UNIVERSITÀ DEGLI STUDI
DI MILANO

DOTTORATO DI MEDICINA MOLECOLARE E TRASLAZIONALE

CICLO XXXIII

MD-PhD

Anno Accademico 2018/2019

TESI DI DOTTORATO DI RICERCA

MED/26 Neurologia

EPIGENETIC MARKS AND PATHOLOGICAL FEATURES ASSOCIATED TO MUTANT *C9ORF72* GENE IN AMYOTROPHIC LATERAL SCLEROSIS: AN *IN VITRO* STUDY IN PATIENT-DERIVED INDUCED PLURIPOTENT STEM CELLS AND MOTOR NEURONS

Dottoranda: Clara VOLPE

Matricola N° R11834

TUTORE: Prof.ssa Antonia RATTI
COORDINATORE DEL DOTTORATO: prof. Michele SAMAJA

Abstract

The expansion of the hexanucleotide repeat sequence GGGGCC (>30 repeats) in the first intron of *C9ORF72* gene is the main genetic cause of two neurodegenerative diseases: amyotrophic lateral sclerosis (ALS) and frontotemporal lobar degeneration (FTLD). The 5' promoter of *C9ORF72* gene has been found hypermethylated in 30% of *C9ORF72* positive ($C9^+$) ALS/FTLD patients and never in unexpanded patients and healthy controls. Promoter methylation seems to have a neuroprotective role from RNA toxicity and RAN translated dipeptide repeats (DPRs). The aim of this study has been to characterize $C9^+$ ALS patient-derived induced pluripotent stem cells (iPSC) and differentiated motor neurons (iPSC-MN) correlating epigenetic marks of gene promoter and of the GC-rich repeat expansion (HRE) to *C9ORF72*-related pathological features (gene expression, RNA foci and DPRs). We initially studied 3 different $C9^+$ iPSC lines which were cultured for up to 40 passages *in vitro* and, at each timepoint (10^{th}, 20^{th}, 30^{th} and 40^{th} passage), DNA was extracted for genetic and epigenetic characterization, RNA was obtained for Q-PCR analyses and slides were fixed for *C9ORF72* RNA foci count. $C9^+$ iPSCs were also differentiated into motor neurons for three times and iPSC-MNs were harvested for the same molecular characterization as for iPSCs. We observed a down-regulation of the two HRE-harboring mRNA isoforms (V1 and V3) both in iPSCs and iPSC-MNs when the promoter was methylated, while RNA foci number showed no correlation with methylation state. Moreover, we found that

the epigenetic pattern of promoter methylation could change after $C9^+$ iPSC reprogramming and through differentiation into iPSC-MNs. When we extended our analysis to a cohort of 8 different $C9^+$ iPSC lines, we observed that both epigenetic marks and HRE length may influence RNA foci formation. Our study reports for the first time in $C9^+$ iPSC and iPSC-MNs that promoter methylation can be considered a possible therapeutic target and corroborates that patient-derived cells represent a suitable model for further studies on *C9ORF72*-related neurodegeneration.

Riassunto

L'espansione di ripetizioni dell'esanucleotide GGGGCC (>30 unità ripetute) situata nel primo introne del gene *C9ORF72* è la causa genetica più frequente di due condizioni neurodegenerative, Sclerosi Laterale Amiotrofica (SLA) e Demenza Frontotemporale (DFT). È stato riscontrato che il promotore al 5' di *C9ORF72*, sempre demetilato nei pazienti con SLA e/o DFT non portatori dell'espansione ($C9^-$) e nei controlli sani, risulta invece ipermetilato nel 30% dei portatori dell'espansione ($C9^+$). La metilazione del promotore di *C9ORF72* sembra essere neuroprotettiva nei confronti della tossicità dovuta all'RNA intronico e ai dipeptidi (DPRs) tossici che ne derivano per un maccanismo di traduzione non canonica (RAN translation). Lo scopo di questo studio è stato quello di correlare lo stato di metilazione del promotore di *C9ORF72* e dell'espansione stessa di esanucleotidi alle caratteristiche patologiche della mutazione (l'espressione genica, la presenza di foci di RNA e di DPRs) in due modelli cellulari derivati da pazienti SLA $C9^+$: cellule staminali pluripotenti indotte (iPSC) e motoneuroni differenziati da iPSC (iPSC-MN). Abbiamo dapprima studiato 3 linee iPSC $C9^+$, tenute in coltura per 40 passaggi. Da queste, a 4 timepoints al 10°, 20°, 30° e 40° passaggio, abbiamo estratto il DNA per le caratterizzazioni genetiche ed epigenetiche e l'RNA per le analisi di Q-PCR e abbiamo fissato alcuni vetrini per la conta dei foci di RNA. Abbiamo poi differenziato per 3 volte ciascuna delle linee iPSC in iPSC-MN, e anche queste cellule sono state raccolte per le medesime caratterizzazioni molecolari. Secondo i

nostri risultati, sia in iPSC sia in iPSC-MN, in presenza di un promotore di C9ORF72 metilato vi è una diminuzione delle isoforme di RNA contenenti l'espansione (V1 e V3). Non abbiamo trovato invece correlazione tra lo stato di metilazione del promotore e il numero di foci di RNA. Abbiamo inoltre riscontrato che lo stato di metilazione del promotore del gene *C9ORF72* può cambiare durante le fasi di riprogrammazione dei fibroblasti $C9^+$ in iPSC e di differenziamento delle iPSC $C9^+$ in iPSC-MN. L'analisi è stata successivamente estesa a una corte di 8 iPSC $C9^+$, in cui abbiamo potuto osservare che le caratteristiche epigenetiche e la dimensione dell'espansione sono in grado di influenzare la formazione di foci di RNA. Il nostro studio mostra per la prima volta che la metilazione del promotore di *C9ORF72* è un possibile bersaglio molecolare nei modelli iPSC e iPSC-MN $C9^+$ e conferma inoltre che tali linee derivate da paziente sono un modello attendibile per studi approfonditi sulla neurodegenerazione legata all'espansione di *C9ORF72*.

Contents

Research integrity disclosure

The study was conducted in full compliance with the principles of Research Integrity: Reliability, Honesty, Respect and Accountability (1). The study was designed based on preliminary data belonging to the Laboratory of Neuroscience of Istituto Auxologico Italiano and carried out during two years (October 2017-September 2019) by Clara Volpe MD-PhD Student of the Doctorate School Medicina Molecolare e Traslazionale of the Università degli Studi di Milano, under the supervision of Professor Antonia Ratti and with the technical and methodological support of the group of Neuroscience headed by Professor Vincenzo Silani. All the parties involved in the procedure declared that no conflict of interest has been arisen or encountered during the investigation.

This thesis work was evaluated by 3 independent reviewers:

- Dr. Valeria Tiranti, Unit of Medical Genetics and Neurogenetics, Fondazione IRCCS Istituto Neurologico C. Besta, Milano

- Professor Alessandro Rosa, Department of “Biology and Biotechnology Charles Darwin”, University of Rome “La Sapienza”, Roma

- Dr. Marcello Ceci, Cytology and comparative Anatomy, Department of Ecological and Biological Science, Tuscia’s University, Viterbo

1. Introduction

1.1. Amyotrophic Lateral Sclerosis

Amyotrophic Lateral sclerosis (ALS) is so far the name of an incurable disease. ALS diagnosis comes while the patient is suffering a progressive impairment of the motor system that has indolently occurred and worsened, diffusing from a single site of onset, usually one limb or bulb, to adjacent regions. A poor prognosis (3-5 years of survival on average, although there are exceptions) follows the diagnosis of such motor neuron disease so that a terrible countdown starts for patients and families. So far there is no cure for ALS and pharmacological, mini-surgical and psychological therapies currently aim to support vital functions and to relieve sufferings. Though the speed of progression is specific for each patient, the last stages imply the generalization of the paralysis until the aggravating failure of respiratory function brings to intubation, to mechanical ventilation support and death.
The incidence of ALS in Europe is 2-3 people / year / 100,000 of the adult population. Male to female ratio stands between 1 and 2 (2). The overall lifetime risk to develop this rare disease is 1:350 for men and 1:400 for women (3).

ALS occurs during the life course of healthy adults between 51 and 66 years of age (2) apparently regardless of a clear environmental risk factor (4,5). The 90% of cases are sporadic since no other cases of

ALS or related neurodegenerative disorders have been previously diagnosed in their pedigree at the time of the diagnosis. The familial cases represent instead the 5-10% of the overall burden of disease and 60-70% of them are known to be carriers of a causative genetic variant, that can be identified through genetic testing (6). The known pathogenic variants are located in an increasing number of genes on the spotlight of current research in neurogenetics worldwide (2,7,8). Even though these ALS genetic cases segregate in families as autosomal dominant or recessive traits, however a proportion of them can be due also (in the remaining 30% of cases) to a misleading occurrence by chance of two “sporadic” cases in the same family or to the co-presence of unknown genetic susceptibility factors producing a higher rate of disorder (3). Since the cause of ALS is still largely unknown and the disease is so terrible, even a first affected individual in a family can naturally awake the doubt of a possible hereditary factor among relatives. Genetic counselling is gaining increasing importance and close clinical relevance in the management of families of patients affected by ALS and is continuously keeping updated with the discoveries in the field of neurodegeneration.

The cause of ALS is nowadays switching from being considered mysterious to become a miscellany of genetic susceptibility factors combined with environmental co-factors that contribute (60% and 40%, respectively), together with the ageing of the body and of the genome, to the onset of a motor and possibly also of a cognitive impairment (3).

Interestingly, as the genetic background of individuals can be thought as the complex soil for multifactorial diseases such as sporadic ALS, the results of the last years of research in the field of ALS led to hypothesize that at least some cases of ALS can be explained as the occurrence of an oligogenic disease, due to the sum of multiple rare variants and benign polymorphisms (9,10). Even more recently, genetic background and environmental factors are supposed to act sequentially and combined in a lifelong multi-step process culminating in the onset of the neurodegenerative disease (11).

1.1.1. Clinical History

The clinical site of onset typically defines the first classification of ALS cases, distinguishing the two main clinical variants:

- Spinal onset
- Bulbar onset

The spinal onset is the most common variant of ALS and also the most diffuse one among males. Early symptoms appear usually asymmetrically involving selectively the motor function of upper or lower limbs and sparing sensitive and autonomic aspects. The bulbar site of onset causes dysarthria, dysphagia, dysphonia, sialorrhea and *ab ingestis*, being epidemiologically related to more rapid disease.

The other main clinical stratification of ALS cases is based on the early involvement of the central or the peripheral motor neuron:

- Upper motor neuron (UMN, in the motor cortex)
- Lower motor neuron (LMN, in the brainstem and spinal cord)

Patients can suffer from weakness, cramps and fasciculation if the LMN is primarily involved or from spasticity and hyperreflexia due to

the UMN, until the anatomical spread of the disease comes to impair all the motor pathway and the progression of the disease brings the paralysis due to muscular atrophy to replace other symptoms. Diagnostic criteria traditionally assume that the disease runs across a determined anatomical trajectory, even if such dynamic description of the disease is consistent with the clinical progression in a limited percentage of patients. Due to the subgroups of ALS patients, nowadays it's known that clinical progression reflects the contiguity progression of neurodegeneration both of UMN (cortical topography) and LMN (latero-lateral and rostro-caudal dissemination).
In fact, under the name of ALS are contained many subtypes of "atypical" ALS, as well as upper or lower restricted phenotypes (Primary Lateral Sclerosis and Progressive Muscular Atrophy respectively) or anatomically limited phenotypes (progressive bulbar palsy, flail arm / leg syndromes) as well as neuropsychiatric overlapping syndromes (with cognitive and behavioral impairment).

Four therapeutics are currently approved by FDA for the treatment of ALS:

* **Rilutek**, 1995 (generic), was the first FDA-approved drug available to treat ALS. Rilutek is the commercial name of riluzole, an inhibitor of glutamate release. It prolongs life approximately three months but has no effect on muscular strength. Despite the apparently low therapeutic efficacy more recent studies based

on real-word evidence support the role of riluzole, with life prolongation over a year.

- **Nuedexta**, 2011 is a combination of dextromethorphan and quinidine and is prescribed for the pseudobulbar symptoms, such as inappropriate crying or laughing, which some people living with ALS can develop.
- **RadicavaTM** (Edaravone), May 2017 (Mitsubishi Tanabe Pharma America), is the first new treatment specifically for ALS after 22 years from Riluzole. It consists of a neuroprotective agent that is thought to act as a free radical scavenger, preventing oxidative stress damage to the neurons. It ameliorates patient's mobility, but the effect on survival is still unknown.
- **Tiglutik**, September 2018 (ITF Pharma), is the only thickened liquid form of riluzole. This formulation is designed to avoid potential problems of crushing tablets.

None of these drugs is able to change the course of the disease. More than a hundred of clinical trials are actively recruiting ALS patients and thousands of laboratories are working on the biological bases of the disease all around the world, aiming to identify new therapeutic targets and drugs.

1.1.2. Pathophysiology of neurodegeneration in ALS

ALS symptoms are consequent to the progressive loss of connection between the central nervous system (CNS) and striated muscles, produced by a relentless degeneration of motor neurons, the cells that

produce, transport and deliver the motor input from the brain to the periphery. The onset of ALS occurs usually in adulthood, while the biological events that contribute to cause neuronal death, still largely undefined, are supposed to occur earlier and act slowly for several years. Collateral reinnervation is able to vehicle electric signal to the periphery to some extent, further delaying the motor impairment, until the degeneration of the axon of surviving motor neurons.

Neuropathology has always assisted researchers the progress achieved towards gaining insight of the pathogenesis of ALS. The name “Amyotrophic Lateral Sclerosis” was indeed attributed from the observation of the substitution of motor neurons in the anterior horns of spinal cord by sclerotic matrix after alpha-motor neurons massive death in ALS autoptic tissues. For decades vacuolization, spongiosis, shrinkage of α–motor neurons and Betz cells of the motor cortex, brainstem nuclei and anterior horns of the spinal cord were characterized, together with typical protein aggregates within and around the spared motor neurons from autoptic ALS specimens. In the ‘90s molecular neuropathology of neurodegenerative disorders allowed deep progresses in the analysis of pathologic inclusions, that were found to contain ubiquitin, alpha-synuclein (Lewi bodies) and tau protein (Pick bodies) (12). The same inclusions were found also in brain samples of patients suffering from other neurodegenerative conditions such as frontotemporal lobar degeneration (FTLD), a neurodegenerative disease that showed a partial clinical overlap with ALS. Many studies of molecular biology, mainly based on functional studies and on candidate gene approach, were conducted to

investigate the molecular pathways involved in neurodegeneration, the role of the genome and of the transcriptome. From these studies together with linkage analyses it was possible to determine that a number of genetic variants segregated with the disease in familial cases, opening the way to the modern genetics of neurodegeneration. So far ALS pathogenesis is known to involve many biological pathways, recognized as impaired cell function and leading to neurodegeneration. (13):

* RNA metabolism;
* protein quality control;
* cytoskeletal integrity;
* cytoskeleton-mediated trafficking;
* mitochondrial function and transport.

Many causative genes associated to familial ALS, including *TARDBP, C9ORF72*, *FUS*, *TIA1*, *MATR3*, *HNRNPA1*, *HNRNPA2/B1*, *EWSR1*, *TAF15*, and *ANG*, are involved in regulating RNA transcription, splicing and metabolism and their impairment in ALS contributes to a defective global RNA processing. Other ALS genes, such as *UBQLN2*, *VCP*, *OPTN*, *VAPB*, *TBK1*, and *SQSTM1*, encode for protein homeostasis factors, that regulate proteasomal and lysosomal degradation mechanisms and that together with stress granules formation and autophagy are involved in response to stress. The *PFN1*, *TUBA4A*, *KIF5A*, *ANXA11* gene mutations suggest the importance of cytoskeleton integrity and axonal transport for neuronal metabolism. The impairment of axonal transport was also proposed

as responsible of the disruption of organelles and mitochondrial trafficking in motor neurons for local translation at the synapses. Lastly, mitochondrial function (*SOD1*, *CHCHD10* genes) is known to be affected in ALS together with an altered calcium homeostasis, abnormal reactive oxide species (ROS production) and defective cell bioenergetics which can also activate autophagy (7).

1.1.3. ALS-FTLD spectrum

Since the last decade of the XX century, the co-occurrence of cognitive/behavioural impairment in some ALS patients became not only clinically recognized but also supported by epidemiological, neuropathological, neuroimaging, neurogenetic and neuropsychological evidence. The studies demonstrated an overlap between motor neuronal degeneration in ALS and the death of cortical neurons in frontotemporal lobar degeneration (FTLD).

The discovery of the convergence of ALS-FTD neuropathology on common TDP-43 protein inclusions (2006) was the first important step. The further discovery that the main component of cytosolic ubiquitinated inclusions in 97% of ALS patients and 45% of FTLD patients was in fact the TAR DNA-binding protein 43 (TDP-43) represented a seminal milestone in comprehending the ALS physiopathology (8,14,15). The mechanism of pathological accumulation of TDP-43 protein into the cytoplasm of motor neurons of sporadic and familial ALS patients seems to be associated to an impairment of the nucleocytoplasmic transport dynamics due to TDP-43 formation of aggregates because of its prion-like domain and to its subsequent nucleus depletion. (16–18).

Once the intronic hexanucleotide repeat expansion in *C9ORF72* gene was identified in families with both ALS and FTLD (19,20), for the first time also a common genetic cause correlated ALS to FTLD. The discovery that the clinical continuum between the two forms of neurodegeneration had a genetic origin confirmed years of clinical, neuropathological and neuroimaging observations.

1.2. *C9ORF72* gene

C9ORF72 gene is located on the reverse strand of the short arm of chromosome 9 (C9p21.2-p13.3). The gene is 27,320 base pairs long and is composed of 12 exons and 11 introns (Figure A).

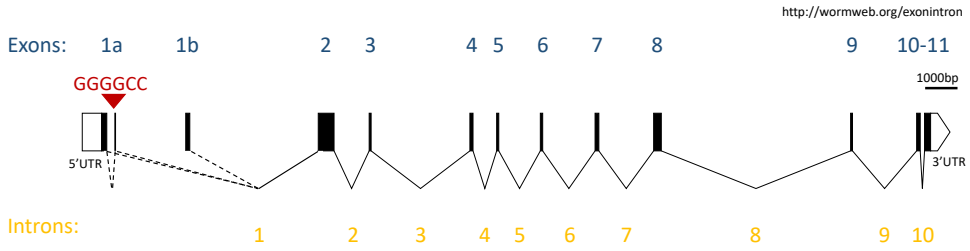


Figure 1-1. Map of *C9ORF72* gene structure. The 11 exons are indicated as black boxes, introns as continuous lines (constitutive splicing) or dashed lines (alternative splicing). The location of the repeat expansion (GGGGCC) in the first non-coding intron is indicated by the red arrow.

The two first exons belong to the non-coding 5'UTR of the gene and are alternative transcription start sites (TSS). There are three different *C9ORF72* mRNA variants (V1, V2 and V3). Two of them, V1 (NCBI Reference sequence: NM_145005.6) and V3 (NM_001256054.2), start from the exon 1a TSS, corresponding to the first non-coding

exon, while the V2 isoform (NM_018325.4), that is the most abundant transcript, is transcribed from 1b TSS, corresponding to the second non-coding exon. V2 and V3 are translated into a long protein isoform of 55 kDa (481 amino acids), encoded by exons 2-11, while the V1 (exons 2-5) is short and encodes for a 25 kDa (222 amino acids) protein.

The function of the protein is still under investigation but, based on structure analysis, C9ORF72 was first defined as a DENN-like (differentially expressed in normal and neoplastic cells) protein and involved in membrane trafficking (21). The DENN-like domain indeed placed it among GDP-GTP exchange factors (GEFs) for Rab GTPases, that are the master regulators of intracellular compartments trafficking. The N-terminal end of the three mRNA isoforms (exons 2-5) harbors a Longin domain, that is known to interact with various types of GTPases involved in the endoplasmic reticulum-to-Golgi transport, while only the long isoform (exons 2-11) contain also the DENN and the C-terminal alpha domain that define a more extended and complex interaction surface involved in GTPase activation (22,23). The long C9ORF72 isoform was then identified as part of the autophagosome assembly machinery by interacting with SMCR8 (Smith-Magenis syndrome chromosome region, candidate 8) and WDR41 (WD-repeat domain containing) scaffolds, (24). These predictions were therefore confirmed by the functional evidence that C9ORF72 takes part in a multiprotein complex that regulates the key autophagy initiation ULK1 complex (25). Furthermore C9ORF72 was demonstrated to be localized at lysosome membranes, where it

seems to recruit mTORC1 targets, contributing also to protein synthesis control by the mTOR pathway (26).

1.2.1. The Hexanucleotide Repeat Expansion in *C9ORF72*

As discovered by linkage analyses on patients suffering from disorders of ALS-FTLD neurodegenerative spectrum, an hexanucleotide repeat expansion (HRE) in the first intron of *C9ORF72* gene is the major cause of familial ALS and FTLD (19,20).

The first intron of *C9ORF72* contains a polymorphic sequence in humans, where the hexanucleotide GGGGCC ranges from 2 to 23 repeats in the normal range, while it expands from 30 up to 4,000 repeats in mutant cases of ALS/FTLD (27).

The age of onset of expansion carriers is slightly earlier than unexpanded ALS/FTLD, survival is reduced and, similarly to other repeat expansion disorders such as Huntington disease, Friedreich ataxia and Myotonic dystrophies, families exhibit genetic anticipation (28,29). *C9ORF72* mutations account for 25.4 - 41.8% of the familial ALS cases, 3.4 - 8.5% of the sporadic ALS cases, 14.3 - 37.5% of the familial FTLD cases and 1.2 - 9.0% of the sporadic FTLD cases depending on the ethnicity. The expanded allele is present also in the 0.03 – 0.3% of control subjects (28).

The penetrance of the HRE is incomplete and age-dependent (nearly complete by 83 years of age) (30) so that apparently sporadic mutated patients indeed represent undiagnosed familiar cases (3,9).

The mutational frequency in individuals with concomitant ALS-FTLD phenotype is very high (33-86%) representing the most probable pathogenic variant in this clinical subgroup (31).

1.2.2. Pathogenic mechanisms of *C9ORF72* HRE

The presence of the HRE in *C9ORF72* is causative of neuronal loss through three main pathogenic mechanisms, including both loss and gain of function. There is evidence that all these pathomechanisms, haploinsufficiency and RNA or protein toxicity, contribute at different extent to induce damage at multiple fundamental cell pathways, culminating in neuronal loss in the motor system (ALS) and in frontal e temporal lobes (FTLD) or both in overlapping ALS-FTLD cases. In the brain of $C9^+$ patients C9ORF72 protein was found reduced by nearly 50% compared to controls (20,32,33). A loss of C9ORF72 protein function is believed to involve the long protein isoform, in a different extent from the short one that can instead be upregulated (34,35). Gene expression analyses of *C9ORF72* showed that the V1 and V3 transcripts, that contain the HRE, are upregulated in brain and blood of $C9^+$ patients, while the main V2 isoform is downregulated (36). The single or double knockout of *C9ORF72* and *SMCR8* in mice produced dysregulation of autophagy and systemic inflammation (25,34), indicating that *C9ORF72* loss of function is responsible for an increase of the autophagic activity. *C9ORF72* knockout mice showed also the impairment of lysosomal function and of mTOR pathway (26). Also stress granules and p-bodies formation is involved in the cascade of adverse events in $C9^+$ neurons in association to C9orf72 protein loss of function (37–39).

On the other hand, and in concomitance, the HRE also generates stable G-quadruplex structures of RNA and R-loops during transcription, that are RNA-DNA hybrids (40,41). These stable RNA structures, forming because of the pairing of the GGGGCC repeat

sequence, promote the formation of truncated RNA transcripts containing the HRE, distributing both in the cytoplasm and in the nucleus as RNA foci (20,42–46). RNA foci were shown to sequester nuclear RNA binding proteins and splicing factors as well as nucleocytoplasmic transport regulators, acting through a RNA gain of function mechanism. From these RNA transcripts containing the repeated sequence, despite the HRE is intronic, a non-canonical mechanism of translation, the repeat-associated non ATG (RAN) translation, occurs and generates Dipeptide Repeats Proteins (DPRs) (47,48). *C9ORF72* DPRs are composed of repeated pairs of only two amino acids that are translated from all the possible reading frames of the sense (GA, GP, GR) and antisense (PA, GP, PR) strands of the $(GGGGCC)_n$ HRE (47,48). DPRs have been shown to act through a gain of function mechanism as they induce nucleolar stress, impair mRNA splicing and ribosomal RNA biogenesis, interfere with stress granules function and also because their injection into cells and animal models is sufficient to cause neurodegeneration (49,50).

1.2.3. Modifying factors of *C9ORF72* ALS-FTLD phenotype

The dual possible clinical phenotype of *C9ORF72* HRE carriers and the variability in the penetrance of the mutation suggest that other concomitant factors may influence the genotype-phenotype correlation. These modifying factors may reside in the patient's genome or come from the environment as well as belong to the genome / environment interface.

As regards the possible environmental modifying factors, no data have been published so far. In defining the role of environment, probably a

methodological bias due to the traditional clinical stratification occurred, together with the difficulty to collect anamnestic, clinical and genetic data on small cohorts of a rare disease such as ALS (3).

1.2.4. Genetic modifying factors

Genetic screenings of ALS-FTLD causative genes in *C9ORF72* carriers showed also a higher incidence of variants in these loci compared to $C9^-$ ALS-FTLD patients so that they were investigated as modifiers of disease phenotype. Many of them were also found among unaffected individuals, where they can be considered benign polymorphisms. However, their incidence was found higher among $C9^+$ (*SOD1*, *TARDBP*, *FUS*, *ANG*, *OPTN*, *UBQLN2*, *VAPB*, *DAO*, *PRPH*, *GRN* and *PSEN2*), indicating a possible role of these variants in disease onset or in modulating the pleiotropic phenotype in $C9^+$ ALS/FTLD carriers (9).

Based on these data supporting a possible oligogenicity in $C9^+$ carriers, some associations were also observed between ALS phenotype and common SNPs in *FUS*, *OPTN*, *ANG* or *SOD1* genes, between pure FTD phenotype and *GRN*, and between ALS-FTLD and *NEK1*, *TARDBP* or *TBK1* (8).

Moreover, *C9ORF72* gene expression was found downregulated in *MAPT* and *GRN* mutated FTLD patients, reinforcing the hypothesis of an interplay between upstream regulators and *C9ORF72* gene (36).

Another possible genetic modifying factor is the presence of small deletions within the GC-rich low complexity sequence (LCS) upstream the gene. The deletions frequency in this region is higher in $C9^+$ compared to $C9^-$ patients (6.0-28.6% vs 0.5-1.2%; $P < 0.001$) (27,51).

It is postulated that this LCS region produces an incomplete GGGGCC-like repeat facilitating the slippage of the DNA polymerase during replication and so possibly contributing to the instability and expansion of the hexanucleotide repeat sequence (51).

Genome-wide screens in *Drosophila melanogaster*, later validated also in *Saccaromyces cerevisiae*, identified that HRE RNA/DPR toxicity was modulated by overexpression of several genes related to regulation of nucleocytoplasmic transport (52–55), starting a new field of investigation on the role of the nuclear pore and envelope in *C9ORF72* pathophysiology.

The length of the HRE in autoptic $C9^+$ brains was investigated as a possible genetic modifier, too, where it was found generally smaller in cerebellum compared to frontal cortex and blood (56,57). A correlation between HRE length and age at onset resulted significant only among FTD patients in the frontal cortex, while survival was worse when the size of the HRE was bigger in the cerebellum (58). *C9ORF72* HRE is also a mosaic above all in the blood of $C9^+$ patients, where the plethora of cells from the bone marrow that compose blood (lymphocytes, polymorphonucleates and rare monocytes and reticulocytes) show high HRE instability (56). These observations also indicate that a clear correlation between HRE size and phenotype based on blood HRE length has a bias and is difficult to establish because it is not representative of the HRE in the $C9^+$ brains.

1.2.5. Epigenetic marks as modifying factors of *C9ORF72*-related pathology

Gene expression dysregulation is a crucial hub of *C9ORF72* pathology. While the presence of the HRE seems to be sufficient to downregulate *C9ORF72* transcription and protein synthesis, transcript-specific differences in gene expression have been associated to DNA hypermethylation of *C9ORF72* locus, as discussed below. Moreover, other repeat expansion disorders, such as Fragile X syndrome and Friedreich ataxia (59,60), already showed that DNA expansions that are enriched in CG sequences are associated to downstream gene silencing and to loss of function mechanisms.

1.2.6. Methylation of C9ORF72 promoter

Human *C9ORF72* promoter is constituted of 26 CpG sites belonging to a 228 bp-long CpG island (Figure 4a). CpG islands are generally highly conserved regions (up to 3000 bps) with a percentage of C and G higher than 50% and more than the 60% of predicted CpG sites.
CpG islands are usually located at gene promoters and in regulatory regions and are actively kept unmethylated by an active enzymatic activity. Methylation / demethylation of cytosines at the CpG sites maintain the chromatin close or open, respectively, to inhibit or promote gene transcription.
C9ORF72 5' CpG island is absent in mouse and zebrafish, while it is completely conserved in chimpanzee (61). The presence of the human CpG island is suggested to be selectively evolved as compensatory to *C9ORF72* gene expression or to the deleterious effects of the $(GGGGCC)_n$ repeats. In support to this phylogenetic hypothesis, while *C9ORF72* gene promoter is always unmethylated in controls (2-23 repeats), in about 30% of HRE ALS and FTD carriers the CpG island

is found hypermethylated in blood and brain tissue (29,62–65). Moreover, *C9ORF72* promoter methylation was demonstrated to be monoallelic, involving only the expanded allele (65). Importantly, correlation studies in familiar versus sporadic $C9^+$ ALS and FTLD and within familial pedigrees showed that *C9ORF72* promoter methylation is higher in familial ALS (62,66).

The methylation of the 5' CpG island at *C9ORF72* gene has been studied as a potential modifying factor and was reported to downregulate *C9ORF72* gene expression in blood and brain tissue of $C9^+$ patients (62) compared to unmethylated $C9^+$. It was also found to associate to a slightly longer disease duration and to a later age at death in a cohort of $C9^+$ unrelated FTLD individuals (66), suggesting a neuroprotective role.

Lymphoblasts from methylated and unmethylated $C9^+$patients were the first patient-derived cell model where *C9ORF72* variant-specific gene expression was assayed in association to promoter methylation state. In this model, the unmethylated $C9^+$cells showed an increase in the usage of the exon 1a TSS, and a subsequent increase of the V3 transcript containing the intronic HRE sequence, while the presence of a methylated promoter was associated to the downregulation of the V3 isoform (65). On the same three lymphoblastoid lines considered in this study a methylated promoter was associated also to a lower RNA foci formation (65). The negative correlation between promoter hypermethylation and RNA foci number was confirmed also in post-mortem cerebellar tissue, together with a concomitant decrease of pathological DPRs (65).

The results of a MRI *in vivo* neuroimaging study reported that *C9ORF72* promoter hypermethylation in patient's cerebellum was related to less atrophy of the gray matter in right hippocampus, thalamus and prefrontal cortex and that the presence of this epigenetic modification was correlated with a reduced longitudinal decline and to better neuropsychological outcome, in association to an overall neuroprotection (67).

A transgenic $C9^+$ animal model was generated resembling also the epigenetic features of *C9ORF72* locus through bacterial artificial chromosomes (BAC) technology (68). In the brains of these C9-BAC mice, promoter hypermethylation, determined by bisulfite pyrosequencing analysis of 8 CpG sites, occurred in a subset of mice and was related to a trend towards decreased DPRs (GPs) levels (69). Promoter methylation state of $C9^+$ fibroblast-derived iPSCs and of iPSC-derived neuronal lines (neural precursors and neural-enriched teratomas) was also compared to the corresponding $C9^+$ human embryonic stem cells (hESC). A difference in *C9ORF72* promoter methylation was found, and iPSCs showed a hypermethylated promoter that was associated to the decreased transcription of V3 isoform from the 1a TSS (70).

1.2.7. Methylation of the C9ORF72 HRE

CpG sites are recognized by DNA methyltransferases (Dnmt) 1/3, that are able to transfer a methyl group from an adenosyl-methionine to the 5th carbon of the target cytosine residue, which is followed in the sequence by a guanosine. There are three main isoforms of Dnmt: Dnmt1 acts on hemimethylated substrates in replication fork and is in

charge of the maintainance of methylation patterns during replication; Dnmt2 is an RNA methyltransferase; Dnmt3a-b are instead *de novo* methyltransferases and work by methylating hemimethylated DNA or completely unmethylated DNA. Besides their essential role in neurodevelopment, Dnmts are known to be expressed in adult brain and in neurons where their impairment was already described in ALS (71).

As the $(GGGGCC)_n$ sequence in *C9ORF72* gene is expanded up to thousand times, together with the enormous genomic expansion at the locus, also the number of CpG sites is immensely increased at this intronic region.

The discovery that the HRE is an active hotspot of DNA methylation belongs to the first years of epigenetic studies of *C9ORF72* locus (63). DNA methylation was initially investigated by restriction enzyme digestion with MspI and HpaII whose recognition site is the CCGG sequence, followed by PCR amplification of the target region. While MspI is not sensitive to DNA methylation, HpaII is methylation-sensitive and it is able to cut only when the site is unmethylated. This kind of analysis in $C9^+$ lymphoblasts initially showed that the HRE was cut by both enzymes differentially, and so that it was unmethylated, at least in some CpG sites (65). A limit of this molecular genetic technique was in fact that, due to the repetitive sequence of the short interspaced CpG sites into the HRE, a single unmethylated CpG would prevent DNA amplification.

An innovative assay was then developed to evaluate HRE methylation in response to the need to detect methylation of one single CpG site, even in the case that some unmethylated cytosines would have been

present in the HRE. This method resulted from the combination of bisulfite conversion and the methylation-specific Repeat Primed PCR already used to detect the presence of HRE as a diagnostic tool (63). The HRE so investigated was found generally methylated (97%) in blood and brain tissue of ALS/FTD patients carrying an expansion size higher than 90 repeats, while short carriers (23-43 repeats) were unmethylated as well as unexpanded controls (63). However, this method has the limit that only the CpG sites at the 5' end of the HRE (40-50 units) can be detected.

1.2.8. Histone modifications related to *C9ORF72* locus

Contemporarily to DNA methylation studies, $C9^+$ brain, blood and fibroblastoid cell lines were found enriched in repressive epigenetic markers such as histone 3 and 4 lysines trimethylation (H3K9, H3K27, H3K79 and H4K20) and this epigenetic modification associated to the downregulation of *C9ORF72* gene expression (64). Besides modifying the DNA structure and increasing Dnmts capacity to catalyze methylation of cytosine residues within promoter and the repeat itself, the presence of the HRE was supposed to facilitate also the modification of the histone tails bound to the locus. This brought to speculate an additional modulation of *C9ORF72* locus and gene expression by histone methylation (64). While DNA methylation at gene promoter was prerogative of a subgroup of $C9^+$, H3 and H4 histone tails trimethylation was described in all $C9^+$ carriers, similarly to methylation of the HRE itself.

Interestingly, modifications of histone residues add a dynamic dimension to the pathogenesis of the disease, since histone epigenetic state is known to change in an age-dependent way (72,73).

1.2.9. Hydroxymethylation of *C9ORF72* promoter

A percentage of cytosines is known to be also hydroxymethylated, consisting in the partial oxidation of the already existing methyl group at the CpG site. This epigenetic modification is known to be present in about 10% of methyl-cytosines in the whole genome, and to be more frequent in the central nervous system (74–77). Hydroxymethylation is known to be caused by ten-eleven translocases (TET), dioxygenases involved in the control and maintenance of the CpG islands in an hypomethylated status. Hydroxymethylation seems to act as a very dynamic regulatory mechanism in differentiated cells to enhance transcription, modifying the existing methylation which, in contrast, inhibits gene transcription. It represents also an intermediate of cytosine demethylation (78).

A preliminary study of *C9ORF72* promoter hydroxymethylation showed that it was increased in the CNS of $C9^{+}$ patients who carried a methylated promoter compared to mesoderm-derived tissues, and in both iPSCs and iPSC-MNs compared to fibroblasts (75). However, the role of *C9ORF72* promoter hydroxymethylation is still poorly understood and needs further investigation, although it is methodologically compelling to discriminate between methylation and hydroxymethylation.

1.3. Patient-derived cell models of *C9ORF72* molecular pathology

The study of neurodegenerative diseases is hampered by the impossibility to obtain patient-derived neurons from the affected brains. However, peripheral cell models, including lymphoblasts and fibroblasts, have been used to study ALS pathomechanisms and they have revealed to be useful *in vitro* models. The quite recent possibility to obtain human neurons from induced pluripotent stem cells (iPSC) has completely changed the experimental approach to study ALS disease.

The advantage of $C9^+$ patient-derived cells is that they contain and express the expanded pathologic sequence without any genetic manipulation and are representative of the genetic background of the patient.

1.3.1. Skin-derived $C9^+$ fibroblasts

Fibroblasts are the most diffused stromal cells of the body. They can be obtained from the skin by excisional or punch biopsy and can be cryopreserved or cultured up to 10 passages *in vitro*. Dermal biopsy is not a very invasive procedure, so that it is not infrequent to voluntarily receive samples also from healthy relative donors or large mutated families.

As regards *C9ORF72* mutation, patient-derived fibroblasts were largely used because they recapitulate one of the most important *C9ORF72*-related trait, such as the formation of pathological RNA foci

easily detectable by RNA *in situ* hybridization in their nucleus and cytoplasm (79,80).

Despite the advantage of using fibroblasts in culture, some features of *C9ORF72* pathology are known to miss in patient's fibroblasts compared to autoptic frontal cortex and spinal cord, such as the absence of both pathological DPRs, which are brain-specific (48), and of TDP-43 inclusions (81). Another issue with fibroblast cell model is that the pathologic HRE they harbor may have a different length compared to neuronal cells (56,75) and may therefore not be completely representative of affected cells in the brain.

1.3.2. Lymphocyte-derived $C9^+$ cultures

Nucleated blood cells have been used by many research groups as suitable patient-derived model of *C9ORF72* pathology. While mononuclear blood cells of a peripheral blood sample show a variable mosaicism in HRE size, lymphoblastoid lines and lymphoblasts, cultured as clonal lines, are generally representative of a unique and generally larger HRE compared to fibroblasts, and resemble HRE sizes of related neuronal cells (82). However, the determination of the *C9ORF72* HRE length in more than one patient's lymphoblast line showed genomic instability in blood-derived cells (57).

Based on CAGE-seq data, *C9ORF72* isoform expression is similar between hematopoietic and neuronal cell lines, differently from fibroblasts (56). These cells show also RNA foci that, together with the characteristic $C9^+$ gene expression, make them a good model of *C9ORF72*-related molecular pathology.

1.3.3. iPSC and iPSC-MNs

Induced Pluripotent Stem Cells (iPSCs) are pluripotent stem cell cultures that are obtained from adult differentiated somatic cells (fibroblasts, blood cells) by genetic reprogramming, aimed at expressing stemness genes and at returning to an uncommitted phenotype (83).

iPSC lines are now considered the frontiers of *in vitro* models of Mendelian diseases and the best suitable model for a personalized medicine. In fact, on one hand the genome of the iPSCs can be targeted or edited obtaining an isogenic control cell, a cell line with the same genome as the patient except for the mutation, and on the other hand the patient's pluripotent rescued cell has the perspective of a transplantable device.

However, many papers have been already published about how much caution is still warranted dealing with patient-derived iPSC, which are difficult to monitor and to correctly use (79,84–87).

Epigenetic modifications in iPSCs have been extensively studied. DNA methylation is the most investigated epigenetic trait. It seems clear that, while pluripotent cell genome is usually demethylated, iPSCs host the trace of the somatic epigenetic memory, consisting in islands of hypermethylation and differences in gene transcription (88,89) .

iPSCs can also be manipulated by sequential addition of cytokines, inhibitors of pluripotency and growth factors to be differentiated into virtually any cell lineage. Working with iPSC-derived neurons has therefore a great advantage to study neurodegeneration, because cell

pathways are the most similar as possible to the affected cells in the disease.

As regards *C9ORF72*-positive iPSCs, the HRE size was found to be unstable after reprogramming and during iPSC culturing *in vitro* (39), while other studies reported that the HRE is maintained stable in these cell models (70).

$C9^+$ iPSCs and iPSC-motor neurons (iPSC-MNs) show the formation of pathological RNA foci at variable extent, histone trimethylation, DNA methylation of *C9ORF72* locus and the typical $C9^+$ transcript variant expression profile observed in blood such as the decrease of the main isoform V2 and the increase of V1 and V3 ones (39,70,90).

One single study so far investigated a serial collection of DNA from $C9^+$ fibroblasts, iPSC, iPSC-neural stem cells (NSC) and iPSC-MN and described that *C9ORF72* promoter changed from a methylated to a hydroxymethylated state during iPSC reprogramming, while both epigenetic modifications increased during iPSC-NSC transition and differentiation (75).

1.3.4. Targeting *C9ORF72* using new molecular strategies

On the evidence that *SOD1* mutations in ALS cause neuronal death through a gain of function mechanism were based the first applications of antisense oligonucleotides (ASOs) to target pathological mRNA and prevent downstream pathogenicity by RNase-H mediated target degradation (91). After years of experiments on animal models indicating the improvement of neuromuscular function and the decrease of mutated SOD1 in cerebrospinal fluid using ASOs, the first phase III trials on *SOD1* mutated patients are now ongoing.

Several groups are working with ASOs targeting also *C9ORF72* HRE (45,79,90,92,93). Not only ASOs, but also small molecules targeting the $(GGGGCC)_n$ sequence, are on the spotlight of research as regards the new molecular therapy for $C9^+$ ALS/FTLD patients (94,95). RNA foci number and DPRs are currently used as readouts of effectiveness of ASOs / small molecules in $C9^+$ iPSC-MNs since, when targeting mRNA transcripts, they are expected to decrease.

Recent findings suggest that also DNA methylation and histone tails modification, the two most studied epigenetic marks related to mutant *C9ORF72* gene, may represent possible targets for novel therapeutic strategies based on epigenetic manipulation. In support of this, the increased genomic DNA methylation at CpG sites, a feature linked to the biological age of the genome, was proved to be associated to earlier age of onset in $C9^+$ patients, and therefore to act as a druggable modifier of *C9ORF72*-related pathology (96).

2. Aim of the study

In ALS and FTLD cases due to the intronic hexanucleotide repeat expansion (HRE) in *C9ORF72* gene, clinical, neuropathological and neuroimaging association studies suggest that DNA methylation of *C9ORF72* promoter seems to act as a neuroprotective factor in *C9ORF72*-related neurodegeneration. Aim of my PhD project was to use mutant *C9ORF72* ($C9^+$) patient-derived iPSC and iPSC-MN to investigate the correlation between DNA methylation in the 5' gene regulatory region and pathological RNA foci formation, gene expression and HRE length, as quantifiable features of *C9ORF72*-related pathology.

iPSC and iPSC-MN represent the best suitable *in vitro* models for *C9ORF72* pathology because pathomechanisms associated to the endogenous mutation, where the HRE can reach up to 4,000 repeat units, and therapeutic strategies can be studied physiologically in the setting of the patient's genetic background and in a neuronal environment.

Whether epigenetic marks in *C9ORF72* gene are able to modify pathological hallmarks associated to the HRE is indeed an indispensable prerequisite to consider for the development of any therapeutic strategy in $C9^+$ iPSC-MN, and the epigenetic manipulation of *C9ORF72* locus may represent itself a novel approach for $C9^+$ ALS/FTLD patients.

3. Materials and methods

3.1. Samples supply

Blood samples, skin biopsies and clinical data for this study were provided by the Neurology Unit at Istituto Auxologico Italiano. The informed consent for research purpose was obtained by each participant according to the Ethic committee of the Institute. $C9^+$ and control fibroblast lines were already available in our biobank, while three commercially available iPSC lines (two $C9^+$ and one isogenic control) were provided by Prof. S. Corti, Ospedale Ca' Granda di Milano, and acquired at the Cedars-Sinai Medical Center (USA).

3.2. iPSC

An experimental flow was designed that followed five iPSC lines (among which three were expanded in the first intron of *C9ORF72* gene and two $C9^-$ controls) along 10 months, in order to observe if and how epigenetic features of *C9ORF72*, repeat expansion length and RNA foci formation would appear to interplay both in patient-derived iPSC and iPSC-derived motor neurons (iPSC-MNs).

3.2.1. Fibroblast reprogramming to iPSC

iPSC clones were obtained by reprogramming of primary fibroblasts, collected from skin biopsies and cultured in RPMI medium (Gibco) with 10% bovine serum (Sigma), 1% L-Glutamine, 1% penicillin/streptomycin (Gibco) and 1% amphotericin. At first passages

3×10^5 fibroblasts were seeded and then transduced using CytoTune-iPS 2.0 Sendai reprogramming kit (Thermo Fisher Scientific) consisting of Sendai virus-mediated induction of stem cell factors (Klf4, Sox2, c-Myc and Oct 4). After 7 days of culture fibroblasts were seeded on Matrigel (Corning)-coated dishes and RPMI medium was replaced with Essential 8 medium (Thermo Fisher Scientific) to allow colony formation. Twenty days post-transduction colonies were manually picked and plated on Matrigel-coated plates for further expansion. (83,97).

3.2.2. iPSC culture and characterization

Colonies were expanded in Essential 8 medium and could be maintained through oner 50 passages, detaching by EDTA 0.5 mM solution. iPSC clones formed round or oval shaped colonies with defined edges and high nucleus/cytoplasmic ratio. For this study iPSCs were collected at the first passage available <3 and then at passages 5, 10, 20, 30 and 40.

To ascertain their normal karyotype, iPSCs were collected and pellets incubated in 5 mL of 0,56% KCl for 6 minutes at room temperature, then of 5% acetic acid was added as pre-fixative for 3 minutes. 5 mL of 3:1 methyl alcohol and acetic acid fixating solution were added to the samples. The dilution was smeared on wet slides. Quality of mitoses and mitotic index were evaluated at 20X Leica phase contrast microscope. Slides were dried at 60°C for 1 hour. Slides were then smeared with quinacrine mustard and mounted with Mc Ilvane pH 7 buffer (82% 0,2 M sodium phosphate and 18% 0,1 M citric acid).

Slides were observed with Leica fluorescence microscope, using the λ 510 nm UV filter. Metaphases were acquired with 100X magnification immersion objective and were analyzed by the CytoVision (Leica) software. At least 20 metaphases were considered for each sample. Chromosome banding resolution was 400 bands. Karyotype was formulated according to International System Cytogenetics Nomenclature (ISCN 2016).

3.2.3. Stem cells markers expression

Immunofluorescence analyses were performed to detect pluripotency markers expression. iPSCs were fixed in 4% paraformaldehyde for 20 minutes and permeabilized using 0,3% triton X-100 and cold 100% methanol. Cells were incubated with 10% normal goat serum (NGS) in PBS for 20 minutes and then with primary antibodies Alkaline Phosphatase (Abcam abID#116592), SSEA4 (MC-813-70 Invitrogen, Carlsbad, CA, USA) and Tra-1-60 (Invitrogen, Carlsbad, CA, USA) were added for 90 minutes at 37°C. After two 5 minutes washes in blocking solution, coverslips were incubated with Alexa 448 or 555 conjugated secondary antibody (1:500 Thermo Fisher Scientific) at room temperature for 45 minutes. DAPI solution was then added for 5 minutes at room temperature, coverslips washed in PBS and mounted with FluorSave (Merck, Darmstadt, Germany). Images were acquired at 40X magnification using a Nikon Eclipse Ti C1 confocal microscope system.

Further characterization of stemness marker expression was performed by RT-PCR, investigating Sox2, Nanog and Oct3/4 mRNA presence. Total RNA was extracted from fresh or -80°C stored iPSCs

pellets using Trizol reagent as described below and reverse transcribed into cDNA with Superscript II reverse transcriptase (Thermo Fisher Scientific) in a mix with oligodT primers.

Primers for RT-PCR Stem cell markers	
Sox2 Forward	TTGCGTGAGTGTGGATGGGATTGGTG
Sox2 Reverse	GGGAAATGGGAGGGGTGCAAAAGAGG
Nanog Forward	CAGCCCTGATTCTTCCACCAGTCC
Nanog Reverse	GTTCTGGAACCAGGTCTTCACCT
Oct3/4 Forward	GACAGGGGGAGGGGAGGAGCTAGG
Oct3/4 Reverse	CTTCCCTCCAACCAGTTGCCCCA

The reaction was performed by a 42°C 40 minutes long and then 70°C 15 minutes long incubation. RT-PCR products were loaded on agarose gel 1,5% and band intensity was detected by ImageJ software.

To test the *in vitro* spontaneous differentiation potential of iPSC, cells were grown in suspension to generate Embryo Bodies (EBs) as described below. After 7 days EBs were plated on Matrigel-coated plates, grown in E8 and allowed to spontaneously differentiate. Expression of thethree germ layer specific markers was evaluated by immunofluorescence by usong the following primary antibodies: beta3-tubulin (Abcam), Desmin (Chemicon) and alpha-fetoprotein (Invitrogen).

3.3. iPSC-MN

3.3.1. iPSC differentiation into iPSC-MN

To obtain motor neurons iPSCs were differentiated as previously described by our group (97). In detail, iPSCs were grown in suspension to generate Embryo bodies (EBs) in HuES medium supplemented with single cell survival enhancer factors, such as fibroblast growth factor (FGF) 20 ng/mL (Peprotech), Rho-associated kinase (ROCK) inhibitor Y27632 20 μM (Selleckchem) until day 7. The third day neuralization was induced by addiction of 10 μM SB431542 and 0,2 μM LDN193189 (both from Stemgent), while the fourth day EBs were switched to a neural induction medium (DMEM/F12, 2 mM L-Glutamine, 0.1% penicillin, 0.1% streptomycin, 1% MEM NEAA, heparin (Sigma-Aldrich), N2 supplement (Thermo Fisher Scientific)), supplemented with 20 μM ROCK inhibitor, 0.4 μg/mL ascorbic acid (AA) (Sigma-Aldrich), 1 μM retinoic acid (RA) (Sigma-Aldrich) and brain-derived neurotrophic factor 10 ng/mL (BDNF) (Peprotech). Starting from day 7, EBs were supplemented with smoothened agonist 1 μM(SAG) (Merck) and purmorphamine (Pur) 0.5 μM (Sigma-Aldrich), and cultured with these factors for ten days. EBs were then dissociated with 0.05% trypsin (Sigma-Aldrich) and plated on poly-lysine / laminin (Lam)-coated (Sigma- Aldrich and Thermo Fisher Scientific) in 24-well plates at a concentration of 5×10^5 cells/well, and cultured during the last 10 days in supplemented neural differentiation medium (Neurobasal medium (Thermo Fisher Scientific), 2 mM L-Glutamine, 0.1% penicillin, 0.1% streptomycin, 1% MEM NEAA and N2 supplement) supplemented with B27 (Thermo Fisher Scientific), 1 μg/mL Lam, 25 μM glutamate (Glu) (Sigma-Aldrich), 0.4 μg/mL AA, 10 ng/mL glial-derived neurotrophic factor (GDNF) and 10 ng/mL ciliary neurotrophic factor (CNTF) (both from Peprotech). At day 27 iPSC-

MNs were fixed for RNA FISH and immunofluorescence or collected for DNA and RNA extraction.
For each iPSC clone motor neuron differentiation step was performed three times.

3.3.2. Immunofluorescence

For immunofluorescence analyses iPSC-MNs were fixed as previously described. To characterize motor neuronal differentiation the following primary antibodies were used: βIII tubulin (Abcam, Cambridge, UK EP1569Y), MNR2/HB9/Mnx1 (DSHB Iowa City, IA, USA 81.5C10), SMI312 (Covance, Princeton, NJ, USA).

3.4. DNA Processing

3.4.1. DNA Extraction

The DNA derived from fibroblasts, iPSCs and iPSC-MNs was isolated using the Wizard® Genomic DNA Purification Kit (Promega). The kit provides a protocol for DNA isolation proceeding with fresh cell pellets from the cultured dish, but cell pellets were cryopreserved at -80°C as needed up to 1 year before DNA extraction.
At the moment of DNA extraction, the pellet was left at room temperature for 10 minutes, the Nuclei Lysis Buffer was added directly and mixed by pipetting. As soon as the pellet was dissolved, 3uL of RNase A were added and mixed by inversion and the tube was incubated at 37°C for 30 minutes. Protein lysis solution was then added at room temperature, the sample was briefly mixed by vortex and put on ice for 5 minutes. The tube was then centrifugated. The

supernatant containing DNA was added to 100% isopropanol and gently inverted. The tube was centrifugated for 5 minutes at 13000 rpm, the supernatant was discarded and replaced with 70% ethanol to clean the DNA from cell debris and inverted. The tube was centrifugated, the supernatant was discarded and the tube containing a pellet of DNA was air dried for 15-20 minutes before resuspended in Rehydratation Solution.

3.4.2. Bisulfite Sequencing

Bisulfite Sequencing (BS-Seq) is a method of DNA sequencing after treatment of genomic DNA with sodium bisulfite ($NaHSO_3$) and Polymerase Chain Reaction (PCR) of the region of interest. It represents the “gold standard” for qualitative analysis of methylation of more than one contiguous cytosines with a single CpG resolution.

3.4.2.1. Bisulfite conversion

Bisulfite conversion was performed using the EZ DNA Methylation kit (Zymo Research). Conversion efficiency was attested higher than 99%. 400-600 ng of genomic DNA (extracted from cell pellet as previously reported) was diluted up to 45 µL with nuclease-free water, and M-dilution Buffer and incubated 15 minutes of 37°C CT Conversion Reagent containing resuspended sodium bisulfite were added and the sample was incubated following the thermic protocol 95°C for 30 seconds and 50°C for 60 minutes, for 16 cycles in a Verity thermocycler. After the incubation samples were put on ice for at least 10 minutes, then they were loaded into a Zymo-Spin IC Column with M-Binding Buffer and were centrifugated. Converted DNA was hold by

the silica matrix of the column while buffer was eluted. Afterwards M-desulphonation Buffer was added to each column and the columns were left at room temperature for 20 minutes. After desulphonation the samples were centrifugated and washed two times with M-Wash Buffer in order to remove desulphonation solution completely. After that, DNA was eluted in M-Elution Buffer from the column by centrifugation and was dosed with spectrophotometer by measuring the 260 nm adsorbance as ssDNA, since once converted the two strands are no longer complementary.

3.4.2.2. Semi-nested PCR of *C9ORF72* promoter

After conversion the amplification of the region of interest was performed with a seminested PCR.

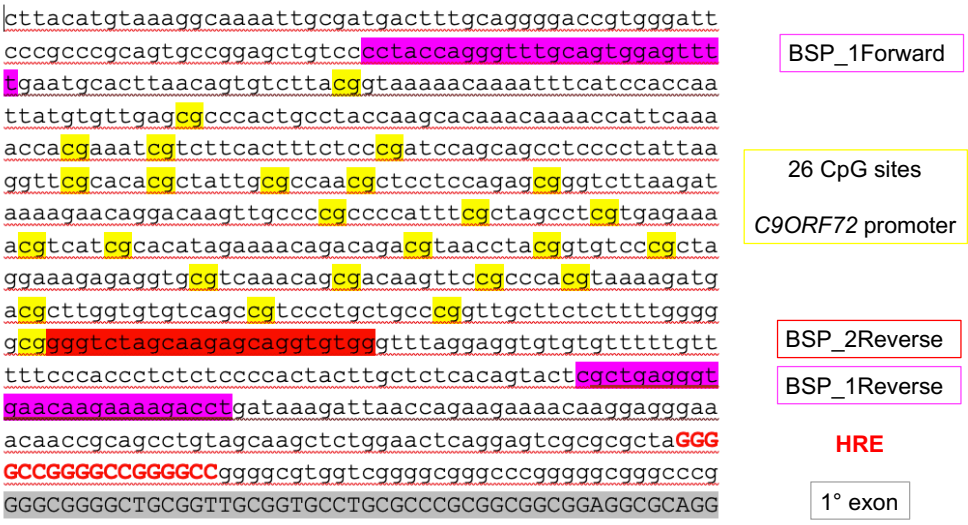


Figure 3-1. Sequence of *C9ORF72* promoter. On the nucleotide sequence the semi-nested PCR oligonucleotides (BSP_1Forward; BSP_1Reverse; BSP_2Reverse), the 26 CpG sites analyzed, the repeat expansion (HRE) location and the beginning of the 1st exon are indicated.

The process including two consecutive PCRs where in the second one the Reverse primer is substituted by another pairing with a sequence that is internal to the region amplified during the first PCR (Figure B). A qualitative analysis on 26 CpG sites of the CpG island at 5' of *C9ORF72* gene was thus performed. In order to obtain a specific amplification product a region of 448 bp was first amplified by PCR using primers for converted DNA. Then a 1:10 of the template DNA from the first PCR was amplified by PCR using the same BSP_1Forward primer and the more internal BSP_2Reverse, obtaining a 431 bp fragment. The mixtures for both the reactions were prepared in a final volume of 25 µL with: 10x Buffer+$MgCl_2$ (Roche), 10 mM dNTPs Mix (Invitrogen), 10 µM primers, 5 U/µL FastStart DNA Taq Polymerase (Promega) and 50-60 µg of template bisulfite converted DNA. In order to avoid no-specific annealing, thermal protocol was designed with Touchdown method, as the annealing temperature of the first 12 cycles was lowered by 0,5°C at each cycle.

Primers for semi-nested PCR *C9ORF72* promoter methylation	
BSP_1Forward	TTTATTAGGGTTTGTAGTGGAGTTTT
BSP_1Reverse	AAATCTTTTCTTATTCACCCTCAAC
BSP_2Reverse	CCACACCTACTCTTACTAAACCC

After the semi-nested PCR, the product of the second reaction was detected by electrophoretic run in a 1,5% agarose gel.

3.4.2.3. Sanger Sequencing

The semi-nested PCR product was purified using Illustra Exostar (GE Healthcare) method. Then, the sequencing kit Big Dye Terminator v3.1 Cycle (Applied Biosystems) was utilized for the sequence reaction. The 20 µL mixture was composed of 10x Buffer, 1x Big-Dye Terminator Solution, 3,2 µM primer BSP_ 1Forward, purified PCR and water up to the final volume. The thermal protocol was: 95°C for 3 minutes; 25 cycles of 95°C for 30 seconds, 54°C for 30 seconds and 60°C for 2 minutes. Fragments derived from the sequencing reaction were then purified by precipitation with 3M sodium acetate and 100% ethanol through the following protocol. For each sample a 2,5x volume of 100% ethanol and resuspended in 15 uL formamide. The purified sequence reaction was finally analyzed by capillary electrophoresis performed on the automatic sequencer Applied Biosystems 3500 Genetic Analyzer. The instrument allows to visualize the single nucleotide-resolution sequence of the peaks of a graph, the electropherogram, whose peaks correspond to the DNA nucleotides. After sodium bisulfite-conversion and semi-nested PCR amplification, unconverted methylated cytosines are still visualized on the electropherogram as cytosines, while converted unmethylated cytosines become thymine and the epigenetic trait can be easily distinguished.

3.4.3. Quantitative Mass Spectrometry

Matrix Assisted Laser Desorption Ionization-Time of Light (MALDI-TOF) mass spectrometry assay is one of the high-throughput methods that allow single-base DNA methylation quantitative detection (98,99).

500 ng of genomic DNA were bisulfite converted by EZ DNA Methylation kit (Zymo Research) as described above and amplified by a T7-promoter-tagged PCR. The PCR was then followed by shrimp alkaline phosphatase treatment (0.3U at 37°C for 20 minutes, then at 85°C for 5 minutes). Single-strand RNA transcription with 20U of T7 DNA polymerase and a special dNTPs mix (2,5 mM dCTP, 1 mM rUTP, rGTP and rATP and 3.14 mM DTT) was performed at the same time of cleavage by RNase A (0.09 mg/mL) at rUTP or rCTP at 37°C for 3 hours. The pool of RNA fragments of different length and of different mass that was obtained was then loaded for the MassARRAY analysis. The signal-to-noise value of the fragments means the abundance of fragments that were cut due to the bisulfite conversion of C into U in the region of interest. The masses of fragments were attributed through the MALDI-TOF-MS analysis and the MassCLEAVE software to a quantitative methylation profile of the template DNA. Partial methylation was detected as a percentage by virtue of different bars reading fully methylated, partially methylated and unmethylated fragments of the same mass.

3.4.4. *C9ORF72* HRE Methylation Assay

In order to determine the methylation state of the repeat expansion in *C9ORF72* gene, we used an already described protocol (63), that combines the methylation-sensitive PCR on bisulfite-converted DNA to the Repeat Prime PCR (rp-PCR) method and that was applied also for the detection of the methylation of the CGG repeat expansion in Fragile X Syndrome. 600ug of genomic DNA were initially converted by sodium bisulfite to change every unmethylated cytosine to uracil,

using the EZ DNA Methylation kit (Zymo Research) as was described above. During conversion uracils were replaced by thymines, the methylated $(GGGGCC)_n$ repeats were converted to $(GGGGTC)_n$ and unmethylated to $(GGGGTT)_n$. Specific primers were designed for methylated and unmethylated DNA, that were differently labeled with Fluorescein (FAM) for the methylated DNA and HEX fluorophore for the unmethylated DNA. Moreover, a reverse primer amplifying from multiple priming sites within the G_4C_2 repeated sequence and an anchor primer to enhance PCR efficiency were used.

Primers for *C9ORF72* $(G_4C_2)_n$ Methylation Assay	
FAM forward	TGTAAAACGACGGCCAGTAGTTTT GGAATTTAGGAGTCGC
HEX forward	TGTAAAACGACGGCCAGTAGTAA GTTTTGGAATTTAGGAGTTGTG
Anchor reverse	CAGGAAACAGCTATGACC
Methylated reverse 1	CAGGAAACAGCTATGACCGAACC CGCCCCGACCACGCCCCGACCC
Methylated reverse 2	CAGGAAACAGCTATGACCGAACC CGCCCCGACCACGCCCCGACCC
Unmethylated reverse 1	CAGGAAACAGCTATGACCAAACC CACCCCAACCACACCCCAACCC
Unmethylated reverse 2	CAGGAAACAGCTATGACCAAACC CACCCCAACCACACCCCAACCC

The mix for the methylation-sensitive rp-PCR was prepared with 10X MgCl2 Buffer (Roche), 5X GC-Rich solution (Roche), deaza GTP/dNTPs (Roche), 100 uM primers FAM, HEX, Anchor reverse,

methylated and unmethylated reverse 1, methylated and unmethylated reverse 2, and FastStart Taq DNA Polymerase (Roche), in a final volume of 50 uL. 6 uL of converted ssDNA were added to each (methylated and unmethylated) mix for parallel amplifications. Both the mixes for methylated and unmethylated amplification were performed in a Veriti 96 Well Termal Cycler (Applied Biosystems) as follows: 13 minutes at 96.5°C; 1 minute at 96.5°C (ramp rate 65%), 1 minute at 70°C (ΔTemperature of -1°C per cycle; ramp rate 25%), 3 minutes at 72°C (ramp rate 65%), 1 minute at 96.5°C (ramp rate 65%), 1 minute at 70°C (ΔTemperature of -1°C per cycle; ramp rate 25%), 3 minutes at 72°C (ramp rate 65%) for 8 cycles; 1 minute at 96.5°C (ramp rate 65%), 1 minute at 61°C (ΔTemperature of -2°C per cycle; ramp rate 25%), 3 minutes at 72°C (ramp rate 65%), 1 minute at 96.5°C (ramp rate 65%), 1 minute at 61°C (ΔTemperature of -2°C per cycle; ramp rate 25%), 3 minutes at 72°C (ramp rate 65%) for 3 cycles; 1 minute at 96.5°C (ramp rate 65%), 1 minute at 55°C (ΔTemperature of -1°C per cycle; ramp rate 25%), 3 minutes at 72°C (ramp rate 65%) for 35 cycles; 10 minutes of elongation at 72°C.

5uL of the so amplified products were added to 15,4 uL Formamide and 0,6 uL of ROX and loaded into ABI DNA analyzer, where a peak ladder of 6 bp increments in amplicon length corresponded to the hexanucleotide repeats. Since the FAM primer was read in blue channel and HEX in green channel, if the G_4C_2 repeat was methylated in all DNA copies, only the blue channel was expected to have signal; and if the G_4C_2-repeat is completely unmethylated, only the green channel is expected to have signal. However, if the G_4C_2-repeat is methylated in some copies and unmethylated in others, both channels

will have products. For large repeat expansions, the peak ladder gradually descends without a clear stop. For small repeat alleles (2-23 repeats), however, the peak ladder was relatively short and ended with a sudden stop.

3.4.5. Hydroxymethylation assay

For quantification of hydroxymethyl-cytosines (5-hmC) relatively to methyl-cytosines (5-mC) levels of the 25th CpG site of *C9ORF72* 5' flanking region the Epimark 5-hmC and 5-mC Analysis Kit (E3317 New England Biolabs) was used. 2,5 ug of genomic DNA were mixed with 80 μM UDP-Glucose, 1X NEBuffer 4 and Nuclease-free water up to 155 μL total volume. Then half volume was glucosylated by 1.5 μL of T4-beta glycosyltransferase (T4-BGT) for 12-18 hours (o/n) incubation at 37°C. During this step only 5-hmC were converted to glycosylmethyl-cytosine (5-gmC), while 5-mC remained unmodified. Both the glycosylated and unglycosylated samples were aliquoted into three tubes and 50 U MspI or 25 U HpaII or no restriction enzyme were, respectively, added to each tube and incubated for 8 hours at 37°C. After incubation 0.5 μL Proteinase K (Molecular Biology Grade) were added and incubated at 40°C for 30 minutes and at 95°C for 10 minutes for deactivation. Uncut fragments only at the 25th CpG site, corresponding to a restriction CCGG site for MspI and HpaII, were then amplified by PCR using the primer pairs for *C9ORF72* promoter 25° CpG (+104 bp from the transcription starting site) and bands were quantified by ImageJ software.

Primers for *C9ORF72* promoter Hydroxymethylation Epimark Kit	
Forward	CCAGCAGCCTCCCCTATTAAGG
Reverse	AGGTCTTTTCTTGTTCACCCTCAG

The PCR protocol included 4 minutes at 95°C; 30 seconds at 95°C, 30 seconds at 67°C, 30 seconds at 72°C for 29 cycles; then 5 minutes at 72°C for elongation. The uncut fragments were therefore amplified by PCR. MspI cuts the methylated and not the hydroxymethylated glycosylated site. HpaII cuts neither the hydroxymethylated and methylated site. All the bands were normalized on the band into the untreated tube, corresponding to the total DNA.

3.4.6. Southern Blot

For SB analysis, the 466bp-long *C9ORF72* probe up-stream the repeat expansion was obtained with the following primers:

Primers for *C9ORF72* Southern Blot probe	
SB Forward	CTTTCTCCAGATCCAGCAGCCTCC
SB Reverse	CTGAGTTCCAGAGCTTGCTACAG

Digestion of a total amount of 9 ug of genomic DNA was performed with 50U of XbaI restriction enzyme and 1% of BSA for 16 hours at 37°C. The digested DNA and molecular weights (1Kb, IV and XV DNA markers) were then loaded on a 0.7% agarose gel in 1X TBE buffer with ethidium bromide and the gel was run at 70 V for the first 2 hours, 40 V 15 hours and 90-100V 5-6 hours, keeping ice blocks around the

tank. The gel was stopped when the wild type band 2000 bps was about 12 cm far from the starting gel front. The gel was then gently shaken in 0.25 N HCl for 20 minutes, in order to break the DNA into smaller fragments and to enhance the transfer from the gel onto the hybridization membrane. After two washes in ddH20, the gel was incubated and gently shaken in a 0.4 M NaOH and 0.6 M NaCl for 30 minutes at room temperature (RT) to make the DNA single-stranded and again washed twice in ddH20. A 1.5 M NaCl + 0.5 M Tris neutralization solution was added, and the gel put in shaker at RT for 30 minutes.

By capillary blotting, DNA was transferred overnight to a positively charged nylon membrane and cross-linking for 2 minutes in a Strata-link oven was performed by UV irradiation. Nylon membrane was incubated into the hybridization glass tube with 25ml of pre-hybridization solution (12.5 mL formamide, 6.25 mL SSC 20X, 1.25 mL Denhardt 100X, 1.25 mL SDS10%, 3.75 mL H_2O) at 42°C for about 6 hours. During pre-hybridization time, the ^{32}P-radioactive labeling of 25 ng of the *C9ORF72* probe was performed with αdCTP by using the Prime-a-Gene Labeling System (Promega) according to the manufacturer instructions. The reaction was stopped at 95-100°C for 2 minutes, put in ice and 5 uL EDTA 0.2 M was added. The labeled probe was purified with ProbeQuant G-50 Micro Columns (GE Healthcare) and the labeling was evaluated by β-Counter (Beckman). After the pre-hybridization of the membrane at 42°C for 6 hours, about 10 mL of the pre-hybridization solution from the glass tube was removed and 10 ug/µL of denatured salmon sperm and the ^{32}P-

radiolabeled probe were added to the glass tube with about 15 mL of the remaining pre-hybridization solution. Hybridization was carried out at 42°C overnight. The membrane was then vigorously shaken in 2X standard sodium citrate (SSC) with 0.1% sodium dodecyl sulfate (SDS) low stringency buffer for 20 minutes at RT. Subsequently incubation in the medium stringency buffer SSC 0.5X/SDS 0.1% (pre-warmed at 60°C) for 15 minutes was performed into a shaking incubator at 60˚C, followed by a brief wash in SSC1X. The membrane was exposed to an X-ray film in an X-ray cassette with intensifying screen and keep at -80˚C for at least 4 days before to develop the film. The size of the expansion of each $C9^{+}$ sample was calculated based on the bp number of the corresponding band of the markers and bp number was converted in number of repeats = (bp – number of bp of wt allele) / 6.

3.5. *C9ORF72* sense foci RNA FISH

Cell cultures were fixed with RNase-free home-made 4% paraformaldehyde. RNA FISH was performed with a 5'TYE 563 $(CCCCGG)_{2.5}$ fluorescent LNA probe (Exiqon, Inc) targeting (GGGGCC)n repeat sequence in RNA foci. The protocol includes cell permeabilization with 0,2% Triton-X100 (Sigma), three fast dry steps in 75-80-100% ethanol, a 1-hour long pre-incubation with the hybridization buffer (50% formamide, 2x SSC, 50 mM Sodium Phosphate Buffer pH 7.0 1M, 10% Dextran Sulfate in DEPC-treated nuclease-free water) at 66°C in a humidity chamber followed by an over-night hybridization with 40 nM fluorescent LNA probe. After

hybridization, samples were washed once in 2X SSC with 0.2% Tween 20 and four times in 0.1X SSC at 65°C to eliminate off-target signals, then dryed, counter-stained with DAPI for 10 minutes and mounted on glass slides using FluorSave (Merck). Images were acquired as confocal stacks at 100X magnification using a Nikon Eclipse Ti confocal microscope system equipped with a Nikon CFI Apo λ 100X oil objective. The images were thenprocessed by ImageJ software (100). Cell counts were made using the "Find Maxima" command after have the noise tolerance set on the negative ($C9^-$) control. RNA foci quantification was reported as two different parameters, such as percentage of foci-containing cells and foci number per cell (reported as average value among the cells that presented at least one focus) (45).

3.6. RNA processing

3.6.1. RNA extraction

Total RNA was isolated from fresh or frozen cell pellets using trizol (Ambion)-chloroform phase separation. Cell pellet was incubated on ice for 10 minutes before adding 800 µL of Trizol reagent. The tube was then left 5 minutes at room temperature, then 160 µL of chlorophorm were added, vortexed for 15 seconds, incubated at room temperature for 5 minutes and centrifuged at 12000 g for 20 minutes at 4°C. The soluble phase was transferred to a new tube containing 400 µL of 100% isopropanol and 1 µg of glycogen for precipitation. After o/n incubation at –20°C, the tube was centrifuged at 12000 g for 30 minutes at 4°C, the pellet washed with 4°C 70% ethanol and

resuspended in nuclease-free water. RNA concentration was measured at Nanodrop Spectrophotometer and its purity was determined through A260/A280 (> 1,8) and A260/A230 (> 2) ratios.

3.6.2. Reverse transcription

One µg of total RNA was first incubated with 1U of DNaseI (Roche) and RNase-inhibitor in 15 µL volume at 37°C for 20 minutes and then at 94°C for 2 minutes to inactivate DNaseI. A15 µL mixture of 5x First Strand Buffer, 3 µL of 0.1M DTT (Invitrogen), 1 µL of Random Primers (Promega), 200 µM dNTPs (Invitrogen), RNase inhibitor was added and incubated at room temperature for 10 minutes in order to allow random primers to anneal. Then 1U Super Script II RT (Invitrogen) and 0.5 µM oligodT were added and the reaction was incubated at 42°C for 40 minutes and at 70°C for 15 minutes. Each sample was then diluted to 40 µL for Real-Time PCR.

3.6.3. Real-time PCR

Real-time PCR was performed using 0.3 µM forward and reverse primers, 2x SYBR Green PCR Master Mix (Applied Biosystems), 2.5 µL of reverse-transcribed cDNA (corresponding to 25 ng) and water up to 25 µL volume. Primer pairs to amplify total *C9ORF72* were designed in a 75 bp region that is common to all RNA isoforms between exons 2 and 3. We also designed specific primer pairs to amplify the three different RNA isoforms of *C9ORF72.* We used Rpl10 as housekeeping gene.

Primers for *C9ORF72* RealTime PCR		
Total Forward		ACACATATAATCCGGAAAGGAAGAAT
Total Reverse		TTCTAAGATAATCTTCTGGACATTTTCTTG
V1 Forward	147 bp	TCATCTATGAAATCACACAGTGTTC
V1 Reverse		GGTATCTGCTTCATCCAGCTT
V2 Forward	76 bp	GCGGTGGCGAGTGGATAT
V2 Reverse		TGGGCAAAGAGTCGACATCA
V3 Forward	86 bp	ACACATATAATCCGGAAAGGAAGAAT
V3 Reverse		As V2 Reverse
Rpl10 Forward		GAAGAAGGTGTTATGTCTGG
Rpl10 Reverse		TCTGTCATCTTCACGTGAC

The amplification protocol included: 2 minutes at 50°C, 10 minutes at 95°C and 45 cycles of 15 seconds at 95°C and 1 minute at 60°C, using the QuantStudio 12K Flex instrument. The specificity of the amplification reaction was checked by analysis of a melting curve at the end of each PCR reaction. Cycle threshold (Ct) were obtained and analysis of data was performed by calculating ΔCt for each sample between the target and the housekeeping Rpl10 gene, and comparing $C9^+$ samples with $C9^-$ control ($\Delta\Delta$Ct) and then obtaining a fold change (FC) value by the formule $2^{(-\Delta\Delta Ct)}$.

3.7. Statistical analysis

Statistical analyses were conducted with PRISM 7.0 (GraphPad) software. Results were considered significative when p-value was <0.05. Student's t-test and one-way ANOVA analyses of variance were used to compare two or multiple groups, respectively. Pearson

correlations were computed while comparing the co-variance of two datasets, assuming a Gaussian distribution from the examined repeated data. The statistical tests that were applied for each experiment and sample replicates are indicated in each figure.

4. Results

4.1. Analysis of *C9ORF72* promoter methylation in C9^{+} fibroblasts

In our Laboratory of Neuroscience methylation of *C9ORF72* promoter, which includes 26 CpG sites, had been already tested by bisulfite sequencing in a cohort of 55 mutant *C9ORF72* (C9^{+}) blood DNA samples. A proportion of methylated (30,9%) versus unmethylated C9^{+} samples was reported in our ALS C9^{+} cohort, similarly to what previously reported in literature (62,66). As the corresponding fibroblasts were also available for some of the C9^{+} patients of this cohort, three different C9^{+} fibroblast lines were selected in our biobank according to the *C9ORF72* promoter methylation state in blood genomic DNA.

The *C9ORF72* promoter methylation status was therefore determined in these C9^{+} fibroblasts and it was consistent with blood data. C9#1 and C9#2 fibroblasts were found methylated at all 26 CpG sites of *C9ORF72* promoter, while C9#3 line was indeed all unmethylated (Table 1).

Sample	Clinical phenotype	Mutation	Site of onset	Age of onset	Disease duration (months)	Sex	Blood Promoter methylation (methyl.CpG/totCpG	Fibroblast Promoter methylation (methyl.CpG/totCpG)
C9#1	ALS	*C9ORF72*	Spinal	62	58	M	26/26	26/26
C9#2	ALS	*C9ORF72*	Bulbar	51	18	F	26/26	26/26
C9#3	ALS	*C9ORF72*	Spinal	47	12	F	0/26	0/26

Table 1. Clinical features of the selected C9+ patients whose fibroblasts were available in our laboratory. Promoter methylation state (26 CpG sites) of *C9ORF72* in peripheral blood mononuclear cells and fibroblasts.

Aim of our study work was to study *C9ORF72* promoter methylation in iPSC and iPSC-derived motor neurons (iPSC-MNs) as a possible epigenetic modifier of *C9ORF72*-related pathology, such as RNA foci formation *in vitro.*

To address the issue, iPSC lines were generated in our lab from the three selected $C9^{+}$ fibroblasts and I analyzed longitudinally (10, 20, 30 and 40 passages *in vitro*) the *C9ORF72* promoter methylation state in the obtained iPSC and in differentiated iPSC-MNs (Figure 1).

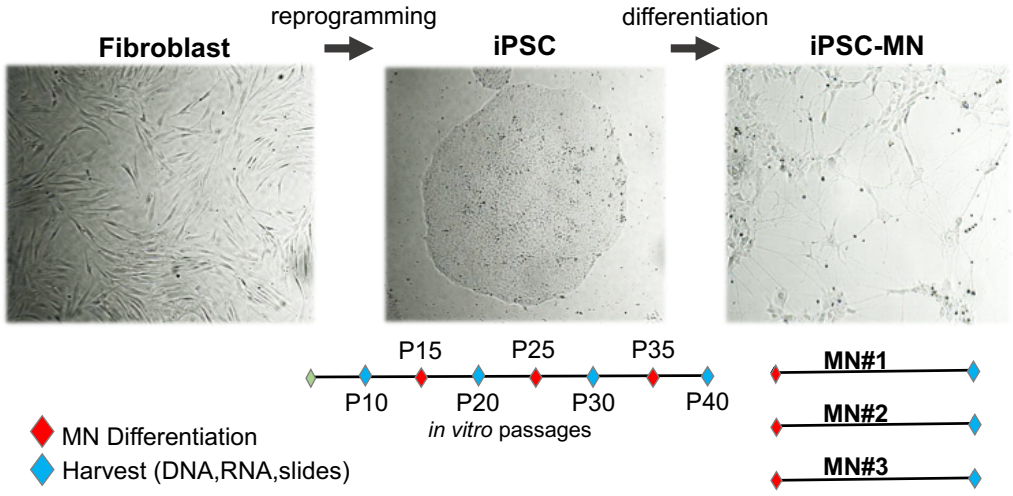


Figure 4-1. Schematic illustration of the workflow and the experimental procedures: skin fibroblasts from *C9ORF72*-mutated patients and controls were reprogrammed into iPSCs, that were cultured longitudinally and harvested at the indicated passages and differentiated for three times into iPSC-motor neurons (iPSC-MNs). The harvested iPSCs and iPSC-MNs were processed for DNA and RNA extraction and for FISH analysis to study *C9ORF72*-related pathology.

4.2. Establishment of $C9^+$ iPSC lines and differentiation into iPSC-MN

iPSCs were generated by reprogramming of the three selected $C9^+$ patient's fibroblasts (Table 1) with the four Yamanaka's factors Klf4, Oct4, Sox2 and c-Myc using the CytoTune-iPS 2.0 Sendai Reprogramming Kit as previously described (97) (Figure 2a). The efficiency of iPSC reprogramming from fibroblasts ranged between 0,01-1% of the total amount of cells. Six different clones were selected and expanded. Nevertheless, some clones were lost during early expansion and just a few of them survived, becoming stable iPSC lines that we later characterized and used.

After reprogramming indeed, we were able to obtain and expand two iPSC clones from C9#1 line, three clones from C9#2 and only one clone from C9#3 fibroblasts. Two $C9^-$ iPSC lines from healthy donors that were already available in our laboratory were used as negative controls in our subsequent analyses.

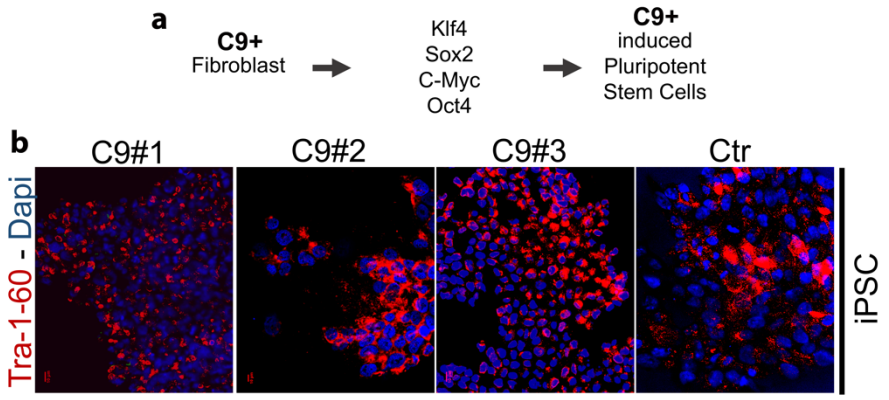


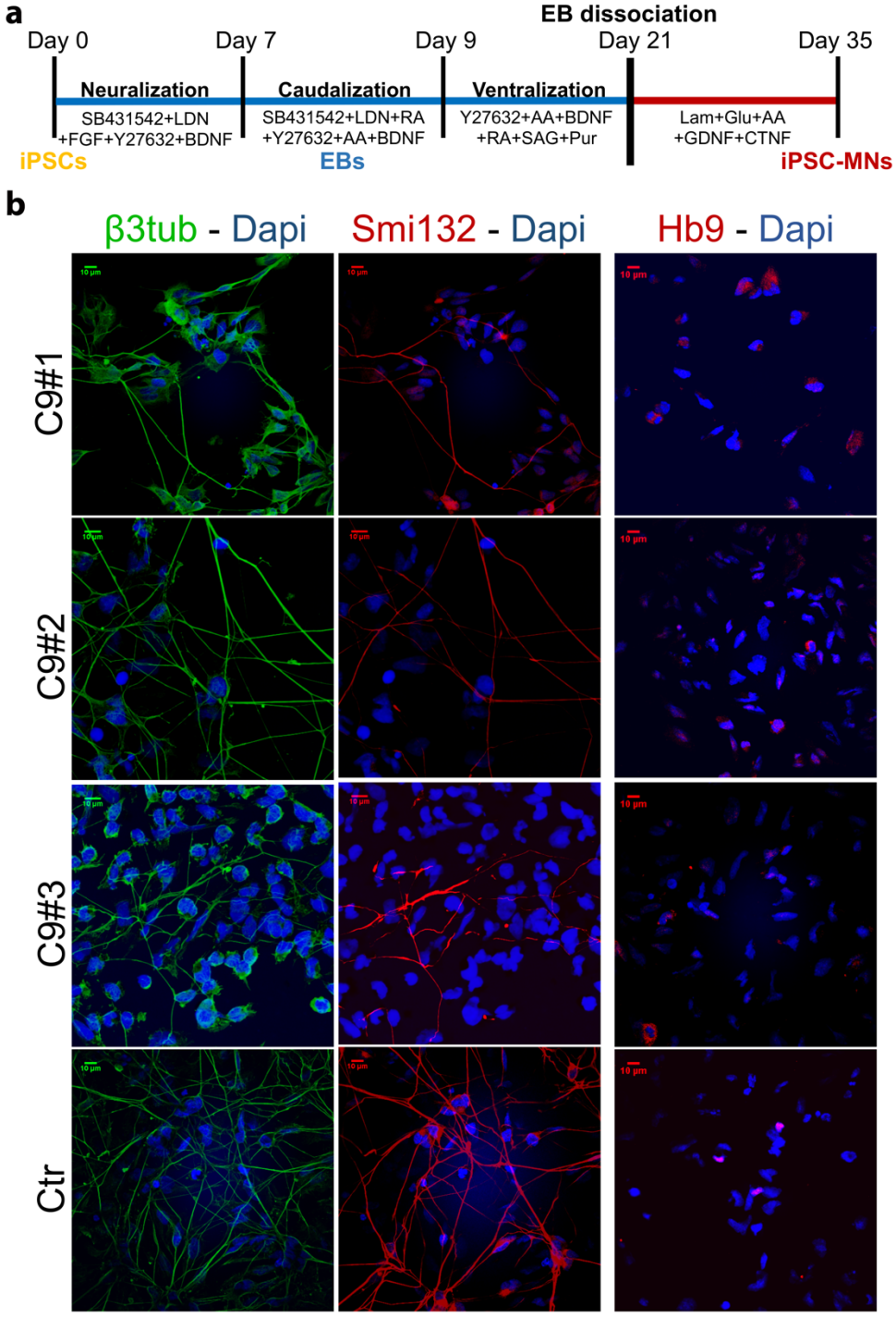
Figure 4-2. iPSC reprogramming and immunofluorescence characterization. a) Fibroblasts were reprogrammed using the Sendai virus (Thermo Fisher Scientific) including the four classic Yamanaka factors: Kruppel Like Factor 4 (Klf4), SRY (sex determining region Y)-box 2 (Sox2), c-Myc and octamer-binding transcription factor 4 (Oct4) to obtain iPSC. b) Representative immunofluorescence image of each C9+ patient's iPSC line (C9#1, C9#2, C9#3) and of a *C9ORF72*-negative healthy control iPSC (Ctr). Stem cell marker TRA-1-60 (red); nuclear staining DAPI (blue).

The expression of different stemness markers including the pluripotency marker podocalyxin TRA-1-60 (Figure 2b), the stage-specific embryonic antigen 4 SSEA4 and alkaline phosphatase (data not shown) was assessed for each $C9^{+}$ iPSC clone by immunofluorescence assays. The expression of the pluripotency markers Oct 3/4, Sox 2 and Nanog was also tested by RT-PCR (data not shown). A further analysis of genomic integrity of the obtained iPSC lines was performed by CytoVision software (Leica) on at least 20 metaphases with a 400 bands resolution (in collaboration with the Laboratory of Cytogenetics, Istituto Auxologico Italiano). The karyotypes of the three $C9^{+}$ iPSC clones excluded any gross genomic translocation or rearrangement (data not shown). More recently, the

minimal quality control of iPSC was extended to mitochondria DNA analysis too (84,85).

Differentiation of iPSCs into motor neurons was obtained through a 5-week protocol that included a 3-week long embryo bodies step (Figure 3a) (97). iPSC cultures were differentiated into motor neurons at passages P15, P25 and P35 (Figure 1). The obtained iPSC-MNs were characterized by IF for the expression of neuronal markers such as β3-tubulin (100% cells) and the axonal neurofilament Smi-312 (80%) (Figure 3b). The efficiency of motor neuron differentiation, measured with the early MN marker Hb9, was about 30% both in $C9^+$ lines and controls, indicating that *C9ORF72* mutation does not affect MN differentiation. Moreover, we evaluated morphologically the formation of β3-tubulin-positive neurites and neural clusters (Figure 3b).

Figure 4-3. a) Schematic representation of the motor neuronal (MN) differentiation protocol of the C9+ iPSC lines through embryo bodies (EB). Neural differentiation factors (SB431542, LDN, Y27632, Retinoic Acid, Ascorbic Acid, SAG and Purmorphamine), and growth factors (FGF, BDNF) were added as indicatedfollowed by terminal differentiation on Lamin (Lam)-coated dishes with motor neuronal factors (GDNF, CTNF), Glutamine and Ascorbic Acid (AA). b) The obtained iPSC-MNs for each C9+ patient (C9#1, C9#2, C9#3) and one healthy control (Ctr) were stained for pan axonal anti-neurofilament (Smi-312, red), neuronal beta3-tubulin (green) and motor neuronal Hb9 (red). Merged images are shown with nuclear staining DAPI (blue).


a
EB dissociation
Day 0
Day 7
Day 9
Day 21
Day 35
Neuralization
SB431542+LDN
+FGF+Y27632+BDNF
Caudalization
SB431542+LDN+RA
+Y27632+AA+BDNF
Ventralization
Y27632+AA+BDNF
+RA+SAG+Pur
Lam+Glu+AA
+GDNF+CTNF
iPSCs
EBs
iPSC-MNs
b
β3tub - Dapi
Smi132 - Dapi
Hb9 - Dapi
C9#1
C9#2
C9#3
Ctr

4.3. Characterization of *C9ORF72* promoter methylation in C9$^+$ iPSC and iPSC-MN

Bisulfite sequencing (BS-Seq) was used to determine qualitatively the methylation state of *C9ORF72* promoter, which is composed of 26 CpG sites (Figure 4a). The methylation pattern of *C9ORF72* promoter was first determined in all the generated clones of the three different C9$^+$ iPSC lines both at very early passages *in vitro* (P<3) and at 5 passages (P5) in order to monitor possible epigenetic changes due to the reprogramming step itself (Figure 4b). The two C9#1 iPSC clones maintained the promoter in a methylated status as in blood and fibroblasts (Table 1) both at P<3 and at later passages P5 (Figure 4b). C9#2 lost promoter methylation compared to blood and to the original fibroblasts, in two out of the three iPSC clones, while the third clone showed a scattered methylation pattern. C9#3 switched from being unmethylated in fibroblasts (Table 1) to acquire a methylated *C9ORF72* promoter at P<3 and switching back to be unmethylated at P5 in the only clone that we obtained after reprogramming (Figure 4b). We decided to follow longitudinally, up to 40 passages *in vitro*, the DNA methylation state of *C9ORF72* promoter in one clone for each C9$^+$ iPSC line (clone#1 of each C9$^+$ iPSC line Figure 4b) in order to define the stability of this epigenetic mark over time in C9$^+$ iPSC cultures (Figure 1b). Although BS-Seq is a qualitative and not a quantitative assay, we selected those clones that were homogeneous in promoter methylation state at all the 26 CpG sites, in order to compare samples that were more representative of being completely unmethylated and methylated.

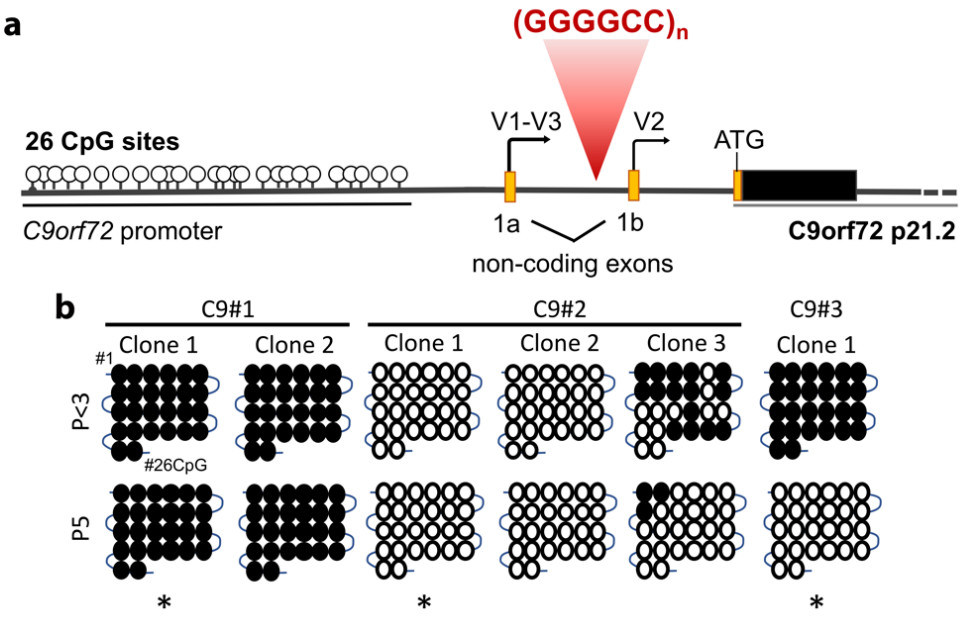


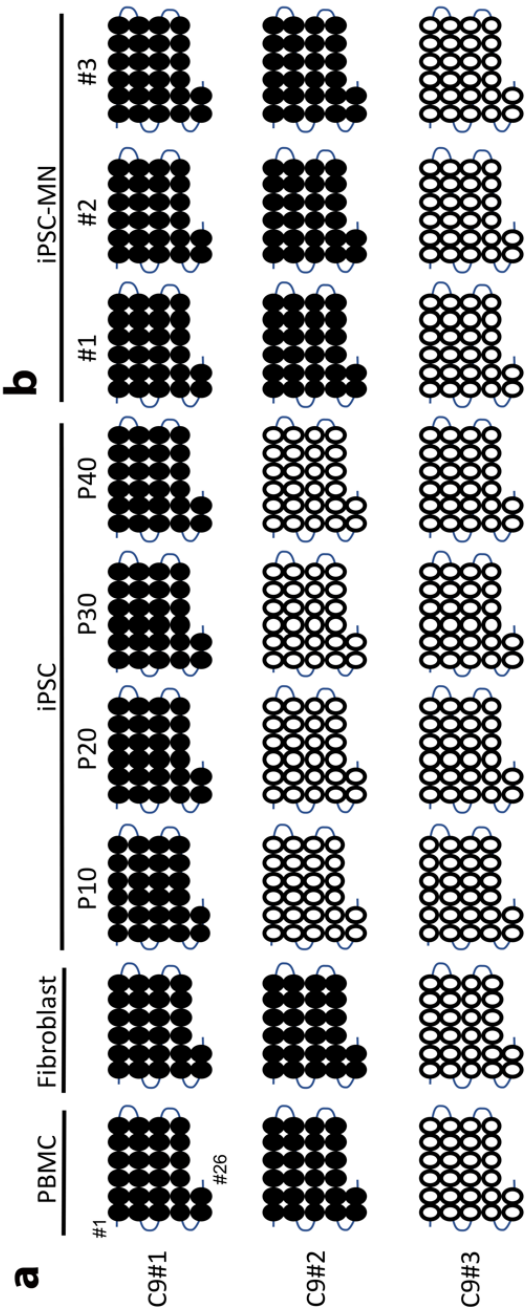
Figure 4-4. a) Scheme of gene promoter, 5'UTR exons (1a, 1b) and coding exon 1 of *C9ORF72* gene at p21.2 locus (MIM: 614260); white pins represent CpG sites of gene promoter; V1 (variant 1), V2 and V3 represent *C9ORF72* transcripts; (GGGGCC)n is the intronic repeat expansion, b) Representation of the methylation state of the single 26 CpG sites of *C9ORF72* promoter in the available C9+ iPSC clones of the three reprogrammed C9+ lines at a low number of passages (P<3) and at the passage 5 (P5) in vitro. Black circle epresents a methylated CpG site, white the unmethylated CpG site. * indicates the clones that were selected for further longitudinal studies.

BS-Seq analysis showed that *C9ORF72* promoter methylation state was stable in each C9^{+} iPSC line over passages *in vitro* at each CpG site by measurements at four different timepoints (P10, P20, P30, P40) (Figure 5a).

The three C9^{+} iPSC lines were then differentiated three times into iPSC-MNs, as described (Figures 1 and 3) and their promoter methylation was therefore determined. Our data showed that C9#1 and C9#3 iPSC-MNs maintained the same promoter methylation

status as iPSCs, while C9#2 iPSC-MN showed a completely methylated *C9ORF72* promoter compared to the unmethylated iPSCs they derive from (Figure 5b). Our results suggest that DNA epigenetic modifications at *C9ORF72* locus can change during both the fibroblast reprogramming into iPSC and the neuronal differentiation steps. Interestingly C9#2 methylation status is shown to change in all iPSC-MNs after the three differentiation experiments, and to return similar to the original fibroblasts before iPSC reprogramming (Figure 5b).

Figure 4-5. Promoter methylation status of the 26 CpG sites of *C9ORF72* promoter in the three C9+ iPSC lines that were a) analyzed at four timepoints at passages P10, P20, P30 and P40 and b) differentiated into motor neurons (iPSC-MNs) for three times (#1, #2 and #3). PBMC are the peripheral blood mononuclear cells of the patent. Black circle represents a methylated site and white unmethylated.



In order to validate BS-Seq data and to obtain a quantitative measure of DNA methylation, *C9ORF72* promoter was then analyzed by high-

throughput mass spectrometry (MassARRAY) assay in the methylated C9#1 iPSC and iPSC-MN and C9#2 iPSC-MN. This method indeed allows to quantify the percentage of DNA methylation at the single CpG sites of the target region but, in the case of *C9ORF72* promoter, the limit of this technique was to be able to analyze only 17 out of the 26 CpG sites. Regardless this limit (bias) the MassARRAY analysis showed that the single CpG sites were methylated from a minimum of 1% to a maximum of 74% in C9#1 iPSC and iPSC-MN while C9#2 iPSC-MN showed a lower degree of methylation at each analyzed CpG site compared to a healthy control iPSC (Figure 6a).

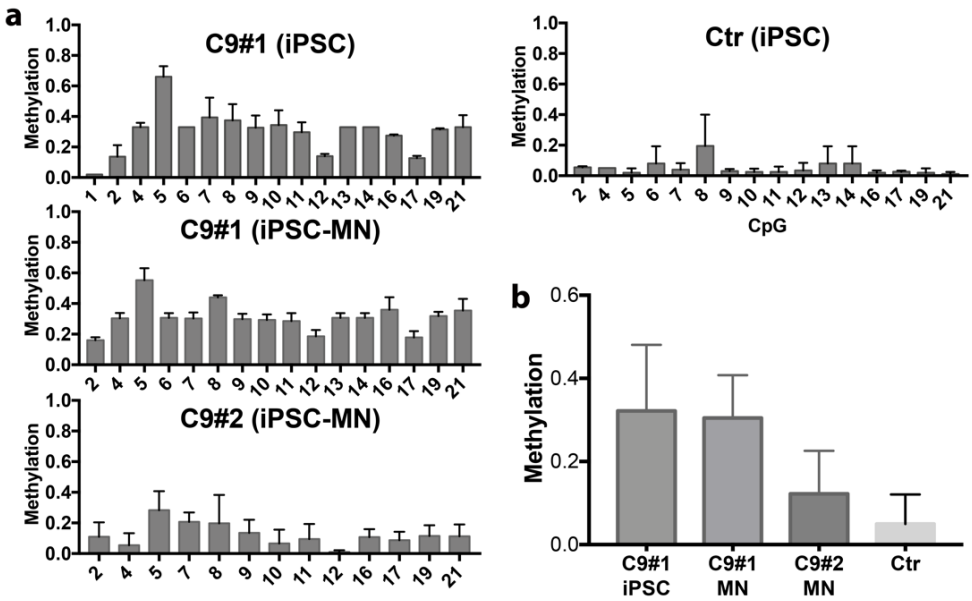


Figure 4-6. a) MassARRAY analyses of C9+ methylated samples (C9#1 iPSC and iPSC-MN, C9#2 iPSC-MN and one C9- control iPSC). Istograms showing the mean methylation at each CpG site (17/26) of *C9ORF72* promoter at three passages in vitro (P10, P20 and P30) for iPSC or differentiation steps for iPSC-MN. b) Total methylation values of *C9ORF72* promoter. (Mean+SEM; one-way ANOVA, n=3 different passages or differentiation steps.)

By considering the mean values of total methylation, C9#1 iPSC showed 33% of promoter methylation similar to the derived iPSC-MN (31%), while C9#2 iPSC-MN promoter was 13% methylated (Figure 6b), even if no significant differences were reported. Since *C9ORF72* promoter methylation was previously demonstrated to be monoallelic in $C9^+$ blood DNA (65) and we also confirmed this observation in our cohort of $C9^+$ blood samples (unpublished results), the quantitative MassARRAY data indicate that this condition is maintained also in $C9^+$ iPSC and iPSC-MN.

Sample	Age of onset	Disease duration (months)	Sex	Promoter methylation (methyl.CpG/totCpG)			
				Peripheral blood	Fibroblasts	iPSC	iPSC-MN
C9#1	62	58	M	26/26	26/26	26/26	26/26
C9#2	51	18	F	26/26	26/26	0/26	26/26
C9#3	47	12	F	0/26	0/26	0/26	0/26

Table 2. Promoter methylation state of *C9ORF72* gene in mononuclear blood cells, fibroblasts, iPSC and iPSC-MNs of the three $C9^+$ samples.

4.4. Relationship between *C9ORF72* promoter methylation and RNA foci formation

C9ORF72 V1 and V3 RNA isoforms originating from the 1a transcription starting site harbor the intronic HRE sequence (Figure 4a) and have the tendency to form very stable G-quadruplex RNA structures in all neuronal and non-neuronal cells of mutant $C9^+$

patients, that when are detected by RNA FISH are called RNA foci and are a hallmark of *C9ORF72* pathology. RNA foci represent a therapeutic target and were shown to decrease when the HRE was targeted by antisense oligonucleotides (ASO) (45) or by small molecules (94) in patient-derived iPSC-MNs.

To analyze RNA foci, we performed RNA FISH assay in all our C9+ iPSC and iPSC-MN at all the timepoints described (Figure 1), in order to assess if the methylation status of *C9ORF72* promoter may act as a modifier of RNA foci formation (Figure 7).

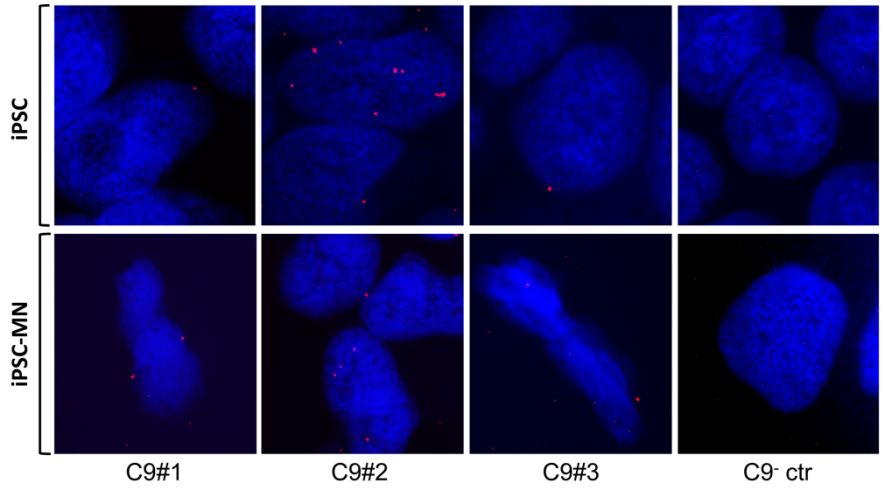
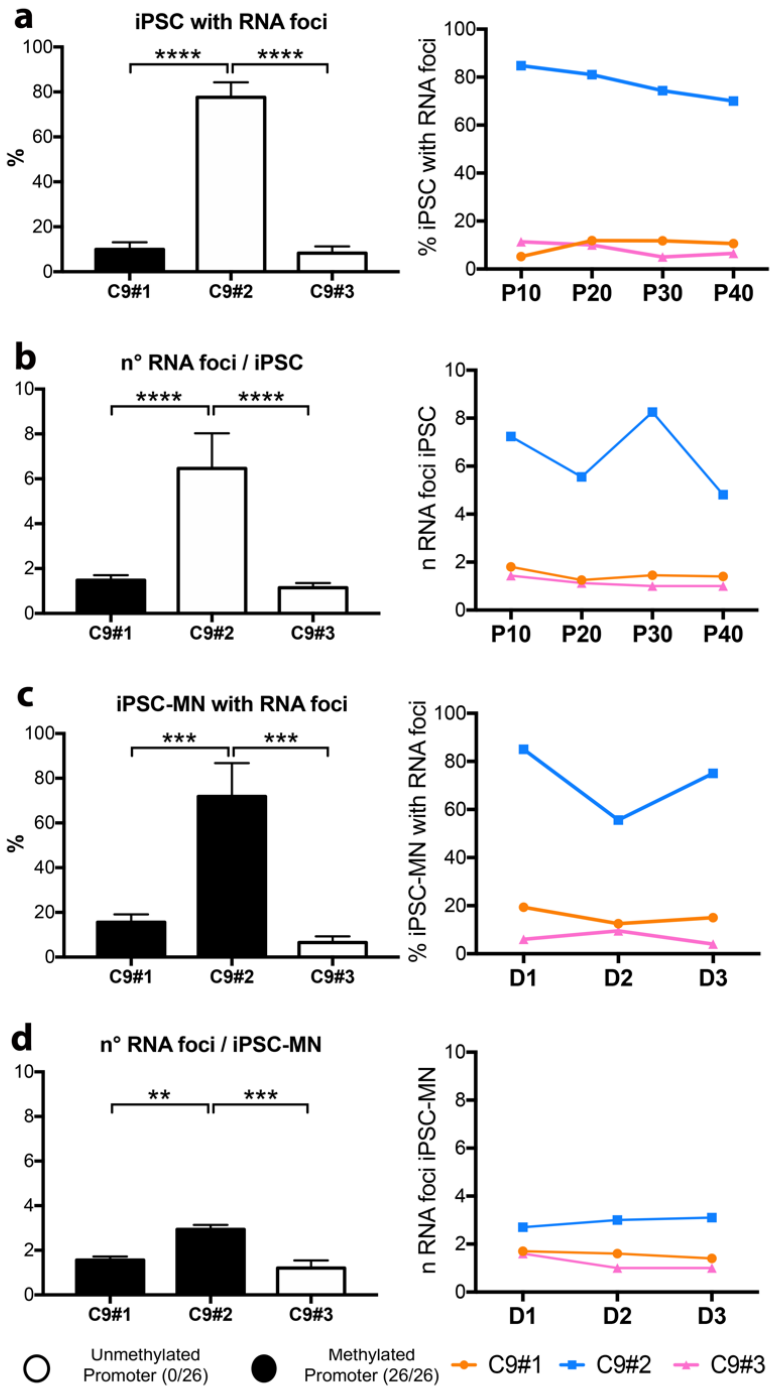


Figure 4-7. Representative images of RNA FISH showing *C9ORF72* RNA foci in C9+ iPSC and iPSC-MN. The TYE563-(GGGGCC)2,5 LNA probe (red) was used to detect sense RNA foci within the nucleus and the cytoplasm of iPSC (upper panel) and iPSC-MN (lower panel). A C9- control iPSC and iPSC-MN was used as a negative control. Nuclear staining is shown in blue (DAPI).

RNA foci quantification in iPSC by ImageJ analysis showed that methylated C9#1 iPSC and unmethylated C9#3 iPSC lines had 7,9% and 9,3% of cells presenting at least one focus, respectively, and 1,5 and 1,3 RNA foci per cell, while the unmethylated C9#2 iPSC showed

RNA foci in 81,5% of cells and 7,1 foci per cell (Figure 8a-b, left panel). The same patient-specific pattern of RNA foci formation was found in iPSC-MNs, where methylated C9#1 and unmethylated C9#3 showed few cells presenting RNA foci, 15,6% and 6,5% respectively, and a low number of RNA foci per cell (1,6 and 1,2), while the methylated C9#2 iPSC-MN showed RNA foci in a high percentage of cells (71,9%) and a higher number of RNA foci per cell (2,9) (Figure 8c-d, left panel). The experimental variability in RNA foci presentation was very limited for each sample both in iPSCs during the different passages *in vitro* (P10, P20, P30, P40) and in iPSC-MNs obtained after differentiation (n=3) (Figure 8, right panels).

Figure 4-8. Correlation between *C9ORF72* promoter methylation and RNA foci formation in three C9+ iPSC lines and the corresponding iPSC-MNs. *C9ORF72* promoter state is indicated with black bar (methylated) and white bar (unmethylated). Percentage of cells presenting RNA foci in a) iPSCs and c) iPSC-MNs (left) and distribution of data during replicate measures at each time-point (P10, P20, P30, P40) (right). Number of RNA foci in b) iPSCs and d) iPSC-MNs (left) and distribution of data during replicate measures at each MN differentiation (D1, D2, D3) (right). (one- way ANOVA and multiple comparisons test. **** p value<0.0001, *** p<0.001, ** p<0.01.


a
iPSC with RNA foci
****
****
%
C9#1
C9#2
C9#3
% iPSC with RNA foci
P10
P20
P30
P40
b
n° RNA foci / iPSC
****
****
C9#1
C9#2
C9#3
n RNA foci iPSC
P10
P20
P30
P40
c
iPSC-MN with RNA foci
***
***
%
C9#1
C9#2
C9#3
% iPSC-MN with RNA foci
D1
D2
D3
d
n° RNA foci / iPSC-MN
**
***
C9#1
C9#2
C9#3
n RNA foci iPSC-MN
D1
D2
D3
Unmethylated Promoter (0/26)
Methylated Promoter (26/26)
C9#1
C9#2
C9#3

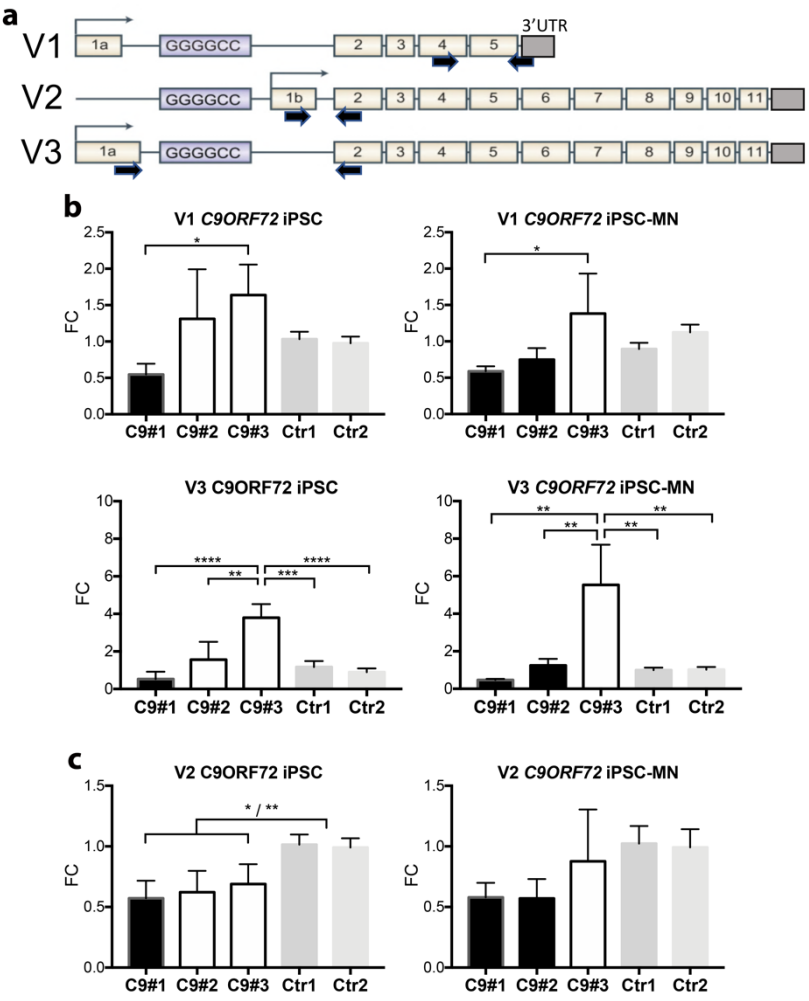


Figure 4-9. a) Scheme of he three different *C9ORF72* transcripts. 1a and 1b are the alternative TSS (transcription start sites). GGGGCC indicates RNA corresponding to the expanded repeat (HRE) in the first intron of transcript V1 and V3 and in promoter region of isoform V2. Boxes from 2 to 11 represent coding exons. Isoform-specific prmers are indicated by arrows. Q-PCR results of: b) the two HRE-containing isoforms V1 and V3 starting from 1a TSS and c) the main isoform V2, starting from 1b TSS. Black bars indicate a methylated promoter and white bars an unmethylated promoter. Gene expression data were normalized on RPL10 housekeeping gene and fold change (FC) (2^-(DDCt)) were referred to the C9- controls. (Mean+SEM; one-way ANOVA and multiple comparisons test on replicate measures at each time-point (P10, P20, P30, P40 and differentiation step (n=3). **** p value <0.0001, *** p<0.001, ** p<0.01, * p<0.05.

4.5. Correlation between *C9ORF72* promoter methylation and its gene expression

Transcription of *C9ORF72* is reported to be altered by the presence of the HRE, which induces the downregulation of V2 mRNA isoform, the most expressed *C9ORF72* transcript, starting from exon 1b downstream the HRE, and the upregulation of the V1 and V3 transcripts that start from exon 1a (38,65,70). Isoforms V1 and V3 harbor the HRE sequence in their pre-mRNAs that form RNA foci and are then translated by RAN translation into toxic DPRs. Since the promoter of *C9ORF72* gene is located upstream the transcription start site (TSS) of V1 and V3 transcripts, its methylation is likely to have an effect on the transcription of these two HRE-containing transcripts but less on the V2 transcript (Figure 4a).

Quantitative-PCR (qPCR) analysis was performed by designing specific primers for the three transcripts of *C9ORF72* V1, V2 and V3 (Figure 9a). As regards the HRE-bearing V1 and V3 transcripts, we could observe that methylated C9#1 is downregulated compared to unmethylated C9#3 both in iPSCs and in iPSC-MNs (Figure 9b). V3 transcript was upregulated in C9#3 also compared to controls and to C9#2, regardless of its switch in promoter methylation state from iPSC to iPSC-MN. In all the experimental conditions the HRE-containing transcripts were similar to controls when the promoter was methylated. In qPCR analyses C9#2 iPSC and iPSC-MN showed no significant difference compared to C9#1 and C9#3 as regards V1 transcript levels while it was different from unmethylated C9#3 in both cell models in V3 gene expression (Figure 9b), regardless its switch in promoter

methylation, suggesting that C9#2 intermediate V1-V3 expression could be due to the different level of promoter methylation from C9#1 detected by MassARRAY (Figure 6a).

The more abundant V2 isoform was found significantly downregulated in all C9^{+} iPSC, regardless of the methylation status of *C9ORF72* promoter, and showed a trend to decrease also in iPSC-MNs although without statistical significance (Figure 9c).

The qPCR analysis of total *C9ORF72* gene expression showed a decrease of the total *C9ORF72* transcript level only in the methylated C9#1 iPSC. In iPSC-MNs no significant differences were observed in gene expression between healthy controls and C9+ cells regardless their promoter methylation state (Figure 10).

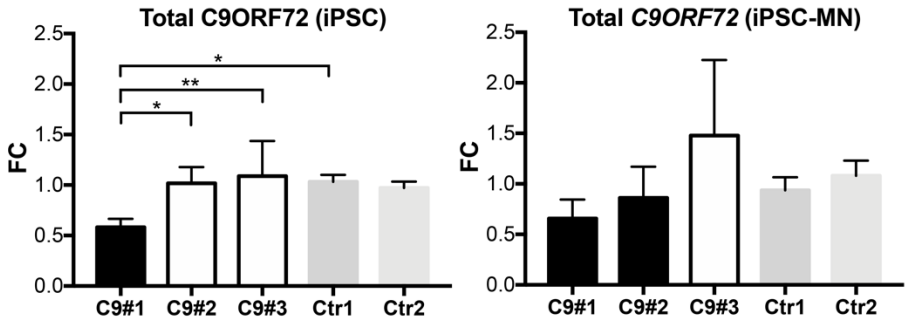


Figure 4-10. 10. Q-PCR results of total *C9ORF72* mRNA levels in iPSC and iPSC-MN. Black bars indicate methylated *C9ORF72* promoter and white bars the unmethylated samples. Gene expression data were normalized on RPL10 housekeeping gene and fold change (FC) values were referred to C9- healthy controls. (Mean+SEM; one-way ANOVA and multiple comparisons test on replicate measures (n=3 iPSC and n=3 iPSC-MN) for each cell line. ** $p<0.01$, * $p<0.05$.

4.6. Characterization of HRE methylation and promoter hydroxy-methylation as additional epigenetic modifiers of *C9ORF72* gene expression

Besides the *C9ORF72* 5' CpG island, the presence of the HRE $(GGGGCC)_n$ introduces a lot of novel clustered CpG sites in the first intron which are described to undergo methylation in different tissues from $C9^+$ patients (63). Since the HRE itself is located in the promoter region of the main V2 isoform, that starts from exon 1b, the methylation of the HRE could influence its transcription. We therefore analyzed the methylation of the HRE by a modified Repeat Primed PCR method as previously described (63) in all the $C9^+$ iPSC and iPSC-MN samples. This method is able to detect the methylation of only the first 35-45 CpG sites in the *C9ORF72* HRE since after this extent the resolution of the chromatograms dramatically decreases. We could observe that in all samples the HRE was always methylated, at least at the beginning of its sequence. The presence of this additional epigenetic modification at the *C9ORF72* locus help may explain why we did not observe any difference in V2 gene expression data among samples with or without methylation at *C9ORF72* promoter (Figure 9c).

We then investigated hydroxy-methylation of *C9ORF72* promoter as another possible confounder of gene expression and/or RNA foci formation. Hydroxy-methylation of cytosines is an epigenetic modification that, in contrast to methylation, seems to promote gene transcription. The bisulfite treatment used in BS-Seq and MassARRAY analyses does not allow to discriminate between the presence of

hydroxy-methylated and methylated cytosines. We therefore used a methylation-sensitive enzymatic assay (Epimark kit), that included the combined digestion of T4-beta glycosyltransferase-treated DNA with MspI and HpaII restriction enzymes and PCR amplification. However, this method, that is sequence dependent, allowed the analysis only of the 25th CpG site of *C9ORF72* promoter. The percentage of hydroxy-methylation (MspI digestion) was quantitatively referred to the total modified cytosines (HpaII digestion) amount, and the value was obtained from the ratio between MspI and HpaII cut amplicons.

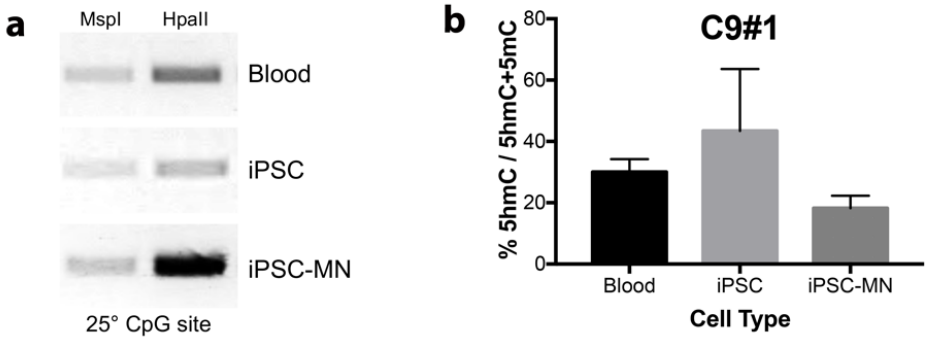


Figure 4-11. Hydroxymethylation assay by Epimark Kit. DNA is glucosylated and cut using glucosylation-sensitive MspI and methylation-sensitive HpaII a) Representative illustration of the Epimark hydroxymethylation assay results as electrophoresis bands after PCR. b) Histogram reports the % of hydroxymethyl-cytosine at #25 CpG site (as cut by MspI after glycosylation treatment) referrend to the amount of hydroxymethyl and methyl-cytosines (as cut by HpaII) of C9#1 sample in blood and fibroblast-derived in vitro models.

Our very preliminary data on the highly methylated C9#1 sample show that hydroxy-methylation is present at the 25th CpG site of *C9ORF72* promoter in blood, iPSC and iPSC-MNs (Figure 11a), although it represents a smaller proportion (18-43%) if compared to DNA

methylation (Figure 11b). However, a limit of this analysis is that we are not able to define the hydroxymethylation status of all the 26 CpG sites endowed in *C9ORF72* promoter.

4.7. Analysis of *C9ORF72* promoter methylation and RNA foci formation in an enlarged cohort of $C9^+$ iPSC lines

The data we obtained from the observational longitudinal study in three different $C9^+$ iPSCs and the corresponding differentiated iPSC-MNs seem to indicate that *C9ORF72* promoter methylation is associated to the regulation of gene expression, in particular of transcripts V1 and V3, while it seems to be not directly related to RNA foci formation.

In order to better assess if promoter methylation behaves as a modifier of RNA foci formation, we extended our analyses to additional $C9^+$ iPSC lines. We selected three additional $C9^+$ iPSC lines already reprogrammed in our laboratory (C9#4, C9#5 and C9#6) and two $C9^+$ commercially available iPSC lines (C9#7 and C9#8) from Cedar Sinai Biobank (West Hollywood, CA, USA). Clinical features of the enlarged $C9^+$ cohort are described in Table 2.

The BS-Seq assay showed that *C9ORF72* promoter was unmethylated in C9#4, C9#5 and C9#6 iPSC lines and methylated in C9#7 and C9#8 ones (Figure 12a). Since for the C9#7 iPSC line the corresponding $C9^-$ isogenic line (C9#7ISO) was also available, the promoter of C9#7ISO line was analyzed and found to be unmethylated, in line with hypothesis that *C9ORF72* promoter can be

hypermethylated only if the HRE is present and attesting that the deletion of the HRE by CRISPR/Cas9 on C9#7 caused also the loss of the $C9^+$-specific epigenetic feature.

We analyzed RNA foci formation by RNA FISH assay and quantitative image analysis on the $C9^+$ iPSC lines of the enlarged cohort.

Table 3. The 8 *C9ORF72* positive iPSC lines were obtained by mutated patient's fibroblasts. Skin fibroblast were withdrawn by excisional or punch biopsy. Clinical phenotype was apparent at skin biobsy but for C9#7. The *C9ORF72* expansion was detected according to diagnostic protocol of the centers of provenience. The presence of at least one ALS/FTLD in the family history was indicated as familiar ALS (FALS) otherwise the case is counted as Sporadic ALS (SALS). Disease duration is expressed as the number of months that elapsed from onset to death. Age at collection is the age at skin biopsy. Patients from Istituto Auxologico Italiano are guaranteed by informed consent for research purpose while cells from iPSC core of Cedars-Sinai are listed as CS52iALS-C9nxx (C9#7) and CS29iALS-C9nxx (C9#8)

iPSC Sample	Clinical phenotype	Mutation	FALS/ SALS	Cognitive impairment	Site of onset	Age of onset	Disease duration	Age at collection	Reprogrammed tissue	Sex	Provenience
C9#1	ALS	C9orf72+	SALS	-	Spinal	62	58	63	fibroblast	M	Auxologico
C9#2	ALS	C9orf72+	FALS	-	Bulbar	51	18	52	fibroblast	F	Auxologico
C9#3	ALS	C9orf72+	SALS	-	Spinal	47	12	48	fibroblast	F	Auxologico
C9#4	ALS	C9orf72+	na	-	Spinal	42	na	43	fibroblast	F	Auxologico
C9#5	ALS/FTD	C9orf72+	FALS	+	-	na	na	na	fibroblast	F	Auxologico
C9#6	ALS	C9orf72+	FALS	-	Bulbar	na	na	na	fibroblast	M	Auxologico
C9#7	ALS	C9orf72+	na	na	Spinal	57	na	49	fibroblast	M	Cedars-Sinai
C9#8	ALS	C9orf72+	na	na	Spinal	46	na	47	fibroblast	M	Cedars-Sinai

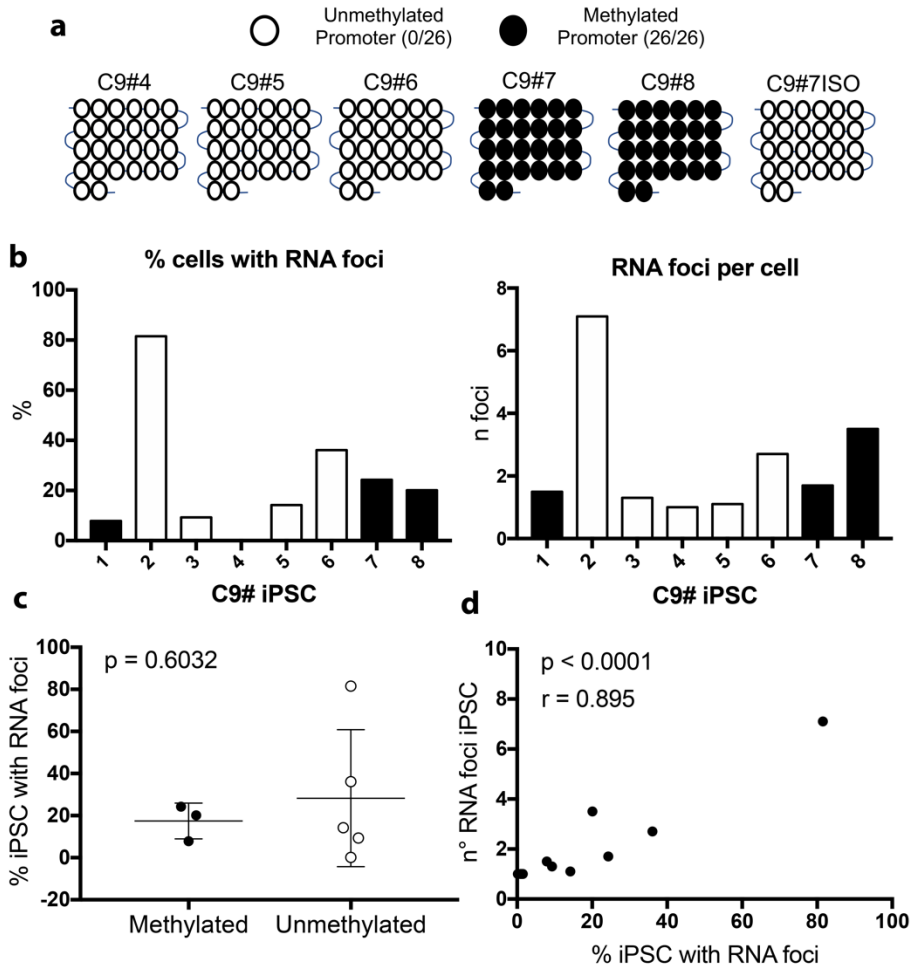


Figure 4-12. Correlation between *C9ORF72* promoter methylation and RNA foci formation in eight C9+ iPSC lines. (Unmethylation is represented as white dots/bars, methylation as black dots/bars.) a) Schematic representation of methylation state of the 26 CpG sites of *C9ORF72* promoter of the 5 novel C9+ iPSC lines and of the C9- isogenic control (C9#7ISO) iPSC line. b) Istograms showing RNA foci count in terms of percentage of cells containing foci and number of foci per cell as detected by RNA FISH quantitative image analysis. c) Correlation of n° of cells presenting RNA foci and *C9ORF72* promoter methylation state (n=1 for each C9+ iPSC line) (Unpaired t-test; p=0.6032). d) Correlation between RNA foci number and number of cells presenting RNA foci (Pearson test, r=0.895; p<0.0001).

Our data showed that both the percentage of cells with RNA foci and number of RNA foci per cell were similar in the methylated (C9#7 and C9#8) and unmethylated (C9#4, C9#5 and C9#6) iPSC groups including the C9#1, 2 and 3 iPSC lines (Figure 12b).By correlating the methylation status of *C9ORF72* promoter to RNA FISH data of the entire $C9^+$ iPSC cohort (C#1-8), we found no significant correlation between promoter methylation state and number of cells forming pathological RNA foci (Figure 12c).
When we considered the percentage of cells presenting RNA foci and the mean number of foci in the cells with at least one focus, we observed a significant positive correlation between these two parameters (Figure 12d).

4.8. Characterization of *C9ORF72* HRE length and possible correlation with RNA foci formation in $C9^+$ iPSC lines

As in mutant *C9ORF72* patients the HRE ranges from >30 to about 4,000 repeat units, which can differ among the different tissues, and as its length is considered a possible modifier factor of disease phenotype, as in other repeat expansion disorders, we wondered whether HRE length would influence also RNA foci formation. We first performed a Southern blot (SB) analysis to determine if *C9ORF72* HRE size is stable during fibroblast reprogramming into iPSC and in iPSC culturing during passages. By SB analyses of C9#1, C9#2, C9#3 and C9#4 iPSC lines, we observed that the HRE has shortened during

iPSC reprogramming only in C9#3 ranging from ~1100 units in skin fibroblast to ~130 units in the selected iPSC clone 1 (Figure 13).

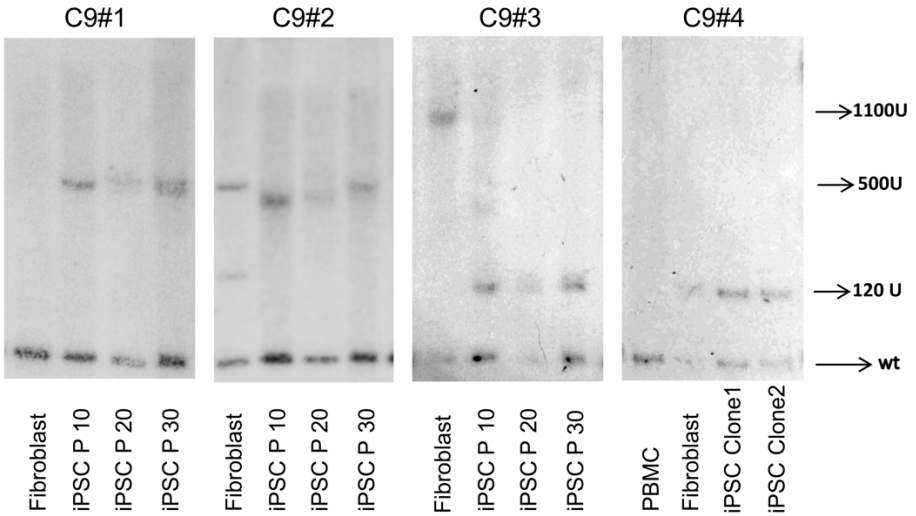


Figure 4-13. Quantification of *C9ORF72* hexanucleotide repeat expansion (HRE) length by Southern Blot analysis in different cell types of C9#1, 2, 3 and 4 patients. The measure of the HRE is reported as GGGGCC hexanucleotide units (U). Wild type allele (wt) (2-23 U). Number of passages in culture are indicated (P10, P20, P30).

The SB analyses, repeated at passages P10, P20 and P30 for all the iPSC lines, showed that the HRE is stable along the *in vitro* passages of the iPSC cultures (Figure 13).

We also observed that the mosaicism of the HRE is higher in fibroblasts, while it is lost with the selection of the single $C9^+$ iPSC clones (Figure 13).

We then analyzed HRE size by SB in the whole cohort of the eight different $C9^+$ iPSC lines and our results showed a high variability in HRE, as expected, ranging from ~120 units (C9#4) and ~130 units (C9#3) to ~1300 units (C9#8) (Figure 14a). The association between

promoter methylation state and HRE length was investigated and no significance was found (p=0.068) (Figure 14b).

When we compared HRE length and RNA foci formation, again we found no significant association (p=0.6671; r=0.181) (Figure 14c).

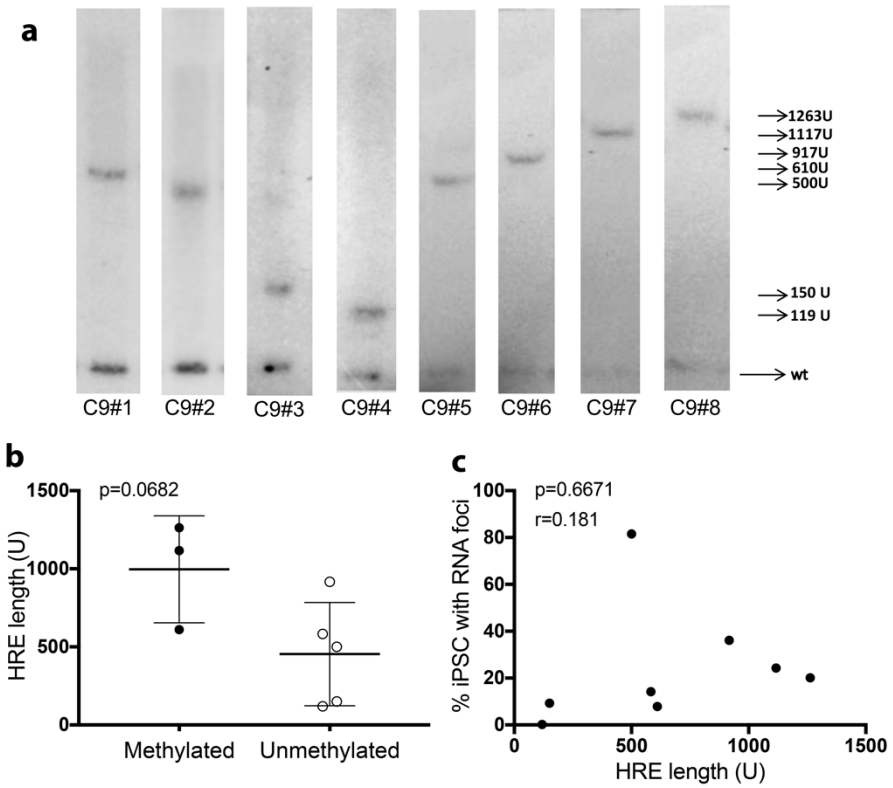


Figure 4-14. Quantification of *C9ORF72* hexanucleotide repeat expansion (HRE) length by SB analysis and correlation with promoter methylation state and RNA foci formation. a) HRE length in the eight C9+ iPSC lines is reported as GGGGCC hexanucleotide units (U). Wild type (wt) allele (2-23 units) is visualized as reference. b) Correlation between HRE length and methylation state in C9+ iPSC is shown. (Unpaired t-test, p=0.0682). c) Correlation between HRE length and RNA foci formation in terms of percentage (%) of cells presenting foci in culture. (Pearson correlation test, p=0.6671; r=0.181)

5. Discussion

Since the identification of *C9ORF72* mutation, the characteristics of the hexanucleotide repeat expansion, generating a novel CpG island in the first non-coding intron of the gene, suggested that epigenetic factors might be involved in disease pathogenesis and that, supported by experimental findings, they would have possibly represented a disease modifier and also a novel therapeutic target.

RNA foci formation and DPR production have been indeed demonstrated to negatively correlate with promoter methylation state, while gene expression changes seem also to confirm a lower transcription of intronic expanded sequence when *C9ORF72* promoter is methylated (65,66,70,75).

Our study aimed to further investigate the role of epigenetic modifications in *C9ORF72* pathology as disease modifiers by studying the correlation of pathological RNA readouts with promoter methylation state in $C9^+$ patient-derived iPSC and iPSC-MN, the most translational *in vitro* models of ALS/FTLD diseases.

Our results in three different $C9^+$ patient-derived samples showed that, while promoter methylation state in fibroblasts was consistent with promoter methylation in blood from in two cases, as previously reported (66), one $C9^+$ iPSC line showed a different methylation state of *C9ORF72* promoter from the original fibroblasts/blood, changing from a methylated to an unmethylated state. We demonstrated that this promoter methylation switch happened during iPSC reprogramming and, contrarily to what supposed by another group (70), we observed that it could reverse. After iPSC differentiation into iPSC-MN, in fact, we observed that also motor neuron differentiation step could cause a switch of *C9ORF72* promoter state in this cell line,

from being unmethylated in iPSC to returning methylated in iPSC-MN, similarly to the original fibroblasts, as was reported by another group in one single iPSC line (75). Interestingly, we also observed that the epigenetic modifications at *C9ORF72* gene promoter were stable during iPSC culturing *in vitro* for 40 passages and even when performing MN differentiation different times on the same iPSC line.

Moreover, using the MassARRAY method, a technique that is largely used in molecular oncology and that is new in this field of application, we could confirm that promoter methylation of the expanded *C9ORF72* is monoallelic (65) and that it occurs only in a percentage of cells in culture, indicating dynamic changes of this epigenetic mark in the same $C9^+$ cell clone. BS-Seq method is not a quantitative assay and gives information about the qualitative state of a methylated promoter either if the whole culture or a sub-population brings a complete promoter methylation. Therefore, the BS-seq analysis alone is not sufficient to study the role of DNA methylation, above all if functional correlations need to be established. We were the first to follow this epigenetic mark over iPSC passages in culture, confirming the stability of mutant *C9ORF72* DNA modifications *in vitro* and that $C9^+$ iPSCs are a suitable model for DNA methylation studies.
Our results showed that the downregulation of the two HRE-bearing V1 and V3 transcripts was the main *C9ORF72* pathologic feature that correlated to promoter methylation state and that therefore would be a reliable readout for promoter methylation-based therapeutic strategies in $C9^+$ cells. In our study we observed an interesting case of intermediate methylation state in one $C9^+$ iPSC line (13%), as

defined by quantitative MassARRAy data, which also showed intermediate levels of V1 and V3 transcripts compared to a more methylated (33%) or an unmethylated $C9^+$ iPSC line, corroborating the tight correlation between promoter methylation and gene expression regulation.

Our results also showed that the HRE itself was always methylated in our $C9^+$ iPSC, as previously reported, and this could well explain why we did not observe any difference in the gene expression level of the V2 isoform between methylated and unmethylated samples. The presence of the methylated HRE, in fact, would act as an additional inhibitory promoter for this specific isoform, leading to a down-regulation of the main V2 transcript, independently on the upstream *C9ORF72* gene promoter which instead regulates the V1 and V3 isoforms.

Besides the association between *C9ORF72* promoter methylation and its gene expression, we also investigated the possible association with other two parameters of $C9^+$ pathology, such as RNA foci formation and HRE size.

RNA foci formation is currently considered an important hallmark, as *C9ORF72* pathological RNA foci are always detectable in $C9^+$ autoptic brain samples as well as in all other $C9^+$ tissues or cultured peripheral samples (20,45,58). Our study, by analyzing 8 different $C9^+$ patient-derived iPSC lines allowed the broadest characterization of RNA foci formation so far in these cell models and in terms of association study with *C9ORF72* promoter methylation state.

We observed a high variability in RNA foci formation in our $C9^+$ iPSC cohort, which seemed to be not directly associated to the DNA

promoter methylation state. However, very preliminary and limited data suggested that *C9ORF72* promoter could be also epigenetically modified by hydroxymethylation which exerts an opposite role compared to methylation on RNA transcription and may therefore also modify RNA foci presentation. Surprisingly, RNA foci formation appeared as a cell line-specific and stable trait since each different iPSC line maintained stable both the percentage of cells containing foci and the number of foci per cell along passages in culture and, similarly, the amount was proportional also after subsequent MN differentiation steps. We found a positive correlation in our extended cohort between the number of foci per cell and the percentage of cells containing foci, suggesting that the higher number of RNA foci $C9^+$ iPSCs show, the more likely they are also surrounded by other cells that contain RNA foci.

Although the groups that are testing gene targeting therapies are currently including the quantification of RNA foci number in the analysis of efficacy (45,92–94), RNA foci formation was never found correlated to a worse clinical phenotype in $C9^+$ patient's tissue or cells (43), neither the topographic quantification of RNA foci led to a clear correlation with a specific neurodegenerative pattern. This lack of correlation supports the observation that the gain of function by toxic RNA aggregation is not the only mechanism involved, but that instead *C9ORF72*-related pathology is more complex. RNA foci are probably the more visible readout *in vitro*, but less RNA foci in $C9^+$ cells does not mean less toxicity, so other factors are necessary to be considered, for example the consequences of gene expression changes. The cascade of effects due to gene expression changes is

related on one hand to the loss of function of *C9ORF72* that likely impairs autophagy and membrane trafficking, but also to the poorly understood effects of higher levels of intronic *C9ORF72* HRE–containing RNAs.

In our study we also assessed the HRE length by SB analysis, looking for its possible implication in RNA foci formation and/or its correlation with promoter hypermethylation state. While no significant correlation was found with neither the number of cells with foci or the number of foci per cell, we could instead confirm that the presence of a HRE is a necessary condition for the formation of RNA foci species. RNA foci are indeed present only in $C9^{+}$ and furthermore we observed a very low number of RNA foci in iPSC and iPSC-MN from the two patients whose HRE size in iPSC was relatively small (120 and 150 repeat units). These findings suggest that the RNA FISH protocol or the probe that we used do not allow to detect the signal from RNA foci with a small repeat, or that RNA foci are fewer or smaller in such iPSC lines. In order to ascertain this hypothesis RNA FISH was performed on the fibroblasts of the same patient that, before reprogramming into iPSC, harbored a longer HRE. These fibroblasts showed much more RNA foci, corroborating the idea that also a link between HRE length and RNA foci formation actually exists.

6. Conclusions

Our results showed that promoter methylation of *C9ORF72* correlates to the downregulation of gene expression of the two HRE-containing transcripts V1 and V3, while promoter methylation is not associated to a lower number of RNA foci. We also found that $C9^+$ iPSC and iPSC-MN are a stable model of *C9ORF72* genetic and epigenetic features and related RNA pathology, even though a proper characterization of these cell lines is needed in order to exclude epigenetic switches and HRE contraction during iPSC reprogramming and neuronal differentiation steps.

Clinical, neuropathological and neuroimaging correlation studies confirmed the hypothesis that cytosine methylation in the promoter of the gene would be an heritable trait correlated to a milder phenotype, (62,63,66,67). The clinical correlation among $C9^+$ ALS/FTLD cohort is weak or particularly difficult to detect, probably also because of heterogeneous phenotypes, variable penetrance and the concurrency of modifier factors in determining the pathogenesis, so if this would represent a biomarker for a subtype of ALS-FTLD with better prognosis is still debated. Nevertheless, results from patient-derived cell models seem to associate promoter hypermethylation to lower pathological HRE-containing RNA content in neurons. Promoter hypermethylation could therefore represent a druggable target in *C9ORF72*-related pathology.

Since hydroxymethylation plays a dynamic role in chromatin remodeling, and since it is differently diffused in iPSCs and motor

neurons, it is likely to represent an additional epigenetic modifier candidate of *C9ORF72* gene expression. Moreover, since hydroxymethylation is the chemical intermediate step between methyl-cytosine and unmethylated cytosine, also its regulation represents a possible therapeutic target.

7. Bibliography

1. All European Academies, European Science Foundation. The European Code of Conduct for Research Integrity: Revised Edition. ALLEA - All Eur Acad. 2017;3–14.

2. Longinetti E, Fang F. Epidemiology of amyotrophic lateral sclerosis. Curr Opin Neurol. 2019;32(5):771–6.

3. Al-Chalabi A, Hardiman O. The epidemiology of ALS: A conspiracy of genes, environment and time. Nat Rev Neurol. 2013;9(11):617–28.

4. Belbasis L, Bellou V, Evangelou E. Environmental risk factors and amyotrophic lateral sclerosis: An umbrella review and critical assessment of current evidence from systematic reviews and meta-analyses of observational studies. Neuroepidemiology. 2016;46(2):96–105.

5. Cicero CE, Mostile G, Vasta R, Rapisarda V, Signorelli SS, Ferrante M, et al. Metals and neurodegenerative diseases. A systematic review. Environ Res [Internet]. 2017;159(August):82–94. Available from: http://dx.doi.org/10.1016/j.envres.2017.07.048

6. Taylor JP, Brown Jr RH, Cleveland DW. Decoding ALS: from genes to mechanism. Nature [Internet]. 2016 Nov 9;539:197. Available from: https://doi.org/10.1038/nature20413

7. Chia R, Chiò A, Traynor BJ. Novel genes associated with amyotrophic lateral sclerosis: diagnostic and clinical implications. Lancet Neurol [Internet]. 2018 Jan 1 [cited 2019 Sep 26];17(1):94–102. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29154141

8. Nguyen HP, Van Broeckhoven C, van der Zee J. ALS Genes in the Genomic Era and their Implications for FTD. Trends Genet [Internet]. 2018;34(6):404–23. Available from: http://dx.doi.org/10.1016/j.tig.2018.03.001

9. Van Blitterswijk M, Dejesus-Hernandez M, Rademakers R. How do *C9ORF72* repeat expansions cause amyotrophic lateral sclerosis and frontotemporal dementia: Can we learn from other noncoding repeat expansion disorders? Curr Opin Neurol. 2012;25(6):689–700.

10. Cooper-Knock J, Shaw PJ, Kirby J. The widening spectrum of *C9ORF72*-related disease; Genotype/phenotype correlations and potential modifiers of clinical phenotype. Acta Neuropathol. 2014;127(3):333–45.

11. Vucic S, Westeneng H-J, Al-Chalabi A, Van Den Berg LH, Talman P, Kiernan MC. Amyotrophic lateral sclerosis as a multi-step process: an Australia population study. Amyotroph Lateral Scler Front Degener [Internet]. 2019 Oct 2;20(7–8):532–7. Available from: https://doi.org/10.1080/21678421.2018.1556697

12. Saberi S, Stauffer JE, Schulte DJ, Ravits J. Neuropathology of Amyotrophic

Lateral Sclerosis and Its Variants. Neurol Clin [Internet]. 2015 Nov;33(4):855–76. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26515626

13. Cook C, Petrucelli L. Genetic Convergence Brings Clarity to the Enigmatic Red Line in ALS. Neuron [Internet]. 2019;101(6):1057–69. Available from: https://doi.org/10.1016/j.neuron.2019.02.032

14. Arai T, Hasegawa M, Akiyama H, Ikeda K, Nonaka T, Mori H, et al. TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Biochem Biophys Res Commun. 2006;351(3):602–11.

15. Neumann M, Sampathu DM, Kwong LK, Truax AC, Micsenyi MC, Chou TT, et al. Ubiquitinated TDP-43 in Frontotemporal Lobar Degeneration and Amyotrophic Lateral Sclerosis. Science (80-) [Internet]. 2006 Oct 6;314(5796):130 LP – 133. Available from: http://science.sciencemag.org/content/314/5796/130.abstract

16. Neumann M, Kwong LK, Lee EB, Kremmer E, Xu Y, Forman M, et al. NIH Public Access. 2010;117(2):137–49.

17. Brettschneider J, Del Tredici K, Toledo JB, Robinson JL, Irwin DJ, Grossman M, et al. Stages of pTDP-43 pathology in amyotrophic lateral sclerosis. Ann Neurol [Internet]. 2013/06/19. 2013 Jul;74(1):20–38. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23686809

18. Braak H, Brettschneider J, Ludolph AC, Lee VM, John Q, Tredici K Del. spread. 2014;9(12):708–14.

19. Renton AE, Majounie E, Waite A, Simón-Sánchez J, Rollinson S, Gibbs JR, et al. A hexanucleotide repeat expansion in *C9ORF72* is the cause of chromosome 9p21-linked ALS-FTD. Neuron. 2011;72(2):257–68.

20. Dejesus-hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, et al. NIH Public Access. 2012;72(2):245–56.

21. Zhang D, Iyer LM, He F, Aravind L. Discovery of novel DENN proteins: Implications for the evolution of eukaryotic intracellular membrane structures and human disease. Front Genet. 2012;3(DEC):1–10.

22. Xiao S, MacNair L, McLean J, McGoldrick P, McKeever P, Soleimani S, et al. *C9orf72* isoforms in Amyotrophic Lateral Sclerosis and Frontotemporal Lobar Degeneration. Brain Res [Internet]. 2016;1647:43–9. Available from: http://dx.doi.org/10.1016/j.brainres.2016.04.062

23. Zhang D, Iyer LM, He F, Aravind L. Discovery of Novel DENN Proteins: Implications for the Evolution of Eukaryotic Intracellular Membrane Structures and Human Disease. Front Genet [Internet]. 2012 Dec 13;3:283.

Available from: https://www.ncbi.nlm.nih.gov/pubmed/23248642

24. Xiao S, MacNair L, McLean J, McGoldrick P, McKeever P, Soleimani S, et al. *C9orf72* isoforms in Amyotrophic Lateral Sclerosis and Frontotemporal Lobar Degeneration. Brain Res [Internet]. 2016;1647:43–9. Available from: http://dx.doi.org/10.1016/j.brainres.2016.04.062

25. Yang M, Chen L, Swaminathan K, Herrlinger S, Lai F, Shiekhattar R, et al. A *C9ORF72*/SMCR8-containing complex regulates ULK1 and plays a dual role in autophagy. Sci Adv. 2016;2(9):1–17.

26. Amick J, Roczniak-Ferguson A, Ferguson SM. *C9orf72* binds SMCR8, localizes to lysosomes, and regulates mTORC1 signaling. Mol Biol Cell. 2016;27(20):3040–51.

27. van der Zee J, Gijselinck I, Dillen L, Van Langenhove T, Theuns J, Engelborghs S, et al. A Pan-European Study of the *C9orf72* Repeat Associated with FTLD: Geographic Prevalence, Genomic Instability, and Intermediate Repeats. Hum Mutat. 2013;34(2):363–73.

28. Ryan M, Heverin M, McLaughlin RL, Hardiman O. Lifetime Risk and Heritability of Amyotrophic Lateral Sclerosis. JAMA Neurol [Internet]. 2019 Jul 22; Available from: https://doi.org/10.1001/jamaneurol.2019.2044

29. Gijselinck I, Van Mossevelde S, Van Der Zee J, Sieben A, Engelborghs S, De Bleecker J, et al. The *C9orf72* repeat size correlates with onset age of disease, DNA methylation and transcriptional downregulation of the promoter. Mol Psychiatry. 2016;21(8):1112–24.

30. Murphy NA, Arthur KC, Tienari PJ, Houlden H, Chiò A, Traynor BJ. Age-related penetrance of the *C9orf72* repeat expansion. Sci Rep. 2017;7(1):1–7.

31. Cruts M, Gijselinck I, Van Langenhove T, van der Zee J, Van Broeckhoven C. Current insights into the *C9orf72* repeat expansion diseases of the FTLD/ALS spectrum. Trends Neurosci [Internet]. 2013 Aug 1 [cited 2019 Sep 30];36(8):450–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23746459

32. Gijselinck I, Van Langenhove T, van der Zee J, Sleegers K, Philtjens S, Kleinberger G, et al. A *C9orf72* promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum: A gene identification study. Lancet Neurol [Internet]. 2012;11(1):54–65. Available from: http://dx.doi.org/10.1016/S1474-4422(11)70261-7

33. Waite AJ, Bäumer D, East S, Neal J, Morris HR, Ansorge O, et al. Reduced *C9orf72* protein levels in frontal cortex of amyotrophic lateral sclerosis and frontotemporal degeneration brain with the *C9ORF72* hexanucleotide repeat

expansion. Neurobiol Aging [Internet]. 2014;35(7):1779.e5-1779.e13. Available from: http://dx.doi.org/10.1016/j.neurobiolaging.2014.01.016

34. Sullivan PM, Zhou X, Robins AM, Paushter DH, Kim D, Smolka MB, et al. The ALS/FTLD associated protein C9orf72 associates with SMCR8 and WDR41 to regulate the autophagy-lysosome pathway. Acta Neuropathol Commun [Internet]. 2016;4(1):51. Available from: http://dx.doi.org/10.1186/s40478-016-0324-5

35. Xiao S, MacNair L, McGoldrick P, McKeever PM, McLean JR, Zhang M, et al. Isoform-specific antibodies reveal distinct subcellular localizations of C9orf72 in amyotrophic lateral sclerosis. Ann Neurol [Internet]. 2015 Oct 1;78(4):568–83. Available from: https://doi.org/10.1002/ana.24469

36. Rizzu P, Blauwendraat C, Heetveld S, Lynes EM, Castillo-Lizardo M, Dhingra A, et al. C9orf72 is differentially expressed in the central nervous system and myeloid cells and consistently reduced in C9orf72, MAPT and GRN mutation carriers. Acta Neuropathol Commun [Internet]. 2016;4(1):37. Available from: http://dx.doi.org/10.1186/s40478-016-0306-7

37. Farg MA, Sundaramoorthy V, Sultana JM, Yang S, Atkinson RAK, Levina V, et al. C9ORF72, implicated in amytrophic lateral sclerosis and frontotemporal dementia, regulates endosomal trafficking. Hum Mol Genet [Internet]. 2014 Feb 18;23(13):3579–95. Available from: https://doi.org/10.1093/hmg/ddu068

38. Haeusler AR, Donnelly CJ, Periz G, Simko EAJ, Shaw G, Kim M, et al. Cascades of Disease. Nature [Internet]. 2014;507(7491):195–200. Available from: http://dx.doi.org/10.1038/nature13124%5Cnpapers3://publication/doi/10.1038/nature13124%5Cnpapers2://publication/doi/10.1038/nature13124

39. Almeida S, Gascon E, Tran H, Chou HJ, Gendron TF, Degroot S, et al. Modeling key pathological features of frontotemporal dementia with C9ORF72 repeat expansion in iPSC-derived human neurons. Acta Neuropathol. 2013;126(3):385–99.

40. Simone R, Fratta P, Neidle S, Parkinson GN, Isaacs AM. G-quadruplexes: Emerging roles in neurodegenerative diseases and the non-coding transcriptome. FEBS Lett. 2015;589(14):1653–68.

41. Fratta P, Mizielinska S, Nicoll AJ, Zloh M, Fisher EMC, Parkinson G, et al. C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. Sci Rep [Internet]. 2012 Dec 21;2:1016. Available from: https://doi.org/10.1038/srep01016

42. Almeida S, Gascon E, Tran H, Chou HJ, Gendron TF, DeGroot S, et al. Modeling key pathological features of frontotemporal dementia with C9ORF72 repeat expansion in iPSC-derived human neurons. Acta Neuropathol [Internet]. 2013 Sep;126(3):385–99. Available from:

https://doi.org/10.1007/s00401-013-1149-y

43. DeJesus-Hernandez M, Finch NCA, Wang X, Gendron TF, Bieniek KF, Heckman MG, et al. In-depth clinico-pathological examination of RNA foci in a large cohort of C9ORF72 expansion carriers. Acta Neuropathol. 2017;134(2):255–69.

44. Lee Y-B, Chen H-J, Peres JN, Gomez-Deza J, Attig J, Štalekar M, et al. Hexanucleotide Repeats in ALS/FTD Form Length-Dependent RNA Foci, Sequester RNA Binding Proteins, and Are Neurotoxic. Cell Rep [Internet]. 2013 Dec 12 [cited 2019 Oct 4];5(5):1178–86. Available from: https://www.sciencedirect.com/science/article/pii/S2211124713006487

45. Lagier-Tourenne C, Baughn M, Rigo F, Sun S, Liu P, Li H-R, et al. Targeted degradation of sense and antisense C9orf72 RNA foci as therapy for ALS and frontotemporal degeneration. Proc Natl Acad Sci [Internet]. 2013;110(47):E4530–9. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.1318835110

46. Mizielinska S, Lashley T, Norona FE, Clayton EL, Ridler CE, Fratta P, et al. C9orf72 frontotemporal lobar degeneration is characterised by frequent neuronal sense and antisense RNA foci. Acta Neuropathol [Internet]. 2013;126(6):845–57. Available from: https://doi.org/10.1007/s00401-013-1200-z

47. Gendron TF, Bieniek KF, Zhang Y-J, Jansen-West K, Ash PEA, Caulfield T, et al. Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. Acta Neuropathol [Internet]. 2013 Dec;126(6):829–44. Available from: https://doi.org/10.1007/s00401-013-1192-8

48. Ash PEA, Bieniek KF, Gendron TF, Caulfield T, Lin WL, DeJesus-Hernandez M, et al. Unconventional Translation of C9ORF72 GGGGCC Expansion Generates Insoluble Polypeptides Specific to c9FTD/ALS. Neuron [Internet]. 2013;77(4):639–46. Available from: http://dx.doi.org/10.1016/j.neuron.2013.02.004

49. Kwon I, Xiang S, Kato M, Wu L, Theodoropoulos P, Wang T, et al. Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. Science (80-) [Internet]. 2014 Sep 5;345(6201):1139 LP – 1145. Available from: http://science.sciencemag.org/content/345/6201/1139.abstract

50. Boeynaems S, Bogaert E, Kovacs D, Konijnenberg A, Timmerman E, Volkov A, et al. Phase Separation of C9orf72 Dipeptide Repeats Perturbs Stress Granule Dynamics. Mol Cell [Internet]. 2017 Mar 16 [cited 2019 Oct 4];65(6):1044-1055.e5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28306503

51. Corrado L, Tiloca C, Locci C, Bagarotti A, Hamzeiy H, Colombrita C, et al. Characterization of the c9orf72 GC-rich low complexity sequence in two cohorts of Italian and Turkish ALS cases. Amyotroph Lateral Scler Front Degener [Internet]. 2018 Jul 3;19(5–6):426–31. Available from: https://doi.org/10.1080/21678421.2018.1440407

52. Jovičič A, Mertens J, Boeynaems S, Bogaert E, Chai N, Yamada SB, et al. Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. Nat Neurosci. 2015;18(9):1226–9.

53. Boeynaems S, Bogaert E, Michiels E, Gijselinck I, Sieben A, Jovičić A, et al. Drosophila screen connects nuclear transport genes to DPR pathology in c9ALS/FTD. Sci Rep [Internet]. 2016 Feb 12;6:20877. Available from: https://doi.org/10.1038/srep20877

54. Freibaum BD, Lu Y, Lopez-Gonzalez R, Kim NC, Almeida S, Lee K-H, et al. GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. Nature [Internet]. 2015 Aug 26;525:129. Available from: https://doi.org/10.1038/nature14974

55. Zhang K, Donnelly CJ, Haeusler AR, Grima JC, Machamer JB, Steinwald P, et al. The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. Nature [Internet]. 2015 Aug 26;525:56. Available from: https://doi.org/10.1038/nature14973

56. van Blitterswijk M, DeJesus-Hernandez M, Niemantsverdriet E, Murray ME, Heckman MG, Diehl NN, et al. Association between repeat sizes and clinical and pathological characteristics in carriers of C9ORF72 repeat expansions (Xpansize-72): a cross-sectional cohort study. Lancet Neurol [Internet]. 2013 Oct 1 [cited 2019 Oct 2];12(10):978–88. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1474442213702102

57. Hübers A, Marroquin N, Schmoll B, Vielhaber S, Just M, Mayer B, et al. Polymerase chain reaction and Southern blot-based analysis of the C9orf72 hexanucleotide repeat in different motor neuron diseases. Neurobiol Aging [Internet]. 2014;35(5):1214.e1-1214.e6. Available from: http://dx.doi.org/10.1016/j.neurobiolaging.2013.11.034

58. Blitterswijk M Van, Dejesus-hernandez M, Niemantsverdriet E, Baker MC, Finch NA, Bauer PO, et al. NIH Public Access. 2014;12(10):1–18.

59. Castaldo I, Pinelli M, Monticelli A, Acquaviva F, Giacchetti M, Filla A, et al. DNA methylation in intron 1 of the frataxin gene is related to GAA repeat length and age of onset in Friedreich ataxia patients. J Med Genet [Internet]. 2008 Dec 1;45(12):808 LP – 812. Available from: http://jmg.bmj.com/content/45/12/808.abstract

60. McLennan Y, Polussa J, Hagerman FT and R. Fragile X Syndrome [Internet]. Vol. 12, Current Genomics. 2011. p. 216–24. Available from:

http://www.eurekaselect.com/node/74124/article

61. Iyer S, Ravi Acharya K, Subramanian V. A comparative bioinformatic analysis of C9orf72. PeerJ. 2018;2018(2).

62. Xi Z, Zinman L, Moreno D, Schymick J, Liang Y, Sato C, et al. Hypermethylation of the CpG island near the G4C2 repeat in ALS with a C9orf72 expansion. Am J Hum Genet [Internet]. 2013;92(6):981–9. Available from: http://dx.doi.org/10.1016/j.ajhg.2013.04.017

63. Xi Z, Zhang M, Bruni AC, Maletta RG, Colao R, Fratta P, et al. The C9orf72 repeat expansion itself is methylated in ALS and FTLD patients. Acta Neuropathol. 2015;129(5):715–27.

64. Belzil V V., Bauer PO, Gendron TF, Murray ME, Dickson D, Petrucelli L. Characterization of DNA hypermethylation in the cerebellum of c9FTD/ALS patients. Brain Res [Internet]. 2014;1584:15–21. Available from: http://dx.doi.org/10.1016/j.brainres.2014.02.015

65. Liu EY, Russ J, Wu K, Neal D, Suh E, McNally AG, et al. C9orf72 hypermethylation protects against repeat expansion-associated pathology in ALS/FTD. Acta Neuropathol. 2014;128(4):525–41.

66. Russ J, Liu EY, Wu K, Neal D, Suh ER, Irwin DJ, et al. Hypermethylation of repeat expanded C9orf72 is a clinical and molecular disease modifier. Acta Neuropathol. 2015;129(1):39–52.

67. McMillan CT, Russ J, Wood EM, Irwin DJ, Grossman M, Mccluskey L, et al. C9orf72 promoter hypermethylation is neuroprotective: Neuroimaging and neuropathologic evidence. Neurology. 2015;84:1622–30.

68. Liu Y, Pattamatta A, Zu T, Reid T, Bardhi O, Borchelt DR, et al. C9orf72 BAC Mouse Model with Motor Deficits and Neurodegenerative Features of ALS/FTD. Neuron [Internet]. 2016;90(3):521–34. Available from: http://dx.doi.org/10.1016/j.neuron.2016.04.005

69. Esanov R, Cabrera GT, Andrade NS, Gendron TF, Brown RH, Benatar M, et al. A C9ORF72 BAC mouse model recapitulates key epigenetic perturbations of ALS/FTD. Mol Neurodegener. 2017;12(1):1–11.

70. Cohen-Hadad Y, Altarescu G, Eldar-Geva T, Levi-Lahad E, Zhang M, Rogaeva E, et al. Marked Differences in C9orf72 Methylation Status and Isoform Expression between C9/ALS Human Embryonic and Induced Pluripotent Stem Cells. Stem Cell Reports [Internet]. 2016;7(5):927–40. Available from: http://dx.doi.org/10.1016/j.stemcr.2016.09.011

71. Chestnut B a., Chang Q, Price A, Lesuisse C, Wong M, Martin LJ. Epigenetic Regulation of Motor Neuron Cell Death Through DNA Methylation. J Neurosci. 2011;31(46):16619–36.

72. Sarg B, Koutzamani E, Helliger W, Rundquist I, Lindner HH. Postsynthetic Trimethylation of Histone H4 at Lysine 20 in Mammalian Tissues Is Associated with Aging. J Biol Chem [Internet]. 2002 Oct 18;277(42):39195–201. Available from: http://www.jbc.org/content/277/42/39195.abstract

73. Fraga MF, Esteller M. Epigenetics and aging: the targets and the marks. Trends Genet [Internet]. 2007 Aug 1 [cited 2019 Oct 4];23(8):413–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17559965

74. Al-Mahdawi S, Virmouni SA, Pook MA. The emerging role of 5-hydroxymethylcytosine in neurodegenerative diseases. Front Neurosci. 2014;8(DEC):1–8.

75. Esanov R, Belle KC, van Blitterswijk M, Belzil V V., Rademakers R, Dickson DW, et al. C9orf72 promoter hypermethylation is reduced while hydroxymethylation is acquired during reprogramming of ALS patient cells. Exp Neurol [Internet]. 2016;277:171–7. Available from: http://dx.doi.org/10.1016/j.expneurol.2015.12.022

76. Kinney SM, Ph D, Chin HG, Pradhan S, Biolabs NE. Analysis of 5-methylcytosine and 5-hydroxymethylcytosine Levels in Human Brain and Liver DNA Samples using the EpiMark 5-hmC and 5-mC Analysis Kit. 11:1–3.

77. Kinney SM, Chin HG, Vaisvila R, Bitinaite J, Zheng Y, Estève PO, et al. Tissue-specific distribution and dynamic changes of 5-hydroxymethylcytosine in mammalian genomes. J Biol Chem. 2011;286(28):24685–93.

78. Ito S, D'Alessio AC, Taranova O V, Hong K, Sowers LC, Zhang Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature [Internet]. 2010;466(7310):1129–33. Available from: https://doi.org/10.1038/nature09303

79. Yanovsky-Dagan S. Modeling diseases of noncoding unstable repeat expansions using mutant pluripotent stem cells. World J Stem Cells [Internet]. 2015;7(5):823. Available from: http://www.wjgnet.com/1948-0210/full/v7/i5/823.htm

80. Cooper-Knock J, Walsh MJ, Higginbottom A, Highley JR, Dickman MJ, Edbauer D, et al. Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. Brain. 2014;137(7):2040–51.

81. Onesto E, Colombrita C, Gumina V, Borghi MO, Dusi S, Doretti A, et al. Gene-specific mitochondria dysfunctions in human TARDBP and C9ORF72 fibroblasts. Acta Neuropathol Commun [Internet]. 2016 May 5;4(1):47. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27151080

82. Zeier Z, Esanov R, Belle KC, Volmar C-H, Johnstone AL, Halley P, et al.

Bromodomain inhibitors regulate the C9ORF72 locus in ALS. Exp Neurol [Internet]. 2015;271:241–50. Available from: http://www.sciencedirect.com/science/article/pii/S0014488615300273

83. Takahashi K, Yamanaka S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell. 2006;126(4):663–76.

84. Sardo V Lo, Ferguson W, Erikson GA, Topol EJ, Baldwin KK, Torkamani A. The effect of aging on human induced pluripotent stem cells. Nat Biotechnol. 2017;35(1):69–74.

85. Kang E, Wang X, Tippner-Hedges R, Ma H, Folmes CDL, Gutierrez NM, et al. Age-related accumulation of somatic mitochondrial DNA mutations in adult-derived human ipscs. Cell Stem Cell [Internet]. 2016;18(5):625–36. Available from: http://dx.doi.org/10.1016/j.stem.2016.02.005

86. Guo W, Fumagalli L, Prior R, van den Bosch L. Current advances and limitations in modeling ALS/FTD in a dish using induced pluripotent stem cells. Front Neurosci. 2017;11(DEC):1–20.

87. Huang K, Shen Y, Xue Z, Bibikova M, April C, Liu Z, et al. A panel of CpG methylation sites distinguishes human embryonic stem cells and induced pluripotent stem cells. Stem Cell Reports [Internet]. 2014;2(1):36–43. Available from: http://dx.doi.org/10.1016/j.stemcr.2013.11.003

88. Nashun B, Hill PW, Hajkova P. Reprogramming of cell fate: epigenetic memory and the erasure of memories past. EMBO J. 2015;34(10):1296–308.

89. Ma H, Morey R, O'Neil RC, He Y, Daughtry B, Schultz MD, et al. Abnormalities in human pluripotent cells due to reprogramming mechanisms. Nature [Internet]. 2014;511(7508):177–83. Available from: https://doi.org/10.1038/nature13551

90. Donnelly CJ, Zhang P, Pham JT, Heusler AR, Mistry NA, Vidensky S, et al. NIH Public Access. 2014;80(2):415–28.

91. Smith RA, Bennett CF, Cleveland DW, Smith RA, Miller TM, Yamanaka K, et al. Antisense oligonucleotide therapy for neurodegenerative disease Find the latest version : Antisense oligonucleotide therapy for neurodegenerative disease. J Clin Invest. 2006;116(8):2290–6.

92. Jiang J, Cleveland DW. Bidirectional Transcriptional Inhibition as Therapy for ALS/FTD Caused by Repeat Expansion in C9orf72. Neuron [Internet]. 2016;92(6):1160–3. Available from: http://dx.doi.org/10.1016/j.neuron.2016.12.008

93. Mis MSC, Brajkovic S, Tafuri F, Bresolin N, Comi GP, Corti S. Development of Therapeutics for C9ORF72 ALS/FTD-Related Disorders. Mol Neurobiol.

2017;54(6):4466–76.

94. Simone R, Balendra R, Moens TG, Preza E, Wilson KM, Heslegrave A, et al. G-quadruplex-binding small molecules ameliorate C9orf72 FTD / ALS pathology in vitro and in vivo . EMBO Mol Med. 2018;10(1):22–31.

95. Su Z, Zhang Y, Gendron TF, Bauer PO, Chew J, Yang WY, et al. Discovery of a Biomarker and Lead Small Molecules to Target r(GGGGCC)-Associated Defects in c9FTD/ALS. Neuron [Internet]. 2014;83(5):1043–50. Available from: http://dx.doi.org/10.1016/j.neuron.2014.07.041

96. Zhang M, Tartaglia MC, Moreno D, Sato C, McKeever P, Weichert A, et al. DNA methylation age-acceleration is associated with disease duration and age at onset in C9orf72 patients. Acta Neuropathol. 2017;134(2):271–9.

97. Bossolasco P, Sassone F, Gumina V, Peverelli S, Garzo M, Silani V. Motor neuron differentiation of iPSCs obtained from peripheral blood of a mutant TARDBP ALS patient. Stem Cell Res [Internet]. 2018;30(January):61–8. Available from: https://doi.org/10.1016/j.scr.2018.05.009

98. Fontana L, Bedeschi MF, Maitz S, Cereda A, Faré C, Motta S, et al. Characterization of multi-locus imprinting disturbances and underlying genetic defects in patients with chromosome 11p15.5 related imprinting disorders. Epigenetics. 2018;13(9):897–909.

99. Coolen MW, Statham AI, Gardiner-Garden M, Clark SJ. Genomic profiling of CpG methylation and allelic specificity using quantitative high-throughput mass spectrometry: Critical evaluation and improvements. Nucleic Acids Res. 2007;35(18).

100. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods [Internet]. 2012;9(7):671–5. Available from: https://doi.org/10.1038/nmeth.2089

Scientific Production

Publications on journals:

Reprogramming fibroblasts and peripheral blood cells from a *C9ORF72* patient: a proof-of-principle study
Bardelli D., Sassone F., Colombrita C., Volpe C., Gumina V., Peverelli S., Catusi I., Ratti A., Silani V., Bossolasco P.
(*Submitted to* Journal of Molecular and Cellular Medicine, under review)

One publication, entirely based on this study, is in preparation and will be submitted to peer-reviewed journal.

Poster presentations:

IPSC-derived motor neurons: an *in vitro* model to study ALS disease. Bossolasco P., Sassone F., Colombrita C., Gumina V., Volpe C., Tiloca C., Peverelli S, Maraschi A.M., Cova L., Ticozzi N., Verde F., Ratti A., Silani V. Neuronest 2° meeting traslazionale del gruppo di ricerca strategico in neuroscienze de "La Statale" Università degli Studi di Milano 27 Mar 2018 Milan

Promoter methylation of *C9orf72*: a disease modifier and a model confounder in *C9orf72* ALS/FTD iPSC and iMNs. Volpe C., Colombrita C., Bossolasco P., Tiloca C., Peverelli S, Gumina V., Maraschi A.M., Sassone F., Silani V., Ratti A.
AMPS/MD-PhD Congress 2018 Paris

TDP-43 protein and SUMOylation. Maraschi A.M., Gumina V., Colombrita C., Volpe C., Feligioni M., Silani V., Ratti A.
ENCALS meeting 20-22 Jun 2018 Oxford

Comparison of iPSC-derived motor neurons from *C9ORF72* patient fibroblasts and peripheral blood cells: a proof-of-principle study. Bardelli D., Sassone F., Volpe C., Colombrita C., Gumina V., Peverelli S., Ratti A., Silani V. and Bossolasco P.
ENCALS meeting 7-9 Dec 2018 Glasgow

Epigenetic features of *C9orf72* gene promoter as RNA foci modifiers in iPSCs and iPSC-MNs. Volpe C., Colombrita C., Bossolasco P., Gumina V., Tiloca C., Peverelli S., Maraschi A., Bardelli D., Silani V., Ratti A.
23rd ESN biennial meeting 1-4 Sep 2019 Milan

Oral presentations:

Epigenetic features of *C9orf72* gene: DNA methylation and hydroxymethylation as RNA foci modifiers in iPSCs. Volpe C., Colombrita C., Bossolasco P., Gumina V., Tiloca C., Peverelli S., Maraschi A., Bardelli D., Silani V., Ratti A.
ENCALS meeting 7-9 Dec 2018 Glasgow

Epigenetic features of *C9ORF72* gene as modifiers in iPSCs and iPSC-derived motor neurons.
Colombrita Claudia, Bossolasco Patrizia, Gumina Valentina, Tiloca Cinzia, Peverelli Silvia, Maraschi AnnaMaria, Bardelli Donatella, Silani Vincenzo, Ratti Antonia
4th BioMetra Workshop 23 Sep 2019 Milan

Acknowledgements

This project was financially supported by the Laboratory of Neuroscience of IRCCS Istituto Auxologico Italiano (Grant Ricerca Finalizzata, Ministero della Salute).
I wish to thank my supervisor Prof. Antonia Ratti and Prof. Vincenzo Silani for the opportunity to work on this PhD project and my colleagues of the Laboratory of Neuroscience of IRCCS Istituto Auxologico Italiano for the scientific support, and also for the extra-patience that pipetting side by side with a physician sometimes requires.
I am especially grateful to Prof. Samaja for inviting MD students, including me, to embark on biomedical research at a time when MD-PhD track was still a pilot program.

More than words of gratitude are owed to the people that supported and inestimably enriched the undertaking and completion of this PhD program and especially to Najwa O., Michela M., Sara O., Federico P., Fabio M. and to my sister Silvia.