

---

Article

Not peer-reviewed version

---

# Exploring SVA Insertion Polymorphisms in Shaping Differential Gene Expression in the Central Nervous System

---

Lauren S Hughes , Alexander Fröhlich , Abigail L Pfaff , [Vivien J Bubb](#) , [John P Quinn](#) , [Sulev Koks](#) \*

Posted Date: 26 December 2023

doi: [10.20944/preprints202312.1981.v1](https://doi.org/10.20944/preprints202312.1981.v1)

Keywords: SINE-VNTR-Alu; transposable elements; whole genome sequencing; RNA-seq; quantitative trait loci; Amyotrophic Lateral Sclerosis; HLA; MAPT



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

## Article

# Exploring SVA Insertion Polymorphisms in Shaping Differential Gene Expression in the Central Nervous System

Lauren S. Hughes <sup>1</sup>, Alexander Fröhlich <sup>1,2</sup>, Abigail L. Pfaff <sup>1,3</sup>, Vivien J. Bubb <sup>2</sup>, John P. Quinn <sup>2</sup> and Sulev Koks <sup>1,3,\*</sup>

<sup>1</sup> Perron Institute for Neurological and Translational Science, Perth, WA 6009, Australia

<sup>2</sup> Department of Pharmacology and Therapeutics, Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool, United Kingdom

<sup>3</sup> Centre for Molecular Medicine and Innovative Therapeutics, Murdoch University, Perth, WA 6150, Australia

\* Correspondence: Sulev.Koks@murdoch.edu.au

**Abstract:** Transposable elements (TEs) are repetitive elements which make up around 45% of the human genome. A class of TEs known as SINE-VNTR-Alu (SVA) demonstrate the capacity to mobilise throughout the genome, resulting in SVA polymorphisms for presence or absence within the population. Although studies have previously highlighted the involvement of TEs within neurodegenerative diseases, such as Parkinson's disease and amyotrophic lateral sclerosis (ALS), however the exact mechanism has yet to be identified. In this study we used whole genome sequencing and RNA sequencing data of ALS patients and healthy controls from the New York Genome Center ALS Consortium, to elucidate the influence of reference SVA elements on gene expression genome-wide within central nervous system (CNS) tissues. To investigate this, we applied matrix expression quantitative trait loci analysis and demonstrated that reference SVA insertion polymorphisms can significantly modulate the expression of numerous genes, preferentially in the *trans* position, and in a tissue-specific manner. We also highlight that SVAs significantly regulate mitochondrial genes as well as genes within the *HLA* and *MAPT* loci, previously associated within neurodegenerative disease. In conclusion, this study continues to bring to light the effects of polymorphic SVAs on gene regulation and further highlights the importance of TEs within disease pathology.

**Keywords:** SINE-VNTR-Alu; transposable element; whole genome sequencing; RNA-seq; gene expression regulation; expression quantitative trait loci; amyotrophic lateral sclerosis; *HLA*; *MAPT*

## 1. Introduction

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease, originally termed by the French neurologist Jean-Martin Charcot in 1869 to describe muscular atrophy (amyotrophic) and tissue scarring and hardening of tissue within the lateral spinal cord [1,2]. ALS is the most common form of motor neuron disease (MND) and is characterized by the progressive deterioration of both upper and lower motor neurons within the brain and spinal cord [2,3]. The incidence of which is 1.75 – 3 per 100 000 people per year and an increased incidence of 4 – 8 per 100 000 people per year in the highest ALS risk age group (45 – 75 years old) [4]. Categorised as 2 groups, ALS can present as familial ALS where at least one family member of the affected individual has ALS, accounting for up to 10% of ALS cases, or sporadic ALS (sALS) where the affected individual has no prior family history which accounts for 90 – 95% of cases [5]. The clinical features of typical ALS patients consist of muscle spasticity, atrophy, muscle wasting, weakness and death due to respiratory failure, with the average survival after symptom onset lying between 3-5 years [1,4,6].

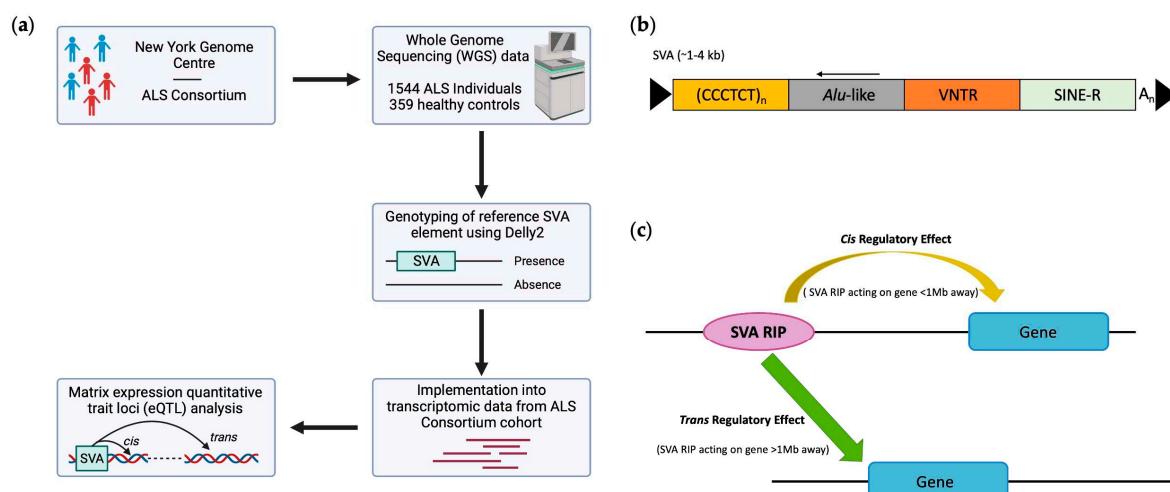
Neurodegenerative diseases are complex disorders, involving both environmental and genetic factor interactions. Therefore, attempts to elucidate disease mechanisms and pathogenic genetic variants are essential. Previous research has identified more than 30 genes associated with ALS, highlighting four genes *SOD1*, *TARDBP* (*TDP-43*), *C90RF72* and *FUS* for harbouring pathogenic mutations which cause the greatest number of ALS cases [7–10]. Although these four genes have been identified as major ALS-associated genes, a 2017 meta-analysis study demonstrated that within European and Asian populations, these genes only contribute to 47.7% and 5.2% of familial and sporadic cases, respectively [11]. Following the identification of these pathogenic mutations, several pathological mechanisms have been implicated in ALS pathogenesis including oxidative stress, mitochondrial dysfunction, axonal transport, inflammation, toxic protein aggregation and RNA metabolism and toxicity [7]. In addition to identified genetic variants only explaining a small fraction of sporadic aetiology, twin studies have highlighted the importance of genetic risk factors within sporadic ALS, estimating a 61% heritability [12]. Therefore, as the exact causation of ALS is still undetermined, a better understanding of disease pathogenesis and identification of genetic biomarkers is essential. To further investigate the missing heritability of ALS, Theunissen et al., proposed structural variations (SVs) as an area of potential significance [13]. SVs are classified as insertions, inversions, deletions and microsatellites, usually of repetitive structure, that are predominantly present within non-coding DNA regions and contribute towards genomic variation [13,14]. Furthermore, SVs have demonstrated the ability to modulate gene expression and have already been implicated in neurodegenerative diseases, such as the *C9orf72* repeat expansion in ALS and frontotemporal dementia (FTD) [15,16]. Hence, as 99% of the human genome is non-coding DNA, continued research into these regions is crucial to provide new insights into disease pathogenesis and to identify new potential targets for therapeutics.

Repetitive DNA is a major contributor to structural genomic variation. One form of repetitive DNA is a group of endogenous transposable elements (TEs), which can exist in both a static form and a mobile form. TEs are categorised into two classes known as DNA transposons and retrotransposons, whereby the later class possesses the ability to propagate throughout the genome via a ‘copy-and-paste’ mechanism, involving an RNA intermediate [17]. This results in the insertion of a new retrotransposon copy at a new locus within the host genome. Although, DNA transposons are capable of mobilisation via a ‘cut-and-paste’ mechanism, they are not currently active within the human genome [8,18]. Originally dismissed as “junk” DNA, TEs are known to drive genetic diversity not only by contributing to regulation and evolution of the genome but also by contributing to genetic instability and disease progression [18,19]. Previous research by Prudencio et al., highlighted the potential implication of retrotransposons in ALS through the analysis of repetitive element expression using RNA sequencing data from both healthy controls as well as *C9orf72* expansion positive carriers and sporadic ALS patients (*C9orf72*-negative) [20]. This research revealed that repetitive element expression, including retrotransposons, was significantly increased in ALS patients with the *C9orf72* expansion in comparison to *C9orf72*-negative patients and healthy controls, thus, suggesting retrotransposon involvement in ALS [20]. However, until recently these elements have been largely overlooked in relation to neurodegenerative diseases, even though TEs constitute to around 45% of the human genome [21].

Retrotransposons are further subdivided into two groups dependent on the presence of long terminal repeats (LTRs), known as LTR and non-LTR retrotransposons [18]. SINE-VNTR-Alu (SVA) elements are a member of the non-LTR retrotransposon family, which are typically 0.7–4kb in length [22]. SVAs are classified by evolutionary age based on their SINE-R region into sub-family groups A–F, whereby subfamily SVA-F is the youngest in evolutionary history [23]. Full-length SVA elements contain a 5' CT element, Alu-like region, GC-rich VNTR (variable number tandem repeat), SINE (short interspersed nuclear element)-R domain and a 3' poly-A tail [23] (Figure 1b). SVAs are major contributors to genetic diversity through a variety of different mechanisms, including acting as transcriptional regulators by providing alternative splice sites, polyadenylation signals and promoters whilst harbouring sites for transcription factor (TF) binding, thus, modulating gene expression [24]. SVAs are not only polymorphic in structure but ongoing mobilisation has resulted

in SVAs being polymorphic for presence or absence within the genome and thus are termed retrotransposon insertion polymorphisms (RIPs) [19]. This adds additional layers of complexity to gene expression dynamics, indicating that SVAs could be associated with predisposition to disease [25].

Previous work within our group has shown that seven SVAs polymorphic for their presence or absence were significantly associated with Parkinson's disease progression and differential gene expression in Parkinson's disease (PD) patients using the Parkinson's Progression Markers Initiative (PPMI) cohort [22,26]. This involved an SVA named SVA\_67 which is located at the MAPT locus, a locus which has been implicated in neurodegenerative disease risk including PD, FTD and Alzheimer's disease [27–29]. By using a clustered regularly interspaced short palindromic repeats (CRISPR) cell line knock-out model for SVA\_67, we demonstrated that this SVA was significantly associated with differential gene expression of three genes at the MAPT locus, associated with neurodegeneration [30].



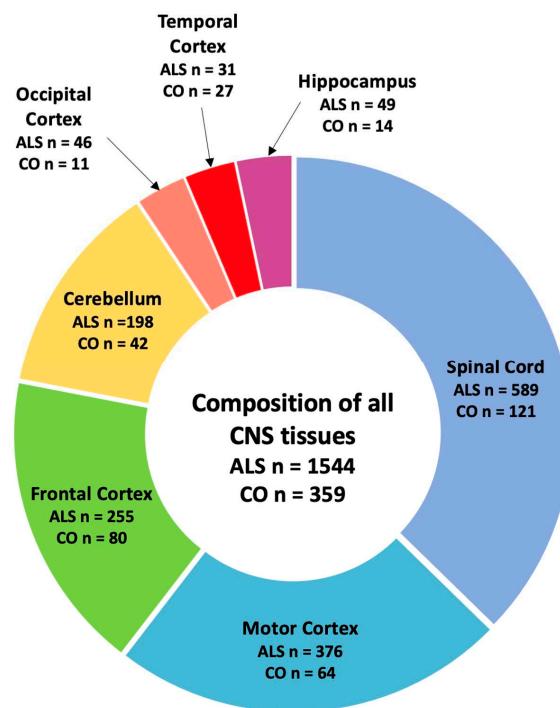
**Figure 1.** General study overview, structure of full-length SVA element and both *cis*- and *trans*-acting mechanisms. (a) This study incorporated whole genome sequencing and transcriptomic data from the New York Genome Centre ALS Consortium cohort, to investigate the ability of SVA retrotransposon insertion polymorphisms (RIPs) to act as expression quantitative trait loci (eQTL) within central nervous system (CNS) tissues. (b) Schematic of SVA structure, displaying a full-length SVA element consisting of a 5' CT rich hexamer repeat, *Alu*-like region, variable number tandem repeat (VNTR), SINE (short interspersed nuclear element)-R domain and a 3' poly-A tail. (c) The mechanism by which SVAs implement a *cis* or *trans* regulatory effect. *Cis*-regulatory effects are defined as effects observed by elements (SVA RIPs) which act to modulate the expression of genes less than 1 Mb away from the element site, whilst *trans* regulatory effects are defined as effects observed by elements which act to modulate the expression of genes greater than 1 Mb away from the element site.

In this study, we analysed SVAs which are present in the human reference genome identified as RIPs in the cohort analysed, herein termed reference SVA RIPs to investigate differential gene expression patterns associated with ALS. For this analysis, we used whole genome sequencing (WGS) and transcriptomic data obtained from the New York Genome Center (NYGC) ALS Consortium, to elucidate the role of SVA insertion polymorphisms on gene expression in central nervous system (CNS) tissues in healthy controls and ALS patients (Figure 1a). Our analysis demonstrated that reference SVA RIPs regulate single and multiple gene targets genome-wide within CNS tissues and display tissue-specific gene modulation by comparing spinal cord, cerebellum, motor cortex and frontal cortex tissue analyses. Furthermore, we discovered that SVA RIPs influence the expression of genes at loci (*HLA* and *MAPT*) previously associated with ALS, highlighting SVA regulation of genes at this locus could be a potential mechanism involved in ALS pathology.

## 2. Results

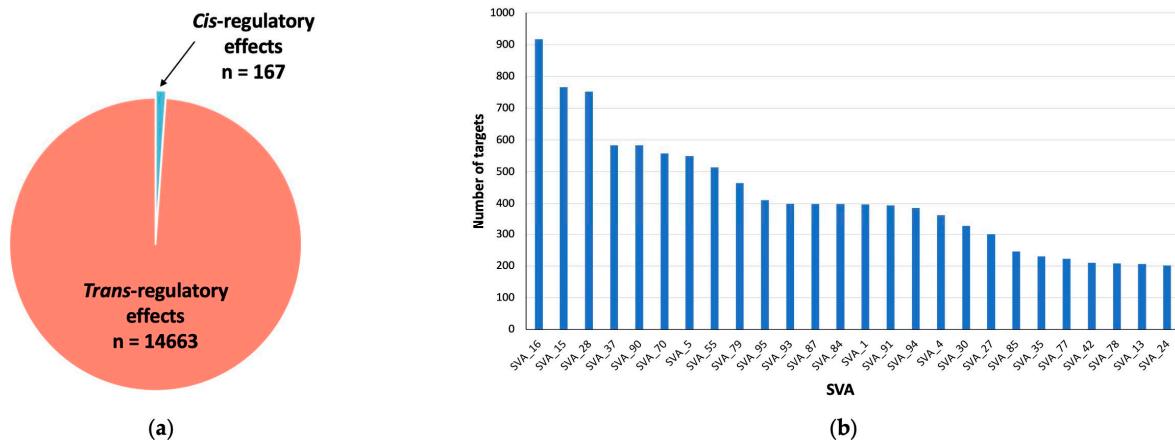
### 2.1. SVA RIPs act as eQTL genome wide in CNS tissues

To determine the impact of reference SVAs RIPs on gene expression profiles, we analysed WGS and transcriptomic data of both ALS patients (n=1544) and healthy controls (n=359) from the ALS Consortium cohort. For this analysis, we initially combined data from different CNS tissues, including motor cortex, frontal cortex, cerebellum, occipital cortex, temporal cortex, hippocampus, and spinal cord tissue (Figure 2).



**Figure 2.** Composition of tissues used in this study. For this study, CNS tissue data from healthy controls (CO) and ALS patients combined (n=1903) composed of spinal cord (n=710), motor cortex (n=440), frontal cortex (n=335), cerebellum (n=240), occipital cortex (n=57), temporal cortex (n=58) and hippocampus (n=63) were included for our analysis.

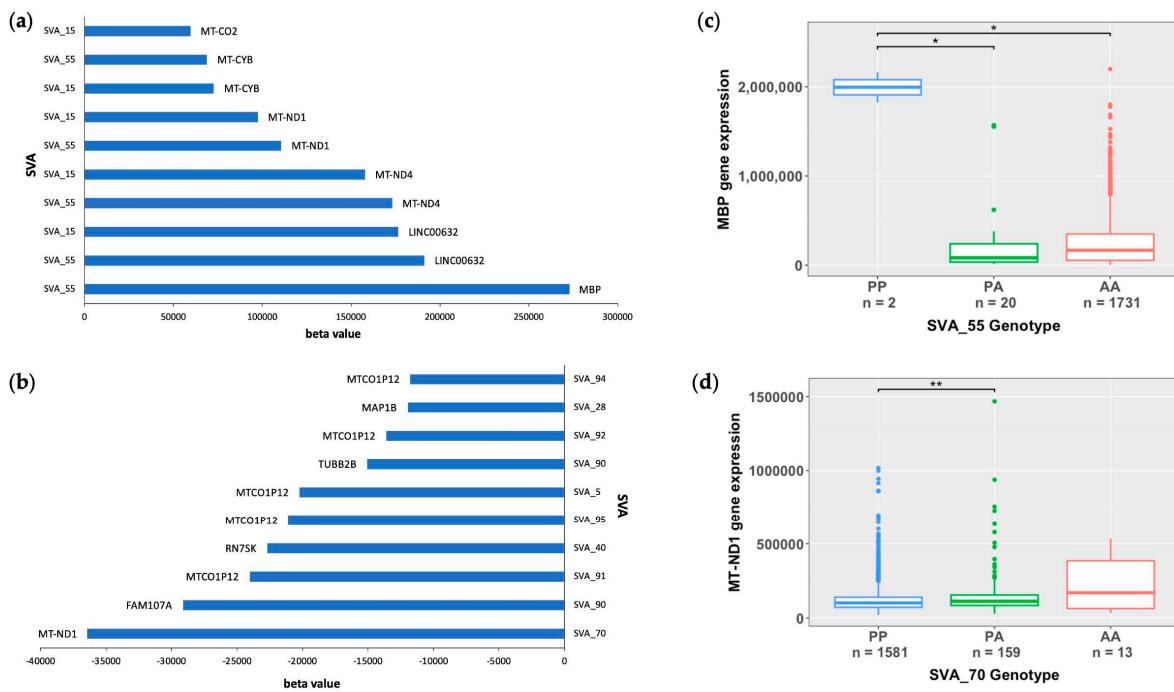
For eQTL analysis we assessed changes at a gene-based level. We demonstrate that, 14,830 genomic loci were significantly differentially modulated by reference SVA insertion polymorphism (FDR  $p < 0.05$ ). Of these effects, 167 were *cis* regulated by SVA RIPs, whilst over 98% (14,663 targets) were *trans*-regulated, where *trans*-regulated is defined as genes regulated greater than 1Mb away from the SVA site (Figure 3a). This indicates that most regulatory effects observed by these SVA RIPs impact target genes >1Mb away from the SVA element itself, these could be both direct and indirect effects (Figure 1c). In addition, single SVA elements had the capacity to regulate either a single gene or multiple gene targets. Our analysis identified that 92 reference SVA RIPs were responsible for the modulation of the 14,830 genomic loci, with SVA\_16 affecting the greatest number of genes (917) (Figure 3b). Of these 92 SVA elements polymorphic for their presence or absence, 26 SVAs affects on more than 200 genes in total, whilst the lowest number of targets for an SVA in this analysis was 6, indicating that all reference SVA RIPs analysed in this study had the ability to modulate expression profiles of multiple gene targets in the CNS (Figure 3b).



**Figure 3.** Overview of the number of genomic loci affected by SVA polymorphism following matrix eQTL analysis of all CNS tissues. **(a)** Pie chart representing the composition of all significant differentially regulated genetic loci (n=14830), displaying the number of *cis*-regulatory (n=167) and *trans*-regulatory (n=14663) effects exhibited by SVAs within all CNS tissues. **(b)** Bar chart displaying reference SVAs and the number of genome wide gene targets. Each of the 92 analysed SVAs had a significant impact on multiple targets, with the lowest number of targets for one SVA being 6 (FDR  $p<0.05$ ). Only SVAs affecting more than 200 targets are displayed (n=26). For this analysis data from ALS individuals and healthy controls were combined.

## 2.2. Mitochondrial genes are significantly modulated by SVA RIs

To further investigate the capacity of reference SVA elements to act as eQTL, we analysed the beta values obtained from the eQTL analysis, to determine the magnitude of the measured effects induced by SVA RIs. The ten reference SVAs RIs with the greatest effect size on gene upregulation (highest beta value) and gene downregulation (lowest beta value) and the affected gene targets are displayed in Figure 4a,b. All top ten hits for the highest and lowest beta values are *trans*-regulatory effects. SVA\_55 demonstrated the greatest increase in activation of the *MBP* gene with a beta coefficient 272,847, whilst SVA\_70 demonstrated the greatest repressive effect of the *MT-ND1* gene with a beta coefficient of -36441 (FDR  $p<0.05$ ) (Figure 4a,b). Interestingly, it was noted that the only two SVAs (SVA\_55 and SVA\_15) were responsible for the top ten hits regarding the effect size for gene activation. Furthermore, although the top hit for effect size for gene repression was SVA\_70 modulating *MT-ND1* expression, our results also demonstrated that the presence of both SVA\_15 and SVA\_55 showed the opposite effect by upregulating *MT-ND1* expression (Figure 4a,b). To determine the impact of SVA presence on gene modulation in more detail, we examined the top hits for SVAs with the greatest effect on gene upregulation (SVA\_55) and downregulation (SVA\_70) independently. For *MBP*, the presence of two copies/alleles (PP genotype) of SVA\_55 significantly upregulated *MBP* gene expression compared to the presence of one copy of SVA\_55 (PA genotype) or complete absence (AA genotype) ( $p=0.0218$ ) (Figure 4c). In comparison to the AA and PA genotypes, individuals with the PP genotype displayed a 9.5-fold and 19.2-fold increase in *MBP* gene expression, respectively. In contrast, individuals homozygous present for SVA\_70 demonstrated significant repression of the *MT-ND1* gene in comparison to individuals heterozygous for SVA\_70, namely down-regulating gene expression of *MT-ND1* by 55% ( $p=0.0098$ ) (Figure 4d). No statistical significance for differential gene expression between the PP and AA genotype was obtained for *MT-ND1* expression using the Wilcoxon pairwise comparison with FDR adjusted  $p$ -values (FDR  $p<0.05$ ).



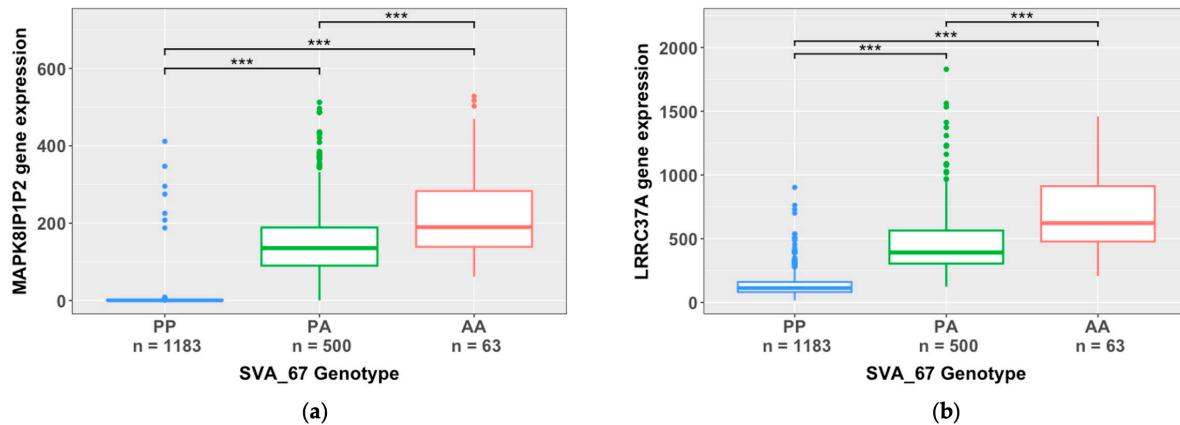
**Figure 4.** Reference SVA RIP elements with the greatest effect size from matrix eQTL analysis of all CNS tissues. (a/b) Clustered bar chart showing the top ten reference SVA RIPs across all CNS tissues with the greatest effect size on gene upregulation (positive beta values) (a) and gene downregulation (negative beta values) (b) from eQTL analysis. SVA\_55 demonstrated the greatest increase in activation of the *MBP* gene with a beta coefficient of 272,847, whilst SVA\_70 demonstrated the greatest repressive effect on the *MT-ND1* gene with a beta coefficient of -36441. (c) Boxplot of SVA\_55 indicating *MBP* gene expression stratified by SVA\_55 genotype. Genotypes PP (n=2), PA (n=20) and AA (n=1731). Significant differences in *MBP* gene expression were observed between the PP and PA group ( $p=0.0218$ ) and PP and AA group ( $p=0.0218$ ). Subjects with the PP genotype displayed a 9.5-fold and 19.2-fold increase in *MBP* gene expression in comparison to subjects with AA and PA genotypes respectively. (d) Boxplot of SVA\_70 displaying *MT-ND1* gene expression, stratified by SVA\_70 genotype. Genotypes PP (n=1581), PA (n=159) and AA (n=13). A significant repression in *MT-ND1* gene expression of 55% was observed between the PP and PA subject group ( $p=0.0098$ ). No statistical significance was obtained for differences between the PP and PA subject groups. For both boxplots the significance of gene expression changes between groups was determined using the Wilcoxon pairwise comparison with FDR adjusted p-values (FDR  $p<0.05$ ). PP, PA, and AA groups represent when there are two copies of the SVA present, one copy of the SVA present and the complete absence of the SVA, respectively. \*  $p<0.05$ , \*\*  $p<0.01$ .

We next analysed the top ten most significant SVA effects based on FDR p-value. The most significant effects observed (lowest FDR p-value) were SVA\_67 acting on the genes *MAPK8IP1P2* and *ENSG00000285668.1*, displaying an FDR p-value of 1.93E-303 (Table 1). Followed by SVA\_67 acting on the gene *LRRC37A*, with an FDR p-value of 5.12E-299 (Table 1). Intriguingly, 6/10 hits for highest significance were *cis*-specific effects on gene upregulation by SVA\_24, SVA\_33 and SVA\_58 and gene downregulation by SVA\_67. This is different to our previous results as the top ten reference SVAs with the greatest effect size on upregulation and downregulation were all *trans*-regulatory effects. Table 1 also shows that the four most significant *trans*-specific effects induced by SVA\_15, SVA\_84, SVA\_87, and SVA\_93 all influenced the same mitochondrial gene *MTND4P24*, whereby SVA\_84, SVA\_87, and SVA\_93 increased gene expression and SVA\_15 repressed gene expression. This demonstrates that multiple SVA RIPs can act on one gene influencing gene regulation, simultaneously activating and repressing gene expression.

**Table 1.** Top ten most significant reference SVA RIP effects from matrix eQTL analysis of all CNS tissues. This analysis included the combination of both *cis* and *trans* effects as well as datapoints from both ALS individuals and healthy controls.

SVA	beta value	False Discovery Rate (FDR)	Target gene	<i>cis/trans</i> effect
SVA_67	-131.2	1.93E-303	MAPK8IP1P2	<i>cis</i>
SVA_67	-5.2	1.93E-303	ENSG00000285668.1	<i>cis</i>
SVA_67	-315.1	5.12E-299	LRRC37A	<i>cis</i>
SVA_87	-126.2	3.02E-224	MTND4P24	<i>trans</i>
SVA_93	-126.2	4.22E-224	MTND4P24	<i>trans</i>
SVA_84	-63.1	4.22E-224	MTND4P24	<i>trans</i>
SVA_58	4.1	3.89E-211	LLPH-DT	<i>cis</i>
SVA_24	12.3	5.06E-201	HLA-K	<i>cis</i>
SVA_15	59.6	3.13E-189	MTND4P24	<i>trans</i>
SVA_33	48.4	8.84E-188	ZFAND2A-DT	<i>cis</i>

To illustrate the impact of SVA\_67 on target gene expression, we investigated the specific influence of SVA\_67 allele dosage on transcriptomic profiles of the respective target gene (Figure 5). As an example, this analysis is shown for the genes *MAPK8IP1P2* and *LRRC37A*, whereby gene expression was stratified by SVA\_67 genotype (PP, PA, and AA). Analysis using the Wilcoxon pairwise comparison with FDR adjusted *p*-values (FDR<0.05) demonstrated that SVA\_67 presence significantly downregulated gene expression of both *MAPK8IP1P2* and *LRRC37A* (*p*<0.001) (Figure 5). Significant differences were observed between all genotype groups PP vs PA (*MAPK8IP1P2* *p*=7.33E-237, *LRRC37A* *p*=1.95E-204), PP vs AA (*MAPK8IP1P2* *p*=4.26E-55, *LRRC37A* *p*=1.06E-39), and PA vs AA (*MAPK8IP1P2* *p*=1.33E-07, *LRRC37A* *p*=2.13E-12) for both genes (Figure 5). For individuals homozygous absent for SVA\_67, a change of 262-fold and 6-fold was observed for gene expression of *MAPK8IP1P2* and *LRRC37A* respectively, in comparison to individuals homozygous present for SVA\_67 (Figure 5).

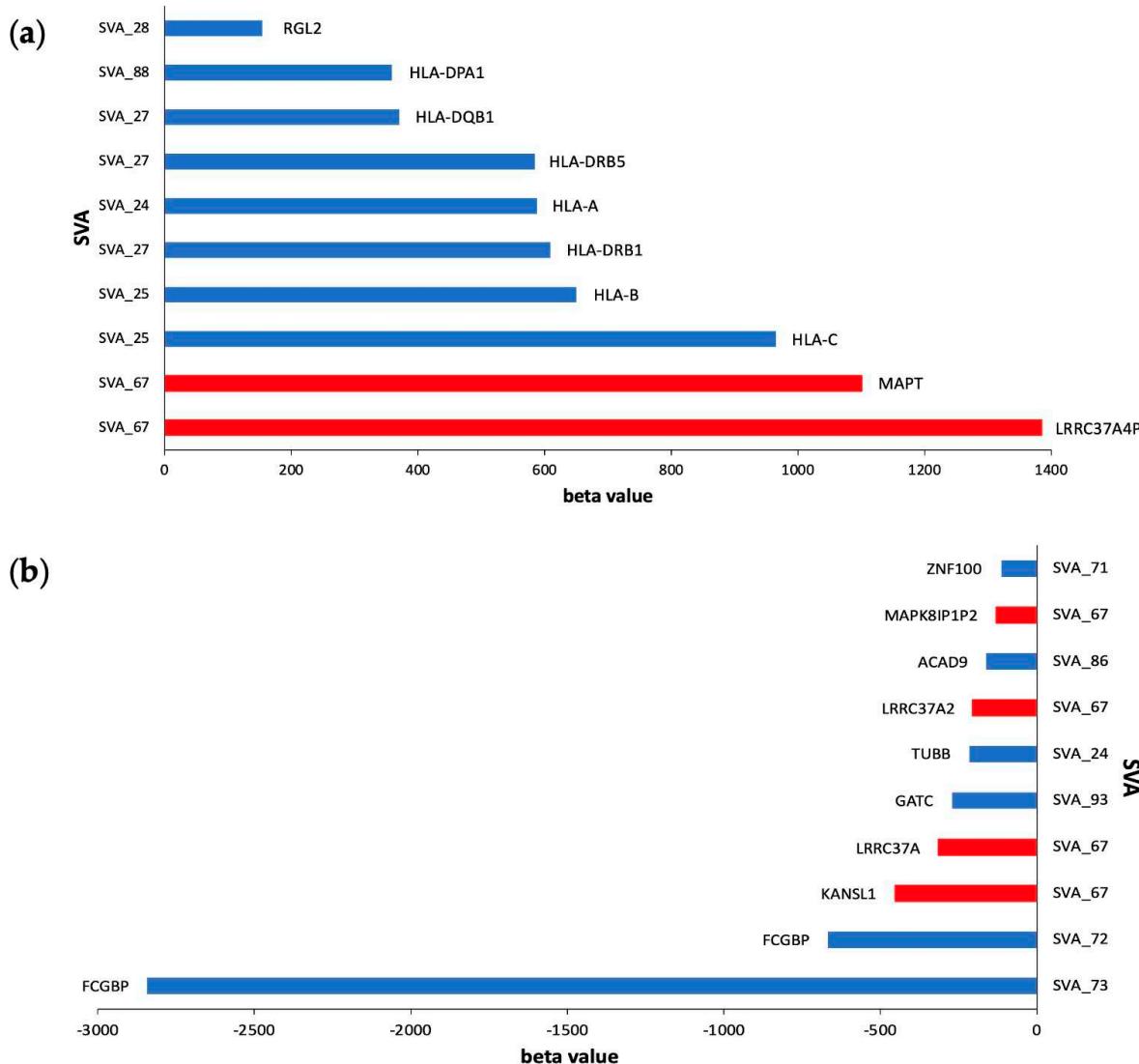


**Figure 5.** Boxplots of the two of the most significant SVA\_67 interactions with *MAPK8IP1P2* and *LRRC37A* obtained from matrix eQTL analysis. Both significant effects are *cis*-regulatory effects. Datapoints from both ALS individuals and healthy controls across all CNS tissues were combined and the significance of gene expression changes between groups was determined using the Wilcoxon pairwise comparison with FDR adjusted *p*-values (FDR<0.05). (a) Boxplot showing the association of SVA\_67 genotype with *MAPK8IP1P2* gene expression. Significant differences were observed between all groups for *MAPK8IP1P2* expression, namely PP and PA (*p*=7.33E-237), PP and AA (*p*=4.26E-55) and PA and AA (*p*=1.33E-07). (b) Boxplot showing the association of SVA\_67 genotype with *LRRC37A* gene expression. Significant differences were observed between all groups for *LRRC37A* gene expression, PP and PA (*p*=1.95E-204), namely PP and AA (*p*=1.06E-39) and PA and AA (*p*=2.13E-12). Fold change in expression of 262-fold and 6-fold for *MAPK8IP1P2* and *LRRC37A*, respectively, was

observed for individuals with AA genotype in comparison to PP genotype. PP (n=1183), PA (n=500) and AA (n=63). \*\*\*  $p<0.001$ .

### 2.3. SVA RIPs demonstrate *cis* effects on HLA and MAPT loci

To elucidate the impact of *cis*-acting SVA RIPs, we explored which SVA elements had the largest significant effects on genes regulated in *cis*, including gene upregulation (positive beta value) and downregulation (negative beta value) using beta values from matrix eQTL analysis (FDR  $p<0.05$ ) (Figure 6). Here, we highlight that SVA\_67 largely influences gene regulation in *cis* of six different genes, including the upregulation of *MAPT* and *LRRC37A4P* and downregulation of *MAPK8IP1P2*, *LRRC37A2*, *LRRC37A*, and *KANSL1* (Figure 6a). SVA\_67 was responsible for the two greatest increases in gene activation of the *MAPT* and *LRRC37A4P* genes with beta coefficients of 1384 and 1101, respectively (Figure 6a). Interestingly, six of the ten most positive hits were effects induced on HLA genes by SVA\_24, SVA\_25, SVA\_27 and SVA\_88 (Figure 6a). Furthermore, we demonstrate that SVA\_73 had the greatest effect out of all *cis*-regulatory effects with a beta coefficient of -2843 for the gene *FCGBP*, indicating a large inhibitory effect of *FCGBP* gene expression (Figure 6b). *FCGBP* gene expression is further repressed by SVA\_72, the second highest hit in our analysis for the top ten SVAs with the greatest effect size on gene downregulation (Figure 6b).



**Figure 6.** Top ten *cis*-acting reference SVA elements with the greatest effects on gene upregulation and downregulation in CNS tissues. (a) Clustered bar chart displaying ten reference SVA RIPs with the greatest *cis*-regulatory effects on gene activation (most positive beta values) from matrix eQTL

analysis (FDR  $p<0.05$ ). SVA\_67 is responsible for the two greatest increases in activation of the genes *LRRC37A4P* and *MAPT* displaying beta coefficients of 1384 and 1101, respectively. SVA\_24 (*HLA-A*), SVA\_25 (*HLA-C* and *HLA-B*), SVA\_27 (*HLA-DRB1*, *HLA-DRB5* and *HLA-DQB1*) and SVA\_88 (*HLA-DPA1*) showed to be responsible for the large increases in activation of a series of *HLA* genes. (b) Clustered bar chart displaying ten reference SVA RIPs with the greatest *cis*-regulatory effects on gene downregulation (most negative beta values) from matrix eQTL analysis (FDR  $p<0.05$ ). The two greatest repressive effects were on the *FCGBP* gene, regulated by both SVA\_73 and SVA\_72 demonstrating a beta coefficient of -2843 and -668, respectively. SVA\_67 was responsible for four of these effects, by showing a downregulating effect, for the genes *KANSL1*, *LRRC37A*, *LRRC37A2* and *MAPK8IP1P2*. This analysis combined ALS individuals and healthy controls datapoints combined.

To determine the influence of *cis*-acting SVA RIPs on gene regulation, we next analysed gene expression patterns of the top two hits for the largest effects on gene upregulation and downregulation (Figure S1). The presence of either one (PA) or two (PP) SVA\_67 alleles significantly upregulated both *LRRC37A4P* and *MAPT* gene expression when tested using the Wilcoxon pairwise comparison with FDR adjusted p-values (FDR  $p<0.05$ ) (Figure S1a,b). In comparison to individuals homozygous absent for SVA\_67, homozygous present and heterozygous individuals displayed a 447.3-fold ( $p=1.04E-40$ ) and a 214.4-fold ( $p=2.55E-38$ ) increase in *LRRC37A4P* expression and a 1.4-fold ( $p=2.53E-09$ ) and 1.2-fold ( $p=2.51E-04$ ) increase in *MAPT* gene expression, respectively (Figure S1a,b). Analysis of the top two hits for the greatest effect on gene downregulation displayed differing gene expression patterns between the influence of SVA\_73 and SVA\_72 on *FCGBP* gene expression (Figure S1c,d). When comparing to the complete absence (AA) of SVA\_73, the PA and PP genotype groups displayed a significant 3.7-fold ( $p=3.61E-02$ ) and 10-fold ( $p=2.61E-03$ ) increase in *FCGBP* gene expression (Figure S1c). Although our analysis displayed that the presence of SVA\_73 upregulates the *FCGBP* gene in comparison to AA individuals, *FCGBP* gene expression is significantly repressed by 63% ( $p=3.29E-08$ ) in the PP genotype group compared to the PA genotype group (Figure S1c). SVA\_72 analysis highlighted that homozygous presence of the SVA\_72 allele significantly downregulated *FCGBP* gene expression by 23% ( $p=5.74E-03$ ) in comparison to homozygous absence (Figure S1d). PP genotype individuals also displayed a significant reduction of 32% ( $p=1.85E-05$ ) in *FCGBP* gene expression in comparison to the PA genotype (Figure S1d). Similar to SVA\_73 gene expression analysis, individuals heterozygous for SVA\_72 indicated a 1.2-fold ( $p=4.23E-01$ ) increase in *FCGBP* expression, however statistical significance was not achieved (Figure S1d).

As a recent genome-wide association study (GWAS) reported the association of the *HLA* locus with ALS risk [31,32], we investigated the influence that SVA RIPs had on *cis* regulated *HLA* genes (Figures S2 and S3). This analysis demonstrated that the top ten hits for greatest effect of SVA RIPs regarding upregulation was for *HLA* genes. We discovered a common gene expression pattern across all *HLA* genes analysed, demonstrating that individuals homozygous present for SVA RIPs (SVA\_24, SVA\_25, SVA\_27, SVA\_88) significantly upregulate *HLA* genes in comparison to homozygous absent individuals ( $p<0.01$ ) (Figures S2 and S3).

#### 2.4. SVA RIPs display tissue-specific modulation of gene expression

To investigate tissue-specific influences of reference SVA polymorphisms, we conducted eQTL analysis on the CNS tissue types spinal cord, cerebellum, motor cortex and frontal cortex individually (Tables 2 and 3). The tissues temporal cortex, occipital cortex and hippocampus were excluded in the analysis due to the low n numbers. The results tables display the eQTL analysis from each individual tissue to allow determination of the top 40 hits for greatest effect size on gene upregulation (positive beta values) (Table 2) and downregulation (negative beta values) (Table 3). From this analysis, we established that a large proportion of the gene regulation effects were present within spinal cord tissue, displaying 27/40 and 26/40 tissue-specific hits for both gene upregulation and downregulation, respectively (Tables 2 and 3). Furthermore, our analysis revealed that five SVAs (SVA\_55, SVA\_15, SVA\_37, SVA\_85, and SVA\_4) were responsible for the 40 most positive beta values and eight SVAs (SVA\_90, SVA\_5, SVA\_87, SVA\_93, SVA\_30, SVA\_70, SVA\_84, SVA\_91, SVA\_16) were responsible

for the 40 most negative beta values, all effects of which were in the *trans* position (Tables 2 and 3). This data demonstrates the capability of SVAs to modulate the expression of multiple gene targets.

**Table 2.** Top 40 significant reference SVA RIPs with the greatest effects on gene upregulation (most positive beta values) from tissue specific matrix eQTL analysis. This analysis included the combination of both *cis* and *trans* effects as well as datapoints from both ALS individuals and healthy controls.

SVA	Gene ID	FDR p-value	Beta value	Gene	Chr	Cis/trans	Tissue
SVA_55	ENSG00000197971.16	5.99E-10	638804.662	MBP	18	<i>trans</i>	Spinal Cord
SVA_15	ENSG00000197971.16	3.59E-07	620026.724	MBP	18	<i>trans</i>	Spinal Cord
SVA_37	ENSG00000197971.16	8.76E-04	585791.502	MBP	18	<i>trans</i>	Spinal Cord
SVA_85	ENSG00000197971.16	2.08E-03	560691.911	MBP	18	<i>trans</i>	Spinal Cord
SVA_37	ENSG00000123560.14	3.72E-12	168441.474	PLP1	X	<i>trans</i>	Spinal Cord
SVA_55	ENSG00000198888.2	1.47E-05	153947.476	MT-ND1	MT	<i>trans</i>	Spinal Cord
SVA_55	ENSG00000203930.12	5.97E-04	148616.177	LINC00632	X	<i>trans</i>	Motor Cortex
SVA_15	ENSG00000198888.2	1.62E-03	142961.051	MT-ND1	MT	<i>trans</i>	Spinal Cord
SVA_55	ENSG00000203930.12	1.25E-16	115977.491	LINC00632	X	<i>trans</i>	Spinal Cord
SVA_15	ENSG00000203930.12	1.86E-08	96910.7666	LINC00632	X	<i>trans</i>	Spinal Cord
SVA_55	ENSG00000180354.16	1.34E-22	90266.4311	MTURN	7	<i>trans</i>	Spinal Cord
SVA_15	ENSG00000123560.14	3.43E-03	80380.5623	PLP1	X	<i>trans</i>	Spinal Cord
SVA_85	ENSG00000203930.12	4.76E-03	78457.5282	LINC00632	X	<i>trans</i>	Spinal Cord
SVA_15	ENSG00000180354.16	1.86E-12	77839.9088	MTURN	7	<i>trans</i>	Spinal Cord
SVA_55	ENSG00000198712.1	3.04E-03	76318.8326	MT-CO2	MT	<i>trans</i>	Spinal Cord
SVA_85	ENSG00000180354.16	1.35E-05	67900.9382	MTURN	7	<i>trans</i>	Spinal Cord
SVA_85	ENSG00000237973.1	4.10E-33	66702.2584	MTCO1P12	1	<i>trans</i>	Motor Cortex
SVA_37	ENSG00000168314.18	3.50E-09	65555.4973	MOBP	3	<i>trans</i>	Spinal Cord
SVA_85	ENSG00000237973.1	1.01E-32	53997.5748	MTCO1P12	1	<i>trans</i>	Frontal Cortex
SVA_15	ENSG00000168314.18	5.02E-09	53362.8451	MOBP	3	<i>trans</i>	Spinal Cord
SVA_55	ENSG00000237973.1	3.61E-22	52048.0183	MTCO1P12	1	<i>trans</i>	Motor Cortex
SVA_55	ENSG00000168314.18	1.38E-09	49216.1425	MOBP	3	<i>trans</i>	Spinal Cord
SVA_15	ENSG00000237973.1	2.83E-26	48449.3666	MTCO1P12	1	<i>trans</i>	Motor Cortex
SVA_85	ENSG00000168314.18	3.76E-04	47350.2242	MOBP	3	<i>trans</i>	Spinal Cord
SVA_85	ENSG00000237973.1	8.82E-42	39521.0532	MTCO1P12	1	<i>trans</i>	Spinal Cord
SVA_55	ENSG00000064787.13	1.62E-30	36504.4209	BCAS1	20	<i>trans</i>	Spinal Cord
SVA_37	ENSG00000091513.16	1.65E-06	35838.5557	TF	3	<i>trans</i>	Spinal Cord
SVA_55	ENSG00000237973.1	1.49E-59	33728.2747	MTCO1P12	1	<i>trans</i>	Spinal Cord
SVA_4	ENSG00000237973.1	3.59E-17	32837.5825	MTCO1P12	1	<i>trans</i>	Frontal Cortex
SVA_37	ENSG00000237973.1	1.52E-23	32783.3654	MTCO1P12	1	<i>trans</i>	Motor Cortex
SVA_37	ENSG00000237973.1	1.72E-31	32728.512	MTCO1P12	1	<i>trans</i>	Frontal Cortex
SVA_15	ENSG00000237973.1	8.40E-41	32193.3558	MTCO1P12	1	<i>trans</i>	Spinal Cord
SVA_37	ENSG00000099194.6	2.82E-04	31333.7253	SCD	10	<i>trans</i>	Spinal Cord
SVA_85	ENSG00000064787.13	3.78E-10	30888.7058	BCAS1	20	<i>trans</i>	Spinal Cord
SVA_15	ENSG00000237973.1	3.46E-14	30602.307	MTCO1P12	1	<i>trans</i>	Frontal Cortex
SVA_15	ENSG00000064787.13	1.97E-15	30356.5865	BCAS1	20	<i>trans</i>	Spinal Cord
SVA_37	ENSG00000237973.1	1.28E-06	30025.949	MTCO1P12	1	<i>trans</i>	Cerebellum
SVA_15	ENSG00000237973.1	5.29E-13	29472.617	MTCO1P12	1	<i>trans</i>	Cerebellum
SVA_55	ENSG00000237973.1	2.00E-08	27693.2393	MTCO1P12	1	<i>trans</i>	Frontal Cortex
SVA_4	ENSG00000237973.1	2.56E-09	27128.1241	MTCO1P12	1	<i>trans</i>	Motor Cortex

**Table 3.** Top 40 significant reference SVA RIPs with the greatest effects on gene downregulation (most negative beta values) from tissue specific matrix eQTL analysis. This analysis included the combination of both *cis* and *trans* effects as well as datapoints from both ALS individuals and healthy controls.

SVA	Gene ID	FDR p-value	Beta value	Gene	Chr	Cis/trans	Tissue
SVA_90	ENSG00000197971.16	3.28E-03	-1337611.1	MBP	18	<i>trans</i>	Spinal Cord
SVA_5	ENSG00000197971.16	3.06E-03	-351064.11	MBP	18	<i>trans</i>	Spinal Cord
SVA_87	ENSG00000123560.14	1.98E-07	-193423.77	PLP1	X	<i>trans</i>	Spinal Cord
SVA_93	ENSG00000123560.14	2.05E-07	-193266.83	PLP1	X	<i>trans</i>	Spinal Cord
SVA_30	ENSG00000198888.2	8.47E-03	-143159.48	MT-ND1	MT	<i>trans</i>	Cerebellum
SVA_70	ENSG00000198886.2	1.029E-02	-133464.74	MT-ND4	MT	<i>trans</i>	Cerebellum
SVA_30	ENSG00000198763.3	7.85E-04	-131182.45	MT-ND2	MT	<i>trans</i>	Motor Cortex
SVA_30	ENSG00000198938.2	6.86E-03	-111339.82	MT-CO3	MT	<i>trans</i>	Cerebellum
SVA_30	ENSG00000198727.2	5.66E-03	-104147.35	MT-CYB	MT	<i>trans</i>	Cerebellum
SVA_84	ENSG00000123560.14	2.11E-07	-96556.036	PLP1	X	<i>trans</i>	Spinal Cord
SVA_30	ENSG00000198786.2	7.25E-03	-94294.133	MT-ND5	MT	<i>trans</i>	Motor Cortex
SVA_70	ENSG00000198763.3	1.29E-02	-82929.066	MT-ND2	MT	<i>trans</i>	Cerebellum
SVA_90	ENSG00000259001.3	6.14E-03	-73740.939	ENSG00000259001	14	<i>trans</i>	Spinal Cord
SVA_91	ENSG00000203930.12	2.02E-03	-71147.071	LINC00632	X	<i>trans</i>	Spinal Cord
SVA_87	ENSG00000168314.18	3.60E-04	-67162.942	MOBP	3	<i>trans</i>	Spinal Cord
SVA_93	ENSG00000168314.18	3.67E-04	-67112.229	MOBP	3	<i>trans</i>	Spinal Cord
SVA_91	ENSG00000180354.16	2.99E-05	-57438.97	MTURN	7	<i>trans</i>	Spinal Cord
SVA_90	ENSG00000168309.18	4.26E-03	-48145.904	FAM107A	3	<i>trans</i>	Spinal Cord
SVA_5	ENSG00000180354.16	1.55E-05	-43141.656	MTURN	7	<i>trans</i>	Spinal Cord
SVA_70	ENSG00000198712.1	6.03E-03	-43049.326	MT-CO2	MT	<i>trans</i>	Cerebellum
SVA_90	ENSG00000168309.18	2.45E-03	-40637.548	FAM107A	3	<i>trans</i>	Motor Cortex
SVA_91	ENSG00000168314.18	5.48E-04	-40457.332	MOBP	3	<i>trans</i>	Spinal Cord
SVA_90	ENSG00000177575.13	2.99E-19	-36552.585	CD163	12	<i>trans</i>	Spinal Cord
SVA_84	ENSG00000168314.18	3.83E-04	-33494.891	MOBP	3	<i>trans</i>	Spinal Cord
SVA_90	ENSG00000087086.15	1.64E-08	-32820.222	FTL	19	<i>trans</i>	Motor Cortex
SVA_5	ENSG00000168314.18	1.72E-04	-31165.253	MOBP	3	<i>trans</i>	Spinal Cord
SVA_16	ENSG00000123560.14	8.97E-07	-31144.963	PLP1	X	<i>trans</i>	Spinal Cord
SVA_91	ENSG00000237973.1	6.16E-30	-29834.037	MTCO1P12	1	<i>trans</i>	Spinal Cord
SVA_87	ENSG00000173786.17	5.88E-04	-28556.333	CNP	17	<i>trans</i>	Spinal Cord
SVA_93	ENSG00000173786.17	5.98E-04	-28536.883	CNP	17	<i>trans</i>	Spinal Cord
SVA_90	ENSG00000137285.11	1.48E-40	-27173.64	TUBB2B	6	<i>trans</i>	Motor Cortex
SVA_5	ENSG00000237973.1	3.92E-10	-24995.474	MTCO1P12	1	<i>trans</i>	Motor Cortex
SVA_5	ENSG00000237973.1	8.77E-13	-24871.705	MTCO1P12	1	<i>trans</i>	Frontal Cortex
SVA_91	ENSG00000064787.13	5.47E-07	-22960.148	BCAS1	20	<i>trans</i>	Spinal Cord
SVA_93	ENSG00000198840.2	3.12E-03	-22049.981	MT-ND3	MT	<i>trans</i>	Spinal Cord
SVA_87	ENSG00000198840.2	3.13E-03	-22046.442	MT-ND3	MT	<i>trans</i>	Spinal Cord
SVA_90	ENSG00000164733.22	1.39E-04	-21828.547	CTSB	8	<i>trans</i>	Spinal Cord
SVA_90	ENSG00000079215.15	1.69E-05	-21020.491	SLC1A3	5	<i>trans</i>	Motor Cortex
SVA_5	ENSG00000237973.1	8.40E-24	-19907.219	MTCO1P12	1	<i>trans</i>	Spinal Cord
SVA_87	ENSG00000136541.15	1.45E-10	-19551.323	ERMN	2	<i>trans</i>	Spinal Cord

Moreover, the *MBP* gene located on chromosome 18 and the *PLP1* gene located on the X chromosome, were the most significantly upregulated and downregulated genes simultaneously (Tables 2 and 3). Here, the gene expression of both genes was modulated in spinal cord tissue (Tables 2 and 3). In addition, the genes *MTURN* and *MOBP*, and mitochondrial genes (*MTCO1P12*, *MT-ND1*, *MT-ND2*, *MT-ND3*, *MT-ND4*, *MT-ND5*, *MT-CO2*, *MT-CO3*, and *MT-CYB*) appear multiple times within both gene upregulation and downregulation hits (Tables 2 and 3). Upon analysis of the whole tissue-specific dataset, the genes *MBP*, *PLP1*, *MTURN*, and *MOBP* were only regulated within spinal

cord tissue, whilst the expression of mitochondrial genes was modulated within spinal cord, cerebellum, and motor cortex while *MTCO1P12* gene expression was regulated in all four tissue types (Tables 2 and 3). Thus, this illustrates that SVAs possess the capacity to influence tissue-specific gene expression.

### 3. Discussion

In this study, we evaluated the role of reference SVAs polymorphic for their presence in the human genome, to modulate gene expression within CNS tissues of ALS patients and healthy controls. Analysis of the NYGC ALS consortium dataset demonstrated that SVA RIPs significantly regulate gene expression genome-wide and in a tissue-specific manner. This study continues to expand on our previous findings, demonstrating the capability of SVAs to differentially regulate gene expression [24,33]. We have previously illustrated the capacity of SVAs to influence gene expression within Parkinson's disease, highlighting the role of SVA\_67 to modulate genes at the *MAPT* locus within the PPMI cohort and a CRISPR deletion model [22,26,30]. This analysis further validates our previous research, illustrating the functional capacity of SVA\_67 within neurodegenerative disease and potentially expanding the importance of SVA\_67 and the *MAPT* locus to ALS (Figures 5 and 6). Therefore, this study not only emphasizes the correlation between SVA presence or absence and differential gene expression but also the involvement of SVAs within disease pathology.

Using WGS data from the ALS consortium for matrix eQTL analysis, we identified that polymorphic SVA RIPs possess the ability to modify expression of multiple target genes, but also many SVAs can regulate gene expression of a single target. Of these SVA RIPs, a greater proportion of *trans* regulatory effects were displayed in comparison to *cis*. Similarly, eQTL studies by both Wang et al. (2017) and Koks et al. (2021) analysing RNA-seq data from the 1000 genome project and the PPMI cohort, respectively, identified the impact of TEs on gene expression. In line with our analysis, these studies highlighted that several TEs simultaneously modulate gene expression of a single gene, an individual TE can regulate the expression of multiple genes, and that a greater number of TEs analysed were in the *trans* position [33,34]. A potential mechanism for *trans*-acting eQTLs could be through the binding of CCCTC-binding factor (CTCF) to SVAs. Through the cooperation with protein complex cohesion, CTCF plays a key role in three-dimensional chromatin regulation and chromatin looping, bringing promoters and regulatory elements within close proximity to activate or repress gene expression [35,36]. An additional mechanism could be through indirect transcription factor (TF) mediated associations. This suggests that SVAs influencing the expression of TF could indirectly regulate one or multiple TF gene targets, ultimately activating or repressing TF target gene expression [34].

Upon the examination of beta values from the eQTL analysis, we determined that within the combined analysis of CNS tissues SVA\_55 demonstrated the greatest effect on myelin basic protein (*MBP*) gene upregulation (Figure 4). Following further analysis, we identified that through individual tissue analysis the top hits for gene upregulation and downregulation, were SVAs influencing *MBP* gene expression (Tables 2 and 3). However, *MBP* expression was only significantly modulated within spinal cord tissue. *MBP* is a key protein involved in the myelination process, whereby myelin sheaths are formed around CNS axons by oligodendrocytes [37,38]. As oligodendrocyte loss and myelin dysfunction has recently been emphasized in neurodegenerative disease, including ALS, it is essential to investigate this relationship between SVAs and *MBP* differential expression [39,40]. Lorente Pons et al. demonstrated the potential significance of *MBP* in ALS through post-mortem analysis of both sporadic ALS and C9orf72-related ALS cases, identifying a significant reduction in *MBP* protein abundance when normalised to proteolipid protein (PLP) in the spinal cord corticospinal tracts in ALS cases in comparison to controls [41]. As *MBP* mRNA is transported to the myelin compartment by the RNA transport granule and PLP is transported as a protein, this suggests that the reduction in *MBP* could be due to impaired mRNA transport [41]. Our data suggests that certain SVAs (SVA\_90 and SVA\_5) act to downregulate *MBP* gene expression, therefore SVAs could play a role in this mechanism leading to the reduction of *MBP* protein levels observed in ALS patients compared to controls.

Our analysis also demonstrated that SVAs can activate (SVA\_37) and repress (SVA\_87 and SVA\_93) *PLP1*, a form of PLP. *PLP1* has previously been implicated in Pelizaeus-Merzbacher disease (PMD), an X-linked neurodegenerative disease whereby mutations within this gene inhibit CNS myelination [42,43]. In addition, we demonstrate that the myelin-associated oligodendrocyte basic protein (*MOBP*) gene, the locus of which has been highlighted for ALS risk, is again potentially regulated by multiple SVAs (SVA\_5, 15, 37, 55, 84, 85, 87, 91 and 93) only within the spinal cord [31]. Hence, SVA regulation could be a potential mechanism involved in ALS risk at the *MOBP* locus. Furthermore, our tissue specific analysis displayed that multiple mitochondrial genes (*MT-ND1,2,3,4,5*, *MT-CYB*, *MT-CO2,3* and pseudogene *MTCO1P12*) were largely activated and repressed within all four tissues analysed. SVA\_30 and SVA\_70 displayed the greatest regulatory effects on mitochondrial genes modulating five and four gene targets, respectively. As mitochondrial dysfunction is known to be implicated in other neurodegenerative diseases including PD and Alzheimer's disease, as well as ALS, SVA regulation resulting in differential expression could be an underlying mechanism involved in disease pathology [44,45].

Upon analysis of CNS tissues from both ALS individuals and healthy controls, seven of the top ten *cis*-acting reference SVA RIPs imposed the greatest effects on the upregulation of genes at the human leukocyte antigen (*HLA*) locus (Figure 6A). Four SVAs are responsible for the effects on *HLA* gene expression, whereby SVA\_24 influences the expression of one gene (*HLA-A*), SVA\_25 two genes (*HLA-B* and *HLA-C*), SVA\_27 two genes (*HLA-DRB1* and *HLA-DRB5*) and SVA\_88 one gene (*HLA-DQB1*). *HLA* also referred to as the major histocompatibility complex (MHC), acts to regulate both innate and adaptive immunity involved in the human immune response [46]. Since, the involvement of the immune response in neurological disease has been recognised the *HLA* locus has been highlighted as a region of importance in numerous neurodegenerative disease, including ALS [46]. Various studies have investigated the significance and mechanism of *HLA* in ALS, demonstrating increased frequencies of *HLA-A*, *HLA-B* and *HLA-C* alleles in ALS cases compared to controls [47–50]. A recent large-scale GWAS conducted by Van Rheenen et al. has identified the *HLA* region as a locus significantly associated with ALS, further highlighting the importance of this locus [31]. Previous studies within our group have highlighted the capability of SVAs to modulate *HLA* gene expression; analysis of whole genome sequencing and transcriptomic data obtained from the whole blood of individuals within the PPMI cohort discovered that SVA\_24, SVA\_25 and SVA\_27 modulate the expression of *HLA-A*, *HLA-B* and *HLA-C*, and *HLA-DRB1* and *HLA-DRB5*, respectively [33,51]. This suggests that the modulation of *HLA* genes by SVAs could be a common mechanism within neurodegenerative disease.

In conclusion, we show that SVAs demonstrate a significant impact on the expression of individual or numerous genes including those previously associated with neurodegenerative diseases, such as ALS. Ultimately, the ability of SVAs to act as a regulatory domain could highlight the importance of TEs in the missing heritability of neurodegenerative disease. However, due to limitations such as low n numbers for some SVA genotypes and CNS tissue types (occipital cortex, temporal cortex, and hippocampus) further research investigating the involvement of TEs in the pathogenesis of neurodegenerative disease, specifically ALS is crucial. In addition, due to low n numbers in the ALS and control groups, individual group analysis was not possible. Therefore, future experiments investigating the influence of SVAs in ALS patient and control groups individually is required. Furthermore, although this analysis continues to demonstrate the potential role of SVAs in neurodegenerative disease further experiments such as CRISPR are essential to validate SVA-specific influences. For example, we have previously shown that SVA\_67 deletion in a CRISPR model resulted in a significant increase in *MAPT* and *LRRC37A* gene expression [30].

#### 4. Materials and Methods

##### 4.1. Genotyping reference SVAs polymorphic for their presence/absence and disease association from whole genome sequencing data from ALS Consortium dataset

Whole genome sequencing (WGS) data from the New York Genome Center (NYGC) as part of the ALS consortium dataset, were obtained in cram file format and aligned to hg38. The ALS consortium dataset contains data from individuals diagnosed with ALS spectrum MND, other neurological disorder (Alzheimer's, Parkinson's disease etc), other MND and ALS with other neurological disorders, along with non-neurological controls (healthy controls). The structural variant caller Delly2 (<https://github.com/dellytools/delly>), with default settings, was used to genotype the ALS consortium cohort (4403 individuals) for previously identified reference SVA sites, as outlined in [52]. In total 92 reference SVAs polymorphic for presence/absence [52] were used for matrix eQTL and differential gene expression analysis.

##### 4.2. RNA-seq differential gene expression analysis

To elucidate the influence of SVA RIP genotypes on differential gene expression, RNA-seq data from CNS tissues obtained from the NYGC ALS consortium were analysed (<https://www.nygenome.org/als-consortium/>). For this analysis both ALS cases and healthy controls were combined. The Salmon quantification tool was used to quantify gene-specific expression levels from previously obtained FASTQ files (<https://salmon.readthedocs.io/en/latest/>). Salmon-generated quant files were imported using the tximport function in R and raw counts were extracted using the DESeqDataSetFromTximport. The DESeq2 R package was implemented to normalise the raw counts through the median-of-ratios method. DESeq2 was also used to detect statistically significant differential gene expression across SVA genotype groups (AA, PA, PP), whereby P represents SVA presence and A represents SVA absence. Results of the analysis were visualised using the ggplot2 package in R. The Kruskal-Wallis multiple comparison test (non-parametric) was applied to determine statistical significance between genotypes. The Wilcoxon test was used for pairwise comparison to assess significant differences between genotypes and obtained *p*-values were corrected for multiple comparisons using false discovery rate (FDR).

##### 4.3. eQTL (expression quantitative trait loci) analysis

Matrix eQTL was applied to evaluate the genetic loci (SVA elements) regulating expression of nearby (*cis*) and distant (*trans*, >1 Mbp away) genes, using the matrixEQTL package in R. For this analysis, the additive linear model was applied with normalised gene expression levels, chromosomal positions of SVAs (hg38) and covariates (sex and age). Effect size estimates were reported as beta values from matrix eQTL. Obtained *p*-values were corrected for multiple testing and the threshold was set to 0.05 for FDR corrected *p*-values.

**Supplementary Materials:** The following supporting information can be downloaded at the website of this paper posted on Preprints.org, Figure S1: Boxplots of *cis*-acting SVA RIPS displaying the top two hits for largest effect size on both gene upregulation and downregulation.; Figure S2: Boxplots of *cis*-acting SVA RIPS displaying four hits for the greatest effect size on gene upregulation. Figure S3: Boxplots of *cis*-acting SVA RIPS displaying three hits for the greatest effect size on gene upregulation.

**Author Contributions:** Conceptualization, L.S.H., A.F. and S.K.; methodology, L.S.H., A.F., A.L.P., and S.K.; formal analysis, L.S.H., A.F. and S.K.; writing—original draft preparation, L.S.H.; writing—review and editing, L.S.H., A.F., A.L.P., V.J.B., J.P.Q., and S.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Andrzej Włodarski Memorial Research Fund (A.F. and J.P.Q.), Andrzej Włodarski Memorial Research PhD scholarship (A.F.), Multiple Sclerosis Western Australia and Perron Institute for Neurological and Translational Science (A.L.P. and S.K.), Darby Rimmer Foundation (V.J.B. and J.P.Q.), and MNDA [ref Quinn/Apr20/875-791] (V.J.B., J.P.Q., A.L.P. and S.K.).

**Informed Consent Statement:** The WGS and RNA sequencing data from the ALS consortium were provided in a de-identified manner with consent obtained by the participating consortia.

**Data Availability Statement:** The sequencing (RNA and WGS) data analysed in this study from the ALS consortium were obtained upon application to the New York Genome Center, and data requests can be made by completing a genetic data request form at [ALSData@nygenome.org](mailto:ALSData@nygenome.org). Additional data from this study will be made available upon reasonable request.

**Acknowledgments:** We would like to acknowledge the ALS Consortium/Target ALS Human Postmortem Tissue Core, New York Genome Center for Genomics of Neurodegenerative Disease, Amyotrophic Lateral Sclerosis Association and TOW Foundation for the RNA-sequencing sequencing data used in this publication. This work was supported by resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Zarei, S.; Carr, K.; Reiley, L.; Diaz, K.; Guerra, O.; Altamirano, P.; Pagani, W.; Lodin, D.; Orozco, G.; Chinea, A. A Comprehensive Review of Amyotrophic Lateral Sclerosis. *Surg Neurol Int* **2015**, *6*, 171. <https://doi.org/10.4103/2152-7806.169561>.
2. Grad, L.I.; Rouleau, G.A.; Ravits, J.; Cashman, N.R. Clinical Spectrum of Amyotrophic Lateral Sclerosis (ALS). *Cold Spring Harb Perspect Med* **2017**, *7*, a024117. <https://doi.org/10.1101/cshperspect.a024117>.
3. Rowland, L.P.; Shneider, N.A. Amyotrophic Lateral Sclerosis. *N Engl J Med* **2001**, *344*, 1688–1700. <https://doi.org/10.1056/NEJM200105313442207>.
4. Masrori, P.; Van Damme, P. Amyotrophic Lateral Sclerosis: A Clinical Review. *Euro J of Neurology* **2020**, *27*, 1918–1929. <https://doi.org/10.1111/ene.14393>.
5. Mejzini, R.; Flynn, L.L.; Pitout, I.L.; Fletcher, S.; Wilton, S.D.; Akkari, P.A. ALS Genetics, Mechanisms, and Therapeutics: Where Are We Now? *Front. Neurosci.* **2019**, *13*, 1310. <https://doi.org/10.3389/fnins.2019.01310>.
6. Brown, R.H.; Al-Chalabi, A. Amyotrophic Lateral Sclerosis. *N Engl J Med* **2017**, *377*, 162–172. <https://doi.org/10.1056/NEJMra1603471>.
7. Van Es, M.A.; Hardiman, O.; Chio, A.; Al-Chalabi, A.; Pasterkamp, R.J.; Veldink, J.H.; Van Den Berg, L.H. Amyotrophic Lateral Sclerosis. *The Lancet* **2017**, *390*, 2084–2098. [https://doi.org/10.1016/S0140-6736\(17\)31287-4](https://doi.org/10.1016/S0140-6736(17)31287-4).
8. Savage, A.L.; Schumann, G.G.; Breen, G.; Bubb, V.J.; Al-Chalabi, A.; Quinn, J.P. Retrotransposons in the Development and Progression of Amyotrophic Lateral Sclerosis. *J Neurol Neurosurg Psychiatry* **2019**, *90*, 284–293. <https://doi.org/10.1136/jnnp-2018-319210>.
9. Gregory, J.M.; Elliott, E.; McDade, K.; Bak, T.; Pal, S.; Chandran, S.; Abrahams, S.; Smith, C. Neuronal Clusterin Expression Is Associated with Cognitive Protection in Amyotrophic Lateral Sclerosis. *Neuropathology Appl Neurobio* **2020**, *46*, 255–263. <https://doi.org/10.1111/nan.12575>.
10. Suzuki, N.; Nishiyama, A.; Warita, H.; Aoki, M. Genetics of Amyotrophic Lateral Sclerosis: Seeking Therapeutic Targets in the Era of Gene Therapy. *J Hum Genet* **2023**, *68*, 131–152. <https://doi.org/10.1038/s10038-022-01055-8>.
11. Zou, Z.-Y.; Zhou, Z.-R.; Che, C.-H.; Liu, C.-Y.; He, R.-L.; Huang, H.-P. Genetic Epidemiology of Amyotrophic Lateral Sclerosis: A Systematic Review and Meta-Analysis. *J Neurol Neurosurg Psychiatry* **2017**, *88*, 540–549. <https://doi.org/10.1136/jnnp-2016-315018>.
12. Al-Chalabi, A.; Fang, F.; Hanby, M.F.; Leigh, P.N.; Shaw, C.E.; Ye, W.; Rijssdijk, F. An Estimate of Amyotrophic Lateral Sclerosis Heritability Using Twin Data. *Journal of Neurology, Neurosurgery & Psychiatry* **2010**, *81*, 1324–1326. <https://doi.org/10.1136/jnnp.2010.207464>.
13. Theunissen, F.; Flynn, L.L.; Anderton, R.S.; Mastaglia, F.; Pytte, J.; Jiang, L.; Hodgetts, S.; Burns, D.K.; Saunders, A.; Fletcher, S.; et al. Structural Variants May Be a Source of Missing Heritability in sALS. *Front. Neurosci.* **2020**, *14*, 47. <https://doi.org/10.3389/fnins.2020.00047>.
14. Roses, A.D.; Akkari, P.A.; Chiba-Falek, O.; Lutz, M.W.; Gottschalk, W.K.; Saunders, A.M.; Saul, B.; Sundseth, S.; Burns, D. Structural Variants Can Be More Informative for Disease Diagnostics, Prognostics and Translation than Current SNP Mapping and Exon Sequencing. *Expert Opinion on Drug Metabolism & Toxicology* **2016**, *12*, 135–147. <https://doi.org/10.1517/17425255.2016.1133586>.
15. GTEx Consortium; Chiang, C.; Scott, A.J.; Davis, J.R.; Tsang, E.K.; Li, X.; Kim, Y.; Hadzic, T.; Damani, F.N.; Ganel, L.; et al. The Impact of Structural Variation on Human Gene Expression. *Nat Genet* **2017**, *49*, 692–699. <https://doi.org/10.1038/ng.3834>.
16. Liscic, R.M. Als and Ftd: Insights into the Disease Mechanisms and Therapeutic Targets. *European Journal of Pharmacology* **2017**, *817*, 2–6. <https://doi.org/10.1016/j.ejphar.2017.10.012>.

17. Elbarbary, R.A.; Lucas, B.A.; Maquat, L.E. Retrotransposons as Regulators of Gene Expression. *Science* **2016**, *351*, aac7247. <https://doi.org/10.1126/science.aac7247>.
18. Ayarpadikannan, S.; Kim, H.-S. The Impact of Transposable Elements in Genome Evolution and Genetic Instability and Their Implications in Various Diseases. *Genomics Inform* **2014**, *12*, 98. <https://doi.org/10.5808/GI.2014.12.3.98>.
19. Gianfrancesco, O.; Geary, B.; Savage, A.L.; Billingsley, K.J.; Bubb, V.J.; Quinn, J.P. The Role of SINE-VNTR-Alu (SVA) Retrotransposons in Shaping the Human Genome. *IJMS* **2019**, *20*, 5977. <https://doi.org/10.3390/ijms20235977>.
20. Prudencio, M.; Gonzales, P.K.; Cook, C.N.; Gendron, T.F.; Daugherty, L.M.; Song, Y.; Ebbert, M.T.W.; Van Blitterswijk, M.; Zhang, Y.-J.; Jansen-West, K.; et al. Repetitive Element Transcripts Are Elevated in the Brain of C9orf72 ALS/FTLD Patients. *Human Molecular Genetics* **2017**, *26*, 3421–3431. <https://doi.org/10.1093/hmg/ddx233>.
21. Hancks, D.C.; Kazazian, H.H. Active Human Retrotransposons: Variation and Disease. *Current Opinion in Genetics & Development* **2012**, *22*, 191–203. <https://doi.org/10.1016/j.gde.2012.02.006>.
22. Fröhlich, A.; Pfaff, A.L.; Bubb, V.J.; Koks, S.; Quinn, J.P. Characterisation of the Function of a SINE-VNTR-Alu Retrotransposon to Modulate Isoform Expression at the MAPT Locus. *Front. Mol. Neurosci.* **2022**, *15*, 815695. <https://doi.org/10.3389/fnmol.2022.815695>.
23. Wang, H.; Xing, J.; Grover, D.; Hedges, D.J.; Han, K.; Walker, J.A.; Batzer, M.A. SVA Elements: A Hominid-Specific Retroposon Family. *Journal of Molecular Biology* **2005**, *354*, 994–1007. <https://doi.org/10.1016/j.jmb.2005.09.085>.
24. Savage, A.L.; Bubb, V.J.; Breen, G.; Quinn, J.P. Characterisation of the Potential Function of SVA Retrotransposons to Modulate Gene Expression Patterns. *BMC Evol Biol* **2013**, *13*, 101. <https://doi.org/10.1186/1471-2148-13-101>.
25. Quinn, J.P.; Bubb, V.J. SVA Retrotransposons as Modulators of Gene Expression. *Mobile Genetic Elements* **2014**, *4*, e32102. <https://doi.org/10.4161/mge.32102>.
26. Pfaff, A.L.; Bubb, V.J.; Quinn, J.P.; Koks, S. Reference SVA Insertion Polymorphisms Are Associated with Parkinson's Disease Progression and Differential Gene Expression. *npj Parkinsons Dis.* **2021**, *7*, 44. <https://doi.org/10.1038/s41531-021-00189-4>.
27. Verpillat, P.; Camuzat, A.; Hannequin, D.; Thomas-Anterion, C.; Puel, M.; Belliard, S.; Dubois, B.; Didic, M.; Michel, B.-F.; Lacomblez, L.; et al. Association Between the Extended Tau Haplotype and Frontotemporal Dementia. *Arch Neurol* **2002**, *59*, 935. <https://doi.org/10.1001/archneur.59.6.935>.
28. Wider, C.; Vilariño-Güell, C.; Jasinska-Myga, B.; Heckman, M.G.; Soto-Ortolaza, A.I.; Cobb, S.A.; Aasly, J.O.; Gibson, J.M.; Lynch, T.; Uitti, R.J.; et al. Association of the MAPT Locus with Parkinson's Disease. *Euro J of Neurology* **2010**, *17*, 483–486. <https://doi.org/10.1111/j.1468-1331.2009.02847.x>.
29. Sánchez-Juan, P.; Moreno, S.; De Rojas, I.; Hernández, I.; Valero, S.; Alegret, M.; Montreal, L.; García González, P.; Lage, C.; López-García, S.; et al. The MAPT H1 Haplotype Is a Risk Factor for Alzheimer's Disease in APOE E4 Non-Carriers. *Front. Aging Neurosci.* **2019**, *11*, 327. <https://doi.org/10.3389/fnagi.2019.00327>.
30. Fröhlich, A.; Hughes, L.S.; Middlehurst, B.; Pfaff, A.L.; Bubb, V.J.; Koks, S.; Quinn, J.P. CRISPR Deletion of a SINE-VNTR-Alu (SVA\_67) Retrotransposon Demonstrates Its Ability to Differentially Modulate Gene Expression at the MAPT Locus. *Front. Neurol.* **2023**, *14*, 1273036. <https://doi.org/10.3389/fneur.2023.1273036>.
31. Van Rheenen, W.; Van Der Spek, R.A.A.; Bakker, M.K.; Van Vugt, J.J.F.A.; Hop, P.J.; Zwamborn, R.A.J.; De Klein, N.; Westra, H.-J.; Bakker, O.B.; Deelen, P.; et al. Common and Rare Variant Association Analyses in Amyotrophic Lateral Sclerosis Identify 15 Risk Loci with Distinct Genetic Architectures and Neuron-Specific Biology. *Nat Genet* **2021**, *53*, 1636–1648. <https://doi.org/10.1038/s41588-021-00973-1>.
32. Nona, R.J.; Greer, J.M.; Henderson, R.D.; McCombe, P.A. HLA and Amyotrophic Lateral Sclerosis: A Systematic Review and Meta-Analysis. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration* **2023**, *24*, 24–32. <https://doi.org/10.1080/21678421.2022.2078665>.
33. Koks, S.; Pfaff, A.L.; Bubb, V.J.; Quinn, J.P. Expression Quantitative Trait Loci (eQTLs) Associated with Retrotransposons Demonstrate Their Modulatory Effect on the Transcriptome. *IJMS* **2021**, *22*, 6319. <https://doi.org/10.3390/ijms22126319>.
34. Wang, L.; Rishishwar, L.; Mariño-Ramírez, L.; Jordan, I.K. Human Population-Specific Gene Expression and Transcriptional Network Modification with Polymorphic Transposable Elements. *Nucleic Acids Res* **2017**, *gkw1286*. <https://doi.org/10.1093/nar/gkw1286>.
35. Pugacheva, E.M.; Teplyakov, E.; Wu, Q.; Li, J.; Chen, C.; Meng, C.; Liu, J.; Robinson, S.; Loukinov, D.; Boukaba, A.; et al. The Cancer-Associated CTCFL/BORIS Protein Targets Multiple Classes of Genomic Repeats, with a Distinct Binding and Functional Preference for Humanoid-Specific SVA Transposable Elements. *Epigenetics & Chromatin* **2016**, *9*, 35. <https://doi.org/10.1186/s13072-016-0084-2>.
36. Sun, X.; Zhang, J.; Cao, C. CTCF and Its Partners: Shaper of 3D Genome during Development. *Genes* **2022**, *13*, 1383. <https://doi.org/10.3390/genes13081383>.

37. Aggarwal, S.; Snaidero, N.; Pähler, G.; Frey, S.; Sánchez, P.; Zweckstetter, M.; Janshoff, A.; Schneider, A.; Weil, M.-T.; Schaap, I.A.T.; et al. Myelin Membrane Assembly Is Driven by a Phase Transition of Myelin Basic Proteins Into a Cohesive Protein Meshwork. *PLoS Biol* **2013**, *11*, e1001577. <https://doi.org/10.1371/journal.pbio.1001577>.

38. Kang, S.H.; Li, Y.; Fukaya, M.; Lorenzini, I.; Cleveland, D.W.; Ostrow, L.W.; Rothstein, J.D.; Bergles, D.E. Degeneration and Impaired Regeneration of Gray Matter Oligodendrocytes in Amyotrophic Lateral Sclerosis. *Nat Neurosci* **2013**, *16*, 571–579. <https://doi.org/10.1038/nn.3357>.

39. Raffaele, S.; Boccazzini, M.; Fumagalli, M. Oligodendrocyte Dysfunction in Amyotrophic Lateral Sclerosis: Mechanisms and Therapeutic Perspectives. *Cells* **2021**, *10*, 565. <https://doi.org/10.3390/cells10030565>.

40. Lubetzki, C.; Zalc, B.; Williams, A.; Stadelmann, C.; Stankoff, B. Remyelination in Multiple Sclerosis: From Basic Science to Clinical Translation. *The Lancet Neurology* **2020**, *19*, 678–688. [https://doi.org/10.1016/S1474-4422\(20\)30140-X](https://doi.org/10.1016/S1474-4422(20)30140-X).

41. Lorente Pons, A.; Higginbottom, A.; Cooper-Knock, J.; Alraifah, A.; Alofi, E.; Kirby, J.; Shaw, P.J.; Wood, J.D.; Highley, J.R. Oligodendrocyte Pathology Exceeds Axonal Pathology in White Matter in Human Amyotrophic Lateral Sclerosis. *The Journal of Pathology* **2020**, *251*, 262–271. <https://doi.org/10.1002/path.5455>.

42. Hübner, C.A.; Orth, U.; Sennig, A.; Steglich, C.; Kohlschütter, A.; Korinthenberg, R.; Gal, A. Seventeen Novel *PLP1* Mutations in Patients with Pelizaeus-Merzbacher Disease: MUTATIONS IN BRIEF. *Hum. Mutat.* **2005**, *25*, 321–322. <https://doi.org/10.1002/humu.9314>.

43. Cloake, N.; Yan, J.; Aminian, A.; Pender, M.; Greer, J. *PLP1* Mutations in Patients with Multiple Sclerosis: Identification of a New Mutation and Potential Pathogenicity of the Mutations. *JCM* **2018**, *7*, 342. <https://doi.org/10.3390/jcm7100342>.

44. Cruz-Rivera, Y.E.; Perez-Morales, J.; Santiago, Y.M.; Gonzalez, V.M.; Morales, L.; Cabrera-Rios, M.; Isaza, C.E. A Selection of Important Genes and Their Correlated Behavior in Alzheimer's Disease. *JAD* **2018**, *65*, 193–205. <https://doi.org/10.3233/JAD-170799>.

45. Wang, Y.; Xu, E.; Musich, P.R.; Lin, F. Mitochondrial Dysfunction in Neurodegenerative Diseases and the Potential Countermeasure. *CNS Neurosci Ther* **2019**, *25*, 816–824. <https://doi.org/10.1111/cns.13116>.

46. Misra, M.K.; Damotte, V.; Hollenbach, J.A. The Immunogenetics of Neurological Disease. *Immunology* **2018**, *153*, 399–414. <https://doi.org/10.1111/imm.12869>.

47. Antel, J.P.; Arnason, B.G.W.; Fuller, T.C.; Lehrich, J.R. Histocompatibility Typing in Amyotrophic Lateral Sclerosis. *Archives of Neurology* **1976**, *33*, 423–425. <https://doi.org/10.1001/archneur.1976.00500060029007>.

48. Behan, P.; Durward, W.; Dick, H. HISTOCOMPATIBILITY ANTIGENS ASSOCIATED WITH MOTOR-NEURONE DISEASE. *The Lancet* **1976**, *308*, 803. [https://doi.org/10.1016/S0140-6736\(76\)90638-3](https://doi.org/10.1016/S0140-6736(76)90638-3).

49. Jokelainen, M.; Tilikainen, A.; Lapinleimu, K. Polio Antibodies and HLA Antigens in Amyotrophic Lateral Sclerosis. *Tissue Antigens* **1977**, *10*, 259–266. <https://doi.org/10.1111/j.1399-0039.1977.tb00755.x>.

50. Kott, E.; Livni, E.; Zamir, R.; Kuritzky, A. Cell-Mediated Immunity to Polio and HLA Antigens in Amyotrophic Lateral Sclerosis. *Neurology* **1979**, *29*, 1040–1040. <https://doi.org/10.1212/WNL.29.7.1040>.

51. Kulski, J.K.; Pfaff, A.L.; Marney, L.D.; Fröhlich, A.; Bubb, V.J.; Quinn, J.P.; Koks, S. Regulation of Expression Quantitative Trait Loci by SVA Retrotransposons within the Major Histocompatibility Complex. *Exp Biol Med (Maywood)* **2023**, *15353702231209411*. <https://doi.org/10.1177/15353702231209411>.

52. Pfaff, A.L.; Bubb, V.J.; Quinn, J.P.; Koks, S. A Genome-Wide Screen for the Exonisation of Reference SINE-VNTR-Alus and Their Expression in CNS Tissues of Individuals with Amyotrophic Lateral Sclerosis. *IJMS* **2023**, *24*, 11548. <https://doi.org/10.3390/ijms241411548>.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.