



# A multifunctional locus controls motor neuron differentiation through short and long noncoding RNAs

Andrea Carvelli<sup>1,†</sup>, Adriano Setti<sup>2,†</sup>, Fabio Desideri<sup>1,†</sup>, Silvia Giulia Galfrè<sup>2</sup> , Silvia Biscarini<sup>1</sup>, Tiziana Santini<sup>2</sup>, Alessio Colantoni<sup>1</sup>, Giovanna Peruzzi<sup>1</sup>, Matteo Jacopo Marzi<sup>4</sup> , Davide Capauto<sup>2,§</sup>, Silvia Di Angelantonio<sup>5</sup> , Monica Ballarino<sup>2</sup> , Francesco Nicassio<sup>4</sup> , Pietro Laneve<sup>3,\*,‡</sup> & Irene Bozzoni<sup>1,2,\*\*,‡</sup>

## Abstract

The transition from dividing progenitors to postmitotic motor neurons (MNs) is orchestrated by a series of events, which are mainly studied at the transcriptional level by analyzing the activity of specific programming transcription factors. Here, we identify a post-transcriptional role of a MN-specific transcriptional unit (MN2) harboring a lncRNA (lncMN2-203) and two miRNAs (miR-325-3p and miR-384-5p) in this transition. Through the use of *in vitro* mESC differentiation and single-cell sequencing of CRISPR/Cas9 mutants, we demonstrate that lncMN2-203 affects MN differentiation by sponging miR-466i-5p and upregulating its targets, including several factors involved in neuronal differentiation and function. In parallel, miR-325-3p and miR-384-5p, co-transcribed with lncMN2-203, act by repressing proliferation-related factors. These findings indicate the functional relevance of the MN2 locus and exemplify additional layers of specificity regulation in MN differentiation.

**Keywords** competing endogenous RNA; long noncoding RNAs; microRNAs; motor neuron differentiation; single-cell RNA sequencing

**Subject Categories** Chromatin, Transcription & Genomics; Neuroscience; RNA Biology

**DOI** 10.15252/embj.2021108918 | Received 8 June 2021 | Revised 29 April 2022 | Accepted 2 May 2022

**The EMBO Journal (2022) 41: e108918**

## Introduction

The development and functional integration of motor neurons (MN) into spinal motor circuits require the establishment of spatiotemporal patterns of signaling factors that activate specific molecular programs (Tanabe & Jessell, 1996; Jessell, 2000; Lu *et al.*, 2015). While many studies underline the relevance of the dynamic interplay between transcriptional events in MN progenitor specification, MN fate acquisition and subtype definition (extensively reviewed in Stifani, 2014; Cave & Sockanathan, 2018; Andrews *et al.*, 2019; Catela & Kratsios, 2021), very little is known about the role of noncoding RNAs (ncRNAs) in the post-transcriptional control of MN differentiation. In the last decades, deep-sequencing methodologies combined with refined computational data analyses have unveiled previously unsuspected features of ncRNAs (Carninci *et al.*, 2005; Birney *et al.*, 2007), and in particular to long noncoding RNAs (lncRNAs; Harrow *et al.*, 2012; Xie *et al.*, 2014), in the metabolism of the nervous system (NS). Indeed, lncRNAs (i) account for more than 40% of the transcriptome (Derrien *et al.*, 2012); (ii) have higher tissue-specificity than mRNAs (Mercer *et al.*, 2008; Molyneaux *et al.*, 2015; Chen *et al.*, 2018; Martone *et al.*, 2020); and (iii) participate in the regulation of physio-pathological neuronal networks (Briggs *et al.*, 2015). Nevertheless, to date, only very few lncRNAs have been mechanistically linked to MN development or identity in vertebrates, and in the few known cases, they were again linked to transcriptional control (Chen & Chen, 2020; Vangoor *et al.*, 2021). Among them, the chromatin-associated lncRNA *Meg3*, expressed from the *Dlk1-Dio3* locus, maintains the Prc2-dependent silenced state of MN progenitor-specific genes and caudal MN genes in rostral MNs (Yen *et al.*, 2018). Similarly, *Cat7* enhances the suppressive activity of the PRC1 suppressive

- 1 Center for Life Nano- & Neuro-Science of Istituto Italiano di Tecnologia (IIT), Rome, Italy
- 2 Department of Biology and Biotechnologies "Charles Darwin", Sapienza University of Rome, Rome, Italy
- 3 Institute of Molecular Biology and Pathology, National Research Council, Rome, Italy
- 4 Center for Genomic Science of Istituto Italiano di Tecnologia (IIT), Milan, Italy
- 5 Department of Physiology and Pharmacology, Sapienza University of Rome, Rome, Italy

\*Corresponding author. Tel: +39 0649912205; E-mail: pieter.laneve@cnr.it

\*\*Corresponding author. Tel: +39 0649912202; E-mail: irene.bozzoni@uniroma1.it

†These authors contributed equally to this work

‡These authors contributed equally to this work

§Present address: Child Study Center, Yale University, New Haven, CT, USA

machinery on the TF *Hb9*, to guarantee its early repression along MN fate determination (Ray *et al*, 2016). Finally, in zebrafish, *lncrps25* is necessary for the expression of the TF *Olig2*, involved in MN and oligodendrocyte differentiation (Gao *et al*, 2020).

We recently described a signature of lncRNAs induced in a model system recapitulating MN differentiation from mouse embryonic stem cells (mESC). Among these transcripts, lncMN2 resulted of specific interest since it was enriched in terminally differentiated MNs, a similar specificity of expression was observed in iPSC-derived human MN, and importantly, its expression levels were affected by pathogenetic mutations in the ALS-associated Fus protein (Biscarini *et al*, 2018).

In this paper, we dissected the complexity of the MN2 locus and studied its molecular mechanism of action; we discovered a complex network of regulatory circuitries operating at the post-transcriptional level. We found that three lncMN2 splicing isoforms (lncMN2-202/204 with predominant nuclear localization and lncMN2-203 mainly cytoplasmic) and two embedded microRNAs (miR-325-3p and miR-384-5p) derive from the same genomic locus. An additional miRNA (miR-466i-5p) was functionally associated with the activity of this locus since it was computationally and biochemically validated to be an interactor of lncMN2-203. CRISPR-Cas9 gene-editing and single-cell RNA sequencing allowed us to assign a role in MN differentiation to the different ncRNAs and to identify the circuitries on which they impinge. A major pathway important for MN differentiation is the miR-466i-5p sponging activity of lncMN2-203, which upregulates several factors involved in neuronal differentiation and function. Additional pathways stand on miR-325-3p and miR-384-5p activities, which were found to repress genes linked to cell proliferation. These data highlight novel circuitries involved in the control of MN maturation and show how the different components of the MN2 locus synergize to control the robustness of this process.

## Results and Discussion

### Transcriptional complexity of the MN2 locus

lncMN2 (5330434G04Rik, according to Ensembl) is a long intergenic noncoding transcript enriched in MNs obtained from differentiation

of mESC-derived embryoid bodies (EB) committed to motoneurogenesis (Biscarini *et al*, 2018). mESCs, carrying the *Hb9::GFP* transgene, were differentiated according to Witcherle and colleagues (Witcherle *et al*, 2002; Witcherle & Peljto, 2008 and Fig EV1A) and correct differentiation was verified by the timing of marker expression (i.e., downregulation of pluripotency genes and concomitant increase in MN-specific factors, Fig EV1B). Three of the four lncMN2 splicing variants annotated on the Ensembl (ENSMUSG00000087620.8, Yates *et al*, 2020) genome browser database (Figs 1A and EV1C), were found to be expressed in MN (lncMN2-202/203/204, Fig 1B, left panel; Biscarini *et al*, 2018). Two miRNA coding regions are also located within the same locus. Specifically, miR-325-3p is harbored in the fourth intron of the gene, whereas miR-384-5p maps downstream of the longest lncMN2-202 and 204 isoforms (Fig 1A). Quantitative real-time PCR (qRT-PCR) analyses revealed that the lncMN2 isoforms and the embedded miRNAs were upregulated in differentiated EB (EB6), as compared to mESC (Fig 1B, right panel). Interestingly, the lncMN2-203 levels increased by day 2-3 of MN differentiation (Fig EV1D, left panel), suggesting a correlation between its expression and the specification of early neuronal progenitors. Consistently, miR-325-3p and 384-5p levels increased with the same trend (Fig EV1D, right panel). Nuclear-cytoplasmic fractionation indicated the prevalent cytoplasmic localization of lncMN2-203, which is opposed to a preferential nuclear accumulation of lncMN2-202 and lncMN2-204 species (Fig EV1E).

Overall, these analyses allowed us to clarify the structure and the expression output of the murine lncMN2 locus and to define the complex array of long and short noncoding RNA species, which are therein transcribed.

### lncMN2 function: CRISPR/Cas9 mediated gene-editing and phenotypic characterization

In order to assess the functional relevance of the lncMN2 locus, a CRISPR/Cas9 gene-editing approach was applied to generate loss-of-function mutant mESCs (KO). *In silico* design of the Cas9 guide RNAs was performed in order to insert, at the beginning of the first lncMN2 exon, a 100 nt-long cassette containing a minimal synthetic poly-A signal, followed by two repetitions of the polymerase destabilizing MAZ sequences (Ballarino *et al*, 2018, Fig EV1F). Parallel

#### Figure 1. lncMN2 KO affects MN differentiation.

- Block-and-line model of lncMN2 gene structure. Exons specific for lncMN2-202/204 isoforms are in yellow; lncMN2-203-specific exon is in blue. Shared exons are gray. miR-325 and miR-384 are represented as red and green blocks, respectively.
- qRT-PCR expression analysis of lncMN2 isoforms (left panel) and embedded miRNAs (right panel) in mESC and EBs differentiated for 6 days (EB6). Data, that are expressed as mean (error bars represent SD), were normalized over *Atp50* mRNA (left) and *snoRNA U25* (right) levels and reported as RNA fold over the control (mESC), set as 1.  $N = 3$  biological replicates.
- qRT-PCR expression analysis of lncMN2-202/203/204 isoforms (left panel), miR-325-3p (middle panel), and miR-384-5p (right panel) in WT, KO1, and KO2 EB6. Details as in B.
- Cytofluorimetric analysis of cells from dissociated EB6. The left diagram reports the distribution of cells (percentage of the peak) according to GFP levels, from WT (gray), KO1 (orange), and KO2 (blue) EB. In the histogram aside, the percentage of GFP<sup>+</sup> cells in WT, KO1, and KO2 is reported. Data are expressed as mean and error bars represent SD.  $N = 3$  biological replicates.
- Left: representative immunofluorescence of the MN marker *Isl1* (red signals) in cells maintained *in vitro* for 14 days after EB6 dissociation (DIV14). DAPI counterstaining in blue. Scale bar = 100  $\mu$ m. Right: box plot shows the distribution of *Isl1*<sup>+</sup> nuclei with respect to the total cell number marked with DAPI staining in WT and in lncMN2 KO mutants (clone 1 and clone 2). Data are expressed as mean and error bars represent SD. The central band represents the median and the whiskers represent the distribution that goes from the minimum value to the lower quartile and from the upper quartile to the maximum value.  $N = 3$  biological replicates and 10 fields were analyzed for each replicate.

Data information: Histogram dots represent single fields replicates. \* $P \leq 0.05$ , \*\* $P \leq 0.01$ , \*\*\* $P \leq 0.001$ , n.s. > 0.05 (two-tailed, unpaired Student's *t*-test).

Source data are available online for this figure.

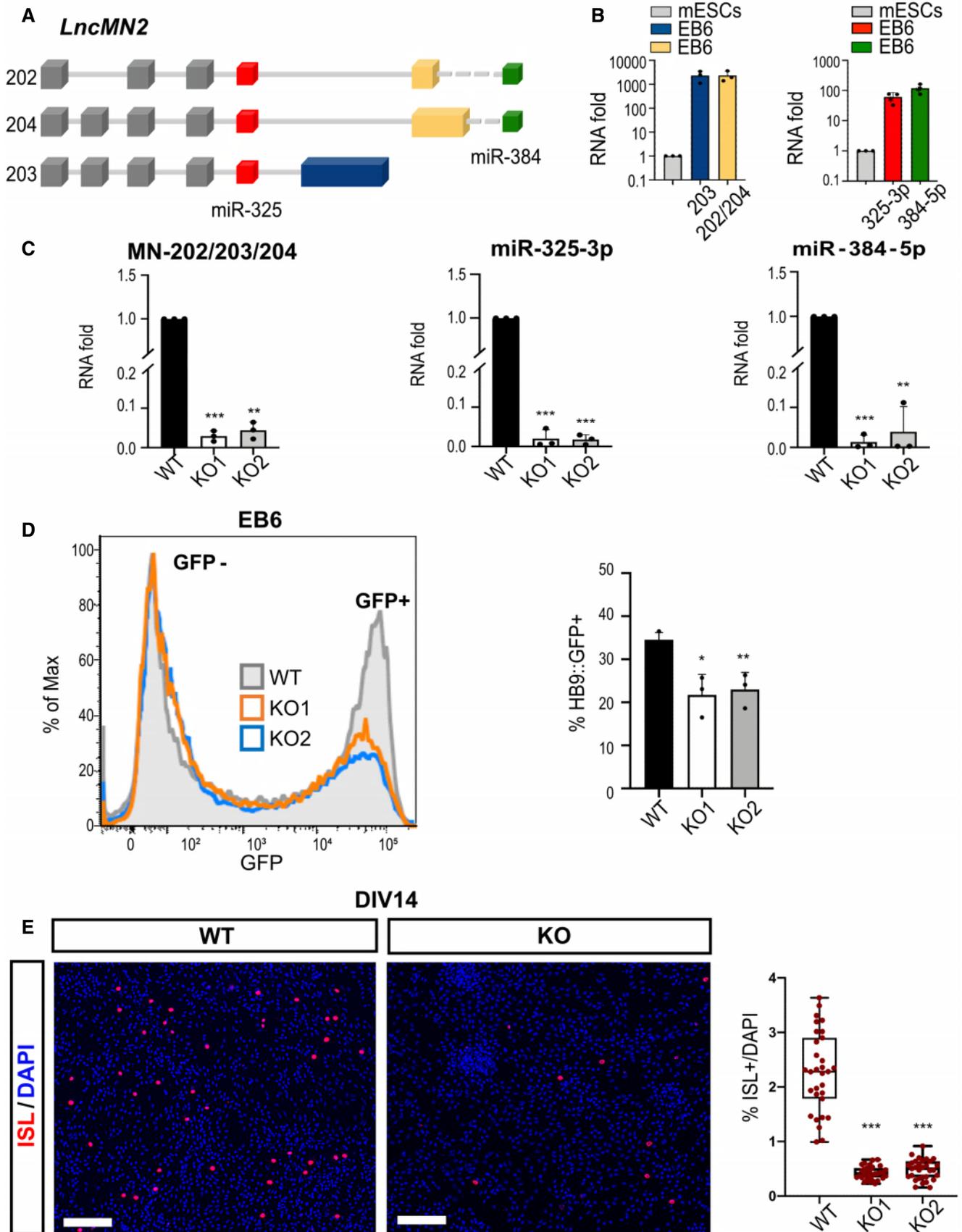


Figure 1.

qRT-PCR analyses performed on RNA extracted from two independent clones (KO1 and KO2) at EB6, revealed a drastic reduction in all the lncMN2 isoforms (Fig 1C, left panel). Notably, the levels of miR-325-3p (Fig 1C, middle panel) and miR-384-5p (Fig 1C, right panel) were also concomitantly reduced in both clones, confirming their common origin from the lncMN2 transcriptional unit.

To investigate the possible impact of lncMN2 ablation on MN production, both WT and lncMN2 KO mESCs were differentiated into MNs and analyzed by flow cytometry (Fig 1D, left panel and Appendix Fig S1A). Interestingly, after 6 days of differentiation (EB6), a 25–30% decrease in (GFP<sup>+</sup>) MNs was observed for the lncMN2 KO (Fig 1D, right panel). We then analyzed postmitotic MNs 14 days (DIV14) after EB6 dissociation, when they have undergone functional and morphological maturation. Immunostaining for Isl1, a major motor neuron marker (Arber *et al*, 1999), proved an even stronger reduction in the amount of *bona fide* MNs (80%) at this time point (Figs 1E and Appendix Fig S1B), strengthening the important role of the MN2 locus in the control of MN differentiation and maturation.

To evaluate the possibility that the miRNAs embedded in the lncMN2 locus are responsible for the observed decrease in MN differentiation, we generated by CRISPR/CAS9 mESC lines carrying the deletion of miR-325 ( $\Delta$ 325, Fig EV2A and B) or miR-384 ( $\Delta$ 384, Fig EV2C and D) coding regions. Correct depletion of the miRNAs in the corresponding clones was confirmed by qRT-PCR (Fig EV2E and F, left panels). Analysis of the *Hb9::GFP<sup>+</sup>* cell populations at day 6 of differentiation (EB6), showed no differences between the miRNA depleted clones and the WT (Fig EV2E and F, right panels), indicating that the depletion of miR-325-5p and miR-384-3p does not affect the MN differentiation potential, as instead observed in the KO clones.

### lncMN2-203 interacts with miR-466i-5p

An *in silico* inspection of lncMN2-203 through the miRanda algorithm (Miranda *et al*, 2006) showed a region embedded in the fifth exon displaying a high density of putative miRNA response elements. Among the predicted miRNAs, we filtered for those with the best Total Energy and with the highest number of binding sites (Appendix Table S1). miR-466i-5p and miR-669a-5p resulted in the best candidates both for the number of interactions and strength of

pairing (Fig 2A). These miRNAs belong to the miR-467 family and are both expressed in mESCs and EB6 (Fig 2B); however, with respect to miR-669a-5p, miR-466i-5p shows a higher degree of complementarity with lncMN2-203 (average  $\Delta$ G of  $-25.6$  Kcal/mole for miR-466i-5p-binding sites and  $-21.2$  Kcal/mole for miR-669a-5p). The presence of a putative region with multiple miRNA-binding sites, together with the enrichment of the lncRNA in the cytoplasm of neural cells (Fig EV1E), raised the intriguing hypothesis that lncMN2-203 and miR-466i-5p/miR-669a-5p may participate into a common competing endogenous RNA regulatory circuitry (Cesana *et al*, 2011; Salmena *et al*, 2011). The occurrence of such interaction is supported by the prediction of a locally unfolded secondary structure surrounding the putative miRNA-binding region by *mfold* (<http://www.unafold.org/>, Zuker, 2003; Fig EV3A).

To demonstrate the ability of lncMN2-203 to bind the predicted miRNAs, a luciferase reporter assay was established in the MN-like NSC-34 cell line (Eggett *et al*, 2000). To this aim, a  $\sim 1,500$  bp-long sequence from the fifth exon of the MN2 gene, including the putative miRNA-binding region, was cloned downstream of the Renilla Luciferase Open Reading Frame (Luc-WT, Fig 2C). In addition, a mutant derivative with a 260 nt-long deletion overlapping the region containing the miRNA-binding sites was similarly generated (Luc-ASP, Fig 2C). Luc-WT and Luc-ASP constructs were ectopically expressed in NSC-34 cells along with locked nucleic acids (LNA) specific for miR-466i-5p and miR-669a-5p. Scrambled LNA (LNA SCR) was used as a control. Cells co-transfected with Luc-WT and LNA against miR-466i-5p (LNA 466i) exhibited  $\sim 50\%$  increase in luciferase activity with respect to scrambled LNA (Fig 2D). Vice versa, no alteration was observed when LNA against miR-669a-5p (LNA 669a) was used. In accordance with the concept that  $\Delta$ SP construct is unable to bind the endogenous miR-466i-5p, cells co-transfected with the Luc-ASP construct and scrambled LNA displayed a significant increase in the basal level of luciferase compared with the WT reporter (Fig 2D). Along this line, the Luc-ASP construct did not show any variation in luciferase activity in the presence of LNAs against miR-466i-5p or miR-669a (Fig 2D). These results indicate that only miR-466i-5p productively interacts with lncMN2-203 in the predicted region.

Along this line, the ability of lncMN2-203 to associate with the miRNA-containing machinery was further tested in EB6 by a

**Figure 2. miR-466i-5p interacts with lncMN2-203.**

- Block-and-line model of the lncMN2-203 structure. Details as in Fig 1A. The miRNA-interacting region in exon5 (dark blue) is represented as a light blue block. Sequences of lncMN2-203 predicted miRNA-binding sites and of miR-466i-5p (above) or miR-669a-5p (below) are indicated in the textbox.
- qRT-PCR expression analysis of miR-466i-5p and miR-669a-5p in mESC and EB6. Data, which are expressed as mean (error bars represent SD), are normalized over *snoRNA U25* levels and reported as RNA fold over the control (mESC), set as 1.  $N = 3$  biological replicates.
- Schematic representation of the luciferase-based reporter constructs. 1,500 bps of WT lncMN2-203 exon5 (dark blue) was cloned downstream to the Renilla Luciferase ORF, represented in gray (Luc-WT). The predicted miRNA-binding region is light blue. In the  $\Delta$ SP mutant (Luc-ASP), a 260 bp deletion was introduced. As indicated, each construct was co-transfected with LNAs targeting miR-466i-5p or miR-669a-5p.
- Quantification of Renilla Luciferase activity in NSC-34 cells co-transfected with the Luc-WT or the Luc-ASP construct and LNA against miR-466i-5p or miR-669a-5p. The histogram represents the mean luciferase activities (error bars represent SD) folded over the control, set as 1 (basal activity from the Luc-WT in the presence of scrambled LNA).  $N = 3$  biological replicates.
- Ago2 cross-linked RNA immunoprecipitation assay performed on total extracts from WT EB6. qRT-PCR quantification of lncMN2 enrichments (isoform  $-203$  or  $-202/204$ ) in the IP and IgG samples, relative to the input is shown. *Ptbp1* and *Gapdh* mRNAs were used as positive and negative controls, respectively. Values are expressed as input percentages and error bars represent SD.  $N = 3$  technical replicates.
- RNA pull-down assay performed on total extracts from differentiated WT EB6. qRT-PCR quantification of lncMN2-203 (F), miR-466i-5p (G), miR-669a-5p (H) or miR-325-3p (I) enrichments in the lncMN2-203 specific (MN2) or in the control (U1) pull-down samples. Values are expressed as input percentages and error bars represent SEM.  $N = 3$  biological replicates.

Data information: Histogram dots represent single replicates. \* $P \leq 0.05$ , \*\* $P \leq 0.01$ , \*\*\* $P \leq 0.001$ , n.s.  $> 0.05$  (two-tailed, unpaired Student's *t*-test).

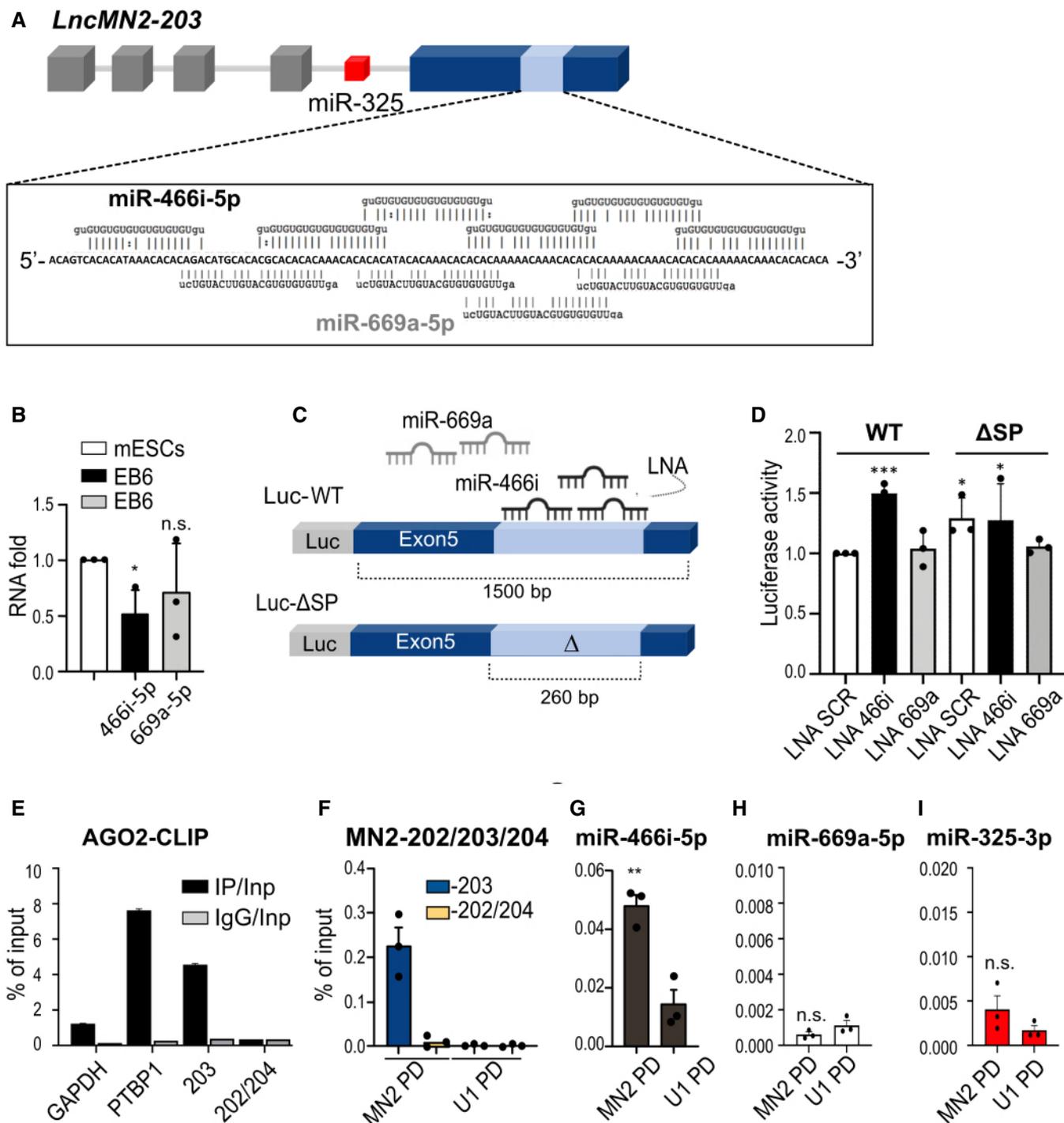


Figure 2.

crosslinking immunoprecipitation (CLIP) analysis for the Argonaute (Ago2) protein (Fig EV3B). Quantification by qRT-PCR of the Ago2 co-precipitated RNAs revealed the specific enrichment for *LncMN2-203*, as compared to *Gapdh* and *Ptpb1* (Engels *et al*, 2012) mRNAs utilized as negative and positive controls, respectively (Fig 2E). Interestingly, we found no enrichment for the 202 and 204 isoforms, indicating that these predominantly nuclear species (Fig EV1E), do not contribute to the interaction with Ago2. Furthermore,

to demonstrate the physical interaction between *LncMN2-203* and miR-466i-5p, we applied an RNA pull-down approach (Ribeiro *et al*, 2018). For this purpose, biotinylated antisense probes (Appendix Table S2) were used to co-purify *LncMN2-203* and its RNA interactors from EB6 total extracts; as a control, we used probes for the U1 snRNA pull-down (Desideri *et al*, 2020). The qRT-PCR analysis confirmed the specific enrichment of *LncMN2-203* (Fig 2F) or U1 snRNA (Fig EV3C) in the respective pull-downs. Notably, upon

lncMN2-203 pull-down a significant enrichment of miR-466i-5p was obtained (Fig 2G) with respect to the U1 pull-down sample. Conversely, no specific enrichment was found for miR-669a-5p (Fig 2H) and miR-325-3p (Fig 2I), used as miRNA specificity controls. Overall, these experiments demonstrate the physical interaction between lncMN2-203 and miR-466i-5p in a RISC complex.

### A CRISPR/Cas9 approach to study the role of the miR-466i-5p interacting region of lncMN2-203

Since the lncMN2-203 isoform showed the interesting feature of interacting with a miRNA expressed in MN, we set up a CRISPR/Cas9 strategy to specifically delete the genomic region carrying the sequence of miR-466i-5p complementarity (Fig 3A and Appendix Fig S2A). Two independent mutants ( $\Delta$ SP1 and  $\Delta$ SP2) were tested for the expression of the different RNAs produced by the locus at EB6. qRT-PCR analysis revealed that such mutation did not affect the expression of any of the three lncMN2 isoforms nor the expression of miR-325-3p, miR-384-5p (Fig 3B), and miR-466i-5p (Appendix Fig S2B).

We then analyzed by cytofluorimetry the percentage of MNs obtained at EB6 and we found a 25–30% decrease in GFP<sup>+</sup> MNs in both  $\Delta$ SP clones, compared with the WT control (Fig 3C). Immunofluorescence analyses of Isl1 at 14 days after EB6 dissociation (DIV14) (Fig 3D, left panel and Appendix Fig S2C) revealed an even stronger effect on the number of MNs (80% reduction (Fig 3D, right panel)). These results show that the deletion of the miR-466i-5p interacting region provides an impairment of MN production similar to the KO clones. Since the knock-out of miR-325 and miR-384 had no effect on the efficiency of MN differentiation (Fig EV2E and F), these data suggest that the interaction of lncMN2-203 with miR-466i-5p plays a crucial function for MN maturation.

Notably, as a further proof that MN production was specifically impaired in both the KO and  $\Delta$ SP clones, we noticed a decrease in the MN markers *Islet1* and *Onecut2* (Roy *et al*, 2012) in the bulk EB6 RNA population (Appendix Fig S2D).

### Single-cell RNA sequencing reveals specific defects in MN lineage commitment

In order to discriminate how the KO and  $\Delta$ SP mutations impinge on MN differentiation and to identify the circuitries that could be

affected in the two cases, we proceeded to a single-cell RNA sequencing (SC-seq) analysis. For this purpose, duplicate cultures of mutant and control mESC were differentiated towards MN and the RNA samples, collected at EB6, were analyzed by 10X Genomics high-throughput single-cell sequencing (Zheng *et al*, 2017). Cluster identity assignment was performed by using the cell subpopulation markers previously described by Rizvi *et al* (2017) (Fig EV4A and B, and Materials and Methods). In agreement with this analysis, we identified, as main cell populations, neural precursors (NP), MN progenitors (MNP), early MN (EMN), and late MN (LMN). In addition, a small class of neurons displaying interneuron (IN) markers was detected. Finally, a group of cells that could not be assigned to any of the above identities (NA) was also found (Fig EV4A and Dataset EV1). Figure 4A shows the differentiation trajectories of WT,  $\Delta$ SP, and KO clones obtained using the Monocle algorithm (Trapnell *et al*, 2014; Haghverdi *et al*, 2018; after dataset integration Appendix Fig S3A and B, and Fig EV4C). A parallel classification of the dataset was performed to distinguish proliferating and postmitotic subpopulations: the first group was identified by the expression of *Mki67* or *Ccnb2* markers, whereas the postmitotic one was assigned through markers of late neuronal differentiation (*Slc18a3*, *Mnx1*, *Gata3*, *Vsx2*, or *En1*) (Fig EV4A and Dataset EV1). The Gene Ontology (GO) terms of enriched genes in these two classes of cells confirmed their proliferative *versus* differentiative conditions (Appendix Fig S4A). The quantification of cells in these subpopulations highlighted an increase in proliferating cells and a decrease in postmitotic ones in  $\Delta$ SP and KO clones (Fig 4B).

Subpopulation analysis (Fig 4C) showed that both mutations strongly affected the number of LMN (red bars) confirming the data obtained by cytofluorimetric analysis on EB6 cells (Figs 1D and 3C) and by immunofluorescence on DIV14 MN (Figs 1E and 3D). In parallel to the LMN decrease, a significative increase in MNP cells (green bars) was observed in KO mutants. Finally, we did not observe any significant variations for all the other populations (Fig 4C).

The heatmap in Fig 4D and the dot plot in Fig EV4D show that the markers used for cluster analysis (in red) enabled to properly classify the different cellular subpopulations since their identity was confirmed by the expression of cell type-specific signatures (Briggs *et al*, 2017; Rizvi *et al*, 2017; Sagner *et al*, 2018).

Overall, these data suggest that the two mutations similarly affect the yield of LMN, while the KO clones display an additional effect on the number of MNP cells.

#### Figure 3. lncMN2 $\Delta$ SP mutation impairs MN production.

- A Block-and-line model of the lncMN2 edited locus. Details as in Figs 1A and 2A. Predicted miRNA-binding site deletion is represented as  $\Delta$ SP.
- B qRT-PCR expression analysis of lncMN2-203 isoform, lncMN2-202/204, miR-325-3p, and miR-384-5p embedded miRNAs in WT,  $\Delta$ SP1, and  $\Delta$ SP2 EB6. Data, which are expressed as mean (error bars represent SD), are normalized over *Atp5o* mRNA levels (for lncRNA isoform quantification) or over *snoRNA U25* levels (for miRNA quantification) and are reported as RNA fold over the control (mESC), set as 1.  $N = 3$  biological replicates.
- C Cytofluorimetric analysis of cells from dissociated EB6. The right diagram reports the cell distribution (percentage of the peak) according to GFP levels, from WT (gray line),  $\Delta$ SP1 (purple line), and  $\Delta$ SP2 (blue line) EB6. In the histogram aside, the percentage of GFP<sup>+</sup> cells in WT,  $\Delta$ SP1, and  $\Delta$ SP2 is reported. Data are expressed as mean, and error bars represent SD.  $N = 3$  biological replicates.
- D Left: representative immunofluorescence of the MN marker *Isl1* (red signals) in cells maintained *in vitro* for 14 days after EB6 dissociation (DIV14). DAPI counterstaining in blue. Scale bar = 100  $\mu$ m. Right: box plot shows the distribution of *Isl1*<sup>+</sup> nuclei with respect to the total cell number marked with DAPI staining in WT and in  $\Delta$ SP mutants (clone 1 and clone 2). The central band represents the median, and the whiskers represent the distribution that goes from the minimum value to the lower quartile than from the upper quartile to the maximum value.  $N = 3$  biological replicates and 10 fields were analyzed for each replicate.

Data information: Histogram dots represent single-field replicates. \* $P \leq 0.05$ , \*\* $P \leq 0.01$ , \*\*\* $P \leq 0.001$ , n.s. > 0.05 (two-tailed, unpaired Student's t-test). Source data are available online for this figure.

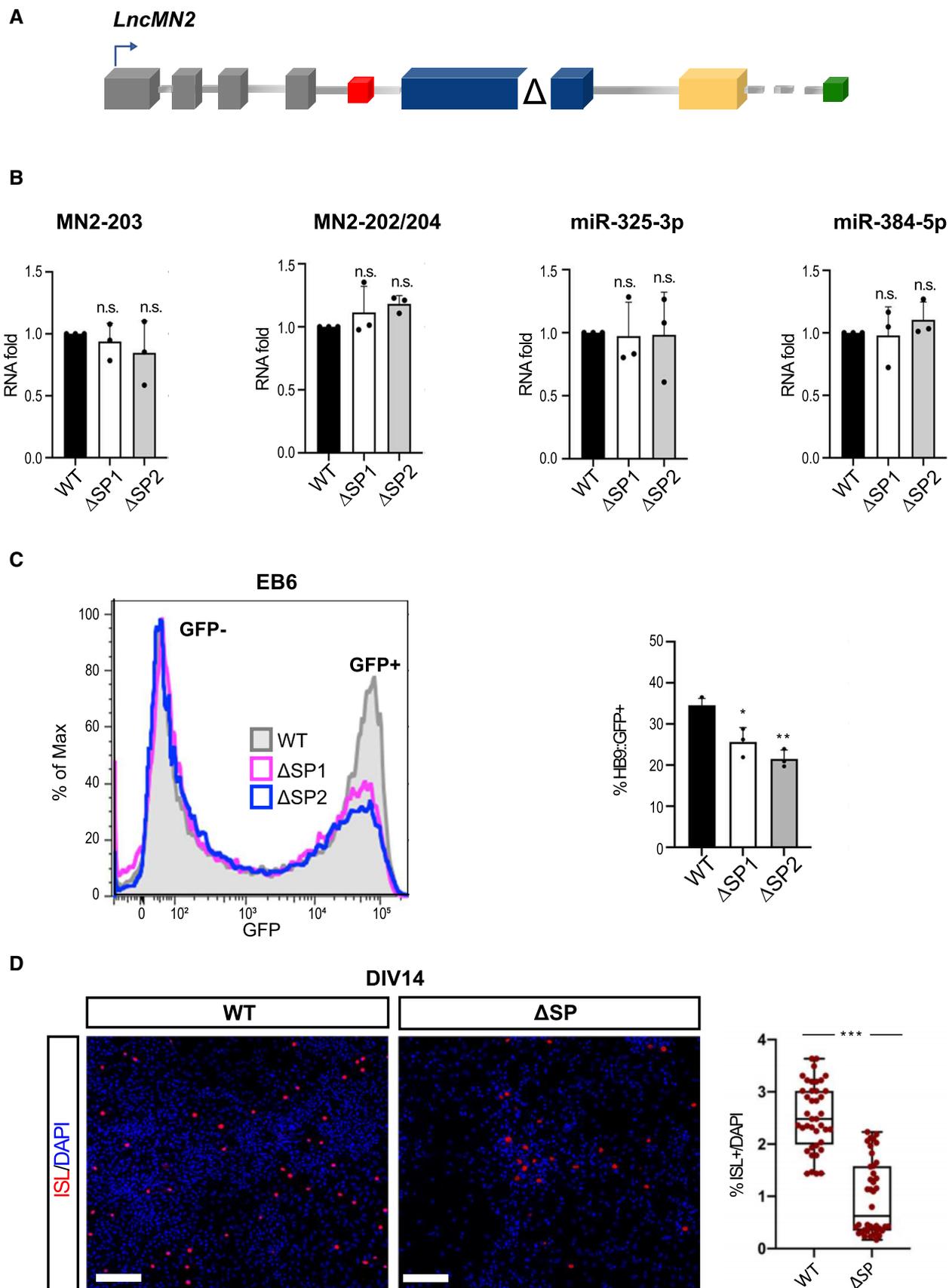


Figure 3.

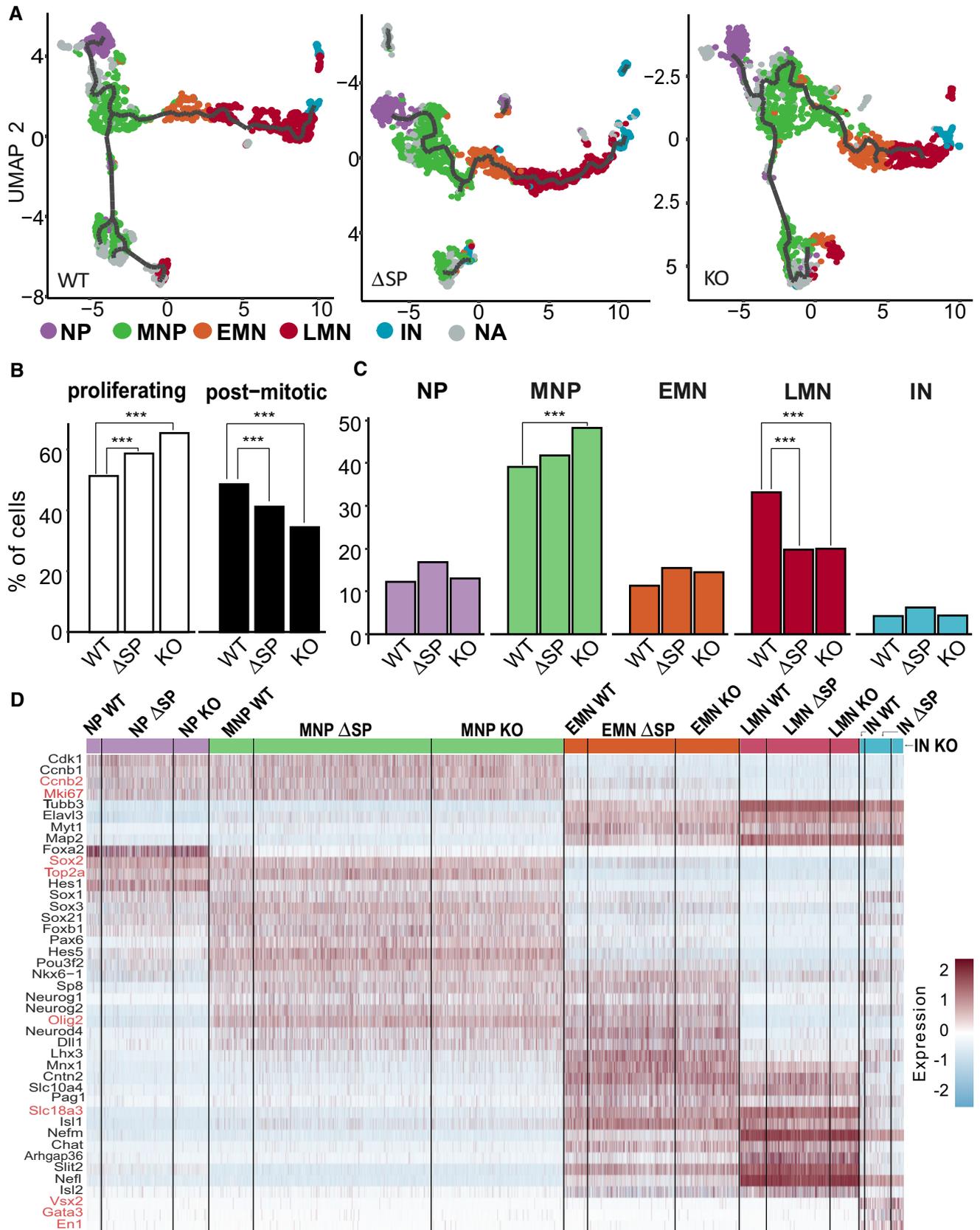


Figure 4.

**Figure 4. SC-seq reveals the implication of lncMN2 in motoneurogenesis.**

- A UMAP visualization plots for WT2 (1,435 cells),  $\Delta$ SP1 (2,039 cells), and KO1 (1,955 cells) produced by Monocle3 (Trapnell *et al.* 2014). The plots present the estimated trajectories (dark gray), and the colors correspond to the cell identity: neural precursors NP cells (purple), motor neurons progenitors (MNP, green), early motor neurons (EMN, orange), late motor neurons (LMN, red), and interneurons (IN, light blue). The cells not assigned to any of the previous groups (NA) are indicated in gray.
- B Histograms reporting the percentage of cells detected as proliferating and postmitotic (Dataset EV1). All percentages of cells were calculated, for each condition, on samples belonging to the same batch (WT2,  $\Delta$ SP1,  $\Delta$ SP2, KO1, and KO2) excluding not assigned cells. The significance of cell subpopulation differences between WT and mutants was calculated with the Fisher's exact test (see Materials and Methods). \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , n.s.  $> 0.05$ ;  $N = 2$  biological replicates.
- C Histograms reporting the percentage of cells assigned to NP, MNP, EMN, LMN, and IN (Dataset EV1). All percentages of cells were calculated, for each condition, on samples belonging to the same batch (WT2,  $\Delta$ SP1,  $\Delta$ SP2, KO1, and KO2) excluding not assigned cells. The significance of cell subpopulation differences between WT and mutants was calculated with the Fisher's exact test (see Materials and Methods). \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , n.s.  $> 0.05$ .  $N = 2$  biological replicates.
- D Heatmap generated by the Seurat *DoHeatmap* function (Stuart *et al.* 2019) represents the expression of cell identity markers (listed on the left) for each cell of the identified subpopulations. Genes used for the assignment of cluster identity are marked in red. Dataset samples are indicated on the top. Red represents the maximum expression value, light blue the minimum.

**Gene Ontology enrichment analysis of WT and mutant transcriptomes**

To identify the genes deregulated in MNP, EMN, and LMN cells, we performed a differential gene expression (DGE) analysis comparing WT and mutant clones ( $\Delta$ SP and KO) using the Seurat software (Stuart *et al.* 2019).

This analysis (Dataset EV2) identified: 1,386 differentially expressed genes in KO (994 up- and 392 downregulated) and 756 in  $\Delta$ SP (462 up- and 294 downregulated) in the MNP subpopulation (Fig 5A); 582 genes deregulated in KO conditions (293 up- and 289 downregulated) and 306 in  $\Delta$ SP (140 up- and 166 down clones) in EMN cells (Fig EV5A), and 1,301 differentially expressed genes (782 up- and 519 downregulated) in KO and 813 (451 up- and 362 downregulated) in  $\Delta$ SP for LMN (Fig 5B).

To identify the affected pathways, a GO enrichment analysis was performed on deregulated mRNAs, using the WebGestalt software (Liao *et al.* 2019). Among the upregulated categories in MNP, we found “cell cycle” genes (Fig 5C, green dots); interestingly, in this class, we found a vast excess of genes specifically deregulated only in the KO mutant (Fig 5D and Dataset EV3). Interestingly, we found the “cell cycle” category also in the upregulated genes in EMN of KO mutant (Fig EV5B).

Among the downregulated genes in LMN of both mutants, we found the “nervous system development” category (Fig 5E, red dots). Conversely, to the previous category, these genes were largely in common between the two mutants (Fig 5F).

These data suggest that the reduced amount of LMN cells in  $\Delta$ SP and KO correlates with the decreased expression of neuronal

differentiation regulators, while the increased number of MNP observed in KO clones could be related to an increased number of cell cycle-related genes specifically altered in this mutant.

These peculiar differences in gene expression inside distinct cell subpopulations would not be underscored on total RNA sequencing.

**Molecular circuitries altered in lncMN2 mutants**

To enter more mechanistically into the molecular pathways affected in the two lncMN2 mutants and to distinguish their contribution to the observed phenotypes, we intersected the differentially expressed genes from the transcriptomes of  $\Delta$ SP and KO clones. Since the KO mutation includes the  $\Delta$ SP one, all the common affected pathways could be ascribed to the ability of lncMN2-203 to bind miR-4661-5p, while alterations observed only in the KO could be attributed to the other components of the lncMN2 transcriptional unit. Even if we cannot exclude a function for the nuclear lncMN2 isoforms (202 and 204), we decided to focus on cytoplasmic activities searching for pathways altered as a consequence of miR-325-3p and miR-384-5p depletion.

In order to identify the pathways regulated by these two miRNAs, we searched for predicted miR-325-3p and miR-384-5p targets among the transcripts significantly upregulated only in the KO clones (Fig 6A, Appendix Fig S5A, and Dataset EV3). Interestingly, in the MNP cluster, we found several genes with a relevant role in sustaining cell proliferation. Six representative species (Ait-Si-Ali *et al.* 2004; Saxe *et al.* 2013; Chiu *et al.* 2016; Wang *et al.* 2018; Zhou *et al.* 2018; Rispoli *et al.* 2019) were selected and validated as targets of either miR-325-3p or miR-384-5p in NSC-34 cells

**Figure 5. GO enrichment analyses in MNP and LMN.**

- A Volcano plots representing differentially expressed genes between WT MNP and  $\Delta$ SP MNP (left) and WT MNP and KO MNP (right). Each dot represents a gene;  $-\log_{10} P$  value is represented in y-axis while average  $\log_2 FC$  in x-axis. Significantly deregulated genes are depicted as black dots.
- B Volcano plots representing differentially expressed genes between WT LMN and  $\Delta$ SP LMN (left) and WT LMN and KO LMN (right). Each dot represents a gene.  $-\log_{10} P$ -value is represented in y-axis while average  $\log_2 FC$  in x-axis. Significantly deregulated genes are depicted as black dots.
- C GO-enriched categories in MNP related to downregulated (left) and upregulated (right) genes in  $\Delta$ SP (top) and KO (bottom) mutants. Green dots indicate “cell cycle” category.
- D Bar plot representing the number of “cell cycle”-related genes upregulated in both  $\Delta$ SP and KO MNP (light green) or specifically in either one of the two mutants (dark green).
- E GO-enriched categories in LMN cells related to downregulated (left) and upregulated (right) genes in  $\Delta$ SP (top) and KO (bottom) mutants. Red dots indicate “nervous system development” category.
- F Bar plot representing the number of “nervous system development”-related genes downregulated in both  $\Delta$ SP and KO (dark red) or specifically in either one of the two mutants (light red).

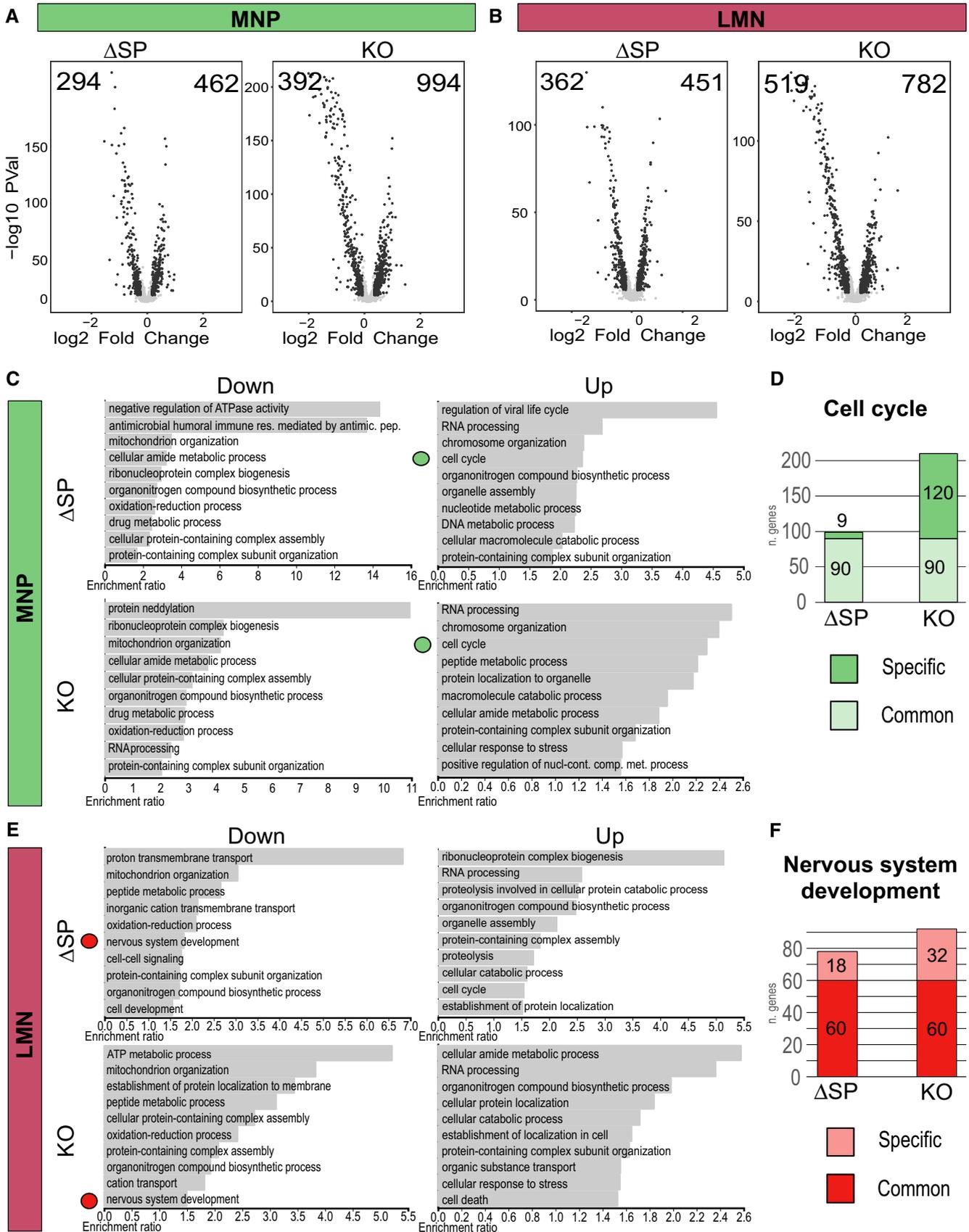


Figure 5.

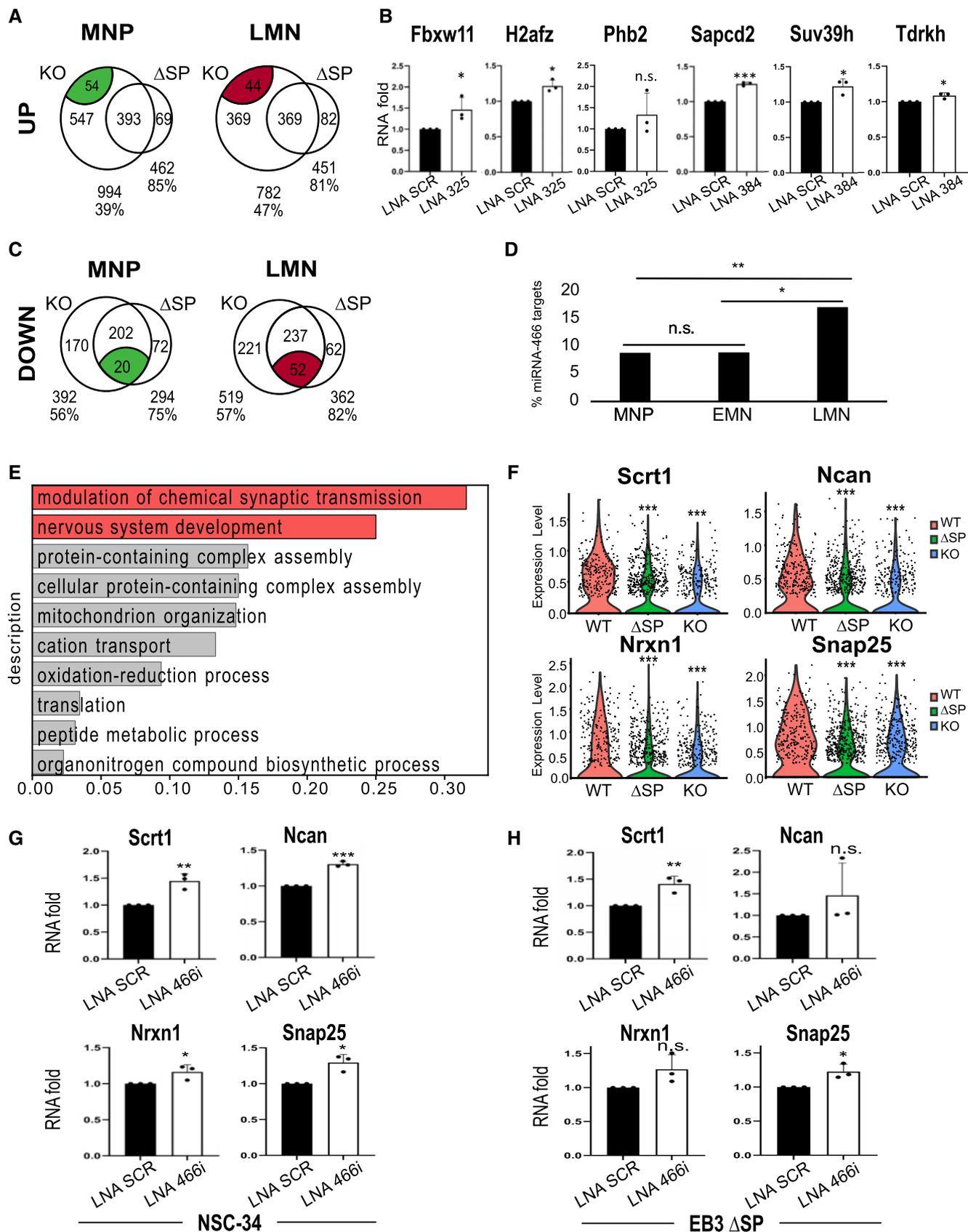


Figure 6.

**Figure 6. Neuronal differentiation genes respond to miR-466i-5p activity.**

- A Venn diagram showing the relationships between genes upregulated in KO and  $\Delta$ SP mutants. Gene numbers and percentages of overlap are shown below diagrams for both groups. The intersection between miR-325-3p and miR-384-5p predicted targets and genes specifically upregulated in KO mutants are indicated in green (MNP, left) or in red (LMN, right).
- B qRT-PCR expression analysis of *Fbxw11*, *H2afz*, and *Phb2* in NSC-34 transfected with miR-325-3p LNA and *Sapcd2*, *Suv39h*, and *Tdrkh* in NSC-34 transfected with miR-384-5p LNA. Data, which are expressed as mean (error bars represent SD), are normalized over *Atp5o* mRNA levels and reported as RNA fold over the control (scrambled LNA, CTRL), set as 1.  $N = 3$  biological replicates.
- C Venn diagram showing the relationships between genes downregulated in KO and  $\Delta$ SP mutants. Gene numbers and percentages of overlap are shown below diagrams for both groups. The intersection between miR-466i-5p-predicted targets and genes downregulated in both mutants are indicated in the MNP (left) and LMN (right) subpopulations as the green and red area, respectively.
- D Bar plot depicting for each analyzed cell subpopulation (MNP, EMN, and LMN) the percentage of miR-466i-5p targets among genes downregulated in both  $\Delta$ SP and KO mutants. Statistical significance was assessed using the Fisher's exact test.  $*P \leq 0.05$ ,  $**P \leq 0.01$ , n.s.  $> 0.05$ .
- E Bar plot depicting the ratio of miR-466i-5p predicted targets (x-axis) in GO-enriched categories of genes downregulated in both  $\Delta$ SP LMN and KO LMN. Neural categories ("modulation of chemical synaptic transmission" and "nervous system development") are colored in red. GO enrichment analysis was performed using WebGestalt software.
- F Violin plots representing for each condition (WT,  $\Delta$ SP, and KO) the expression of *Scrt1*, *Ncan*, *Nrxn1*, and *Snap25* in LMN. Plots were produced by Seurat "VlnPlot" function.  $N = 2$  biological replicates. Statistical significance was assessed using the Wilcoxon rank-sum test (default Seurat FindMarkers function).  $***P \leq 0.001$ .
- G qRT-PCR expression analysis of *Scrt*, *Ncan* (up) and *Nrxn1*, *Snap25* (down) in NSC-34 cells transfected with miR-466i-5p LNA. Data, which are expressed as mean (error bars represent SD), are normalized over *Atp5o* mRNA levels and reported as RNA fold over the control (scrambled LNA, CTRL), set as 1.  $N = 3$  biological replicates.
- H qRT-PCR expression analysis of *Scrt*, *Ncan* (up) and *Nrxn1*, *Snap25* (down) in  $\Delta$ SP mESCs transfected with miR-466i-5p LNA and differentiated to early EBs (day 3). Data, which are expressed as mean (error bars represent SD), are normalized over *Atp5o* mRNA levels and reported as RNA fold over the control (scrambled LNA, CTRL), set as 1.  $N = 3$  biological replicates.
- Data information: Histogram dots represent single replicates.  $*P \leq 0.05$ ,  $**P \leq 0.01$ ,  $***P \leq 0.001$ , n.s.  $> 0.05$  (two-tailed, unpaired Student's *t*-test).

treated with LNAs against either one or the other miRNA (Fig 6B). These data allowed us to conclude that the role of miR-325-3p and miR-384-5p is mainly related to the repression of pathways that control proliferation more than differentiation, in agreement with the phenotypic data of the corresponding KO clones.

We next proceeded to the analysis of the pathways controlled by lncMN2-203. According to its ability to bind and in turn to compete for miR-466i-5p, the genes found to be downregulated should trace back to the activity of this miRNA; moreover, they should be commonly downregulated in  $\Delta$ SP and KO mutants.

At first, in order to identify the subpopulation displaying higher responsiveness to miR-466i-5p, we compared the number of miR-466i-5p predicted targets in downregulated genes common to MNP, EMN, and LMN (Fig 6C and Appendix Fig S5B).

We noticed that LMN shows a significantly larger overlap (17.9%) if compared with MNP (9%) and EMN (9%) (Fig 6D). GO enrichment analysis of commonly downregulated genes in LMN revealed as enriched categories those linked to neuron physiology ("modulation of chemical synaptic transmission") and development ("nervous system development"); moreover, both categories displayed the highest proportion of bona fide miR-466i-5p targets (Fig 6E and Dataset EV3).

Among the genes deregulated in this group and harboring binding sites for miR-466i-5p, we found interesting regulators of neuronal functions: *Scrt1*, is a transcription factor with a prominent role in neuronal fate commitment and in neuron generation and maintenance (Nakakura *et al*, 2001; Itoh *et al*, 2013; Matsuda *et al*, 2019), *Ncan* is a member of the CSPG (chondroitin sulfate proteoglycan) family, which modulates neuronal adhesion, axon guidance and neurite growth (Schmidt *et al*, 2020), *Nrxn1* mediates efficient neurotransmission and formation of synaptic contacts (Lam *et al*, 2019) and *Snap25* is a presynaptic plasma membrane SNARE protein involved in the regulation of neurotransmitter release (Roballo *et al*, 2019).

All four transcripts resulted to be significantly downregulated in  $\Delta$ SP and KO clones (Fig 6F and Dataset EV2). To validate the

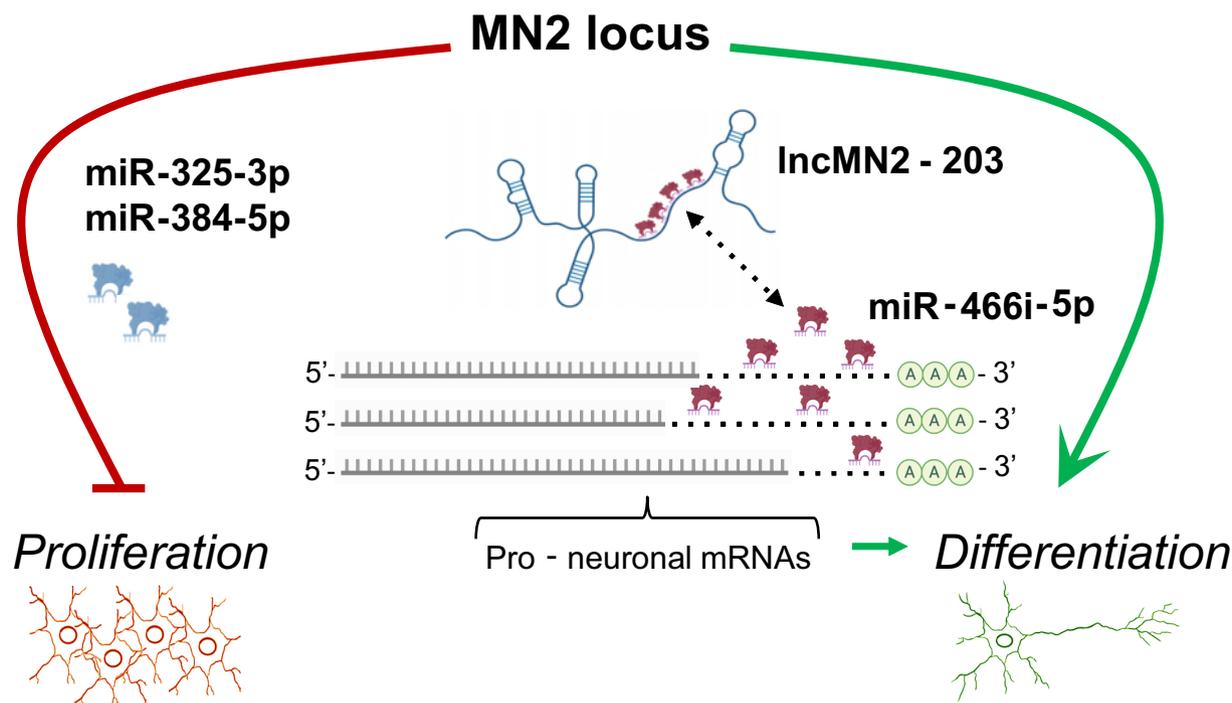
responsiveness of these factors to miR-466i-5p, we transfected the MN-like NSC-34 cell line with LNAs against miR-466i-5p and tested their expression levels. Fig 6G shows that all four mRNAs were upregulated in a significant manner upon inhibition of miR-466i-5p. We then transfected the same LNAs in  $\Delta$ SP mESCs and induced the cells to differentiate to EB3. Also in this case we observed the upregulation of *Scrt1*, *Ncan*, *Nrxn1*, and *Snap25* mRNAs (Fig 6H). Transfection of NSC-34 cells with an LNA targeting mir-669a-5p, which is unable to interact with lncMN2-203 (Fig 2H), even though its sequence largely overlaps with that of miR-466i-5p (Fig 2A), did not produce any variation in the expression levels of the target mRNAs (Appendix Fig S5C), thus supporting the specificity of miR-466i-5p LNAs effect. As a further control, we performed a complementary experiment by using miRNA Mimics to overexpress miR-466i-5p in NSC-34 cells. In these conditions, we observed a significant reduction in *Scrt1*, *Ncan*, and *Nrxn1* (Appendix Fig S5D).

Finally, the ectopic overexpression in NSC-34 cells of lncMN2-Ex5, containing the miR-466i-5p sponging region, showed for *Scrt1* and *Ncan* a clear trend of upregulation (Appendix Fig S5E).

Altogether, these data support a mechanism by which lncMN2-203 could act as a sponge for miR-466i-5p thus controlling its target mRNAs; moreover, they indicate that the ability of lncMN2-203 to sequester miR-466i-5p plays alone an important function in the control of MN differentiation mainly by impacting on regulators of nervous system development and function.

Interestingly, RNAi-mediated downregulation of *Scrt1* in mouse ESC-derived EBs (day 3) caused a strong and significant downregulation of genes crucially associated with MN differentiation and function (i.e., *Onecut2* and *Islet1*) (Appendix Fig S5F).

These results highlight the activity of *Scrt1* in MN generation and provide additional evidence on the relevant regulatory function of lncMN2-203 in MN development. Analysis of the relative abundance of the different transcripts in the SC-seq data, combined with the number of their miRNA-binding sites, indicates a stoichiometry compatible with the proposed sponging mechanism (Dataset EV3).



**Figure 7.** Schematic representation of the circuitry linking lncMN2-203, miR-466i-5p, miR-325-3p, miR-384-5p, and MN differentiation. The model was created with BioRender.com

In addition to this pathway, further ones which explain the differential KO mutant phenotype act through miR-325-3p and miR-384-5p in controlling cell proliferation-related genes. Even though we may hypothesize additional roles (He *et al*, 2021) for lncMN2 in the nucleus, more than 90% of the genes included in the MN2 topologically associated domain were not differentially expressed upon lncMN2 KO (Bonev *et al*, 2017 and Dataset EV3).

In conclusion, the short and long noncoding RNAs embedded in the MN2 locus cooperate to confer robustness to the MN differentiation process by allowing, on one side, the repression of proliferative pathways and, on the other, the activation of differentiative ones (Fig 7).

## Materials and Methods

### Cell culture

mESC were cultured with Embryomax MEM (Chemicon, cat. no. SLM-220-B), ES cell tested fetal bovine serum (FBS; HyClone, cat. no. SH30070.03), 100× Nucleosides (Chemicon, cat. no. ES-008-D), 100× nonessential amino acids (Chemicon, cat. no. TMS-001-C), 10 ng/ml of Leukemia Inhibitory Factor (Chemicon, cat. no. ESG1107); 1 μM FGFR inhibitor (Merk PD173074), 0.1 μM GSK-3 inhibitor (Merck 361559), 1% GlutaMAX (Sigma-Aldrich), 1% 2-mercaptoethanol, 1% Pen/Strep (Sigma-Aldrich).

NSC-34 cells were cultured in a growth medium with DMEM F12 Ham (Sigma-Aldrich cat. no. D6421), 10% FBS (Sigma-Aldrich), 1 mM Sodium pyruvate, 1% GlutaMAX (Sigma-Aldrich), and 1% Pen/Strep (Sigma-Aldrich).

### mESC to MN differentiation protocol

*Hb9::GFP* mESC were cultured and differentiated into MN as described in Wichterle and Peljto (2008), Caputo *et al* (2018). Generation of EB was obtained by culturing mESC in ADNFK medium (1:1 Advanced DMEM/F12: Neurobasal medium, 10% Knockout Serum Replacement (Gibco, 10828028), 1% GlutaMAX, 1% 2-mercaptoethanol, 1% Pen/Strep). On day 2, the ADNFK medium was complemented with 2% B27 Supplement (Gibco, 17504-044), 1 mM RA (Sigma-Aldrich, R2625), and 0.5 mM SAG (Merck Millipore, 566660). EB was expanded and, on day 6, disrupted with Papain dissociation system (Worthington Biochemical Corporation) following the manufacturer's instructions. The mixed population of MNs was plated on Poly-L-ornithine (Sigma P3655) 1% and Laminin, 1% (Sigma L2020), eventually on a glass slide for experimental analysis. Mixed populations were kept in culture with N2B27 medium with 1% N2 Supplement, 2% B27 Supplement medium, 1% nonessential amino acids; neurotrophic factor BDNF (20 ng/ml), GDNF (10 ng/ml), CNTF (10 ng/ml), 200 ng/ml L-ascorbic acid (Sigma-Aldrich) (TMS-001-C) (1% GlutaMAX, 1% 2-mercaptoethanol, 0.5% Pen/Strep).

### Plasmid construction

Luciferase constructs: a 1,500 nt-long region of lncMN2-203 exon 5 was PCR-amplified and cloned downstream of the luciferase stop codon in the psiCHECK 2 vector (Luc-WT). From this construct, a mutant derivative, lacking a 260 nt-long region corresponding to the miR-466i-5p-binding sites, was generated by the CRISP/Cas9 strategy (Luc-ΔSP).

LncMN2-Ex5 construct: the entire sequence of exon 5 (2,844 nt) was PCR-amplified from genomic DNA of WT mESCs using Clone AMP PCR HIFI (Takara-Clontech). The obtained fragment was then cloned into a pcDNA 3.1 + plasmid (previously linearized with inverse PCR using Clone AMP PCR HIFI, Takara-Clontech) through In-Fusion cloning (Takara-Clontech) for expression in mammalian cells. See Appendix Table S2 for oligo sequences.

### Cell transfection

NSC-34 cells were plated (200,000 cells for each well of a 12 well plate) and transfected 24 h later with 100 nM of each miRNA LNA (Appendix Table S2) using Lipofectamine 2000 (Thermo Fisher Scientific) according to the manufacturer's specifications. 48 h after transfections cells were lysed, and RNA was collected. For the luciferase assays, cells were co-transfected with 100 nM of each miRNA LNA and 20 ng of psiCheck2 plasmid Luciferase fused or not to the miRNA-binding region. Luciferase activity was measured in GloMax-Multi<sup>+</sup> Detection System (Promega), using the Dual-Luciferase Reporter Assay System (Promega). For LncMN2-Ex5 overexpression, cells were transfected with 1 µg of pcDNA 3.1 + plasmid or of LncMN2-Ex5 construct. mESCs were thawed, plated, and transfected 24 h later with 100 nM of each miRNA LNA or siRNA (Appendix Table S2) using Lipofectamine 2000 (Thermo Fisher Scientific) according to the manufacturer's specifications. 24 h after transfections MN differentiation was induced.

### RNA preparation and analysis

Total RNA from cells was extracted with the E.Z.N.A Total RNA Kit I (Omega) and retro-transcribed with PrimerScript RT reagent Kit (Takara). For mRNAs, real-time qRT-PCR analysis was performed with SYBR Green Power-UP (Life Technologies), using the housekeeping *Atp5o* (ATP synthase, H<sup>+</sup> transporting, mitochondrial F1 complex, O subunit) gene as an internal control.

For miRNAs, real-time qRT-PCR analysis was performed with SYBR Green PCR Master Mix (Qiagen) for microRNAs, using the housekeeping gene *Sno25* (Small Nucleolar RNA, C/D Box 25) as the internal control.

### Protein analysis

For Western Blot analysis, proteins were collected in RIPA Protein buffer, loaded on 4–12% bis-tris-acrylamide gel (Thermo Fisher Scientific), and transferred to a nitrocellulose membrane. The membrane was blocked in 10% milk and hybridized with the specific antibodies overnight at 4°C at the appropriate dilutions, according to manufacturers' instructions (see Appendix Table S2 for details). After three washes in TBST, the filter was hybridized with the corresponding secondary antibody for one hour at room temperature. Protein detection was carried out with the Long-Lasting Chemiluminescent Substrate (EuroClone) using ChemiDoc MP System Images were analyzed using Image Lab Software (BioRad).

### Cell fractionation

EB6 were washed with PBS (without Ca<sup>2+</sup> and Mg<sup>2+</sup>) and lysed in buffer A (Tris 20 mM pH 8.0, NaCl 10 mM, MgCl<sub>2</sub> 3 mM, NP40

(IGEPAL) 0.10%, EDTA 0.2 mM, DTT 1 mM, Protease inhibitor cocktail 1×, Ribolock 1×), and incubated on ice for 5 min. Nuclei were pelleted 400 g for 5 min at 4°C. The supernatant, corresponding to the cytoplasmic fraction, was stored on ice. The nuclear pellet was washed twice with 50 µl of buffer A. RNA was then extracted adding 600 µl of RNA Lysis buffer (QuickRNA Miniprep) to both fractions as described in "RNA preparation and analysis".

### Crosslinking Immunoprecipitation (CLIP) assay

Dissociated-EB6 were resuspended in PBS (without Ca<sup>2+</sup> and Mg<sup>2+</sup>) and plated in dishes (10 × 10 cm<sup>2</sup>). Cells were UV cross-linked with 4,000 µJoules/cm<sup>2</sup> energy using a stratalinker and centrifuged at 200 g for 5 min at 4°C. Cell pellets were resuspended in 3 volumes of NP40 lysis buffer pH 7.5 (50 mM Hepes-KOH; 150 mM KCl; 2 mM EDTA; 1 mM NaF; 0.5% NP40; 0.5 mM DTT; 1× PIC) and incubated on ice for 10–15 min followed by centrifugation at 18,000 g for 10 min at 4°C. Resulting cellular lysates were incubated (overnight on a rotating wheel, at 4°C) with 30 µl of Dynabeads Protein G magnetic particles (Invitrogen) preincubated with either 10 µg of Ago2 Antibody (018-22021, Wako) or mouse IgG (sc-2025, Santa Cruz). Beads were washed with a High-Salt buffer (50 mM Hepes-KO; 500 mM KCl; 0.5 mM DTT; 0.05% NP40). Before RNA extraction, 1/5 of the cell lysate was heated for 5 min at 95°C, and the supernatant was collected and resuspended in Protein elution buffer (4× Laemmli sample buffer [BioRad]) with DTT 50 mM and analyzed by Western blot. RNA fraction was treated with Proteinase K (AM2546, Thermo Fisher Scientific) for 30 min at 50°C; the samples were then placed for 10 min at 95°C, and finally, the RNA was extracted using miRNeasy Mini Kit with on-column DNase treatment, according to the manufacturer's instructions (Qiagen).

### RNA pull-down assay

Native RNA pull-down was performed on total extract from EB6 cells. Cells were harvested in PBS and centrifuged at 400 g for 5 min. Cell pellets were lysed in a buffer containing Tris-HCl pH 7.5 50 mM, NaCl 150 mM, MgCl<sub>2</sub> 3 mM, NP40 0.5%, EDTA 2 mM, DTT 1 mM; 1× PIC, and RNase inhibitors. After lysis and clearing by centrifugation, 1 mg of extract was diluted in a 1:2 ratio with hybridization buffer containing Tris-HCl pH 7.5 100 mM, NaCl 300 mM, MgCl<sub>2</sub> 1 mM, SDS 0.2%, Formamide 15%, NP40 0.5%, EDTA 10 mM, DTT 1 mM, 1× PIC, and RNase inhibitors.

10% of the total extract was collected for Input (INP). 100 pmol of previously heat-denatured biotinylated probes were added (see Appendix Table S2 for details). After a 4-h incubation at 4°C, 0.1 ml of streptavidin Magnasphere paramagnetic beads (Promega) were added to pull down the complex, and the mixture was incubated for 1 h at room temperature. After pull-down, beads were washed 4 times with hybridization buffer and RNA was extracted and DNase treated. Pull-down (PD) qRT-PCR results were represented as a percentage of PD/input signal (% of input).

### Immunofluorescence

The mixed population from EB6 was cultured on precoated glass coverslips (0.01% poly-L-ornithine/Murine Laminin 20 µg/ml, Sigma) and then fixed in 4% paraformaldehyde (Electron

Microscopy Sciences, Hatfield, PA) for 15 min at room temperature, washed with PBS and then permeabilized and blocked with 0.2% Triton X-100/1% donkey serum/PBS. Subsequently, cells were incubated with anti-Islet 1/2 primary antibody (clone 39.4D5, DSHB) in 1% BSA/1% donkey serum/PBS overnight at 4°C. After washing with PBS, cells were labeled with secondary antibody (Donkey anti-mouse 647, Invitrogen, A32787) 1% donkey serum/PBS for 45 min at room temperature. Nuclei were counterstained with DAPI solution (Sigma, D9542; 1 µg/ml/PBS) for 5 min at room temperature, and the coverslips were mounted using ProLong Diamond Antifade Mountant (Thermo Fisher Scientific, P-36961). Cells were imaged using an inverted confocal Olympus IX73 microscope equipped with a Crestoptics X-LIGHT V3 spinning disk system and a Prime BSI Express Scientific CMOS camera. The images were acquired as 16-bit 2,048 × 2,048 pixel files by using a Plan CN 10× (NA 0.25) objective and were collected with the MetaMorph software (Molecular Devices).

The percentage of Is1<sup>+</sup> nuclei with respect to the total cell number was performed by using ImageJ software for automatic and manual counting of nuclei. Total cell number examined for each condition was: 180,000 (WT); 79,800 (KO1) and 95,800 (ΔSP), respectively. See Appendix Table S2 for details.

### Flow cytometry

MN was analyzed based on GFP expression levels from the *Hb9::GFP* reporter using a FACSAriaIII (BD Biosciences, San Jose, CA) equipped with a 488 nm laser and FACSDiva software (BD Biosciences version 6.1.3). Data were analyzed using the FlowJo software (Tree Star, Ashland, OR). Briefly, cells first gated based on forward and side scatter area (FSC-A and SSC-A) plots were then detected in the green fluorescence channel for GFP expression (530/50 nm filter).

### CRISPR/Cas9 IncMN2 KO generation

sgRNAs were designed using Zhang design tools at <http://crispr.mit.edu/>. The regions for sgRNA design were selected taking into consideration FANTOM5 TSS CAGE data (<http://fantom.gsc.riken.jp/5/>) and PolII ChIP (ENCODE/LICR TFBS data from Ren Lab). PX330 plasmid, encoding WT Cas9 protein, was purchased from Addgene while sgRNAs were ordered as single-strand DNA oligos (Biofab) and cloned as previously recommended resulting in a vector identified as PX330-sgRNA. HR110PA-1 (System Biosciences) was used as a backbone to create the donor vector (DONOR). Poly (A)/2×MAZ sequence (PAS; Ballarino *et al*, 2018) was cloned into the donor vector. Homology arms (HA) were designed to be longer than 500 nt and amplified by PCR on mESC gDNA (Kapa HiFi, Takara) and cloned in MCS1 (upstream of PAS) and MSC2 for the left and right arms, respectively. mESC cells were transfected on gelatin-coated dishes using reverse transfection with lipofectamine2000 (Life Technologies) following the manufacturer's instruction and transfecting a total of 500 ng of DNA (PX330+sgRNA and DONOR) for 250,000 cells. Two negative controls (PX330 and DONOR; DONOR only) were used. Cells were let grow and then split into 2 sister dishes for further selection. One dish was maintained for one week and then sorted for RFP<sup>+</sup> using FACS with the positivity threshold set on the two negative controls; the sister dish cells

were selected with 1 µg/ml Puromycin until the negative controls died. Single clones were amplified and genotyped.

For the excision of the loxP-flanked selection cassette, mutant IncMN2 KO1/2 mESC were transfected with 2.5 µg of Cre-recombinase-expressing plasmid. Cells were split into two sister dishes (1 µg/ml Puromycin-treated and not treated). Once all the cells on the Puromycin-treated dish died, the sister dish was genotyped to further confirm the removal of the cassette.

### miR-466i-binding region, miR-325- and miR-384-KO generation by CRISPR/Cas9

sgRNAs were designed and cloned as described above. mESC cells were transfected on gelatin-coated dishes using reverse transfection with lipofectamine2000 (Life Technologies) following the manufacturer's instruction and transfecting a total of 500 ng of DNA for 250,000 cells.

An average of 0.5 cells were plated in each well of a 96-well, maintained and passaged for 2–3 weeks, and then screened for the lack of miR-466i-5p-binding region through PCR analysis.

### Single-cell transcriptomics

EB6 were papain-dissociated to single-cell, filtered, and fixed in 80% methanol. Cells were subsequently rehydrated using the 10× Genomics Demonstrated Protocol: briefly, samples were thawed on ice, pelleted at 4° 3,000 g for 10', and washed twice with 1 ml of cold resuspension buffer (PBS-BSA 1% + 0.5 U/ml of RNase inhibitor). Afterwards, cells were resuspended in 200 µl of resuspension buffer, filtered (40 µm flowmi filters, Bel Art), and manually counted in presence of trypan blue. The volume was adjusted to achieve a cell concentration of 1,000 cells/µl and 4,000 cells were immediately subjected to the 10X Single Cell Protocol for transcriptome determination through droplet-based single-cell RNA-seq methodology (10× Genomics Chromium, Chromium Single Cell 3' Solution V3.1, Klein and Macosko, 2017). Cells were separated into droplet emulsion using the Chromium Single Cell 3' Solution (V3.1) and Single-cell RNA-seq libraries were prepared according to the Single Cell 3' Reagent Kits User Guide (V3.1). Libraries were sequenced on a Novaseq 6000 flow cell (Illumina), with a minimum depth of 50K reads/cell. FASTQ reads were aligned, filtered, and counted through the Cell Ranger pipeline (v4.0) using standard parameters. The Mm10 genome version was used in the alignment step, and annotation refers to Ensembl Release93.

### Dataset cleaning

Original datasets were cleaned using Seurat V3.2 (Stuart *et al*, 2019) standard workflows after visualizing the QC metrics, (nCount\_RNA, nFeature\_RNA and *percent.mt* plots) and setting specific “*ad hoc*” thresholds for each dataset to remove possible duplets, multiplets, and dying cell variables among samples. The histogram of *nFeature\_RNA/nCount\_RNA* was also used for the cleaning procedure. For some samples, in fact, this plot showed a clear bimodal distribution with the higher number of cells below a 0.6 threshold. So, datasets were cleaned using the following parameters: WT1 – nFeature\_RNA/nCount\_RNA ≤ 0.6, nCount\_RNA > 200, nFeature\_RNA between 350 and 3,000, percent.mt < 10; WT2 – nFeature\_RNA/

nCount\_RNA < 0.6, nCount\_RNA < 40,000, nFeature\_RNA between 200 and 7,500, percent.mt < 7.5;  $\Delta$ SP1 – nFeature\_RNA/nCount\_RNA < 0.6, nCount\_RNA < 35,000, nFeature\_RNA between 200 and 7,000, percent.mt < 7.5;  $\Delta$ SP2 – nFeature\_RNA/nCount\_RNA < 0.6, nCount\_RNA < 30,000, nFeature\_RNA between 200 and 7,000, percent.mt < 7.5; KO1 – nFeature\_RNA/nCount\_RNA < 0.6, nCount\_RNA < 40,000, nFeature\_RNA between 200 and 7,500, percent.mt < 7.5; KO2 – nFeature\_RNA/nCount\_RNA < 0.6, nCount\_RNA < 30,000, nFeature\_RNA between 200 and 7,000, percent.mt < 7.5.

### Cell clustering and cluster identity assignment

For each sample, cells were clustered by Seurat V3.2 (Stuart et al, 2019). Cluster uniformity was then checked using COTAN (Galfrè et al, 2021) evaluating whether < 1% of genes were over the threshold of 1.5 of GDI. If a cluster result was not uniform, with more than 1% of genes above 1.5, a new round of clustering was performed. After this iterative procedure, the few remaining cells not fitting any cluster were discarded. Using the COTAN function *cotan\_on\_cluster\_DE*, cell identity was assigned to all uniform clusters. This function produces a correlation coefficient and a *P*-value for each gene and each cluster. These data were automatically checked for the two cluster identity assignments performed in which cells were classified into proliferating or postmitotic groups and NP, MNP, MN, or IN groups. For the first one, clusters with a significant positive correlation score ( $P < 0.05$ ) for *Mki67* or *Ccnb2* were identified as “mitotic”, while clusters with the same correlation for *Slc18a3*, *Mnx1*, *Vsx2*, *Gata3*, or *En1* were defined as “postmitotic”. Any clusters not belonging to these two groups were assigned to the NA group (Figs EV4A, right, and EV4B).

For the second classification, clusters with a significant positive correlation score for *Sox2* or *Top2a* and for proliferation genes (*Ccnb2* or *Mki67*) and a significant negative correlation for *Olig2* were assigned to the NP group; clusters with positive correlation for motoneuronal markers (*Slc18a3* or *Mnx1*), for *Olig2* and negative correlation for proliferation genes (*Mki67*, *Ccnb2*) were assigned to the EMN group; while clusters with the same attributes but not displaying significant positive correlation for *Olig2* were assigned to LMN group.

Finally, clusters with a positive and significant correlation for *Vsx2*, *En1*, or *Gata3* and a negative correlation for proliferation genes (*Mki67*, *Ccnb2*) were assigned to IN group (Figs EV4A, left and EV4B). Assignment of clusters was manually inspected and refined by checking marker expression using Seurat *FeaturePlot* function (Fig EV4B).

All subpopulation markers were previously described by Rizvi et al (2017) exception of the additional proliferative marker the cyclin *Ccnb2*.

To evaluate the proliferating/postmitotic assignment a differential expression was performed between the mitotic group and the postmitotic one (Seurat *FindMarkers* function—Dataset EV2) followed by a gene ontology done on significantly differentially expressed genes (adjusted *p*-value lower than 0.05—Dataset EV3). Categories related to cell cycle can be observed in genes enriched in the proliferating cell group (Appendix Fig S4A, left panel, and Dataset EV3), while categories associated with neural differentiation can be observed in the GO for the genes enriched in the postmitotic group (Appendix Fig S4B, right panel and Dataset EV3). To further

test this classification, some known markers of neuronal and proliferative conditions were considered: *Tubb3*, *Elavl3*, *Myt*, and *Map2* for the neuronal condition; while *Cdk1* and *Ccnb1* for proliferative one (Figs 4D and EV4D). To evaluate the NP, MNP, EMN, and LMN classification were considered also: *Foxa2*, *Hes1*, *Sox1*, *Sox3*, *Sox21*, *Foxb1*, *Pax6*, *Hes5*, *Pou3f2*, *Nkx6-1*, *Sp8*, *Neuog1*, *Neurog2*, *Neurod4*, *Dill1*, *Lhx3*, *Mnx1*, *Cntn2*, *Slc10a4*, *Pag1*, *Isl1*, *Nefm*, *Chat*, *Arhgap36*, *Slit2*, *Nefl*, and *Isl2* (Figs 4D and EV4B and D). All these markers were previously described by Briggs et al (2017), Sagner et al (2018), or Rizvi et al (2017).

Gene Ontology over representation analyses were performed with WebGestalt tool (Liao et al, 2019). Categories significantly enriched (FDR < 0.05) from Biological Function database were shown with weighted set cover dimensionality reduction.

### Datasets integration

To perform the datasets integration in Monocle3 (v.0.2.3.0) the official documentation using the *combine\_cds* and *align\_cds* functions was followed. Data was also checked for batch effects. Appendix Fig S3A shows the UMAP plot before (left) and after (right) batch effect correction. Fig EV4C shows the estimated trajectory for the whole dataset and, on top, the cell density plot.

For data visualization, datasets were integrated by Seurat (from the official web page “Introduction to scRNA-seq integration”): each dataset was normalized, and variable features were identified; then, features repeatedly variable across datasets were selected for integration using anchors (with the *FindIntegrationAnchors* function). Appendix Fig S3B shows the UMAP plot for the whole dataset before (left) and after (right) integration.

### Cell type composition analysis and differential gene expression

Differences in cell type composition were assessed using Fisher’s exact test. For each cell type, a contingency table with the number of cells belonging or not to the analyzed group and to each compared condition was evaluated. In the analysis, all the cells belonging to the same batch (WT2,  $\Delta$ SP1,  $\Delta$ SP2, KO1, KO2) were taken into account. Only significant differences between WT and  $\Delta$ SP or WT and KO conditions that were consistent between replicates ( $P$ -value  $\leq 0.05$ ) were considered relevant (Fig 4A).

For MNP, EMN, and LMN subpopulations, differential gene expression analysis was performed comparing WT (2) to  $\Delta$ SP mutants (1,2) or KO mutants (1,2) (Seurat *FindMarkers* function—Dataset EV2) followed by a gene ontology done on significantly differentially expressed genes (adjusted *p*-value lower than 0.05 and | average  $\log_2FC$  | > 0.2—Dataset EV3). Gene Ontology over representation analyses were performed on protein-coding genes with WebGestalt tool (Liao et al, 2019). Categories significantly enriched (FDR < 0.05) from the Biological Function database were shown with weighted set cover dimensionality reduction.

### miRNA-binding site detection

Miranda tool v3.3a (Miranda et al, 2006) was used for microRNA-binding site predictions. In order to have more stringent predictions, we filtered base pairings with  $\Delta G > -15$  kcal/mol. MicroRNA sequences were obtained from miRBase22 (Kozomara et al, 2019).

For the identification of miRNA accessible regions on lncMN2-203, predictions between mouse miRNAs and lncMN2-203 were ranked by “Tot Energy” and the best 25 miRNAs with the lower energy were selected. This list was further filtered by selecting the 9 miRNAs with the highest number of binding sites on lncMN2-203 repeated accessible regions (more than 5). Among them, miR-466i-5p and miR-669a-5p resulted to have the best interaction site with the stronger thermodynamic affinity.

Target predictions in NP, MNP, EMN, and LMN subpopulations were performed for miR-466i-5p on protein-coding genes commonly downregulated between  $\Delta$ SP and KO mutants, while for miR-325-3p and miR-384-5p were selected mRNAs specifically upregulated in KO mutants. For each protein-coding gene, we selected the most representative isoforms with assigned more than 25% of Unique Molecular Identifiers (UMI) of the locus of origin in MNP and MN subpopulations (union of EMN and LMN cells). Transcripts with at least one prediction localized in 3'UTR were considered as putative targets (Dataset EV3).

For UMIs assignment to isoforms, we proceeded as follows: barcoded BAM files obtained from the Cell Ranger pipeline were deduplicated using `umi_tools` (Smith *et al*, 2017) with the following parameters `--extract-umi-method tag --umi-tag 'UB' --per-cell --cell-tag 'CB' --per-gene --gene-tag 'GX' --multimapping-detection-method=NH --mapping-quality=255`. The deduplicated BAM was split into bam relative to each subpopulation with Sinto (<https://timoast.github.io/sinto/>) using cell barcodes assigned to each cluster. For each splitted BAM file, UMIs error-corrected (tag: UB) assigned to each gene (tag: GX) and to each related isoform (tag: TX) were counted, and the ratio was calculated.

### Statistical analysis

Data are expressed as mean values, and error bars represent SD or SEM. Statistical differences were analyzed by a two-tailed unpaired Student's *t*-test. A *P*-value < 0.05 was considered as statistically significant. Differences in cell number among clusters were assessed using the Fisher's exact test. A *P*-value < 0.05 was considered as statistically significant.

## Data availability

Single-cell RNA-Seq data from this publication have been deposited to the GEO database and assigned to the identifier GSE174671 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE174671>).

**Expanded View** for this article is available online.

### Acknowledgements

We thank M. Marchioni, M. Arcenci, and N. Pediconi for technical support and M. Caruso for assistance. We also thank R. Pizzarelli for the helpful discussion on neuronal physiology and M. Cozzolino (IFT-CNR) for providing cells. The authors wish to thank the Imaging Facility at the Center for Life Nano- & Neuro-Science (IIT) and the Genomics Unit at the CGS of IIT@SEMM for single-cell sequencing. This work was partially supported by grants from: ERC-2019-SyG (855923-ASTRA), AIRC (IG 2019 Id. 23053), EC-H2020 Marie Skłodowska-Curie Action ITN 2016 PN 721890; PRIN 2017 (2017P352Z4) and H2020 Program “Sapienza Progetti Collaborativi” to IB; Sapienza University

(prot. RM12117A5DE7A45B and RM11916B7A39DCE5) and POR FESR Lazio 2020-T0002E0001 to MB; AIRC (IG18774 and IG22851), Fondazione Cariplo (2015-0590) and M.J.M. (2016-0615) to FN. Open Access Funding provided by Università degli Studi di Roma La Sapienza within the CRUI-CARE Agreement.

### Author contributions

**Andrea Carvelli:** Conceptualization; Validation; Investigation (genome editing, cell differentiations and RNA analyses). **Adriano Setti:** Conceptualization; Data curation; Software; Formal analysis (SC-seq analyses). **Fabio Desideri:** Conceptualization; Validation; Investigation (molecular experiments). **Silvia Giulia Galfre:** Data curation; Software; Formal analysis; Visualization (SC-seq analyses). **Silvia Biscarini:** Investigation (genome editing). **Tiziana Santini:** Investigation (imaging experiments). **Alessio Colantoni:** Formal analysis (miRNA predictions). **Giovanna Peruzzi:** Investigation (flow cytometry). **Matteo Jacopo Marzi:** Formal analysis; Investigation (SC-seq analyses). **Davide Capauto:** Investigation (experimental activity). **Silvia Di Angelantonio:** Investigation (experimental activity). **Monica Ballarino:** Conceptualization; Funding acquisition (transcriptomics and biochemical assays). **Francesco Nicassio:** Resources; Funding acquisition (SC-seq analyses). **Pietro Laneve:** Conceptualization; Supervision; Writing—original draft; Writing—review and editing (wrote the manuscript and coordinated the work). **Irene Bozzoni:** Conceptualization; Funding acquisition; Writing—original draft; Project administration; Writing—review and editing (wrote the manuscript and coordinated the work).

### Disclosure and competing interests statement

IB is an editorial advisory board/EMBO Member.

## References

- Ait-Si-Ali S, Guasconi V, Fritsch L, Yahi H, Sekhri R, Naguibneva I, Robin P, Cabon F, Poleskaya A, Harel-Bellan A (2004) A Suv39h-dependent mechanism for silencing S-phase genes in differentiating but not in cycling cells. *EMBO J* 23: 605–615
- Andrews MG, Kong J, Novitch BG, Butler SJ (2019) New perspectives on the mechanisms establishing the dorsal-ventral axis of the spinal cord. *Curr Top Dev Biol* 132: 417–450
- Arber S, Han B, Mendelsohn M, Smith M, Jessell TM, Sockanathan S (1999) Requirement for the homeobox gene Hb9 in the consolidation of motor neuron identity. *Neuron* 23: 659–674
- Ballarino M, Cipriano A, Tita R, Santini T, Desideri F, Morlando M, Colantoni A, Carrieri C, Nicoletti C, Musarò A *et al* (2018) Deficiency in the nuclear long noncoding RNA Charmé causes myogenic defects and heart remodeling in mice. *EMBO J* 37: e99697
- Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE *et al* (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447: 799–816
- Biscarini S, Capauto D, Peruzzi G, Lu L, Colantoni A, Santini T, Shneider NA, Caffarelli E, Laneve P, Bozzoni I (2018) Characterization of the lncRNA transcriptome in mESC-derived motor neurons: Implications for FUS-ALS. *Stem Cell Res* 27: 172–179
- Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y, Xu X, Lv X, Hugnot J-P, Tanay A *et al* (2017) Multiscale 3D genome rewiring during mouse neural development. *Cell* 171: 557–572
- Briggs JA, Li VC, Lee S, Woolf CJ, Klein A, Kirschner MW (2017) Mouse embryonic stem cells can differentiate via multiple paths to the same state. *Elife* 6: e26945

- Briggs JA, Wolvetang EJ, Mattick JS, Rinn JL, Barry G (2015) Mechanisms of long non-coding RNAs in mammalian nervous system development, plasticity, disease, and evolution. *Neuron* 88: 861–877
- Caputo D, Colantoni A, Lu L, Santini T, Peruzzi G, Biscarini S, Morlando M, Shneider NA, Caffarelli E, Laneve P et al (2018) A regulatory circuitry between Gria2, miR-409, and miR-495 is affected by ALS FUS mutation in ESC-derived motor neurons. *Mol Neurobiol* 55: 7635–7651
- Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C et al (2005) The transcriptional landscape of the mammalian genome. *Science* 309: 1559–1563
- Catela C, Kratsios P (2021) Transcriptional mechanisms of motor neuron development in vertebrates and invertebrates. *Dev Biol* 475: 193–204
- Cave C, Sockanathan S (2018) Transcription factor mechanisms guiding motor neuron differentiation and diversification. *Curr Opin Neurobiol* 53: 1–7
- Cesana M, Cacchiarelli D, Legnini I, Santini T, Sthandier O, Chinappi M, Tramontano A, Bozzoni I (2011) A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* 147: 358–369
- Chen KW, Chen JA (2020) Functional roles of long non-coding RNAs in motor neuron development and disease. *J Biomed Sci* 27: 38
- Chen L, Zhang YH, Pan X, Liu M, Wang S, Huang T, Cai YD (2018) Tissue expression difference between mRNAs and lncRNAs. *Int J Mol Sci* 19: 3416
- Chiu CW, Monat C, Robitaille M, Lacomme M, Daulat AM, Macleod G, Helen McNeill H, Cayouette M, Angers S (2016) SAPCD2 controls spindle orientation and asymmetric divisions by negatively regulating the Gxi-LGN-NuMA ternary complex. *Dev Cell* 36: 50–62
- Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG et al (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution and expression. *Genome Res* 22: 1775–1789
- Desideri F, Cipriano A, Petrezselyova S, Buonaiuto G, Santini T, Kasperek P, Prochazka J, Janson G, Paiardini A, Calicchio A et al (2020) Intronic determinants coordinate lncRNA nuclear activity through the interaction with MATR3 and PTBP1. *Cell Rep* 33: 108548
- Eggett CJ, Crosier S, Manning P, Cookson MR, Menzies FM, McNeil CJ, Shaw PJ (2000) Development and characterisation of a glutamate-sensitive motor neuron cell line. *J Neurochem* 74: 1895–1902
- Engels B, Jannot G, Remenyi J, Simard MJ, Hutvagner G (2012) Polypyrimidine tract binding protein (hnRNP I) is possibly a conserved modulator of miRNA-mediated gene regulation. *PLoS One* 7: e33144
- Galfre SG, Morandini F, Pietrosanto M, Cremisi F, Helmer-Citterich M (2021) COTAN: scRNA-seq data analysis based on gene co-expression. *NAR Genom Bioinform* 3: lqab072
- Gao T, Li J, Li N, Gao Y, Yu L, Zhuang S, Zhao Y, Dong X (2020) Incrps25 play an essential role in motor neuron development through controlling the expression of olig2 in zebrafish. *J Cell Physiol* 235: 3485–3496
- Haghverdi L, Lun ATL, Morgan MD, Marioni JC (2018) Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat Biotechnol* 36: 421–427
- Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S et al (2012) GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* 22: 1760–1774
- He D, Wu D, Muller S, Wang L, Saha P, Ahanger SH, Liu SJ, Cui M, Hong SJ, Jain M et al (2021) miRNA-independent function of long noncoding pri-miRNA loci. *Proc Natl Acad Sci USA* 118: e2017562118
- Itoh Y, Moriyama Y, Hasegawa T, Endo TA, Toyoda T, Gotoh Y (2013) Scratch regulates neuronal migration onset via an epithelial-mesenchymal transition-like mechanism. *Nat Neurosci* 16: 416–425
- Jessell TM (2000) Neuronal specification in the spinal cord: inductive signals and transcriptional codes. *Nat Rev Genet* 1: 20–29
- Klein AM, Macosko E (2017) InDrops and Drop-seq technologies for single-cell sequencing. *Lab Chip* 17: 2540–2541
- Kozomara A, Birgaoanu M, Griffiths-Jones S (2019) miRBase: from microRNA sequences to function. *Nucleic Acids Res* 47(D1): D155–D162
- Lam M, Moslem M, Bryois J, Pronk RJ, Uhlin E, Ellström ID, Laan L, Olive J, Morse R, Rönholm H et al (2019) Single cell analysis of autism patient with bi-allelic NRXN1-alpha deletion reveals skewed fate choice in neural progenitors and impaired neuronal functionality. *Exp Cell Res* 383: 111469
- Liao Y, Wang J, Jaehng E, Shi Z, Zhang B (2019) WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res* 47: W199–W205
- Lu DC, Niu T, Alaynick WA (2015) Molecular and cellular development of spinal cord locomotor circuitry. *Front Mol Neurosci* 8: 25
- Martone J, Mariani D, Desideri F, Ballarino M (2020) Non-coding RNAs shaping muscle. *Front Cell Dev Biol* 7: 394
- Matsuda T, Irie T, Katsurabayashi S, Hayashi Y, Nagai T, Hamazaki N, Adefuin AMD, Miura F, Ito T, Kimura H et al (2019) Pioneer factor NeuroD1 rearranges transcriptional and epigenetic profiles to execute microglia-neuron conversion. *Neuron* 101: 472–485
- Mercer TR, Dinger ME, Sunkin SM, Mehler MF, Mattick JS (2008) Specific expression of long noncoding RNAs in the mouse brain. *Proc Natl Acad Sci USA* 105: 716–721
- Miranda KC, Huynh T, Tay Y, Ang YS, Tam WL, Thomson AM, Lim B, Rigoutsos I (2006) A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell* 126: 1203–1217
- Molyneux BJ, Goff LA, Brettler AC, Chen HH, Brown JR, Hrvatin S, Rinn JL, Arlotta P (2015) DeCoN: genome-wide analysis of *in vivo* transcriptional dynamics during pyramidal neuron fate selection in neocortex. *Neuron* 85: 275–288
- Nakamura EK, Watkins DN, Schuebel KE, Sriuranpong V, Borges MW, Nelkin BD, Ball DW (2001) Mammalian Scratch: a neural-specific Snail family transcriptional repressor. *Proc Natl Acad Sci USA* 98: 4010–4015
- Ray MK, Wiskow O, King MJ, Ismail N, Ergun A, Wang Y, Plys AJ, Davis CP, Kathrein K, Sadreyev R et al (2016) CAT7 and cat7l long non-coding RNAs tune polycomb repressive complex 1 function during human and zebrafish development. *J Biol Chem* 291: 19558–19572
- Ribeiro DM, Zanzoni A, Cipriano A, Delli Ponti R, Spinelli L, Ballarino M, Bozzoni I, Tartaglia GG, Brun C (2018) Protein complex scaffolding predicted as a prevalent function of long non-coding RNAs. *Nucleic Acids Res* 46: 917–928
- Rispol J, Baron L, Beaulieu JF, Chevillard-Briet M, Trouche D, Escaffit F (2019) The H2A.Z histone variant integrates Wnt signaling in intestinal epithelial homeostasis. *Nat Commun* 10: 1–14
- Rizvi AH, Camara PG, Kandror EK, Roberts TJ, Schieren I, Maniatis T, Rabadan R (2017) Single-cell topological RNA-seq analysis reveals insights into cellular differentiation and development. *Nat Biotechnol* 35: 551–560
- Roballo KCS, da Silveira JC, Bressan FF, de Souza AF, Pereira VM, Porras JEP, Rós FA, Pulz LH, Strefezzi RDF, Martins DDS et al (2019) Neurons-derived extracellular vesicles promote neural differentiation of ADSCs: a model to prevent peripheral nerve degeneration. *Sci Rep* 9: 11213
- Roy A, Francius C, Rouso DL, Seuntjens E, Debruyjn J, Luxenhofer G, Huber AB, Huylebroeck D, Novitsch BG, Clotman F (2012) Onecut transcription factors act upstream of Isl1 to regulate spinal motoneuron diversification. *Development* 17: 3109–3119

- Sagner A, Gaber ZB, Delile J, Kong JH, Rousso DL, Pearson CA, Weicksel SE, Melchionda M, Mousavy Gharavy SN, Briscoe J et al (2018) Olig2 and Hes regulatory dynamics during motor neuron differentiation revealed by single cell transcriptomics. *PLoS Biol* 16: e2003127
- Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP (2011) A ceRNA hypothesis: the Rosetta stone of a hidden RNA language? *Cell* 146: 353–358
- Saxe JP, Chen M, Zhao H, Lin H (2013) Tdrkh is essential for spermatogenesis and participates in primary piRNA biogenesis in the germline. *EMBO J* 32: 1869–1885
- Schmidt S, Arendt T, Morawski M, Sonntag M (2020) Neurocan contributes to perineuronal net development. *Neuroscience* 442: 69–86
- Smith T, Heger A, Sudbery I (2017) UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* 27: 491–499
- Stifani N (2014) Motor neurons and the generation of spinal motor neuron diversity. *Front Cell Neurosci* 8: 293
- Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck III WM, Hao Y, Stoeckius M, Smibert P, Satija R (2019) Comprehensive integration of single-cell data. *Cell* 177: 1888–1902
- Tanabe Y, Jessell TM (1996) Diversity and pattern in the developing spinal cord. *Science* 274: 1115–1123
- Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* 32: 381–386
- Vangoor VR, Gomes-Duarte A, Pasterkamp RJ (2021) Long non-coding RNAs in motor neuron development and disease. *J Neurochem* 156: 777–801
- Wang L, Feng W, Yang X, Yang F, Wang R, Ren Q, Zhu X, Zheng G (2018) Fbxw11 promotes the proliferation of lymphocytic leukemia cells through the concomitant activation of NF- $\kappa$ B and  $\beta$ -catenin/TCF signaling pathways. *Cell Death Dis* 9: 427
- Wichterle H, Lieberam I, Porter JA, Jessell TM (2002) Directed differentiation of embryonic stem cells into motor neurons. *Cell* 110: 385–397
- Wichterle H, Peljto M (2008) Differentiation of mouse embryonic stem cells to spinal motor neurons. *Curr Protoc Stem Cell Biol* Chapter 1, Unit 1, H.1.1-1H.1.9
- Xie C, Yuan J, Li H, Li M, Zhao G, Bu D, Zhu W, Wu W, Chen R, Zhao Y (2014) NONCODEv4: exploring the world of long non-coding RNA genes. *Nucleic Acids Res* 42: D98–D103
- Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Azov AG, Bennett R et al (2020) Ensembl 2020. *Nucleic Acids Res* 48: D682–D688
- Yen Y-P, Hsieh W-F, Tsai Y-Y, Lu Y-L, Liao ES, Hsu H-C, Chen Y-C, Liu T-C, Chang M, Li J et al (2018) Dlk1-Dio3 locus-derived lncRNAs perpetuate postmitotic motor neuron cell fate and subtype identity. *Elife* 7: e38080
- Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J et al (2017) Massively parallel digital transcriptional profiling of single cells. *Nat Commun* 8: 14049
- Zhou Z, Ai H, Li K, Yao X, Zhu W, Liu L, Yu C, Song Z, Bao Y, Huang Y et al (2018) Prohibitin 2 localizes in nucleolus to regulate ribosomal RNA transcription and facilitate cell proliferation in RD cells. *Sci Rep* 8, 1–12
- Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31: 3406–3415



**License:** This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.