Published in final edited form as:

Acta Neuropathol. 2017 November ; 134(5): 715–728. doi:10.1007/s00401-017-1760-4.

Conserved DNA methylation combined with differential frontal cortex and cerebellar expression distinguishes *C9orf72*-associated and sporadic ALS, and implicates *SERPINA1* in disease

Mark T. W. Ebbert, Ph.D.¹, Christian A. Ross, Ph.D.², Luc J. Pregent¹, Rebecca J. Lank¹, Cheng Zhang, Ph.D.², Rebecca B. Katzman, Ph.D.¹, Karen Jansen-West¹, Yuping Song¹, Edroaldo Lummertz da Rocha, Ph.D.², Carla Palmucci³, Pamela Desaro³, Amelia E. Robertson³, Ana M. Caputo³, Dennis W. Dickson, M.D.¹, Kevin B. Boylan, M.D.³, Rosa Rademakers, Ph.D.¹, Tamas Ordog, M.D.^{4,5}, 4-5, Hu Li, Ph.D.^{2,*}, and Veronique V. Belzil, Ph.D.^{1,*}

¹Department of Neuroscience, Mayo Clinic, 4500 San Pablo Road, Jacksonville, Florida 32224, USA

²Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, 200 First Street SW, Rochester, Minnesota 55905, USA

³Department of Neurology, Mayo Clinic, 4500 San Pablo Road, Jacksonville, Florida 32224, USA

⁴Epigenomics Program, Center for Individualized Medicine, Mayo Clinic, 200 First Street SW, Rochester, Minnesota 55905, USA

⁵Department of Physiology and Medical Engineering, Mayo Clinic, 200 First Street SW, Rochester, Minnesota 55905, USA

Abstract

We previously found *C9orf72*-associated (c9ALS) and sporadic amyotrophic lateral sclerosis (sALS) brain transcriptomes comprise thousands of defects, among which, some are likely key contributors to ALS pathogenesis. We have now generated complementary methylome data and combine these two data sets to perform a comprehensive “multi-omic” analysis to clarify the molecular mechanisms initiating RNA misregulation in ALS. We found that c9ALS and sALS patients have generally distinct but overlapping methylome profiles, and that the c9ALS- and sALS-affected genes and pathways have similar biological functions, indicating conserved pathobiology in disease. Our results strongly implicate *SERPINA1* in both *C9orf72* repeat expansion carriers and non-carriers, where expression levels are greatly increased in both patient groups across the frontal cortex and cerebellum. *SERPINA1* expression is particularly pronounced in *C9orf72* repeat expansion carriers for both brain regions, where *SERPINA1* levels are strictly down regulated across most human tissues, including the brain, except liver and blood, and are not

*Corresponding authors: Lead contact: Veronique Belzil, Ph.D., 4500 San Pablo Road, Jacksonville, Florida, 32224 USA, Tel.: 1-904-953-6905, Belzil.Veronique@mayo.edu; Hu Li, Ph.D., 200 First Street SW, Rochester, Minnesota, 55905, USA, Tel.: 1-507-293-1182, Li.Hu@mayo.edu.

Conflict of interest: The authors declare that they have no conflict of interest.

measurable in E18 mouse brain. The altered biological networks we identified contain critical molecular players known to contribute to ALS pathology, which also interact with *SERPINA1*. Our comprehensive combined methylation and transcription study identifies new genes and highlights that direct genetic and epigenetic changes contribute to c9ALS and sALS pathogenesis.

Keywords

Amyotrophic lateral sclerosis; *C9orf72*; DNA methylation; epigenetic modification; *SERPINA1*; transcriptome regulation

Introduction

Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are devastating neurodegenerative disorders seen comorbidly in up to 50% of patients, with only a small proportion of ALS patients reporting a family history of either disease[8]. Genetic studies have implicated more than two dozen genes in disease pathogenesis[36,40], the most common mutation being a GGGGCC (G₄C₂) hexanucleotide repeat expansion in the *C9orf72* gene that is carried by 6% and 34% of sporadic (sALS) and familial (fALS) cases, respectively[34]. Since sALS accounts for up to 95% of all ALS cases[7,13], and only ~11% of these cases can be attributed to known genetic risk factors, there is an intensifying need to reveal sALS etiology. The motor cortex is the brain region primarily affected in ALS, though increasing evidence suggests the frontal cortex and cerebellum play a role in both c9ALS and sALS[33,12]. Our recent RNA sequencing (RNAseq) study demonstrated distinct misregulated brain transcriptome profiles in the cerebellum and frontal cortex for both sALS patients, and ALS patients carrying the *C9orf72* G₄C₂ repeat expansion (c9ALS)[33]. Transcriptome defects, including differential expression and alternative splicing, were abundant in both the frontal cortex and cerebellum[33]. While transcriptome data provide critical disease insight, understanding the mechanism underlying these distinct transcriptome profiles may be even more important. Here, we contrast epigenetic modifications in frontal cortex and cerebellum between genetically unexplained sALS with c9ALS cases to: (1) further understand the mechanisms driving the distinct transcriptome profiles and disease, and (2) filter and identify new genes driving disease development and progression.

Substantial evidence suggests epigenetic changes may be regulating the changes observed in ALS. Xi et al. previously reported ALS-discordant monozygotic twins carrying the G₄C₂ *C9orf72* repeat expansion[43], suggesting the *C9orf72* repeat expansion, alone, may not have driven disease in the affected twin. Similarly, Lam et al. recently demonstrated significant DNA methylation differences between 50-year-old, ALS-discordant monozygotic twins[21]. Both studies suggest epigenetics may be a catalyst for disease, even with the highly penetrant *C9orf72* genetic variant. Recent ALS epigenetic studies also showed patients with c9ALS have expression-altering epigenetic modifications of the *C9orf72* locus, including repressive histone marks and DNA methylation in promoter and repeat regions[4,45,42,44]. Liu et al. further demonstrated that hyper-methylation in the *C9orf72* promoter region moderates: (1) RNA foci accumulation, (2) *C9orf72* protein burden, and (3) vulnerability to oxidative stress, supporting an interaction between aberrant epigenetic

control, RNA misregulation and protein toxicity in c9ALS[23]. Additionally, hairpin and G-quadruplex structures adopted by the *C9orf72* repeat expansion RNA[11,38,15] may interact with chromatin remodeling factors, as similar secondary structures have been shown to attract several chromatin-modifying complexes[17]. In fact, there is now compelling evidence that RNA not only serves as an intermediary between DNA and protein, but also mediates chromatin remodeling and nuclear architecture through epigenetic regulation[2,24,16,26,6]. Evidence further supports altered RNA-mediated regulation as a central pathological epigenetic mechanism in many diseases[14,31,33,35,5]. While some studies have demonstrated epigenetic involvement in ALS[25], identifying the *C9orf72* repeat expansion provided even more support, as other repeat expansion disorders lead to chromatin remodeling [4,3,42,45,44,5].

To clarify the molecular mechanisms initiating RNA misregulation in ALS, we sought to begin characterizing the relationship between methylation and disease by: (1) evaluating whether aberrant DNA methylation contributes to the distinct c9ALS and sALS transcriptomes previously described[33], and (2) identifying interesting epigenetic mediators of RNA regulation in ALS, and (3) assess the role of ALS-associated genes in an interconnected gene network context. We generated DNA methylome profiles for frontal cortex and cerebellum from c9ALS, sALS, and non-disease control individuals from our initial RNAseq study[33] and found that genes in critical ALS pathways have both transcriptome defects (e.g., differential expression and alternative splicing) and differential methylation. We present a valuable characterization of methylation profiles in c9ALS and sALS across the frontal cortex and cerebellum, and we validate new gene targets that may be critical in disease. This is the first full methylome characterization across the frontal cortex and cerebellum in ALS, and we combine it with our previous RNASeq data for a comprehensive “multi-omic” approach to better understand what drives disease.

Materials and Methods

Human Tissue

Human cerebellum and frontal cortex tissues were obtained with written informed consent from participants or authorized family members and processed under the approved Mayo Clinic Institutional Review Board and Ethics Committee on Human Experimentation 09-008148 and 12-007795 protocols. Information about ALS subject recruitment, diagnosis, genotyping, pathological assessment procedure and brain tissue sampling were previously described[33]. Among the c9ALS, sALS and control subjects used for the RNAseq study, four subjects per group were matched based on clinical information and selected for methylome profiling. We validated differential gene expression using a total of 32 cases carrying a pathogenic *C9orf72* repeat expansion (16 ALS, 3 ALS-FTD, 11 FTD), 28 cases carrying a normal *C9orf72* repeat size (17 ALS, 3 ALS-FTD, 8 FTD), and 8 disease controls. Some samples included in cerebellum were excluded in frontal cortex because of RNA quality. We included FTD cases to increase sample sizes during validation. The controls were confirmed pathologically normal, and have no history of motor neuron disease or dementia, as previously described[33]. Patients were classified into ALS, ALS-FTD, and FTD groups based on the neuropathological diagnosis. Information on all subjects used for

this study can be found in online resource, Table 12. All subjects were confirmed negative for protein-coding mutations in *SOD1*, *TARDBP*, and *FUS*.

Next Generation RNA sequencing (RNAseq)

All RNASeq results are available online at the NCBI Gene Expression Omnibus under accession number GSE67196 [33].

DNA and RNA Preparation

DNA was isolated from 20–40mg of sampled frontal cortex and cerebellum frozen tissue using the Wizard genomic DNA purification kit (Promega, Madison, WI, USA) per manufacturers' instructions, and re-suspended in TE buffer. Concentration was assessed using a Qubit 3.0 Fluorometer (ThermoFisher Scientific, Waltham, MA), and 50ng/ul of DNA in a 10ul volume was sent to the Mayo Clinic Medical Genome Facility in Rochester (Minnesota, USA) for processing and sequencing.

RNA was isolated from 20–40mg of sampled frontal cortex and cerebellum frozen tissue using a two-step tissue homogenization method. Total RNA was extracted using miRNeasy Mini Kit (QIAGEN, Hilden, Germany), per manufacturer's instructions. RNA integrity (RIN>6.6) was verified on an Agilent 2100 bioanalyzer (Santa Clara, CA, USA).

Reduced Representation Bisulfite Sequencing (RRBS)

Reduced representation bisulfite sequencing (RRBS) library construction and sequencing were conducted at the Mayo Clinic Medical Genome Facility in Rochester (Minnesota, USA). Briefly, genomic DNA (250ng) was digested with the methylation insensitive restriction digest enzyme, MspI, which cleaves the DNA at CCGG, creating fragments high in CpG content. The DNA was then purified using QIAGEN (Hilden, Germany) MiniElute columns. Samples were end-repaired and A-tailed using a Klenow fragment. The end repair A-tailing allowed for selective ligation of the Illumina (San Diego, CA, USA) TruSeq™ indexed adaptors. TruSeq™ adaptors were ligated to the sample DNA using T4 DNA ligase. Size selection was performed with Ampure Xp™ beads. Selection range was between 150 and 400bp, allowing for insert size of 40 to 280bp and including many CpG islands and shores. Each sample was bisulfite modified using the Zymo EX DNA Methylation Kit (Zymo Research Corp, Irvine, CA; PN D5001) under the following conditions: 55 cycles of 95°C for 30 seconds followed by 50°C for 15 minutes, then held at 4°C. The samples were then PCR amplified (15 cycles) to enhance for CpG regions. Concentration and size distribution of the libraries were determined using Agilent (Santa Clara, CA, USA) Bioanalyzer DNA 1000 chip, Qubit fluorometry (Invitrogen, Carlsbad, CA), and qPCR. Libraries were pooled at equimolar concentrations and loaded onto a paired end flow cell at concentrations of 7-8pM to generate cluster densities of 600,000-700,000/mm² following Illumina's standard protocol, using the Illumina cBot and HiSeq Paired end cluster kit version 3 (San Diego, CA, USA). Because bisulfite treated samples are unbalanced due to an under representation of C bases, a lane of PhiX genomic DNA was run to allow the Illumina data collection software to call bases correctly. The flow cells were sequenced as 51×2 paired end reads on an Illumina HiSeq 2000 using TruSeq SBS sequencing kit version 3 and HiSeq data collection version 2.0.12.0 software. Base-calling was performed using

Illumina's RTA version 1.17.21.3. We assessed bisulfite conversion success rate by calculating the percentage of non-CpG cytosines across all samples, where we would expect nearly all said cytosines to convert. Conversion percentages were >99% for all samples in the CHG context, and ranged from 98.6% to 99.5% in the CHH context, where 'H' represents any non-G nucleotide.

RRBS Alignment, Annotation, Duplicate Removal, and Bioinformatic Analysis

Paired-end reads were aligned to the hg19 reference genome and cytosine counts were determined using Bismark (v0.13.1)[20] and Bowtie 2 (v2.2.3)[22]. We tested for differential methylation between sample groups with methylKit's[1] calculateDiffMeth function, using logistic regression and a sliding linear model for false discovery rate adjustment. Sites were filtered for minimum read coverage of 10, and to be within the bottom 99.9% of coverage for all sites. Each differentially methylated cytosine (DMC) was then annotated using methylKit's *get Association With TSS* function according to (1) the nearest downstream transcription start site or (2) the "gene body" (located between the transcription start site and end of the 3' UTR) in which the DMC was located, according to UCSC genome browser hg19 refseq annotations (<https://genome.ucsc.edu/>). A DMC was considered significant if the q-value was less than 0.01 and the estimated methylation in the c9ALS or sALS group was greater than 25% different from the control group. Duplicate removal was conducted in the following way: (1) if a DMC was located upstream from the TSS of more than one gene, it was counted as multiple events; (2) if a DMC was located upstream or in the gene body of different transcript variants of the same gene, it was counted as one single event; (3) if a DMC was located one base pair away from another methylation event on the opposite strand and annotated to the same gene, it was considered a paired symmetrical methylation and was counted as one single event; (4) if a DMC was located one base pair away from another DMC on the opposite strand and annotated to a known antisense gene, it was counted as two different events.

RT-PCR and qRT-PCR

Using the High Capacity cDNA Transcription Kit (Applied Biosystems, Foster City, CA, USA) per manufacturer's instructions, 500ng of RNA was used to generate cDNA by reverse transcription polymerase chain reaction (RT-PCR). To quantify the RNAs, quantitative real-time PCRs (qRT-PCRs) were conducted using cDNA generated from the High Capacity cDNA Transcription Kit and the SYBR Green PCR Mastermix (Applied Biosystems, Foster City, CA, USA), per manufacturer's instructions. All primer sequences are available in online resource, Table 13. Samples were run in triplicate on a QuantStudio 7 Flex PCR System (Applied Biosystems, Foster City, CA, USA). Relative quantification was determined using the $\Delta\Delta C_t$ method and the median for each sample was normalized to the endogenous control *RPLP0*. The non-parametric Kruskal-Wallis test was used to test for significant differences between cases and controls. The false discovery rate was controlled using the two-stage linear step-up procedure of Benjamini, Krieger and Yukutieli. Comparisons reached significance with p-values <0.05 (*p<0.05; **p<0.01; ***p<0.001).

Combined RRBS and RNASeq analysis and validation

Because our data are a comprehensive methylation profile in ALS, we filtered out many genes to identify specific targets to validate. Specifically, we focused on genes that were differentially expressed ($p < 0.05$, $|\log_2FC| > 1$) and had differential methylation patterns ($p < 0.05$) with $>25\%$ difference between c9ALS and controls, or between sALS and controls. From that list, we looked for genes that were common between either c9ALS and sALS, or between cerebellum and frontal cortex, and tested for association between expression and methylation in these genes using logistic regression (glm, family = "binomial") in R (v. 3.3.2). To preserve power, samples were sorted into two groups, fully methylated and partial/fully unmethylated, and tested for association with expression levels. Genes that are common across these groups may be critical to disease development and progression. We also used public data sets from The Genotype-Tissue Expression (GTEx) Project (<https://gtexportal.org/home/>) and 10× Genomics (https://support.10xgenomics.com/single-cell-gene-expression/datasets/1.3.0/1M_neurons) to assess gene expression for top gene(s) across human tissues and across individual mouse brain cell types. GTEx is supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from the GTEx Portal on July 12, 2017.

Hierarchical clustering

Methylation differences from the top 2,000 sites in c9ALS and sALS frontal cortex and cerebellum contrast by magnitude were clustered using complete-linkage clustering and Euclidian distance.

Data Availability

We deposited the RRBS data in the NCBI Gene Expression Omnibus (GEO) database under accession number # GSE97106.

Results

Differentially methylated cytosines are abundant in ALS patient brains

We found thousands of differentially methylated cytosines (DMCs) in the frontal cortex and cerebellum of c9ALS and sALS cases, when compared to control brains. Overall, there were 4,430 total frontal cortex DMCs in c9ALS patients, where 1,487 were in a gene body and 2,943 were upstream of a transcription start site (TSS) (Fig. 1a). There were 3,721 frontal cortex DMCs in sALS patients, where 1,267 were in a gene body and 2,454 were upstream of a TSS (Fig. 1a). In the cerebellum, there were 4,454 (1,330 + 3,124) and 6,412 (2,100 + 4,312) across c9ALS and sALS (Fig. 1b), respectively. In strong contrast, DMCs were least abundant in sALS frontal cortex and most abundant in sALS cerebellum. DMC abundance was similar between c9ALS frontal cortex and cerebellum.

Approximately 66% of DMCs were located upstream of a TSS in the frontal cortex for both c9ALS and sALS, while the remaining 34% were located in a gene body. While proportions were similar in sALS cerebellum, 70% of c9ALS cerebellum DMCs were upstream of a TSS. Of the 66-70% of DMCs located upstream of the TSS, we observed that DMCs in the

frontal cortex were more frequent in c9ALS (2,943) than sALS (2,454; Fig. 1c), while DMCs in the cerebellum were much more abundant in sALS (4,312) than c9ALS (3,124; Fig. 1d). A majority of DMCs, ranging from 95-96%, were located >3kb upstream of the TSS, and DMCs were predominantly located in the cerebellum for both disease groups. Among the upstream DMCs identified, more than 100 were located within 3kb of a TSS for each of the four groups (Fig. 1c and 1d). Among these, more than half were found less than 1.5kb away, a region known to play a crucial role in gene regulation.

c9ALS and sALS patients may have distinct, but overlapping brain methylome profiles

While most upstream DMCs were unique to a certain disease group or brain region, some were also shared between c9ALS and sALS in both frontal cortex (n=385; Fig. 1e) and cerebellum (n=576; Fig. 1h), affecting 687 and 1,014 genes respectively. Some DMCs were also shared between the two brain regions in c9ALS (n=737; Fig. 1f) and sALS (n=673; Fig. 1g), affecting 973 and 954 genes respectively.

Most of the genes with upstream DMCs encode proteins (69-70%), while 13-14% of the total number of genes affected comprise non-coding RNAs (ncRNAs; Fig. 1i and 1j). Approximately 5-6% of the genes consisted of small RNA transcripts (sRNAs) such as miRNAs and PIWI-associated RNAs (piRNAs), whereas the remaining 11-12% of genes for each disease group and brain region mostly consisted of other unclassified genes. A small proportion of the affected regions were in transposable elements and genes encoding RNA-binding proteins (Fig 1i and 1j).

To evaluate the similarity of the methylome profiles included in each disease group, we conducted hierarchical clustering of the top 2,000 upstream DMCs found in c9ALS and the top 2,000 upstream DMCs identified in sALS in both frontal cortex and cerebellum. We observed that samples clustered perfectly per group (Fig. 1k and 1l), suggesting that between-group methylomes are distinct from one another and within-group methylomes are highly similar.

Combined methylation and transcriptome data identify genes with both DMCs and differential transcriptome defects

Combining data between the RRBS and our previous RNASeq data[33], we identified differentially-expressed genes in ALS that also have DMCs. DMCs were less frequent in differentially-expressed genes for c9ALS frontal cortex (n=77, Fig. 2a), and sALS frontal cortex (n=21, Fig. 2b), compared to c9ALS cerebellum (n=91, Fig. 2c) and sALS cerebellum (n=43, Fig. 2d). For both c9ALS and sALS frontal cortex and sALS cerebellum, the most common types of transcriptome changes in genes with DMCs were differential expression and alternative poly-A events, while alternative cassette exon splicing was the most common in c9ALS cerebellum (Fig. 2a-2d).

We also evaluated the position of DMCs affecting genes with transcriptome defects (Fig. 2a-2d). While many DMCs were located in regions predicted to modulate transcription, most DMCs were located farther away or outside these regions. Regions typically predicted to modulate transcription include regions within 3kb of a TSS or in 5'UTR for differentially expressed genes (Fig. 2a-2d, light blue), in the 3'UTR for alternative poly-A (light purple),

or within an alternatively spliced region for cassette exons (light green) and intronic region retention (light orange).

***SERPINA1* and other genes are differentially expressed and methylated in the cerebellum and frontal cortex**

We selected top differentially expressed genes with significant DMCs that were common across brain regions or disease states to validate (Fig. 2a-2d, protruding pie slices). The gene *SERPINA1* had a significant *increase* in differential expression (DE) and significant *decrease* in differential methylation (DM) in the frontal cortex for sALS (Fig. 3a; DE vs. control: $p = 9.14 \times 10^{-3}$, $\log_2FC = 1.69$; DM vs. control: $q = 2.49 \times 10^{-6}$, percent difference = -25.54), and in the cerebellum for both c9ALS (DE vs. control: $p = 6.38 \times 10^{-4}$, $\log_2FC = 2.29$; DM vs. control: $q = 1.31 \times 10^{-8}$, percent difference = -29.40) and sALS (DE vs. control: $p = 4.67 \times 10^{-2}$, $\log_2FC = 1.26$; DM vs. control: $q = 2.37 \times 10^{-10}$, percent difference = -33.39). We confirmed *SERPINA1* expression is increased in the frontal cortex for sporadic cases ($p = 0.036$, Fig. 3b), when compared to controls. Additionally, we found *C9orf72* expansion carriers also have increased *SERPINA1* expression ($p = 0.001$, Fig. 3b), when compared to controls. We also confirmed increased *SERPINA1* expression in the cerebellum for both sporadic ($p = 0.048$) and *C9orf72* expansion carriers ($p = 0.007$), when compared to controls in our expanded validation cohort (Fig. 3c). *SERPINA1* methylation at position 94,849,201 on chromosome 14 is also correlated with *SERPINA1* expression ($p = 0.051$) in our expanded validation cohort.

After bisulfite DNA amplification of the *SERPINA1* region of interest, we observed fully unconverted methylated cytosines in all but one control samples, whereas fully converted, unmethylated cytosines were observed in *C9orf72* expansion carriers (Figure 3d). Both methylation states were found in eight *C9orf72* expansion carriers and non-carriers.

To test whether *SERPINA1* is commonly expressed in mouse brain, we investigated several existing public data sets that demonstrate *SERPINA1* is specifically expressed in the liver and blood (Fig. 3e; <https://gtexportal.org/home/gene/SERPINA1>, accessed July 2017). *SERPINA1* is largely unexpressed in both the human (Fig. 3e) and mouse brain (online resource Fig. 1); mouse homologue of the *SERPINA1* transcripts (*Serpina1a* through *Serpina1f*) were measured in only approximately four E18 mouse brain cells in the 1.3million cell 10× Genomics single-cell expression data set (18,500 reads per cell; https://support.10xgenomics.com/single-cell-gene-expression/datasets/1.3.0/1M_neurons; accessed July 2017), while *SERPINA1* transcripts were measured in zero E18 mouse brain cells in the 9k single-cell expression dataset (42,000 reads per cell; https://support.10xgenomics.com/single-cell-gene-expression/datasets/1.3.0/neuron_9k).

We also attempted to replicate twelve other genes that were both differentially expressed and methylated, but were not in common across brain regions or disease groups (online resource Table 1). Three of the twelve genes could not be assessed because of very low expression. Four genes were differentially expressed in our expanded cohort (online resource Fig. 2a-2d). As expected, *DCAF4L1* ($p = 0.017$) and *VAMP5* ($p = 0.026$) were over expressed in the cerebellum for *C9orf72* expansion carriers when compared to controls, and *COL27A1* ($p = 0.017$) and *PODNL1* ($p = 0.003$) were over expressed for *C9orf72* expansion carriers in

the frontal cortex, when compared to controls. *DCAF4L1* and *COL27A C9orf72* expansion carriers were also over expressed compared to sporadic cases in cerebellum and frontal cortex, respectively ($p = 0.026$ and 0.035 ; online resource Fig. 2a and 2c), while *VAMP5* and *PODNL1* sporadic cases were also over expressed when compared to controls in cerebellum and frontal cortex, respectively ($p = 0.040$ and 0.011 ; online resource Fig. 2b and 2d). The other five genes (*H19*, *CDKN1A*, *CIRBP-AS1*, *TJP3*, and *LRG1*) did not reach significance in our expanded validation cohort.

Genes exhibit alternative cassette exon splicing and methylation

We also tested five genes that exhibited significantly different alternative cassette exon splicing and methylation: *FBLN2*, *C14orf80*, *IFI27*, *TRPM2*, and *NAGPA*. Two of the genes, *C14orf80* and *TRPM2*, replicated. As expected, both *C14orf80* and *TRPM2* demonstrated significantly less inclusion of exons 6 ($p = 0.001$; transcript variant 1, NM_001134875.1; online resource Table 2 and online resource Fig. 2e) and 27 ($p = 0.007$; transcript variant 1, NM_003307.3; online resource Table 2 and online resource Fig. 2f), respectively, in the frontal cortex for *C9orf72* expansion carriers, when compared to controls. Exon 6 of *C14orf80* was also included less in sporadic cases ($p = 0.042$; online resource Fig. 2e), when compared to controls.

Common biological functions in ALS

Considering the distinct methylomes identified in c9ALS and sALS, we assessed whether differentially-expressed genes found with DMCs upstream of TSS or within the gene-body in each group (1) have similar or distinct cellular and molecular gene functions and (2) are part of similar or distinct biological networks. Using Ingenuity Pathway Analysis (IPA) software, we evaluated the cellular and molecular functions of differentially-expressed genes ($p < 0.05$) that also have DMCs, and the functions of their associated biological networks (online resource, Tables 3-6) in c9ALS and sALS frontal cortex and cerebellum. Of the top five most significantly affected cellular and molecular gene functions in each group, 2-4 were shared with at least one other brain region or disease group, including “cell cycle” and cellular function and maintenance, movement, morphology, assembly, and organization.

Top affected biological networks were identified in each disease group and brain region based on the significance of aberrant methylation upstream of genes involved in associated networks. The commonly top-affected network functions, considering all affected genes together, were highly similar to the individual molecular and cellular gene functions previously identified (online resource, Tables 3-6). Of the top ten hyper- and hypomethylated genes for each disease group, 60% and 40% were part of the top five affected networks for c9ALS and sALS across both brain regions, respectively (online resource, Table 7). These networks are connected to other networks containing genes with methylation and expression changes. Therefore, a change in a gene can affect not only members of the network to which it belongs, but also members of connected networks, expanding the impact range of a given defect (online resource, Tables 4 and 6).

***SERPINA1* contributes to many critical cellular and molecular functions, and interacts with ALS-associated molecules**

The most significant network in c9ALS cerebellum includes *SERPINA1*, and is also part of networks 2 and 4 (Fig. 3a, online resource Table 4 and 6). All three networks are interconnected through *SERPINA1* and have “cellular movement” as a common biological function. Within network 1, there are seven molecules previously associated with ALS, including *FUS*, VEGF, histones H3 and H4, ACTIN, estrogen receptor, and NFκB (Fig. 4). The network also includes *KHDRBS1*, a gene encoding SAM68, which is sequestered by CGG-containing RNA foci in Fragile X-associated tremor/ataxia syndrome (FXTAS). Among members of the same network, *CDC25B*, *FBLN2*, and *PPP2R2A* were part of the top ten hyper-methylated genes, while *ALDH1A1*, *COL4A2*, *LIG1*, and *PRMT6* were part of the top hypo-methylated genes. Five additional genes in network 1 (*FBLN2*, *GSN*, *H19*, *POLE*, and *TNFAIP8*) were differentially expressed and differentially methylated within the region expected to modulate transcription (Fig. 2a-2d, part of light-blue protruding pie slices).

Key ALS-associated genes appear to accumulate transcriptome and methylome changes

While additional studies are necessary to clarify, we nonetheless observed transcriptome and methylome changes in key ALS genes. We analyzed DNA methylation and transcriptome defects in 172 candidate genes (76 ALS-associated and 96 neuro-associated genes) carefully selected based on their involvement in ALS and other neurological diseases (full list of genes and complete results available in online resource, Tables 8-11). We found that, while transcriptome defects were not common in these selected genes, except for cassette exons in c9ALS cerebellum, DNA methylation events were much more frequent (Table 1). We identified an upstream DMC in *C9orf72* for both c9ALS frontal cortex and cerebellum. We also identified upstream DMCs for *TARDBP* in both c9ALS and sALS, in frontal cortex and cerebellum. *RANGAP1* also has upstream DMCs in c9ALS and sALS frontal cortex, and in sALS cerebellum, as well as in the *RANGAP1* gene body in c9ALS frontal cortex (Table 1). *IGHMBP2*, involved in Charcot-Marie-Tooth disease, and *KHDRBS1*, involved in FXTAS, were also found to have upstream DMCs in frontal cortex and cerebellum in both c9ALS and sALS. Gene body DMCs were also found in *IGHMBP2* for c9ALS and sALS frontal cortex and cerebellum, further supporting a role for this gene in ALS (Table 1).

Discussion

The methylation data presented herein provides a comprehensive methylation overview between controls, sALS, and c9ALS patients across the frontal cortex and cerebellum. We also combined it with our transcriptome data to identify and validate new genes that may be heavily involved in ALS. To our knowledge, no other report broadly addresses methylation and transcriptome effects in sALS and c9ALS across two important brain regions. While the sample sizes for our methylation data are small, the general trends are still valuable, and assessing both overlapping and distinct DMCs, combined with transcriptome data, will help clarify what drives sALS vs c9ALS. Our data also reports individual cytosine methylation events, where most studies combine methylation events across large regions. Characterizing

individual methylation events will be critical to determining what role each methylation event plays in gene regulation, as some events are likely far more influential than others.

Increasing evidence suggests cerebellar dysfunction plays a role in ALS

The broad methylation and expression overview we presented suggests generally distinct, but overlapping methylome profiles between sALS and c9ALS, where disease phenotypes may converge on common biological mechanisms or pathways, and also supports increasing evidence that the cerebellum plays a role in ALS[33,12]. We identified top DMCs and differentially expressed genes for sALS and c9ALS when compared to controls that distinguish between groups, and we identified approximately 800 more DMCs in the cerebellum of sALS cases compared to c9ALS patients—nearly twice as many DMCs as was found in sALS frontal cortex (Fig. 1c and 1d). There were many overlapping DMCs and differentially expressed genes, however. These data support cerebellar involvement in disease, and further suggest that, while sALS and c9ALS methylome and expression profiles are largely distinct, there may be some genes in common that play a central role in disease.

The cerebellum has not generally been considered important in ALS since there is no substantial neuronal loss[32,39], but cell death may not be the only mechanism driving ALS phenotypes. For example, cellular functions may be inhibited without actually killing the cells, and the excessive number of DMCs in the cerebellum may simply indicate general gene dysregulation in the cerebellum that might modulate downstream effects in other brain regions.

Aberrant DMCs may be a common link amongst sALS cases

While it is unsurprising that sALS had many more DMCs than c9ALS, given its sporadic nature, it was unexpected that many common DMCs were within sALS. These data suggest that while there is currently no known common cause for sALS, there may actually be a core set of biological pathways that explain this group of patients (Fig. 1k and 1l). More studies are necessary to fully assess these common features, but these results may have big implications in sALS disease therapies.

Both proximal and distal DMCs may affect the transcriptome

Surprisingly, approximately 95% of all DMCs were located >3kb upstream of the affected gene's TSS, providing some support that distal DMCs may regulate transcriptome modifications. While proximal epigenetic modifications are the most likely cause for a given gene's altered transcription, DMCs that modulate transcription have been reported as distant as 1Mbp[37], which our data further supports. Just as a few critical DMCs may be more influential than an abundance of trivial DMCs to modulate expression, methylation of key sites located more than 3kb away from the TSS may be as imperative as changes closer to the TSS. Precisely, our results support a role for DNA methylation at distant regulatory regions, such as at enhancer sites, which are responsible for transcription regulation through transcription factor-binding[30].

Epigenetics is still a relatively young field, however, and additional work is necessary to clarify what role distant DMCs may have on transcriptome regulation. Similarly, most work

to date has focused on DMCs within a region as a whole, rather than characterizing how influential *individual* DMCs (proximal or distal) are on a gene's transcription regulation; more work is needed to clarify the individual DMC contributions.

Brain cells may not tolerate increased *SERPINA1* expression

While the comprehensive methylation and transcription data are individually valuable, their true utility is best exploited together as a combined “multi-omics” analysis. As described, we filtered our candidate gene list by including only those that were both differentially expressed and methylated, and identified *SERPINA1* as a potentially critical disease gene. Our data suggests *SERPINA1* is involved across both sporadic cases and *C9orf72* expansion carriers in frontal cortex and cerebellum (Fig. 3a), where overexpression is particularly pronounced in *C9orf72* expansion carriers (Fig. 3b and 3c). While most genes are expressed in easily measureable abundance across body tissues, *SERPINA1* expression is strictly down regulated throughout the body (including all major brain regions) except in the liver and blood (Fig. 3e). We further verified this finding in public 10x Genomics single-cell expression experiments in E18 mouse brain (online resource Fig 1). We verified this finding in both the 1.3 million and 9,000 single-cell 10x Genomics mouse data sets with 18,500 and 42,000 reads per cell, respectively. While the single-cell sequencing experiments may not sequence enough reads to accurately quantify the low *SERPINA1* levels, it does emphasize how tightly *SERPINA1* is regulated in the brain. Given how strictly *SERPINA1* is regulated throughout the brain and body, over expression may be detrimental to normal neuronal function.

SERPINA1 encodes a serine protease inhibitor that interacts primarily with elastase, but also with trypsin, chymotrypsin, thrombin, and bacterial proteases[9]. Our previous data demonstrated members of the same gene family are also differentially expressed, namely *SERPINH1* (increased in c9ALS frontal cortex, $p = 0.0017$, and cerebellum, $p = 0.0250$) and *SERPINA3* (increased in c9ALS frontal cortex, $p = 0.0085$, and trending towards significance in cerebellum, $p = 0.0527$)[33]. All three genes are serine protease inhibitors, and *SERPINA3* has also been implicated in Alzheimer's[18,19] and Parkinson's disease[41]. Given that all three genes have increased expression may indicate serine protease inhibitors are obstructing neuronal function in disease.

DCAF4L1, *VAMP5*, *COL27A1*, and *PODNL1* are also implicated in disease

Differential expression for genes *DCAF4L1*, *VAMP5*, *COL27A1*, and *PODNL1* was also previously implicated in our RNASeq data, and we found DMCs in regions predicted to regulate their expression. Differential expression replicated for *DCAF4L1*, *VAMP5*, *COL27A1*, and *PODNL1* (online resource Fig. 2a-2d). *DCAF4L1* and *VAMP5* were differentially expressed in the cerebellum, while *COL27A1* and *PODNL1* were implicated and validated in the frontal cortex. Little is known about *DCAF4L1* and *PODNL1*, but *VAMP5* is associated with muscle formation (myogenesis) and vesicle tracking[46]. *COL27A1* is in the fibrillar collagen family[29] which is important in tissue growth and repair. *COL27A1* is also highly expressed in the brain, when compared to other body tissues, and most highly expressed in the cerebellum, compared to other brain tissues (<https://gtexportal.org/home/gene/COL27A1>, accessed July 2017)[27]. Nagase et al. further

demonstrated that *COL27A1* (referred to as *KIAA1870* in their paper) has a primary function in cellular structure and motility[27]. Collagen binding was also one of the top upregulated molecular functions identified in our RNAseq Gene Ontology study for c9ALS frontal cortex[33].

Methylation modifications may be driving *TARDBP* and *RANGAP1* dysregulation

We identified *TARDBP* and *RANGAP1* as two commonly differentially methylated ALS associated genes. *TARDBP* encodes TDP-43, a key RNA-binding protein known to aggregate in ALS neurons after mislocalizing from the nucleus to the cytoplasm[28]. RanGAP physically interacts with *C9orf72* expanded RNA[10] and is believed to be a key regulator of nucleocytoplasmic transport in c9ALS[47]. Since approximately 97% of ALS cases have TDP-43 pathology[28], future studies should assess whether modulating aberrant *TARDBP* methylation has functional consequences on downstream TDP-43 cytoplasmic aggregation. Also, while alterations in RanGAP function have been reported in c9ALS[47], our data suggest an unpredicted role for RanGAP in sALS as well, both in frontal cortex and cerebellum. In addition to these ALS associated genes, we also found DMCs in *IGHMBP2*, a gene associated with other neurological diseases, in both cerebellum and frontal cortex tissue from both c9ALS and sALS patients. *IGHMBP2* encodes the DNA-binding protein SMBP2, a transcription regulator that predominantly localizes to the cytoplasm of neurons and associates with ribosomes. SMBP2 is an ATP-dependent 5' to 3' helicase, responsible for unwinding RNA and DNA duplexes. Our analyses of ALS and neuro-associated genes also further support our findings from the upstream and gene body analyses. Our data points to a complex biological network as being affected in both c9ALS and sALS in which genes found with both altered transcription and aberrant methylation can modulate multiple interconnected genes and molecules within individual networks, all together modulating the biological function of the network as a whole.

Conclusions

ALS is a devastating and fatal disease with no effective therapy to prevent, stop or decelerate neurodegeneration, making comprehensive “multi-omic” approaches critical to identifying key molecular players in disease etiology. Our study is the first to demonstrate that c9ALS and sALS patients have distinct methylomes, with members within each disease group sharing thousands of DMCs, and the first to combine methylome and expression data for a “multi-omic” analysis in ALS. Our results are particularly important for genetically unexplained sALS cases, as there is currently no known common cause. Aberrant methylation may be key to understanding the disease.

Our data also strongly implicates *SERPINA1* in both c9ALS and sALS, as it was differentially methylated and expressed in both disease groups across the frontal cortex and cerebellum. It also interacts with many ALS-associated genes and is strictly down regulated throughout the body, under normal conditions. We propose that functional studies in mouse and cell models will be necessary to determine whether *SERPINA1* is driving ALS phenotypes, or whether its dysregulation is one of many downstream effects of the underlying cause.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are extremely grateful to all individuals who agreed to donate their brains to research. This study was supported by the National Institutes of Health/National Institute on Aging [AG16574-17J PILOT (V.V.B.)]; National Institutes of Health/National Institute of Neurological Disorders and Stroke [R21NS074121 (K.B.B.), P01NS084974 (D.W.D., K.B.B., and R.R.)]; Mayo Clinic Center for Individualized Medicine (V.V.B., K.B.B., and H.L.); ALS Association (K.B.B.), Donors Cure Foundation (H.L.), and the Robert Packard Center for ALS Research at Johns Hopkins. V.V.B. is recipient of the Career Transition Award from ALS Canada and Brain Canada, the Milton Safenowitz Post-Doctoral Fellowship from the Amyotrophic Lateral Sclerosis Association, the Post-Doctoral Fellowship from the Canadian Institutes of Health Research, the Career Development Award for Young Investigators in Neurosciences from the Siragusa Foundation, and the Research Fellowship from the Robert and Clarice Smith & Abigail Van Buren Alzheimer's Disease Research Foundation. M.T.W.E received the P/RMA Foundation Research Starter grant.

References

1. Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* 2012; 13:R87.doi: 10.1186/gb-2012-13-10-r87[pil] [PubMed: 23034086]
2. Amaral PP, Dinger ME, Mercer TR, Mattick JS. The eukaryotic genome as an RNA machine. *Science.* 2008; 319:1787–1789. DOI: 10.1126/science.1155472.319/5871/1787[pil] [PubMed: 18369136]
3. Belzil VV, Bauer PO, Gendron TF, Murray ME, Dickson D, Petrucelli L. Characterization of DNA hypermethylation in the cerebellum of c9FTD/ALS patients. *Brain Res.* 2014; 1584:15–21. DOI: 10.1016/j.brainres.2014.02.015 [PubMed: 24530272]
4. Belzil VV, Bauer PO, Prudencio M, Gendron TF, Stetler CT, Yan IK, et al. Reduced C9orf72 gene expression in c9FTD/ALS is caused by histone trimethylation, an epigenetic event detectable in blood. *Acta Neuropathol.* 2013; 126:895–905. DOI: 10.1007/s00401-013-1199-1 [PubMed: 24166615]
5. Belzil VV, Gendron TF, Petrucelli L. RNA-mediated toxicity in neurodegenerative disease. *Mol Cell Neurosci.* 2012; 56C:406–419. doi:S1044-7431(12)00224-2[pil]10.1016/j.mcn.2012.12.006.
6. Belzil VV, Katzman RB, Petrucelli L. ALS and FTD: an epigenetic perspective. *Acta Neuropathol.* 2016; 10.1007/s00401-016-1587-4[pil]. doi: 10.1007/s00401-016-1587-4
7. Byrne S, Heverin M, Elamin M, Bede P, Lynch C, Kenna K, et al. Aggregation of neurologic and neuropsychiatric disease in amyotrophic lateral sclerosis kindreds: a population-based case-control cohort study of familial and sporadic amyotrophic lateral sclerosis. *Ann Neurol.* 2013; 74:699–708. DOI: 10.1002/ana.23969 [PubMed: 23836460]
8. Byrne S, Walsh C, Lynch C, Bede P, Elamin M, Kenna K, et al. Rate of familial amyotrophic lateral sclerosis: a systematic review and meta-analysis. *J Neurol Neurosurg Psychiatry.* 2011; 82:623–627. DOI: 10.1136/jnnp.2010.224501 [PubMed: 21047878]
9. Cox, DW. Alpha-1-antitrypsin. In: Scriver, CR.B, AL.Sly, WS., Valle, D., editors. *The Metabolic and Molecular Bases of Inherited Disease*. 8th. Vol. IV. McGraw-Hill; New York: 2001. p. 5559-5584.
10. Donnelly CJ, Zhang PW, Pham JT, Haeusler AR, Mistry NA, Vidensky S, et al. RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. *Neuron.* 2013; 80:415–428. DOI: 10.1016/j.neuron.2013.10.015 [PubMed: 24139042]
11. Fratta P, Mizielińska S, Nicoll AJ, Zloh M, Fisher EM, Parkinson G, et al. C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. *Sci Rep.* 2012; 2:1016.doi: 10.1038/srep01016 [PubMed: 23264878]
12. Gendron TF, van Blitterswijk M, Bieniek KF, Daugherty LM, Jiang J, Rush BK, et al. Cerebellar c9RAN proteins associate with clinical and neuropathological characteristics of C9ORF72 repeat

- expansion carriers. *Acta Neuropathol.* 2015; 130:559–573. 10.1007/s00401-015-1474-4 [pii]. DOI: 10.1007/s00401-015-1474-4 [PubMed: 26350237]
13. Gibson SB, Figueroa KP, Bromberg MB, Pulst SM, Cannon-Albright L. Familial clustering of ALS in a population-based resource. *Neurology.* 2014; 82:17–22. DOI: 10.1212/01.wnl.0000438219.39061.da [PubMed: 24306004]
 14. Guerreiro R, Bras J, Hardy J. SnapShot: Genetics of ALS and FTD. *Cell.* 2015; 160:798e791.doi: 10.1016/j.cell.2015.01.052S0092-8674(15)00130-0[pii] [PubMed: 25679767]
 15. Haeusler AR, Donnelly CJ, Periz G, Simko EA, Shaw PG, Kim MS, et al. C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature.* 2014; 507:195–200. nature13124 [pii]. DOI: 10.1038/nature13124 [PubMed: 24598541]
 16. Holloch D, Moazed D. RNA-mediated epigenetic regulation of gene expression. *Nat Rev Genet.* 2015; 16:71–84. nrg3863 [pii]. DOI: 10.1038/nrg3863 [PubMed: 25554358]
 17. Jung J, Bonini N. CREB-binding protein modulates repeat instability in a Drosophila model for polyQ disease. *Science.* 2007; 315:1857–1859. doi:1139517[pii]10.1126/science.1139517. [PubMed: 17332375]
 18. Kalsheker NA. Alpha 1-antichymotrypsin. *Int J Biochem Cell Biol.* 1996; 28:961–964. [PubMed: 8930118]
 19. Kamboh MI, Minster RL, Kenney M, Ozturk A, Desai PP, Kammerer CM, et al. Alpha-1-antichymotrypsin (ACT or SERPINA3) polymorphism may affect age-at-onset and disease duration of Alzheimer's disease. *Neurobiol Aging.* 2006; 27:1435–1439. DOI: 10.1016/j.neurobiolaging.2005.07.015 [PubMed: 16137793]
 20. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics.* 2011; 27:1571–1572. DOI: 10.1093/bioinformatics/btr167 [PubMed: 21493656]
 21. Lam L, Chin L, Halder RC, Sagong B, Famenini S, Sayre J, et al. Epigenetic changes in T-cell and monocyte signatures and production of neurotoxic cytokines in ALS patients. *FASEB J.* 2016 doi:fj.201600259RR[pii]10.1096/fj.201600259RR.
 22. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012; 9:357–359. DOI: 10.1038/nmeth.1923 [PubMed: 22388286]
 23. Liu EY, Russ J, Wu K, Neal D, Suh E, McNally AG, et al. C9orf72 hypermethylation protects against repeat expansion-associated pathology in ALS/FTD. *Acta Neuropathol.* 2014; 128:525–541. DOI: 10.1007/s00401-014-1286-y [PubMed: 24806409]
 24. Mattick JS, Amaral PP, Dinger ME, Mercer TR, Mehler MF. RNA regulation of epigenetic processes. *Bioessays.* 2009; 31:51–59. DOI: 10.1002/bies.080099 [PubMed: 19154003]
 25. Morahan JM, Yu B, Trent RJ, Pamphlett R. A genome-wide analysis of brain DNA methylation identifies new candidate genes for sporadic amyotrophic lateral sclerosis. *Amyotroph Lateral Scler.* 2009; 10:418–429. 10.3109/17482960802635397 [pii]. DOI: 10.3109/17482960802635397 [PubMed: 19922134]
 26. Morris KV, Mattick JS. The rise of regulatory RNA. *Nat Rev Genet.* 2014; 15:423–437. nrg3722 [pii]. DOI: 10.1038/nrg3722 [PubMed: 24776770]
 27. Nagase T, Nakayama M, Nakajima D, Kikuno R, Ohara O. Prediction of the coding sequences of unidentified human genes. XX. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. *DNA Res.* 2001; 8:85–95. [PubMed: 11347906]
 28. Neumann M, Sampathu DM, Kwong LK, Truax AC, Micsenyi MC, Chou TT, et al. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science.* 2006; 314:130–133. [PubMed: 17023659]
 29. Pace JM, Corrado M, Missero C, Byers PH. Identification, characterization and expression analysis of a new fibrillar collagen gene, COL27A1. *Matrix Biol.* 2003; 22:3–14. [PubMed: 12714037]
 30. Pennacchio LA, Bickmore W, Dean A, Nobrega MA, Bejerano G. Enhancers: five essential questions. *Nat Rev Genet.* 2013; 14:288–295. DOI: 10.1038/nrg3458nrg3458[pii] [PubMed: 23503198]
 31. Polymenidou M, Lagier-Tourenne C, Hutt KR, Bennett CF, Cleveland DW, Yeo GW. Misregulated RNA processing in amyotrophic lateral sclerosis. *Brain Res.* 2012; 1462:3–15. DOI: 10.1016/j.brainres.2012.02.059S0006-8993(12)00391-5[pii] [PubMed: 22444279]

32. Prell T, Grosskreutz J. The involvement of the cerebellum in amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Frontotemporal Degener.* 2013; 14:507–515. DOI: 10.3109/21678421.2013.812661 [PubMed: 23889583]
33. Prudencio M, Belzil VV, Batra R, Ross CA, Gendron TF, Prent J, et al. Distinct brain transcriptome profiles in C9orf72-associated and sporadic ALS. *Nat Neurosci.* 2015; 18:1175–1182. DOI: 10.1038/nn.4065 [pii] [PubMed: 26192745]
34. Rademakers R, van Blitterswijk M. Motor neuron disease in 2012: Novel causal genes and disease modifiers. *Nat Rev Neurol.* 2013; 9:63–64. nrneurol.2012.276 [pii]. DOI: 10.1038/nrneurol.2012.276 [PubMed: 23318296]
35. Renoux AJ, Todd PK. Neurodegeneration the RNA way. *Prog Neurobiol.* 2012; 97:173–189. doi:S0301-0082(11)00193-6 [pii]10.1016/j.pneurobio.2011.10.006. [PubMed: 22079416]
36. Renton AE, Chio A, Traynor BJ. State of play in amyotrophic lateral sclerosis genetics. *Nat Neurosci.* 2014; 17:17–23. nn.3584 [pii]. DOI: 10.1038/nn.3584 [PubMed: 24369373]
37. Spitz F. Gene regulation at a distance: from remote enhancers to 3D regulatory ensembles. *Semin Cell Dev Biol.* 2016 doi:S1084-9521(16)30178-1 [pii]10.1016/j.semcdb.2016.06.017.
38. Su Z, Zhang Y, Gendron TF, Bauer PO, Chew J, Yang WY, et al. Discovery of a biomarker and lead small molecules to target r(GGGGCC)-associated defects in c9FTD/ALS. *Neuron.* 2014; 83:1043–1050. DOI: 10.1016/j.neuron.2014.07.041 [PubMed: 25132468]
39. Tan RH, Kril JJ, McGinley C, Hassani M, Masuda-Suzukake M, Hasegawa M, et al. Cerebellar neuronal loss in amyotrophic lateral sclerosis cases with ATXN2 intermediate repeat expansions. *Ann Neurol.* 2016; 79:295–305. DOI: 10.1002/ana.24565 [PubMed: 26599997]
40. Therrien M, Dion PA, Rouleau GA. ALS: Recent Developments from Genetics Studies. *Curr Neurol Neurosci Rep.* 2016; 16:59. 10.1007/s11910-016-0658-1 [pii]. doi: 10.1007/s11910-016-0658-1 [PubMed: 27113253]
41. Wang YC, Liu HC, Liu TY, Hong CJ, Tsai SJ. Genetic association analysis of alpha-1-antichymotrypsin polymorphism in Parkinson's disease. *Eur Neurol.* 2001; 45:254–256. doi:52138. [PubMed: 11385264]
42. Xi Z, Rainero I, Rubino E, Pinessi L, Bruni AC, Maletta RG, et al. Hypermethylation of the CpG-island near the C9orf72 G4C2-repeat expansion in FTLN patients. *Hum Mol Genet.* 2014 doi:ddu279 [pii]10.1093/hmg/ddu279.
43. Xi Z, Yunusova Y, van Blitterswijk M, Dib S, Ghani M, Moreno D, et al. Identical twins with the C9orf72 repeat expansion are discordant for ALS. *Neurology.* 2014; 83:1476–1478. DOI: 10.1212/WNL.0000000000000886 [PubMed: 25209579]
44. Xi Z, Zhang M, Bruni AC, Maletta RG, Colao R, Fratta P, et al. The C9orf72 repeat expansion itself is methylated in ALS and FTLN patients. *Acta Neuropathol.* 2015; 129:715–727. DOI: 10.1007/s00401-015-1401-8 [PubMed: 25716178]
45. Xi Z, Zinman L, Moreno D, Schymick J, Liang Y, Sato C, et al. Hypermethylation of the CpG island near the G4C2 repeat in ALS with a C9orf72 expansion. *Am J Hum Genet.* 2013; 92:981–989. DOI: 10.1016/j.ajhg.2013.04.017 [PubMed: 23731538]
46. Zeng Q, Subramaniam VN, Wong SH, Tang BL, Parton RG, Rea S, et al. A novel synaptobrevin/VAMP homologous protein (VAMP5) is increased during in vitro myogenesis and present in the plasma membrane. *Mol Biol Cell.* 1998; 9:2423–2437. [PubMed: 9725904]
47. Zhang K, Donnelly CJ, Haeusler AR, Grima JC, Machamer JB, Steinwald P, et al. The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature.* 2015; 525:56–61. DOI: 10.1038/nature14973 [PubMed: 26308891]

Fig. 1. c9ALS and sALS have distinct brain methylome profiles

(a, b) Pie charts representing the proportion of DMCs identified upstream of a TSS (darker color) or in the body of a gene (lighter color) in c9ALS (blue; n=4) and sALS (yellow; n=4), for both frontal cortex (a) and cerebellum (b). (c, d) Pie charts representing the total number of DMCs located upstream of a TSS in the frontal cortex (c) and cerebellum (d) for c9ALS (n=4) and sALS (n=4), and their distance from the nearest TSS. (e, h) Pie charts representing DMCs upstream of the TSS unique to c9ALS (blue; n=4) and sALS (yellow; n=4), or shared between the two disease groups, in frontal cortex (e) and cerebellum (h). (f, g) Pie charts representing upstream of TSS DMCs unique to frontal cortex (left) and cerebellum (right), or shared between the two brain regions, in c9ALS (n=4) and sALS (n=4). (i, j) Pie charts representing the type of genes affected by upstream of TSS DMCs in frontal cortex (i) and cerebellum (j). (k, l) Hierarchical clustering of genes with the top 2,000 upstream of TSS DMCs in c9ALS (n=4) and sALS (n=4) (4,000 total), in frontal cortex (k) and cerebellum (l).

Fig. 2. DMCs and transcriptome defects are abundant in ALS

(a-d) The darker left pie charts show the total number of genes found with differential expression (DE) (blue; $p < 0.05$; $|\log_2FC| \geq 1$), alternative cassette exon (CE) splicing (green; $FDR < 0.05$; $|dI| \geq 0.1$), intron retention (IR) (orange; $FDR < 0.05$; $|dI| \geq 0.1$), and alternative polyadenylation (APA) (purple; $FDR < 0.05$; $|PDUI| \geq 0.2$; $|dPDUI| \geq 0.2$) in our RNAseq for c9ALS (a, c; $n=8$) and sALS (b, d; $n=10$) when compared to control subjects ($n=9$ in frontal cortex; $n=8$ in cerebellum), in both frontal cortex (a, b), and cerebellum (c, d). The lighter right pie charts (a-d) represent the proportion of genes from each transcriptome defect category with DMCs. For each of these events, the protruding slices of each color represent DMCs in the predicted effective region. Of note, numbers for genes with DE combine both upstream of TSS and 5'UTR DMCs, while numbers for genes with CE, IR and APA are for gene body DMCs.

Fig. 3. *SERPINA1* is over-expressed in *C9orf72* repeat carriers and non-carriers brains
(a) Summary of data obtained after RNAseq, RRBS and systems biology analyses for *SERPINA1* in c9ALS and sALS frontal cortex and cerebellum. **(b-c)** Validation of *SERPINA1* over-expression in frontal cortex (b) and cerebellum (c). Validation cohort included ALS, ALS-FTD, and FTD cases for both disease groups. Quantifications were performed for *C9orf72* repeat carriers (*C9orf72*, blue) and non-carriers (sporadic, orange) and disease control cases (white). The non-parametric Kruskal-Wallis test was used to test for significant differences between cases and controls. The false discovery rate was controlled using the two-stage linear step-up procedure of Benjamini, Krieger and Yukutieli. Stars represent comparisons with p-values reaching significance (*p-value<0.05; **p-value<0.01); red stars denote comparisons with p-values reaching significance and also qualifying as a true discovery. Results are shown using box-and-whiskers plots. Expression levels are shown as medians normalized to *RPLPO* endogenous control. **(d)** Chromatogram showing sequences of a complete bisulfite conversion (Refseq, two unmethylated alleles), completely unconverted cytosines (two methylated alleles, found in control cases,), combination of converted and unconverted cytosines (one methylated and one unmethylated alleles, found in patients from both *C9orf72* and sporadic groups), and completely converted cytosines (two unmethylated alleles, found in patients from *C9orf72* group only) at position

chr. 14:94,849,201 (h19). (e) Graph showing that *SERPINA1* is largely unexpressed in most human tissues, except in liver and blood. Information has been retrieved from <https://gtexportal.org/home/gene/SERPINA1> in July 2017.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 4. Schematic representation of network 1 in c9ALS cerebellum after conducting IPA network analysis of genes with expression changes and DMCs

Nodes in green represent hypomethylated genes with expression p -value <0.05 , and nodes in red represent hypermethylated genes with expression p -value <0.05 . Nodes in gray represent upstream regulators. Some of the top 10 hypermethylated and 10 hypomethylated genes identified in our IPA analysis were part of network 1 (dark blue font). Genes or molecules associated with ALS (red font) and another neurodegenerative disease (purple font) are also part of network 1. Network 1 also include genes with differential expression (p -value <0.05 ; $|\log_2FC| \geq 1$) or alternative cassette exon splicing ($FDR<0.05$; $|d| \geq 0.1$) and DMCs in regions predicted to modulate transcription (light blue font). Of note, *FBLN2* was both part of the top 10 hypermethylated group and the alternative cassette exon splicing ($FDR<0.05$; $|d| \geq 0.1$) and DMCs in regions predicted to modulate transcription group. *FBLN2* is shown in light blue font.

Table 1
ALS and neuro-associated genes with DMCS and/or transcriptome defects in ALS

Frontal Cortex	DE p<0.05 log2FC	1	CE FDR<0.05 d	0.1	IR FDR<0.05 d	0.1	APA FDR<0.05 dPDI	0.2	dPDI	0.2	Upstream/ promoter q- value<0.01 methy diff >25%	Gene Body q-value<0.01 methy diff >25%
c9ALS	ALS-associated	PRPH, SERPINH1, SUSD2	ATXN2	n/a	n/a	CYP27A1, DNMT3A, DNMT3B				C9orf72, CAMTA1, RANGAPI, SPAST, TARDBP, UNC13A	ATXN2, CAMTA1, GLE1, RANGAPI, VAPB	
	neuro-associated	PFN1P2, TBPL2	MAPT	n/a	n/a	KHDRBS1				ADARB1, CACNA1C, ELAVL4, ENOX1, IGHMBP2, KHDRBS1, PLA2G6, PSEN2, RALY, SRSF1	ADARB1, ELAVL4, ENOX1, IGHMBP2, NOVA2, PARK2, PARK7, PLA2G6	
	ALS-associated	C7orf57	EWSR1, SLC1A2	n/a	n/a	DNMT3A, DNMT3B, SUSD2				C21orf2, CAMTA1, DAO, RANGAPI, TARDBP	ATXN2, C21orf2, CAMTA1, MOB3B	
sALS	neuro-associated	BCL2L2-PABPN1, TBPL2	HNRNPA1L2	n/a	n/a	ATXN2L, CACNA1C, ENOX1, HNRNPA1L2, KHDRBS1, RBFOX1, SRSF10				ADARB1, ADARB2, ATXN808, CACNA1C, CELF1, ENOX1, IGHMBP2, JPH3, KCNMA1, KHDRBS1, NRXN1, NRXN3, PARK2, PARK7, PURA	ADARB2, BCL2L2-PABPN1, CACNA1C, CELF1, CELF5, ENOX1, IGHMBP2, NRXN2, PABPN1, PARK2, PSEN1	
Cerebellum	ALS-associated	C9orf72, IFIH1, SERPINH1	ANG, ATXN2, C21orf2, CAMK1G, CAMTA1, DCTN1, DNMT1, ELP3, EWSR1, FUS, KIFAP3, MATR3, SCFD1, SPAST	DNMT1, EWSR1, SIGMAR1, TAF15	ALS2, ATXN2, CCNF, VAPB	C9orf72, FUS, HNRNPA2B1, OPTN, TARDBP	MOB3B, ZNF512B					

Acta Neuropathol. Author manuscript; available in PMC 2018 November 01.

DE: differential expression; CE: alternative cassette exon splicing; IR: retention of intronic region; APA: alternative polyadenylation.