


STAT3-dependent long non-coding RNA Lncenc1 contributes to mouse ES cells pluripotency via stabilizing Klf4 mRNA

Emanuele Monteleone , Paola Corrieri, Paolo Provero, Daniele Viavattene, Lorenzo Pulvirenti, Laura Raggi, Elena Carbognin, Marco E. Bianchi, Graziano Martello, Salvatore Oliviero, Pier Paolo Pandolfi and Valeria Poli 

*Corresponding authors: Emanuele Monteleone, Università Vita-Salute San Raffaele, Division of Genetics and Cell Biology Via Olgettina, 60, 20132 Milano, Italy. Tel: +39 02 2643 4756; E-mail: monteleone.emanuele@hsr.it, Valeria Poli, Department of Molecular Biotechnology and Health Science, University of Torino, Via Nizza 52, 10126 Torino, Italy. Tel: +39-0116706428; E-mail: valeria.poli@unito.it

Emanuele Monteleone is an assistant professor in bioinformatics at the University Vita-Salute San Raffaele in Milan, Italy. He received his Ph.D. in Molecular Medicine at the University of Torino. Dr. Monteleone's current research focuses in the organization and function of chromatin in cell identity transitions in both physiological and pathological conditions. In particular, he is interested in nucleosome dynamics and positioning in key cellular processes, including reprogramming to iPS cells, Epithelial-To-Mesenchymal Transition and cellular senescence.

Paola Corrieri after a Master Degree in Preclinical and Clinical Research and Drug development, she won a fellowship and worked as study coordinator. She is currently a Clinical Research Associate, supporting sites into clinical trials development. She is working on studies in the following fields: oncology, infectious diseases and hemoglobinopathies.

Paolo Provero is a professor of Molecular Biology in the Department of Neurosciences of the University of Turin. He received his PhD degree in Theoretical Physics from the University of Genoa in 1992. Since 2000 his research has focused on computational genomics, and in particular on the development of analytical methods for the study of gene regulation at the transcriptional and post-transcriptional levels.

Daniele Viavattene is a PhD student in Molecular Medicine at the Molecular Biotechnology Center, University of Turin, Italy. He received his M.S. degree in Medical Biotechnology at University of Turin and his B.S. degree in Biotechnology at University of Messina, Italy. His research interests lie in cancer and molecularbiology.

Lorenzo Pulvirenti is a clinical data manager at AOU S. Luigi Gonzaga at the department of oncology. He received his M.D. in molecular biology at the University of Turin. His research intersects lie in clinical data management in the field of oncology.

Laura Raggi is a third year PhD student at San Raffaele Telethon Institute for gene therapy (SR-TIGET) in Milan, Italy. Her research interests focus on hematopoietic stem cell-niche interactions in beta-thalassemia with potential translational relevance for improving gene therapy in congenital anemia.

Elena Carbognin is an Assistant Professor at the Department of Biology, University of Padua, Italy. She received her Ph.D. degree in Biosciences at Padua University, Italy. Her research focuses on the molecular mechanisms controlling maintenance of pluripotency in embryonic stem cells and early differentiation.

Graziano Martello is Associate Professor at the Department of Biology, University of Padua. He received in 2009 a Ph.D. in Genetics and Developmental Biology studying the role of non-coding RNAs during early development. He then moved to Cambridge University (UK) to study pluripotent stem cells from 2010 to 2014. He set up his independent laboratory of pluripotent stem cell biology (martellolab.org) at University of Padua in 2014, where he was Associated Professor of Molecular Biology in 2017. His research is focussed on understanding the mechanisms controlling pluripotent maintenance and induction, and to model neurodegenerative diseases with pluripotent cells. His work is supported by the Armenise-Harvard foundation, Telethon foundation, Microsoft Research and the European Research Council.

Salvatore Oliviero is a Professor in Molecular Biology at the University of Torino and head of the Laboratory of Epigenetics at the Italian Institute for Genomic Medicine (IIGM). Between 1985 and 1988 he was at EMBL (Heidelberg, Germany) as a PhD student. From 1988 to 1992 he was postdoctoral fellow at Harvard Medical School (Boston, USA). His laboratory identified and characterized a new vascular endothelial angiogenic growth factor (VEGF-D); unveiled the molecular mechanism by which MYC induces the elongation step of the transcription. More recently his laboratory is focused on the function of epigenetic modifications that regulate the early steps of embryonic development.

Pier Paolo Pandolfi is an Italian scientist renowned for his groundbreaking contributions to the field of cancer research and molecular medicine. Pandolfi completed his medical degree at the University of Perugia in Italy and subsequently pursued a Ph.D. in Molecular and Cellular Biology at the European Molecular Biology Laboratory (EMBL) in Germany. He became PI at the Memorial Sloan-Kettering Cancer Center, NY, USA in 1994, and then joined the Harvard Medical School to become Director of the BIDMC Cancer Center in 2013. His research focuses on the genetic factors underlying human cancers.

Valeria Poli is Professor in Molecular Biology at the University of Torino, Italy since 2001. She was Post-doctoral fellow at the EMBL, Heidelberg, Germany in 1988-1990 and at Columbia University, NY, USA in 1990-1992. She was Principal Investigator at IRBM, Pomezia, Rome, Italy in 1992-1997, and at the University of Dundee, Scotland, UK in 1997-2001. She has cloned the IL-6-activated transcription factor C/EBPbeta and generated and characterized several genetically modified mouse strains including IL-6 and C/EBPbeta null, and conditional and isoform-specific STAT3 null. Her research focuses on the regulation of gene expression during inflammation, auto-immunity and cancer.

Received: June 1, 2023. Revised: August 28, 2023. Accepted: September 12, 2023.

© The Author(s) 2023. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

according to manufacturer's instructions. To score AP staining and colonies circularity, imaged cells were analyzed using a custom macro in the FIJI software (<https://imagej.net/software/fiji>).

Plasmid constructs

Custom shRNAs were designed using the TRC hairpin design tool (<http://www.broadinstitute.org/mnai/public/seq/search>). shRNAs with more than four consecutive matches to non-target transcripts were avoided. Hairpins were cloned into the pLKO.1 vector (Addgene, Watertown, MA, USA: Cat. No. 10878) and controlled by sequencing. The pLKO.1 non-targeting control vector was purchased from Addgene (Cat. No. 136035). The MREs enriched region of *Lncenc1* was amplified by PCR from E14 genomic DNA purified using QIAquick Gel Extraction Kit_(Qiagen, Hilden, Germany, Cat. No. 28706X4) and cloned in the pLVX-Tight-Puro vector. The pPyCAGSTAT3ERT2iresZeo construct (STAT3ERT) was kindly provided by Austin Smith [7]. For luciferase constructs, the *Klf4* 3'UTR was amplified by PCR from E14 genomic DNA and cloned in the pMIR-REPORT vector (Thermo Fisher Scientific, Cat. No. AM5785). Site-directed mutagenesis on the resulting plasmid was performed using the Quick-Change kit (Stratagene, La Jolla, CA, USA, Cat. No. #200523). All primers used for cloning and mutagenesis are listed in Supplementary Table 1.

Transfection

Transfection of E14 cells was performed using the Lipofectamine™ 2000 Transfection Reagent (Thermo Fisher Scientific), using 10 μl of transfection reagent, 4 μg of plasmid DNA and 800 000 cells in 800 μl of OPTIMEM (Thermo Fisher Scientific) for each well of a six-well plate. After 6 hours, cells were rinsed with 1.2 ml of complete ESCs medium, incubated O/N, detached and seeded into a 10 cm plate. Unless otherwise noted, shRNA transfected cells were selected with 1 μg/ml of puromycin for 48 hours, followed by cell lysis at 72 hours, whereas ASOs-treated cells were analyzed 36 hours after removal of the transfection medium.

RNA extraction and real time PCR analysis

Total RNA was extracted by using TRIzol reagent (Thermo Fisher Scientific), according to manufacturer's protocol, cDNA generated using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Waltham, MA, USA, Cat. No. #4375575), and Real-time PCR performed using the Fast SYBR™ Green Master Mix (Thermo Fisher Scientific), with the primers listed in Supplementary Table 5. For miRNAs quantification, cDNAs were generated using the Taqman Advanced miRNA cDNA Synthesis Kit (Applied Biosystems, Cat. No. A28007), and Real-time PCR performed on total RNA with the indicated TaqMan MicroRNA Assay (Applied Biosystems), according to manufacturer's instructions, and normalized on U6 RNA levels.

Lentivirus production and titration

Lentivirus production was performed using the Lenti-XTM Lentiviral Expression Systems (Clontech Laboratories, Mountain View, CA, USA). LentiX 293 T-Rex cell lines were transfected using Effectene Transfection Reagent (Qiagen) according to the manufacturer's protocol. Lentiviral particles were harvested at 24 and 48 hours, filtered through a 0.22-μm pore cellulose acetate filters, concentrated by ultracentrifugation for 2 hours at 22,000 g and resuspended in 1× PBS, 1% BSA. Vector infectivity was evaluated by transducing ES cells with serial 4-fold dilutions: undiluted, 1/4, 1/16, 1/64, 1/256 and 1/1024. After 72 h, the titer was estimated by real-time quantitative RT-PCR of a common

lentiviral genome region (WPRE), upon collection of the SN and RNA extraction.

Chromatin immunoprecipitation assay

Chromatin immunoprecipitation was performed as described in Avalle et al. (2022) [22], with modifications. Briefly, 2*10⁷ cells were cross-linked with 1% formaldehyde for 10 minutes at RT, quenched with 0.125 M glycine for 5 minutes, and then washed twice in cold PBS. The cells were suspended in Lysis Buffer 1 (50 mM Hepes-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% Glycerol, 0.5% NP-40, 0.25% Triton X-100 and protease inhibitor) to disrupt the cell membrane, followed by Lysis Buffer 2 (10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA and protease inhibitor) to isolate nuclei. The isolated nuclei were then resuspended in SDS ChIP Buffer (20 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS and protease inhibitors). Extracts were sonicated using the Bioruptor H Twin (Diagenode, Liege, Belgium) for two runs of 10 cycles [30 sec 'ON', 30 sec 'OFF'] at high power setting. Nuclear lysates were centrifuged at 12,000 g for 10 minutes at 4°C. The supernatant was diluted with ChIP Dilution Buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 1% Triton) before the immunoprecipitation step. Streptavidin beads (Dynabeads® Protein G, Thermo Fisher Scientific) were saturated with PBS/1% BSA and samples were incubated with 5 μg of antibody overnight at 4 °C on a rotator. The antibodies used were anti-STAT3 antibodies (Cell Signaling Technology, Danvers, USA, Cat. No. 9132), phospho (S5) RNA polymerase II, ab5408 (Abcam, Cambridge, UK) or rabbit IgG (Thermo Fisher Scientific, Cat. No. 31235). Samples were then incubated with saturated beads for 2 hours at 4 °C on a rotator. Immunoprecipitated complexes were washed five times with RIPA buffer (50 mM Hepes-KOH pH 7.6, 500 mM LiCl, 1 mM EDTA, 1% NP-40, 0.7% Na-Deoxycholate) at 4 °C for 5 minutes each on a rotator. Elution Buffer was added and incubated at 65°C for 15 minutes. The de-crosslinking was performed at 65 °C overnight, followed by standard phenol:chloroform:isomyl alcohol (25:24:1) DNA purification and real time PCR (see 'RNA extraction and Real Time PCR analysis'). Primers for SYBR green qPCR reactions are listed in Supplementary Table 5.

RNA fluorescence in situ hybridization

A single long RNA biotinylated probe targeting *Lncenc1* RNA was obtained as follows: a region of 542 bp was PCR amplified from genomic DNA with primers listed in Supplementary Table 5. The resulting DNA fragment was cloned into the pGEM T-easy vector (Promega Corporation, Madison USA, Cat. No. A1360). The resulting construct was used as a template for T7-based *in vitro* transcription using the MEGAscript™ T7 Transcription Kit (Thermo Fisher Scientific, Cat. No. AMB13345). *In situ* hybridization was performed as follows. Briefly, mESCs were seeded (3x10⁵) onto poly-L-lysine and gelatin coated glass coverslips, rinsed with PBS and fixed in 4% PFA. Upon incubating in 2× SSC, 1% BSA saturation buffer for 3 hours at room temperature, 10 ng of biotin-labeled probes diluted in hybridization buffer (2× SSC, 10% Deionized Formamide, 50% Dextran Sulfate) were heat denatured and incubated with the cells at 37°C O/N in a humidified chamber. All buffers used prior to hybridization were supplemented with RNase inhibitors (SUPERaseIn, Thermo Fisher Scientific, Cat. No. A2694) at a concentration of 100 U/ml. Cells were then rinsed twice in either Wash buffer (2% SSC, 50% Deionized Formamide) or PBS. Primary antibodies conjugated or not with streptavidin-PE (BioLegend, San Diego, CA, USA) were diluted 1:200 in blocking buffer (PBS 1X, 1% BSA). Biotinylated Anti-Streptavidin antibodies

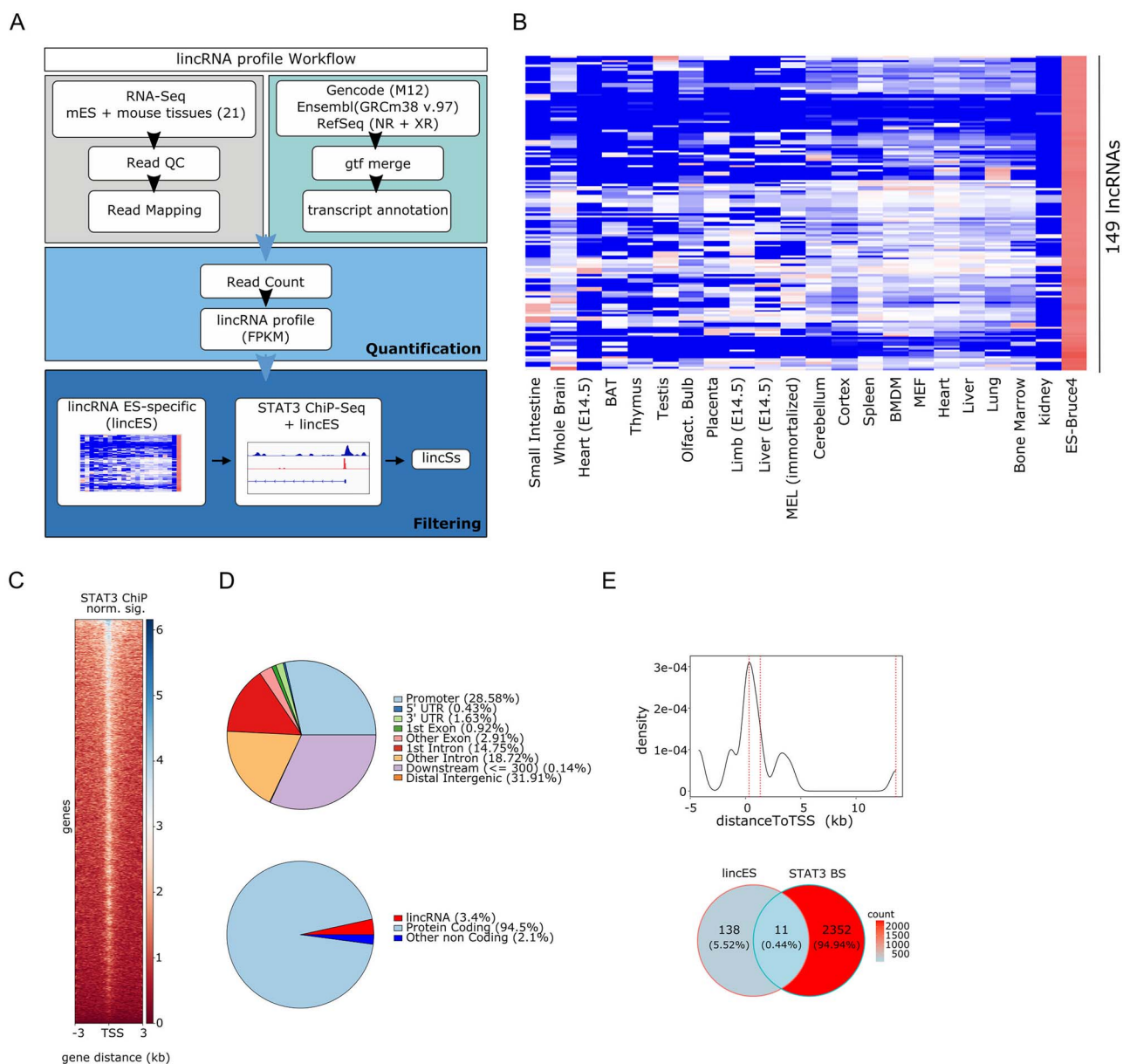


Figure 1. Identification of putative STAT3-dependent ESC-associated long intergenic non-coding RNAs (lincS). **(A)** Bioinformatics pipeline for ESC and mouse tissues transcriptome analysis. **(B)** Heatmap (GeneWise Z-score) representing the expression levels of the 149 lincRNAs identified as most specifically expressed in mESCs, in the indicated tissues or in Bruce ESCs (ENCODE RNA-sequencing data). **(C)** Heatmap showing the normalized STAT3-binding signal around the TSS (−3 to 3 kb) of genes expressed in mESC. **(D)** Genomic distribution of STAT3. Top panel: percentage of STAT3 binding at the indicated genomic locations. Bottom panel: STAT3 locations at promoters (~28.6% of all peaks) was further dissected according to the corresponding gene biotype (coding or non-coding). **(E)** Top panel: density distribution of the distance of STAT3-binding sites to the TSS of associated genes (see materials and methods for the peak–gene association rule). The distance between three STAT3-binding peaks and the TSS of lincS3 lincRNA is indicated by dotted lines intersecting the curve. Bottom panel: overlap between STAT3 peaks and putative regulatory regions of ESC-associated lincRNAs, identifying 11 lincSs.

a detected STAT3 binding site, within ± 5 kb from the peak summit. Interestingly, the majority of peaks mapped relatively near the corresponding TSS, with only lincS3 displaying three STAT3 binding sites, at both proximal and distal sites (Figure 1E, top panel). We identified 11 lincRNAs displaying *in vivo* STAT3 binding at their putative regulatory regions, which we named STAT3-dependent lincRNAs (lincS) (Supplementary Table 1 and Figure 1E, bottom panel).

We focused our attention on LincS3, on the ground of its high expression levels and highest enrichment of STAT3 *in vivo* binding. This lincRNA (Lncenc1) was previously identified as a putative

regulator of stem cell functions in several high throughput RNAi screenings [3, 28–30]. However, nothing is known about its functions along the LIF–STAT3 axis.

The ESC-specific LincS3/Lncenc1 RNA is a STAT3 target

Lncenc1 ESC-specific expression was confirmed by RT-qPCR analysis, confirming that this RNA is undetectable in adult tissues or in EBs but highly expressed in mESCs (Figure 2A, B). We next sought to characterize the Lncenc1 locus, of which Figure 2C shows the structure and epigenetic status in ESCs. Three different

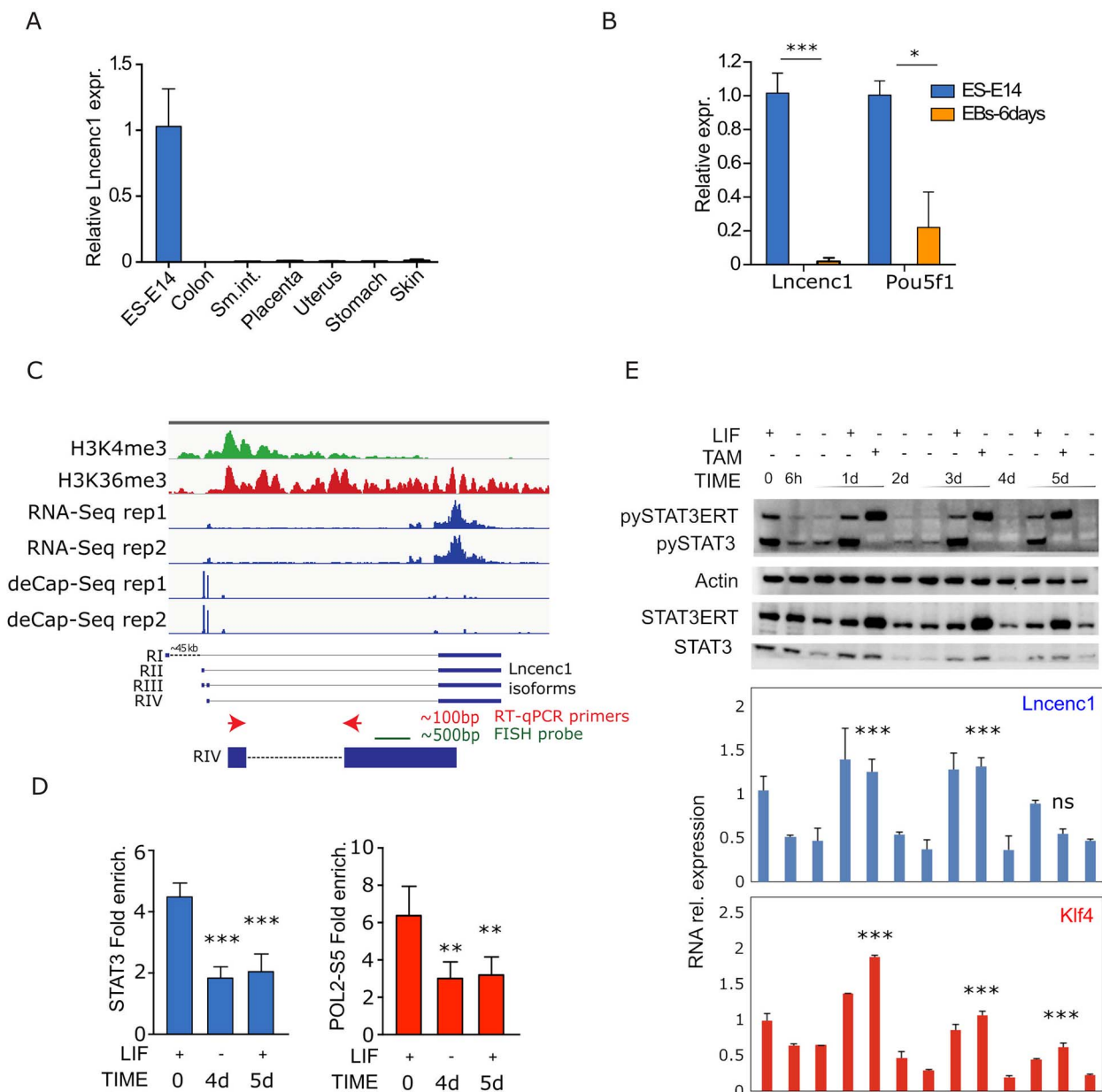


Figure 2. Lncenc1 is a STAT3-dependent ESC-specific lncRNA. **(A, B)** Lncenc1 and/or Pou5f1 (Oct4) RNA levels were measured by RT-qPCR in E14 mESCs and (A) the indicated adult mouse tissues, or (B) in derived EBs. Data are represented as the mean \pm SD of the values normalized to the 18S rRNA internal control. $N = 3$. The asterisks indicate statistically significant differences. $***P < 0.0001$; $*P < 0.01$. **(C)** Integrative Genomics Viewer genome browser picture at the Lncenc1 locus showing the distribution of the H3K4 and H3K36 me3 chromatin marks, RNA-Seq profiles and deCAP-Seq data in E14 ES cells. Below, the RefSeq Lncenc1 isoforms RI–RIII are shown, together with the novel RIV isoform predicted by the deCap experiments. The RT-qPCR primers (arrows) and the probe used in the RNA FISH and Northern blot experiments are also shown. **(D)** ChIP-qPCR analysis showing STAT3 or PolII-PS5 relative enrichment on the Lncenc1 regulatory region and promoter, respectively, performed in E14 ESCs before or after LIF withdrawal (4 days) and upon 24 hours LIF re-supplementation (5 days). Data are mean \pm SD of three independent experiments. $***P < 0.001$; $**P < 0.01$. **(E)** STAT3-ERT E14 ESCs were deprived of LIF for 6 hours, 2 or 4 days, followed by 24-hour LIF or Tamoxifen re-stimulation prior to protein and RNA extraction. Top panel: Western blot obtained with either total or pYSTAT3 antibodies at the indicated time points. Actin was used as an internal control. The bottom panels show Lncenc1 and Klf4 RNA expression levels measured by RT-qPCR in the same samples, as the mean \pm SD of three independent experiments. The asterisks indicate statistically significant difference between LIF-deprived samples, either treated or not treated with TAM at the indicated time points; $***P < 0.001$; ns, not significant.

Lncenc1 isoforms are reported in the RefSeq database, RI, RII and RIII, all sharing the last and largest exon that indeed shows the highest RNA-Seq coverage. However, analysis of mESCs deCAP-Seq [31], RNA-Seq, H3K4me3 and H3K36me3 histones methylation ChIP-Seq data (The ENCODE Project Consortium 2012) revealed a fourth prevalent isoform, RIV, with an alternative transcription start site but sharing the common largest exon. Targeted RT-PCR

in E14 mESCs failed to detect the RI isoform and confirmed the prevalence of the RIV isoform (Supplementary Figure 1A). Accordingly, Northern blot analysis detected a single band, compatible with the predicted 3.3 kb molecular weight of the RIV isoform (Supplementary Figure 1B). The RT-PCR primers and probe used in the following experiments were therefore designed on this isoform (Figure 2C, bottom).

Two of the three STAT3 binding sites detected at the Lncenc1 locus (Figure 1E) mapped near the first exon of the RI isoform, while the third one was within the largest exon. Given that we could not detect any expression of the RI transcript and that the proximal STAT3 peaks are ~45 kb distal with respect to the RIV isoform, we focused on this latter binding site, where we could confirm LIF-dependent STAT3 *in vivo* binding by Chromatin Immunoprecipitation (ChIP) (Figure 2D). Further confirming STAT3-mediated transcriptional regulation, also active Serine 5 phosphorylated RNA Polymerase II binding was detected at the Lncenc1 promoter region, and was impaired upon LIF withdrawal (Figure 2D). Of note, the Lncenc1 locus is embedded in a previously reported super-enhancer region in mESCs [32], suggesting that STAT3 might orchestrate DNA looping to regulate Lncenc1 transcription.

To further assess transcriptional dependency on STAT3, we took advantage of previously generated E14 mESCs stably expressing a tamoxifene-dependent STAT3-ERT fusion protein [7]. Total RNA and protein extracts were obtained from these cells upon LIF deprivation for 6 hours, 2, 3 or 4 days, followed by supplementation with either LIF or Tamoxifene for 24 hours. As expected, tyrosine phosphorylation of both endogenous and recombinant STAT3 dramatically dropped after 6 hours of LIF withdrawal, became undetectable after 4 days and increased again upon either LIF or tamoxifen treatment, respectively (Figure 2E, top panel). Lncenc1 RNA levels closely paralleled STAT3-YP, being strongly decreased already six hours after LIF withdrawal and induced again by either LIF or Tamoxifen treatment up to 3 days after LIF deprivation (Figure 2E, bottom panel). At day 4, however, despite equivalent STAT3 phosphorylation levels, Lncenc1 expression could be no longer rescued, suggesting that differentiation had reached a point of no return. Similarly, neither LIF nor Tamoxifene re-supplementation could rescue STAT3 and PolII binding, suggesting an epigenetic mechanism impeding STAT3 binding and transcription (Figure 2D). Remarkably, the mRNA for the core pluripotency factor Klf4, a well-known LIF-STAT3 direct transcriptional target, closely paralleled Lncenc1 expression patterns (Figure 2E, bottom panel). Thus, both Lncenc1 and Klf4 mRNA levels similarly respond to STAT3, independently of LIF.

Lncenc1 is associated with naïve pluripotency

To dissect its biological role, we depleted Lncenc1 by transfecting E14 ESCs with two independent shRNAs or a non-targeting control (Supplementary Figure 2A). Lncenc1 silencing triggered ESCs differentiation, as shown by disrupted colony morphology, quantified as colony circularity, and the significant reduction of Alkaline Phosphatase (AP) activity, a well-recognized pluripotency marker (Figure 3A). Further supporting its role in the maintenance/induction of pluripotency, Lncenc1 levels were dramatically induced upon reprogramming of epiblast stem cells toward a naïve, ground state via culturing in 2i medium plus LIF (Figure 3B).

We then compared the RNA expression profiles of E14 ESCs silenced or not for Lncenc1 for 72 or 36 hours, respectively, by means of either shRNAs or Antisense Oligonucleotides (ASO) (Supplementary Figure 2B). Since ASOs are known to elicit quicker but more transient silencing as compared to shRNAs (Watts and Corey 2012, and data not shown), we reasoned that this strategy would allow us to assess both early and late silencing effects. Comparing two separate time points and two distinct silencing methods may in fact allow to better discern direct phenotypic effects. Bioinformatic analysis of RNAs dysregulated by Lncenc1 down-modulation across the different experimental conditions revealed a similar, significant impact on the expression

of about 1000 genes (LRT and FDR < 0.05) (Figure 3C). Of note, higher variations were elicited by the shRNAs as compared to the ASO, as expected from the relative silencing strength and supporting results specificity. In agreement with the observed phenotype, upregulated genes were mostly related to embryonic development and cell fate commitment (Figure 3D and Supplementary Table 2), including the mesoderm and neuroectoderm markers Sox1, Nestin, Fgf5 and Foxd3 (Supplementary Table 2). In contrast, many down-regulated genes belonged to metabolic categories, including the glycolytic process (Figure 3D and Supplementary Table 2).

Lncenc1 predominantly localizes to the cytosol and acts via miRNA-mediated mechanisms

LincRNA functions are partly dictated by their cellular localization. To determine Lncenc1 localization, cytosolic, nuclear or chromatin E14 subcellular fractions were analyzed by RT-qPCR, showing prevalent cytosolic localization (Figure 4A). The predominantly chromatinic Meg3 lincRNA was used as a control. Confirming the fractionation data, FISH experiments showed a punctate, mostly cytosolic pattern (Figure 4B).

A potential mode of action of cytoplasmic lincRNAs is that of acting via miRNA-mediated mechanisms. Indeed, Lncenc1 exon 2 is highly enriched for canonical miRNA responsive elements (MREs). In particular, 22 and 17 elements were detected for microRNAs 128 and 138 MREs, respectively, with a highly significant enrichment ($P < 1E^{-150}$, Supplementary Figure 3 and Supplementary Table 3). Of note, 21 out of the 22 miR-128 MREs are of the 7mer-8 type, containing a perfect match to the 6 nucleotides of the miRNA seed plus a further match at position 8. Conversely, 12 out of the 17 miR-138 MREs are of the 8mer-1a type, showing the seed match flanked by both a match at position 8 and an A at position 1, while 5 are 7mer-1a, with a seed match and an A at target position 1b [33]. Interestingly, 20 out of the 22 miR-128 MREs display a regular 22 nucleotides spacing, suggestive of a functional role. Both miRNAs are expressed in naïve mESCs but also at different stages of differentiation toward the neuronal lineage, to become then enriched in the brain [34]. Together with its cytoplasmic localization, the Lncenc1 high MRE content suggests that its functions in maintaining ESCs naïve state could be at least partly due to buffering the activities of microRNAs, in particular miR-128 and miR-138, protecting their target mRNAs from silencing. Supporting this idea, we could show that the MRE-containing region is responsible for ESCs pluripotency maintenance, since its ectopic expression in E14 cells was sufficient to completely rescue the differentiation triggered by Lncenc1 silencing (Figure 4C, D and Supplementary Figure 4). Of note, miR138 levels were significantly affected upon Lncenc1 depletion, while miR-128 expression was not affected (Supplementary Figure 5). This is not surprising, since sponges are known to inhibit miRNAs activity by sequestering them from their endogenous targets, not necessarily implicating their degradation/downregulation.

Lncenc1 acts by preventing miRNA-mediated suppression of Klf4 protein

In order to identify candidate Lncenc1 partner mRNAs, we took advantage of the prediction algorithm described in Karreth et al. (2015) [35], which is based on the relative expression of the different MRE-containing RNAs. For each mRNA sharing at least one MRE with Lncenc1, a score was computed that indicated the probability of being susceptible to Lncenc1 expression fluctuations. Only mRNAs and microRNAs actually expressed

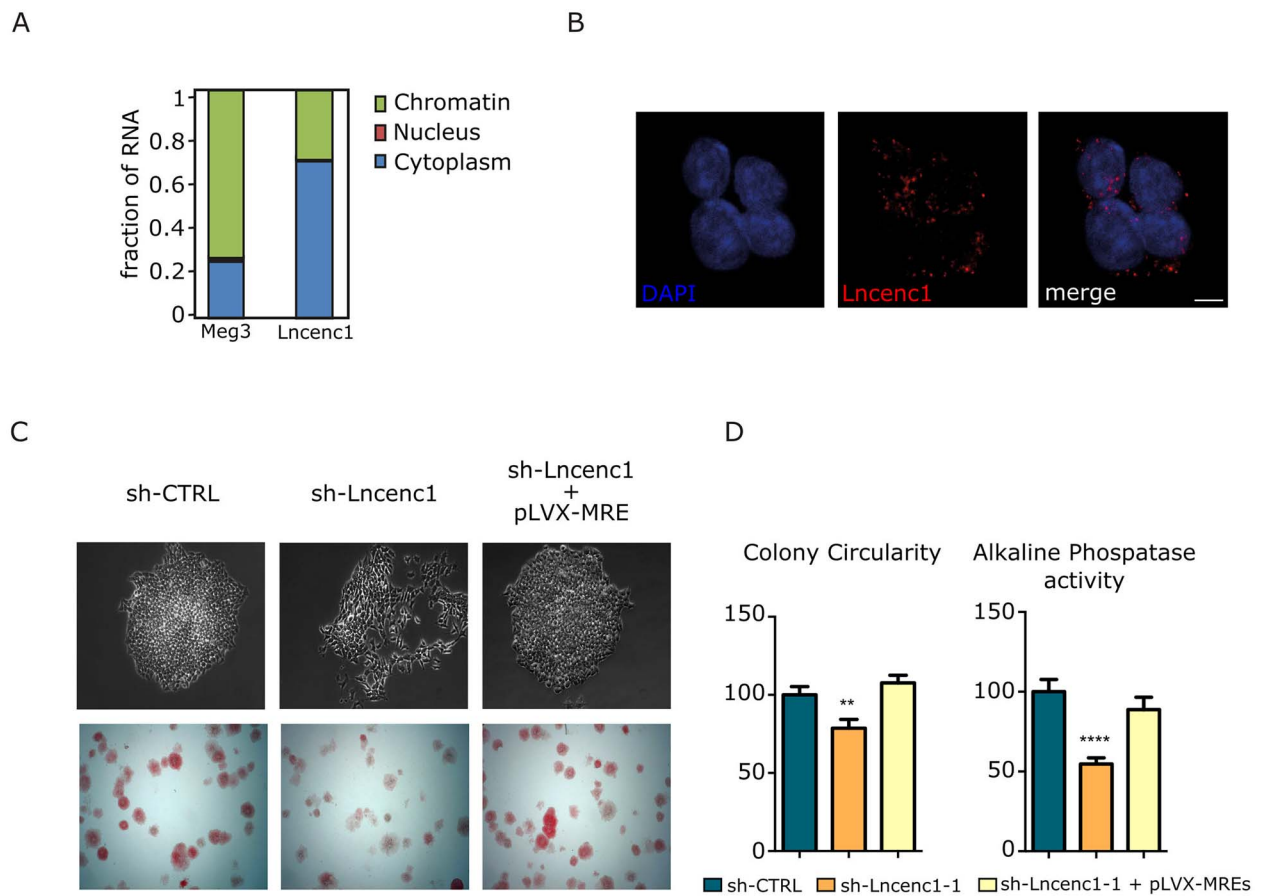


Figure 4. Lncenc1 localization and functional assessment. **(A)** Lncenc1 RNA was quantified by RT-qPCR on total RNA extracted from the indicated subcellular fractions of E14 cells, and the percentage of RNA localized to the indicated fractions is shown. The prevalently chromatinic Meg3 lincRNA and the cytosolic Actin mRNA (not shown) were also measured as controls. **(B)** Lncenc1 RNA-FISH and DAPI staining in E14 cells. **(C)** The differentiation triggered by Lncenc1 silencing in E14 cells was rescued by co-transfecting the pLVX-MRE construct, overexpressing the Lncenc1 MRE region, together with a Lncenc1 shRNA. Top panel: phase contrast imaging of representative colonies. Bottom panel: staining for alkaline phosphatase activity. **(D)** Quantification of morphology disruption (Colony Circularity) and of alkaline phosphatase staining. Data are shown as mean \pm SEM of at least three independent experiments. ** $P < 0.001$, **** $P < 0.0001$.

disrupted, supporting the idea of Klf4 mRNA being a direct target for this microRNA, the availability of which is, in turn, regulated by Lncenc1 levels. Further confirming an RNA interference-based mechanism, Lncenc1-mediated modulation of luciferase activity was partially abolished by Dicer1 silencing (Figure 5D and Supplementary Figure 7A). Accordingly, the expression of an miR-128 mimic could recapitulate the effects of Lncenc1 silencing on luciferase expression (Figure 5D). Finally, overexpression of the MRE region could also rescue the downregulation of Klf4 protein levels triggered by Lncenc1 silencing (Supplementary Figure 7B). Taken together, these data demonstrate that Lncenc1 contributes to the maintenance of ESCs identity at least partly by protecting Klf4 mRNA from miR-128-mediated silencing. Although our results do not exclude the implication of other target mRNAs in addition to Klf4 in the Lncenc1-microRNA-mRNA network, downregulation of this key pluripotency transcription factor may well explain the differentiation phenotype triggered by Lncenc1 silencing. In perfect agreement, the phenotypic effects of Lncenc1 silencing were completely rescued by Klf4 ectopic expression (Figure 5E). Moreover, among all transcription factors known to be active in ES cells, Klf4 is the one able to bind the highest number of genes down-regulated upon Lncenc1 interference, as shown by ChIP-Seq analysis (Figure 5F), supporting a central role for Klf4 downregulation in determining the gene expression patterns following Lncenc1 silencing.

DISCUSSION

Pluripotent ESCs represent valid systems to model mechanisms controlling stemness *versus* differentiation, with important implications in development, regenerative biology and cancer. To fully exploit their potential, however, the complete understanding of the molecular mechanisms underlying stemness is crucial. ESCs pluripotency can be maintained thanks to an intricate network of signaling pathways and transcription factors. Among these, the LIF-STAT3 pathway plays a crucial role in both human and mouse ESCs to elicit a ground state of pluripotency akin to the early epiblast state of the pre-implantation embryos [13, 15].

Our bioinformatic approach has allowed us to identify a set of STAT3-dependent lincRNAs abundantly and specifically expressed in mESCs (LincS). One of them, LincS3/Lncenc1, had already been described as involved in maintaining ESCs pluripotency [28, 30]. Our data indicate that Lncenc1 is indeed a direct STAT3 transcriptional target and has a functional impact on the maintenance of the pluripotent state of ESCs, as shown by cell differentiation upon Lncenc1 silencing. Notably, in addition to being required for LIF-dependent Lncenc1 expression, STAT3 is also able to drive Lncenc1 transcription independently of LIF.

We could show that Lncenc1 localizes predominantly to the cytoplasm by both cell fractionation and *in situ* hybridization experiments. Nevertheless, a small fraction could also be

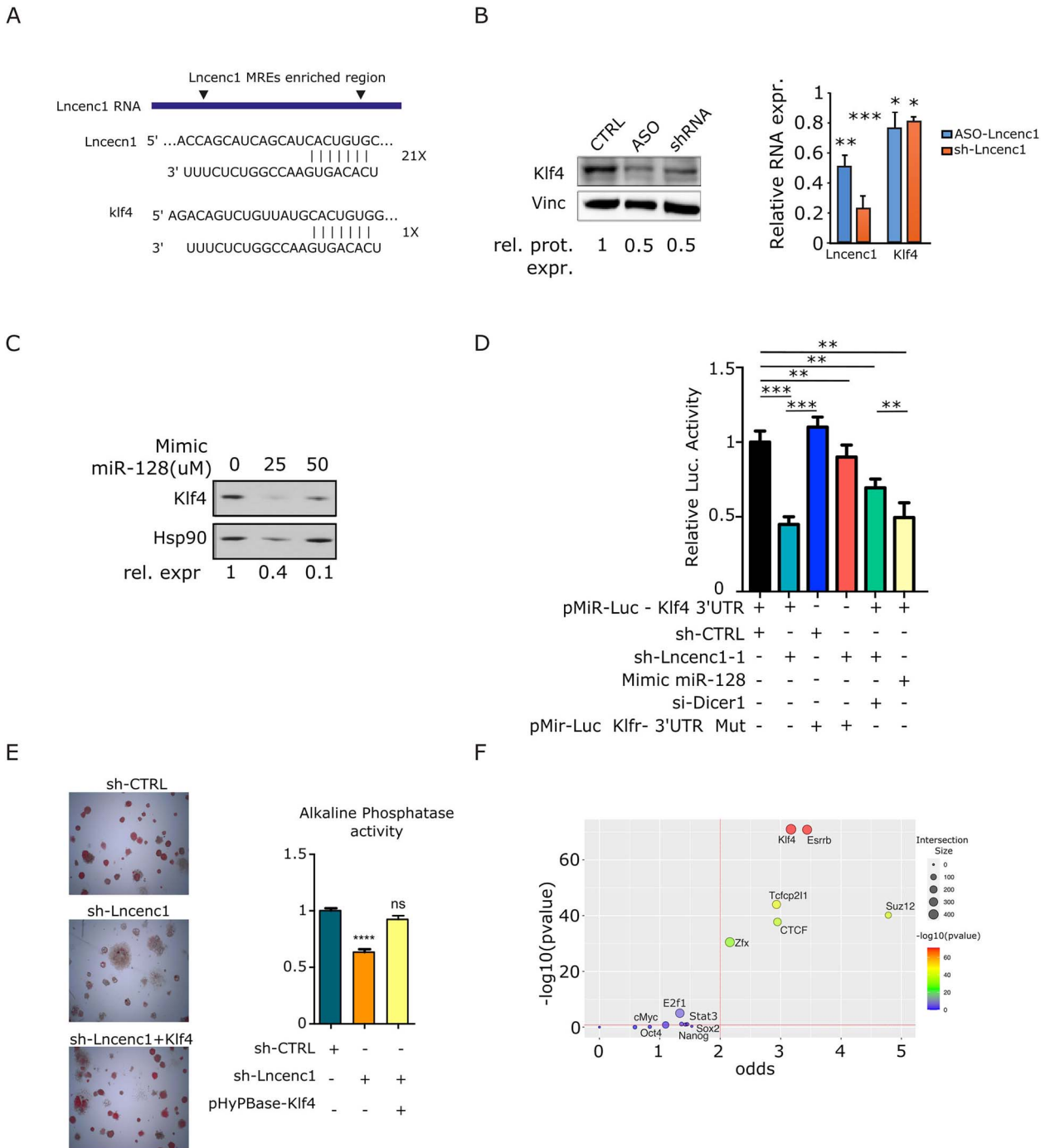


Figure 5. Lncenc1 regulates KLF4 expression by sequestering miR-128. (A) Schematic of Lncenc1 RNA, showing the MRE-enriched region between the arrowheads. Below, representative seed-pairing sites for miR-128 in the Lncenc1 RNA and the Klf4 3'UTR are shown. (B) Klf4 protein and RNA levels were measured in E14 ESCs, silenced or not for Lncenc1 by the indicated treatments. Numbers below the Western blot panel represent the relative Klf4 expression upon normalization to the Vinculin internal control, representative of three independent experiments. Fold reduction was 50% +/- 5% (ASO) and 50% +/- 4% (shRNA). The histograms represent Lncenc1 and Klf4 RNA levels measured by RT-qPCR in the same samples. Data are mean ± SD of three independent replicates. Values were normalized to the 18S internal control and represented as fold change relative to control cells. The asterisks indicate statistically significant differences (two-tailed paired t-test) between the indicated groups and corresponding control; *P < 0.01; **P < 0.001; ***P < 0.0001. (C) Western blot showing Klf4 protein downregulation upon miR-128 mimic transfection in E14 ESCs. Numbers below the panel represent the relative Klf4 expression upon normalization to the HSP90 internal control. Representative of three independent experiments. (D) Dual luciferase assay in E14 cells transiently co-transfected with the Renilla and the pMIRLuc-Klf4 3'UTR/Mut vectors, together with either control or Lncenc1 shRNA, the miR-128 mimic or Dicer1/control siRNA. Data are mean ± SD of at least three independent experiments. Asterisks indicate statistically significant differences between the indicated groups. ***P < 0.0001; **P < 0.001. (E) E14 cells stably expressing Klf4 and silenced or not for Lncenc1 were stained for Alkaline Phosphatase activity. The quantification shows mean ± SEM of at least three independent experiments. ****P < 0.0001; ns, not significant. (F) Scatter plot showing the *in vivo* binding enrichment of transcription factors known to be active in mES cells (see Materials and Methods) at promoters of genes downregulated upon Lncenc1 silencing. Each dot represents a TF. The Y-axis shows the log transformed (-log10) significance of enrichment compared to random peaks computed using one-sided Fishers exact test. The X-axis represents the odds of finding a given number of promoters overlapping with a given TF. The size of the dots represents the intersection size. The indicated TFs were considered enriched if falling within the top right area delineated by dotted lines.

