Cell Death & Differentiation (2020) 27:345–362
https://doi.org/10.1038/s41418-019-0359-9

Cell Death & Differentiation

ARTICLE

A novel member of Prame family, *Gm12794c*, counteracts retinoic acid differentiation through the methyltransferase activity of PRC2

Giuliana Napolitano[1] · Daniela Tagliaferri[2] · Salvatore Fusco[1] · Carmine Cirillo[1] · Ilaria De Martino[1] · Martina Addeo[1] · Pellegrino Mazzone[2] · Nicola Antonino Russo[2] · Francesco Natale[3,6] · Maria Cristina Cardoso[3] · Luciana De Luca[4] · Daniela Lamorte[4] · Francesco La Rocca[4] · Mario De Felice[5] · Geppino Falco[1,2,5]

Received: 11 September 2018 / Revised: 11 February 2019 / Accepted: 3 April 2019 / Published online: 11 June 2019

Abstract

Embryonic stem cells (ESCs) fluctuate among different levels of pluripotency defined as metastates. Sporadically, metastable cellular populations convert to a highly pluripotent metastate that resembles the preimplantation two-cell embryos stage (defined as 2C stage) in terms of transcriptome, DNA methylation, and chromatin structure. Recently, we found that the retinoic acid (RA) signaling leads to a robust increase of cells specifically expressing 2C genes, such as members of the Prame family. Here, we show that *Gm12794c*, one of the most highly upregulated Prame members, and previously identified as a key player for the maintenance of pluripotency, has a functional role in conferring ESCs resistance to RA signaling. In particular, RA-dependent expression of *Gm12794c* induces a ground state-like metastate, as evaluated by activation of 2C-specific genes, global DNA hypomethylation and rearrangement of chromatin similar to that observed in naive totipotent preimplantation epiblast cells and 2C-like cells. Mechanistically, we demonstrated that *Gm12794c* inhibits *Cdkn1A* gene expression through the polycomb repressive complex 2 (PRC2) histone methyltransferase activity. Collectively, our data highlight a molecular mechanism employed by ESCs to counteract retinoic acid differentiation stimuli and contribute to shed light on the molecular mechanisms at grounds of ESCs naive pluripotency-state maintenance.

These authors contributed equally: Giuliana Napolitano, Daniela Tagliaferri

Edited by Y. Shi

Supplementary information The online version of this article (https://doi.org/10.1038/s41418-019-0359-9) contains supplementary material, which is available to authorized users.

✉ Giuliana Napolitano
giuliana.napolitano@unina.it

✉ Geppino Falco
geppino.falco@unina.it

[1] Department of Biology, University of Naples 'Federico II', Naples, Italy

[2] Biogem, Istituto di Biologia e Genetica Molecolare, Via Camporeale, Ariano Irpino (AV), Italy

[3] Department of Biology, Technische Universität Darmstadt, 64287 Darmstadt, Germany

[4] IRCCS-CROB, Referral Cancer Center of Basilicata, Rionero in Vulture, Italy

[5] IEOS, CNR, Naples, Italy

[6] Present address: IRBM Science Park S.p.A., Biology Unit, Naples, Italy

Introduction

ESCs colonies are characterized by different metastable sub-populations (metastates) marked by the presence of specific factors whose expression profoundly affects the state of cells pluripotency [1–6]. Interestingly, metastable sub-populations may occasionally convert to a highly pluripotent metastate that resembles the two-cell stage (2C) embryos during which an enormous rearrangement of chromatin takes place causing a reprogramming of transcription called zygotic genome activation (ZGA) [7–9]. The transcriptional signature of cells at the 2C stage is characterized by the expression of murine endogenous retrovirus genes with Leucine tRNA primer (MuERV-L, MERVL, or ERV4) [10–12] and members of the Zscan4, Prame, Thoc, and Tcstv gene families [13–16].

ESCs that in vitro sporadically convert to the 2C-like metastate are functionally ground state (hereafter 2C-like cells), represent a small fraction of the cellular population and closely resemble cells at the 2C-stage embryos, both in terms of global gene expression and epigenetic

modifications [8, 9, 17]. Recently, we [18] and others [19] demonstrated that retinoic acid (RA) induces a metastable population of cells (*Zscan4c* metastate) that specifically expresses genes of the 2-cell embryos developmental stage. Among these, there are genes of the Prame family that encode for leucine-repeat rich (LRR) proteins as their peptide sequences contain LXXLL motifs, also called nuclear receptors boxes (NR boxes) [20]. Interestingly, the action of RA hinges on nuclear receptors (NRs), a family of ligand-regulated transcription factors that control a wide range of developmental processes, called retinoic acid receptors (RARs). RARs have modular structures and exploit their functions by homo- or hetero-dimerization [21]. However, a number of co-regulators control the transcriptional activity of RARs in a ligand-dependent manner, either acting as corepressors or coactivators. LRR proteins directly interact with NRs through LXXLL motifs, and indeed many of them are RARs co-regulators [20]. Accordingly, human PRAME has been shown to modulate the activity of RAR alpha [22].

Here, we present data showing that *Gm12794c*, a member of the Prame family, is a key player in counteracting RA-dependent differentiation. We found that, in the presence of RA, *Gm12794c* led to high levels of 2C-specific genes transcription and contributed to the overall DNA hypomethylation and global increase of H3K27 acetylation levels. Mechanistically, we highlighted a RA-dependent molecular mechanism at the basis of naive pluripotency maintenance, whereas *Gm12794c* enables ESCs to overcome RA-dependent differentiation by inducing 2C-like cellular metastate throughout the PRC2-mediated transcriptional repression of the RA-responsive *Cdkn1A* gene expression.

Experimental procedures

Cell cultures, treatments, transient transfections, and Luciferase assay

E14 Rosa26^{Tm_Gm12794c} ES cells, derived from strain 129P2/OlaHsd, were cultured in gelatin-coated dishes in complete ES medium: DMEM (Dulbecco's Modified Eagle's Medium, Gibco), 15% fetal bovine serum FBS EuroClone), 1000 U/ml leukemia inhibitory factor (LIF) (EuroClone), 1.0 mM sodium pyruvate (Invitrogen), 0.1 mM nonessential amino acids (Invitrogen), 2.0 mM Glutamax (Invitrogen), 0.1 mM β-mercaptoethanol, and 500 U/ml penicillin/streptomycin (Invitrogen).

Where indicated, doxycycline (Dox) has been used for 3 days at 1.5 μg/ml final concentration.

$_{pGm12794c}$-Strawberry and $_{pGm12794c}$*FLAG-Gm12794c* ES cells were cultured in gelatin-coated dishes in complete ES medium: GMEM (Glasgow Minimum Essential Medium, Gibco), 15% fetal bovine (FBS EuroClone), 1000 U/ml leukemia inhibitory factor (LIF) (EuroClone), 1.0 mM sodium pyruvate (Invitrogen), 0.1 mM nonessential amino acids (Invitrogen), 2.0 mM L-glutamine (Invitrogen), 0.1 mM β-mercaptoethanol, and 500 U/ml penicillin/streptomycin (Invitrogen).

For experiments in *2i* medium, E14Tg2a.4 and Rosa26^{Tm_Gm12794c} ES cells were maintained in serum-free N2B27-based medium supplemented with *2i*. In particular, medium was generated by inclusion of the following: Knockout DMEM (Gibco), N2 supplement (Invitrogen), B27 supplement (Invitrogen), 1 mM glutamine (Invitrogen), 1% nonessential amino acids (Invitrogen), 0.1 mM β-mercaptoethanol (Gibco), penicillin–streptomycin (Invitrogen), 1000 U ml-1 leukemia inhibitory factor (LIF) (EuroClone), and the following small molecules: PD0325901 (0.5 μM, Stemgen), and CHIR99021 (3.0 μM, Stemgen).

ESCs were incubated at 37 °C in 6% CO_2; medium was changed daily and cells were split every 2 to 3 days routinely.

Where indicated, cells were treated with RA for 4 days at 2 μM final concentration.

NIH3T3 cells were cultured in DMEM containing 10% FBS (EuroClone) and 1% penicillin/streptomycin at 37 °C with 5% CO_2. Transient transfections and luciferase assays have been performed as previously described [23]. Briefly, NIH3T3 cell were transiently transfected with the LipofectAMINE 2000 reagent (Gibco-BRL, Carlsbad, CA) in 35 -mm/dish in multiwells, using 100 ng of reporter DNA (pr21-Luc) with increasing amount of FLAG-*Gm12794c*, as indicated in the text and 20 ng of Renilla luciferase expression plasmid (pRL-CMV, Promega Corporation, Madison, WI) for normalization of transfections efficiencies. Cells were harvested 48 h after transfections, and extracts were assayed for luciferase activity using Dual-Luciferase Reporter assay (Promega). Where indicated, RA was added to cells at final concentration of 10 μM for 24 h.

Cells were checked routinely for the absence of mycoplasma. All experiments were performed at least three times.

Generation of doxycycline-inducible Rosa26^{Tm_Gm12794c} cell line

A2lox.Cre mouse ESCs (a gift of Prof. Kyba) were routinely cultured in DMEM (Invitrogen) supplemented with 15% ES-certified FBS (Invitrogen), 0.1 mM nonessential

amino acids (Invitrogen), 1 mM sodium pyruvate (Invitrogen), 0.1 mM β-mercaptoethanol (Sigma), 50 U ml^{-1} penicillin/50 µg ml^{-1} streptomycin (Invitrogen) and 1000 U ml^{-1} LIF (ESGRO).

The tetracycline-inducible *Gm12794c* ESC line was generated as previously described [24]. Briefly, the coding sequence of *Gm12794c* was amplified from an available plasmid and cloned into p2Lox targeting vector. In total, 5×10^6 mESCs were electroporated with the *Gm12794c*-p2Lox vector allowing the unidirectional recombination of the transgene into the HPRT locus. Positive clones were isolated using 275 µg/ml neomycin (Invitrogen) selection. Clones were screened by qRT-PCR after 72 h in the absence or presence of 1.5 µg/ml doxycycline to verify transgene expression.

$pr_{Gm12794c}$FLAG-*Gm12794c* vector construction

To generate the $pr_{Gm12794c}$FLAG-*Gm12794c* vector, a DNA fragment containing the *Gm12794c* coding sequence was amplified from an available plasmid with primers NotI-RNIc3F (5′-gcggccgctatgagcacctacaaccct cc-3′) and BamHI-RNIc4R (5′-ggatccaacttctctttgctgcca ac-3′), and then cloned into 3xFlag-CMV-10 vector using NotI and BamHI restriction sites. 3xFlag-Gm12794 was amplified with the couple of primers EcoRV-RNIc (5′-GATATCGACTACAAAGACCATGACGG-3′) and Xho1-RNIc (5′-CTCGAGAATTCAACAGGCATCTACT G-3′); this fragment was inserted in the available $pr_{Gm12794c}$-PCR-XL-TOPO vector using EcoRV- and XhoI-restriction sites. All the passages were verified by sequence analysis.

Generation of E14tg2$^{prGm12794c_FLAG\text{-}Gm12794c}$ and E14tg2^{prGm12794c}–Strawberry cell lines

To generate the $pr_{Gm12794c}$*FLAG-Gm12794c* and $pr_{Gm12794c}$Strawberry cell lines, 5×10^6 cells in suspension were electroporated with 3 µg of linearized $p_{Gm12794c}$FLAG-*Gm12794c* or $pr_{Gm12794c}$-Strawberry vectors and plated in gelatin-coated 100-mm dishes. Cells were selected with 250 µg/ml G418. ES cell colonies were picked on the 8th day, expanded, and frozen for further analysis.

p21-Luc vector construction

To generate the p21-Luc plasmid vector, a 150-bp DNA fragment encompassing the RARE element and the TSS of the p21/*Cdkn1A* promoter was amplified from the mouse genomic DNA and inserted into pGL3 plasmid vector (Promega) using HindIII and SacI restriction sites. All the passages were verified by sequence analysis.

pcDNA3_pr*Gm12794c*_LNGFR plasmid vector construction

For the construction of the plasmid pcDNA3_pr*Gm12794c*_LNGFR, *Gm12794c* promoter (5080 bp) was amplified by PCR from p*Gm12794c*-Strawberry vector and inserted into KpnI/EcoRV sites of pcDNA3 vector (Invitrogen). Subsequently, LNGFR fragment (874 bp) was amplified by PCR from pPRIME-CMV-LNGFR and inserted into EcoRV site of pcDNA3_p*Gm12794c* vector. The construct was verified by sequencing.

Generation of E14tg2^{prGm12794c_LNGFR} stable cell line

To generate the stably transfected ES cell line, pr*Gm12794c*_LNGFR was linearized with KpnI and transfected into wild-type E14tg2. Forty-eight hours after transfection, the cells were splitted, and positive clones selected for Neo resistance. After 1 week of G418 (Gibco) treatment, Neo^r clones were picked and propagated.

E14tg2^{prGm12794c_LNGFR} culture and magnetic separation

The stably transfected ESCs were cultured for 3 days on gelatin-coated dishes in ES complete medium. The cells were then trypsinized and plated onto gelatin-coated dishes in the N2B27-VitA medium: knockout DMEM high glucose (Gibco) supplemented with L-glutamine 2 mM (Gibco), penicillin/streptomycin 100 U-µg/ml (Gibco), B27-VitA Supplement 1 × 10(Gibco), N2 Supplement 1× (Gibco), 2(β)Mercaptoethanol 0.1 mM (Gibco), LIF 1000 U/ml (Millipore), G418 137.5 µg/ml, with or without 2 µM RA for 72 h. For magnetic labeling, single-cell suspensions were centrifuged, resuspended in PBS supplemented with 5 mM EDTA and 0.5% BSA, and incubated with MACSelect (TM) LNGFR MicroBeads for 15 min on ice. Magnetically labeled cells were isolated over the AutoMACS Pro Separator (Miltenyi Biotec) with "posseld2" program, according to the manufacturer's protocol. For purity assessment, aliquots of original cell population (magnetically labeled cells before separation), eluted positive (enriched target cells) and negative cell populations (untargeted cells collected in the flow-through fraction), were fluorescently stained with MACSelect control FITC antibody (Miltenyi Biotec) that specifically stains MACSelect MicroBead-labeled cells and analyzed by Navios Flow Cytometer (Beckman Coulter).

Phylogenetic analysis and sequences alignment

Peptide sequences were aligned using the ClustalW tool in the Molecular Evolutionary Genetic Analysis version

7 software package (MEGA7; www.megasoftware.net/). Phylogenetic trees were constructed using the maximum-parsimony method with a bootstrap test of 1000 replicates and the Subtree–Pruning–Regrafting (SPR) algorithm.

Multiple sequence alignments were performed with ClustalW program (NCBI) using default parameters, except that each alignment step was iterated.

Sequences compared for both analyses were as follows:
hPRAME:NM_006115
Gm12794a:NM_001033790.3
Gm12794b:NC_000070.6
Gm12794c:NM_001085516
Pramef25:NM_001126315
Gm7971:XM_982184.4
Gm16429:NC_000071.6
Gm10424:XM_001476548.4
BC061212:NP_796161.3
Gm13119:NM_001034101
Gm13057:NM_001113735
BC080695:NM_001007579

Data set analysis

The preimplant (GSE936 and GSE1749) and the Emerald (GSE75977) data sets were downloaded from GEO [18, 25–27]. In order to merge the GSE936 and GSE75977 data sets, considering they belong to two different platforms of the same producer (Agilent), intersection of the common genes and batch-effect removal was necessary. Only interesting cell samples were retrieved. Multidimensional scaling plot [28] of the merged data sets has been performed to depict the relative distances between the different cell samples. Most variable genes have been detected using the interquartile range (IQR), and the Top 100 has been used to test the correlation between the different cell samples. The correlation matrix was graphically displayed [29], and the genes reported in an excel file (Supplementary Table 1). In addition, a merge of data sets was necessary for the GSE1749 data set, considering it consists of two different platforms of the same producer (Affymetrix) covering half of the genes each. Differentially expressed genes (2C vs the others) have been obtained by fitting a linear model for each gene, applying empirical Bayes smoothing to the standard errors and summarizing and selecting the top-ranked genes for any given contrast [29]. A threshold based on the lowest expression internal chip control (BioB) has been defined. Differentially expressed genes have been retained if above threshold in 20% of the samples. A selection of genes, both characterizing *Zscan4*$^{+}$/*Gm12974*$^{+}$ cells and differentially expressed in 2-cell stage cells, were graphically displayed. All the analysis have been performed in R [30].

RNA extraction and qRT-PCR quantification

RNA extraction and qRT-PCR analyses have been performed as previously described [31, 32]. Briefly, RNA was extracted from cells using EuroGold Trifast (EuroClone). cDNA was generated using Quantitec Reverse Transcription Kit (Qiagen), according to the manufacturer's protocol. Quantitative analysis was performed using SYBR Green 2X PCR Master Mix (Applied Biosystem). Each sample was run in triplicate and normalized to the expression of housekeeping glyceraldehyde 3-phosphate dehydrogenase (GAPDH) gene. Primers are listed in Supplementary Table 1.

Immunofluorescence (anti-H3K27ac, anti-H3K27me3, anti-5mC, and anti-p21)

Immunofluorescence analyses have been performed as previously described [33, 34].

Briefly, cells were fixed and stained with primary antibodies for 30 min at 37 °C after brief permeabilization (0.1%Triton X-100/PBS) and pre-blocking (2% BSA–3% NRS) steps. After extensive washes, anti-rabbit-Cy3 and anti-rabbit-FITC secondary antibodies (Jackson Immunotech, West Grove, PA) have been used at the dilution of 1:400 for 30′ at room temperature. Nuclei were stained with DAPI (Invitrogen Corporation).

For anti-5mC immunofluorescence, cells were treated with 4 N HCl for 10 min before the permeabilization step. Images were acquired using Nikon Eclipse TE 2000-U or using a confocal ZEISS LSM 700 microscope (individual planes; $z = 700$–800 nm), and digitally processed using FIJI image analysis software.

3D-RNA FISH and image analysis

RNA FISH analysis protocol has been already described [35, 36] and adapted here using few modifications. Briefly, cells grown on gelatin-coated coverslips were fixed and treated with 0.5% Triton X-100/PBS for 10 min on ice followed by denaturation with 2xSSC/50% formamide 10 min at 80 °C. Hybridization was carried out using heat-denatured fluorescently labeled LNA probe directed toward the forward major satellite transcript (0.2 µM, Exiqon) in 40% formamide (Sigma), 2x saline sodium citrate (SSC) (Sigma), 10% dextran sulfate (Fluka), 10 mM vanadyl ribonucleoside complex, and 2 mg/ml BSA (NEB) for 35 min at 37 °C and washed with 0.1 × SSC at 60 °C. Nuclei were stained with DAPI (Invitrogen Corporation). Images were acquired using a confocal ZEISS LSM

700 microscope (individual planes; $z = 700–800$ nm) and digitally processed using FIJI image analysis software.

Western blot (WCE and fractionated extracts)

Whole-cell extracts (WCE) for western blot analyses have been performed as previously described [37]. Briefly, WCE were obtained using buffer F (10 mMTris–HCl pH 7.5, 150 mM NaCl, 30 mM $Na_4O_7P_2$, 50 mM NaF, 5 mM $ZnCl_2$, 0.1 mM Na_3VO_4, 1% Triton, and 0.1 mM PMSF). In total, 50 µg of protein extracts were loaded and separated by SDS–PAGE.

Cell fractionation into cytoplasmic and nuclear fractions was performed as described [35]. Cells are collected, washed, and then resuspendend in buffer A (10 mM HEPES pH 7.9, 10 mM KCl, 0.1 mM EGTA, 0.1 mM EDTA, 1 mM DTT, and 0.5 mM PMSF) by gentle pipetting. Then NP-40 is added, and the omogenate is centrifuged. The supernatant containing the cytolplasmic extracts is transferred into fresh tubes and stored at −80 °C. Nuclei are resuspendend in 50 µl of ice-cold buffer C (20 mM HEPES pH 7.9, 0.4 M NaCl, 1 mM EDTA, 1 mM EGTA, 1 mM DTT, 1 mM PMSF). After centrifugation, supernatants are stored at −80 °C.

Equal cell volume from nuclear and cytoplasmic fractions were resolved by SDS–PAGE. Antibodies used are listed in Supplementary Table 2. Binding was visualized by enhanced chemiluminescence (ECL-plus Kit, Amersham, Buckinhamshire, UK).

Chromatin immunoprecipitation

ChIP analyses have been performed as previously described [31]. After cross-linking (formaldehyde 1% 10 min at room temperature followed by 0.125 M glycine), cells were harvested and pelleted for lysis in the following buffer: Pipes 5 mM pH 8, KCl 85 mM, and NP-40 0.5%. Then nuclei were harvested by centrifugation, incubated in nuclear lysis buffer (50 mM Tris pH 8.1, 10 mM EDTA, and 1% SDS), and sonicated so as to obtain DNA fragments of about 300–700 bp. Samples were diluted in dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris pH 8.1, and 167 mM NaCl), and 10 µg of chromatin were incubated overnight at 4 °C with specific antibody or without antibody as a negative control. Immune complexes were then recovered by using 30 µl of blocked protein A magnetic beads (Dynabeads Protein A, Invitrogen by Thermo Fisher Scientific) for 2 h at 4 °C on a rotating wheel. Beads were washed once in dialysis buffer (2 mM EDTA, 50 mM Tris pH 8, 0.2% Sarkosyl) and four times in wash buffer (100 mM Tris pH 8.8, 500 mM LiCl, 1% NP-40, 1% NaDoc). Elution from the beads was achieved by incubation in elution buffer (1% SDS, 100 mM $NaHCO_3$) for 15 min. Cross-link was reversed by adding NaCl and RNase A to the samples and incubating overnight at 62 °C. After a 2 h of proteinase K treatment, DNA was purified. Immunoprecipitated DNA and input DNA were analyzed in triplicate by qRT-PCR (primer sequences are provided in Supplementary Table 1).

Results

Phylogenetic analysis of Prame family genes expressed in the RA-induced *Zscan4* metastate

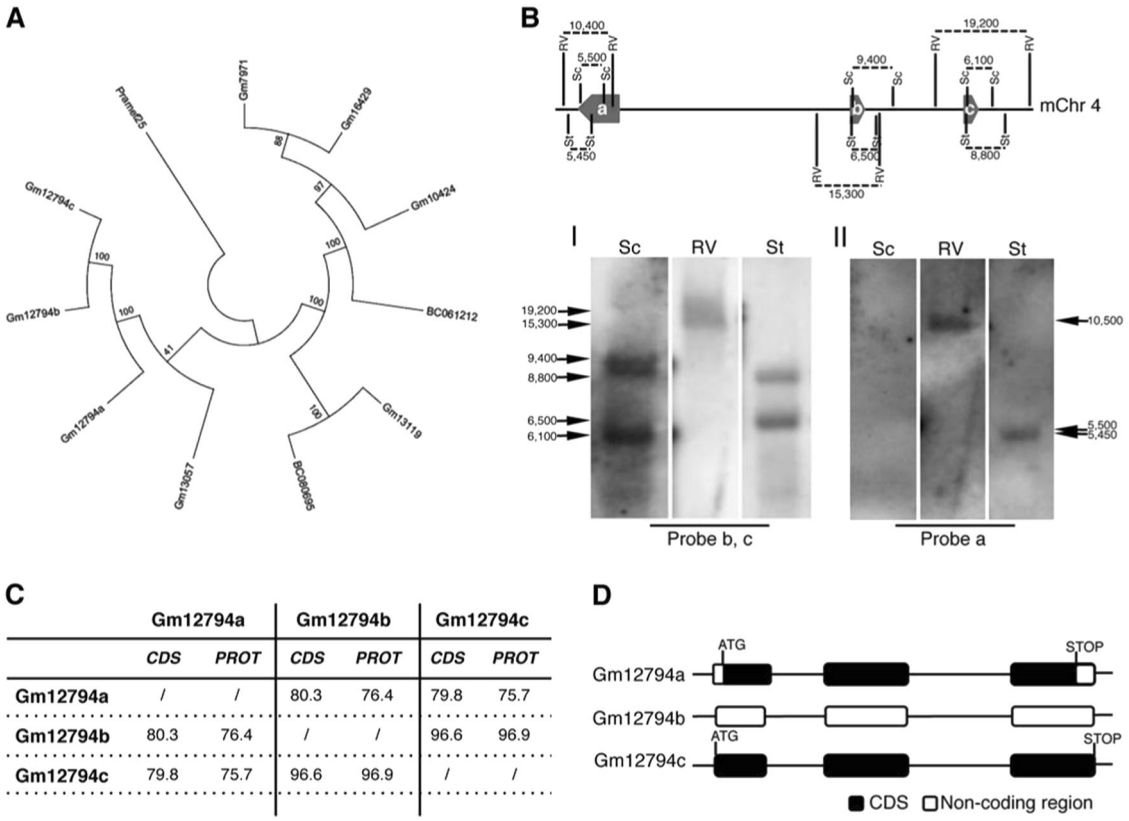
Recently, we demonstrated that RA treatment induces the *Zscan4c* metastate up to 20% of cellular population (in absence of RA *Zscan4c* metastate cells are less than 5% of total) [6, 18], as later confirmed by others [19]. The global gene expression profile of the *Zscan4c* metastate significantly overlapped with the 2-cell embryos transcriptional signature [18]. Interestingly, particularly enriched were *Pramef25* and ten predicted gene models (*Gm12794c*, *Gm12794a, Gm12794b*, *Gm7971*, *Gm16429, Gm10424, BC061212, Gm13119, Gm13057*, and BC080695) encoding for RNI-like LRR-proteins and belonging to the Prame family genes (Supplementary Fig. 1). As they have very similar DNA (data not shown) and putative peptide sequences (Supplementary Fig. 1), we sought to determine their phylogeny. All RA-induced members of the Prame family originated from the ancestor gene *Pramef25* (Fig. 1a).

As *Gm12794c*-expressing cells are a limiting factor for the pluripotency of mESCs colonies [6, 18], we characterized its genetic locus. *Gm12794c* is located on murine chromosome 4 in close proximity to *Gm12794a* and *Gm12794b* forming a cluster (Fig. 1b) spanning from nt 101834969 to nt 101943183. *Gm12794a* and *Gm12794b* have similar coding (79.8% and 96.6%, respectively) and similar putative protein sequences (75.7% and 96.9%, respectively) to *Gm12794c* (Fig. 1c). Although RNA is expressed from the three genes in RA-treated mESCs, *Gm12794b* does not show an open-reading frame (Fig. 1d). *Gm12794a* is the only one having both 5′- and 3′-UTR sequences suggesting that *Gm12794c* and *Gm12794b* derived from retrotransposition events.

Collectively, RA signaling induces a 2C-like transcriptional signature that is enriched with genes belonging to the *Prame* family among which particularly expressed is *Gm12794c*.

Gm12794c induces 2C-like mESCs in response to retinoic acid

To elucidate the role of *Gm12794c* during RA signaling, we employed mESCs to generate a recombinant Rosa26 locus



	Gm12794a		Gm12794b		Gm12794c	
	CDS	*PROT*	*CDS*	*PROT*	*CDS*	*PROT*
Gm12794a	/	/	80.3	76.4	79.8	75.7
Gm12794b	80.3	76.4	/	/	96.6	96.9
Gm12794c	79.8	75.7	96.6	96.9	/	/

Fig. 1 Phylogenetic analysis of RA-induced *Prame* genes and characterization of *Gm12794c* genetic locus. **a** A maximum parsimony tree reflecting similarities among the RA-induced murine *Prame* genes has been generated using MEGA7. The tree is based on the bootstrap method for phylogenetic testing with a number of bootstrap replications equal to 1000. The Subtree–Pruning–Regrafting (SPR) validated the tree topology. Numbers indicate phylogenetic distance. **b** (Up) A schematic representation of the *Gm12794* genetic locus on murine chromosome 4 is shown. Relative positions of EcoRV (RV), SacI (Sc) and StuI (St) DNA sites, and digestion fragments lengths are indicated. Thick arrows labeled as a, b, and c represent *Gm12794a*, *Gm12794b*, and *Gm12794c* genes, respectively. (Down) mESCs genomic DNA has been digested with the indicated restriction enzymes and probed with probes b and c to identify *Gm12794b* and *Gm12794c* gene loci (I) and probe a to validate *Gm12794a* locus (II). Arrows indicated the expected fragments. **c** Similarities among coding sequences (CDS) and putative protein sequences of *Gm12794a*, *Gm12794b*, and *Gm12794c* genes are shown as percentage. **d** Schematic representation of coding and non-coding sequences of *Gm12794a*, *Gm12794b* and *Gm12794c* genes

knock-in cell line (Rosa26^{Tm_Gm12794c}), in which the overexpression of *Gm12794c* is inducible by doxycycline (Dox) [24]. Transgene mRNA levels increased up to fivefold after 3 days treatment with Dox, showing that *Rosa26*^{Tm_Gm12794c} ESCs are responsive to Dox induction (Fig. 2a). Of note, in these culture conditions cells showed the classical morphology of pluripotent ESCs (Supplementary Fig. 2).

To highlight the contribution of *Gm12794c* to RA signaling, 2 µM RA was added to *Rosa26*$^{Tm\_Gm12794c}$ cells in the presence (3 days treatment) or absence of Dox (Dox+ and Dox−, respectively). After addition of RA, cells were cultured for 4 days according to the scheme in Fig. 2b.

Remarkably, after RA treatment, the majority of Dox+ *Rosa26*$^{Tm\_Gm12794c}$ cell colonies remained completely undifferentiated. In sharp contrast, *Rosa26*$^{Tm\_Gm12794c}$ cells not overexpressing *Gm12794c* (Dox−) underwent differentiation, although we were able to observe the appearance of a considerable number of secondary colonies (Fig. 2b). It is noteworthy that RA-induced secondary colonies were enriched with cells expressing *Gm12794c* (Supplementary Fig. 3).

Interestingly, we observed that in response to RA, relative expression of the key ground state genes *Eif1a, Gm4340, Tcstv1*, and *Zscan4* increased significantly (about fourfold) in *Gm12794c*-overexpressing cells. In contrast, mRNA levels of crucial pluripotency regulator genes as *Nanog*, *Oct3/4* (*Pou5f1*), and *Rex1* remained quite unchanged in the same experimental conditions (Fig. 2c). Collectively, these findings demonstrated that *Gm12794c* confers mESC resistance to retinol-induced differentiation, as highlighted by the formation of tightly packed and domed colonies. Importantly, the finding that *Gm12794c* favored the expression of ground state and not of pluripotency genes suggests a link between *Gm12794c* expression levels and the induction of 2C-like cells. To address whether *Gm12794c*-expressing cells resemble, at the transcriptomic level, 2-cell stage mouse

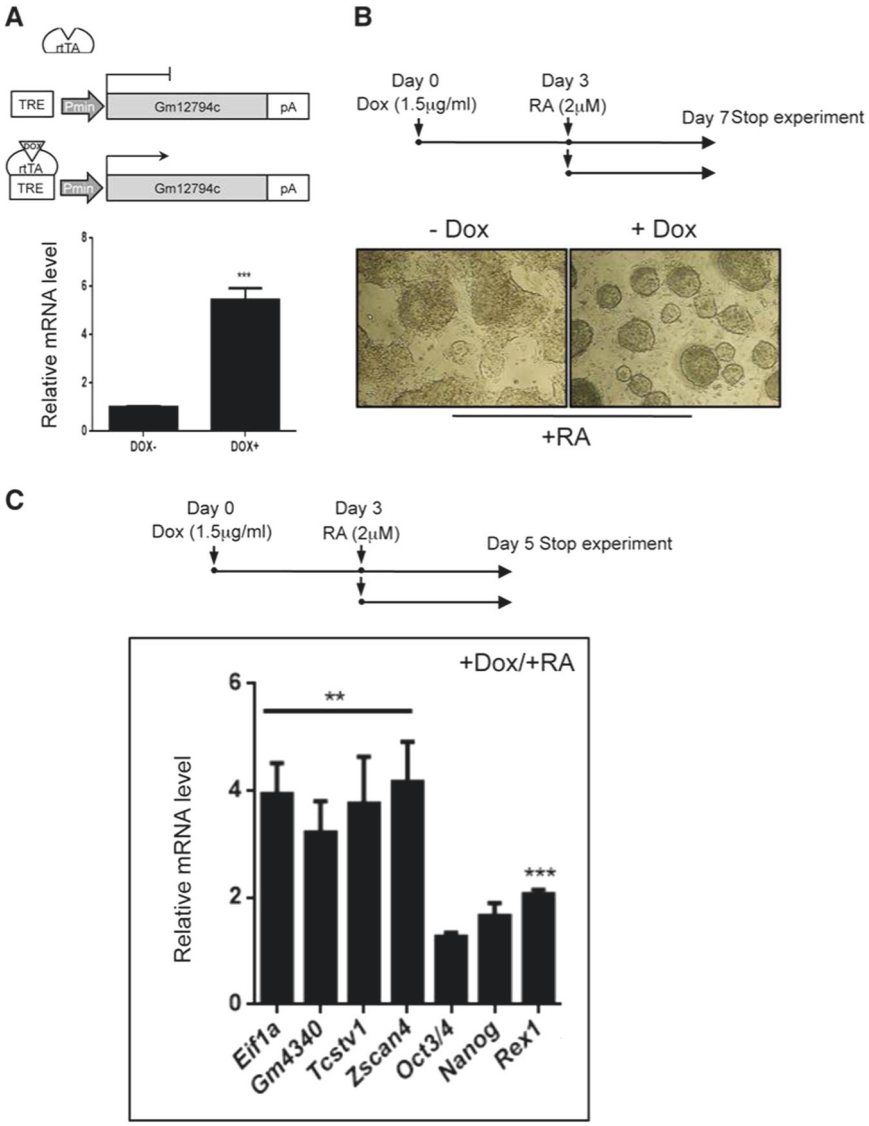


Fig. 2 In response to retinoic acid (RA), *Gm12794c* enhances the expression of 2C genes. **a** (Up) Schematic representation of the minimal Dox-inducible TRE promoter driving the expression of *Gm12794c* in the Rosa26^{Tm_Gm12794c} mESC cell line. (Down) qRT-PCR analysis showing relative *Gm12794c* mRNA levels of Rosa26^{Tm_Gm12794c} mESC cells treated or not with Dox (1.5 μg/ml) for 72 h. **b** (Up) Graphic representation of the experimental design. Rosa26^{Tm_Gm12794c} cells were treated with 2 μM retinoic acid (RA) alone for 96 h or grown in the presence of 1.5 μg/ml doxycycline (Dox) for 72 h and then supplemented with 2 μM RA and cultured for 96 more hours. (Down) Contrast-phase light microscopy images of Rosa26^{Tm_Gm12794c} cells, treated as indicated, showing that *Gm12794c* overexpression confers Rosa26^{Tm_Gm12794c} cells resistance to RA-induced differentiation. **c** (Up) Schematic representation of the experimental design. Rosa26^{Tm_Gm12794c} cells were treated as in **b**), except that cells were grown in the presence of 2 μM RA for 48 h. Rosa26^{Tm_Gm12794c} cells were treated with both Dox/RA according to the scheme in the upper panel then relative mRNAs levels of genes have been analyzed by qRT-PCR and normalized to mRNAs levels of cells grown in RA alone (not shown), indicating that *Gm12794c* expression levels positively correlate with the induction of 2C-genes transcriptional activation. Experiments have been performed at least four times, and each sample has been prepared in triplicate. The data are presented as the mean of independent experiments, and standard deviations are shown. Statistical analysis was performed using Student's *t* test: $^{**}p \le 0.01$, $^{***}p \le 0.001$

embryos, we compared the gene expression profiles of preimplantation mouse embryo stages to *Gm12794c*-enriched cells transcriptome (*Zscan4c*$^+$/*Gm12794c*$^+$ subpopulation) by analyzing publicly released data sets [18, 25, 26] through principal component analysis (PCA) and Correlation Plot. Two-dimensional PCA analysis grouped *Zscan4c*$^+$/*Gm12794c*$^+$ mESCs close to two-cell-stage mouse embryos (Supplementary Fig. 4A). Furthermore, Correlation Plot analysis showed that among preimplantation stages, *Gm12794c*-enriched cells robustly correlated with 2-cell stage embryos (Supplementary Fig. 4B), strongly supporting the PCA data.

In addition, we analyzed publicly available 2-cell stage data sets [25] and checked the expression of six among the top ten highly expressed genes (listed in Supplementary Table 1) in *Gm12794c*$^{+}$ cells homogeneously isolated from RA-treated *mESC*$^{pr\_Gm12794c\_Strawberry}$ cells through FACS [18]. Overall bioinformatics studies were consistent with quantitative transcript analyses revealing that *Gm12794c*$^{+}$ cells express a significant number of 2C-exclusive transcripts (Supplementary Fig. 5A, B). Collectively, these data strongly suggest that RA-induced *Gm12794c*$^{+}$ cells resemble, at the transcriptome level, preimplantation 2-cell stage embryo.

Gm12794c favors acetylation of H3K27 and enhances 2C-like cells having global DNA hypomethylation

To evaluate the epigenetic landscape of *Gm12794c*-expressing mESC cells in response to RA, we analyzed global levels of H3K27ac and H3K27me3 employing an E14 cell line bearing a transgenic Strawberry reporter under the control of *Gm12794c* promoter that, as the endogenous promoter, is responsive to retinoic acid, and—upon stimulation—allows cells activating *Gm12794c* promoter to express the Strawberry reporter and fluoresce in red.

Confocal immunofluorescence analysis showed that *Gm12794c*-expressing cells (*Gm*$^{+}$) presented higher levels of H3K27ac compared with the *Gm12794c*-negative cells (*Gm*$^{-}$), both in the presence (Fig. 3a, I) and the absence of RA (Supplementary Fig. 6, full bars), while no significant difference was observed when we investigated H3K27me3 levels, despite the trend favoring a retinoic acid-dependent increased tri-methylation (Fig. 3a, II; Supplementary Fig. 6). Both the levels of H3K27ac and H3K27me3 overall increased after RA treatment independently of the expression of *Gm12794c* (Supplementary Fig. 6, empty bars). Altogether, our results showed a link between the presence of *Gm12794c* and overall acetylation levels of H3K27 that resembled the in vivo ZGA and paralleled the global increase of transcriptional activity typical of 2C-like cells. As ground-state cells are also characterized by a global hypomethylated DNA and dispersed pericentromeric chromatin (chromocenters) [1, 38–41], we overexpressed *Gm12794c* in *Rosa26*$^{Tm\_Gm12794c}$ cells by Dox treatment and analyzed DNA methylation and clustering of chromocenters. Global DNA methylation content was evaluated by analyzing intensity levels of 5mC staining [42]. In normal growing conditions, 5mC exclusively co-localized with DAPI-dense foci, indicative of chromocenters, and no diffuse 5mC staining was observed, showing that in standard culture conditions, methylated DNA is mostly located at clustered heterochromatin. On the contrary, cells treated with RA showed a more diffuse 5mC fluorescence signal, meaning that DNA was globally methylated and that this methylation did not induce a comprehensive DNA compaction. Remarkably, when *Gm12794c* was overexpressed, a global DNA hypomethylation was observed, as evaluated by low levels of 5mC staining. Furthermore, we found that overexpression of *Gm12794c* dramatically affects pericentromeric DNA compaction as we observed a dispersion of DAPI-dense chromocenter-associated foci in the majority of the analyzed cells (Fig. 3b, I and II). These results were confirmed when we evaluated DNA methylation content by using the methylation-sensitive restriction enzyme HpaII. CpG methylation was reduced in cells overexpressing *Gm12794c* if compared with *Rosa26*$^{Tm\_Gm12794c}$ cells treated with RA alone and, strikingly, also if it was compared with the level of methylated DNA in mock-treated cells (Fig. 3b, III). Although we cannot address a direct functional role of *Gm12794c* in determining DNA hypomethylation and dispersion of chromocenters, our findings showed that expression levels of *Gm12794c* were directly related to global mESCs DNA hypomethylation and increased H3K27ac levels. Notably, as we observed an increase of acetylation at H3K27 also in *Gm12794c*-expressing cells independently from RA stimuli (Supplementary Fig. 6), we conclude that *Gm12794c* is a key factor for the maintenance of 2C-like cells in mESC colonies.

Gm12794c counteracts RA signaling through the repression of p21/*Cdkn1A* transcription

Morphogenetic activity of RA is achieved throughout the induction of differentiation and apoptosis while halting cell cycle progression [ref. 43 and references therein]. Among key players of cell cycle arrest is the CDK inhibitor p21/*Cdkn1A* whose ectopic expression was recently shown to induce ESCs differentiation [44]. Interestingly, p21/*Cdkn1A* presents RARE elements in its own promoter and consequently is a RA responsive gene [45]. Consistent with these observations, endogenous p21/*Cdkn1A* mRNA and protein levels are induced by RA in both NIH3T3 and *Rosa26*$^{Tm\_Gm12794c}$ mESC cells (Supplementary Fig. 7, I and II, respectively).

Strikingly, overexpression of *Gm12794c* in RA-treated *Rosa26*$^{Tm\_Gm12794c}$ cells resulted in a severe reduction of p21/*Cdkn1A* mRNA and protein levels (Fig. 4a). Consistently, overexpression of FLAG-*Gm12794c* in NIH3T3 inhibited expression of p21-Luc in a dose-dependent manner and determined the downregulation of endogenous p21/*Cdkn1A* mRNA and protein levels (Fig. 4b). In line with these results, in nearly 80% of cells overexpressing FLAG-*Gm12794c*, p21 is barely detectable, suggesting a direct role of FLAG-*Gm12794c* in the downregulation of p21/*Cdkn1A* (Fig. 4c).

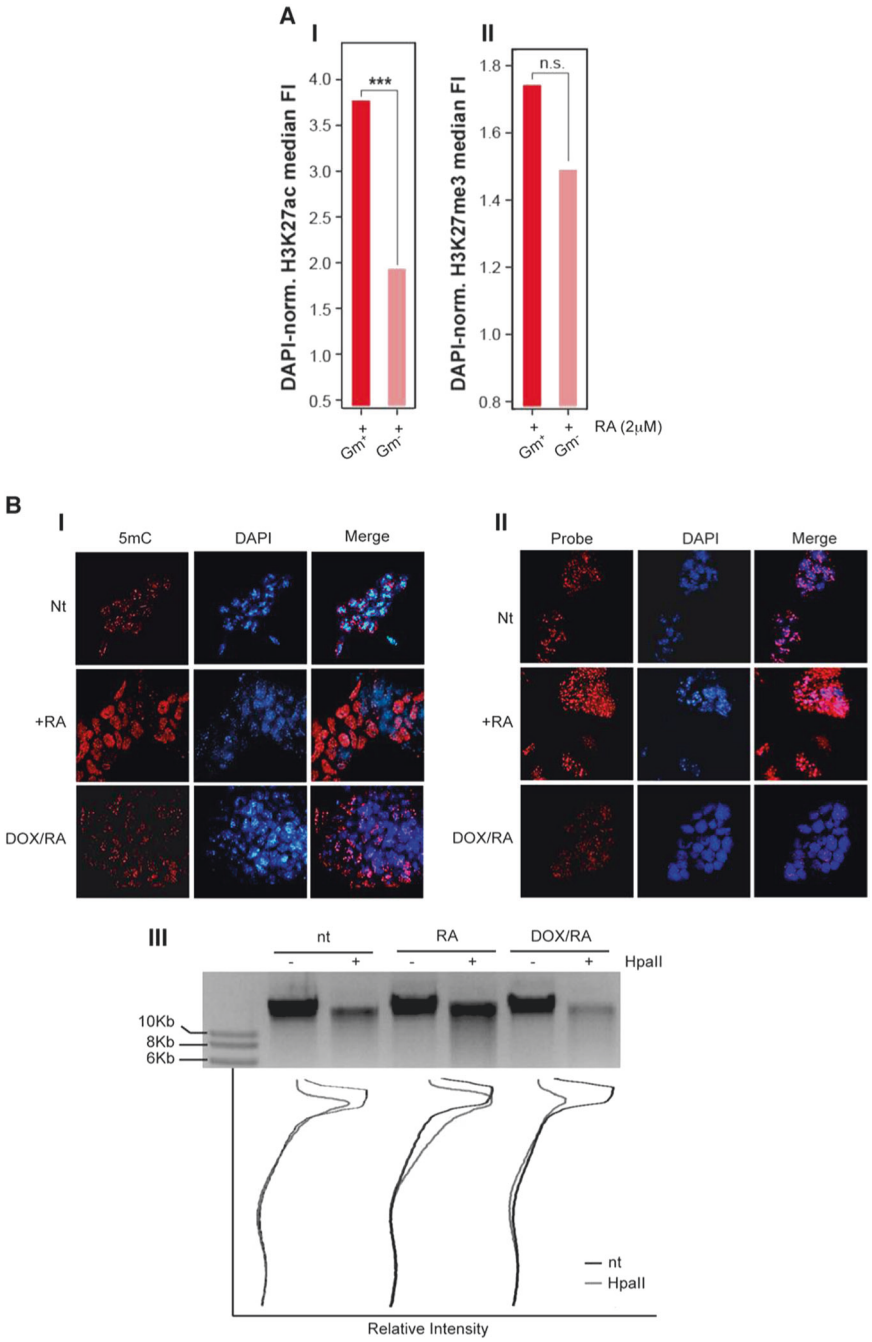
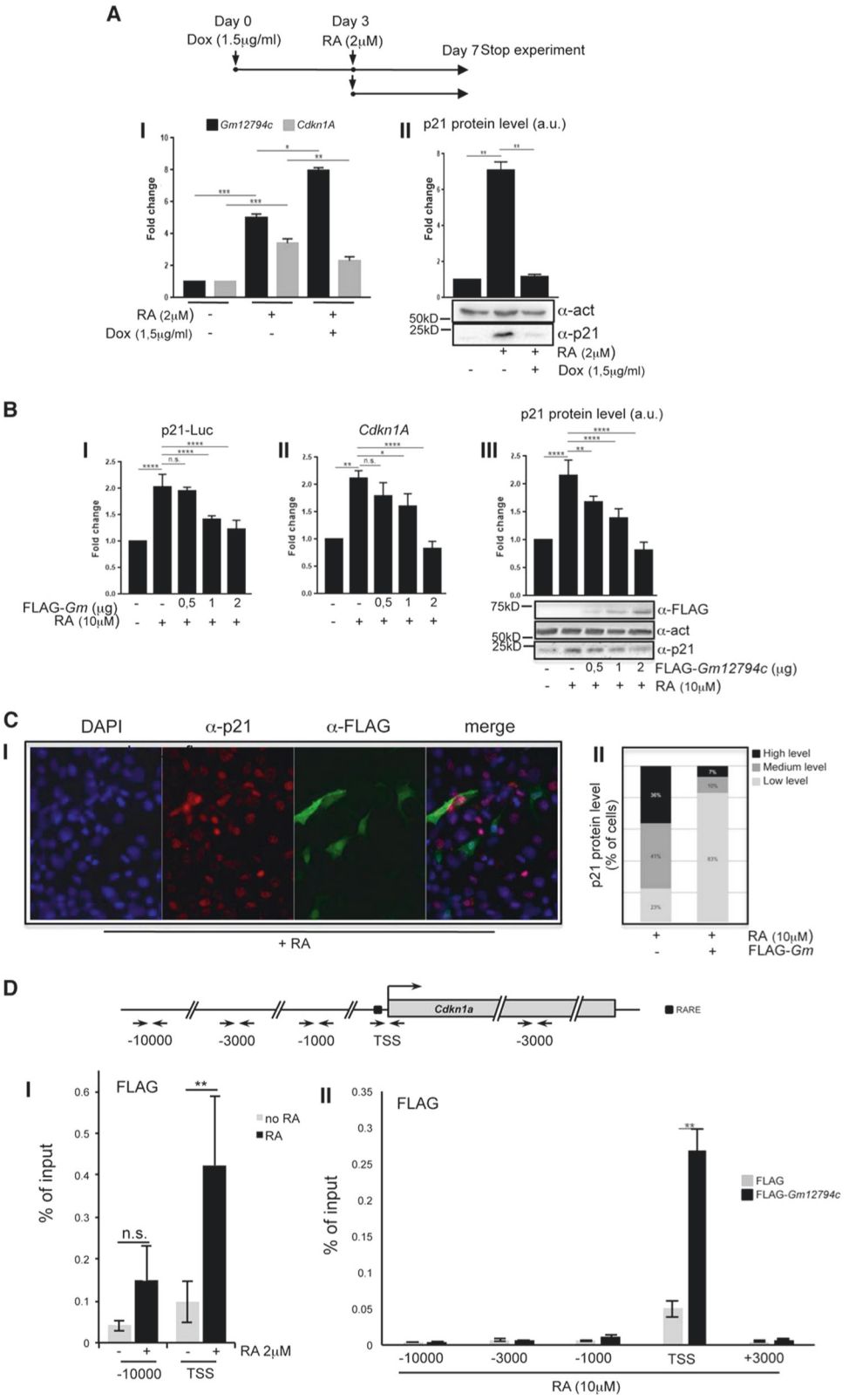


Fig. 3 *Gm12794c* contributes to global DNA hypomethylation and favors acetylation of lysine 27 of histone H3 (H3K27ac). **a** $pr_{Gm12794c}$ Strawberry cells were treated or not with 2 µM retinoic acid (RA) for 96 h before fixation. H3K27ac or H3K27me3 were then probed and revealed by immunofluorescence. Confocal images (individual planes; $z = 700–800$ nm) were analyzed by FIJI image analysis software. The integrated (total) nuclear fluorescence intensity (FI) signal from H3K27ac/me3 was normalized over the DNA (DAPI) or the histone content (H3), to take into account the cell cycle, the chromatin compaction and the substrate histone abundance. **I** Median H3K27ac FI. **II** Median H3K27me3 FI. Statistical significance was assessed by Wilcoxon rank sum test. ***$p < 10–6$; n.s.: not significant. Independent RA treatments: $n = 3$. **b** (**I**) $Rosa26^{Tm\_Gm12794c}$ cells were treated as indicated, and then fixed and probed with anti-5mC antibody. Confocal images have been captured using a confocal ZEISS LSM 700 microscope (individual planes; $z = 700–800$ nm). The number of nuclei analyzed: 100. Representative images are shown. **II** $Rosa26^{Tm\_Gm12794c}$ cells were treated as indicated, and then fixed and probed with a locked nucleic acid (LNA) fluorescent probe complementary to the major DNA satellite transcript (Exiqon) to identify chromocenters [27]. Confocal images have been captured using a confocal ZEISS LSM 700 microscope (individual planes; $z =$ 700–800 nm). The number of nuclei analyzed: 100. Representative images are shown. **III** $Rosa26^{Tm\_Gm12794c}$ cells were treated as indicated, and then CpG methylation levels were assayed by digestion of genomic DNA with the methylation-sensitive HpaII-restriction enzyme. Quantification of DNA signal was measured using FIJI image analysis software



As we observed a reduction of p21/*Cdkn1A* at the transcriptional level both in mESCs and NIH3T3 *Gm12794c*-overexpressing cells, we asked whether *Gm12794c* might contribute to the p21/*Cdkn1A* transcriptional repression by binding to its promoter. Since specific antibodies against *Gm12794c* are not commercially available, we performed

◀ **Fig. 4** *Gm12794c* determines p21/*Cdkn1A* transcriptional repression by directly binding to its promoter region. **a** (Up) Scheme depicting the experimental design. Rosa26^{Tm_Gm12794c} cells were treated with 2 µM retinoic acid (RA) alone for 96 h or grown in the presence of 1.5 µg/ml doxycycline (Dox) for 72 h and then supplemented with 2 µM RA and cultured for 96 more hours. **I** Rosa26^{Tm_Gm12794c} cells have been treated as indicated, and relative amount of *Gm12794c* (black bars) and p21/*Cdkn1A* (grey bars) mRNAs have been assessed by qRT-PCR. **II** Representative image showing p21 protein levels in cells treated as in (**I**) and assayed by WB. Anti-actin has been used as a loading control. WB results have been quantified and graphed as fold change to the mock-treated control cells (a.u.: arbitrary units). **b** (**I**) NIH3T3 were transfected with p21-Luc reporter gene alone or in combination with increasing amounts of FLAG-*Gm12794c* and luciferase activity has been evaluated. Where indicated, transfected cells were treated with 10 µM RA for 24 h. Values in the graph are shown as fold changes normalized to the p21-Luc transfected sample. **II** NIH3T3 cells were transfected as indicated and treated as in (**I**). Relative mRNA levels of p21/*Cdkn1A* were assayed by qRT-PCR and normalized to the mock transfected control cells. **III** Total cell extracts from NIH3T3 in (**I**) were assayed by western blot (WB) to evaluate endogenous p21 protein level and expression of transfected FLAG-*Gm12794c*; anti-actin was used as a loading control. Representative images are shown. WB results have been quantified and graphed as fold changes to the mock-transfected control cells (a.u.: arbitrary units). **c** (**I**) anti-p21 (red) and anti-FLAG (green) co-immunofluorescence has been performed on NIH3T3 cells transfected with FLAG-*Gm12794c* and treated with 10 µM RA for 24 h. Representative images are shown. **II** Fluorescence intensity signal of p21 immunostaining (**I**) has been evaluated in at least 300 FLAG-*Gm12794c*-expressing or not expressing cells treated with 10 µM RA for 24 h. The data are reported as percentage of cells. **d** Schematic representation of the murine p21/*Cdkn1A* genetic locus is shown in the upper panel. Amplicons detected in ChIP analysis are indicated (arrows). Black box indicates the RARE element present upstream the *Cdkn1A* TSS. **I** pr$_{Gm12794c}$*FLAG-Gm12794c* mESCs were treated or not (black bars and grey bars respectively) for 96 h with 2 µM RA and processed for ChIP using anti-FLAG. FLAG-*Gm12794c* enrichment has been evaluated by qRT-PCR using primers located at the RARE element and at distal site (−10000) as a negative control. **II** Control (FLAG) and FLAG-*Gm12794c*-transfected NIH3T3 cells were grown for 24 h with 10 µM RA and processed for ChIP experiments using anti-FLAG. FLAG-*Gm12794c* enrichment has been evaluated by qRT-PCR using primers located at the RARE element (TSS) and at distal sites (−10000; −3500, −1000, and +3000) as a negative control. Experiments have been performed at least four times, and each sample has been prepared in duplicate (luciferase assays, immunofluoresce, ChIP) or in triplicate (qRT-PCR). The data are presented as the mean of independent experiments, and standard deviations are shown. Statistical analysis was performed using Student's *t* test or one-way analysis of variance (ANOVA) followed by Tukey's HSD multiple comparison post hoc tests [61]. $^{**}p \le 0.01$; $^{***}p \le 0.001$; $^{****}p \le 0.0001$; n.s.: not significant.

anti-FLAG ChIP analyses on both E14tg2$^{prGm12794c_FLAG\text{-}Gm12794c}$ and FLAG-*Gm12794c* transiently transfected NIH3T3 cells. FLAG-*Gm12794c* accumulates on the TSS of *Cdkn1A* promoter upon RA treatment in both cell lines, while accumulation in distal upstream regions or in the gene body was not observed (Fig. 4d I, II). These data demonstrated that the presence of *Gm12794c* on the *Cdkn1A* locus is concomitant with p21/*Cdkn1A* transcriptional repression. Although we cannot address a direct role of *Gm12794c* in the transcriptional repression of p21/*Cdkn1A*, our data strongly suggest that its presence on the *Cdkn1A* TSS is instrumental in repressing the developmental gene.

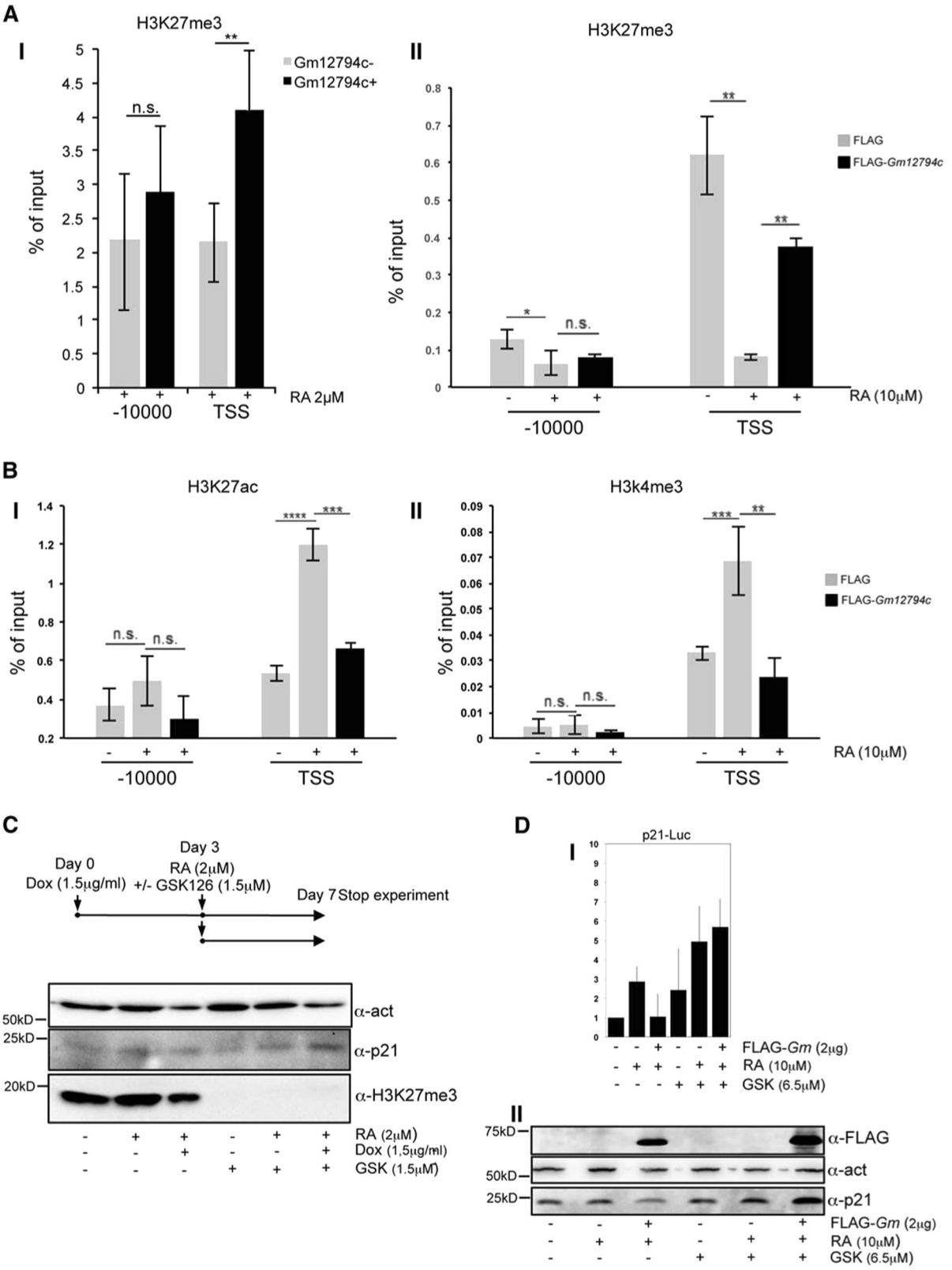
Gm12794c-dependent inhibition of p21/*Cdkn1A* transcription involves polycomb repressive complex 2

Both human and mouse p21/*Cdkn1A* genes have been demonstrated to harbor bivalent promoters that are silent on the majority of developmental genes. They are poised for sudden transcription, depending on specific stimuli, and are characterized by the presence of both activating H3K4me3 and repressive H3K27me3 histone marks. The maintenance of correct balance between activating H3K4me3 and repressive H3K27me3 marks, in mammalian cells mediated by SETD1A, SETD1B, and MLL complexes and by the polycomb repressive complex 2, PRC2, respectively, is of key importance for the transcriptional tuning of developmentally regulated bivalent promoters [46–48].

To address whether *Gm12794c*-dependent p21/*Cdkn1A* repression might be mediated by PRC2, we performed ChIP assays. Furthermore, to assess the bivalent nature of p21/*Cdkn1A* promoter, we also evaluated the presence of H3K4me3 and H3K27Ac, epigenetic modifications associated with transcriptionally active chromatin.

As shown earlier, in absence of RA, p21/*Cdkn1A* is expressed at very low levels in cells, and this is consistent with the presence of H3K27me3 on the TSS, indicating that p21/*Cdkn1A* was in the physiological PRC2-dependent poised status in our experimental conditions (Fig. 5a). We found p21/*Cdkn1A* expression concomitant with the increase of H3K4me3 and H3K27ac marks at its promoter region with the parallel reduction of the H3K27me3 species following RA treatment (Fig. 5b, II). Strikingly, the presence of FLAG-*Gm12794c* in RA-treated NIH3T3 cells led to the persistence of H3K27me3 on *Cdkn1A* TSS, suggesting its role in p21/*Cdkn1A* transcriptional repression. Consistently, levels of H3K4me3/H3K27ac were reduced, thus clearly showing that accumulation of *Gm12794c* at the p21/*Cdkn1A* TSS was concomitant with a repressive chromatin state. In *Gm12794c*$^{+}$ cells purified by magnetic separation of transgenic E14tg2^{prGm12794c_LNGFR} mESC line, upon RA treatment, the presence of *Gm12794c* determined a robust persistence of the repressive H3K27me3 mark on *Cdkn1A* TSS when compared with *Gm12794c*$^{-}$ cells (Fig. 5a, I).

These results, in line with what we observed earlier, demonstrated that *Gm12794c* determines the transcriptional repression of *Cdkn1A* through PRC2 histone methyltransferase activity. Consistently, treatment of *Rosa26*$^{Tm\_Gm12794c}$ mESC cells with the EZH2-specific



inhibitor GSK126 [49] abolished *Gm12794c* repressive effects demonstrating that functional PRC2 complex is required for the *Gm12794c*-mediated p21/*Cdkn1A* transcriptional repression (Fig. 5c). These data were confirmed also in NIH3T3, where we observed that in the presence of GSK126, FLAG-*Gm12794c* had no effect on

◀ **Fig. 5** *Gm12794c* induces p21/*Cdkn1A* transcriptional repression through PRC2-mediated H3K27 trimethylation. **a** (**I**) E14tg2^{pcDNA3_prGm12794c_LNGFR} mESCs were treated with 2 µM RA for 72 h, and *Gm12794c*-expressing cells (black bars, *Gm12794c*$^{+}$) were magnetically separated by *Gm12794c*-not expressing mESCs (grey bars, *Gm12794c*$^{-}$) using the AutoMACS Pro Separator (Miltenyi Biotec). Cells were processed for ChIP using H3K27me3 antibody. **II** Control (FLAG) and FLAG-*Gm12794c*-transfected NIH3T3 were grown for 24 h with 10 µM RA and processed for ChIP experiments using anti-H3K27me3 antibody. **b** Control (FLAG) and FLAG-*Gm12794c*-transfected NIH3T3 were treated with 10 µM RA for 24 h and processed for ChIP experiments using anti-H3K27ac (**I**) and H3K4me3 (**II**). H3K27me3, H3K27ac, and H3K4me3 enrichment have been evaluated by qRT-PCR using the indicated primers. Values reported were calculated as percentage of input. Error bars indicate standard deviations for at least three independent experiments. **c** Schematic representation of the experimental procedure is shown in the upper panel. Rosa26^{Tm_Gm12794c} cells were treated as indicated in the scheme. 1,5µg/ml doxycycline (Dox) was added for 72 hours and then cells were supplemented or not with 2µM RA with or without 1,5µg/ml GSK and cultured for 96 more hours. Where indicated, cells were treated with 1,5µg/ml GSK alone and grown for 96 hours. Whole-cell extracts were prepared and assayed using indicated antibodies. WB anti-H3K27me3 has been used to show EZH2 inhibition by GSK126 treatment. Anti-actin has been shown as a loading control. Representative images are shown (number of biological replicates is at least three). **d** (**I**) NIH3T3 were transfected with p21-Luc reporter gene alone or in combination with increasing amounts of FLAG-*Gm12794c*, and luciferase activity has been evaluated. Where indicated, transfected cells were treated with 10 µM RA for 24 h and/or with (6.5 µM) GSK126 for 40 h. Values in the graph are shown as fold changes normalized to the p21-Luc transfected sample. **II** Total cell extracts from NIH3T3 in (**I**) were assayed by WB to evaluate endogenous p21 protein level and expression of transfected FLAG-*Gm12794c*; anti-actin was used as a loading control. Representative images are shown. Statistical analysis was performed using one-way analysis of variance (ANOVA) followed by Tukey's HSD multiple

p21-Luc and on endogenous p21 protein accumulation (Fig. 5d).

Taken together, our findings showed that the binding of *Gm12794c* to the *Cdkn1A* promoter region in response to RA was concomitant with the repression of p21/*Cdkn1A* transcription, and clearly demonstrated that RA-dependent *Gm12794c*-mediated inhibition of p21/*Cdkn1A* involves PRC2 histone methyltransferase (Fig. 5).

In the light of these results, we speculate that as other LRR-rich proteins, *Gm12794c* might form a ligand-regulated complex with a nuclear receptor that in turn is responsible for *Gm12794c*:PRC2 recruitment on the p21/*Cdkn1A* locus.

LKDLL motif in Gm12794c is required to repress p21/Cdkn1A

Analysis of *Gm12794c* peptide sequence led us to identify four putative NR boxes (LXXLL motifs), as schematically reported in the Fig. 6a, I. As nicely shown in the study of Sheppard et al. [50], depending on residues in the neighborhood, NR boxes show different levels of interaction to nuclear receptors. More precisely, LXXLL motifs presenting a hydrophobic residue at position -1 (where the first leucine of the box is signed as +1 residue) exhibit a strong propensity to interact with NRs [50, 51]. Analyzing the *Gm12794c* peptide sequence, we found that the LKDLL domain located at aa 395–400 presented a valine at −1. Conversely, the other three NR boxes (aa 142–147, aa 342–347, and aa 368–373, respectively) presented small or polar residues that determine weak interactions to NRs or are not involved in the binding of NRs (Fig. 6a, II) [50, 51]. We generated a FLAG-*Gm12794c* mutant, in which the conserved leucines were changed into valines (LKDVV, FLAG-*Gm12794c*LDKVV, hereafter FLAG-*Gm*LDKVV). Expression levels of FLAG-*Gm*LDKVV in NIH3T3 were comparable with those of FLAG-*Gm12794c* (compare Figs. 4b, 6b). Strikingly, we found that mutant FLAG-*Gm*LDKVV failed to repress p21-Luc expression (Fig. 6b, I). Of note, the expression levels of endogenous p21/*Cdkn1A* were also unaffected (Fig. 6b, II and III). As we only observed FLAG-*Gm*LDKVV expression levels in whole-cell extracts, we aimed to determine its cellular sublocalization. We found that mutant FLAG-*Gm*LDKVV showed a cellular localization comparable with that of wt FLAG-*Gm12794c* (compare Figs. 4c, 6c), suggesting that the failure to repress p21/*Cdkn1A* expression was not due to the inability of the protein to enter the nuclei of cells. Furthermore, colocalization of FLAG-*Gm*LKDVV with p21/*Cdkn1A* clearly demonstrates that mutant FLAG-*Gm*LKDVV did not affect p21 protein levels (Fig. 6c). Taken together, these findings demonstrated that the wt LKDLL motif is functional to repress p21/*Cdkn1A*. Moreover, FLAG-*Gm*LDKVV failed to accumulate on the p21/*Cdkn1A* promoter region (compare Fig. 4d, II and Fig. 7a) and in cells overexpressing mutant FLAG-*Gm*LDKVV, we were unable to detect significant levels of H3K27me3 on the TSS of *Cdkn1A* in the presence of RA (Fig. 7b, I). Accordingly, chromatin at the *Cdkn1A* promoter was retained open in FLAG-*Gm*LKDVV-overexpressing cells, as evaluated by the presence of H3K27ac and H3K4me3 (Fig. 7b, II and III). Collectively, our findings demonstrated that through the LKDLL motif at the C terminus of the protein FLAG-*Gm12794c* is recruited on the *Cdkn1A* TSS to mediate PRC2-dependent p21/*Cdkn1A* transcriptional repression.

Discussion

Precisely orchestrated and dynamic processes define early phases of animal development that upon specific differentiation stimuli finally lead to cell fate specification. Nonetheless, if majority of the cells undergo cell priming, some of them must shield their stemness and self-renewal

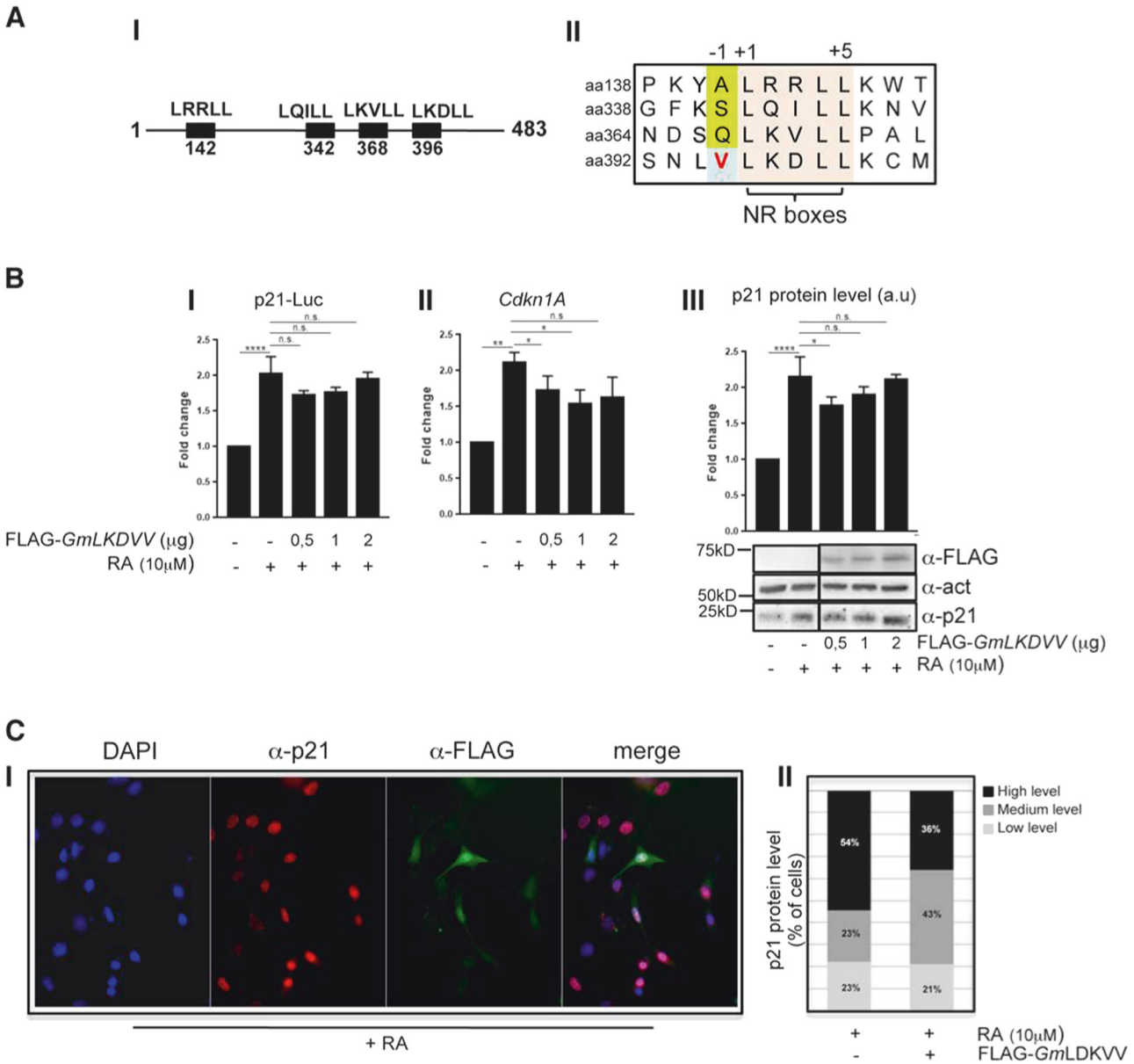


Fig. 6 Mutant *Gm12794c*LKDVV fails to repress p21/*Cdkn1A* transcription. **a** (**I**) Schematic representation of the NR motifs in the *Gm12794c* peptide sequence. **II** Amino acid sequence alignment of the four NR motifs identified in the peptide sequence of *Gm12794c* showing that only the C-terminal LKDLL box located at aa396 has a hydrophobic residue in the −1 position (bold, light blue box). **b** (**I**) NIH3T3 cells were transfected with p21-Luc reporter gene alone or in combination with increasing amounts of FLAG-*Gm12794c*LKDVV and luciferase activity has been evaluated. Where indicated, transfected cells were treated with 10 μM RA for 24 h. Values in the graph are shown as fold changes normalized to the p21-Luc transfected sample. **II** NIH3T3 cells were transfected and treated as in (**I**). Relative mRNA levels of p21/*Cdkn1A* were assayed by qRT-PCR and normalized to the mock-transfected control cells. **III** Total cell extracts from NIH3T3 in (**I**) were assayed by western blot (WB) to evaluate endogenous p21 protein level and expression of transfected FLAG-*Gm12794c*LKDVV; anti-actin was used as a loading control. Representative images are shown. WB results have been quantified and graphed as fold changes to the p21-Luc-transfected control cells (a.u.: arbitrary units). **c** (**I**) anti-p21 (red) and anti-FLAG (green) co-immunofluorescence has been performed on NIH3T3 cells transfected with FLAG-*Gm12794c*LKDVV and treated with 10 μM RA for 24 h. Representative images are shown. **II** Fluorescence intensity signal of p21 immunostaining has been evaluated in FLAG-*Gm12794c*LKDVV-expressing or not expressing cells. The data are reported as percentage of cells. The number of nuclei analyzed: 300 in total (150 + 150). Experiments have been performed at least four times, and each sample has been prepared in duplicate (luciferase assays, immunofluoresce) or in triplicate (qRT-PCR). The data are presented as the mean of independent experiments, and standard deviations are shown. Statistical analysis was performed using one-way analysis of variance (ANOVA) followed by Tukey's HSD multiple comparison post hoc tests. *$p \le 0.05$; **$p \le 0.01$; ****$p \le 0.0001$; n.s.: not significant.

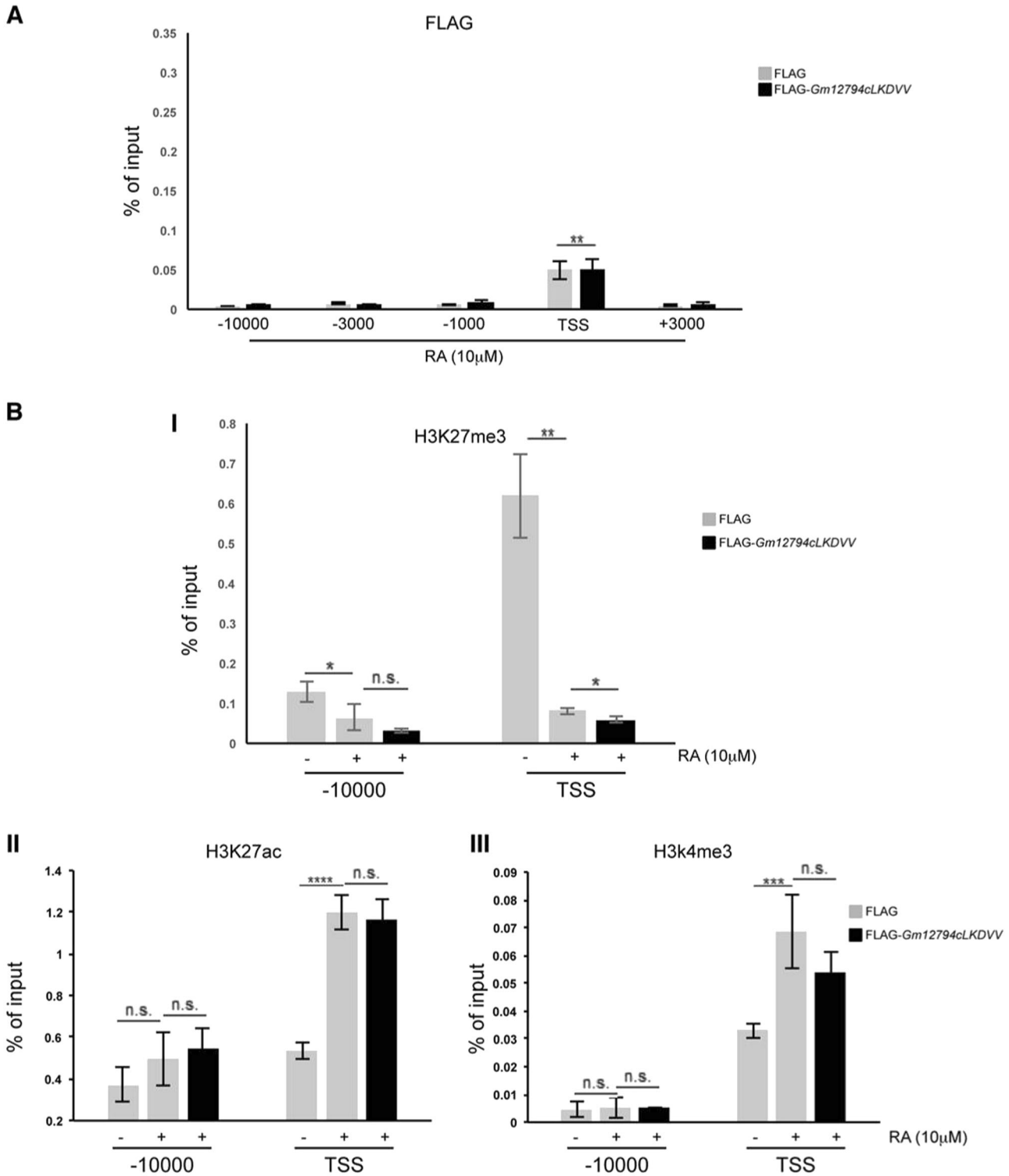


Fig. 7 The LKDLL motif at the C-terminus of *Gm12794c* is required for its accumulation on the p21/*Cdkn1A* promoter region. **a** Control (FLAG) and FLAG-*Gm12794c*LKDVV-transfected NIH3T3 cells were treated as indicated (24 h treatment), and processed for ChIP experiments using anti-FLAG. FLAG-*Gm12794c*LKDVV enrichment has been evaluated by qRT-PCR using primers as in Fig. 5a. **b** Control (FLAG) and FLAG-*Gm12794c*-transfected NIH3T3 were treated as in **a** and processed for ChIP experiments using anti-H3K27me3 (**I**), anti-H3K27ac (**II**), and H3K4me3 (**III**). H3K27me3, H3K27ac, and H3K4me3 enrichment have been evaluated as in Fig. 5b. The values reported were calculated as percentage of input. Error bars indicate standard deviations for at least three independent experiments. Statistical analysis was performed using one-way analysis of variance (ANOVA) followed by Tukey's HSD multiple comparison post hoc tests. $*p \le 0.05$; $**p \le 0.01$; $***p \le 0.001$; $****p \le 0.0001$; n.s.: not significant.

capabilities from extrinsic stimuli to keep the stem cell reservoir. Even though in recent years much has been clarified regarding how and when cells respond to differentiation signals in early stage of development, the question remains open of how in tissues, stem cells preserve their characteristics indefinitely in time.

In this work, we presented data reporting that *Gm12794c*, a RA-induced member of the murine Prame family, counteracts RA-dependent differentiation while specifically contributing to the induction of 2C-like cells resembling the naive pluripotency of preimplantation embryos.

Following RA treatment of mESCs, we detected the induction of 11 *Prame* genes, among which we found *Gm12794c* expression significantly enhanced. Importantly, we observed that high expression levels of *Gm12794c* confer mESCs resistance to the RA-induced differentiation. In particular, *Gm12794c* overexpression is responsible for: (i) high expression of 2C-specific genes, (ii) open-chromatin structure, (iii) low levels of methylated DNA, and (iv) chromocenters dispersion. These findings altogether strongly support the hypothesis that *Gm12794c* might have a role in the mechanisms at grounds of pluripotency maintenance in the ESCs reservoir.

As is known, *2i* medium culture condition affects naive ground-state genes expression [40, 52]. Nonetheless, we showed that RA still significantly induced *Gm12794c* and 2C transcription signature in *2i* (N2B27*2i*/RA). Interestingly, in this condition, *Gm12794c* ectopic overexpression did not affect 2C genes (Supplementary Fig. 9), possibly because N2B27*2i* medium, supplemented with RA, determines 2C-like genes expression plateau.

At the mechanistic level, we reported that *Gm12794c* counteracts antiproliferative effects of RA recruiting PRC2 on the promoter region of *Cdkn1A* to mediate its transcriptional repression and showed that PRC2 enzymatic activity is required for this function. These findings have important significance not only for a better comprehension of ESCs differentiation mechanisms but also in respect to tumors. In cancer cells, it has been shown that PRC2 is responsible for de novo repression of crucial developmental genes that are associated with PRC2 complexes in ESCs [53, 54]. In this respect, our data indicate that *Gm12794c* acts as a "reprogrammer" of PRC2 in the presence of RA, by recruiting the repressive complex on a key developmental gene ensuring its silencing to promote resistance to differentiation, a mechanism that we speculate might be conserved in malignancies that show resistance to RA.

Notably, we found positive feedback between *Gm12794c* expression levels and RARs transcription activation (Supplementary Fig. 8), suggesting that it might have a broader role in the RA-dependent transcriptional program defining early steps of cell fate decision.

It is important to note that we recently showed that RA-dependent induction of highly pluripotent cells is reliant on PI3K signaling [18]. In fact, in mESCs, the primary role of PI3K signaling is the maintenance of undifferentiated cellular conditions that is achieved through the inhibition of the MAPK/ERK and the GSK3 pathways (see ref. [55] and references therein [56, 57]). Interestingly, MAPK/ERK signaling must be inhibited to not induce cellular differentiation while GSK3 inhibition ensures activation of *Nanog* and *Tbx3* allowing induction of pluripotency [58–60].

Hence, our findings support the hypothesis that the *Gm12794c*-dependent inhibition of p21/*Cdkn1A* expression functionally contributes to the PI3K-dependent role of inducing mESCs pluripotency through the blockage of differentiation.

As we highlighted a functional role of the LDKLL NR box of *Gm12794c* in the aberrant recruitment of PRC2 on the *Cdkn1A* TSS, our findings suggest LRR-rich members of the Prame family as suitable therapeutic targets to treat malignancies that show stem cells characteristics and overexpression of PRAME and collectively support the hypothesis that pharmacological inhibition of functional NR boxes of Prame members might represent an interesting strategy to prevent aberrant interactions in malignant cells.

Finally, our data showed that *Gm12794c* was present both in the nuclear and cytoplasmic compartments of cells. Therefore, despite its role as PRC2 "reprogrammer" and co-repressor of p21/*Cdkn1A* transcription, we foresee other *Gm12794c* cellular functions contributing to the resistance to RA-dependent differentiation and to the induction of 2C-like cells.

Acknowledgements We thank Dr. D. Antonini for valuable discussion, Dr. Pietro Zoppoli for supporting statistical analysis, Dr. Floriana Della Ragione for the kind gift of the LNA probe and Prof. Enrico Avvedimento for the kind gift of the anti-5mC antibody. This work was supported by Biogem, Istituto di Biologia e Genetica Molecolare, Via Camporeale, Ariano Irpino (AV), STAR Linea 1, 2014 (University of Naples Federico II). InterOmics 2017 "PROPAGA" (IEOS, CNR) to GF.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Akiyama T, Xin L, Oda M, Sharov AA, Amano M, Piao Y, et al. Transient bursts of Zscan4 expression are accompanied by the rapid derepression of heterochromatin in mouse embryonic stem cells. DNA Res. 2015;22:307–18.

2. Amano T, Hirata T, Falco G, Monti M, Sharova LV, Amano M, et al. Zscan4 restores the developmental potency of embryonic stem cells. Nat Commun. 2013;4:1966.
3. Jiang J, Lv W, Ye X, Wang L, Zhang M, Yang H, et al. Zscan4 promotes genomic stability during reprogramming and dramatically improves the quality of iPS cells as demonstrated by tetraploid complementation. Cell Res. 2013;23:92–106.
4. Hisada K, Sánchez C, Endo TA, Endoh M, Román-Trufero M, Sharif J, et al. RYBP represses endogenous retroviruses and preimplantation- and germ line-specific genes in mouse embryonic stem cells. Mol Cell Biol. 2012;32:1139–49.
5. Falco G, Lee S-L, Stanghellini I, Bassey UC, Hamatani T, Ko MSH. Zscan4: a novel gene expressed exclusively in late 2-cell embryos and embryonic stem cells. Dev Biol. 2007;307:539–50.
6. Cerulo L, Tagliaferri D, Marotta P, Zoppoli P, Russo F, Mazio C, et al. Identification of a novel gene signature of ES cells self-renewal fluctuation through system-wide analysis. PLoS ONE 2014;9:e83235.
7. Dahl JA, Jung I, Aanes H, Greggains GD, Manaf A, Lerdrup M, et al. Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. Nature. 2016;537:548–52.
8. Wu J, Greely HT, Jaenisch R, Nakauchi H, Rossant J, Belmonte JCI. Stem cells and interspecies chimaeras. Nature. 2016;540:51–9.
9. Hamatani T, Falco G, Carter MG, Akutsu H, Stagg CA, Sharov AA, et al. Age-associated alteration of gene expression patterns in mouse oocytes. Hum Mol Genet. 2004;13:2263–78.
10. Peaston AE, Evsikov AV, Graber JH, de Vries WN, Holbrook AE, Solter D, et al. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. Dev Cell. 2004;7:597–606.
11. Evsikov AV, de Vries WN, Peaston AE, Radford EE, Fancher KS, Chen FH, et al. Systems biology of the 2-cell mouse embryo. Cytogenet Genome Res. 2004;105:240–50.
12. Kigami D, Minami N, Takayama H, Imai H. MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. Biol Reprod. 2003;68:651–4.
13. Zhang W, Walker E, Tamplin OJ, Rossant J, Stanford WL, Hughes TR. Zfp206 regulates ES cell gene expression and differentiation. Nucleic Acids Res. 2006;34:4780–90.
14. Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature. 2012;487:57–63.
15. Hendrickson PG, Doráis JA, Grow EJ, Whiddon JL, Lim J-W, Wike CL, et al. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. Nat Genet. 2017;49:925–34.
16. De Iaco A, Planet E, Coluccio A, Verp S, Duc J, Trono D. DUX-family transcription factors regulate zygotic genome activation in placental mammals. Nat Genet. 2017;49:941–5.
17. Hackett JA, Surani MA. Regulatory principles of pluripotency: from the ground state up. Cell Stem Cell. 2014;15:416–30.
18. Tagliaferri D, De Angelis MT, Russo NA, Marotta M, Ceccarelli M, Del Vecchio L, et al. Retinoic Acid Specifically Enhances Embryonic Stem Cell Metastate Marked by Zscan4. PLoS ONE 2016;11:e0147683.
19. Sharova LV, Sharov AA, Piao Y, Stagg CA, Amano T, Qian Y, et al. Emergence of undifferentiated colonies from mouse embryonic stem cells undergoing differentiation by retinoic acid treatment. Vitr Cell Dev Biol Anim. 2016;52:616–24.
20. Heery DM, Kalkhoven E, Hoare S, Parker MG. A signature motif in transcriptional co-activators mediates binding to nuclear receptors. Nature. 1997;387:733–6.
21. McKenna NJ, O'Malley BW. Combinatorial control of gene expression by nuclear receptors and coregulators. Cell. 2002;108:465–74.
22. Epping MT, Wang L, Edel MJ, Carlée L, Hernandez M, Bernards R. The human tumor antigen PRAME is a dominant repressor of retinoic acid receptor signaling. Cell. 2005;122:835–47.
23. Napolitano G, Mazzocco A, Fraldi A, Majello B, Lania L. Functional inactivation of Cdk9 through oligomerization chain reaction. Oncogene. 2003;22:4882–8.
24. Iacovino M, Bosnakovski D, Fey H, Rux D, Bajwa G, Mahen E, et al. Inducible cassette exchange: a rapid and efficient system enabling conditional gene expression in embryonic stem and primary cells. Stem Cells. 2011;29:1580–8.
25. Hamatani T, Carter MG, Sharov AA, Ko MS. Dynamics of global gene expression changes during mouse preimplantation development. Dev Cell. 2004;6:117–31.
26. Zeng F, Baldwin DA, Schultz RM. Transcript profiling during preimplantation mouse development. Dev Biol. 2004;272:483–96. Aug 15PubMed PMID: 15282163
27. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res. 2013;41:D991–5.
28. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47.
29. Taiyun Wei and Viliam Simko R package "corrplot": Visualization of a Correlation Matrix (Version 0.84). 2017. Available from: https://github.com/taiyun/corrplot
30. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2017. https://www.R-project.org/
31. Ambrosio S, Di Palo G, Napolitano G, Amente S, Dellino GI, Faretta M, et al. Cell cycle-dependent resolution of DNA double-strand breaks. Oncotarget. 2016;7:4949–60.
32. Falco G, Stanghellini I, Ko MS. Use of Chuk as an internal standard suitable for quantitative RT-PCR in mouse preimplantation embryos. Reprod Biomed Online. 2006;13:394–403.
33. Al-Anee RS, Sulaiman GM, Al-Sammarrae KW, Napolitano G, Bagnati R, Lania L, et al. Chemical characterization, antioxidant and cytotoxic activities of the methanolic extract of Hymenocrater longiflorus grown in Iraq. Z Nat C. 2015;70:227–35.
34. Napolitano G, Amente S, Lavadera ML, Di Palo G, Ambrosio S, Lania L, et al. Sequence-specific double strand breaks trigger P-TEFb-dependent Rpb1-CTD hyperphosphorylation. Mutat Res. 2013;749:21–7.
35. Napolitano G, Amente S, Castiglia V, Gargano B, Ruda V, Darzacq X, et al. Caffeine prevents transcription inhibition and P-TEFb/7SK dissociation following UV-induced DNA damage. PLoS ONE 2010;5:e11245.
36. Probst AV, Okamoto I, Casanova M, El Marjou F, Le Baccon P, Almouzni G. A strand-specific burst in transcription of pericentric satellites is required for chromocenter formation and early mouse development. Dev Cell. 2010;19:625–38.
37. Ambrosio S, Amente S, Napolitano G, Di Palo G, Lania L, Majello B. MYC impairs resolution of site-specific DNA double-strand breaks repair. Mutat Res. 2015;774:6–13.
38. Leitch HG, McEwen KR, Turp A, Encheva V, Carroll T, Grabole N, et al. Naive pluripotency is associated with global DNA hypomethylation. Nat Struct Mol Biol. 2013;20:311–6.
39. Habibi E, Brinkman AB, Arand J, Kroeze LI, Kerstens HHD, Matarese F, et al. Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. Cell Stem Cell. 2013;13:360–9.
40. Eckersley-Maslin MA, Svensson V, Krueger C, Stubbs TM, Giehr P, Krueger F, et al. MERVL/Zscan4 Network Activation Results in Transient Genome-wide DNA Demethylation of mESCs. Cell Rep. 2016;17:179–92.

41. Ishiuchi T, Enriquez-Gasca R, Mizutani E, Bošković A, Ziegler-Birling C, Rodriguez-Terrones D, et al. Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. Nat Struct Mol Biol. 2015;22:662–71.
42. Graf U, Casanova EA, Wyck S, Dalcher D, Gatti M, Vollenweider E, et al. Pramel7 mediates ground-state pluripotency through proteasomal-epigenetic combined pathways. Nat Cell Biol. 2017; 19:763–73.
43. di Masi A, Leboffe L, De Marinis E, Pagano F, Cicconi L, Rochette-Egly C, et al. Retinoic acid receptors: from molecular mechanisms to cancer therapy. Mol Asp Med. 2015;41:1–115.
44. Itahana Y, Zhang J, Göke J, Vardy LA, Han R, Iwamoto K, et al. Histone modifications and p53 binding poise the p21 promoter for activation in human embryonic stem cells. Sci Rep. 2016;6:28112.
45. Liu M, Iavarone A, Freedman LP. Transcriptional activation of the humanp21(WAF1/CIP1) gene by retinoic acid receptor. Correlation with retinoid induction of U937 cell differentiation. J Biol Chem. 1996;271:31723–8.
46. Pan G, Tian S, Nie J, Yang C, Ruotti V, Wei H, et al. Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. Cell Stem Cell. 2007;1:299–312.
47. Sachs M, Onodera C, Blaschke K, Ebata KT, Song JS, Ramalho-Santos M. Bivalent chromatin marks developmental regulatory genes in the mouse embryonic germline in vivo. Cell Rep. 2013;3:1777–84.
48. Harikumar A, Meshorer E. Chromatin remodeling and bivalent histone modifications in embryonic stem cells. EMBO Rep. 2015;16:1609–19.
49. Gherardi S, Bovolenta M, Passarelli C, Falzarano MS, Pigini P, Scotton C, et al. Transcriptional and epigenetic analyses of the DMD locus reveal novel cis-acting DNA elements that govern muscle dystrophin expression. Biochim Biophys Acta. 2017;1860: 1138–47.
50. Heery DM, Hoare S, Hussain S, Parker MG, Sheppard H. Core LXXLL motif sequences in CREB-binding protein, SRC1, and RIP140 define affinity and selectivity for steroid and retinoid receptors. J Biol Chem. 2001;276:6695–702.
51. Plevin MJ, Mills MM, Ikura M. The LxxLL motif: a multifunctional binding sequence in transcriptional regulation. Trends Biochem Sci. 2005;30:66–9.
52. Guo R, Ye X, Yang J, Zhou Z, Tian C, Wang H, et al. Feeders facilitate telomere maintenance and chromosomal stability of embryonic stem cells. Nat Commun. 2018;9:2620. 5
53. Gal-Yam EN, Egger G, Iniguez L, Holster H, Einarsson S, Zhang X, et al. Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line. Proc Natl Acad Sci USA. 2008;105:12979–84.
54. De Carvalho DD, Mello BP, Pereira WO, Amarante-Mendes GP. PRAME/EZH2-mediated regulation of TRAIL: a new target for cancer therapy. Curr Mol Med. 2013;13:296–304.
55. Yu JSL, Cui W. Proliferation, survival and metabolism: the role of PI3K/AKT/ mTORsignalling in pluripotency and cell fate determination. Development. 2016;143:3050–60.
56. Storm MP, Kumpfmueller B, Thompson B, Kolde R, Vilo J, Hummel O, et al. Characterization of the phosphoinositide 3-kinase-dependent transcriptome in murine embryonic stem cells: identification of novel regulators of pluripotency. Stem Cells. 2009;27:764–75.
57. Singh AM, Bechard M, Smith K, Dalton S. Reconciling the different roles of Gsk3β in "naïve" and "primed" pluripotent stem cells. Cell Cycle. 2012;16:2991–299. 11
58. Paling NRD, Wheadon H, Bone HK, Welham MJ. Regulation of embryonic stem cell self-renewal by phosphoinositide 3-kinase-dependent signaling. J Biol Chem. 2004;279:48063–70.
59. Niwa H, Ogawa K, Shimosato D, Adachi K. A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. Nature. 2009;460:118–22.
60. Wray J, Kalkan T, Gomez-Lopez S, Eckardt D, Cook A, Kemler R, et al. Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. Nat Cell Biol. 2011;13:838–45.
61. Vivo M, Fontana R, Ranieri M, Capasso G, Angrisano T, Pollice A, et al. p14ARF interacts with the focal adhesion kinase and protects cells from anoikis. Oncogene. 2017;36:4913–28.