Supplementary Materials for  
**Functional identification of protocadherin alpha 9 (*PCDHA9*)**

**as a causative gene for amyotrophic lateral sclerosis**

**Table S1** **Demographic data of the study cohort**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Clinical features** | **Whole-exome sequencing (WES)** | |  | **Targeted sequencing**  **(besides WES)**\* | |
| **ALS**  **(n=154)** | **Control (n=102)** |  | **ALS (n=238)** | **Control**  **(n=226)** |
| Sex, female (%) | 28.6 | 35.2 |  | 29.8 | 61.1 |
| Age (year, mean ± SD) | 52.8±9.6 | 66.5±10.0 |  | 54.5±10.9 | 69.1±7.4 |
| Age at onset (year, mean ±SD) | 50.4±10.5 | - |  | 51.9±10.8 | - |
| Site of onset, bulbar (%) | 24.7 | - |  | 16.4 | - |

\*The sample set for targeted sequencing included these and those for whole-exome sequencing

**Table S2 Statistics of the whole-exome sequencing data**

|  |  |  |
| --- | --- | --- |
| **Terms** | **Case (n=154)** | **Control (n=102)** |
| Target region size | 60.46 Mb | 60.46 Mb |
| Number of sequencing reads  (s.d.) | 94,674,936  (12,307,549) | 73,451,673  (8,495,361) |
| Size of sequencing data  (s.d.) | 16,736.40 Mb  (2,484.11 Mb) | 14,785.51 Mb  (1,798.53 Mb) |
| Mapping rate (s.d.) | 99.91% (0.03%) | 99.78% (0.18%) |
| On target rate (s.d.) | 71.38% (3.13%) | 63.84% (2.05%) |
| Mismatch rate in target region (s.d.) | 0.19% (0.05%) | 0.42% (0.10%) |
| Mean sequencing depth in target region(s.d.) | 174.49 (22.00) | 176.82 (27.43) |
| 10×coverage in target region (s.d.) | 98.77% (0.76%) | 99.58% (0.26%) |
| 20× coverage in target region (s.d.) | 97.12% (1.01%) | 99.18% (0.36%) |

**Table S3 Genes selected for customized panel sequencing**

|  |  |
| --- | --- |
| **Category of evidence**  **(No. gene)** | **Gene list** |
| Known ALS gene (25) | ALS2, ANG, ATXN2, C9orf72, CHMP2B, DAO, DCTN1, ERBB4, EWSR1, FIG4, FUS, NEFH, OPTN, PFN1, PRPH, SETX, SIGMAR1, SOD1, SPG11, SQSTM1, TAF15, TARDBP, UBQLN2, VAPB, VCP |
| Other association study including GWAS (86) | AGT, ALAD, APEX1, APOE, AR, ARHGEF28, B4GALT6, BCL11B, BCL6, C1orf27, CABIN1, CAMK1G, CCS, CDH13, CDH22, CNTF, CNTN4, CNTN6, CRIM1, CRYM, CST3, CYP2D6, DIAPH3, DISC1, DOC2B, DPP6, DYNC1H1, EFEMP1, ELP3, FEZF2, FGGY, GARS, GRB14, GRN, HEXA, HFE, ITPR2, KDR, KIF13A, KIFAP3, LIF, LIPC, LOX, LUM, MAOB, NAIP, NEK1, NETO1, NT5C1A, OGG1, OMA1, PCP4, PFN2, PFN3, PON1, PON2, PON3, PSEN1, PVR, RAB25, RAMP3, RNF19A, SCN7A, SELL, SEMA6A, SLC1A2, SLC39A11, SLITRK6, SMN1, SMN2, SNCG, SOD2, SOX5, SPAST, SPG7, SUSD1, SYT9, TBK1, TIA1, UBQLN1, VDR, VEGFA, VPS54, ZFP64, ZNF512B, ZNF746 |
| Whole-exome sequencing (36) | AIM1L, CHRM1, CNOT1, CSNK1G3, DENND2C, ELL, FAM151A, FOLR4, FOXA1, FOXK1, GPR132, HDAC10, HOXD8, HS3ST2, KRTAP21, KTI12, LPHN3, MLL3, NCKAP5, NLRC5, NTM, OR5B3, PLEKHO2, PSMB7, SRCAP, SS18L1, STARD13, TRPM4, TRRAP, UBQLN3, UNC13A, UTP6, VCL, WDR1, ZNF410, ZNF778 |
| Functional study (RRM) (31) | CELF4, CSTF2, CSTF2T, DAZ1, DAZ2, DAZ3, DAZ4, DAZAP1, ELAVL1, ELAVL2, ELAVL3, ELAVL4, G3BP1, G3BP2, HNRNPA0, HNRNPA1, HNRNPA3, HNRNPAB, HNRNPD, HNRNPH1, HNRNPH2, HNRNPH3, HNRPDL, MSI2, PSPC1, RBM14, RBM33, RBMS1, SFPQ, SSB, TIAL1 |
| The current study (110) | ABCA2, ABCA5, AHNAK2, AKD1, ANAPC7, ARNTL2, ARPP21, ASPM, ATP8B3, BAGE2, BIRC6, BRIP1, C20orf26, C5orf42, C9orf11, CACNA1H, CAMP, CCDC141, CCDC75, CEP70, CHCHD10, CHGB, CHRNA3, CHRNA4, CHRNB4, CMIP, CNGA4, COL19A1, CRLF3, CST7, CUBN, DGKK, DNAH10, DNAH2, DNAH9, DNMT3A, EIF4E1B, EIF5, EHMT1, ENAH, EPHA4, ERICH1, F11R, FAT4, FGF23, FKBP5, FLG, FOXN3, FOXR1, FRAS1, GLE1, GMPR, GOLGA5, GOLGA6B, GPR158, GPR98, GTF2H4, GZMH, ITPR1, ITPR3, KCNA5, KLHL6, KRIT1, KRTAP5, LAMC3, LBP, LGALSL, LIMD1, MATR3, METTL16, METTL22, MKI67, MUC4, MXRA5, MYO3B, MYOM1, NCOA6, NEB, NIPA1, NR2E3, NTE, OR4N2, PCDH19, PCDHA8, PCDHA9, PDE2A, PLEKHG5, PRTG, RAB38, RAI2, RINL, RP1L1, SLC22A3, SMG1, SND1, SPTA1, SPTB, SRSF8, STK36, SV2A, SYNE1, SYNE2, TAS2R31, TBC1D30, TCHH, TRIP10, TRPM7, TSSK1B, TUBA4A, USH2A, WDR60 |

GWAS: genome-wide association study; RRM: RNA-recognition motif

**Table S4** **Statistics of the targeted gene sequencing data**

|  |  |  |
| --- | --- | --- |
| **Terms** | **Case (n=392)** | **Control (n=328)** |
| Target region size | 0.95Mb | 0.95Mb |
| Number of sequencing reads  (s.d.) | 18,018,674 (4,538,421) | 16,586,023  (4,026,816) |
| Size of sequencing data  (s.d.) | 2,598.12Mb  (634.36 Mb) | 2,372.85 Mb  (553.52 Mb) |
| Mapping rate (s.d.) | 99.63% (0.48%) | 99.74% (0.24%) |
| On target rate (s.d.) | 48.35% (3.42%) | 50.46% (4.41%) |
| Mismatch rate in target region (s.d.) | 0.72% (0.16%) | 0.68% (0.13%) |
| Mean sequencing depth in target region(s.d.) | 1320.88 (346.59) | 1252.65 (303.75) |
| 10X coverage in target region (s.d.) | 97.73% (1.31%) | 98.70% (0.40%) |
| 20X coverage in target region (s.d.) | 97.50% (1.49%) | 98.64% (0.40%) |

**Table S5 Statistics for variants identified by targeted gene sequencing**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variants** | **Total** | **Known** | **Novel** |
| Total variants | 12,898 | 7,818 | 5,080 |
| Common (MAF ≥ 0.05) | 1,113 | 1,107 | 6 |
| Low (0.01 ≤ MAF < 0.05) | 509 | 502 | 7 |
| Rare (MAF < 0.01) | 11,276 | 6,209 | 5,067 |
| Total SNVs | 11,076 | 6,295 | 4,781 |
| Non-coding SNVs | 166 | 68 | 98 |
| Splicing sites | 5 | 2 | 3 |
| Coding SNVs | 10,905 | 6,225 | 4,680 |
| Synonymous | 3,776 | 2,547 | 1,229 |
| Nonsynonymous | 7,129 | 4,784 | 3,789 |
| Common (MAF>0.05) | 528 | 527 | 1 |
| Low-frequency (0.01≤MAF≤0.05) | 266 | 264 | 2 |
| Rare (MAF<0.01) | 6,335 | 2,887 | 3,448 |
| Rare damaging | 3,139 | 1,300 | 1,839 |

**Table S6 Mutations identified in known ALS genes**

|  |  |  |
| --- | --- | --- |
| **Gene** | **No. mutations** | **No. Carriers (n=392)** |
| TARDBP | 2 | 2 |
| FUS | 3 | 2 |
| SOD1 | 4 | 5 |
| ALS2 | 3 | 3 |
| ANG | 1 | 1 |
| EWSR1 | 3 | 2 |
| FIG4 | 2 | 2 |
| NEFH | 2 | 2 |
| OPTN | 3 | 3 |
| PRPH | 2 | 3 |
| SETX | 1 | 1 |
| SPG11 | 3 | 3 |
| DCTN1 | 3 | 3 |
| ERBB4 | 3 | 3 |
| SQSTM1 | 1 | 1 |
| HNRNPA2B1 | 1 | 2 |
| SS18L1 | 1 | 1 |
| MATR3 | 2 | 2 |
| Total | 39 | 41 |

**Table S7 Clinical features of patients with the homozygous *PCDHA9* L700P mutation**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Family 1 II-4** | **Family 2 II-2** | **Family 3 II-5** |
| Zygosity | Homozygous | Homozygous | Homozygous |
| Sex | Male | Female | Male |
| Age at onset (years) | 38 | 42 | 36 |
| Onset symptoms | Weakness in left lower limb | Weakness in right upper limb | Weakness in left upper limb |
| Diagnostic delay (months) | 6 | 8 | 10 |
| Phenotypes (p-UMN/Classic） | Classic | Classic | Classic |
| Duration (years) | 2.8 | 3.8 | 3.6 |

**Table S8. The raw electromyogram (EMG) data of the mutant and wildtype mice**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No. mice | Fib | | PSW | | Fas | |  | Fib | | PSW | | Fas | |  | Fib | | PSW | | Fas | |  | Fib | | PSW | | Fas | |  | Fib | | PSW | | Fas | |  | Fib | | PSW | | Fas | |
| L | R | L | R | L | R |  | L | R | L | R | L | R |  | L | R | L | R | L | R |  | L | R | L | R | L | R |  | L | R | L | R | L | R |  | L | R | L | R | L | R |
| Mut 1 | 1 | 1 | 2 | 0 | 0 | 0 |  | 1 | 4 | 1 | 0 | 0 | 0 |  | 0 | 0 | 0 | 0 | 0 | 0 |  | 0 | 2 | 1 | 1 | 0 | 0 |  | 1 | 4 | 1 | 0 | 0 | 0 |  | 0 | 0 | 0 | 0 | 0 | 0 |
| Mut 2 | 0 | 2 | 2 | 0 | 0 | 0 |  | 0 | 0 | 4 | 2 | 0 | 1 |  | 4 | 0 | 1 | 1 | 0 | 0 |  | 1 | 0 | 1 | 1 | 0 | 0 |  | 0 | 0 | 0 | 0 | 0 | 0 |  | 0 | 0 | 0 | 0 | 0 | 0 |
| Mut 3 | 2 | 0 | 1 | 0 | 0 | 0 |  | 0 | 0 | 0 | 0 | 0 | 0 |  | 3 | 1 | 1 | 0 | 0 | 0 |  | 1 | 0 | 0 | 0 | 0 | 0 |  | 1 | 4 | 2 | 0 | 0 | 0 |  | 0 | 0 | 0 | 0 | 0 | 0 |
| Mut 4 | 2 | 1 | 1 | 3 | 0 | 0 |  | 0 | 0 | 0 | 0 | 0 | 0 |  | 1 | 0 | 0 | 2 | 0 | 0 |  | 0 | 0 | 1 | 0 | 0 | 0 |  | 0 | 0 | 0 | 0 | 0 | 0 |  | 3 | 0 | 1 | 1 | 0 | 0 |
| Con 1 | 0 | 0 | 0 | 0 | 0 | 0 |  | 1 | 0 | 0 | 0 | 0 | 0 |  | 0 | 0 | 0 | 0 | 0 | 0 |  | 0 | 0 | 0 | 0 | 0 | 0 |  | 1 | 0 | 1 | 0 | 0 | 0 |  | 0 | 0 | 1 | 0 | 0 | 0 |
| Con 2 | 0 | 0 | 0 | 0 | 0 | 0 |  | 0 | 0 | 0 | 0 | 0 | 0 |  | 0 | 0 | 0 | 0 | 0 | 0 |  | 0 | 0 | 0 | 0 | 0 | 0 |  | 0 | 0 | 0 | 0 | 0 | 0 |  | 0 | 0 | 0 | 0 | 0 | 0 |
| Con 3 | 0 | 0 | 0 | 0 | 0 | 0 |  | 0 | 0 | 0 | 0 | 0 | 0 |  | 0 | 0 | 0 | 0 | 0 | 0 |  | 0 | 0 | 0 | 0 | 0 | 0 |  | 0 | 0 | 0 | 0 | 0 | 0 |  | 0 | 0 | 0 | 0 | 0 | 0 |
| Con 4 | 0 | 0 | 0 | 0 | 0 | 0 |  | 0 | 0 | 0 | 0 | 0 | 0 |  | 0 | 0 | 0 | 0 | 0 | 0 |  | 0 | 0 | 0 | 0 | 0 | 0 |  | 0 | 0 | 0 | 0 | 0 | 0 |  | 0 | 0 | 0 | 0 | 0 | 0 |

Needles were entered in the four quadrants of each muscle. 0, no spontaneous potential is found in all four quadrants; 1, spontaneous potential is found in one quadrant; 2, spontaneous potential found in two quadrants, and so on. Anesthesia method :10%chloral hydrate 0.3mg/kg intraperitoneal injection. Mut: mutant; Con: control; L: left; R: right; Fib: fibrillation; PSW: positive sharp wave; Fas: fasculation

**图表

描述已自动生成**

**Supplementary Fig. 1. Generation of *Pcdha9* L729P (Mut) and M723fs (Del) mice.**

**(A)** Conservation between mouse and human in the transmembrane domain of PCDHA9.

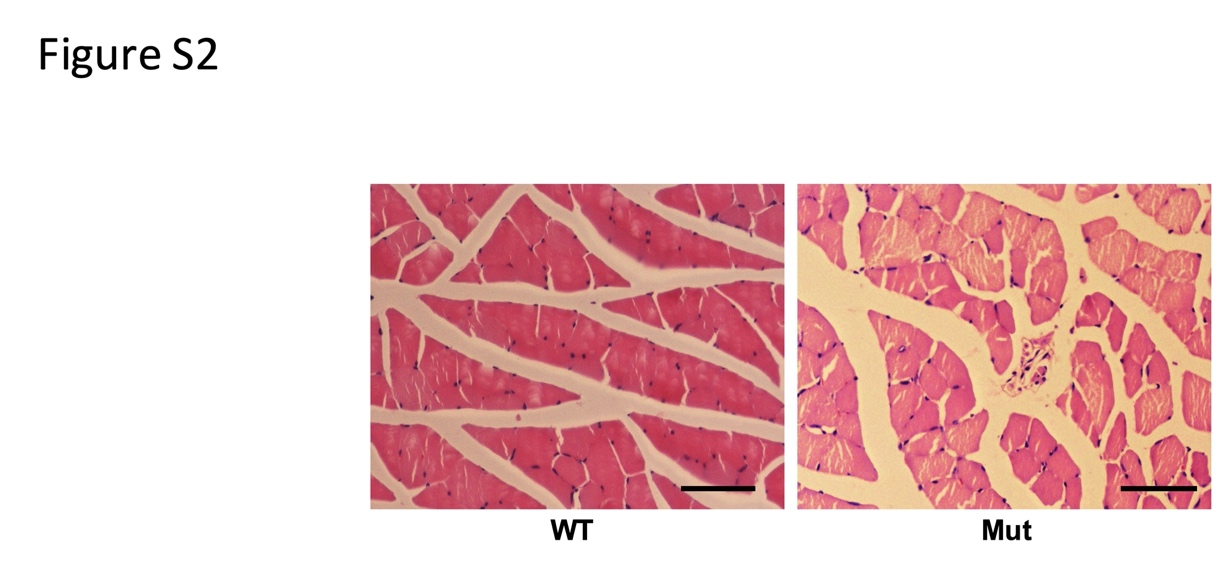
**(B)** Schematic of genomes for Mut and Del mice.

**(C)** Sequencing confirmation of T2185C (left panel) and deletion of 2166A (right panel) in mice.

**(D)** Relative mRNA levels of *Pcdha9* in the spinal cord of 6-month old female WT, Mut and Del mice by quantitative real time PCR (n=3).

**(E)** Body weight of mice at different age. Mut and Del mice exhibited body weight loss at 12-month old (7-month old: wt, n=11; mut, n=13; del, n=4. 10-month old: wt, n=12; mut, n=10; del, n=6. 12-month old: wt, n=18; mut, n=12; del, n=15). All data represent mean ± SEM.\*P < 0.05, \*\*P < 0.01, one-way ANOVA.

(Related to Fig. 2)

****

**Supplementary Fig. 2. Skeletal muscular atrophy in *Pcdha9* Mut mice.** H&E staining of gastrocnemius muscles from 12-month old WT and Mut mice. Scale bar = 50 μm.

(related to Fig. 3)

**图示

描述已自动生成**

**Supplementary Fig. 3. Abnormal neuromuscular junctions (NMJs) in both *Pcdha9* Mut and Del mice.**

**(A)** Statistic analysis of CHAT+ motor neurons in 10 month old mice spinal cord. N=3 mice. t test. Error bars represent ± SEM. **(B)** Confocal images of NeuN-positive neurons in the ventral horn of lumbar spinal cord from 12-month old WT and Mut mice (related to Fig. 4A. Right panel: Statistical analyses of NeurN positive cells (N=3 mice; n=15 slice. \*p<0.05, \*\*\*p<0.001, t test. Error bars represent ± SEM). **(C)** The area analysis of single NMJs. Scale bar = 30 μm. N=3 mice. \*\*p<0.01, \*\*\*p<0.001, one-way ANOVA. Error bars represent ± SEM. **(D)** Confocal images of AChR in NMJs and analyses of AChR-faint-innervated NMJs (realated to Fig. 4B). NMJs were double stained with α-BTX (red) and anti-neurofilament (NF, presynaptic, green). The white arrow indicated faintly labeled NF-200 and AChRs (with decreased intensity of α-BTX staining). Scale bar = 30 μm. (3 mice for each group. NMJs: n=452 in WT, n=736 in Mut, n=620 in Del mice).

(related to Fig. 4)

**图形用户界面, 应用程序

描述已自动生成**

**Supplementary Fig. 4. Myelination is not affected significantly in the spinal cord and sciatic nerve in Mut mice.**

**(A)** Transmission electron microscopy of sciatic nerve from 13-month old mice.

**(B)** Statistic analysis of the g-ratio in sciatic nerve from (A). N=3 mice. one-way ANOVA. Error bars represent ± SEM.

**(C)** Black gold II staining of the spinal cord from 12-month old mice.

(related to Fig. 4)

**图示

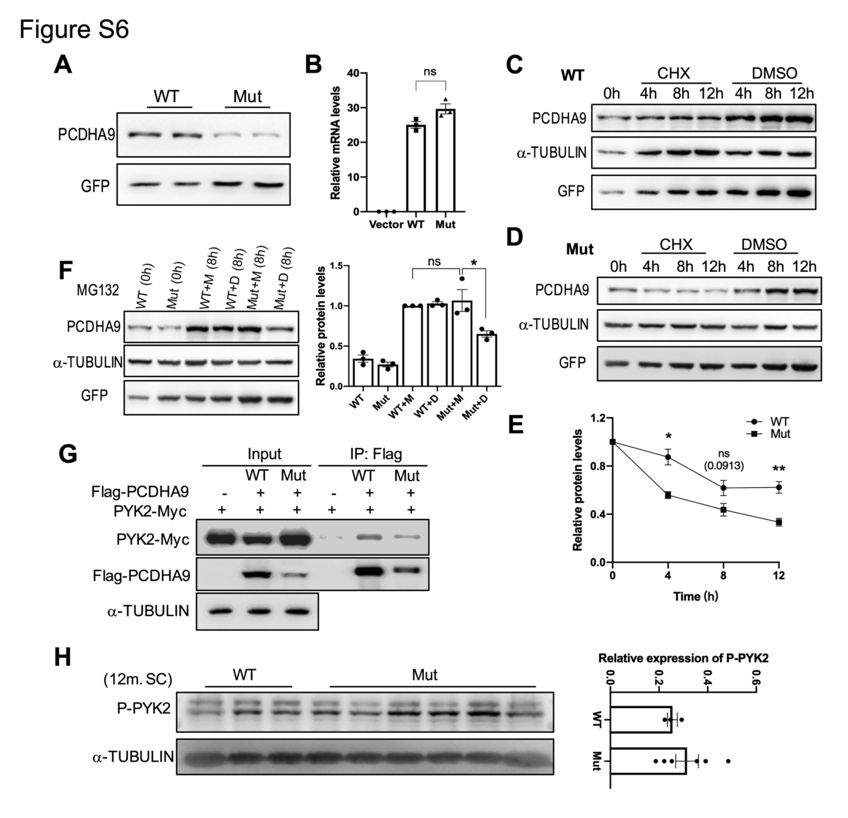
描述已自动生成**

**Supplementary Fig. 5. Tandem mass tag (TMT)- labeled quantitative proteomic analysis of the ventral horn of spinal cord from Mut and littermate control mice.**

**(A)** GO analysis of quantitative proteomic data of the ventral horn of L3-L5 spinal cord from the 12-month-old WT and Mut littermates. Top 4 are labelled with rectangular box.

**(B)** Network graph of the significantly changed proteins with their functions. The graph was generated by the Cytoscape software. Red color indicated the increased expression levels in Mut mice compared with WT mice, while blue color indicated the decreased ones. The depth of the color represents the extent of changes. Log2 FD > 1.25 and Log2 FD < 0.75.

(related to Fig. 5)

****

**Supplementary Fig. 6. Mutation of PCDHA9 leads to its instability and the hyper-phosphorylation of PYK2. (A)** Western blot of overexpressed WT and mutant hPCDHA9 in HEK293 cells. GFP was used as transfection efficiency control. **(B)** Relative mRNA levels of transfected WT and mutant hPCDHA9 analyzed by real time PCR and normalized by GAPDH. **(C, D)** Expression of WT (C) or mutant (D) hPCDHA9 in transfected 293 cells treated with/without CHX for different time. GFP and α-TUBULIN were used as transfection efficiency and loading controls. **(E)** Statistic analysis of the relative expression levels of hPCDHA9 in C and D. t-test. The data represent mean ± SEM. \*P<0.05. \*\*P<0.01. **(F)** Expression levels of WT and mutant hPCDHA9 in transfected 293 cells treated with/without the proteasome inhibitor MG132 for 8 hours. Expression of mutant hPCDHA9 was enhanced by MG132. (M for “MG132”, D for “DMSO”). **(G)** Anti-FLAG immunoprecipitation was performed with 293 cells expressing the indicated constructs. The input and immunoprecipitated proteins were detected by Western blots. **(H)** Phosphorylation levels of PYK2 in the spinal cord of 12-month old WT and Mut mice by Western blot. Statistic analysis the expression levels were normalized by α-TUBULIN, unpaired t-test. The data represent mean ± SEM. No significant difference.

(related to Fig. 5)

**图示

描述已自动生成**

**Supplementary Fig. 7. The protein levels of NKA-β1 and NKA-α3 in the spinal cord of 12-month-old WT and Mut mice were not affected. (A)** Western blot of NKA-β1. **(B)** Western blot of NKA-α3. N=4 mice for each group.

(related to Fig. 5)

**图示

描述已自动生成**

**Supplementary Fig. 8. Quality control and integration efficiency of snRNA-seq and snATAC-seq. (A)** UMAP plot of the percentage of contamination in both WT and Mut snRNA-seq data. **(B)** The relationships of the features of cells from spinal cord snRNA-seq, such as the number of UMI detected, percentage of cell counts mapping to mitochondrial genes and the number of genes detected. (C) Visualization of cells in WT and Mut snRNA-seq data in tSNE plots and no obvious distribution differences were found between the two datasets. **(D)** The simulated doublet enrichments of snATAC-seq data from WT and Mut mice in UMAP plots. **(E)** Fragment size distributions for the cells in snATAC-seq. **(F)** Aggregate TSS insertion profiles centered at all TSS regions in snATAC-seq. **(G)** UMAP plot of integrated snATAC-seq dataset from WT and Mut mice. **(H)** Visualization of cells predicted scores that represent the similarity of snATAC-seq and snRNA-seq data. Vertical red line marked the predicted scores 0.8. **(I)** Sankey plots illustrating the efficiency of integration between snATAC-seq and snRNA-seq data from WT and Mut mouse spinal cord.

(related to Fig. 6)

**图示, 示意图

描述已自动生成**

**Supplementary Fig. 9. Analysis of snRNA-seq and snATAC-seq. (A)** Unsupervised cell clusters visualized by UMAP in integrated WT and Mut snATAC-seq data, and major cell types was identified by known markers. **(B)** Heatmap for conserved gene expression enriched in different cell types based on gene score. **(C)** Pie chart comparing proportion of gene function in DEGs.

**(D)** The depictions of the linkages of genes and biological concepts of interested. **(E)** GO analysis was performed to illuminate biological functions of genes associated with significant DARs in neuron cluster. **(F)** Go analysis of the downstream genes of Rxra according to Mut mouse GRN. **(G)** In neuron cluster, the specific cis-regulatory interactions of Malat1 in Mut mouse and WT mouse. According to the GRN, these peaks were predicted to be the binding sites of Rxra. **(H)** Overview of specific ligand–receptor interactions using CellPhoneDB on Mut mouse compared with WT mouse snRNA-seq.

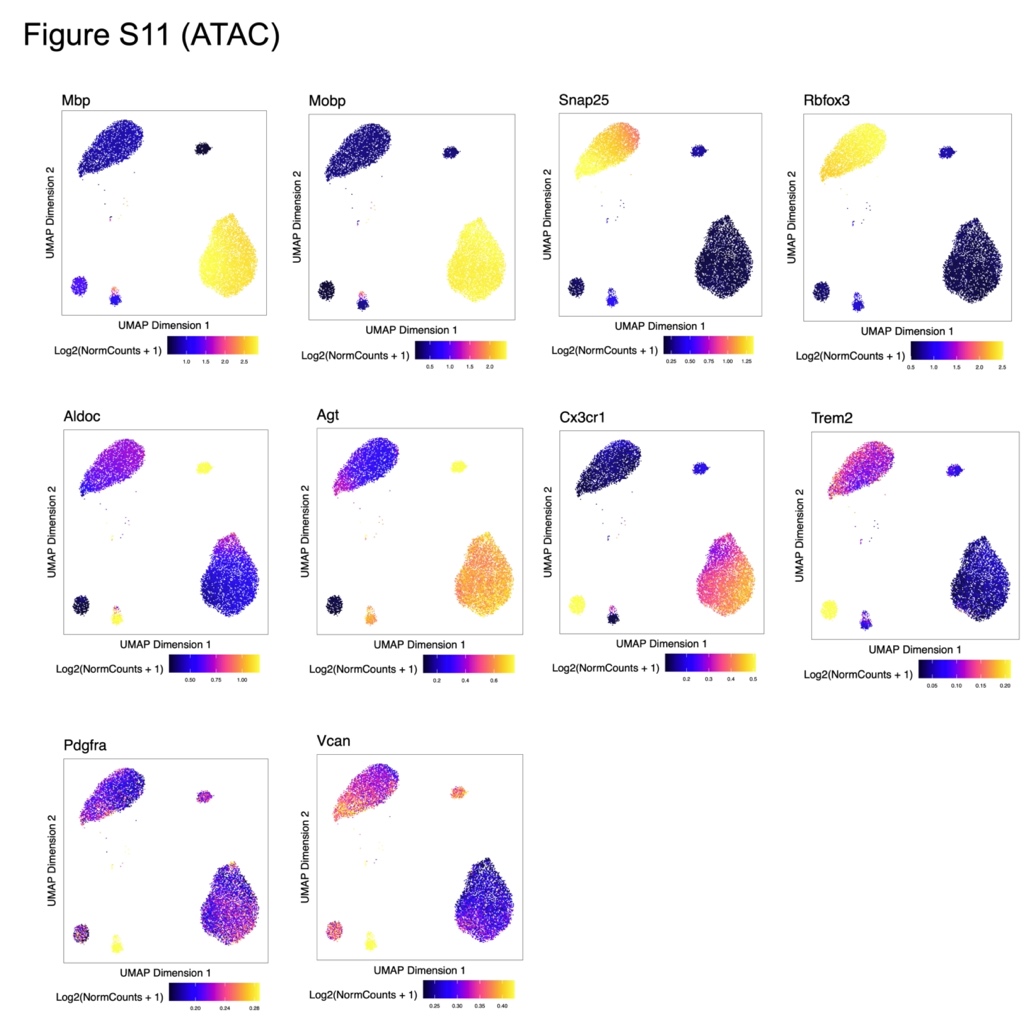
(related to Fig. 6)

**图示

描述已自动生成**

**Supplementary Fig. 10. Feature plot of marker genes in snRNA-seq.** *Mbp* and *Mobp* for oligodendrocytes, *Snap25* and *Rbfox3* for neurons, *Aldoc* and *Agt* for astrocytes, *Cx3cr1* and *Trem2* for microglias, *Pdgfra* and *Vcan* for OPCs.

(related to Fig. 6)

****

**Supplementary Fig. 11. Feature plot of marker genes in snATAC-seq.** *Mbp* and *Mobp* for oligodendrocytes, *Snap25* and *Rbfox3* for neurons, *Aldoc* and *Agt* for astrocytes, *Cx3cr1* and *Trem2* for microglias, *Pdgfra* and *Vcan* for OPCs.

(related to Fig. 6)