Stem Cell Reports
Article

OPEN ACCESS

Marked Differences in *C9orf72* Methylation Status and Isoform Expression between C9/ALS Human Embryonic and Induced Pluripotent Stem Cells

Yaara Cohen-Hadad,[1] Gheona Altarescu,[2] Talia Eldar-Geva,[3] Ephrat Levi-Lahad,[2] Ming Zhang,[4] Ekaterina Rogaeva,[4] Marc Gotkine,[5] Osnat Bartok,[6] Reut Ashwal-Fluss,[6] Sebastian Kadener,[6] Silvina Epsztejn-Litman,[1] and Rachel Eiges[1,*]

[1]Stem Cell Research Laboratory
[2]Zohar PGD Laboratory, Medical Genetics Institute
[3]IVF Unit
Shaare Zedek Medical Center, Hebrew University School of Medicine, Jerusalem 91031, Israel
[4]Tanz Centre for Research in Neurodegenerative Diseases, University of Toronto, Toronto, ON M5S 3H2, Canada
[5]Department of Neurology, Hadassah Medical Center, Hebrew University, Jerusalem 91120, Israel
[6]Biological Chemistry Department, Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem 91904, Israel
*Correspondence: rachela@szmc.org.il
http://dx.doi.org/10.1016/j.stemcr.2016.09.011

SUMMARY

We established two human embryonic stem cell (hESC) lines with a GGGGCC expansion in the *C9orf72* gene (C9), and compared them with haploidentical and unrelated C9 induced pluripotent stem cells (iPSCs). We found a marked difference in C9 methylation between the cells. hESCs and parental fibroblasts are entirely unmethylated while the iPSCs are hypermethylated. In addition, we show that the expansion alters promoter usage and interferes with the proper splicing of intron 1, eventually leading to the accumulation of repeat-containing mRNA following neural differentiation. These changes are attenuated in C9 iPSCs, presumably owing to hypermethylation. Altogether, this study highlights the importance of neural differentiation in the pathogenesis of disease and points to the potential role of hypermethylation as a neuroprotective mechanism against pathogenic mRNAs, envisaging a milder phenotype in C9 iPSCs.

INTRODUCTION

Amyotrophic lateral sclerosis (ALS, OMIM #105400) is characterized by progressive muscle weakness and atrophy due to the degeneration of upper and lower motor neurons in the brain and spinal cord, while frontotemporal degeneration (FTD, OMIM #600274) affects behavior and cognition, and is caused by the preferential loss of neurons in the frontal and temporal lobe cortices. Most people who develop ALS/ALS-FTD are between the ages of 40 and 70 years, and die within 2–5 years from diagnosis. Currently there is no cure for ALS-FTD.

The leading known cause of ALS-FTD is a GGGGCC repeat expansion in the first intron of the *C9orf72* gene (termed C9 mutation), between noncoding exons 1a and 1b (DeJesus-Hernandez et al., 2011; Dols-Icardo et al., 2014). This mutation can manifest as ALS, FTD, or a combined phenotype, and accounts for 20%–80% of familial and 5%–15% of sporadic ALS and FTD cases (reviewed by Cruts et al., 2013). While in most people the number of GGGGCC repeats is steady and varies between 2 and 19 units, in ALS-FTD it abnormally expands to more than 30 copies and becomes increasingly unstable (Dols-Icardo et al., 2014). The mechanism by which the C9 mutation leads to selective death of neurons is unknown, and the normal function of *C9orf72* is just beginning to be defined. Multiple mechanisms for C9/ALS-FTD have been suggested, including haploinsufficiency, RNA toxicity, and abnormal translation of expanded repeat sequences by RAN translation (reviewed by Gendron et al., 2014). However, whether the C9 related neurodegeneration is initiated via a gain-of-function (toxic RNA and/or unconventional dipeptide translation) or a loss-of-function mechanism is still under investigation in animal and cellular models.

The GGGGCC repeat sequence is flanked by two CpG islands (CGIs) within a ~1-kb region that spans from the promoter sequence into intron 1 of *C9orf72*. Like many CGIs in the genome, this region typically remains free of DNA methylation in subjects with 2–90 repeats. Yet large expansions lead to the formation of one large CGI with abnormal CpG methylation at the repeats (Xi et al., 2015b), which could spread to the 5′ upstream CGI in up to 37% of C9 cases (Belzil et al., 2014; Gijselinck et al., 2015; Liu et al., 2014; Russ et al., 2015; Xi et al., 2013, 2014, 2015a, 2015b). Hypermethylation is suggested to be coupled with the local gain of repressive histone modifications (H3K9me3 and H3K27me3) (Belzil et al., 2013; Zeier et al., 2015). It is still unclear how, or whether, hypermethylation contributes to disease pathogenesis. While some reports demonstrate an overall decrease in *C9orf72* transcription, others show a change in the relative distribution between the three different mRNA isoforms, favoring transcription from exon 1a (V1 and V3, NM_145005.5 and NM_001256054.1, respectively) over exon 1b (V2, NM_018325.3) (Donnelly

et al., 2013; Haeusler et al., 2014; Lee et al., 2013). While previous reports failed to detect a correlation between hypermethylation and ALS versus FTD phenotype (Xi et al., 2015b), experimental evidence demonstrates that *C9orf72* haploinsufficiency affects cell morphology and function of motor neurons in zebrafish (Ciura et al., 2013). On the other hand, hypermethylation protects against the accumulation of pathogenic RNA foci and dipeptides, caused by the repeat-containing mRNA variants 1 and 3 (Bauer, 2016; Day and Roberson, 2015; Liu et al., 2014). These conflicting results warrant further investigation regarding the contribution and timing of *C9orf72* hypermethylation in ALS-FTD pathogenesis, and the discrepancies may be resolved by the use of in vitro derived neurons from C9/ALS-FTD pluripotent cells. Indeed, induced pluripotent stem cells (iPSCs) from C9/ALS patient fibroblasts have already been used to generate motor neurons in culture that recapitulate the key neuropathological features of FTD-ALS (Almeida et al., 2013; Cooper-Knock et al., 2014, 2015; Devlin et al., 2015; Donnelly et al., 2013; Li et al., 2015; Peters et al., 2015; Rossi et al., 2015; Sareen et al., 2013; Satoh et al., 2014; Wainger et al., 2014). Nevertheless, the epigenetic aspects of the disease have never been addressed using this model system. The aim of this study is to characterize the methylation state of the expanded region and explore its effect on *C9orf72* variant transcription in C9/ALS human embryonic stem cells (hESCs), and compare them with that of their haploidentical (mother-to-child genetic identity) and unrelated C9 iPSCs before and after differentiation.

RESULTS

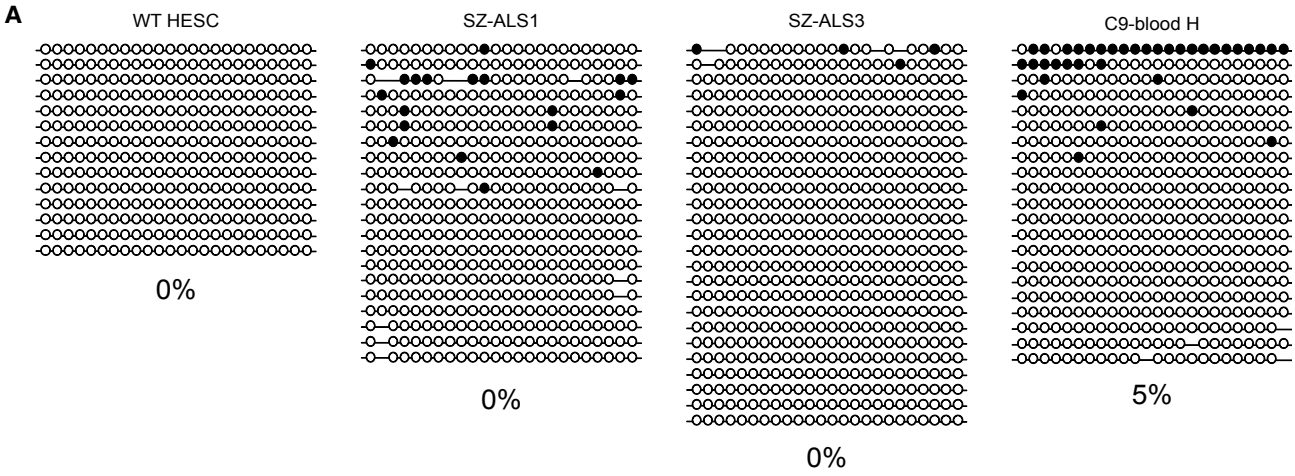
Derivation and Characterization of C9/hESC Lines

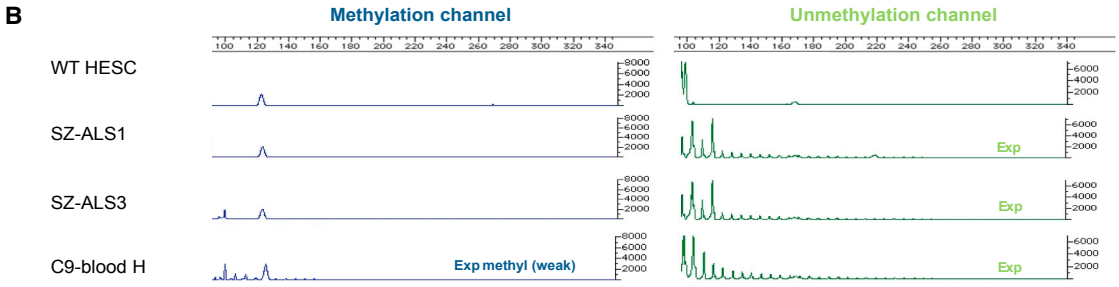
We established two hESC lines with a C9 mutation (SZ-ALS1 and SZ-ALS3) from embryos, which were obtained through preimplantation genetic diagnosis (PGD) and donated for cell line derivation by a family in which the mother was an expansion carrier (patient H, 30 years old, originally diagnosed as a carrier of an expansion with >40 repeats in blood by a repeat primed PCR (rp-PCR); data not shown). Our newly established C9 hESC lines display the key features of pluripotent cells, namely unrestricted growth in culture, expression of undifferentiated cell-specific markers, and potential to differentiate into a wide range of cell types by forming teratomas (Figure S1A, B, D). Chromosome analysis by Giemsa staining demonstrated a 46(XX) karyotype for SZ-ALS1 and a 45(X0) for SZ-ALS3 (Figure S1C). Southern blot analysis identified a GGGGCC expansion of at least ~270 repeats in both cell lines (Figure S1E).

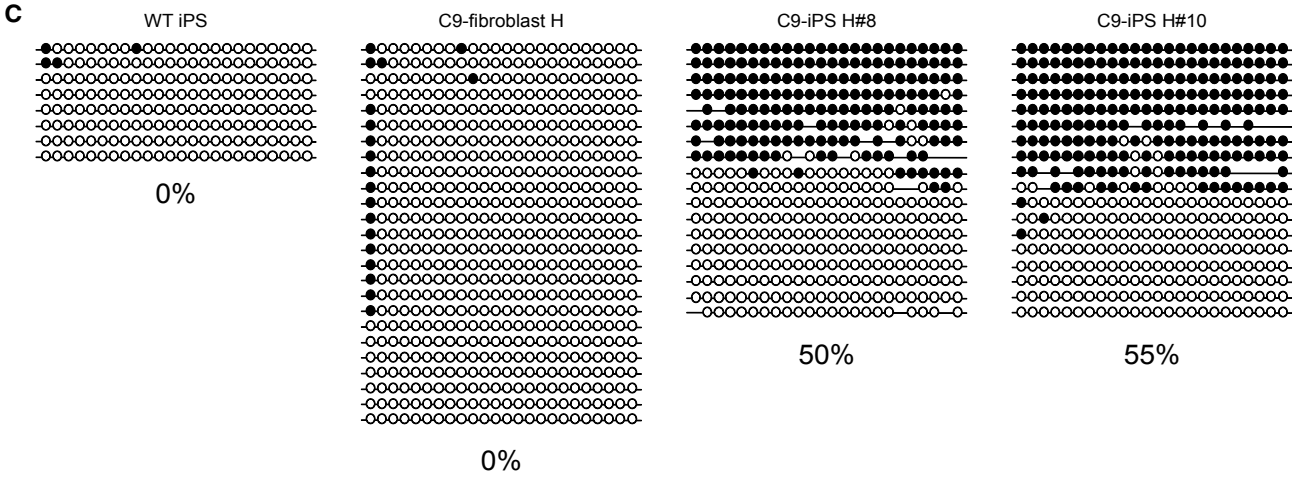
Analysis of *C9orf72* Methylation in C9 hESCs and Their Haploidentical iPSCs

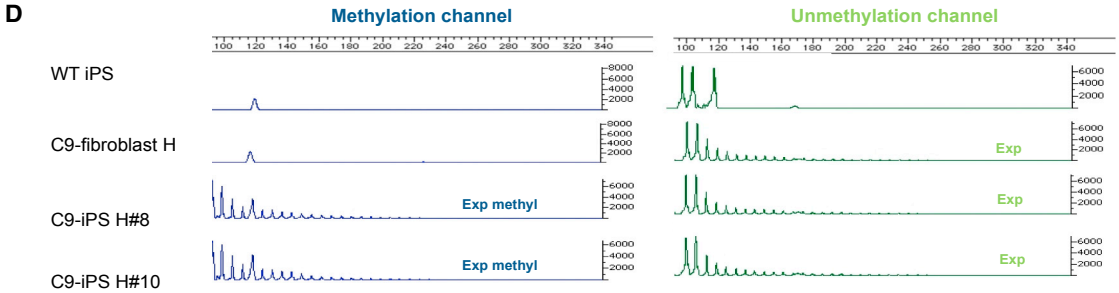
Considering the accumulated data regarding hypermethylation in C9 carriers, we aimed to determine whether hypermethylation is already established in the undifferentiated state. Therefore we examined methylation levels, 200 bp upstream of the 5′ end of the GGGGCC repeats, by bisulfite DNA colony sequencing in the C9 hESCs (24 CpG sites). Interestingly, despite the presence of a large expansion, methylation was almost 0% in both cell lines (Figure 1A). To exclude the possibility that methylation had already begun, but failed to spread further upstream to the 5′ CGI, we searched for methylation at the 5′ end of the repeats by carrying out a qualitative $(G_4C_2)_n$-methylation assay that is sensitive enough to detect repeat methylation in the mixture containing only 2%–5% highly methylated DNA (Xi et al., 2015b). Here again, methylation was undetectable in the mutant hESCs (Figure 1B). In contrast, the expanded allele in whole blood cells of the mother revealed methylation, although weak, in the expansion itself and the 5′ flanking region (estimated as 5% by bisulfite sequencing) (Figures 1A and 1B).

To generate a complement model system, we established haploidentical (genetically half-identical) iPSCs (more than ten different clones) from skin fibroblasts of the currently asymptomatic mother with a C9 expansion, who donated the embryos for hESC line derivation. The primary fibroblasts, which had an unmethylated ~700-repeat expansion (Figure S2D), were reprogrammed by standard protocol using Yamanaka's four transcription factors (Takahashi et al., 2007). The newly established C9 iPSCs with typical ESC morphology demonstrated unrestricted self-renewal, expressed undifferentiated cell-specific markers, and presented a normal karyotype (Figure S2A–C). By comparing their methylation status with that of the C9 hESCs and parental fibroblasts, we found that the C9 iPSC clones were unusual in their methylation levels. Unlike hESCs and primary fibroblasts, methylation was exceptionally high in the iPSCs. Bisulfite DNA colony sequencing of the region upstream of the expansion identified 50% fully methylated DNA copies in all iPSC clones, suggesting complete methylation of the mutant allele (Figure 1C). The $(G_4C_2)_n$-methylation assay revealed hypermethylation of the repeats at the 5′ end of the expansion exclusively in C9 iPSCs (Figure 1D). However, while the upstream flanking region of the mutant allele was completely methylated based on bisulfite DNA sequencing (50%), the GGGGCC repeats themselves were not methylated in all DNA molecules (evidenced by signals of repeats beyond the normal range in both the green and blue channels in the $(G4C2)_n$-methylation assay). Importantly, differential methylation between the iPSCs and their parental fibroblasts could not be attributed to the increase in repeat









(legend on next page)

number, since expansion size remained steady following cell reprogramming, as determined by Southern blot analysis (Figure S2D). In addition, as expansion size in the C9 hESCs is well above the threshold necessary to elicit hypermethylation in *C9orf72* in any other cell type thus far examined, it is very unlikely that the striking differences in hypermethylation between the C9 iPSCs and their C9 hESCs counterparts stem from a difference in expansion size. We further confirmed the fundamental difference in the epigenetic state of the C9 mutation between both cell types by demonstrating significant enrichments for the repressive histone modification H3K9me3 by chromatin immunoprecipitation (ChIP) analysis exclusively in mutant iPSCs (Figure 2A). No enrichments for H3K27me3 could be detected in wild-type (WT) or affected hESCs and iPSCs (Figure 2B). Taken together, these findings suggest that reprogramming alters the epigenetic state of this region as a consequence of the expansion in iPSCs, likely spreading from the 5′ border of the CGI toward the repeats.

Methylation Analysis in C9 iPSCs Derived from an Unrelated Symptomatic ALS Patient

To further corroborate our findings and examine whether methylation is affected by age or disease symptoms, we also generated iPSCs from a skin biopsy of a 65-year-old C9 ALS patient, 2 years following disease onset (patient M). Southern blot analysis demonstrated the presence of a ~2,700 repeat expansion in the primary fibroblasts of the patient (Figure S2D). Reprogramming of these fibroblasts led to the establishment of more than ten different C9 iPSC clones. The newly established C9 ALS patient-derived iPSCs had the typical characteristics of hESCs (Figure S2). By comparing expansion size and methylation levels between the iPSCs and their parental fibroblasts we found that, although expansion size remained the same, methylation levels at the 5′ CGI dramatically increased from 0% in parental fibroblasts to 50% in affected iPSCs, suggesting methylation levels of 100% on the expanded allele as determined by bisulfite colony sequencing (Figure 3A). The $(G_4C_2)_n$-methylation assay indicated methylation of the repeats at the 5′ end of the expansion, but not in the parental fibroblasts (Figure 3B). These results provide further evidence that somatic cell reprogramming excessively hypermethylates the 5′ UTR of the *C9orf72* locus. In addition, we examined the methylation status of a different gene, *SIGLEC6*, as a reference locus for aberrant de novo methylation by transcription factor reprogramming, independent of the method used (integratable and nonintegratable vectors) (Huang et al., 2014). Here again, methylation was exclusively acquired in all iPSCs (WT and C9-iPSCs), but was completely absent in primary fibroblasts (patients H and M) or hESCs (SZ-ALS1 and SZ-ALS3) (Figure S3). However, unlike in *C9orf72*, hypermethylation in *SIGLEC6* is not conditioned by a change in the DNA sequence.

The Effect of Differentiation on the Methylation Status of *C9orf72*

To explore the effect of differentiation on the methylation status of the C9 mutation, we induced the unmethylated mutant hESCs (SZ-ALS1 and SZ-ALS3) and C9 hypermethylated iPSC clones (H#8 and M#9) to differentiate into disease-relevant cell types, i.e., neural precursor cells (NPCs) and neural-enriched teratomas. For NPCs, we applied a commonly used and highly efficient differentiation protocol that relies on the generation of neural rosettes by the use of two inhibitors (dorsomorphin and SB431542) (Figure S4A) (Kim et al., 2010). Differentiation efficiency into NPCs was assessed by fluorescence-activated cell sorting (FACS) analysis for NCAM1-positive cells (above 90%, Figure S4B), and by monitoring for the expression of early neural differentiation markers *SOX2*, *Nestin*, and *PAX6* by RT-PCR (Figure S4C). In addition we took advantage of the teratomas, which are highly enriched for mature neurons (assessed by histological examination [H&E staining, Figures S1] and the presence of Tuj1-expressing cells [Figure S4D]), to explore the effect of differentiation on the methylation levels of the mutation. Interestingly, methylation levels remained unchanged as determined by bisulfite colony sequencing. In the NPCs and teratomas generated from C9 hESCs, methylation remained at 0%, while in the NPCs and teratomas produced from C9 iPSCs

Figure 1. Analysis of Methylation Levels in the Upstream Region of the G_4C_2 Repeats and in the G_4C_2 Repeat Itself in C9 hESCs, Maternal Blood, Primary Fibroblasts, and iPSCs Derived from Them

(A) Methylation levels upstream of the repeats were determined by bisulfite sequencing of a region 200 bp upstream of the repeats (altogether 24 CpG sites) in SZ-ALS1, SZ-ALS3, and wild-type (WT) (SZ-13) hESCs and whole blood from the mother (patient H). Each line represents a single DNA molecule (upstream region of the repeats), with methylated and unmethylated CpGs designated by black and white circles, respectively.

(B) Results of the $(G_4C_2)_n$-methylation assay in WT and affected (C9/ALS) hESCs and blood cells of the mother. The left panel (blue channel) represents methylated alleles, while the right panel (green channel) represents unmethylated alleles.

(C) Methylation levels upstream to the repeats as determined by single-molecule bisulfite sequencing of the same region shown in (A) in WT iPSCs, C9 parental fibroblasts (from patient H) and iPSCs derived from them (C9-iPS#H8 and iPS#H10).

(D) Results of the $(G_4C_2)_n$-methylation assay in WT and C9 affected cells described in (C).

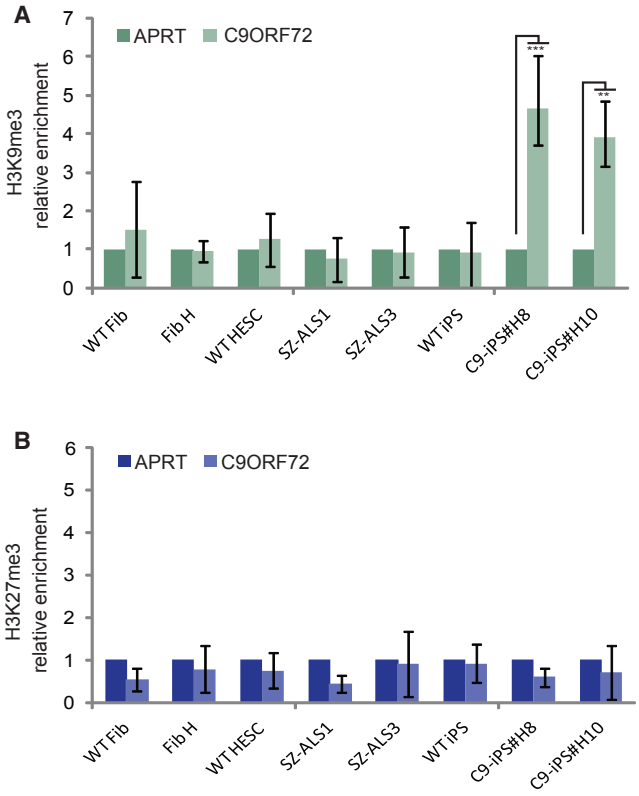


Figure 2. ChIP Analysis for H3K9me3 and H3K27me3 in WT and C9 Fibroblasts, Undifferentiated hESCs, and iPSCs

Real-time PCR ChIP analysis for (A) H3K9me3 and (B) H3K27me3 in WT and C9 affected hESCs (SZ-ALS1 and SZ-ALS3), haploidentical parental fibroblasts (Fib-H), and iPSCs derived from them (C9-iPS #H8 and #H10). *APRT* was used as a negative control for both modifications. Negative controls were set to 1. The data in each panel represent an average of three to five independent ChIP experiments. Error bars represent SE (paired t test, **p < 0.01, ***p < 0.001).

methylation remained at 50% (Figures 4A and 4C). These results were further validated by the $(G_4C_2)_n$-methylation assay (Figures 4B and 4D).

The Effect of Hypermethylation on *C9orf72* Expression

To explore whether the expansion alters *C9orf72* gene expression and to examine whether it corresponds with hypermethylation, we determined the mRNA levels of the three isoforms (see schematic illustration in Figure 5A) in WT as well as C9 hESCs and iPSCs. Targeting transcript variants 1 (V1), 2 (V2), and 3 (V3) individually (amplicons spanning exon boundaries 1a-2 and 1b-2), and all together (amplicon spanning exon boundaries 2–3) with TaqMan probes, we show that while the general levels of *C9orf72* (V1 + V2 + V3) as well as of V2 and V3 remain unchanged between WT and C9 cells in hESCs and iPSCs (Figure 5B), V1 alone is significantly higher in mutant versus WT undifferentiated cells (t test for equal variances, $p < 0.05$) (Figure 5B). Nevertheless, upregulation of V1 is at least 2.5-fold higher in C9 hESCs relative to C9 iPSCs (t test for equal variances, $p < 0.05$) (Figure 5B). Given the high abundance of V2 relative to the other variants (cycle threshold [Ct] values of 26–27 versus 27–34), the change in V1, even though significant, reflects minute differences overlooked when measuring overall levels of *C9orf72* mRNA. We conclude that the GGGGCC expansion alters the region leading to the increased expression of *C9orf72* V1 transcripts in C9 iPSCs.

To explore whether neural differentiation contributes to the change in *C9orf72* transcription, we also assayed the various isoforms upon differentiation into NPCs and teratomas. Interestingly, we found that differentiation of C9 hESCs into both cell types further enhanced the transcription of exon 1a-bearing transcripts by increasing the expression of both V3 and V1 (Figure 5C). This is in contrast to the upregulation of V2 (the most abundant transcript) in NPCs and teratomas of C9 iPSCs, contributing to the overall increase in *C9orf72* mRNA levels (Figure 5C). Jointly, these findings demonstrate that the C9 mutation, together with neural differentiation, favors transcription from the upstream promoter. This effect is largely restricted in iPSCs, presumably due to hypermethylation.

Considering the central role of intron 1-retaining transcripts in the pathology of C9/ALS-FTD disease (Niblock et al., 2016), we aimed to search for the expression of repeat-containing mRNAs in C9 cells. First we generated rRNA-depleted RNA sequencing (RNA-seq) libraries from WT as well as C9-mutant undifferentiated hESCs (two cell lines) and iPSCs (five clones), and RNA deep-sequenced them utilizing next-generation sequencing (see Figures S5A and S5B for general coverage and amount of reads in intron 1 relative to exon 2 of *C9orf72* and the RNA-seq data at GEO: GSE87273). As anticipated and in agreement with the report of Niblock et al. (2016), by zooming in to the 5′ part of *C9orf72* we found preferential retention of intron 1 in both C9 hESCs and iPSCs (Figure 6A), indicating the propensity of the mutation to interfere with proper splicing of this region. However, quantifying this change between mutant and unaffected cells demonstrates a greater effect in hESCs (4.5-fold increase) relative to iPSCs (only 2-fold increase) (Figure 6B). Considering that C9 cells carry both the allele with expansion and a normal allele, these fold changes are predicted to reflect even greater differences in expression between WT and mutant alleles. We conclude that the shift in promoter usage is coupled with the general tendency to retain intron 1 by the expansion. Next, we aimed to explore whether this may promote accumulation of repeat-containing mRNA transcripts (exon

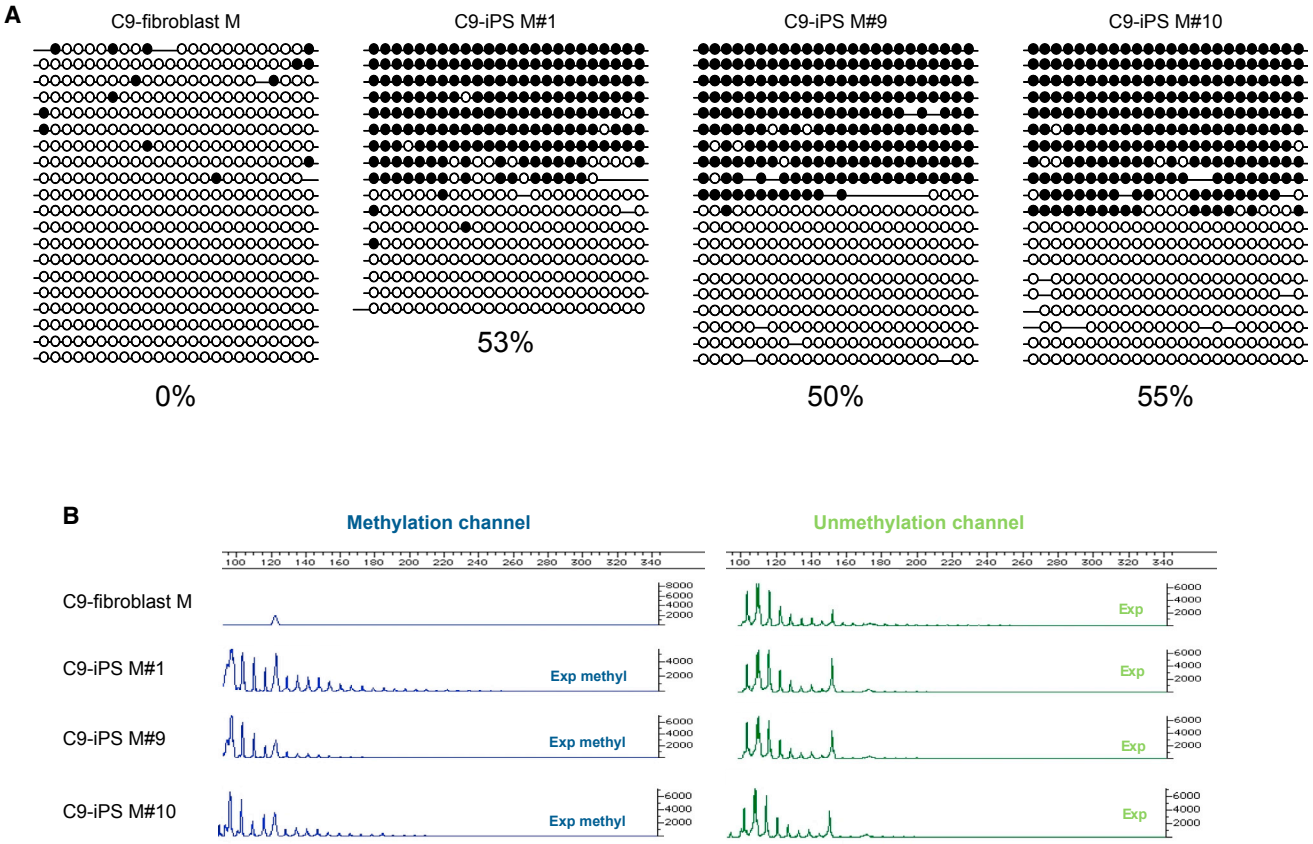


Figure 3. Analysis of Methylation Levels in the Region Upstream of the G_4C_2 Repeats and in the G_4C_2 Repeat Itself in Primary Fibroblasts and iPSCs Derived from Patient M

(A) Methylation levels upstream of the repeats were determined by bisulfite sequencing of the same region as in Figure 1A in WT and C9 affected parental fibroblasts derived from a 65-year-old C9/ALS-manifesting patient (patient M) and iPSCs derived from them (C9 iPS#M1, iPS#M9, and iPS#H10). Each line represents the upstream region of the repeats, with methylated and unmethylated CpGs designated by black and white circles, respectively.

(B) Results of the $(G_4C_2)_n$-methylation assay in primary fibroblasts and iPSC clones derived from patient M. Left panel (blue channel) represents methylated alleles while the right panel (green channel) represents unmethylated alleles.

1a-initiating transcripts that retain repeat expansion), which form the underlying mechanism for protein/RNA gain of function in C9/ALS-FTD. cDNA sequencing at the boundary between intron 1 and exon 5 confirmed the existence of such transcripts in our cells (Figure S5). Using primers located upstream of the repeats that specifically target repeat-containing mRNAs in undifferentiated (hESCs and iPSCs, Figure 6C) and differentiated (NPCs and teratomas, Figure 6D) cells, we monitored for the existence of these unusual mRNAs by real-time RT-PCR. Importantly, while no difference could be detected between C9 and WT undifferentiated cells (hESCs and iPSCs), their level became significantly higher upon differentiation into NPCs and teratomas exclusively in C9 hESCs (7.8- and 2.6-fold change, respectively) (Figures 6D and S5), and not in C9 iPSCs. Hence, our findings provide evidence that the C9 mutation interferes with proper splicing of intron 1, thereby enhancing the formation of potentially pathogenic repeat-containing mRNAs upon differentiation. Importantly, this effect is missing in C9 iPSCs, presumably due to hypermethylation.

DISCUSSION

We report on the derivation and full characterization of two hESC lines (SZ-ALS1 and SZ-ALS3) with a GGGGCC expansion of approximately 270 repeats. Our C9 hESC lines were established from embryos obtained through PGD from a woman with a C9 mutation with >40 repeats in her peripheral blood. Interestingly, despite the sufficiently large expansion, both cell lines were completely

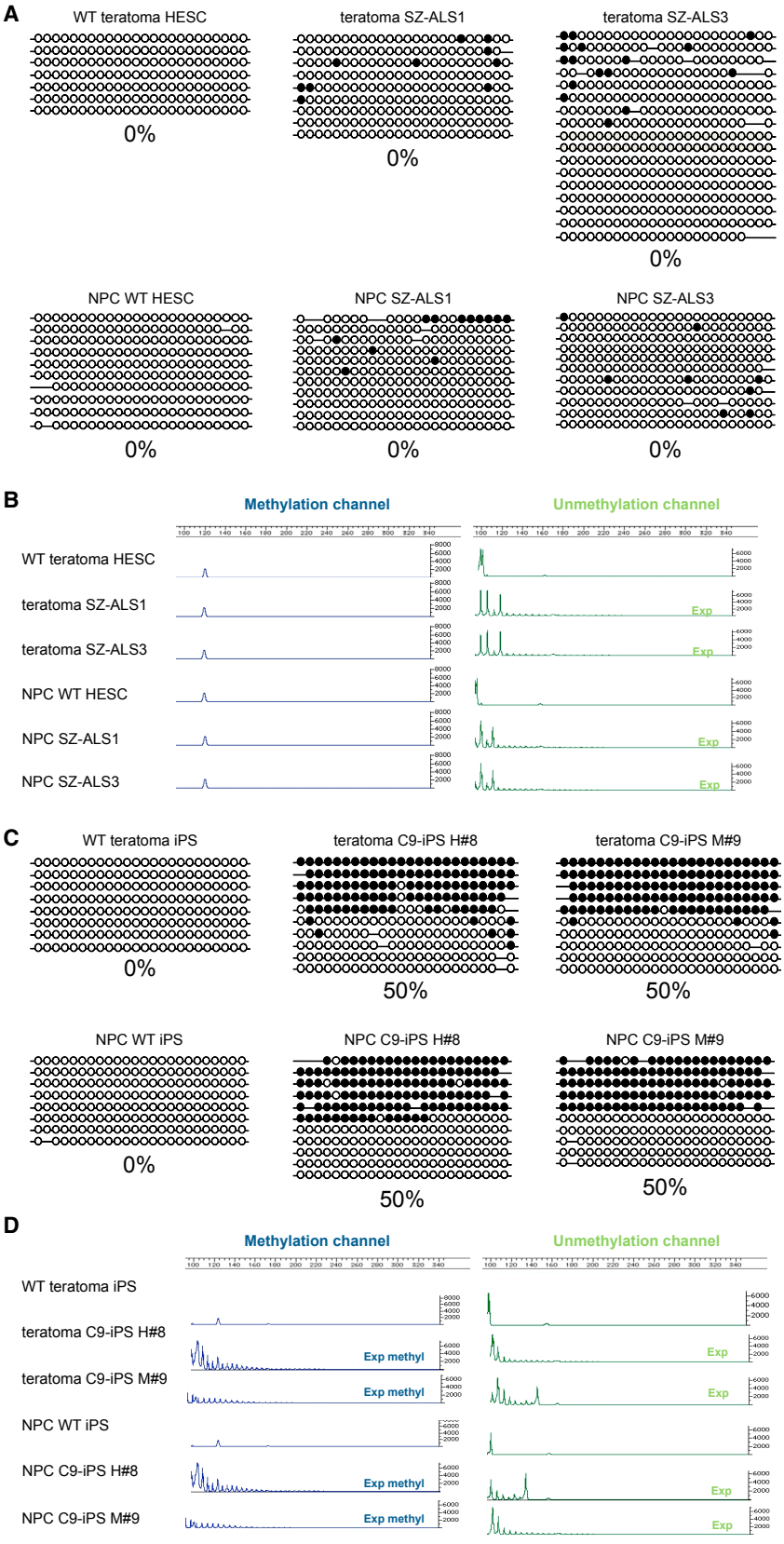


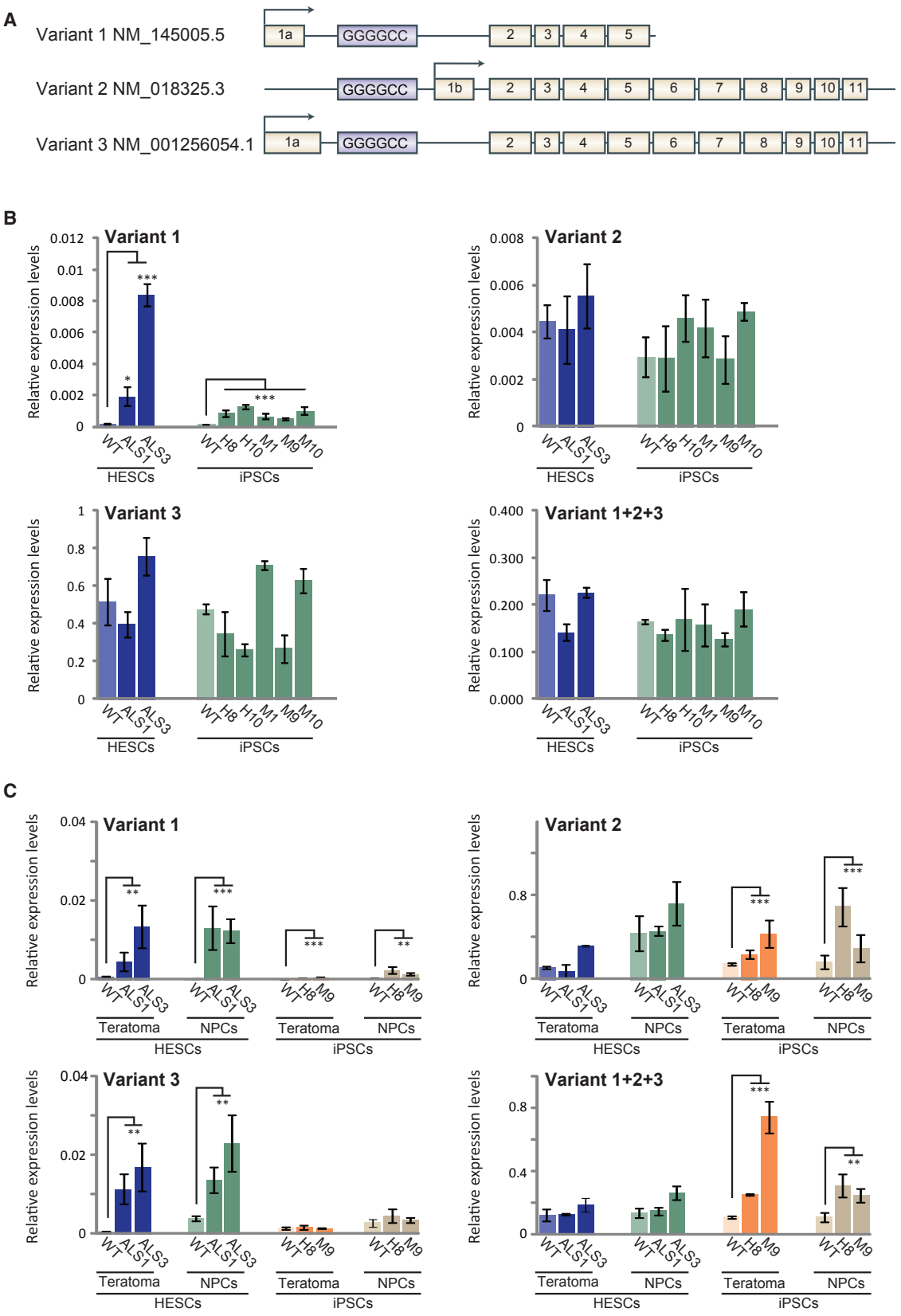
Figure 4. Analysis of Methylation Levels in Neural Precursors and Teratomas Derived from C9 and WT hESCs and iPSCs

(A) Methylation levels upstream to the repeats were determined by bisulfite sequencing in teratomas and neural precursor cells (NPC) derived from WT and C9 mutant hESCs (SZ-ALS1 and SZ-ALS3). Each line represents a single DNA molecule (upstream region of the repeats), with methylated and unmethylated CpGs designated by black and white circles, respectively.

(B) Results of the $(G_4C_2)_n$-methylation assay in WT and affected (C9) teratomas and NPCs derived from hESCs. Left panel (blue channel) represents methylated alleles while the right panel (green channel) represents unmethylated alleles.

(C) Methylation levels upstream of the repeats were determined by bisulfite sequencing in teratomas and NPCs derived from WT and C9 iPSCs from patient H (C9-iPSC H#8, 700 repeats) and patient M (C9-iPSC M#9, 2,700 repeats). Each line represents a single DNA molecule (upstream region of the repeats), with methylated and unmethylated CpGs designated by black and white circles, respectively.

(D) Results of the $(G_4C_2)_n$-methylation assay in WT and affected (C9/ALS) teratomas and NPCs derived from iPSCs. Left panel (blue channel) represents methylated alleles while right panel (green channel) represents unmethylated alleles.



(legend on next page)

unmethylated at the repeats (based on $(G_4C_2)_n$-methylation assay [Xi et al., 2015a, 2015b]) and upstream of the repeats (based on bisulfite colony sequencing). In addition, we generated iPSCs clones that are haploidentical to the mutant hESCs from skin fibroblasts of the asymptomatic C9-carrier mother (700 repeats, patient H), and from an unrelated 65-year-old ALS-manifesting patient (2,700 repeats, patient M). Unexpectedly, we found a striking difference in methylation levels at the 5′ UTR of *C9orf72* between the C9 iPSCs and all other cell types examined. Unlike in the C9 hESCs and parental fibroblasts, methylation was detected at the expanded repeats, and reached almost 100% at the upstream CGI in all iPSC clones. Although the difference in *C9orf72* hypermethylation between the C9 hESCs and iPSCs could, in theory, result from a difference in expansion size, this is very unlikely since the number of GGGGCC repeats in the hESCs is well above the threshold for methylation to be triggered in any cell type examined thus far (Xi et al., 2015b). In addition, methylation levels in the iPSCs are much higher than those observed in somatic cells of patients (where it does not exceed 20% [Liu et al., 2014]), and is consistent with the exclusive enrichment of the repressive histone mark H3K9me3 and not H3K27me3. This is somewhat different from the data obtained by Belzil et al. (2013), who demonstrated enrichments for H3K27me3 in addition to H3K9me3, although they analyzed different cell types from those described in this study. While we used undifferentiated cells, they looked at frontal cortex and cerebellum. Indeed, our findings in undifferentiated cells cast doubt on the importance of H3K27me3 modification for setting the methylation state and transcriptionally inactive chromatin configuration at the C9 expansion.

We argue that the C9 mutation acts as a hotspot for de novo methylation by transcription factor reprogramming considering the dramatic rise in DNA methyltransferase 3B (DNMT3B) expression levels during the reprogramming procedure (Huang et al., 2014). In fragile X syndrome it was also shown that reprogramming of fibroblasts with an unmethylated full CGG expansion results in hypermethylation and complete inactivation of the *FMR1* gene in mutant iPSCs (de Esch et al., 2014). Thus, perhaps hypermethylation of the GGGGCC repeats in our C9 iPSCs represents a much wider phenomenon, whereby untranslated repeat expansions that reside within CGIs provide a "sink" for DNMTs during transcription factor reprogramming (rather than first de novo methylate and then failure to demethylate). This may also explain why we and others generally fail to demethylate and reactivate the *FMR1* gene by somatic cell reprogramming when producing iPSCs from cells of fragile X-affected patients (Avitzour et al., 2014; Sheridan et al., 2011; Urbach et al., 2010). It should be noted that our results contradict the report of Esanov et al. (2016), who showed demethylation (rather than de novo methylation) of the C9 mutation in C9 iPSCs. The discrepancy could result from the different cell states employed. We used primary skin fibroblasts from two unrelated patients, whereas they used an immortalized cell line from blood cells of a single patient.

To explore the effect of differentiation on the methylation status of the C9 mutation, we induced the unmethylated mutant hESCs (SZ-ALS1 and SZ-ALS3) and C9 hypermethylated iPSC clones (from C9 individuals H and M) to differentiate into disease-relevant cell types (NPCs and teratomas), which are generally enriched in neural cells. Interestingly, methylation levels remained unchanged at the repeats and upstream of them in both hESCs and iPSCs, following in vitro and in vivo differentiation (0% and 50%, respectively). Perhaps extending differentiation length to fully matured neurons, or to specific subtypes of neurons, would have elicited C9 methylation in the hESCs.

Finally, to associate hypermethylation with disease pathogenesis, we analyzed the expression of transcript variants 1, 2, and 3 individually. We show that the GGGGCC expansion alters the region to allow the enhancement of V1 transcripts, albeit with much lower levels in iPSCs relative to hESCs. In addition, we demonstrate that the C9 mutation, together with neural differentiation, favors transcription from an upstream promoter (exon 1a-initiating transcripts, V1 and V3) over a downstream promoter (exon 1b-initiating transcript, V2), and that this effect is largely restricted in iPSCs. Furthermore, we found preferential retention of

Figure 5. Analysis of *C9orf72* Variant Expression Levels in Undifferentiated and Differentiated Derivatives of hESCs and iPSCs by qRT-PCR

(A) Schematic illustration of the three *C9orf72* mRNA isoform variants (V1, V2, and V3).

(B and C) Mean value of qRT-PCR for *C9orf72* transcription in both C9 hESC lines (SZ-ALS1 and SZ-ALS3), C9 iPSC clones derived from patients H (H8, H10) and M (M1, M9, M10), and their WT controls from the TaqMan gene expression assay for *C9orf72* transcript variants 1, 2, and 3. mRNA transcription levels were determined in (B) undifferentiated (hESCs and iPSCs) and (C) differentiated cell derivatives (NPCs and teratomas) of C9 mutation carrying cells and appropriate controls (WT hESCs and WT iPSCs). Using TaqMan probes targeting transcript variants 1, 2, and 3 individually (V1, V2, V3) as well as altogether (V1 + V2 + V3), we determined the relative abundance of each transcript variant. The expression level in each cell type represents an average of three to six independent experiments. Cycle threshold (Ct) values were normalized to the corresponding Ct value of *GUS*. WT hESC line is SZ-13. Error bars represent SE (t test for equal variances, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$).

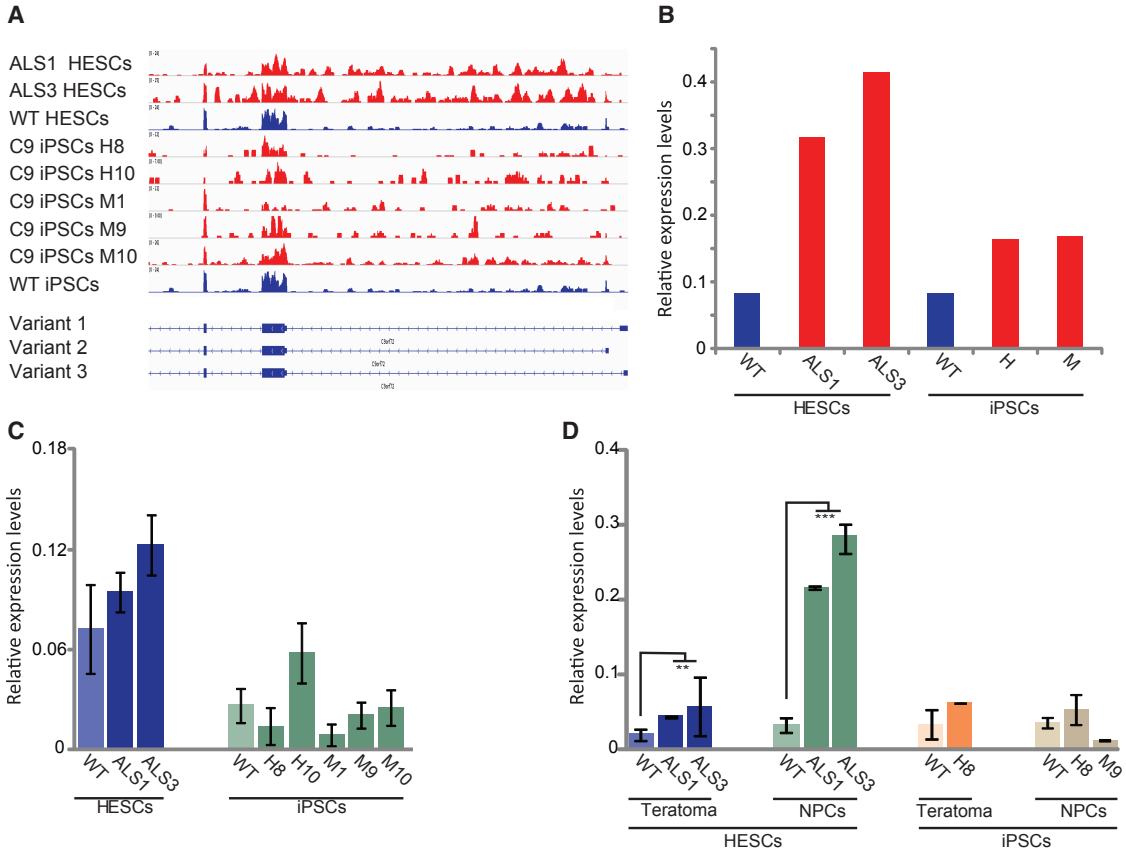


Figure 6. Expression of Intron 1-Retaining *C9orf72* Transcripts in Undifferentiated and Differentiated hESCs and iPSCs

(A) Genebrowser view of the first three exons (exons 1a, 1b, 2, and 3) and intronic sequences of *C9orf72*.

(B) Quantification of the intron 1 retention data presented in (A) in undifferentiated WT and C9 hESCs and iPSCs is based on RNA-seq data. C9 iPSC clones from patient H and M were averaged.

(C) Relative expression levels of intron 1-retaining exon 1a-initiating transcripts as determined by real-time RT-PCR in undifferentiated C9 hESCs (SZ-ALS1 and SZ-ALS3), C9 iPSC clones, and their appropriate WT controls.

(D) Relative expression levels of intron 1-retaining exon 1a-initiating transcripts as determined by real-time RT-PCR in teratomas and NPCs derived from C9 hESCs (SZ-ALS1 and SZ-ALS3), C9 iPSC clones, and their appropriate WT controls. The expression level in each cell type represents an average of three to six independent experiments.

In (C) and (D), Ct values were normalized to the corresponding Ct value of *GUS*. WT hESC line is SZ-13. Error bars represent SE (t test for equal variance, $^{**}p < 0.01$, $^{***}p < 0.001$).

intron 1 in both C9 hESCs and iPSCs by RNA deep sequencing, illustrating the propensity of the mutation to interfere with the proper splicing of this region in both exon 1a- and 1b-initiating transcripts. However, this change between WT and C9 cells demonstrates a greater effect in hESCs (4.5-fold increase) relative to iPSCs (only a 2-fold increase). More importantly, by monitoring for the expression of potentially pathogenic mRNA transcripts (i.e., exon 1a-initiating transcripts that retain intron 1), we find no difference between mutant and WT undifferentiated cells (hESCs and iPSCs). However, their level becomes significantly higher upon differentiation into disease-relevant cell types (NPCs and teratomas) exclusively in C9 hESCs, and not in C9 iPSCs. We propose that methylation counteracts the effect of the expansion by downregulating exon 1a-initiating mRNA species. In line with this idea are previous reports by Liu et al. (2014) and others (Bauer, 2016; Day and Roberson, 2015), which point to a mechanistic link between hypermethylation and reduced accumulation of RNA foci and dipeptide inclusions in patient cell lines, brain samples, and HEK293T transgenic cell lines.

Our data related to *C9orf72* mRNA levels conflict with the reports of others, who demonstrate a general reduction in *C9orf72* in diseased cells (Ciura et al., 2013; Cooper-Knock et al., 2013; Donnelly et al., 2013; Waite et al., 2014; Xi et al., 2013). These discrepancies may stem from the different types of samples (single versus mixed type of cells) employed. Nevertheless, even if our mutant NPCs do not

reflect the physiological levels observed in fully mature disease-relevant cells, this model system could help to interpret the role of methylation in disease pathogenesis because it features both an extreme C9 hypermethylation state and an extreme C9 hypomethylation state.

To summarize, this study clearly demonstrates how reprogramming excessively hypermethylates the C9 expanded locus, and how the C9 mutation alters *C9orf72* variant transcription and processing. In addition, the current work highlights the importance of neural differentiation in the pathogenesis of ALS-FTD and points to the potential role of C9 hypermethylation as a neuroprotective mechanism that attenuates the accumulation of potentially toxic repeat-containing mRNAs in C9 neurons.

Altogether, this implies that the difference between C9 hESCs and iPSCs may be crucial for investigating the neural phenotype of the C9/ALS-FTD disease, given that mutant hESCs are likely to present a more accurate and more severe phenotype than comparable iPSCs. Recognizing that C9 hypermethylation limits the increase in intron 1-retaining transcripts, it may be possible in the future (with the advent of gene editing) to modify the C9 expanded locus by specifically targeting chromatin silencing factors/DNMTs to the region. This would affect epigenetic inactivation of exon 1a-initiating variants so as to reverse/minimize the adverse effect of the C9 mutation in disease-affected tissues.

EXPERIMENTAL PROCEDURES

hESC Cell Line Derivation and Characterization

The use of embryos carrying the C9 mutation, derived from PGD treatment, for hESC derivation was performed in compliance with protocols approved by the National Ethics Committee. All cell lines were established at the Shaare Zedek Medical Center (87/07). Cell line derivation was carried out as previously described (Eiges et al., 2007). All hESC lines were examined for all typical characteristics of hESCs (for primers and conditions see Supplemental Experimental Procedures).

C9 iPSC Derivation

For transcription reprogramming, Yamanaka's four retroviral vectors expressing OCT3/4, SOX2, KLF4, and c-MYC were individually packaged in 293T cells. Infectious viruses were collected 24 and 48 hr post transfection and immediately added to primary fibroblasts (88/11). Four days following infection the cells were placed on mitomycin C-treated mouse embryonic fibroblasts (MEFs) and maintained in hESC medium. Manual isolation of single clones was carried out approximately 30 days post transfection, resulting in stable cell lines with hESC-like morphology.

Teratoma Induction

Cells (2.5–5 × 10^6) were harvested, diluted 1:1 in medium/Matrigel, and injected subcutaneously to both sides of the back of NOD-SCID IL2R$\gamma^{-/-}$ mice. Six to eight weeks later the mice were euthanized and tumors were isolated, sectioned, and assessed for differentiation by H&E staining.

Neuronal Precursor Differentiation of HESCs and iPSCs

The cells were grown on MEFs to similarly sized, defined colonies using hESC medium. Induced embryoid body (EB) formation was achieved by detaching colonies with collagenase IV for 30 min. The cells were centrifuged and transferred as colonies to a Petri dish containing hESC medium without basic fibroblast growth factor (bFGF) + 5 μM dorsomorphin (catalog #11967, Cayman) and 5 μM SB431542 (SB #13031, Cayman). The EBs were cultured for 4–5 days with medium replacement every other day. For the NPC expansion, the cells were plated on Matrigel-coated plates (without dissociation) and cultured with KO DMEM medium + 1XN2 supplement (stock X100) and 20 ng/mL bFGF for 8–10 days with medium replacement every other day.

Chromosome Analysis

Karyotype analysis was carried out by Giemsa staining according to standard protocol.

Southern Blot Analysis

Genomic DNAs (10–25 μg) were digested with EcoRI and HindIII (Fermentas) restriction endonucleases, separated on 0.8% agarose gels, blotted onto Hybond N+ membranes (Amersham), and hybridized with a PCR Dig-labeled 576-bp probe (primer F: TTG CGA TGA CTT TGC AGG GGA CC and primer R: CAG CGA GTA CTG TGA GAG).

RT-PCR

Total RNA was isolated from the cells by TRI Reagent extraction, then 1 μg of RNA was reverse transcribed by random hexamer priming and Multi Scribe reverse transcriptase (ABI). Amplification was performed using the primers listed in Supplemental Experimental Procedures using Super-Therm Taq DNA polymerase (Jain Biologicals).

Expression of Undifferentiated Cell-Specific Markers

Undifferentiated cell cultures were examined for the expression of undifferentiated cell-specific markers by immunostaining using monoclonal mouse OCT3/4 (Santa Cruz Biotechnology #sc-5279, 1:50 dilution) or TRA-1-60 (Santa Cruz #sc-21705, 1:50 dilution), together with Cy3-conjugated goat anti-mouse polyclonal antibodies (Jackson Immunostaining #115-035-062, 1:100 dilution). Nuclear staining was performed with Hoechst 33258 (Sigma #861405). Staining for alkaline phosphatase was carried out using an Alkaline Phosphatase Kit (Sigma Diagnostics #86R-1KT) according to the manufacturer's protocol.

Real-Time TaqMan PCR

TaqMan gene expression assays for *C9orf72* were carried out according to Belzil et al. (2013). qRT-PCR experiments were conducted in triplicate using custom-made TaqMan-based expression assays for transcript variants 1 (NM_145005.5), 2 (NM_018325.3), and 3 (NM_001256054.1) individually, and altogether (adopted

from DeJesus-Hernandez et al., 2011); for primers and probe sequences see Supplemental Experimental Procedures). *GUS* transcript (NM_000181.3) was used as housekeeping gene for normalization of ΔΔCt mean values.

Expression of Intron 1-Retaining Transcripts Upstream of the Repeats by qRT-PCR

Total RNA was isolated from cells by TRI reagent extraction. RNA (1 μg) was reverse transcribed (Multi Scribe RT, ABI) with random hexamer primers. Real-time PCR was performed using Power SYBR Green Master Mix (ABI) on an ABI 7900HT instrument. Primers are listed in Supplemental Experimental Procedures.

FACS Analysis for NCAM1-Positive Cells

NPCs were washed with PBS and dissociated to single cells using preheated (to 37°C) TrypLE Selected (Life Technologies #12563-011) for a minimal amount of time. After cell dissociation, the cells were resuspended with cold sterile PBS-FACS medium (Commercial PBS without Ca^{+2} Mg^{+2} to prevent adhesion with FCS 10%), and filtered through a mesh tube (#352235). From this point onward everything was carried out on ice or under cold centrifugation (4°C). After centrifugation 1 × 10^6 cells were resuspended into 200 μL of PBS-FCS. The cells were stained directly with primary anti-NCAM1 (R&D #af2408, 1:150) and incubated on ice for 1 hr. After staining the cells were washed two times by adding 5 mL of PBS-FCS and centrifuged at 150 × *g* for 5 min. The cells were then resuspended into the initial volume and incubated with a secondary antibody (1:200, fluorescein isothiocyanate-conjugated donkey anti-goat immunoglobulin G [IgG], #705095147) for an additional 1 hr. Control cells were stained for secondary antibody only. The cells were washed twice and assayed by FACS.

Immunostaining for Tuj1-Positive Cells in Teratoma Sections

Paraffin-embedded tissues were sectioned (4 μm) using a microtome (Leica RM2255) and transferred to silane-coated slides. After incubation in a 60°C in a dry oven for 1 hr, paraffin-embedded sections were deparaffinized in xylene and rehydrated through graded ethanol (100% to 50%). For antigen retrieval, sections were microwaved in 10 mM sodium citrate buffer (pH 6.0) for 5 min. Nonspecific antibody binding was blocked in PBS with 10% fetal bovine serum and 0.3% Triton X-100 (Sigma) for 45 min. Sections were then incubated with anti-Tuj1 antibody (rabbit IgG 1:400, Biolegend #845501) at 4°C overnight. The secondary antibody Alexa anti-rabbit (1:400, Life Technologies #A31572) was applied for 1 hr at room temperature. The sections were nuclear stained by DAPI, mounted with Vectashield (Vector Laboratories #H1000), and examined under a fluorescent microscope.

Bisulfite Sequencing

Genomic DNA (1 μg) was modified by bisulfite treatment (EZ DNA methylation Kit, Zymo Research) and amplified by FastStart DNA polymerase (Roche). Amplified products were cloned and single colonies were analyzed for CpG methylation by direct sequencing (ABI 3130). Primers are listed in Supplemental Experimental Procedures.

$(G_4C_2)_n$-Methylation Assay

Methylation of the G_4C_2-repeat itself was studied using a reported $(G_4C_2)_n$-methylation assay (Xi et al., 2015b). In brief, this assay combines rp-PCR with methylation-specific PCR, where each sample was amplified by rp-PCR using primers specific for methylated versus unmethylated DNA after bisulfite conversion. The primers were labeled by fluorescein amidite for methylated DNA amplification (blue channel) and HEX for unmethylated DNA amplification (green channel). Data were visualized by Genotyper software (version 3.6, Applied Biosystems).

Chromatin Immunoprecipitation Assay

ChIP was performed using the Upstate EZTM ChIP kit, according to the manufacturer's protocol with slight modifications. In brief, cells were harvested and then fixed, quenched, and washed in 50-mL tubes. Sonication was carried out using a Vibra Cell VCX130 with a 3-mm microtip and 30% amplitude, in five cycles of 10 s and 30 s rest on ice. Immunoprecipitation was performed using an anti-H3K27me3 (Abcam #6002) and anti-H3K9me3 (Abcam #8898) antibody. Immunoprecipitation efficiency was evaluated by *HOXA9* (enriched in H3K9me3 and H3K27me3 in pluripotent stem cells). Real-time PCR was carried out on an ABI 7900HT instrument (primers are listed in Supplemental Experimental Procedures). ΔΔCt values were normalized according to a negative control (*APRT*) to account for histone modification enrichment.

RNA Deep Sequencing

RNA was extracted using TRI Reagent (Sigma). Depletion of rRNA was performed using the Ribominus kit (Invitrogen). RNase-R treatments were performed by adding 3 U of RNase-R (Epicenter Biotechnologies) per milligram of RNA and 15-min incubation at 37°C. cDNA libraries were generated using the TruSeq RNA sample preparation kit and protocol (Illumina), and stranded, ligation-based libraries were sequenced as previously described (Engreitz et al., 2013). RNA-seq reads were aligned to the genome (hg19) using STAR. Coverage of *C9orf72* second exon (chr9:27566674-27567162) and first intron (chr9:27567163-27573426) were calculated using the samtools depth tool. We normalized the total coverage values to the length of the exon/intron to obtain average coverage per base.

ACCESSION NUMBERS

The accession number for the RNA-seq data reported in this paper is GEO: GSE87273.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and five figures and can be found with this article online at http://dx.doi.org/10.1016/j.stemcr.2016.09.011.

AUTHOR CONTRIBUTIONS

Y.C.-H., M.Z., E.R., and S.E.-L. contributed to the conception and design of the study, the collection and assembly of data, data analysis and interpretation, and manuscript writing. G.A., T.E-.G.,

E.L.-L., M.G., O.B., R.A.-F., and S.K. contributed to the collection of data. R.E. contributed to the conception and design of the study, financial support, data analysis and interpretation, and manuscript writing.

ACKNOWLEDGMENTS

We thank the family who donated the C9 embryos for hESC line derivation, and the C9 mutation carriers (patients H and M) who made donations of skin for the generation of C9 iPSCs. We would also like to thank Dr. David Zeevi for critical reading of the manuscript and Tamar Golan-Lev for the graphic illustrations. This research was partly supported by the ALS Association initiator grant (grant no. 16-IIP-256, R.E. and S.K.), the Legacy Heritage Biomedical program of the Israel Science Foundation (grant no. 0621/02), Mirsky foundation (S.E.-L.), the Canadian Consortium on Neurodegeneration in Aging and Weston Brain Institute (E.R., M.Z.).

Received: April 2, 2015
Revised: September 25, 2016
Accepted: September 26, 2016
Published: October 20, 2016

REFERENCES

Almeida, S., Gascon, E., Tran, H., Chou, H.J., Gendron, T.F., Degroot, S., Tapper, A.R., Sellier, C., Charlet-Berguerand, N., Karydas, A., et al. (2013). Modeling key pathological features of frontotemporal dementia with C9ORF72 repeat expansion in iPSC-derived human neurons. Acta Neuropathol. *126*, 385–399.

Avitzour, M., Mor-Shaked, H., Yanovsky-Dagan, S., Aharoni, S., Altarescu, G., Renbaum, P., Eldar-Geva, T., Schonberger, O., Levy-Lahad, E., Epsztejn-Litman, S., et al. (2014). FMR1 epigenetic silencing commonly occurs in undifferentiated fragile X-affected embryonic stem cells. Stem Cell Rep. *3*, 699–706.

Bauer, P.O. (2016). Methylation of C9orf72 expansion reduces RNA foci formation and dipeptide-repeat proteins expression in cells. Neurosci. Lett. *612*, 204–209.

Belzil, V.V., Bauer, P.O., Prudencio, M., Gendron, T.F., Stetler, C.T., Yan, I.K., Pregent, L., Daughrity, L., Baker, M.C., Rademakers, R., et al. (2013). Reduced C9orf72 gene expression in c9FTD/ALS is caused by histone trimethylation, an epigenetic event detectable in blood. Acta Neuropathol. *126*, 895–905.

Belzil, V.V., Bauer, P.O., Gendron, T.F., Murray, M.E., Dickson, D., and Petrucelli, L. (2014). Characterization of DNA hypermethylation in the cerebellum of c9FTD/ALS patients. Brain Res. *1584*, 15–21.

Ciura, S., Lattante, S., Le Ber, I., Latouche, M., Tostivint, H., Brice, A., and Kabashi, E. (2013). Loss of function of C9orf72 causes motor deficits in a zebrafish model of amyotrophic lateral sclerosis. Ann. Neurol. *74*, 180–187.

Cooper-Knock, J., Higginbottom, A., Connor-Robson, N., Bayatti, N., Bury, J.J., Kirby, J., Ninkina, N., Buchman, V.L., and Shaw, P.J. (2013). C9ORF72 transcription in a frontotemporal dementia case with two expanded alleles. Neurology *81*, 1719–1721.

Cooper-Knock, J., Walsh, M.J., Higginbottom, A., Robin Highley, J., Dickman, M.J., Edbauer, D., Ince, P.G., Wharton, S.B., Wilson, S.A., Kirby, J., et al. (2014). Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. Brain *137*, 2040–2051.

Cooper-Knock, J., Bury, J.J., Heath, P.R., Wyles, M., Higginbottom, A., Gelsthorpe, C., Highley, J.R., Hautbergue, G., Rattray, M., Kirby, J., et al. (2015). C9ORF72 GGGGCC Expanded repeats produce splicing dysregulation which correlates with disease severity in amyotrophic lateral sclerosis. PLoS One *10*, e0127376.

Cruts, M., Gijselinck, I., Van Langenhove, T., van der Zee, J., and Van Broeckhoven, C. (2013). Current insights into the C9orf72 repeat expansion diseases of the FTLD/ALS spectrum. Trends Neurosci. *36*, 450–459.

Day, J.J., and Roberson, E.D. (2015). DNA methylation slows effects of C9orf72 mutations: an epigenetic brake on genetic inheritance. Neurology *84*, 1616–1617.

de Esch, C.E., Ghazvini, M., Loos, F., Schelling-Kazaryan, N., Widagdo, W., Munshi, S.T., van der Wal, E., Douben, H., Gunhanlar, N., Kushner, S.A., et al. (2014). Epigenetic characterization of the FMR1 promoter in induced pluripotent stem cells from human fibroblasts carrying an unmethylated full mutation. Stem Cell Rep. *3*, 548–555.

DeJesus-Hernandez, M., Mackenzie, I.R., Boeve, B.F., Boxer, A.L., Baker, M., Rutherford, N.J., Nicholson, A.M., Finch, N.A., Flynn, H., Adamson, J., et al. (2011). Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron *72*, 245–256.

Devlin, A.C., Burr, K., Borooah, S., Foster, J.D., Cleary, E.M., Geti, I., Vallier, L., Shaw, C.E., Chandran, S., and Miles, G.B. (2015). Human iPSC-derived motoneurons harbouring TARDBP or C9ORF72 ALS mutations are dysfunctional despite maintaining viability. Nat. Commun. *6*, 5999.

Dols-Icardo, O., Garcia-Redondo, A., Rojas-Garcia, R., Sanchez-Valle, R., Noguera, A., Gomez-Tortosa, E., Pastor, P., Hernandez, I., Esteban-Perez, J., Suarez-Calvet, M., et al. (2014). Characterization of the repeat expansion size in C9orf72 in amyotrophic lateral sclerosis and frontotemporal dementia. Hum. Mol. Genet. *23*, 749–754.

Donnelly, C.J., Zhang, P.W., Pham, J.T., Haeusler, A.R., Mistry, N.A., Vidensky, S., Daley, E.L., Poth, E.M., Hoover, B., Fines, D.M., et al. (2013). RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. Neuron *80*, 415–428.

Eiges, R., Urbach, A., Malcov, M., Frumkin, T., Schwartz, T., Amit, A., Yaron, Y., Eden, A., Yanuka, O., Benvenisty, N., et al. (2007). Developmental study of fragile X syndrome using human embryonic stem cells derived from preimplantation genetically diagnosed embryos. Cell Stem Cell *1*, 568–577.

Engreitz, J.M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., Kadri, S., Xing, J., Goren, A., Lander, E.S., et al. (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science *341*, 1237973.

Esanov, R., Belle, K.C., van Blitterswijk, M., Belzil, V.V., Rademakers, R., Dickson, D.W., Petrucelli, L., Boylan, K.B., Dykxhoorn,

D.M., Wuu, J., et al. (2016). C9orf72 promoter hypermethylation is reduced while hydroxymethylation is acquired during reprogramming of ALS patient cells. Exp. Neurol. *277*, 171–177.

Gendron, T.F., Belzil, V.V., Zhang, Y.J., and Petrucelli, L. (2014). Mechanisms of toxicity in C9FTLD/ALS. Acta Neuropathol. *127*, 359–376.

Gijselinck, I., Van Mossevelde, S., van der Zee, J., Sieben, A., Engelborghs, S., De Bleecker, J., Ivanoiu, A., Deryck, O., Edbauer, D., Zhang, M., et al. (2015). The C9orf72 repeat size correlates with onset age of disease, DNA methylation and transcriptional downregulation of the promoter. Mol. Psychiatry *21*, 1112–1124.

Haeusler, A.R., Donnelly, C.J., Periz, G., Simko, E.A., Shaw, P.G., Kim, M.S., Maragakis, N.J., Troncoso, J.C., Pandey, A., Sattler, R., et al. (2014). C9orf72 nucleotide repeat structures initiate molecular cascades of disease. Nature *507*, 195–200.

Huang, K., Shen, Y., Xue, Z., Bibikova, M., April, C., Liu, Z., Cheng, L., Nagy, A., Pellegrini, M., Fan, J.B., et al. (2014). A panel of CpG methylation sites distinguishes human embryonic stem cells and induced pluripotent stem cells. Stem Cell Rep. *2*, 36–43.

Kim, D.S., Lee, J.S., Leem, J.W., Huh, Y.J., Kim, J.Y., Kim, H.S., Park, I.H., Daley, G.Q., Hwang, D.Y., and Kim, D.W. (2010). Robust enhancement of neural differentiation from human ES and iPS cells regardless of their innate difference in differentiation propensity. Stem Cell Rev. *6*, 270–281.

Lee, Y.B., Chen, H.J., Peres, J.N., Gomez-Deza, J., Attig, J., Stalekar, M., Troakes, C., Nishimura, A.L., Scotter, E.L., Vance, C., et al. (2013). Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic. Cell Rep. *5*, 1178–1186.

Li, Y., Balasubramanian, U., Cohen, D., Zhang, P.W., Mosmiller, E., Sattler, R., Maragakis, N.J., and Rothstein, J.D. (2015). A comprehensive library of familial human amyotrophic lateral sclerosis induced pluripotent stem cells. PLoS One *10*, e0118266.

Liu, E.Y., Russ, J., Wu, K., Neal, D., Suh, E., McNally, A.G., Irwin, D.J., Van Deerlin, V.M., and Lee, E.B. (2014). C9orf72 hypermethylation protects against repeat expansion-associated pathology in ALS/FTD. Acta Neuropathol. *128*, 525–541.

Niblock, M., Smith, B.N., Lee, Y.B., Sardone, V., Topp, S., Troakes, C., Al-Sarraj, S., Leblond, C.S., Dion, P.A., Rouleau, G.A., et al. (2016). Retention of hexanucleotide repeat-containing intron in C9orf72 mRNA: implications for the pathogenesis of ALS/FTD. Acta Neuropathol. Commun. *4*, 18.

Peters, O.M., Cabrera, G.T., Tran, H., Gendron, T.F., McKeon, J.E., Metterville, J., Weiss, A., Wightman, N., Salameh, J., Kim, J., et al. (2015). Human C9ORF72 hexanucleotide expansion reproduces RNA foci and dipeptide repeat proteins but not neurodegeneration in BAC transgenic mice. Neuron *88*, 902–909.

Rossi, S., Serrano, A., Gerbino, V., Giorgi, A., Di Francesco, L., Nencini, M., Bozzo, F., Schinina, M.E., Bagni, C., Cestra, G., et al. (2015). Nuclear accumulation of mRNAs underlies G4C2-repeat-induced translational repression in a cellular model of C9orf72 ALS. J. Cell Sci. *128*, 1787–1799.

Russ, J., Liu, E.Y., Wu, K., Neal, D., Suh, E., Irwin, D.J., McMillan, C.T., Harms, M.B., Cairns, N.J., Wood, E.M., et al. (2015). Hypermethylation of repeat expanded C9orf72 is a clinical and molecular disease modifier. Acta Neuropathol. *129*, 39–52.

Sareen, D., O'Rourke, J.G., Meera, P., Muhammad, A.K., Grant, S., Simpkinson, M., Bell, S., Carmona, S., Ornelas, L., Sahabian, A., et al. (2013). Targeting RNA foci in iPSC-derived motor neurons from ALS patients with a C9ORF72 repeat expansion. Sci. Transl. Med. *5*, 208ra149.

Satoh, J., Yamamoto, Y., Kitano, S., Takitani, M., Asahina, N., and Kino, Y. (2014). Molecular network analysis suggests a logical hypothesis for the pathological role of c9orf72 in amyotrophic lateral sclerosis/frontotemporal dementia. J. Cent. Nerv Syst. Dis. *6*, 69–78.

Sheridan, S.D., Theriault, K.M., Reis, S.A., Zhou, F., Madison, J.M., Daheron, L., Loring, J.F., and Haggarty, S.J. (2011). Epigenetic characterization of the FMR1 gene and aberrant neurodevelopment in human induced pluripotent stem cell models of fragile X syndrome. PLoS One *6*, e26203.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861–872.

Urbach, A., Bar-Nur, O., Daley, G.Q., and Benvenisty, N. (2010). Differential modeling of fragile X syndrome by human embryonic stem cells and induced pluripotent stem cells. Cell Stem Cell *6*, 407–411.

Wainger, B.J., Kiskinis, E., Mellin, C., Wiskow, O., Han, S.S., Sandoe, J., Perez, N.P., Williams, L.A., Lee, S., Boulting, G., et al. (2014). Intrinsic membrane hyperexcitability of amyotrophic lateral sclerosis patient-derived motor neurons. Cell Rep. *7*, 1–11.

Waite, A.J., Baumer, D., East, S., Neal, J., Morris, H.R., Ansorge, O., and Blake, D.J. (2014). Reduced C9orf72 protein levels in frontal cortex of amyotrophic lateral sclerosis and frontotemporal degeneration brain with the C9ORF72 hexanucleotide repeat expansion. Neurobiol. Aging *35*, 1779.e5–1779.e13.

Xi, Z., Zinman, L., Moreno, D., Schymick, J., Liang, Y., Sato, C., Zheng, Y., Ghani, M., Dib, S., Keith, J., et al. (2013). Hypermethylation of the CpG island near the G4C2 repeat in ALS with a C9orf72 expansion. Am. J. Hum. Genet. *92*, 981–989.

Xi, Z., Rainero, I., Rubino, E., Pinessi, L., Bruni, A.C., Maletta, R.G., Nacmias, B., Sorbi, S., Galimberti, D., Surace, E.I., et al. (2014). Hypermethylation of the CpG-island near the C9orf72 G(4)C(2)-repeat expansion in FTLD patients. Hum. Mol. Genet. *23*, 5630–5637.

Xi, Z., van Blitterswijk, M., Zhang, M., McGoldrick, P., McLean, J.R., Yunusova, Y., Knock, E., Moreno, D., Sato, C., McKeever, P.M., et al. (2015a). Jump from pre-mutation to pathologic expansion in C9orf72. Am. J. Hum. Genet. *96*, 962–970.

Xi, Z., Zhang, M., Bruni, A.C., Maletta, R.G., Colao, R., Fratta, P., Polke, J.M., Sweeney, M.G., Mudanohwo, E., Nacmias, B., et al. (2015b). The C9orf72 repeat expansion itself is methylated in ALS and FTLD patients. Acta Neuropathol. *129*, 715–727.

Zeier, Z., Esanov, R., Belle, K.C., Volmar, C.H., Johnstone, A.L., Halley, P., DeRosa, B.A., Khoury, N., van Blitterswijk, M., Rademakers, R., et al. (2015). Bromodomain inhibitors regulate the C9ORF72 locus in ALS. Exp. Neurol. *271*, 241–250.

Stem Cell Reports, Volume *7*

Supplemental Information

Marked Differences in *C9orf72* Methylation Status and Isoform Expression between C9/ALS Human Embryonic and Induced Pluripotent Stem Cells

Yaara Cohen-Hadad, Gheona Altarescu, Talia Eldar-Geva, Ephrat Levi-Lahad, Ming Zhang, Ekaterina Rogaeva, Marc Gotkine, Osnat Bartok, Reut Ashwal-Fluss, Sebastian Kadener, Silvina Epsztejn-Litman, and Rachel Eiges

SUPPLEMETAL FIGURES AND LEGENDS

Fig S1. Characterization of ALS HESC lines, Related to section "*Derivation and characterization of C9/HESC lines*". (A) Staining for OCT4, Tra 1-60 and alkaline phosphatase activity. Scale bars stand for 200μm. (B) Expression of *OCT4, NANOG, SOX2* and *REX1*, by RT-PCR. (C) Karyotype analysis of ALS HESC lines by Giemsa staining. (D) Teratoma sections stained by H&E derived from SZ-ALS1 and SZ-ALS3. Scale bars stand for 130 μm. (E) Southern blot analysis identified a ~270 repeat expansion in both C9-HESC lines.

Fig S2. Characterization of C9-iPSC clones, Related to sections "*Analysis of C9orf72 methylation in C9 HESCs and their haplo-identical iPSCs" and "Methylation Analysis in C9 iPSCs derived from an unrelated symptomatic ALS patient*". (A) Staining for OCT4, Tra 1-60 and alkaline phosphatase activity. Scale bars stand for 200μm. (B) Expression of *OCT4, NANOG, SOX2* and *REX1*, by RT-PCR. (C) Karyotype analysis of C9 iPSC clones by Giemsa staining. (D) Southern blot analysis identified a ~700 and ~2,700 repeat expansions in C9 iPSCs derived from patients H and M, respectively.

Fig S3. Methylation analysis at the promoter of *SIGLEC6* in primary fibroblasts, HESCs and iPSCs, Relates to Fig 1 and 3. Methylation levels in C9 primary fibroblasts (C9-fibroblasts H and M), iPSC clones derived from them (C9-iPSC H#8, H#10, M#1, M#9, M#10), and C9 HESCs (SZ-ALS1 and SZ-ALS3) by bisulfite DNA colony sequencing demonstrates *de novo* methylation of *SIGLEC6* exclusively in iPSCs.

Fig S4. Differentiation of HESCs and iPSCs into neural precursors (NPCs) and teratomas, Related to section "*The effect of differentiation on the methylation status of C9orf72*". (A) Schematic illustration of differentiation protocol into NPCs using 2 inhibitors. (B) FACs analysis from NCAM1-positive cells in NPCs from wild type (WT) and C9 HESCs (SZ-ALS1 and SZ-ALS3), WT and C9 iPSC clones derived from patient H (C9-iPS#H8) and patient M (C9-iPS#M9). For each cell sample, unstained (left panel) and stained cells (right panel) are presented. (C) RT-PCR analysis for the expression of NPC-specific markers *SOX2, PAX6 and Nestin*; the undifferentiated cell-specific marker *OCT4*; and a housekeeping gene *GAPDH* in undifferentiated and NPCs of WT, C9 HESCs (SZ-ALS1 and SZ-ALS3), and mutant iPSCs (C9-iPS#H8 and C9-iPS#M9). (D) Enrichment for mature neurons in teratomas from C9 HESCs and iPSCs. As determined by immunostaining for Tuj1-positive cells (red) and DAPI (blue) staining in teratoma sections from C9 HESCs (SZ-ALS1 and SZ-ALS3) and iPSCs (C9-iPS#H8 and C9-iPS#M9).

Fig S5. Expression of intron 1 retaining *C9orf72* transcripts in undifferentiated and differentiated HESCs and iPSCs, Related to Fig 6. (A) Description of the samples utilized to generate RNA-seq libraries. Samples 1-2 were derived from C9 HESCs (SZ-ALS1 and SZ-ALS3), sample 3 from wild type HESCS, samples 4-8 from C9 iPSCs clones (C9-iPS#H8, -iPS#H10, and -iPS#M1, -iPS#M9 and -iPS#H10), and sample 9 from wild type iPSCs. RNA extracted from these samples was DNAse1 treated and we generated rRNA-depleted libraries utilizing a standard protocol. (B) Average coverage across intron 1 relative to exon 2 in all undifferentiated cells samples; wild type and affected HESCs and iPSCs. The data presented in fig 6B is intron/exon ratio. (C) Validation of intron 1 retaining transcripts in undifferentiated C9 mutant HESCs (SZ-ALS1 and SZ-ALS3) and iPSCs (C9-iPS#H8 and iPS#M9) and their differentiated cell counterparts; NPCs (NPC C9-iPS M#9 and H#8) and teratomas (teratoma C9-iPS M#9 and H#8) by RT-PCR. (D) Sanger sequencing of PCR products validated the existence of intron 1 retaining transcripts using primers that span over intron 1-exon 5.

Fig S1

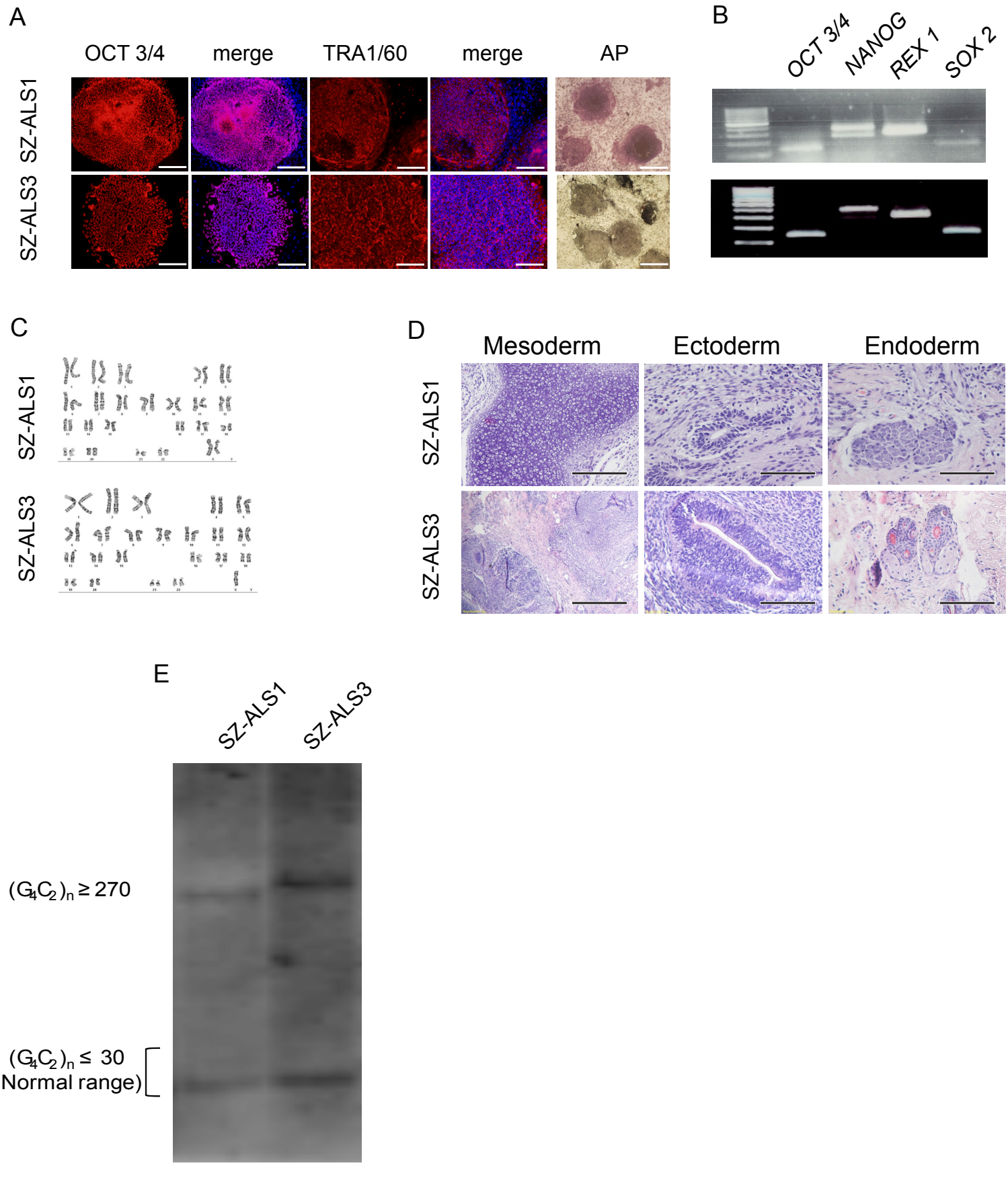
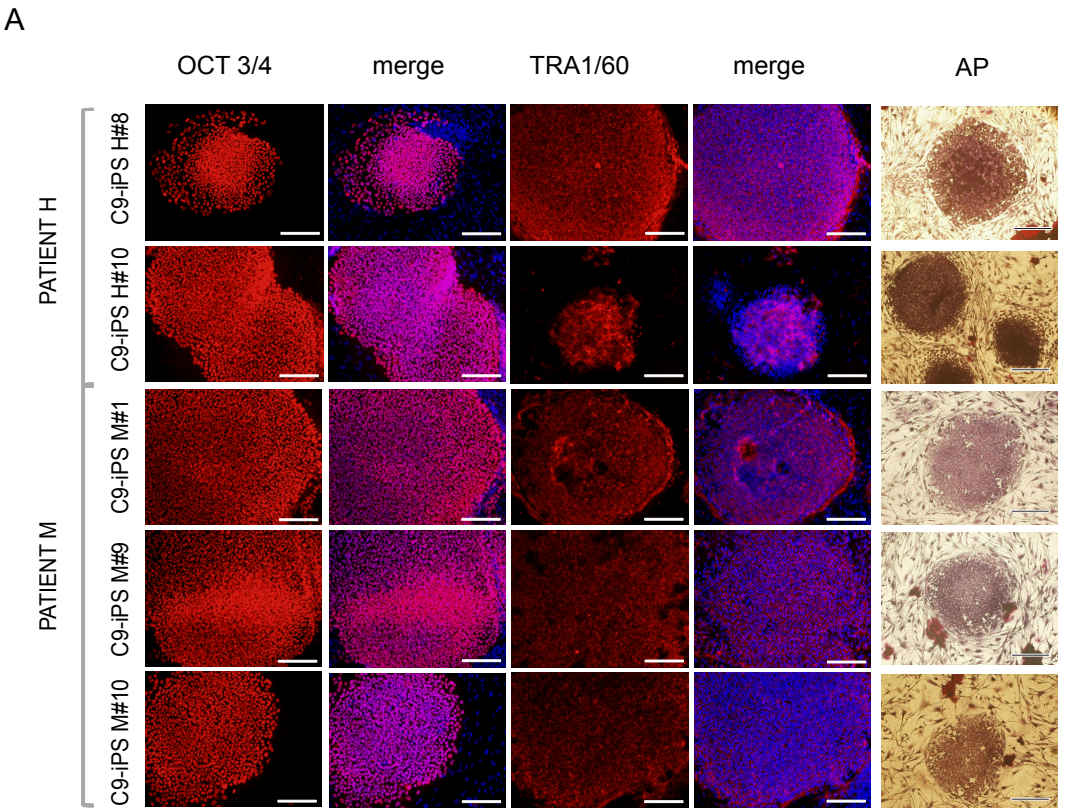
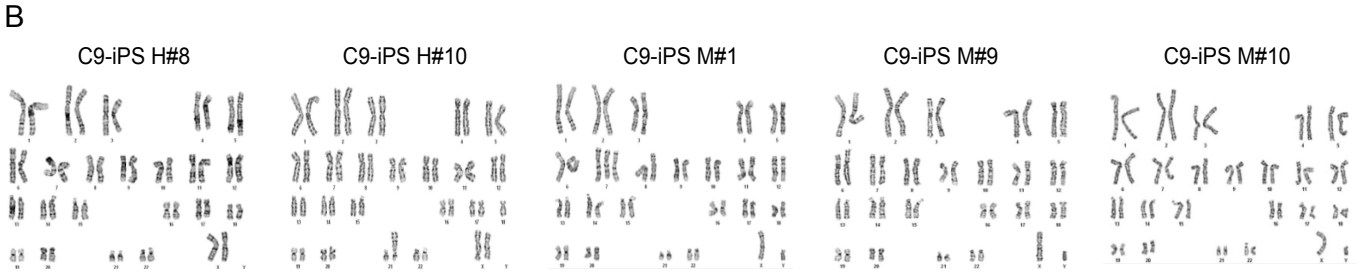
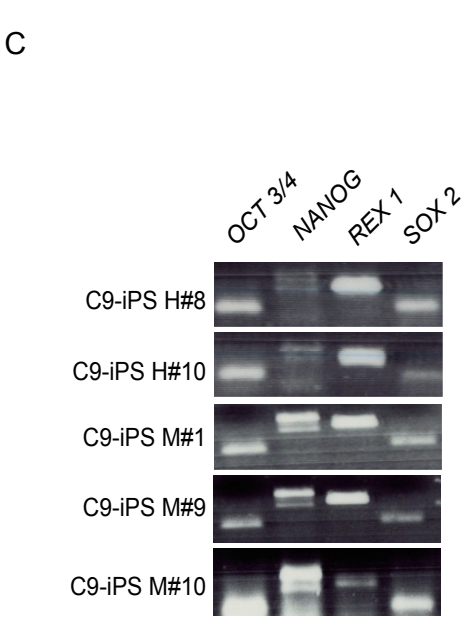

A
OCT 3/4
merge
TRA1/60
merge
AP
SZ-ALS1
SZ-ALS3
B
OCT 3/4
NANOG
REX 1
SOX 2
C
SZ-ALS1
SZ-ALS3
D
Mesoderm
Ectoderm
Endoderm
SZ-ALS1
SZ-ALS3
E
SZ-ALS1
SZ-ALS3
(G4C2)n ≥ 270
(G4C2)n ≤ 30
(Normal range)

Fig S2


A
OCT 3/4
merge
TRA1/60
merge
AP
PATIENT H
C9-iPS H#8
C9-iPS H#10
PATIENT M
C9-iPS M#1
C9-iPS M#9
C9-iPS M#10


B
C9-iPS H#8
C9-iPS H#10
C9-iPS M#1
C9-iPS M#9
C9-iPS M#10


C
OCT 3/4
NANOG
REX 1
SOX 2
C9-iPS H#8
C9-iPS H#10
C9-iPS M#1
C9-iPS M#9
C9-iPS M#10

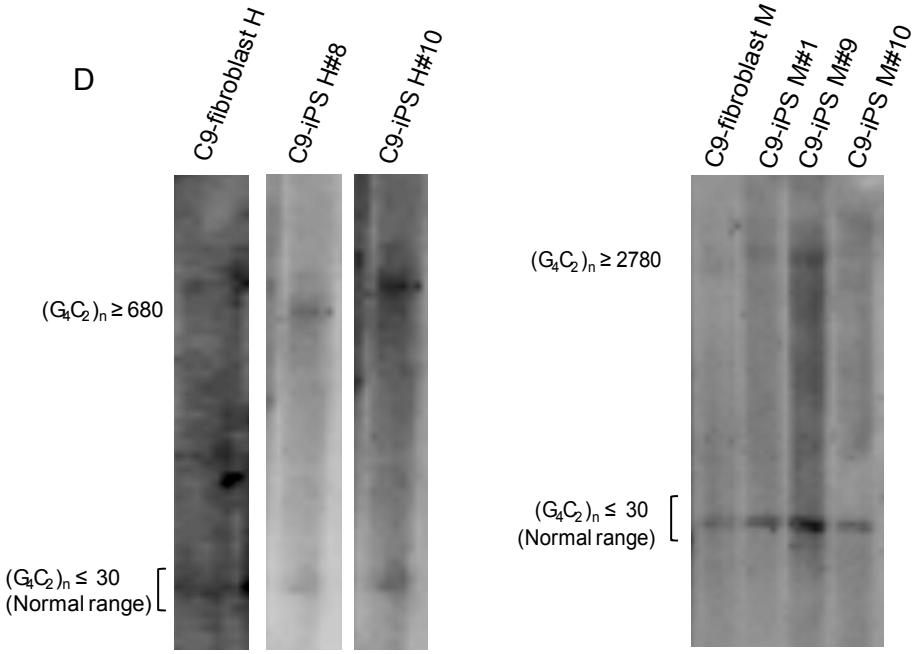

D
C9-fibroblast H
C9-iPS H#8
C9-iPS H#10
C9-fibroblast M
C9-iPS M#1
C9-iPS M#9
C9-iPS M#10
(G4C2)n ≥ 680
(G4C2)n ≤ 30
(Normal range)
(G4C2)n ≥ 2780
(G4C2)n ≤ 30
(Normal range)

Fig S3

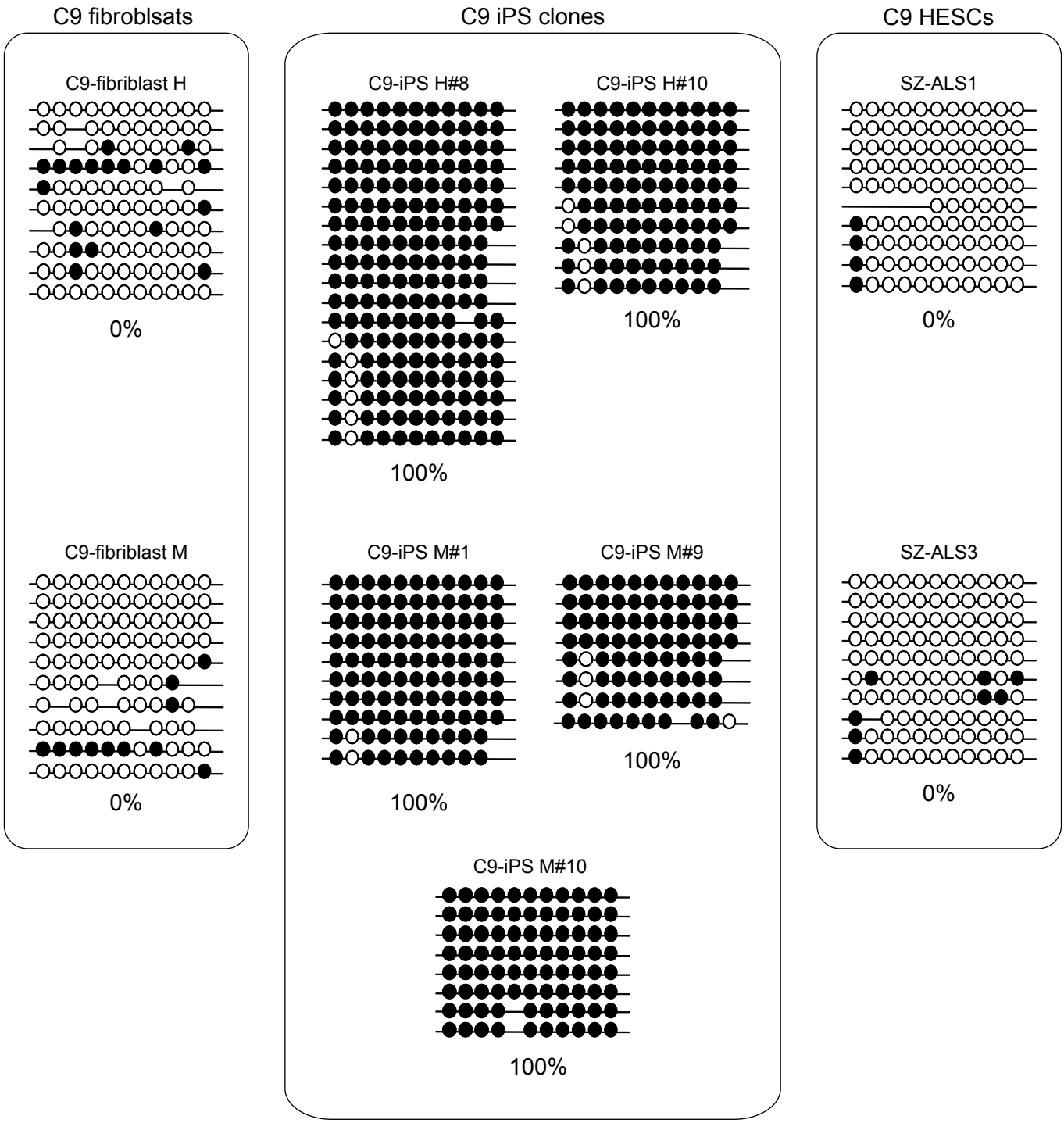
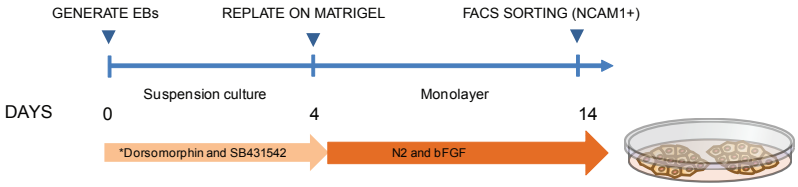
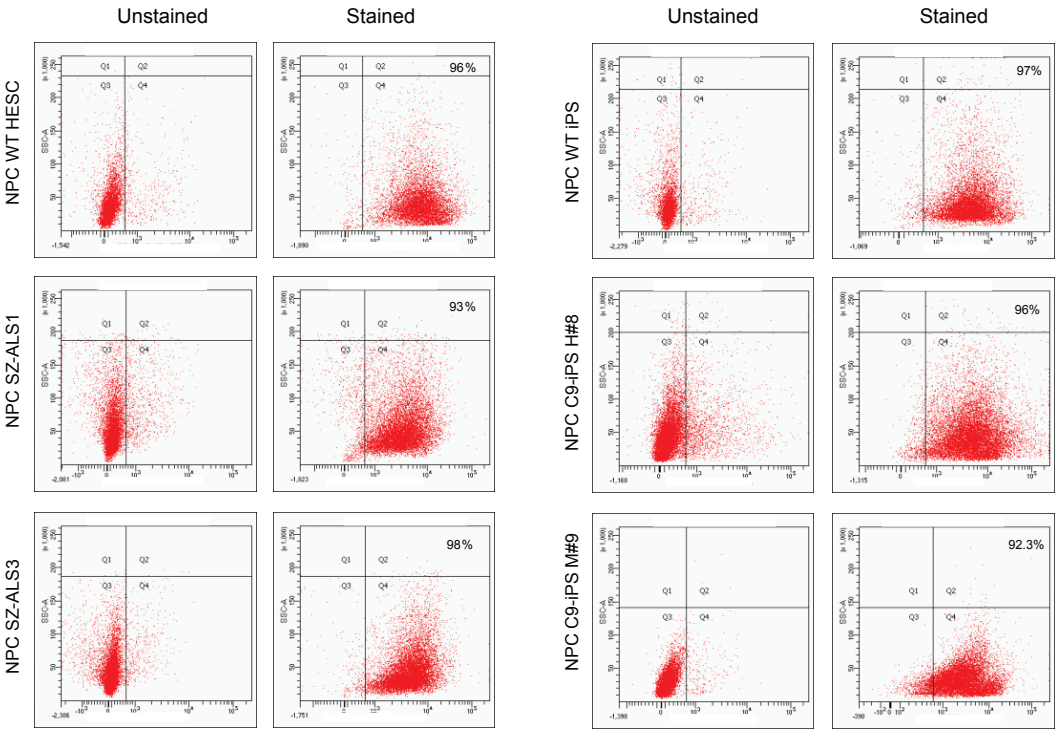

C9 fibroblsats
C9 iPS clones
C9 HESCs
C9-fibriblast H
0%
C9-iPS H#8
100%
C9-iPS H#10
100%
SZ-ALS1
0%
C9-fibriblast M
0%
C9-iPS M#1
100%
C9-iPS M#9
100%
SZ-ALS3
0%
C9-iPS M#10
100%

Fig S4

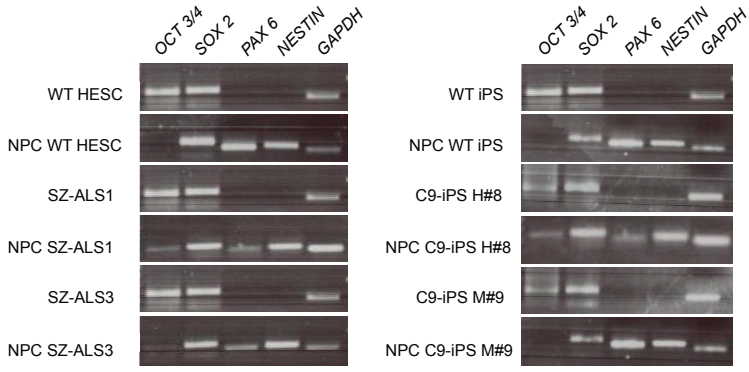
A


GENERATE EBs
REPLATE ON MATRIGEL
FACS SORTING (NCAM1+)
Suspension culture
Monolayer
DAYS
0
4
14
*Dorsomorphin and SB431542
N2 and bFGF

B


Unstained
Stained
Unstained
Stained
NPC WT HESC
96%
NPC WT iPS
97%
NPC SZ-ALS1
93%
NPC C9-iPS H#8
96%
NPC SZ-ALS3
98%
NPC C9-iPS M#9
92.3%

C


OCT 3/4
SOX 2
PAX 6
NESTIN
GAPDH
WT HESC
NPC WT HESC
SZ-ALS1
NPC SZ-ALS1
SZ-ALS3
NPC SZ-ALS3
WT iPS
NPC WT iPS
C9-iPS H#8
NPC C9-iPS H#8
C9-iPS M#9
NPC C9-iPS M#9

D

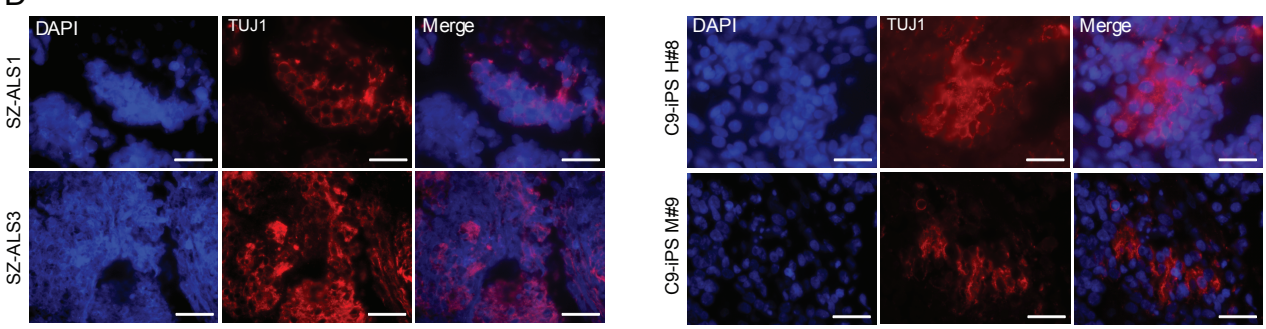

DAPI
TUJ1
Merge
SZ-ALS1
SZ-ALS3
C9-iPS H#8
C9-iPS M#9

Fig S5

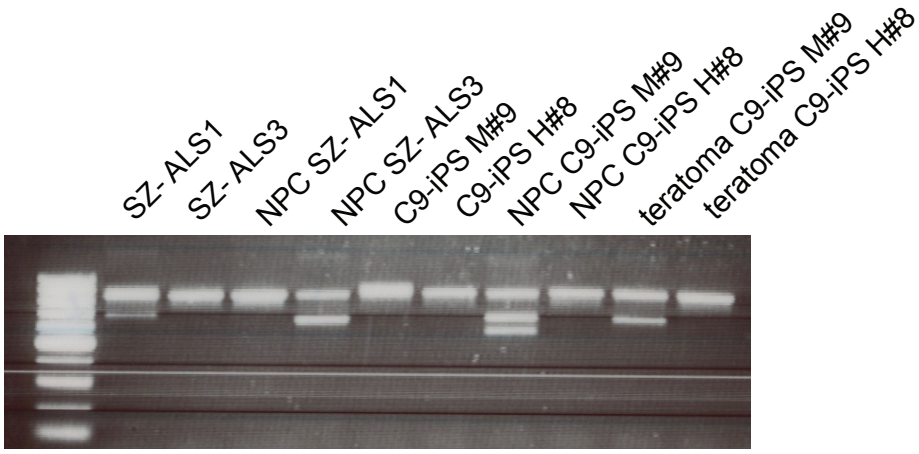
A

CELL TYPE	NUMBER OF NON-rRNA READS
SZ-ALS 1 HESCs	27,855,214
SZ-ALS 3 HESCs	26,362,223
WT HESCs	60,412,437
C9-iPSCs M#1	21,860,174
C9- iPSCs M#9	26,015,064
C9-iPSCs M#10	44,822,844
C9-iPSCs H#8	28,726,995
C9-iPSCs H#10	28,038,769
WT iPSCs	60,412,437

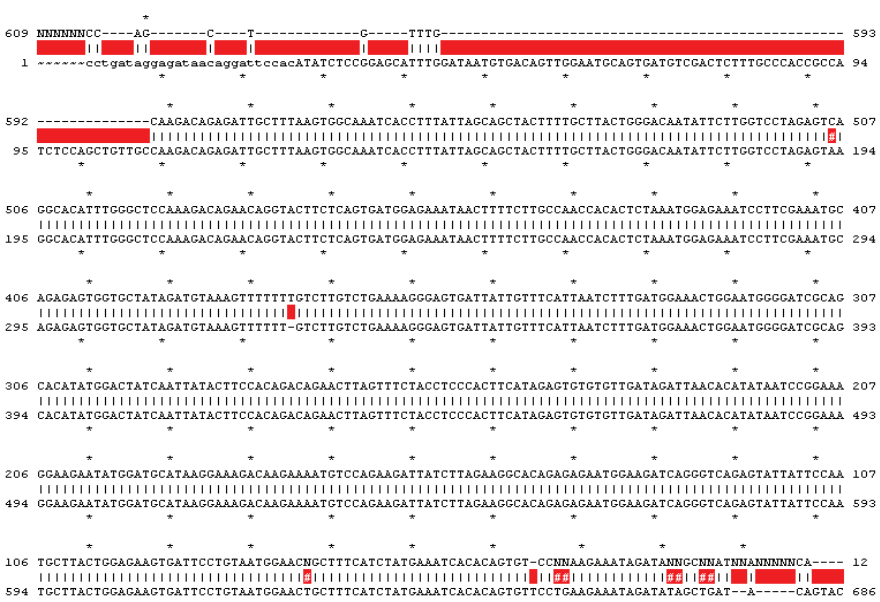
B

	SZ-ALS1 HESCs	SZ-ALS3 HESCs	WT HESCs	C9-iPSCs -M	C9 iPSCs -H	WT iPSCs
INTRON 1	3.928	3.470	1.284	1.697	0.685	1.404
EXON 2	12.367	8.385	15.285	8.926	3.944	16.797
INTRON 1/EXON 2	0.317584964	0.413791435	0.083997598	0.169390848	0.163897482	0.083583307

C



D



SUPPLEMETAL EXPERIMENTAL PROCEDURES

	5' Primer (sequence 5'-3')	3' Primer (sequence 5'-3')	Annealing Temp ^{0}C	Product Size (bp)
Primer Sets for RT-PCR for endogenous pluripotent genes				
OCT4	GACAGGGGGAGGGGAGGAGCTAGG	CTTCCCTCCAACCAGTTGCCCCAAAC	60	144
NANOG	CAGCCCCGATTCTTCCACCAGTCCC	CGGAGATTCCCAGTCGGGTTCACC	55	342, 390
REX1	CAGATCCTAAACAGCTCGCAGAAT	GCGTACGCAAATTAAAGTCCAGA	60	306
SOX2	GGGAAATGGGAGGGGTGCAAAAGAGG	TTGCGTGAGTGTGGATGGGATTGGTG	55	151
GAPDH	CCACTCCTCCACCTTTGAC	ACCCTGTTGCTGTAGCCA	62	102
Primer Sets for early neuronal cell markers				
PAX6	GCGGAAGCTGCAAAGAAATA	TTTGGCTGCTAGTCTTTCTCG	58	118
Nestin	TGCGGGCTACTGAAAAGTTC	AGGCTGAGGGACATCTTGAG	60	130
Primer Sets for ChIP Analysis				
HOXA9	CTCAGGAGCCTCGTGTCTTT	GTGACCAGGTGGAGGTGTGT	60	82
APRT	GCCTTGACTCGCACTTTTGT	TAGGCGCCATCGATTTTAAG	60	85
C9ORF72	AGGAAAGAGAGGTGCGTCAA	CAGGTGTGGGTTTAGGAGGT	60	138
Primer Sets for Southern blot Analysis				
Probe	TTGCGATGACTTTGCAGGGGACC	CAGCGAGTACTGTGAGAG	60	576
Primer Sets for Bisulfite Analysis				
BSP 1	TTTATTAGGGTTTGTAGTGGAGTTTT	AAATCTTTTCTTATTCACCCTCAAC	58	554
BSP 2	TATTAGGGTTTGTAGTGGAGTTTT	CCACACCTACTCTTACTAAACCC	58	504
SIGELC6	TTGTGTAGAGGGAGTGGAGTT	TCCTAAACCAAAACCCCTATAA	60	284
Primer Sets for methylation assay				
Methyl-specific primers	FAM-Methyl-F	FAM-TGTAAAACGACGGCCAGTAGTTTTGGAATTTAGGAGTCGC		
	Methyl-R1	CAGGAAACAGCTATGACCGAACCCGCCCCGACCACGCCCCGACCCCGACCCCG		
	Methyl-R2c	CAGGAAACAGCTATGACCGAACCCGCCCCGACCACGCCCCGACCCCG		
Unmethyl-specific primers	HEX-Unmethyl-F	HEX-TGTAAAACGACGGCCAGTAGTAAGTTTTGGAATTTAGGAGTTGTG		
	Unmethyl-R1	CAGGAAACAGCTATGACCAAACCCACCCCAACCACACCCCAACCCCAACCCCAA		
	Unmethyl-R2c	CAGGAAACAGCTATGACCAAACCCACCCCAACCACACCCCAACCCCAA		
Anchor primer		CAGGAAACAGCTATGACC		
Taqman primers and probes				
Variant 2	primer F	CGGTGGCGAGTGGATATCTC		
	primer R	TGGGCAAAGAGTCGACATCA		
	Probe	TAATGTGACAGTTGGAATGC		
Variant 1 **(NM_145005.5)**	Hs00331877 (Applied Biosystems)			
Variant 3 (**NM_001256054.1)**	Hs00948764 (Applied Biosystems)			
Variants (1+2+3)	Hs00376619 (Applied Biosystems)			
GUS **(NM_000181.3)**	Hs99999908 (Applied Biosystems)			
Real Time Primer Sets for Intron 1 retaining product				
Exon 1a-Intron 1 Real time	GGTGCGTCAAACAGCGACAAGTTC	GGAAACAACCGCAGCCTGTAGC		
GUS	CTCATTTGGAATTTTGCCGATT	CCGAGTGAAGATCCCCTTTTTA		
Intron 1- Exon 5	CCTGATAGGAGATAACAGGATTCCAC	GGTGACAGCTGTCATGAAGGC		