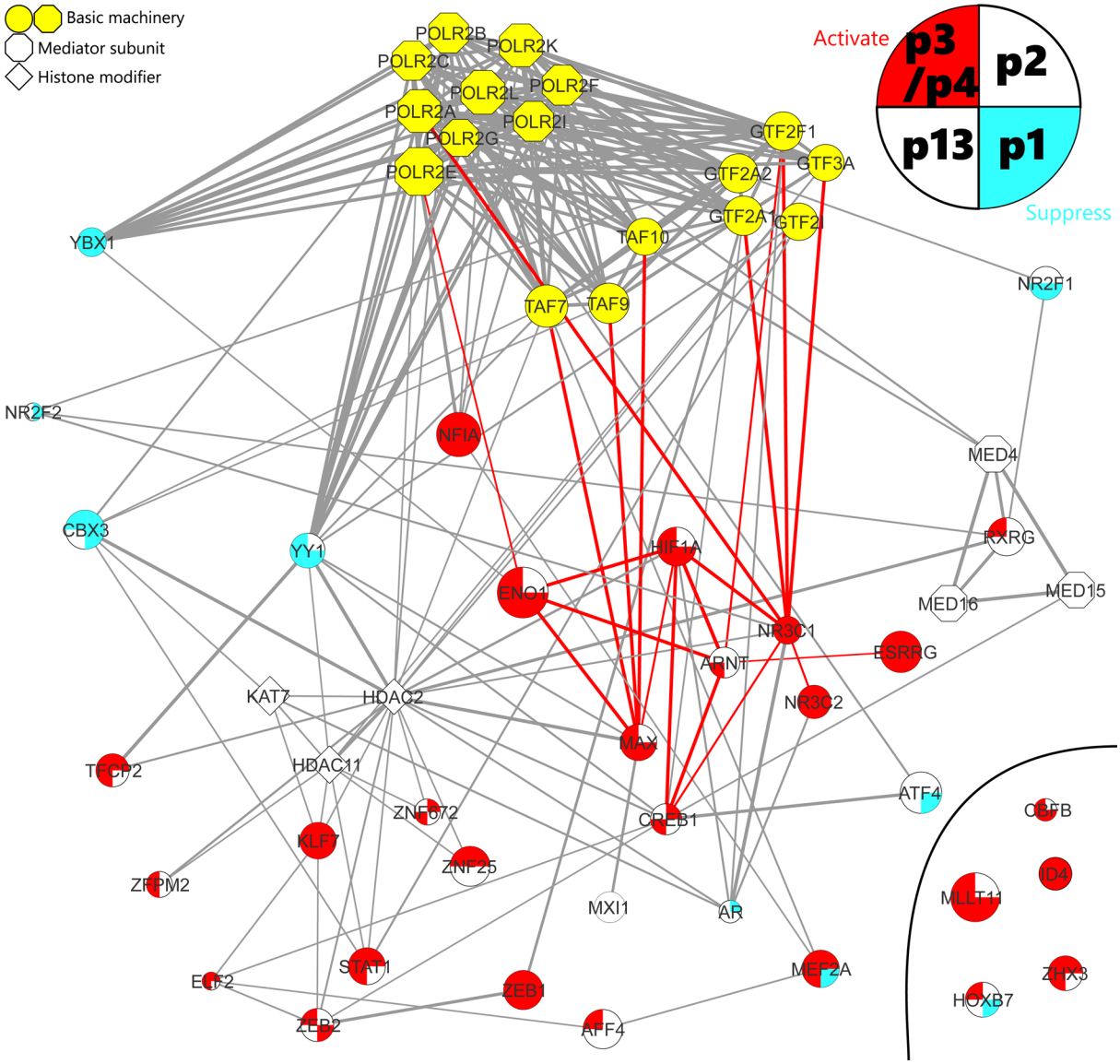
Discussion Figure 1A Expression levels of *ADARB1* and the examined motoneuronal TFs.

Expression levels of *ADARB1* and the motoneuronal TFs examined by luciferase assay are shown. The duplicated tissues between the Data Set 1 and 2 are not shown. The color key at the top in each panel directs the expression values. Cereb, cerebellum; Frontal, frontal cortex; hippo, hippocampus; Mot, motor cortex; Tem, temporal cortex; MN; motor neuron; sALS, sporadic ALS; iMN, motor neurons differentiated from induced pluripotent stem cells. a, adipose; b, adrenal gland; c, bone marrow; d, duodenum; e, esophagus; f, gallbladder; g, gastrocnemius muscle; h, heart; i, kidney; j, liver; k, lung; l, lymph node; m, ovary; n, pancreas; o, placenta; p, prostate; q, sigmoid colon; r, spleen; s, stomach; t, testis; u, thyroid gland; v, urinary bladder; w, uterus; x, whole blood; 1, astrocyte; 2, oligodendrocyte-precursor cells; 3-4, oligodendrocytes (replicates); 5, microglia; 6 and 7, grey and white matter of frontal cortex, respectively.



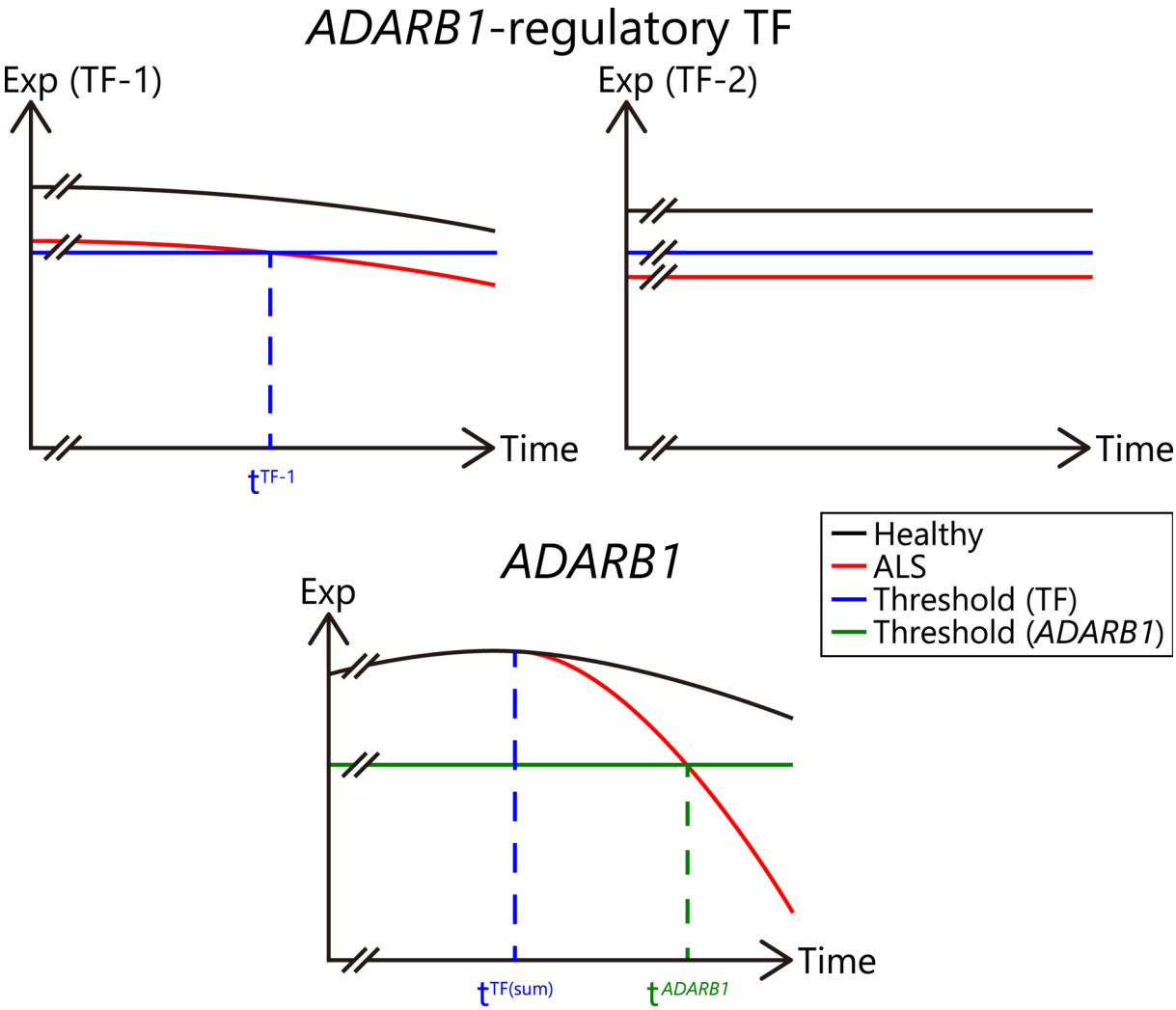
Discussion Figure 1B Expression levels of *ADARB1* and the examined non-motoneuronal TFs.

Expression levels of *ADARB1* and the non-motoneuronal TFs examined by luciferase assay are shown. The duplicated tissues between the Data Set 1 and 2 are not shown. The color key at the top in each panel direct the expression values. Cereb, cerebellum; Frontal, frontal cortex; hippo, hippocampus; Mot, motor cortex; Tem, temporal cortex; MN; motor neuron; sALS, sporadic ALS; iMN, motor neurons differentiated from induced pluripotent stem cells. a, adipose; b, adrenal gland; c, bone marrow; d, duodenum; e, esophagus; f, gallbladder; g, gastrocnemius muscle; h, heart; i, kidney; j, liver; k, lung; l, lymph node; m, ovary; n, pancreas; o, placenta; p, prostate; q, sigmoid colon; r, spleen; s, stomach; t, testis; u, thyroid gland; v, urinary bladder; w, uterus; x, whole blood; 1, astrocyte; 2, oligodendrocyte-precursor cells; 3-4, oligodendrocytes (replicates); 5, microglia; 6 and 7, grey and white matter of frontal cortex, respectively.



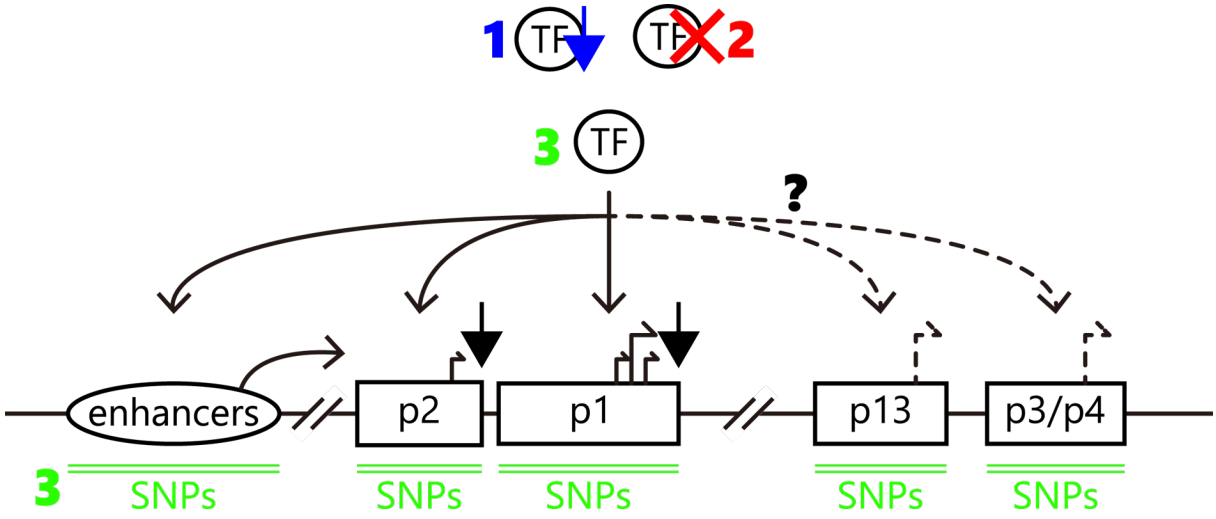
Discussion Figure 2 Protein-protein interactions among the examined TFs and other proteins in the transcriptional machinery.

Interactome retrieved from STRING v10.5 based on the experiments, databases and co-expression with > 0.15 confidence is shown. Nodes of the examined TFs are colored, based on their activating or suppressing effect on the four *ADARB1* promoters, respectively (see the example chart at the top right). The legend at the top left directs the nodes of the transcriptional machinery. The size of nodes and the thickness of edges are in accord with the expression level of those proteins and the confidence of the interactions, respectively. Some interactions mentioned in the text are highlighted with red edges. TFs separated by line at the bottom right show no reported interaction with the others.



Discussion Figure 3 Implication for the pathogenesis of sporadic ALS based on the working hypothesis in this study with some modifications.

ADARB1 expression is cooperatively regulated by multiple TFs classified as TF-1 (top left) and TF-2 (top right) groups. The TF-1 decreases with aging even in the motor neurons (MNs) of the healthy (e.g., MLLT11), while the TF-2 is constantly expressed throughout the life (e.g., ENO1). In the MNs of patients with sporadic ALS (red), such *ADARB1*-regulatory TFs are inherently fewer compared to the healthy (black) due to alternative alleles at expression quantitative trait loci of these TFs decreasing their basal expression. Thereby, the TF-1 in the patients' MNs may decrease below the threshold for maintaining the required expression of *ADARB1* (blue horizontal line) much earlier than the healthy ($t^{TF\text{-}1}$). When expression levels of key *ADARB1*-regulatory TFs are all below the threshold ($t^{TF(sum)}$), *ADARB1* may begin to be down-regulated in the patients' MNs. Exp, expression level; TF, transcription factor.



Discussion Figure 4 Possible causes down-regulating *ADARB1* expression in the MNs of patients with sporadic ALS.

There are three possible causes down-regulating *ADARB1* expression in the patients' MNs: 1, decrease of the *ADARB1*-regulatory TFs; 2, qualitative abnormality of the TFs; 3, SNPs within the *cis*-regulatory regions of *ADARB1* (i.e., enhancers and the four alternative promoters) that weaken binding affinity of the TFs to these regions. As indicated in the question mark, however, it is unclear whether or not transcription at the p3/p4 and/or the p13 promoters is activated for compensation, when transcriptional activity at the dominant p1 and supportive p2 promoters decreases. SNPs, single nucleotide polymorphisms; TF, transcription factor.

Discussion Table 1. SNPs at eQTL of the tested TFs found in the patients.

TF	dbSNP rs ID	Effect of Alt	Ref	Alt	#Genotype			Avg depth	A/A >R/R	Src
					R/R	R/A	A/A			
AFF4	rs803223	D	G	A	3	0	10	1.1	✔	V
	rs708458	D	A	G	1	0	19	1.2	✔	
	rs803221	D	T	C	32	20	97	2.6	✔	
	rs708457	D	A	C	2	0	24	1.1	✔	
	rs6596093	D	A	G	4	34	227	4.8	✔	
	rs12654316	U	C	T	20	0	2	1.2		
ARNT	rs10847	D	C	T	13	0	9	1.2		V
	rs11552229	U	T	C	16	0	17	1.2	✔	
	rs2228099	U	C	G	99	131	47	57.6		V/db
	rs10305672	U	G	A	20	1	10	1.1		V
ATF4	rs1048310	U	T	G	10	2	8	1.4		V
	rs710192	U	C	T	25	21	11	4.2		
	rs10854722	U	G	A	33	25	10	10.3		
	rs146172204	D	T	TC	29	2	21	1.3		
	rs4894	U	A	C	122	115	40	38.2		V/db
ENO1	rs2781060	D	T	C	16	92	165	7.4	✔	V
	rs2765511	D	G	A	13	95	168	10.6	✔	
	rs1006950	D	T	C	4	0	23	1.1	✔	
	rs61784221	D	T	G	8	0	12	1.2	✔	
	rs11811795	D	T	C	4	0	29	1.2	✔	
	rs1359591	D	G	A	12	0	7	1.1		
	rs1325920	D	G	A	2	0	7	1.3	✔	
KLF7	rs2269077	U	C	T	14	0	9	1.3		V
	rs2543596	U	C	T	9	0	7	1.1		
	rs2284934	D	C	T						db
	rs2037893	U	C	G	12	25	27	5.2	✔	V
	rs1920048	U	G	A						G
MAX	rs4902357	D	C	G	17	0	6	1.2		V
	rs1957949	D	C	T	18	1	12	1.3		
	rs1957948	D	C	T	20	1	8	1.3		
	rs4902359	D	G	A	35	1	12	1.3		
	rs7159443	D	T	A	124	125	28	54.1		
MLLT11	rs3806387	D	T	A	6	0	23	1.1	✔	V
	rs3806386	D	T	C	9	0	21	1.1	✔	
	rs3738481	D	A	G	13	83	181	29.4	✔	
	rs15740	D	G	A	6	19	36	20.0	✔	
	rs2864699	D	G	C	0	0	14	1.1	✔	

Discussion Table 1 (cont.). SNPs at eQTL of the tested TFs found in the patients.

TF	dbSNP rs ID	Effect of Alt	Ref	Alt	#Genotype			Avg depth	A/A >R/R	Src
					R/R	R/A	A/A			
MXI1	rs35741140	U	G	C	177	59	16	11.5		V
	rs111446990	D	A	T	39	0	5	1.3		
	rs113623728	D	G	A	29	0	3	1.3		
	rs112148911	D	C	T	29	0	3	1.2		
	rs112927374	D	C	G	31	0	3	1.3		
	rs201839727	D	C	*1	99	3	3	1.5		
	rs113238292	D	A	T	161	8	6	1.9		
	rs113138616	D	T	C	200	20	9	2.8		
	rs61745504	D	C	T	224	48	5	58.8		V/db
	rs10884942	U	A	G	46	16	30	2.5		V
	rs16410	D	*2	G	50	1	3	1.4		
	rs1133195	D	A	G	28	1	2	1.1		
	rs14401	D	C	T	12	1	17	1.2	✔	
TFCP2	rs1056897	U	G	A	29	3	7	1.4		V
	rs11169736	U	G	T	30	0	11	1.4		
	rs11738	U	G	A						G
ZNF25	rs13503	D	C	A	7	0	15	1.1	✔	V
	rs1208604	U	A	G	9	0	1	1.3		
	rs1208605	U	G	A	12	0	1	1.3		
	rs201863197	U	*3	T	21	0	3	1.1		
	rs1208606	U	A	C	225	48	4	74.1		V/db

*1, CAAAA; *2, GTTGA; *3, TAA.
Ref, reference allele; Alt, alternative allele; Avg depth, Avgerage sample read depth; Src, data source; V, ALS Variant Server; db, ALS Data Browser; G, ALS Gene. Blank, data not available. Effect of alternative allele was defined by the GTEx Consortium. The number of allele genotypes and read depth are quoted from the ALS Variant Server. SNPs whose genotype count in Alt/Alt is larger than Ref/Ref are indicated with a check mark in the A/A > R/R column.

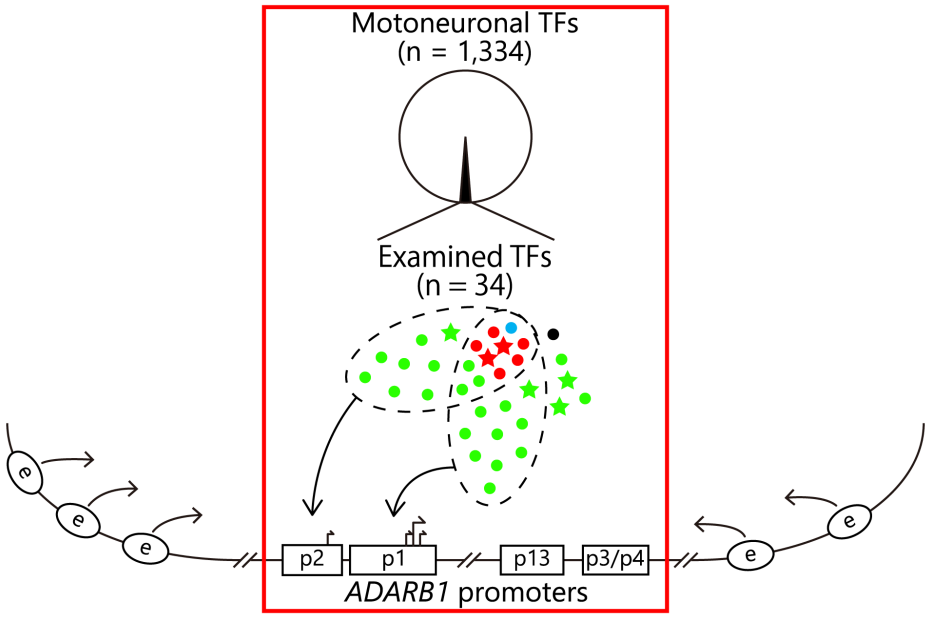
Discussion Table 2. Assessment of the predicting approaches in this study.

Predicting approach		Data availability	Efficiency in narrowing down the candidates	Consistency with luciferase assay of the 34 mnTFs				Possible bias
				p2	p1	p13	p3/p4	
SingleCor	[4-4-1]	H	H	7/10 (10/14)	3/10 (5/14)	7/10 (10/14)	9/10 (12/14)	Difference between the total expression level and individual transcriptional activity atl the promoters of *ADARB1*.
DB	[4-4-2]	L	H	1/1	0/1	0/1	0/1	Definition of the promoter region.
WGCNA	[4-4-3]	H	H	0/5	1/5	3/5	3/5	The same as SingleCor.
ChIP	[4-4-4]	L	H	12/23	10/23	8/23	12/23	Intracellular environment; non-functional TF binding.
Motif	[4-4-4]	L	H	12/25	14/25	17/25	11/25	Variability in the consensus motifs.
Pat. tissues	[4-4-5]	M	L	7/19 (2/6)	8/19 (3/6)	11/19 (4/6)	13/19 (5/6)	The original statistical method; sample collection; responsive change; dismissed change in individuals.
Put. links	[4-4-6]	M	L	13/30	12/30	19/30	20/30	Unrevealed molecular links.

Ratio of the observed to the predicted change for each of the tested *ADARB1* promoter is shown in the "Consistency" columns. Parentheses in the "Consistency" columns of the "SingleCor" row show the consistency at the lower threshold (|Spearman' rho| > 0.4). For the consistency of "ChIP" and "Motif" approaches, when a TF fulfills the following condition, the prediction and the observed luciferase assay results are regarded as consistent, respectively: a TF activated or suppressed the promoter within which its ChIP-seq singals and consensus motifs existed, or a TF was inactive for the promoter within which its ChIP-seq signal and consensus motifs were not found. Parentheses in the "Consistency" columns of the "Pat.tissues" row show the number of the TFs that activated the promoter and were decreased in the patients' motor neurons and/or the whole spinal cords.
ChIP, ChIP-seq data of TFs; DB, the database ORTI for target genes of TFs; Motif, consensus motifs of TFs; Pat. tissues, significantly decreased TFs in tissues of patients with sporadic ALS; Put. links, putative molecular links between *ADARB1* regulation and FUS-linked ALS pathogenesis; SingleCor, correlation between a single TF and *ADARB1* expression; WGCNA, correlation between a TF in the module and *ADARB1* expression.
H, high; L, low; M, modest; mnTF, motoneuronal transcription factor. The number of the Results Section of each predicting approach is shown in brackets.

6. CONCLUSION

Based on the CAGE analysis on the laser-captured human MNs, the present study defined two out of the four promoters as active *ADARB1* promoters in the MNs, predicted the *ADARB1*-regulatory TFs in the MNs by *in silico* analyses on the multiple public data, and then surveyed their regulatory roles for each of the *ADARB1* promoters by *in vitro* luciferase assay for the first time [**Conclusion Figure**]. The present study lists the putative *ADARB1*-regulatory motoneuronal TFs and enables to further investigate the regulatory mechanism of *ADARB1* expression in the MNs. This knowledge is vital to uncover the abnormal molecular cascade underlying the pathogenic down-regulation of *ADARB1* in the MNs of patients with sporadic ALS. Hence, even though further studies should be conducted, I believe that the present study propels one step forward for understanding the whole image of the pathogenesis of this fatal disease.



Conclusion Figure The inferred regulatory mechanism of *ADARB1* expression in the human motor neurons (MNs).

Based on the present CAGE analysis on the human MNs and a number of the previous data, the present study defined the four *ADARB1* promoters and 1,334 TFs expressed in the MNs (the motoneuronal TFs). The present *in vitro* luciferase assay of the selected 34 motoneuronal TFs on the *ADARB1* promoters showed that the p2 and p1 promoters active in the MNs could be regulated by a different set of multiple motoneuronal TFs, respectively (two dotted ellipses within the red rectangle). Circles colored in red, blue, green and black denote the universal activators for all the four promoters, the universal suppressor for all the four promoters, the selective activators/suppressors and the inactive TF, respectively. Stars colored in red and green denote the TFs whose expression was decreased in tissues of patients with sporadic ALS. The 1,300 motoneuronal TFs remains unexamined by luciferase assay, and any regulatory effects from enhancers onto the promoter activity are beyond the scope of this study (outside the red rectangle).

7. ACKNOWLEDGEMENT

First, I would like to express my deepest gratitude to Dr. Shin Kwak at Department of Neurology (currently Kwak Laboratory), Graduate School of Medicine, The University of Tokyo. It is his perseverant and accurate supervision over nine years that has educated me who had had no idea even how to search research articles at first. I strongly believe that somebody has to tackle the theme of this study, because this theme is really challenging but definitely essential to understand the pathogenesis of sporadic ALS more deeply. I am proud of his decision to let me struggle with such an important theme.

I am deeply grateful to Dr. Takenari Yamashita at Kwak Laboratory not only for his invitation to be a member of Kwak Laboratory when I was a second-year undergraduate student at The University of Tokyo, but also for his huge and valuable advice on my *in vitro* experimental procedures. I also appreciate Ms. Sayaka Teramoto for her great and accurate assists in the *in vitro* experiments. Besides, I would like to extend my appreciation for all the other members at Kwak Laboratory, Mr. Takashi Hosaka, Dr. Megumi Akamatsu and Ms. Keiko Izumi, for their supports with the *in vitro* experiments and encouragement.

Also I really appreciate Dr. Yoshihide Hayashizaki and Dr. Hideya Kawaji at RIKEN Preventive Medicine and Diagnosis Innovation Program and Dr. Yasuhiro Murakawa at RIKEN Innovation Center for their frequent key advice on my research hurdles.

I would like to thank Dr. Takeshi Iwatsubo and Dr. Tadafumi Hashimoto at Department of Neuropathology, Graduate School of Medicine, The University of Tokyo for supervising me on activities as a graduate student and kindly providing me with research facilities.

I am also thankful Dr. Shoji Tsuji, the ex-professor at Department of Neurology, Graduate School of Medicine, The University of Tokyo, for kindly providing me with research facilities.

I appreciate all the laboratories producing the data that I used in this study.

Finally, here I would like to express my thankfulness to Dr. Susumu Tamai, Emeritus Professor at Nara Medical University, for suggesting me to become a human doctor and research on hand surgery when I was 16 years old. Although I consequently modified his suggestion from "human" to "animal" and from "hand surgery" to "neurology", I believe that he realized to where he should have connected me as he had known how to connect everything to successfully replant a completely amputated thumb for the first time in the world.

8. REFERENCES

1. Duttke S. H. C., Lacadie S. A., Ibrahim M. M., Glass C. K., Corcoran D. L., Benner C., Heinz S., Kadonaga J. T. and Ohler U. Human promoters are intrinsically directional. *Mol. Cell* **57,** 674–684 (2015).

2. Chen Y., Pai A. A., Herudek J., Lubas M., Meola N., Järvelin A. I., Andersson R., Pelechano V., Steinmetz L. M., Jensen T. H. and Sandelin A. Principles for RNA metabolism and alternative transcription initiation within closely spaced promoters. *Nat. Genet.* **48,** 984–994 (2016).

3. Andersson R., Chen Y., Core L., Lis J. T., Sandelin A. and Jensen T. H. Human gene promoters are intrinsically bidirectional. *Mol. Cell* **60,** 346–347 (2015).

4. Kadonaga J. T. Perspectives on the RNA polymerase II core promoter. *Wiley Interdiscip. Rev. Dev. Biol.* **1,** 40–51 (2012).

5. Danino Y. M., Even D., Ideses D. and Juven-Gershon T. The core promoter: At the heart of gene expression. *Biochim. Biophys. Acta* **1849,** 1116–1131 (2015).

6. Carninci P., Sandelin A., Lenhard B., Katayama S., Shimokawa K., Ponjavic J., Semple C. A., Taylor M. S., Engström P. G., Frith M. C., Forrest A. R., Alkema W. B., Tan S. L., Plessy C., Kodzius R., Ravasi T., Kasukawa T., Fukuda S., Kanamori-Katayama M. *et al.* Genome-wide analysis of mammalian promoter

architecture and evolution. *Nat. Genet.* **38,** 626–635 (2006).

7. Frith M. C. and the FANTOM Consortium. Explaining the correlations among properties of mammalian promoters. *Nucleic Acids Res.* **42,** 4823–32 (2014).

8. Andersson R., Refsing Andersen P., Valen E., Core L. J., Bornholdt J., Boyd M., Heick Jensen T. and Sandelin A. Nuclear stability and transcriptional directionality separate functionally distinct RNA species. *Nat. Commun.* **5,** 5336 (2014).

9. Core L. J., Martins A. L., Danko C. G., Waters C. T., Siepel A. and Lis J. T. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat. Genet.* **46,** 1311–1320 (2014).

10. Forrest A. R. R., Kawaji H., Rehli M., Kenneth Baillie J., de Hoon M. J. L., Haberle V., Lassmann T., Kulakovskiy I. V., Lizio M., Itoh M., Andersson R., Mungall C. J., Meehan T. F., Schmeier S., Bertin N., Jørgensen M., Dimont E., Arner E., Schmidl C. *et al.* A promoter-level mammalian expression atlas. *Nature* **507,** 462–470 (2014).

11. Hoskins R. A., Landolin J. M., Brown J. B., Sandler J. E., Takahashi H., Lassmann T., Yu C., Booth B. W., Zhang D., Wan K. H., Yang L., Boley N., Andrews J., Kaufman T. C., Graveley B. R., Bickel P. J., Carninci P., Carlson J. W. and Celniker S. E. Genome-wide analysis of promoter architecture in Drosophila melanogaster. *Genome Res.* **21,** 182–192 (2011).

12. Irizarry R. A., Ladd-Acosta C., Wen B., Wu Z., Montano C., Onyango P., Cui H., Gabo K., Rongione M., Webster M., Ji H., Potash J. B., Sabunciyan S. and Feinberg A. P. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.* **41,** 178–186 (2009).

13. Rauch T. A., Wu X., Zhong X., Riggs A. D. and Pfeifer G. P. A human B cell methylome at 100-base pair resolution. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 671–678 (2009).

14. Suzuki Y., Tsunoda T., Sese J., Taira H., Mizushima-Sugano J., Hata H., Ota T., Isogai T., Tanaka T., Nakamura Y., Suyama A., Sakaki Y., Morishita S., Okubo K. and Sugano S. Identification and characterization of the potential promoter regions of 1031 kinds of human genes. *Genome Res.* **11,** 677–684 (2001).

15. Scruggs B. S., Gilchrist D. A., Nechaev S., Muse G. W., Burkholder A., Fargo D. C. and Adelman K. Bidirectional transcription arises from two distinct hubs of transcription factor binding and active chromatin. *Mol. Cell* **58,** 1101–1112 (2015).

16. Trinklein N. T., Force Aldred S., Hartman S. J., Schroeder D. I., Otillar R. P. and Myers R. M. An abundance of bidirectional promoters in the human genome. *Genome Res.* **14,** 62–66 (2004).

17. Preker P., Nielsen J., Kammler S., Lykke-Andersen S., Christensen M., Mapendano C.,

Schierup M. and Jensen T. RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322,** 1851–1854 (2008).

18. Seila A. C., Calabrese J. M., Levine S. S., Yeo G. W., Rahl P. B., Flynn R. A., Young R. A. and Sharp P. A. Divergent transcription from active promoters. *Science* **322,** 1849–1851 (2008).

19. Core L. J., Waterfall J. J. and Lis J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322,** 1845–1848 (2008).

20. Ntini E., Järvelin A. I., Bornholdt J., Chen Y., Boyd M., Jørgensen M., Andersson R., Hoof I., Schein A., Andersen P. R., Andersen P. K., Preker P., Valen E., Zhao X., Pelechano V., Steinmetz L. M., Sandelin A. and Jensen T. H. Polyadenylation site–induced decay of upstream transcripts enforces promoter directionality. *Nat. Struct. Mol. Biol.* **20,** 923–928 (2013).

21. Almada A. E., Wu X., Kriz A. J., Burge C. B. and Sharp P. A. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* **499,** 360–363 (2013).

22. Andersson R., Gebhard C., Miguel-Escalada I., Hoof I., Bornholdt J., Boyd M., Chen Y., Zhao X., Schmidl C., Suzuki T., Ntini E., Arner E., Valen E., Li K., Schwarzfischer L., Glatz D., Raithel J., Lilje B., Rapin N. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507,** 455–461 (2014).

23. De Santa F., Barozzi I., Mietton F., Ghisletti S., Polletti S., Tusi B. K., Muller H., Ragoussis J., Wei C. L. and Natoli G. A large fraction of extragenic RNA Pol II transcription sites overlap enhancers. *PLoS Biol.* **8,** e1000384 (2010).

24. Yanez-Cuna J. O., Arnold C. D., Stampfel G., Boryn L. M., Gerlach D., Rath M. and Stark A. Dissection of thousands of cell type-specific enhancers identifies dinucleotide repeat motifs as general enhancer features. *Genome Res.* **24,** 1147–1156 (2014).

25. Fukaya T., Lim B. and Levine M. Enhancer control of transcriptional bursting. *Cell* **166,** 1–11 (2016).

26. Jin F., Li Y., Dixon J. R., Selvaraj S., Ye Z., Lee A. Y., Yen C.-A., Schmitt A. D., Espinoza C. a and Ren B. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503,** 290–294 (2013).

27. Banerji J., Rusconi S. and Schaffner W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27,** 299–308 (1981).

28. Kim T. K., Hemberg M., Gray J. M., Costa A. M., Bear D. M., Wu J., Harmin D. A., Laptewicz M., Barbara-Haley K., Kuersten S., Markenscoff-Papadimitriou E., Kuhl D., Bito H., Worley P. F., Kreiman G. and Greenberg M. E. Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465,** 182–187 (2010).

29. Darrow E. M. and Chadwick B. P. Boosting transcription by transcription:

enhancer-associated transcripts. *Chromosom. Res.* **21,** 713–724 (2013).

30. Arner E., Daub C. O., Vitting-Seerup K., Andersson R., Lilje B., Drablos F., Lennartsson A., Rönnerblad M., Vitezic M., Freeman T. C., Alhendi A., Arner P., Axton R., Baillie J. K., Beckhouse A., Bodega B., Brombacher F., Davis M., Detmar M. *et al.* Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science* **347,** 1010–1014 (2015).

31. Schoenfelder S., Furlan-magaril M., Mifsud B., Tavares-cadete F., Sugar R., Javierre B., Nagano T., Katsman Y., Sakthidevi M., Wingett S. W., Dimitrova E., Dimond A., Edelman L. B., Elderkin S., Tabbada K., Darbo E., Andrews S., Herman B., Higgs A. *et al.* The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res.* **25,** 1–16 (2015).

32. Joo J.-Y., Schaukowitch K., Farbiak L., Kilaru G. and Kim T.-K. Stimulus-specific combinatorial functionality of neuronal c-fos enhancers. *Nat. Neurosci.* **19,** 75–83 (2016).

33. Whyte W. A., Orlando D. A., Hnisz D., Abraham B. J., Lin C. Y., Kagey M. H., Rahl P. B., Lee T. I. and Young R. A. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153,** 307–319 (2013).

34. Hnisz D., Abraham B. J., Lee T. I., Lau A., Saint-André V., Sigova A. A., Hoke H. A.

and Young R. A. Super-enhancers in the control of cell identity and disease. *Cell* **155,** (2013).

35. Parker S. C. J., Stitzel M. L., Taylor D. L., Orozco J. M., Erdos M. R., Akiyama J. A., van Bueren K. L., Chines P. S., Narisu N., Black B. L., Visel A., Pennacchio L. A., Collins F. S., Becker J., Benjamin B., Blakesley R., Bouffard G., Brooks S., Coleman H. *et al.* Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 17921–17926 (2013).

36. Hnisz D., Shrinivas K., Young R. A., Chakraborty A. K. and Sharp P. A. A phase separation model for transcriptional control. *Cell* **169,** 13–23 (2017).

37. Pott S. and Lieb J. D. What are super-enhancers? *Nat. Genet.* **47,** 8–12 (2014).

38. Kundaje A., Meuleman W., Ernst J., Bilenky M., Yen A., Heravi-Moussavi A., Kheradpour P., Zhang Z., Wang J., Ziller M. J., Amin V., Whitaker J. W., Schultz M. D., Ward L. D., Sarkar A., Quon G., Sandstrom R. S., Eaton M. L., Wu Y.-C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518,** 317–330 (2015).

39. Dunham I., Kundaje A., Aldred S. F., Collins P. J., Davis C. A., Doyle F., Epstein C. B., Frietze S., Harrow J., Kaul R., Khatun J., Lajoie B. R., Landt S. G., Lee B.-K., Pauli F., Rosenbloom K. R., Sabo P., Safi A., Sanyal A. *et al.* An integrated

encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).

40. Wamstad J. A., Wang X., Demuren O. O. and Boyer L. A. Distal enhancers: new insights into heart development and disease. *Trends Cell Biol.* **24,** 294–302 (2013).

41. Fulco C. P., Munschauer M., Anyoha R., Munson G., Grossman S. R., Perez E. M., Kane M., Cleary B., Lander E. S. and Engreitz J. M. Systematic mapping of functional enhancer-promoter connections with CRIPR interference. *Science* **354,** 769–773 (2016).

42. Tehranchi A. K., Myrthil M., Martin T., Hie B. L., Golan D. and Fraser H. B. Pooled ChIP-seq links variation in transcription factor binding to complex disease risk. *Cell* **165,** 730–741 (2016).

43. Quinn J. J. and Chang H. Y. Unique features of long non-coding RNA biogenesis and function. *Nat. Rev. Genet.* **17,** 47–62 (2015).

44. Bose D. A., Donahue G., Reinberg D., Shiekhattar R., Bonasio R. and Berger S. L. RNA binding to CBP stimulates histone acetylation and transcription. *Cell* **168,** 135–149.e22 (2017).

45. Werner M. S., Sullivan M. A., Shah R. N., Nadadur R. D., Grzybowski A. T., Galat V., Moskowitz I. P. and Ruthenburg A. J. Chromatin-enriched lncRNAs can act as cell-type specific activators of proximal gene transcription. *Nat. Struct. Mol. Biol.* **24,**

596–603 (2017).

46. Hon C., Ramilowski J., Harshbarger J., Bertin N., Rackham O., Gough J., Denisenko E., Schmeier S., Poulsen T., Severin J., Lizio M., Kawaji H., Kasukawa T., Itoh M., Burroughs A., Noma S., Djebali S., Alam T., Medvedeva Y. *et al.* An atlas of human long non-coding RNAs with accurate 5‘ ends. *Nature* **543,** 199–204 (2017).

47. Vaquerizas J. M., Kummerfeld S. K., Teichmann S. A. and Luscombe N. M. A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.* **10,** 252–263 (2009).

48. Ravasi T., Suzuki H., Cannistraci C. V., Katayama S., Bajic V. B., Tan K., Akalin A., Schmeier S., Kanamori-Katayama M., Bertin N., Carninci P., Daub C. O., Forrest A. R. R., Gough J., Grimmond S., Han J. H., Hashimoto T., Hide W., Hofmann O. *et al.* An atlas of combinatorial transcriptional regulation in mouse and man. *Cell* **140,** 744–752 (2010).

49. Wingender E., Schoeps T. and Dönitz J. TFClass: an expandable hierarchical classification of human transcription factors. *Nucleic Acids Res.* **41,** D165–D170 (2013).

50. Ehsani R., Bahrami S. and Drabløs F. Feature-based classification of human transcription factors into hypothetical sub-classes related to regulatory function. *BMC*

Bioinformatics **17,** 459 (2016).

51. Weirauch M. T., Yang A., Albu M., Cote A. G., Montenegro-Montero A., Drewe P., Najafabadi H. S., Lambert S. A., Mann I., Cook K., Zheng H., Goity A., van Bakel H., Lozano J.-C., Galli M., Lewsey M. G., Huang E., Mukherjee T., Chen X. *et al.* Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158,** 1431–1443 (2014).

52. Nitta K. R., Jolma A., Yin Y., Morgunova E., Kivioja T., Akhtar J., Hens K., Toivonen J., Deplancke B., Furlong E. E. M. and Taipale J. Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. *Elife* **4,** e04837 (2015).

53. Yin Y., Morgunova E., Jolma A., Kaasinen E., Sahu B., Khund-Sayeed S., Das P. K., Kivioja T., Dave K., Zhong F., Nitta K. R., Taipale M., Popov A., Ginno P. A., Domcke S., Yan J., Schübeler D., Vinson C. and Taipale J. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356,** eaaj2239 (2017).

54. Levo M. and Segal E. In pursuit of design principles of regulatory sequences. *Nat. Rev. Genet.* **15,** 453–468 (2014).

55. Jolma A., Yin Y., Nitta K. R., Dave K., Popov A., Taipale M., Enge M., Kivioja T.,

Morgunova E. and Taipale J. DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature* **527,** 384–388 (2015).

56. Cusanovich D. A., Pavlovic B., Pritchard J. K. and Gilad Y. The functional consequences of variation in transcription factor binding. *PLoS Genet.* **10,** e1004226 (2014).

57. Pierson E., The GTEx Consortium, Koller D., Battle A. and Mostafavi S. Sharing and Specificity of Co-expression Networks across 35 Human Tissues. *PLOS Comput. Biol.* **11,** e1004220 (2015).

58. Guo Y., Alexander K., Clark A. G., Grimson A. and Yu H. Integrated network analysis reveals distinct regulatory roles of transcription factors and microRNAs. *RNA* **22,** 1663–1672 (2016).

59. Gerstein M. B., Kundaje A., Hariharan M., Landt S. G., Yan K.-K., Cheng C., Mu X. J., Khurana E., Rozowsky J., Alexander R., Min R., Alves P., Abyzov A., Addleman N., Bhardwaj N., Boyle A. P., Cayting P., Charos A., Chen D. Z. *et al.* Architecture of the human regulatory network derived from ENCODE data. *Nature* **489,** 91–100 (2012).

60. Stampfel G., Kazmar T., Frank O., Wienerroither S., Reiter F. and Stark A. Transcriptional regulators form diverse groups with context-dependent regulatory functions. *Nature* **528,** 147–151 (2015).

61. Morgunova E. and Taipale J. Structural perspective of cooperative transcription factor binding. *Curr. Opin. Struct. Biol.* **47,** 1–8 (2017).

62. Spitz F. and Furlong E. E. M. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* **13,** 613–626 (2012).

63. Spivakov M. Spurious transcription factor binding: Non-functional or genetically redundant? *BioEssays* **36,** 798–806 (2014).

64. Li J., Hua X., Haubrock M., Wang J. and Wingender E. The architecture of the gene regulatory networks of different tissues. *Bioinformatics* **28,** i509–i514 (2012).

65. Saint-André V., Federation A. J., Lin C. Y., Abraham B. J., Reddy J., Lee T. I., Bradner J. E. and Young R. A. Models of human core transcriptional regulatory circuitries. *Genome Res.* **26,** 385–396 (2016).

66. Zhang S., Tian D., Tran N. H., Choi K. P. and Zhang L. Profiling the transcription factor regulatory networks of human cell types. *Nucleic Acids Res.* **42,** 12380–12387 (2014).

67. Sherwood R. I., Hashimoto T., O'Donnell C. W., Lewis S., Barkal A. A., van Hoff J. P., Karun V., Jaakkola T. and Gifford D. K. Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nat. Biotechnol.* **32,** 171–178 (2014).

68. Cirillo L. A., Lin F. R., Cuesta I., Friedman D., Jarnik M. and Zaret K. S. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol. Cell* **9,** 279–289 (2002).

69. Iwafuchi-Doi M. and Zaret K. S. Cell fate control by pioneer transcription factors. *Development* **143,** 1833–1837 (2016).

70. Swinstead E. E., Miranda T. B., Paakinaho V., Baek S., Goldstein I., Hawkins M., Karpova T. S., Ball D., Mazza D., Lavis L. D., Grimm J. B., Morisaki T., Grontved L., Presman D. M. and Hager G. L. Steroid receptors reprogram FoxA1 occupancy through dynamic chromatin transitions. *Cell* **165,** 593–605 (2016).

71. Swinstead E. E., Paakinaho V., Presman D. M. and Hager G. L. Pioneer factors and ATP-dependent chromatin remodeling factors interact dynamically: A new perspective. *Bioessays* **38,** 1150–1157 (2016).

72. Barski A., Cuddapah S., Cui K., Roh T.-Y., Schones D. E., Wang Z., Wei G., Chepelev I. and Zhao K. High-resolution profiling of histone methylations in the human genome. *Cell* **129,** 823–837 (2007).

73. Heintzman N. D., Stuart R. K., Hon G., Fu Y., Ching C. W., Hawkins R. D., Barrera L. O., Van Calcar S., Qu C., Ching K. A., Wang W., Weng Z., Green R. D., Crawford G. E. and Ren B. Distinct and predictive chromatin signatures of transcriptional promoters

and enhancers in the human genome. *Nat. Genet.* **39,** 311–318 (2007).

74. Heintzman N. D., Hon G. C., Hawkins R. D., Kheradpour P., Stark A., Harp L. F., Ye Z., Lee L. K., Stuart R. K., Ching C. W., Ching K. A., Antosiewicz-Bourget J. E., Liu H., Zhang X., Green R. D., Lobanenkov V. V, Stewart R., Thomson J. A., Crawford G. E. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459,** 108–112 (2009).

75. Creyghton M. P., Cheng A. W., Welstead G. G., Kooistra T., Carey B. W., Steine E. J., Hanna J., Lodato M. A., Frampton G. M., Sharp P. A., Boyer L. A., Young R. A. and Jaenisch R. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 21931–21936 (2010).

76. Pekowska A., Benoukraf T., Zacarias-Cabeza J., Belhocine M., Koch F., Holota H., Imbert J., Andrau J.-C., Ferrier P. and Spicuglia S. H3K4 tri-methylation provides an epigenetic signature of active enhancers. *EMBO J.* **30,** 4198–4210 (2011).

77. Maunakea A. K., Nagarajan R. P., Bilenky M., Ballinger T. J., D'Souza C., Fouse S. D., Johnson B. E., Hong C., Nielsen C., Zhao Y., Turecki G., Delaney A., Varhol R., Thiessen N., Shchors K., Heine V. M., Rowitch D. H., Xing X., Fiore C. *et al.* Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **466,** 253–257 (2010).

78. Ball M. P., Billy Li J., Gao Y., Lee J.-H., LeProust E. M., Park I.-H., Xie B., Daley G. Q. and Church G. M. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat. Biotechnol.* **27,** 361–368 (2009).

79. Jones P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13,** 484–492 (2012).

80. Dermitzakis E. T. and Clark A. G. Evolution of transcription factor binding sites in mammalian gene regulatory regions: conservation and turnover. *Mol. Biol. Evol.* **19,** 1114–11121 (2002).

81. Odom D. T., Dowell R. D., Jacobsen E. S., Gordon W., Danford T. W., MacIsaac K. D., Rolfe P. A., Conboy C. M., Gifford D. K. and Fraenkel E. Tissue-specific transcriptional regulation has diverged significantly between human and mouse. *Nat. Genet.* **39,** 730–732 (2007).

82. Zeng J., Konopka G., Hunt B. G., Preuss T. M., Geschwind D. and Yi S. V. Divergent whole-genome methylation maps of human and chimpanzee brains reveal epigenetic basis of human regulatory evolution. *Am. J. Hum. Genet.* **91,** 455–465 (2012).

83. Zhang Y., Sloan S. A., Clarke L. E., Caneda C., Plaza C. A., Blumenthal P. D., Vogel H., Steinberg G. K., Edwards M. S. B., Li G., Duncan J. A., Cheshier S. H., Shuer L. M., Chang E. F., Grant G. A., Gephart M. G. H. and Barres B. A. Purification and

characterization of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. *Neuron* **89,** 37–53 (2016).

84. Shay T., Jojic V., Zuk O., Rothamel K., Puyraimond-Zemmour D., Feng T., Wakamatsu E., Benoist C., Koller D. and Regev A. Conservation and divergence in the transcriptional programs of the human and mouse immune systems. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 2946–2951 (2013).

85. Hawrylycz M., Miller J. A., Menon V., Feng D., Dolbeare T., Guillozet-Bongaarts A. L., Jegga A. G., Aronow B. J., Lee C.-K., Bernard A., Glasser M. F., Dierker D. L., Menche J., Szafer A., Collman F., Grange P., Berman K. A., Mihalas S., Yao Z. *et al.* Canonical genetic signatures of the adult human brain. *Nat. Neurosci.* **18,** 1832–1844 (2015).

86. Mele M., Ferreira P. G., Reverter F., DeLuca D. S., Monlong J., Sammeth M., Young T. R., Goldmann J. M., Pervouchine D. D., Sullivan T. J., Johnson R., Segre a. V., Djebali S., Niarchou A., The GTEx Consortium, Wright F. A., Lappalainen T., Calvo M., Getz G. *et al.* The human transcriptome across tissues and individuals. *Science* **348,** 660–665 (2015).

87. Zhu J., Chen G., Zhu S., Li S., Wen Z., Bin Li, Zheng Y. and Shi L. Identification of tissue-specific protein-coding and noncoding transcripts across 14 human tissues using

RNA-seq. *Sci. Rep.* **6,** 28400 (2016).

88. Fagerberg L., Hallstrom B. M., Oksvold P., Kampf C., Djureinovic D., Odeberg J., Habuka M., Tahmasebpoor S., Danielsson A., Edlund K., Asplund A., Sjostedt E., Lundberg E., Szigyarto C. A., Skogs M., Takanen J. O., Berling H., Tegel H., Mulder J. *et al.* Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteomics* **13,** 397–406 (2014).

89. Mills J. D., Kavanagh T., Kim W. S., Chen B. J., Kawahara Y., Halliday G. M. and Janitz M. Unique transcriptome patterns of the white and grey matter corroborate structural and functional heterogeneity in the human frontal lobe. *PLoS One* **8,** e78480 (2013).

90. Oldham M. C., Konopka G., Iwamoto K., Langfelder P., Kato T., Horvath S. and Geschwind D. H. Functional organization of the transcriptome in human brain. *Nat. Neurosci.* **11,** 1271–1282 (2008).

91. Taylor J. P., Brown Jr. R. H. and Cleveland D. W. Decoding ALS: from genes to mechanism. *Nature* **539,** 197–206 (2016).

92. Leblond C. S., Kaneb H. M., Dion P. A. and Rouleau G. A. Dissection of genetic factors associated with amyotrophic lateral sclerosis. *Exp. Neurol.* **262,** 91–101 (2014).

93. Manjaly Z. R., Scott K. M., Abhinav K., Wijesekera L., Ganesalingam J., Goldstein L. H., Janssen A., Dougherty A., Willey E., Stanton B. R., Turner M. R., Ampong M.-A., Sakel M., Orrell R. W., Howard R., Shaw C. E., Leigh P. N. and Al-Chalabi A. The sex ratio in amyotrophic lateral sclerosis: A population based study. *Amyotroph. Lateral Scler.* **11,** 439–442 (2010).

94. Mitchell J. and Borasio G. Amyotrophic lateral sclerosis. *Lancet* **369,** 2031–2041 (2007).

95. Chiò A., Logroscino G., Traynor B. J., Collins J., Simeone J. C., Goldstein L. A. and White L. A. Global epidemiology of amyotrophic lateral sclerosis: A systematic review of the published literature. *Neuroepidemiology* **41,** 118–130 (2013).

96. Rosen D., Siddique T., Patterson D., Figlewicz D., Sapp P., Hentati A., Donaldson D., Goto J., O'Regan J., Deng H., Rahmani Z., Krizus A., McKenna-Yasek D., Cayabyab A., Gaston S., Berger R., Tanzi R., Halperin J., Herzfeldt B. *et al.* Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* **362,** 59–62 (1993).

97. Kwiatkowski T. J., Bosco D., Leclerc A., Tamrazian E., Vanderburg C., Russ C., Davis A., Gilchrist J., Kasarskis E., Munsat T., Valdmanis P., Rouleau G., Hosler B., Cortelli P., de Jong P., Yoshinaga Y., Haines J., Pericak-Vance M., Yan J. *et al.* Mutations in

the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science* **323,** 1205–1208 (2009).

98. Vance C., Rogelj B., Hortobágyi T., De Vos K., Nishimura A., Sreedharan J., Hu X., Smith B., Ruddy D., Wright P., Ganesalingam J., Williams K., Tripathi V., Al-Saraj S., Al-Chalabi A., Leigh P., Blair I., Nicholson G., de Belleroche J. *et al.* Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science* **323,** 1208–1211 (2009).

99. Kabashi E., Valdmanis P. N., Dion P., Spiegelman D., McConkey B. J., Vande Velde C., Bouchard J.-P., Lacomblez L., Pochigaeva K., Salachas F., Pradat P.-F., Camu W., Meininger V., Dupre N. and Rouleau G. A. TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. *Nat. Genet.* **40,** 572–574 (2008).

100. Sreedharan J., Blair I. P., Tripathi V. B., Hu X., Vance C., Rogelj B., Ackerley S., Durnall J. C., Williams K. L., Buratti E., Baralle F., de Belleroche J., Mitchell J. D., Leigh P. N., Al-Chalabi A., Miller C. C., Nicholson G. and Shaw C. E. TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science* **319,** 1668–1672 (2008).

101. Maruyama H., Morino H., Ito H., Izumi Y., Kato H., Watanabe Y., Kinoshita Y., Kamada M., Nodera H., Suzuki H., Komure O., Matsuura S., Kobatake K., Morimoto

N., Abe K., Suzuki N., Aoki M., Kawata A., Hirai T. *et al.* Mutations of optineurin in amyotrophic lateral sclerosis. *Nature* **465,** 223–226 (2010).

102. Johnson J. O., Mandrioli J., Benatar M., Abramzon Y., Van Deerlin V. M., Trojanowski J. Q., Gibbs J. R., Brunetti M., Gronka S., Wuu J., Ding J., McCluskey L., Martinez-Lage M., Falcone D., Hernandez D. G., Arepalli S., Chong S., Schymick J. C., Rothstein J. *et al.* Exome sequencing reveals VCP mutations as a cause of familial ALS. *Neuron* **68,** 857–864 (2010).

103. Deng H.-X., Chen W., Hong S.-T., Boycott K. M., Gorrie G. H., Siddique N., Yang Y., Fecto F., Shi Y., Zhai H., Jiang H., Hirano M., Rampersaud E., Jansen G. H., Donkervoort S., Bigio E. H., Brooks B. R., Ajroud K., Sufit R. L. *et al.* Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia. *Nature* **477,** 211–215 (2011).

104. Wu C.-H., Fallini C., Ticozzi N., Keagle P. J., Sapp P. C., Piotrowska K., Lowe P., Koppers M., McKenna-Yasek D., Baron D. M., Kost J. E., Gonzalez-Perez P., Fox A. D., Adams J., Taroni F., Tiloca C., Leclerc A. L., Chafe S. C., Mangroo D. *et al.* Mutations in the profilin 1 gene cause familial amyotrophic lateral sclerosis. *Nature* **488,** 499–503 (2012).

105. Kim H. J., Kim N. C., Wang Y.-D., Scarborough E. A., Moore J., Diaz Z., MacLea K.

S., Freibaum B., Li S., Molliex A., Kanagaraj A. P., Carter R., Boylan K. B., Wojtas A. M., Rademakers R., Pinkus J. L., Greenberg S. A., Trojanowski J. Q., Traynor B. J. *et al.* Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* **495,** 467–473 (2013).

106. Renton A. E., Majounie E., Waite A., Simón-Sánchez J., Rollinson S., Gibbs J. R., Schymick J. C., Laaksovirta H., van Swieten J. C., Myllykangas L., Kalimo H., Paetau A., Abramzon Y., Remes A. M., Kaganovich A., Scholz S. W., Duckworth J., Ding J., Harmer D. W. *et al.* A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72,** 257–268 (2011).

107. DeJesus-Hernandez M., Mackenzie I. R., Boeve B. F., Boxer A. L., Baker M., Rutherford N. J., Nicholson A. M., Finch N. C. A., Flynn H., Adamson J., Kouri N., Wojtas A., Sengdy P., Hsiung G. Y. R., Karydas A., Seeley W. W., Josephs K. A., Coppola G., Geschwind D. H. *et al.* Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72,** 245–256 (2011).

108. Williams K. L., Topp S., Yang S., Smith B., Fifta J. A., Warraich S. T., Zhang K. Y., Farrawell N., Vance C., Hu X., Chesi A., Leblond C. S., Lee A., Rayner S. L., Sundaramoorthy V., Dobson-Stone C., Molloy M. P., van Blitterswijk M., Dickson D.

W. *et al.* CCNF mutations in amyotrophic lateral sclerosis and frontotemporal dementia. *Nat. Commun.* **7,** 11253 (2016).

109. Shaw C. E., Enayat Z. E., Chioza B. A., Al-Chalabi A., Radunovic A., Powell J. F. and Leigh P. N. Mutations in all five exons of SOD-1 may cause ALS. *Ann. Neurol.* **43,** 390–394 (1998).

110. Aizawa H., Hideyama T., Yamashita T., Kimura T., Suzuki N., Aoki M. and Kwak S. Deficient RNA-editing enzyme ADAR2 in an amyotrophic lateral sclerosis patient with a FUS P525L mutation. *J. Clin. Neurosci.* **32,** 128–129 (2016).

111. Rademakers R. and van Blitterswijk M. Motor neuron disease in 2012: Novel causal genes and disease modifiers. *Nat. Rev. Neurol.* **9,** 63–64 (2013).

112. Majounie E., Renton A. E., Mok K., Dopper E. G. P., Waite A., Rollinson S., Chiò A., Restagno G., Nicolaou N., Simon-Sanchez J., van Swieten J. C., Abramzon Y., Johnson J. O., Sendtner M., Pamphlett R., Orrell R. W., Mead S., Sidle K. C., Houlden H. *et al.* Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: A cross-sectional study. *Lancet Neurol.* **11,** 323–330 (2012).

113. Nakamura R., Sone J., Atsuta N., Tohnai G., Watanabe H., Yokoi D., Nakatochi M., Watanabe H., Ito M., Senda J., Katsuno M., Tanaka F., Li Y., Izumi Y., Morita M.,

Taniguchi A., Kano O., Oda M., Kuwabara S. *et al.* Next-generation sequencing of 28 ALS-related genes in a Japanese ALS cohort. *Neurobiol. Aging* **39,** 219.e1-219.e8 (2016).

114. Ikenaka K., Kawai K., Katsuno M., Huang Z., Jiang Y.-M., Iguchi Y., Kobayashi K., Kimata T., Waza M., Tanaka F., Mori I. and Sobue G. Dnc-1/dynactin 1 knockdown disrupts transport of autophagosomes and induces motor neuron degeneration. *PLoS One* **8,** e54511 (2013).

115. Hirano A., Donnenfeld H., Sasaki S. and Nakano I. Fine structural observations of neurofilamentous changes in amyotrophic lateral sclerosis. *J. Neuropathol. Exp. Neurol.* **43,** 461–70 (1984).

116. Rothstein J., Martin L. and Kuncl R. Decreased glutamate transport by the brain and spinal cord in amyotrophic lateral sclerosis. *N. Engl. J. Med.* **326,** 1464–1468 (1992).

117. Bowling A., Schulz J., Brown, Jr. R. and Beal M. Superoxide dismutase activity, oxidative damage, and mitochondrial energy metabolism in familial and sporadic amyotrophic lateral sclerosis. *J. Neurochem.* **61,** 2322–2325 (1993).

118. Sasabe J., Chiba T., Yamada M., Okamoto K., Nishimoto I., Matsuoka M. and Aiso S. D-serine is a key determinant of glutamate toxicity in amyotrophic lateral sclerosis. *EMBO J.* **26,** 4149–59 (2007).

119. Hideyama T. and Kwak S. When does ALS start? ADAR2-GluA2 hypothesis for the etiology of sporadic ALS. *Front. Mol. Neurosci.* **4,** 33 (2011).

120. Hideyama T., Yamashita T., Aizawa H., Tsuji S., Kakita A., Takahashi H. and Kwak S. Profound downregulation of the RNA editing enzyme ADAR2 in ALS spinal motor neurons. *Neurobiol. Dis.* **45,** 1121–1128 (2012).

121. Kawahara Y., Ito K., Sun H., Aizawa H., Kanazawa I. and Kwak S. RNA editing and death of motor neurons. *Nature* **427,** 801 (2004).

122. Feldmeyer D., Kask K., Brusa R., Kornau H. C., Kolhekar R., Rozov A., Burnashev N., Jensen V., Hvalby O., Sprengel R. and Seeburg P. H. Neurological dysfunctions in mice expressing different levels of the Q/R site-unedited AMPAR subunit GluR-B. *Nat. Neurosci.* **2,** 57–64 (1999).

123. Brusa R., Zimmermann F., Koh D.-S., Feldmeyer D., Gass P., Seeburg P. H. and Sprengel R. Early-onset epilepsy and postnatal lethality associated with an editing-deficient GluR-B allele in mice. *Science* **270,** 1677–80 (1995).

124. Higuchi M., Maas S., Single F. N., Hartner J., Rozov A., Burnashev N., Feldmeyer D., Sprengel R. and Seeburg P. H. Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2. *Nature* **406,** 78–81 (2000).

125. Sasaki S. Autophagy in spinal cord motor neurons in sporadic amyotrophic lateral sclerosis. *J. Neuropathol. Exp. Neurol.* **70,** 349–359 (2011).

126. Yamashita T., Hideyama T., Hachiga K., Teramoto S., Takano J., Iwata N., Saido T. C. and Kwak S. A role for calpain-dependent cleavage of TDP-43 in amyotrophic lateral sclerosis pathology. *Nat. Commun.* **3,** 1307 (2012).

127. Yamashita T., Aizawa H., Teramoto S., Akamatsu M. and Kwak S. Calpain-dependent disruption of nucleo-cytoplasmic transport in ALS motor neurons. *Sci. Rep.* **7,** 39994 (2017).

128. Arai T., Hasegawa M., Akiyama H., Ikeda K., Nonaka T., Mori H., Mann D., Tsuchiya K., Yoshida M., Hashizume Y. and Oda T. TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Biochem. Biophys. Res. Commun.* **351,** 602–11 (2006).

129. Neumann M., Sampathu D. M., Kwong L. K., Truax A. C., Micsenyi M. C., Chou T. T., Bruce J., Schuck T., Grossman M., Clark C. M., McCluskey L. F., Miller B. L., Masliah E., Mackenzie I. R., Feldman H., Feiden W., Kretzschmar H. A., Trojanowski J. Q. and Lee V. M.-Y. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314,** 130–133 (2006).

130. Kinoshita Y., Ito H., Hirano A., Fujita K., Wate R., Nakamura M., Kaneko S., Nakano

S. and Kusaka H. Nuclear contour irregularity and abnormal transporter protein distribution in anterior horn cells in amyotrophic lateral sclerosis. *J. Neuropathol. Exp. Neurol.* **68,** 1184–1192 (2009).

131. Aizawa H., Sawada J., Hideyama T., Yamashita T., Katayama T., Hasebe N., Kimura T., Yahara O. and Kwak S. TDP-43 pathology in sporadic ALS occurs in motor neurons lacking the RNA editing enzyme ADAR2. *Acta Neuropathol.* **120,** 75–84 (2010).

132. Hideyama T., Yamashita T., Suzuki T., Tsuji S., Higuchi M., Seeburg P. H., Takahashi R., Misawa H. and Kwak S. Induced loss of ADAR2 engenders slow death of motor neurons from Q/R site-unedited GluR2. *J. Neurosci.* **30,** 11917–25 (2010).

133. Sasaki S., Yamashita T. and Shin K. Autophagy in spinal motor neurons of conditional ADAR2-knockout mice: An implication for a role of calcium in increased autophagy flux in ALS. *Neurosci. Lett.* **598,** 79–84 (2015).

134. Moore S., Mendez E., Starr A., Lorenzini I., Nelson A., Ghaffari L., Levy J., Burciu C., Chew J., Belzil V., Robertson J., Contente-Cuomo T., Alsop E., Petrucelli L., Murtaza M., Jensen K., Kalb R. and Sattler R. Nucleo-cytoplasmic mislocalization of ADAR2 in C9orf72-ALS and FTD. *Abstr. Soc. Neurosci. 2017* **760.14,** (2017).

135. Slavov D. and Gardiner K. Phylogenetic comparison of the pre-mRNA adenosine

deaminase ADAR2 genes and transcripts: conservation and diversity in editing site sequence and alternative splicing patterns. *Gene* **299,** 83–94 (2002).

136. Kawahara Y., Ito K., Ito M., Tsuji S. and Kwak S. Novel splice variants of human ADAR2 mRNA: Skipping of the exon encoding the dsRNA-binding domains, and multiple C-terminal splice sites. *Gene* **363,** 193–201 (2005).

137. Mittaz L., Scott H. S., Rossier C., Seeburg P. H., Higuchi M. and Antonarakis S. E. Cloning of a human RNA editing deaminase (ADARB1) of glutamate receptors that maps to chromosome 21q22.3. *Genomics* **41,** 210–217 (1997).

138. Maas S. and Gommans W. M. Novel exon of mammalian ADAR2 extends open reading frame. *PLoS One* **4,** e4225 (2009).

139. Rueter S. M., Dawson T. R. and Emeson R. B. Regulation of alternative splicing by RNA editing. *Nature* **399,** 75–80 (1999).

140. Lai F., Chen C. X., Carter K. C. and Nishikura K. Editing of glutamate receptor B subunit ion channel RNAs by four alternatively spliced DRADA2 double-stranded RNA adenosine deaminases. *Mol. Cell. Biol.* **17,** 2413–2424 (1997).

141. Gerber A., O'Connell M. and Keller W. Two forms of human double-stranded RNA-specific editase 1 (hRED1) generated by the insertion of an Alu cassette. *RNA* **3,** 453–463 (1997).

142. Agranat L., Sperling J. and Sperling R. A novel tissue-specific alternatively spliced form of the A-to-I RNA editing enzyme ADAR2. *RNA Biol.* **7,** 253–262 (2010).

143. Fu Y., Zhao X., Li Z., Wei J. and Tian Y. Splicing variants of ADAR2 and ADAR2-mediated RNA editing in glioma (Review). *Oncol. Lett.* **12,** 788–792 (2016).

144. Shiraki T., Kondo S., Katayama S., Waki K., Kasukawa T., Kawaji H., Kodzius R., Watahiki A., Nakamura M., Arakawa T., Fukuda S., Sasaki D., Podhajska A., Harbers M., Kawai J., Carninci P. and Hayashizaki Y. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 15776–15781 (2003).

145. Hideyama T., Teramoto S., Hachiga K., Yamashita T. and Kwak S. Co-occurrence of TDP-43 mislocalization with reduced activity of an RNA editing enzyme, ADAR2, in aged mouse motor neurons. *PLoS One* **7,** e43469 (2012).

146. Nicholas A., de Magalhaes J. P., Kraytsberg Y., Richfield E. K., Levanon E. Y. and Khrapko K. Age-related gene-specific changes of A-to-I mRNA editing in the human brain. *Mech. Ageing Dev.* **131,** 445–447 (2010).

147. Hwang T., Park C.-K., Leung A. K. L., Gao Y., Hyde T. M., Kleinman J. E., Rajpurohit A., Tao R., Shin J. H. and Weinberger D. R. Dynamic regulation of RNA editing in human brain development and disease. *Nat. Neurosci.* **19,** 1093–1099 (2016).

148. Tomso D. J., Inga A., Menendez D., Pittman G. S., Campbell M. R., Storici F., Bell D. A. and Resnick M. A. Functionally distinct polymorphic sequences in the human genome that are targets for p53 transactivation. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 6431–6436 (2005).

149. Sawada J., Yamashita T., Aizawa H., Aburakawa Y., Hasebe N. and Kwak S. Effects of antidepressants on GluR2 Q/R site-RNA editing in modified HeLa cell line. *Neurosci. Res.* **64,** 251–258 (2009).

150. Yang L., Huang P., Li F., Zhao L., Zhang Y., Li S., Gan Z., Lin A., Li W. and Liu Y. c-Jun amino-terminal kinase-1 mediates glucose-responsive upregulation of the RNA editing enzyme ADAR2 in pancreatic beta-cells. *PLoS One* **7,** e48611 (2012).

151. Peng P. L., Zhong X., Tu W., Soundarapandian M. M., Molner P., Zhu D., Lau L., Liu S., Liu F. and Lu Y. ADAR2-dependent RNA editing of AMPA receptor subunit GluR2 determines vulnerability of neurons in forebrain ischemia. *Neuron* **49,** 719–33 (2006).

152. Bonini D., Filippini A., La Via L., Fiorentini C., Fumagalli F., Colombi M. and Barbon A. Chronic glutamate treatment selectively modulates AMPA RNA editing and ADAR expression and activity in primary cortical neurons. *RNA Biol.* **12,** 43–53 (2015).

153. Yamashita T., Chai H. L., Teramoto S., Tsuji S., Shimazaki K., Muramatsu S. and

Kwak S. Rescue of amyotrophic lateral sclerosis phenotype in a mouse model by intravenous AAV9-ADAR2 delivery to motor neurons. *EMBO Mol. Med.* **5,** 1710–1719 (2013).

154. Akamatsu M., Yamashita T., Hirose N., Teramoto S. and Kwak S. The AMPA receptor antagonist perampanel robustly rescues amyotrophic lateral sclerosis (ALS) pathology in sporadic ALS model mice. *Sci. Rep.* **6,** 28649 (2016).

155. R Core Team. R: A language and environment for statistical computing. (2015).

156. RStudio Team. RStudio: Integrated Development for R. (2015).

157. Washietl S., Findeiss S., Muller S. A., Kalkhof S., von Bergen M., Hofacker I. L., Stadler P. F. and Goldman N. RNAcode: robust discrimination of coding and noncoding regions in comparative sequence data. *RNA* **17,** 578–594 (2011).

158. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348,** 648–660 (2015).

159. Li H. and Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).

160. Kim D., Pertea G., Trapnell C., Pimentel H., Kelley R. and Salzberg S. L. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14,** R36 (2013).

161. Haberle V., Forrest A. R. R., Hayashizaki Y., Carninci P. and Lenhard B. CAGEr: precise TSS data retrieval and high-resolution promoterome mining for integrative analyses. *Nucleic Acids Res.* **43,** e51 (2015).

162. Quinlan A. R. and Hall I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).

163. Prudencio M., Belzil V. V, Batra R., Ross C. A., Gendron T. F., Pregent L. J., Murray M. E., Overstreet K. K., Piazza-Johnston A. E., Desaro P., Bieniek K. F., DeTure M., Lee W. C., Biendarra S. M., Davis M. D., Baker M. C., Perkerson R. B., van Blitterswijk M., Stetler C. T. *et al.* Distinct brain transcriptome profiles in C9orf72-associated and sporadic ALS. *Nat. Neurosci.* **18,** 1175–1182 (2015).

164. Magistri M., Velmeshev D., Makhmutova M. and Faghihi M. A. Transcriptomics profiling of Alzheimer's disease reveal neurovascular defects, altered Amyloid-β homeostasis, and deregulated expression of long noncoding RNAs. *J. Alzheimer's Dis.* **48,** 647–665 (2015).

165. Lin L., Park J. W., Ramachandran S., Zhang Y., Tseng Y.-T., Shen S., Waldvogel H. J., Curtis M. A., Faull R. L. M., Troncoso J. C., Ross C. A., Davidson B. L. and Xing Y. Transcriptome sequencing reveals aberrant alternative splicing in Huntington's disease HMG Advance Access. *Hum. Mol. Genet.* **25,** 3454–3466 (2016).

166. Batra R., Hutt K., Vu A., Rabin S. J., Baughn M. W., Libby R. T., Hoon S., Ravits J. and Yeo G. W. Gene expression signatures of sporadic ALS motor neuron populations. *bioRxiv* (2016).

167. Zhao S., Zhang Y., Gordon W., Quan J., Xi H., Du S., von Schack D. and Zhang B. Comparison of stranded and non-stranded RNA-seq transcriptome profiling and investigation of gene overlap. *BMC Genomics* **16,** 675 (2015).

168. Kapeli K., Pratt G. A., Vu A. Q., Hutt K. R., Martinez F. J., Sundararaman B., Batra R., Freese P., Lambert N. J., Huelga S. C., Chun S. J., Liang T. Y., Chang J., Donohue J. P., Shiue L., Zhang J., Zhu H., Cambi F., Kasarskis E. *et al.* Distinct and shared functions of ALS-associated proteins TDP-43, FUS and TAF15 revealed by multisystem analyses. *Nat. Commun.* **7,** 12143 (2016).

169. Bolger A. M., Lohse M. and Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120 (2014).

170. Dobin A., Davis C. A., Schlesinger F., Drenkow J., Zaleski C., Jha S., Batut P., Chaisson M. and Gingeras T. R. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).

171. Harrow J., Frankish A., Gonzalez J. M., Tapanari E., Diekhans M., Kokocinski F., Aken B. L., Barrell D., Zadissa A., Searle S., Barnes I., Bignell A., Boychenko V.,

Hunt T., Kay M., Mukherjee G., Rajan J., Despacio-Reyes G., Saunders G. *et al.* GENCODE: The reference human genome annotation for The ENCODE Project. *Genome Res.* **22,** 1760–1774 (2012).

172. Pertea M., Pertea G. M., Antonescu C. M., Chang T.-C., Mendell J. T. and Salzberg S. L. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33,** 290–295 (2015).

173. Frazee A. C., Pertea G., Jaffe A. E., Langmead B., Salzberg S. L. and Leek J. T. Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nat. Biotechnol.* **33,** 243–246 (2015).

174. Pertea M., Kim D., Pertea G. M., Leek J. T. and Salzberg S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11,** 1650–1667 (2016).

175. Hebenstreit D., Fang M., Gu M., Charoensawan V., van Oudenaarden A. and Teichmann S. A. RNA sequencing reveals two major classes of gene expression levels in metazoan cells. *Mol. Syst. Biol.* **7,** 497–497 (2014).

176. Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G. and Durbin R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

177. Ramírez F., Dündar F., Diehl S., Grüning B. A. and Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42,** 187–191 (2014).

178. Noguchi S., Arakawa T., Fukuda S., Furuno M., Hasegawa A., Hori F., Ishikawa-Kato S., Kaida K., Kaiho A., Kanamori-Katayama M., Kawashima T., Kojima M., Kubosaki A., Manabe R., Murata M., Nagao-Sato S., Nakazato K., Ninomiya N., Nishiyori-Sueki H. *et al.* FANTOM5 CAGE profiles of human and mouse samples. *Sci. Data* **4,** 170112 (2017).

179. Revelle W. psych: Procedures for Personality and Psychological Research. (2017).

180. Vafaee F., Krycer J. R., Ma X., Burykin T., James D. E. and Kuncic Z. ORTI: an open-access repository of transcriptional interactions for interrogating mammalian gene expression data. *PLoS One* **11,** 1–21 (2016).

181. Winden K. D., Oldham M. C., Mirnics K., Ebert P. J., Swan C. H., Levitt P., Rubenstein J. L., Horvath S. and Geschwind D. H. The organization of the transcriptional network in specific neuronal classes. *Mol. Syst. Biol.* **5,** 291 (2009).

182. Konopka G., Friedrich T., Davis-Turak J., Winden K., Oldham M. C., Gao F., Chen L., Wang G. Z., Luo R., Preuss T. M. and Geschwind D. H. Human-specific transcriptional networks in the brain. *Neuron* **75,** 601–617 (2012).

183. Ghazalpour A., Doss S., Zhang B., Wang S., Plaisier C., Castellanos R., Brozell A.,

Schadt E. E., Drake T. A., Lusis A. J. and Horvath S. Integrating genetic and network analysis to characterize genes related to mouse weight. *PLoS Genet.* **2,** e130 (2006).

184. Stacklies W., Redestig H., Scholz M., Walther D. and Selbig J. pcaMethods - a bioconductor package providing PCA methods for incomplete data. *Bioinformatics* **23,** 1164–1167 (2007).

185. Leek J. T., Johnson E., Parker H. S., Fertig E. J., Jaffe A. E. and Storey J. D. sva: Surrogate variable analysis.

186. Langfelder P. and Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9,** 559 (2008).

187. Warnes G. R., Bolker B., Bonebakker L., Gentleman R., Huber W., Liaw A., Lumley T., Maechler M., Magnusson A., Moeller S., Schwartz M. and Venables B. gplots: Various R programming tools for plotting data. (2016).

188. ChIP-Atlas. Available at: http://chip-atlas.org/.

189. Wang Z., Martins A. L. and Danko C. G. RTFBSDB: an integrated framework for transcription factor binding site analysis. *Bioinformatics* **32,** 3024–3026 (2016).

190. Mathelier A., Fornes O., Arenillas D. J., Chen C. Y., Denay G., Lee J., Shi W., Shyr C., Tan G., Worsley-Hunt R., Zhang A. W., Parcy F., Lenhard B., Sandelin A. and Wasserman W. W. JASPAR 2016: A major expansion and update of the open-access

database of transcription factor binding profiles. *Nucleic Acids Res.* **44,** D110–D115 (2016).

191. Tan G. and Lenhard B. TFBSTools: An R/bioconductor package for transcription factor binding site analysis. *Bioinformatics* **32,** 1555–1556 (2016).

192. Rabin S. J., Hugo Kim J. M., Baughn M., Libby R. T., Kim Y. J., Fan Y., Libby R. T., La Spada A., Stone B. and Ravits J. Sporadic ALS has compartment-specific aberrant exon splicing and altered cell-matrix adhesion biology. *Hum. Mol. Genet.* **19,** 313–328 (2009).

193. D'Erchia A. M., Gallo A., Manzari C., Raho S., Horner D. S., Chiara M., Valletti A., Aiello I., Mastropasqua F., Ciaccia L., Locatelli F., Pisani F., Nicchia G. P., Svelto M., Pesole G. and Picardi E. Massive transcriptome sequencing of human spinal cord tissues provides new insights into motor neuron degeneration in ALS. *Sci. Rep.* **7,** 10046 (2017).

194. Aronica E., Baas F., Iyer A., ten Asbroek A. L. M. A., Morello G. and Cavallaro S. Molecular classification of amyotrophic lateral sclerosis by unsupervised clustering of gene expression in motor cortex. *Neurobiol. Dis.* **74,** 359–376 (2015).

195. Lederer C. W., Torrisi A., Pantelidou M., Santama N. and Cavallaro S. Pathways and genes differentially expressed in the motor cortex of patients with sporadic

amyotrophic lateral sclerosis. *BMC Genomics* **8,** 26 (2007).

196. Nakaya T., Alexiou P., Maragkakis M., Chang A. and Mourelatos Z. FUS regulates genes coding for RNA-binding proteins in neurons by binding to their highly conserved introns. *RNA* **19,** 498–509 (2013).

197. Zhou Y., Liu S., Liu G., Öztürk A. and Hicks G. G. ALS-associated FUS mutations result in compromised FUS alternative splicing and autoregulation. *PLoS Genet.* **9,** e1003895 (2013).

198. Hoell J. I., Larsson E., Runge S., Nusbaum J. D., Duggimpudi S., Farazi T. a, Hafner M., Borkhardt A., Sander C. and Tuschl T. RNA targets of wild-type and mutant FET family proteins. *Nat. Struct. Mol. Biol.* **18,** 1428–1431 (2011).

199. Sun S., Ling S.-C., Qiu J., Albuquerque C. P., Zhou Y., Tokunaga S., Li H., Qiu H., Bui A., Yeo G. W., Huang E. J., Eggan K., Zhou H., Fu X.-D., Lagier-Tourenne C. and Cleveland D. W. ALS-causative mutations in FUS/TLS confer gain and loss of function by altered association with SMN and U1-snRNP. *Nat. Commun.* **6,** 6171 (2015).

200. Yoshino H., Enokida H., Itesako T., Tatarano S., Kinoshita T., Fuse M., Kojima S., Nakagawa M. and Seki N. Epithelial–mesenchymal transition-related microRNA-200s regulate molecular targets and pathways in renal cell carcinoma. *J. Hum. Genet.* **58,**

508–516 (2013).

201. Park S. M., Gaur A. B., Lengyel E. and Peter M. E. The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Genes Dev.* **22,** 894–907 (2008).

202. Agarwal V., Bell G. W., Nam J.-W. and Bartel D. P. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* **4,** (2015).

203. Schmidl C., Hansmann L., Lassmann T., Balwierz P. J., Kawaji H., Itoh M., Kawai J., Nagao-Sato S., Suzuki H., Andreesen R., Hayashizaki Y., Forrest A. R. R., Carninci P., Hoffmann P., Edinger M. and Rehli M. The enhancer and promoter landscape of human regulatory and conventional T-cell subpopulations. *Blood* **123,** e68-78 (2014).

204. Szklarczyk D., Morris J. H., Cook H., Kuhn M., Wyder S., Simonovic M., Santos A., Doncheva N. T., Roth A., Bork P., Jensen L. J. and von Mering C. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* **45,** D362–D368 (2017).

205. Shannon P., Markiel A., Owen Ozier 2, Baliga N. S., Wang J. T., Ramage D., Amin N., Schwikowski B. and Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13,** 2498–2504 (2003).

206. Kenna K. P., van Doormaal P. T. C., Dekker A. M., Ticozzi N., Kenna B. J., Diekstra F.

P., van Rheenen W., van Eijk K. R., Jones A. R., Keagle P., Shatunov A., Sproviero W., Smith B. N., van Es M. A., Topp S. D., Kenna A., Miller J. W., Fallini C., Tiloca C. *et al.* NEK1 variants confer susceptibility to amyotrophic lateral sclerosis. *Nat. Genet.* **48,** 1037–1042 (2016).

207. Cirulli E. T., Lasseigne B. N., Petrovski S., Sapp P. C., Dion P. A., Leblond C. S., Couthouis J., Lu Y., Wang Q., Krueger B. J., Ren Z., Keebler J., Han Y., Levy S. E., Boone B. E., Wimbish J. R., Waite L. L., Jones A. L., Carulli J. P. *et al.* Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science* **347,** 1436–1441 (2015).

208. Lill C. M., Abel O., Bertram L. and Al-Chalabi A. Keeping up with genetic discoveries in amyotrophic lateral sclerosis: the ALSoD and ALSGene databases. *Amyotroph. Lateral Scler.* **12,** 238–249 (2011).

209. Emmert-Buck M. R., Bonner R. F., Smith P. D., Chuaqui R. F., Zhuang Z., Goldstein S. R., Weiss R. A. and Liotta L. A. Laser capture microdissection. *Science* **274,** 998–1001 (1996).

210. Suzuki H., Forrest A. R. R., van Nimwegen E., Daub C. O., Balwierz P. J., Irvine K. M., Lassmann T., Ravasi T., Hasegawa Y., de Hoon M. J. L., Katayama S., Schroder K., Carninci P., Tomaru Y., Kanamori-Katayama M., Kubosaki A., Akalin A., Ando Y.,

Arner E. *et al.* The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nat. Genet.* **41,** 553–562 (2009).

211. Kundaje A., Kyriazopoulou-panagiotopoulou S., Libbrecht M., Smith C. L., Raha D., Winters E. E., Johnson S. M., Snyder M., Batzoglou S. and Sidow A. Ubiquitous heterogeneity and asymmetry of the chromatin environment at regulatory elements. *Genome Res.* **22,** 1735–1747 (2012).

212. Hoffmann P., Eder R., Kunz-schughart L. A., Andreesen R. and Edinger M. Large-scale in vitro expansion of polyclonal human CD4+ CD25 high regulatory T cells. *Blood* **104,** 895–903 (2004).

213. Ringner M. What is principal component analysis? *Nat. Biotechnol.* **26,** 303–304 (2008).

214. Lagier-Tourenne C., Polymenidou M., Hutt K. R., Vu A. Q., Baughn M., Huelga S. C., Clutario K. M., Ling S.-C., Liang T. Y., Mazur C., Wancewicz E., Kim A. S., Watt A., Freier S., Hicks G. G., Donohue J. P., Shiue L., Bennett C. F., Ravits J. *et al.* Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. *Nat. Neurosci.* **15,** 1488–1497 (2012).

215. Fujioka Y., Ishigaki S., Masuda A., Iguchi Y., Udagawa T., Watanabe H., Katsuno M., Ohno K. and Sobue G. FUS-regulated region- and cell-type-specific transcriptome is

associated with cell selectivity in ALS/FTLD. *Sci. Rep.* **3,** 2388 (2013).

216. Honda D., Ishigaki S., Iguchi Y., Fujioka Y., Udagawa T., Masuda A., Ohno K., Katsuno M. and Sobue G. The ALS/FTLD-related RNA-binding proteins TDP-43 and FUS have common downstream RNA targets in cortical neurons. *FEBS Open Bio* **4,** 1–10 (2014).

217. Masuda A., Takeda J. ichi, Okuno T., Okamoto T., Ohkawara B., Ito M., Ishigaki S., Sobue G. and Ohno K. Position-specific binding of FUS to nascent RNA regulates mRNA length. *Genes Dev.* **29,** 1045–1057 (2015).

218. Masuda A., Takeda J. and Ohno K. FUS-mediated regulation of alternative RNA processing in neurons: insights from global transcriptome analysis. *Wiley Interdiscip. Rev. RNA* **7,** 330–340 (2016).

219. Reber S., Stettler J., Filosa G., Colombo M., Jutzi D., Lenzken S. C., Schweingruber C., Bruggmann R., Bachi A., Barabino S. M., Mühlemann O. and Ruepp M. Minor intron splicing is regulated by FUS and affected by ALS - associated FUS mutants. *EMBO J.* **35,** 1504–1521 (2016).

220. Dini Modigliani S., Errichelli L., Sabatelli M., Morlando M. and Bozzoni I. An ALS-associated mutation in the FUS 3'-UTR disrupts a microRNA-FUS regulatory circuitry. *Nat. Commun.* **5,** 4335 (2014).

221. Kawahara Y., Kwak S., Sun H., Ito K., Hashida H., Aizawa H., Jeong S.-Y. and Kanazawa I. Human spinal motoneurons express low relative abundance of GluR2 mRNA: an implication for excitotoxicity in ALS. *J. Neurochem.* **85,** 680–689 (2003).

222. Yu N. Y. L., Hallström B. M., Fagerberg L., Ponten F., Kawaji H., Carninci P., Forrest A. R. R., Fantom Consortium T., Hayashizaki Y., Uhlén M. and Daub C. O. Complementing tissue characterization by integrating transcriptome profiling from the Human Protein Atlas and from the FANTOM5 consortium. *Nucleic Acids Res.* **43,** 6787–6798 (2015).

223. Yamada M., Clark J. and Iulianella A. MLLT11/AF1q is differentially expressed in maturing neurons during development. *Gene Expr. Patterns* **15,** 80–87 (2014).

224. Postigo A. A. and Dean D. C. Differential expression and function of members of the zfh-1 family of zinc finger/homeodomain repressors. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 6391–6396 (2000).

225. Tan M. H., Li Q., Shanmugam R., Piskol R., Kohler J., Young A. N., Liu K. I., Zhang R., Ramaswami G., Ariyoshi K., Gupte A., Keegan L. P., George C. X., Ramu A., Huang N., Pollina E. A., Leeman D. S., Rustighi A., Goh Y. P. S. *et al.* Dynamic landscape and regulation of RNA editing in mammals. *Nature* **550,** 249–254 (2017).

226. Zervos A. S., Gyuris J. and Brent R. Mxi1, a protein that specifically interacts with

Max to bind Myc-Max recognition sites. *Cell* **72,** 223–232 (1993).

227. Bruson A., Sambataro F., Querin G., D'Ascenzo C., Palmieri A., Agostini J., Gaiani A., Angelini C., Galbiati M., Poletti A., Pennuto M., Pegoraro E., Clementi M. and Soraru G. CAG repeat length in androgen receptor gene is not associated with amyotrophic lateral sclerosis. *Eur. J. Neurol.* **19,** 1373–1375 (2012).

228. Garofalo O., Figlewicz D. A., Leigh P. N., Powell J. F., Meininger V., Dib M. and Rouleau G. A. Androgen receptor gene polymorphisms in amyotrophic lateral sclerosis. *Neuromuscul. Disord.* **3,** 195–199 (1993).

229. Saunderson R. B., Yu B., Trent R. J. A. and Pamphlett R. A comparison of the lengths of androgen receptor triplet repeats in brain and blood in motor neuron diseases. *J. Neurol. Sci.* **267,** 125–128 (2008).

230. Inukai S., Kock K. H. and Bulyk M. L. Transcription factor–DNA binding: beyond binding site motifs. *Curr. Opin. Genet. Dev.* **43,** 110–119 (2017).

231. Ho R., Sances S., Gowing G., Amoroso M. W., O'Rourke J. G., Sahabian A., Wichterle H., Baloh R. H., Sareen D. and Svendsen C. N. ALS disrupts spinal motor neuron maturation and aging pathways within gene co-expression networks. *Nat. Neurosci.* **19,** 1256–1267 (2016).

232. Aguet F., Ardlie K. G., Cummings B. B., Gelfand E. T., Getz G., Hadley K.,

Handsaker R. E., Huang K. H., Kashin S., Karczewski K. J., Lek M., Li X., MacArthur D. G., Nedzel J. L., Nguyen D. T., Noble M. S., Segrè A. V., Trowbridge C. A., Tukiainen T. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550,** 204–213 (2017).