Nucleic Acids Research, 2025, **53**, gkaf197
https://doi.org/10.1093/nar/gkaf197
Gene regulation, Chromatin and Epigenetics

DGCR8 haploinsufficiency leads to primate-specific RNA dysregulation and pluripotency defects

Ana Colomer-Boronat[1,2,†], Lisanne I. Knol[3,†], Guillermo Peris[2,4], Laura Sanchez[2], Silvia Peluso[5], Pablo Tristan-Ramos[1,2], Ana Gazquez-Gutierrez[1,2], Priscilla Chin[3], Katrina Gordon[3], Guillermo Barturen[2,6], Robert E. Hill[5], Francisco J. Sanchez-Luque[7], Jose Luis Garcia-Perez[2], Alasdair Ivens[3], Sara Macias[3,*], Sara R. Heras[1,2,*]

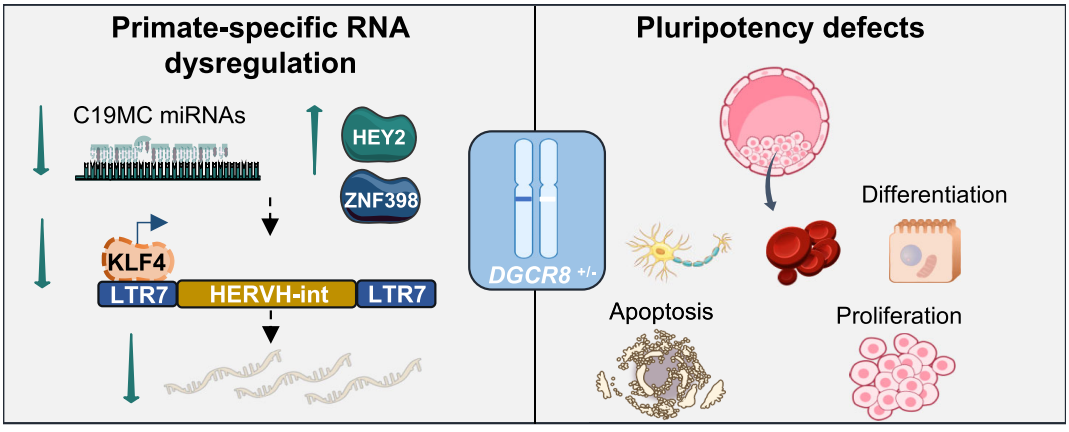
[1]Department of Biochemistry and Molecular Biology II, Faculty of Pharmacy, University of Granada, 18071 Granada, Spain
[2]GENYO – Centre for Genomics and Oncological Research – Pfizer, University of Granada, Andalusian Regional Government, PTS, 18016 Granada, Spain
[3]Institute of Immunology and Infection Research, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3FL, United Kingdom
[4]Department of Computer Languages and Systems, Universitat Jaume I, 12071 Castellon de la Plana, Spain
[5]MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh EH4 2XU, United Kingdom
[6]Department of Genetics, Faculty of Science, University of Granada, 18071 Granada, Spain
[7]Institute of Parasitology and Biomedicine Lopez-Neyra (IPBLN), Spanish National Research Council (CSIC), PTS Granada, 18016 Granada, Spain

[*]To whom correspondence should be addressed. Email: sararheras@ugr.es
Correspondence may also be addressed to Sara Macias. Email: smacias@ed.ac.uk
[†]These authors contributed equally to this work.

Abstract

The 22q11.2 deletion syndrome (22qDS) is a human disorder where the majority of clinical manifestations originate during embryonic development. 22qDS is caused by a microdeletion in one chromosome 22, including *DGCR8*, an essential gene for microRNA (miRNA) production. However, the impact of *DGCR8* hemizygosity on human development is still unclear. In this study, we generated two human pluripotent cell models containing a single functional *DGCR8* allele to elucidate its role in early development. *DGCR8*$^{+/-}$ human embryonic stem cells (hESCs) showed increased apoptosis as well as self-renewal and differentiation defects in both the naïve and primed states. The expression of primate-specific miRNAs was largely affected, due to impaired miRNA processing and chromatin accessibility. *DGCR8*$^{+/-}$ hESCs also displayed a pronounced reduction in human endogenous retrovirus class H (HERVH) expression, a primate-specific retroelement essential for pluripotency maintenance. The reintroduction of miRNAs belonging to the primate-specific C19MC cluster as well as the miR-371-3 cluster rescued the defects of *DGCR8*$^{+/-}$ cells. Mechanistically, downregulation of HERVH by depletion of primate-specific miRNAs was mediated by KLF4. Altogether, we show that *DGCR8* is haploinsufficient in humans and that miRNAs and transposable elements may have co-evolved in primates as part of an essential regulatory network to maintain stem cell identity.

Graphical abstract



Received: June 3, 2024. **Editorial Decision:** February 13, 2025. **Accepted:** March 3, 2025

Introduction

The 22q11.2 deletion syndrome (22qDS) is a human genetic disorder caused by a heterozygous microdeletion at chromosome 22. It is the most common human chromosomal deletion, with an incidence of 1/3000 to 1/6000 in live births [1]. Although the deletion is variable in size (1.5–3 Mb), the largest and most frequent deletion (85% patients) affects around 40 protein-coding genes [1]. The major clinical manifestations include developmental disabilities, congenital heart disease, palatal abnormalities, immune deficiency, and increased risk of autoimmune diseases and psychiatric illnesses, such as autism and schizophrenia [1].

The precise relationship between the deletion of specific genes and the subsequent clinical symptoms remains to be fully elucidated. Among the 40 genes affected by the microdeletion, the *DGCR8* gene has received much attention, due to its essential role in microRNA (miRNA) biogenesis. MiRNAs are small non-coding RNAs that negatively regulate messenger RNA (mRNA) stability and translation by imperfect base-pairing to complementary sequences [2]. Most miRNAs are transcribed as long primary transcripts (pri-miRNAs) that fold into hairpin structures. These are recognized and cleaved in the nucleus by the Microprocessor complex, which is composed of the dsRNA-binding protein DGCR8 and the RNAse III endonuclease Drosha [3–6]. Next, the excised hairpin is exported to the cytoplasm and further processed by the RNAse III endonuclease Dicer. Finally, one of the strands of the mature miRNA duplex is incorporated into the RNA-induced silencing complex (RISC) to guide repression of the target mRNA [7]. In addition to miRNA production, the Microprocessor can directly control the levels of mRNAs by cleaving stem-loop structures that resemble pri-miRNAs, including the *DGCR8* transcript itself and retrotransposon-derived RNAs among others [8–12].

Mouse models of *Dgcr8* heterozygosity have led to contrasting results. While losing one copy of *Dgcr8* in mouse embryonic stem cells (mESCs) was not sufficient to observe significant alterations in the expression of miRNAs by microarrays [13], analyses of miRNA expression by reverse transcription-quantitative polymerase chain reaction (RT-qPCR) or deep sequencing from brain tissue of *Dgcr8* heterozygous mice showed dysregulation of a subgroup of miRNAs [14–18]. These mice also displayed behavioural changes and cognitive defects, which have been attributed to changes in the structure of neuronal dendrites and their synaptic properties [14–16]. These studies suggest that the alteration of the structure and function of neuronal circuits in *Dgcr8* heterozygous mice could provide a genetic explanation for the neuropsychiatric manifestations in 22qDS. Although the complete deficiency of *Dgcr8* is embryonically lethal in mice [13], cell-specific *Dgcr8* ablation has revealed that this gene is also necessary for optimal function of mouse immune cells (including helper T cells, B cells, and NK cells) and thymic architecture as well as reproductive function, female fertility, and spermatogenesis [19–24].

Here, we have developed two different human cell models of *DGCR8* heterozygosity, in the embryonic stem cell line H9 and teratocarcinoma cell line PA-1, to investigate if *DGCR8* is haploinsufficient in humans and to understand the impact of hemizygosity on early development. Our results show that inactivating one *DGCR8* allele results in haploinsufficiency as manifested by dysregulation of miRNA biogenesis but also changes in chromatin accessibility. Despite the functional conservation of *DGCR8* throughout evolution, we found that a high proportion of the affected mature miRNAs are primate specific (PS), mostly belonging to the large miRNA cluster C19MC [25]. As a result, *DGCR8* heterozygous cells display alterations in the gene expression profile associated with pluripotency maintenance and embryonic development, including the human endogenous retrovirus class H (HERVH) family, a crucial transposable element for stem cell identity [26–29]. Interestingly, the alterations caused by *DGCR8*$^{\pm}$ are conserved in both the naïve and primed stages of pluripotency. Consistent with these findings, *DGCR8* heterozygous cells display defects in self-renewal and impaired differentiation into the three primary major germ layers. We demonstrate that the main molecular and cellular defects are restored upon reintroduction of the miR-371-372 cluster and members of the PS C19MC cluster. Mechanistically, we show that HERVH downregulation is mediated by KLF4, as this key pluripotency factor is controlled by important targets of the C19MC cluster, including *HEY2* and the human-specific *ZNF398*. Altogether, these data indicate that *DGCR8* has a critical role in human development and potentially a more relevant role in the aetiology of 22qDS than previously suggested by the mouse models of this disorder.

Materials and methods

Cell lines

All cell lines were grown at 37°C and 5% CO_2. H9 human embryonic stem cells (hESCs) were obtained from WiCell and cultured in mTeSR1 media (STEMCELL Technologies) in plates coated with Matrigel (Corning). Human PA-1 cells were cultured in Minimum Essential Medium (MEM) (Gibco) supplemented with GlutaMAX, 20% heat-inactivated fetal bovine serum (FBS; Gibco), 100 U/ml penicillin–streptomycin (P/S; Invitrogen), and 0.1 mM Non-Essential Amino Acids (Gibco). iROCK (10 μM, Y-27632; Sigma) was added to the medium during the first 24 h after splitting, followed by replacement with fresh media. HEK293T cells were obtained from ATCC and cultured in high-glucose Dulbecco's modified Eagle's medium (DMEM; Gibco) supplemented with GlutaMAX, 10% FBS (HyClone), and 100 U/ml P/S. Karyotyping for both H9 and PA-1 was performed at the Andalusian Biobank (Centre for Biomedical Research, Granada). Short tandem repeat analysis was carried out at the Genomic Unit (GENYO, Granada).

Generation of CRISPR-edited clonal cell lines

Guide RNAs (gRNAs) A (GCACCACTGGACGTTTGCAG) and B (GAGGTAATGGACGTTGGCTC) were designed to target exon 2 of *DGCR8*, after the start codon, using the double nickase design from CRISPR Design Tool (https://chopchop.cbu.uib.no/). gRNAs were cloned into pX461 [pSpCas9n(BB)-2A-GFP; Addgene ID #48140] as in [30]. For CRISPR targeting, H9 hESCs were maintained in E8 media [DMEM/F12, L-ascorbic acid-2-phosphate magnesium (64 mg/l), sodium selenium (14 μg/l), FGF2 (100 μg/l), insulin (19.4 mg/l), sodium bicarbonate (1.74 g/l), NaCl (5 mM), holo-transferrin (10.6 mg/l), TGFβ1 (1.8 μg/l)]. This parental line was used as a control for comparison with the genome-edited cells and was referred to as wild-type (WT) in the experiments. Approximately 2 × 10^6 hESCs were nucleofected

with 2 μg each of pX461-sgRNA(A) and (B) using the V-Kit solution (Amaxa) and the A-23 program and seeded at a density of 2×10^3 in a Matrigel-coated plate, as in [31]. Control transfection with plasmid pMAX-EGFP (Amaxa) revealed that ∼50% of cells were GFP+ by fluorescence microscopy. Five days after nucleofection, cells were dissociated with a recombinant enzyme (TrypLE™) and single-cell clonal cell lines generated by limited dilution in 96-well coated plates. iROCK inhibitor was added to the media during passaging to increase cell survival. After two passages, a fraction of the cells was used for genomic DNA extraction using QuickExtract DNA Solution (Lucigen). Polymerase chain reaction (PCR) from genomic DNA was performed using KAPA2G Fast Hotstart Ready Mix PCR Kit (Kapa Biosystems). PCR products were cloned in pGEM-T Easy Vector (Promega), and at least 10 clones were sent for Sanger sequencing for each cell line. Two *DGCR8* heterozygous clones (HET) with a frameshift mutation in one allele were selected for further studies. These clones were named H9 hESCs HET(1) and HET(2). For CRISPR targeting of PA-1 cells, 1.25 μg of each pX461-sgRNA(A) and (B) were co-transfected using Lipofectamine 2000. GFP^+ cells were sorted 48 h post-transfection using a FACSAria Cell Sorter (BD) and seeded in 96-well plates. iROCK was added during passaging. Genomic DNA sequencing of PA-1 clones was performed as described for H9 hESC clones. As no heterozygote clones were obtained during the first round of targeting, one clone containing frameshift mutations in both *DGCR8* alleles, *DGCR8knockout* (KO) was used to generate $DGCR8^{+/-}$ (HET) cells by repairing one of the mutated alleles. To this end, 245 pmols of CRISPR RNA (crRNA) (AGGTAATGGACGTTGGACGT), complementary to only one of the mutated alleles in the KO cells, and 245 pmols of trans-activating CRISPR RNA (tracrRNA) (IDT) were incubated in 25 μl nuclease-free buffer for 5 min at 95°C and allowed to anneal at room temperature (RT). The resulting gRNA was incubated with 25 μg of Cas9 protein (IDT) for 15–25 min at 37°C prior to transfection. In brief, 1.2×10^6 PA-1 $DGCR8^{-/-}$ (KO) cells were resuspended in 80 μl of buffer T (Neon Transfection System), and the Cas9/gRNA mixture and 300 pmol of the repair template were added to be electroporated using three pulses of 1600 V and 10 ms. Single-cell clones were obtained by limiting dilution in 96-well plates. To test successful gene editing, genomic DNA was extracted by incubating cells in lysis buffer [30 mM Tris–HCl (pH 8.0), 10 mM ethylenediaminetetraacetic acid (EDTA), 0.1% sodium dodecyl sulfate, 0.5% Tween-20, 10 ug/ml Proteinase K] for 15 min at RT. Next, lysate was transferred to 57°C for 10 min, followed by 98°C for 10 min. From this mix, amplification of the sequence of interest was performed by PCR, and products were cloned in pGEM-T Easy Vector. Successful gene editing was confirmed by Sanger sequencing. Two clones were selected for further studies and named PA-1 HET(1) and HET(2). All primers used are listed in Table 1.

Lentiviral transduction of hESCs

The lentiviral particles overexpressing human DGCR8—pLV-EF1α:hDGCR8—were purchased from Vector Builder. As a lentiviral empty control vector, the promoter EF1α and human *DGCR8* sequences were removed using the restriction enzymes FseI and BsBtI (NEB). After ligation, the right se-

Table 1. Oligonucleotides used in this study

H9 cells	
Primer name	**Sequence (5′ to 3′)**
DGCR8 gRNA(A) Fw	CACCGCACCACTGGACGTTTGCAG
DGCR8 gRNA(A) Rv	AAACCTG CAA ACG TCC AGT GGT GC
DGCR8 gRNA(B) Fw	CACCGAGGTAATGGACGTTGGCTC
DGCR8 gRNA(B) Rv	AAAC GAG CCA ACG TCC ATT ACC TC
DGCR8 PCR Fw	ACTCGCTTAGTCGCCAGTCA
DGCR8 PCR Rv	TTACTCCTGCAGCTCTCGGT
DGCR8 Fw	TGCAAAGATGAATCCGTTGA
DGCR8 Rv	AGTAACTTGCTCAAAGTCAAA
NANOG Fw	TGCAGTTCCAGCCAAATTCTC
NANOG Rv	CCTAGTGGTCTGCTGTATTACATTAAGG
SOX2 Fw	TCAGGAGTTGTCAAGGCAGAGAAG
SOX2 Rv	CTCAGTCCTAGTCTTAAAGAGGCAGC
KLF4 Fw	GCTGCCGAGGACCTTCTG
KLF4 Rv	GCGAACGTGGAGAAAGATGG
OCT4 Fw	AGTGAGAGGCAACCTGGAGA
OCT4 Rv	ACACTCGGACCACATCCTTC
HERVH Fw	ACGCTTTACAGCCCTAGACC
HERVH Rv	GTCGGGAGCAGATTGGGTA
LTR7 Fw	GGCCAGTCCTTGCCTTAACT
LTR7 Rv	GCTCAGTGGGGGTGCTTTT
OTX2 Fw	GACCCGGTACCCAGACATC
OTX2 Rv	GCTCTTCGATTCTTAAACCATACC
SOX1 Fw	CACAACTCGGAGATCAGCAA
SOX1 Rv	GGTACTTGTAATCCGGGTGC
PAX6 Fw	CCGGCAGAAGATTGTAGAGC
PAX6 Rv	CGTTGGACACGTTTTGATTG
HAND1 Fw	AACCTCAGCCCTATCTCC
HAND1 Rv	CTTTCATCTTCCTGCGTC
TBXT Fw	GATGATCGTGACCAAGAACG
TBXT Rv	CTTCCAGCGGTGGTTGTC
CD34 Fw	CCGCGCTTTGCTTGCTGAG
CD34 Rv	TCTGGGGTAGCAGTACCGT
FOXA2 Fw	CGCCCTACTCGTACATCTCG
FOXA2 Rv	AGCGTCAGCATCTTGTTGG
GATA6 Fw	AATACTTCCCCCACAACACAA
GATA6 Rv	ACTCTCCCGCACCAGTCAT
HNF3 Fw	GTGGCTCCAGGATGTTAGGA
HNF3 Rv	GCCTGAGTTCATGTTGCTGA
SOX7 Fw	ACGCCGAGCTCAGCAAGAT
SOX7 Rv	TCCACGTACGGCCTCTTCTG
SOX17 Fw	CAGAATCCAGACCTGCACAACGC
SOX17 Rv	CTTCAGCCGCTTCACCTGCTTG
KLF17 Fw	GGGATGGTGCGATAGATTCA
KLF17 Rv	GCCTCACCCTCACCTAACAA
DPPA3 Fw	ATCGGAAGCTTTACTCCGTCGAG
DPPA3 Rv	CCCTTAGGCTCCTTGTTTGTTGG
DPPA5 Fw	ACATCGAGCAGGTGAGCAAGG
DPPA5 Rv	CATGGCTTCGGCAAGTTTGAG
DNMT3L Fw	GGACCCTTCGATCTTGTGTA
DNMT3L Rv	ACCAGATTGTCCACGAACAT
DUSP6 Fw	GCTGTGGCACCGACACAGT
DUSP6 Rv	ACTCGCCGCCCGTATTCT
GAPDH Fw	TGCACCACCAACTGCTTAGC
GAPDH Rv	GGCATGGACTGTGGTCATGAG
Actin Fw	AGAGCTATGAGCTGCCTGACG
Actin Rv	TGTGTTGGCATAGAGGTCTTTACG
Human miR371a-5pAAA	ACUCAAACUGUGGGGGCACUAAA
Human miR373-3pAAA	GAAGUGCUUCGAUUUUGGGGUGUAAA
Human miR372-3pAAA	AAAGUGCUGCGACAUUUGAGCGUAAA
Human miR515-5pAAA	UUCUCCAAAAGAAAGCACUUUCUGAAA
Human miR519c-3pAAA	AAAGUGCAUCUUUUUAGAGGAUAAA
Human miR520c-3pAAA	AAAGUGCUUCCUUUUAGAGGGUAAA
Human miR520d-3pAAA	AAAGUGCUUCUCUUUGGUGGGUAAA
Human miR520g-3pAAA	ACAAAGUGCUUCCCUUUAGAGUGUAAA
Human miR520f-3pAAA	AAGUGCUUCCUUUUAGAGGGUUAAA
Viral titration ΔU3_Fw	GACGGTACAGGCCAGACAA
Viral titration PBS_Rv	TGGTGCAAATGAGTTTTCCA
Human pri-miR-371a Fw	GCCTCTTCTGATGGGTAAG
Human pri-miR-371a Rv	TCTGCTCTCTGGTGAAAG
Human pri-miR-372 Fw	GCCCTAGACCCTGCCTATG
Human pri-miR-372 Rv	CTGATGTCCAAGTGGAAAGTGC
Human pri-miR-373 Fw	GAAAGTCACAGTGATGGCAG
Human pri-miR-373 Rv	GAGTAGGGCATCACGAAC

Table 1. Continued

H9 cells	
Primer name	**Sequence (5′ to 3′)**
Human pri-miR-512 Fw	TGGCACTCAGCCTTGAGGGCACTT
Human pri-miR-512 Rv	TGGCGCAGAACAAGCACCACGG
Human pri-miR-302a Fw	CTTAAACGTGGATGTACTTG
Human pri-miR-302a Rv	CTGCGGTCAATACAATAAAG
Human pri-miR-522 Fw	GCAAGAAGATCTCAGGCTGTGTCCC
Human pri-miR-522 Rv	ACCGCACTCCAGTTTGGGCAGC
ESGR_Fw	TGGGATGGAGCCATAGAAGT
ESGR_Rv	TGGGTCTTTCAAGAAGTTCCTC
LIN00458_Fw	CTGCTCTTTGCTCCGTGAGA
LIN00458_Rv	AGTCCGATTTTCAGTGGGGT
psiCHECK_hsa-miR-515-5p_Fw	TCGAGCAGAAAGTGCTTTCTTTTGGAGAAC
psiCHECK_hsa-miR-515-5p_Rv	GGCCGTTCTCCAAAAGAAAGCACTTTCTGC
psiCHECK_hsa-miR-519c-3p_Fw	TCGAGATCCTCTAAAAAGATGCACTTTC
psiCHECK_hsa-miR-519c-3p_Rv	GGCC GAAAGTGCATCTTTTTAGAGGATC
psiCHECK_hsa-miR-372-3p_Fw	TCGAGACGCTCAAATGTCGCAGCACTTTC
psiCHECK_hsa-miR-372-3p_Rv	GGCCGAAAGTGCTGCG ACATTTGAGCGTC
PA-1 cells	
Primer name	Sequence (5′ to 3′)
Repair template for one of the alleles in *DGCR8*$^{-/-}$ PA-1 cells	TTAGAGAAGGATCCTTTGGAGAGAAGAG AAGCTCCGTAGAAGTTGAAGGGGTCCTC AGCAGGGAGTTCGGACTGTCCATCACCA CCAGAGCCAACGTCCATTACCTCTGCAC CACTGGAC
exon 2 human *DGCR8* Fw	AGGAGAAGCGGTGATGGAG
exon 2 human *DGCR8* Rv	CATCCACTCTGTCTCTCTGAAC
Human pri-miR-302a Fw	CGTGGATGTACTTGCTTTGAAAC
Human pri-miR-302a Rv	GCTGCGGTCAATACAATAAAGT
Human pri-miR-302b Rv	CCTTCAAATGAGGTTAGCGTGT
Human pri-miR-21 Fw	TCTCATGGCAACACCAGTCG
Human pri-miR-21 Rv	AAGTGCCACCAGACAGAAGG
Human pri-miR-92a-2 Fw	AGTATTGCACTTGTCCCGGC
Human pri-miR-92a-2 Rv	TGACTAAATATCAGAACTTACGGCT
Human pri-miR-135b Fw	TCGCTTCCCTATGAGATTCCT
Human pri-miR-135b Rv	TGGGACAGCAATCACATAGGA
Human pri-miR-767 Fw	TGCACCATGGTTGTCTGAG
Human pri-miR-767 Rv	GACAATGAAGGTTCCTGCTCA
Human miR-302a-3p	TAAGTGCTTCCATGTTTTGGTGA
Human miR-302b-3p	TAAGTGCTTCCATGTTTTAGTAG
Human miR-21-5p	TAGCTTATCAGACTGATGTTGA
Human miR-92a-3p	TATTGCACTTGTCCCGGCCTGT
Human miR-135b-5p	TATGGCTTTTCATTCCTATGTGA
Human miR-767-5p	TGCACCATGGTTGTCTGAGCATG

quence was confirmed by Sanger sequencing. Lentiviral control particles were generated as in [32]. Viral titres (transduction units per millilitre) were calculated using quantitative polymerase chain reaction (qPCR). To this end, 1×10^5 HEK293T cells were transduced with different volumes of the viral supernatant (1, 5, and 10 µl). Seventy-two hours post-transduction, genomic DNA was isolated (QiAamp DNA miniKit, Qiagen) and the lentiviral copy number integrated per cell was calculated using a standard curve method. Primers used are listed in Table 1. For transduction, hESCs were dissociated and mixed with lentiviruses at a multiplicity of infection of 5 and seeded on 24-well plates coated with Matrigel in mTeSR1 medium supplemented with iROCK. After 24 h, the medium was replaced. Three days later, 2 µg/ml puromycin selection was initiated, replacing media every 2 days. After 5 days, transduced hESCs were grown under normal conditions.

Immunofluorescence

hESCs were seeded on Matrigel-coated coverslips and fixed with 4% paraformaldehyde (PFA) for 5 min, followed by permeabilization with 0.1% Triton X-100 in phosphate-buffered saline (PBS) for 5 min at RT. Blocking was performed with 10% donkey serum (Merck Life Science) in PBS containing 0.5% Triton X-100, for 1 h at RT. Cells were incubated overnight at 4°C with primary antibodies diluted in PBS containing 1% donkey serum and 0.5% Triton X-100, followed by three washes with 1% donkey serum in PBS. Incubation with secondary antibodies was performed for 30 min at RT, followed by three additional washes with 1% donkey serum in PBS. Cells were counterstained with DAPI (ProLong Gold antifade; Invitrogen), and Zeiss LSM 710 Confocal Microscopy with a Plan-Apochromat 63×/1.40 Oil DIC M27 was used for imaging. ImageJ was used for quantification. Primary antibodies against Nanog (1:1000, 500-P236; Prepotech), Tra-1-60 (1:100, 09-0068; Stemgent), Brachyury (1:1600, #81694; CST), Sox17 (1:200, AF1924; R&D Systems), and Pax6 (1:200, #60433; CST) were used. Secondary antibodies included an anti-mouse IgG Alexa Fluor 488 (1:1000, A21202; Invitrogen), anti-rabbit IgG Alexa Fluor 555 (1:1000, A31572; Invitrogen), and anti-goat IgG Alexa Fluor 555 (1:1000, A21432; Invitrogen).

Western blot

Cells were lysed in RIPA buffer supplemented with 1× EDTA-free Protease Inhibitor cocktail (Roche), 1 mM phenylmethylsulfonyl fluoride (PMSF), and 35 nM β-mercaptoethanol (Sigma). Protein was quantified using the Micro BCA Protein Assay Kit (Thermo Fisher Scientific). Lysates were subjected to sodium dodecyl sulfate–polyacrylamide gel electrophoresis and transferred to polyvinylidene fluoride (PVDF) membranes using the Trans-Blot Turbo Transfer System (Bio-Rad). Membranes were blocked using 5% non-fat dried milk in tris-buffered saline (TBS) (0.1% Tween-20 in 1× TBS) (TBS-T) or PBS-T (0.1% Tween-20 in 1× PBS) and incubated with primary and secondary antibodies. Images were acquired using an ImageQuant LAS4000 or Chemidoc Imaging System (Bio-Rad), and fluorescent signal was quantified using ImageJ software. Antibodies used include anti-DGCR8 (1:1000, ab90579; Abcam), anti-DROSHA (1:1000, NBP1-03349; NovusBio), anti-SOX2 (1:2000, AB5603; Merck), anti-NANOG (1:500, 500-P236; Peprotech), anti-OCT4 (1:300, sc-8628; Santa Cruz Biotechnology), and anti-KLF4 (1:1000, 4038; Cell Signaling Technology). α-Tubulin (1:1000, sc-23948; Santa Cruz Biotechnology) and β-actin (1:15000, A1978; Sigma) antibodies were used as loading controls. As secondary antibodies, anti-rabbit HRP (1:1000, 7074S; Cell Signaling Technology), anti-mouse HRP (1:1000, 7076S; Cell Signaling Technology), and anti-goat HRP (1:10000, 305-035-003; Jackson ImmunoResearch) were used.

Naïve-like hESCs induction

To revert the primed state of HET and WT H9 hESCs to a naïve-like state, RSeT™ Feeder-Free Medium (STEMCELL Technologies) was used. Briefly, hESCs were cultured in a T25 flask to 90% confluency. Next, cells were dissociated and detached into aggregates of ∼100–200 µm in diameter. A dilution (1:20) of the hESCs aggregates was seeded in mTeSR1 medium supplemented with iROCK on Matrigel-coated six-well plates. One day after seeding, medium was replaced with

RSeT™ Feeder-Free Medium and hESCs were grown under hypoxic conditions (37°C, 5% CO_2, and 5% O_2). A full-medium change was done every 2 days. After 5 days, hESCs were reverted into naïve-like hESCs.

Clonal expansion and cell proliferation assays

For clonal expansion assays, 13×10^3 WT, HET(1) and HET(2) H9 hESCs were seeded in 12-well plates coated with Matrigel in mTeSR1 medium. For naïve-like hESCs, 8×10^3 WT, HET(1) and HET(2) were plated in six-well plates coated with Matrigel in RSeT™ Feeder-Free Medium. After 10 days for hESCs and 5 days for naïve-like hESCs, cells were fixed (37% formaldehyde, 50% glutaraldehyde in 10× PBS) for 30 min and stained with 0.5% crystal violet for 40 min at RT. Stained areas and number of colonies were quantified using ImageJ software. The area of naïve-like hESC colonies was measured using cellSens Entry software. For cell proliferation assays, hESCs were seeded at a density of 36 500 cells per well in 12-well plates coated with Matrigel and maintained for 10 days. Cells were counted at days 3, 5, 7, and 10 using a Neubauer chamber. For PA-1 cells, growth rates of WT, HETs, and KO cells were compared by seeding 1×10^5 cells in six-well plates. Cells were harvested and counted using a hemocytometer before reseeding at day 2, 4, and 7.

Alkaline phosphatase staining analysis

H9 hESCs were seeded at a density of 13 000 cells per well in a 12-well plate coated with Matrigel for 6 days. Cells were fixed in 4% PFA for 2 min and stained with Alkaline Phosphatase Detection Kit (Sigma–Aldrich). Colonies were manually counted, distinguishing between differentiated, mixed, and undifferentiated colonies depending on the staining grade and morphology. For single-cell assay, cells were seeded at a density of 0.5 cells/well in a 96-well plate coated with Matrigel and supplemented with cloneR (STEMCELL Technologies) during the first 96 h after plating. After 10 days, colonies were stained with alkaline phosphatase detection kit. Next, colonies were stained with crystal violet to visualise negative alkaline phosphatase cells.

Cell-cycle and apoptosis analyses

hESCs were detached using TrypLE and fixed in 70% cold ethanol, washed three times with ice-cold PBS, and centrifuged at 450 × *g* for 5 min at 4°C. Cell pellets were resuspended in propidium iodide staining buffer [1× PBS, 0.05% NP-40, 3 mM EDTA (pH 8), 1 mg/ml RNaseA, 0.05 mg/ml propidium iodide] for 10 min at RT followed by 20 min on ice. Cells were analysed by flow cytometry using FlowJo software. For apoptosis, cells were labelled with the PE Annexin V Apoptosis kit (BD Biosciences) to distinguish between early (EA) and late apoptosis (LA) by flow cytometry. Results were represented using BD FACSDiva™ software.

H9 hESC differentiation

For embryoid body (EB) differentiation, WT and HET H9 hESCs were cultured to 60% confluency. Next, mTeSR1 medium containing Matrigel (1:6 ratio) was added to increase the thickness of the colonies. At 80% confluency, cells were gently detached and cultured in suspension in ultra-low-attachment plates, allowing for EB formation during 21 days in medium, consisting of DMEM Knockout (Gibco) supplemented with 20% FBS (Hyclone), 0.1 mM Non-Essential Amino Acids (Gibco), 1 mM L-glutamine, and 0.1 mM β-mercaptoethanol. Medium was replaced every 2 days. For guided differentiations to ectoderm, mesoderm, and endoderm lineages STEMdiff Trilineage differentiation kit was used (STEMCELL Technologies). Briefly, H9 hESCs were seeded in mTeSR1 medium supplemented with iROCK on Matrigel-coated coverslips in 24-well plates for immunofluorescence assay. In brief, 400 000 cells per well were used for ectoderm and endoderm differentiation, and 100 000 for mesoderm differentiation. One day after seeding, medium was replaced with lineage-specific medium for 5 days (mesoderm and endoderm), or 7 days (ectoderm). Medium was replaced every day. Cells were next fixed and processed for immunofluorescence analyses as described above. For RNA extraction, cells were seeded in six-well plates at the same confluence and the same protocol was followed.

RT-qPCR

For RT-qPCR, total RNA was extracted from cells using Trizol, followed by RQ1 DNAse treatment and phenol/chloroform purification. Next, 1 μg of total RNA was further treated with DNase I (Invitrogen), and complementary DNA (cDNA) was synthesized using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) and used for qPCR (GoTaq qPCR Mix; Promega). Alternatively, cDNA was synthesized using Transcriptor Universal cDNA Master (Roche), and qPCR was carried out with LightCycler 480 SYBR Green I Master Mix (Roche). *GAPDH* or *ACTB* was used as normalizer. Gene expression levels were quantified using the second derivative method. Primers used are listed in Table 1.

miRNA mimics transfection

H9 WT and HETs (7×10^5) hESCs were seeded in six-well plates. After 24 h, cells were transfected with 60 nM of control mimic (4464058; ThermoFisher), 30 nM of each miR-372-3p and 373-3p mimics (MC10165 and MC11024), or 15 nM of each miR-520g-3p, miR-520d-3p, miR-519c-3p, and miR-515-5p mimics (MC10365, MC12807, MH10575, and MC10387; ThermoFisher) using Lipofectamine 2000 (Life Technologies). At 48 h post-transfection, total RNA or protein was extracted for downstream analysis. Alternatively, for clonal expansion assays, cells were counted and seeded at a density of 13 000 cells per well in a 12-well plate coated with Matrigel 48 h post-transfection. Cells were fixed and stained after 6 days.

siRNAs depletion

Knockdown of *DGCR8*, *KLF4*, *HEY2*, and *ZNF398* in hESCs was achieved with two rounds of small interfering RNA (siRNAs) transfection using Dharmafect 4 solution (Dharmacon) according to the manufacturer's instructions. Briefly, cells (7×10^5) were seeded in six-well plates and, after 24 h, were transfected with 25 nM of each siRNA or non-targeting siRNA control. The transfection medium was replaced after 24 h, and cells were grown for another 24 h before a second transfection with the same amount of siRNA. For miRNA transfection assays, hESCs were co-transfected with miRNA mimics during the second round of siRNA transfection. Twenty-four hours after second transfection, cells were collected for analyses or re-seeded at a density of 13 000 cells

per well in 12-well plates coated with Matrigel during 6 days for clonal expansion assays. SiRNAs against *DGCR8*, *KLF4*, *HEY2*, and *ZNF398* and siRNA control were purchased from Dharmacon (L-015713–00-0005, L-005089–00-0005, L-008223–00-0005, L-013163–00-0005, and D-001810–02-05).

Generation of psiCHECK variants and luciferase assay

The fully complementary sequences of hsa-miR-515-5p, hsa-miR-519c-3p, and hsa-miR-372-3p were ordered as oligos flanked by XhoI and NotI sites. First, 10 μM of each forward and reverse primer were annealed and phosphorylated with T4 Polynucleotide Kinase (NEB) using the following program: 30 min at 37°C, 5 min at 95°C, and ramp down to 25°C at −5°C/min. Hybridized primers were cloned in psiCHECK2 (Promega), fused to the luciferase reporter gene, using the XhoI and NotI sites, generating psiCHECK2-515-5p/519c-3p/372-3p. All plasmids were confirmed by Sanger sequencing. For luciferase assays, (1.6×10^5) WT and HET were seeded in 24-well plates coated with Matrigel. After 24 h, cells were transfected with psiCHECK2-515-5p, psiCHECK2-519c-3p, or psiCHECK2-372-3p, using Lipofectamine 2000 (Invitrogen). Firefly and Renilla luciferase quantification was performed 48 h post-transfection using the Dual-Luciferase Reporter Assay System (Promega) in a GloMax Luminometer (Promega), following the manufacturer's protocol.

Chromatin RNA sequencing and Microprocessor processing index

PA-1 cells were fractionated similar to [33, 34]. In brief, 8×10^6 cells were lysed in mild buffer [10 mM Tris (pH 7.4), 150 mM NaCl, 0.075% NP-40] and the nuclei and cytoplasm were separated by a sucrose gradient. The nucleic fraction was subsequently separated into nucleoplasmic and chromatin-associated fractions as described before, and the chromatin-associated fraction was sonicated on a Bioruptor (five times 20 s on/off intervals) prior to DNase treatment using RQ1 DNase (Promega) and RNA extraction using Trizol. Four biological replicate samples for each of the three PA-1 cell lines (WT, HET, and KO) were prepared and sent to BGI Genomics for library preparation and high-throughput sequencing after ribosomal RNA (rRNA) depletion. Obtained reads were aligned to the human genome (GRCh38.p13) using HISAT2 (v2.1.0) with the options –no-discordant –no-mixed –no-unal [35]. Human pre-miRNA locations were determined by aligning human precursor sequences, obtained using the mature miRNA and hairpin sequences from miRBase (v22.1) against the same genome using Bowtie2 with options –very-sensitive –no-unal [36]. The genomic locations of the pre-miRNAs plus 100 nt on each side were determined using bedtools getfasta -s. The read depth for each nucleotide in the alignment of chromatin-associated reads on the appropriate strand for each pre-miRNA was extracted using R. A 10-nt gap between each pre-miRNA and its flanking regions was created by excluding the outer 5 nt from the pre-miRNA and flanking regions, creating leeway for potential alternative Drosha cleaving. For each region, the average read depth of all included nucleotides was used for Microprocessor processing index (MPI) calculation. MPI was defined as the negative $\log_2$-transformed ratio between the mean read depth (RD) of the pre-miRNA region (hairpin) to the mean read depth of the regions flanking the pre-miRNA. In this manner, a high MPI and >0 indicates efficient processing, and an MPI close to 0 or negative, indicates inefficient or absence of processing.

$$MPI_{\text{sample}} = -\log_2 \frac{\overline{\text{pre-miRNA } RD + 1}}{\overline{(\text{flank } RD) + 1}},$$

$$MPI_{\text{cell line}} = \overline{MPI_{\text{samples}}}, \text{ and}$$

$$\Delta MPI = \overline{MPI_{\text{mutant}}} - \overline{MPI_{\text{mutant}}}.$$

To filter out Microprocessor-independent pri-miRNAs, reduce noise, and exclude artefacts, pri-miRNAs were only included if (i) they were no mirtrons or tailed mirtrons; (ii) they produced miRNAs that were detected and included in statistical analysis for small RNA sequencing (RNA-seq); (iii) on average, the flank depth in WT samples is ≥2; and (iv) the ratio between the two flanks did not differ more than four-fold. Differences in processing for individual pri-miRNA between cell lines (ΔMPI) were computed by subtracting MPI values in *DGCR8*(WT) versus (HET) or (KO) cells [$\Delta MPI = MPI$(HET/KO) – MPI(WT)]. A negative ΔMPI indicates that the pri-miRNA is less processed in HET/KO compared with WT cells.

For motif enrichment of PS miRNAs, motifs were extracted from mirGeneDB 3.0 [37].

In vitro miRNA processing assays

Templates for RNA synthesis and radiolabelling of pri-miRNA substrates were obtained by PCR of human genomic DNA with the following primers: pri-let-7f-2 (Fw: ACTCTCCTTCCCTTTCTCCC and Rv: TGAACAAGACACATGACCTCA) to clone in pGEMt- Easy. Pri-miR-23b was amplified from genomic DNA with the following primers: (Fw-containing T7 promoter: taatacgactcactatagggGTGAAGGCGGCAGTGTCGCCGG and Rv: TGGTGGCTTCTTTTGTTTCCA). *In* vitro transcription reactions were carried out with T7-polymerase, and *in vitro* processing reactions with extracts from PA-1 WT, HET, and KO were performed as previously described in [33].

MiRNA quantification and small RNA high-throughput sequencing

MiRNA quantification in H9 hESCs was performed as described [38]. For PA-1 cells, 200 ng of total RNA was used for the retrotranscription reaction using the miRCURY LNA miRNA PCR system (Qiagen). After dilution of the resulting cDNA, qPCR was performed using the miRCURY LNA SYBR Green PCR Kit (Qiagen) with corresponding miRNA primers on a LightCycler 480 Instrument (Roche). Both H9 and PA-1 miRNA quantification data were normalized to the DGCR8-independent miRNA, *hsa-miR-320a*. For miRNA-specific primers, see Table 1.

For small RNA high-throughput sequencing, RNA from three different biological replicates for WT, HET(1), and HET(2) hESCs was extracted using mirVana microRNA isolation kit (ThermoFisher). Small RNA libraries were generated using NEXTFlex Small RNA Library Prep Kit v3 (Catalogue #NOVA-5132-06) and sequenced on the NextSeq 500 system (Illumina, CA, USA) by the Genomic Unit at GENYO. For PA-1 cells, RNA from four different biological replicates from WT and HET cells was extracted using the miRNeasy

Mini Kit (Qiagen) followed by on-column DNAse digestion. Small RNA-seq libraries using unique molecular identifiers were sequenced using DNA-nanoball sequencing (DNB-seq) [39] by BGI. For each sample, identical small RNA-seq reads were collapsed and only reads that were present more than once were included in further analysis. MiRNAs were identified and counted using miRDeep2, using the quantification function with human hairpin and mature miRNA sequence files from miRBase (v22.1) and allowing no mismatches [40–42]. DESeq2 was used for statistical analysis, comparing WT versus HET in H9 and PA-1 separately, and using apeglm as a log fold-change shrinkage model [43, 44].

In silico miRNA target prediction and pathway analysis

Differentially expressed miRNAs ($P_{adj} \leq .05$) that are common to both H9 hESCs HET(1) and HET(2) (85 miRNAs) and differentially expressed miRNAs in PA-1 HET (155 miRNAs) underwent functional enrichment analysis with DIANA-mirPATH v3 software [45] (for PA-1 HET, only the top 100 miRNAs with the highest absolute $\log_2$FC were used). The MicroT-CDS prediction algorithm was used to identify putative mRNA targets of miRNAs and associated significantly enriched (FDR ≤ 0.05) KEGG pathways [46] identified. Dot plots were generated using ggplot2 (v3.3.5) and DOSE v3.14.0 [47] R packages.

Total RNA high-throughput sequencing and Gene Ontology analyses

Total RNA from three biological replicates of H9 [WT, HET(1), and HET(2)] and four biological replicates of PA-1 (WT and HET) cells was extracted using the miRNeasy Mini Kit (Qiagen) followed by on-column DNAse digestion. Purified RNA was rRNA-depleted prior to sequencing (DNB-seq) by BGI. For PA-1 RNA-seq analyses, paired-end reads were aligned to the human genome (GRCh38.p13) using HISAT2 with options –no-discordant –no-mixed –no-unal [35]. Transcript counts for each sample were created with featureCounts (Rsubread v2.2.6), using reverse counts and excluding reads that aligned to the genome multiple times [48]. For H9 RNA-seq analyses, qualities of individual H9 WT and HET sequences were evaluated using FastQC v0.11.5 software (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) [49]. Paired-end reads were aligned to GRCh38.p13 human genome assembly with STAR v2.7.6a [50] and quantified with featureCounts v2.0.1 [48] using NCBI Annotation Release 109. For both H9 and PA-1, differential expression analysis and count normalization were performed with the R package DESeq2 v1.28 [44]. After differential analysis, DESeq2's apeglm lfcShrink was applied to shrink $\log_2$ fold changes [43, 44].

Functional enrichment analysis was carried out for differentially expressed genes ($P_{adj} \leq .05$) using the enrichGO function in R package clusterProfiler v3.16.1 [51] and using the Gene Ontology (GO) biological processes gene sets.

Analysis of PS C19MC cluster

A list of PS miRNAs was obtained from MirGeneDB 2.1 [25]. R package miRBaseConverter v1.14.0 [52] was used to adapt this list to miRBase v22.1 used for differential expression analysis. All miRNAs with $P_{adj} \leq .05$ were considered as dysregulated. Common PS dysregulated miRNAs in H9 HET (34 miRNAs) and PA-1 HET cells (47 miRNAs) were analysed using DIANA-mirPATH v3 software. Statistical significance for the enrichment of PS dysregulated miRNAs was calculated using Fisher's exact test (phyper function of R package stats, v4.2.1).

Transposable element expression analyses

The Software for Quantifying Interspersed Repeat Elements (SQuIRE) v0.9.9.92 pipeline [53] was used to measure transposable elements (TEs)' expression changes using default parameters. SQuIRE downloads TE annotations from RepeatMasker and uses an expectation-maximization algorithm to assign multimapping reads. Next, it performs TE differential expression using DESeq2, either by grouped TE subfamilies or by analysing individual loci. Expression of individual TEs was represented using ggplot2 (v3.3.5) and gghalves (v0.1.1) R packages (https://cran.rproject.org/web/packages/gghalves/index.html) [54]. hESCs-specific chimeric transcripts and long non-coding RNAs (lncRNAs) derived from HERVH were extracted from [26] and all gene symbols were updated using the R package HGNChelper v0.8.1 [55]. This gene list was compared with differentially expressed genes ($P_{adj} \leq .05$) from hESCs HET(1) and HET(2). Heatmap was generated using the R package ComplexHeatmap v2.4.3 [56] applied to the *Z*-score for each gene and volcano graphs using ggplot2 (v3.3.5). Normalized bigwig files were generated using the Draw tool from SQuIRE and visualized using IGV.

Assay for transposase-accessible chromatin with sequencing

Assays for transposase-accessible chromatin coupled to high-throughput sequencing (ATAC-seq) libraries were prepared as previously described [57]. Around 50 000 cells were used for each biological replicate ($n = 3$). Cells were lysed in 50 µl cold lysis buffer [10 mM Tris–HCl (pH 7.4), 10 mM NaCl, 3 mM $MgCl_2$, 0.1% IGEPAL CA-360] and pelleted at 500 × *g* for 10 min at 4°C. Pellets were resuspended in 50 µl transposition reaction mix as follows: 2× Tagment DNA buffer (Illumina 15027866), 20× Tagment DNA enzyme (Illumina, 15027865), and then incubated at 37°C for 30 min. Transposed samples were purified using the MinElute PCR purification kit (Qiagen, 28204), and eluted in 10 µl. Transposed DNA samples were amplified by PCR by setting up a 50 µl reaction as follows: 10 µl transposed DNA, 9.7 µl ddH_2O, 2.5 µl 25 µM customized Nextera PCR primer FW, 2.5 µl 25 µM customized Nextera PCR primer RV, 0.3 µl 100× SYBR Green I (Invitrogen, S-7563), and 25 µl 2× NEBNext high-fidelity PCR master mix (NEB, M0541). To reduce GC/size bias and overamplification of libraries, the reaction was monitored by qPCR to stop amplification prior to saturation. A 15-µl qPCR side-reaction was set up as follows: 5 µl of five cycles of PCR-amplified DNA, 4.44 µl ddH_2O, 0.25 µl 25 µM customized Nextera PCR primer FW, 0.25 µl 25 µM customized Nextera PCR primer RV, 0.06 µl 100× SYBR Green I, and 5 µl 2× NEBNext high-fidelity PCR master mix. The additional number of cycles required for each sample was determined by plotting a linear run versus the cycle number. The number of cycles that correspond to one-fourth maximum fluorescent intensity was calculated for each sample. The remaining 45 µl five cycles PCR-amplified DNA was run as before (omitting 72°C initial step and modifying the number of cycles to the calculated amount). Amplified DNA samples were subjected to double size selection to remove DNA fragments <150

and >1000 bp that would hinder sequencing reactions. Samples were diluted up to 90 µl with ddH_2O and purified using the SPRIselect beads (Beckman Coulter, B23317). For removal of large DNA fragments, a ratio of 0.55 DNA/beads slurry was used, and for removal of small DNA fragments, a ratio of 1.8 was used. Purified DNA samples were quantified using the Qubit dsDNA high sensitivity assay (ThermoFisher, Q32854) with the Qubit 4 fluorometer. The quality of DNA samples was analysed using the high-sensitivity DNA kit (Agilent 5067-4626) in Bioanalyzer. Samples with the nucleosomal fragment distribution profile expected for ATAC-seq libraries (mono-, di-, and trisomal fragments) were sent for sequencing.

Samples were sequenced on an Illumina HiSeq 4000 platform to obtain 50-bp paired-end reads at the Wellcome Trust Clinical Research Facility (University of Edinburgh). Reads were trimmed using cutadapt v3.5 paired-end trimming and aligned to the hg38 human genome using bowtie2 v2.4.4 [36] paired-end alignment with options –very-sensitive –no-mixed –no-discordant -X 2000. Unmapped reads and those mapping to the mitochondrial genome were removed and duplicate reads were filtered out using Picard 2.27.5 MarkDuplicates (http://broadinstitute.github.io/picard/). Reads were shifted by +4 bp for those mapping to the positive strand and −5 bp for those mapping to the negative strand using alignmentSieve tool from deepTools package. Broad peaks were called using MACS2 2.2.7.1 callpeak [58] with options –g hs –f BED –keep-dup all –*q* 0.01 –nomodel –shift –75 –extsize 150 –broad. Peaks overlapping blacklisted regions were removed using bedtools. A union peak set across all samples was obtained following the iterative overlap peak merging procedure described in [59] (code provided in https://github.com/corceslab/ATAC_IterativeOverlapPeakMerging). A count matrix over the union peak set was computed using featureCounts 2.0.1 [48] and differently expressed peaks were obtained using DESeq2 1.28 R package [44] and selecting dysregulated peaks using $P_{adj} \leq .05$ and $abs(\log_2 FC) \geq 1$. Peaks were annotated using annotatePeaks.pl from Homer v4.11 package [60]. Functional analysis of differential ATAC-seq regions was carried out with rGREAT package [61]. Motif analysis was performed with findMotifsGenome.pl tool from Homer v4.11 package using default parameters and random background selection. Metagene plots were generated using computeMatrix and plotProfile tools from deepTools. Genes contained in an interval of ±10 kb from differentially expressed PA-1 HET ATAC-seq peaks were obtained using annotatePeakInBatch function in ChIPpeakAnno R package [62].

Results

Characterization of *DGCR8* heterozygosity in human pluripotent cellular models

22qDS is caused by a microdeletion in one chromosome 22, resulting in the hemizygosity of around 40 protein-coding genes [1]. It is still unclear if the disease originates from the haploinsufficiency of a small subset of these genes or from the absence of the entire region. To investigate this, we assessed if the genes affected by the microdeletion were predicted to be haploinsufficient by comparing their natural variation in the human population using the Genome Aggregation Database (gnomAD) [63]. Essential genes are predicted to have a very low frequency of loss-of-function (LoF) mutations in the general population, as these may be incompatible with life. When the frequency of observed LoF mutations is lower than expected [observed/expected (obs/exp) ≤ 0.089], the gene is considered haploinsufficient [63]. Obs/exp LoF ratios were plotted for each of the genes affected by the most common microdeletion in 22qDS, and only five genes were predicted to be haploinsufficient, including *DGCR8* (Fig. 1A). Similar conclusions were previously reported by [64]. To validate this prediction, we generated H9 hESCs and human embryonic carcinoma cells (PA-1 hECCs), where a single copy of the *DGCR8* gene was inactivated using the CRISPR/Cas9 nickase system. PA-1 cells are a diploid human embryonic teratocarcinoma cell line, which has retained limited pluripotent capacity [65, 66]. After targeting, two different *DGCR8* heterozygote clones [HET(1) and HET(2)] for each cell line were selected for further studies (Supplementary Fig. S1A–D). Inactivation of a single *DGCR8* allele resulted in reduced DGCR8 protein expression in both H9 and PA-1 cells, but also of DROSHA, as it requires DGCR8 interaction for stabilization [8] (Fig. 1B and Supplementary Fig. S1E–G). Despite the reduction in DGCR8 expression, H9 HET hESCs did not display obvious changes in colony morphology (Fig. 1C) or in the expression of the typical pluripotency markers, NANOG and TRA-1–60 (Fig. 1D). Consistently, no differences were observed in the expression of the other pluripotency markers, OCT4 and SOX2 (Fig. 1E). Only KLF4 was less abundant in HET H9-HESCs, at both protein and RNA levels (Fig. 1E and Supplementary Fig. S1G–H). The proportion of alkaline phosphatase-expressing colonies was also similar between WT and HET H9-hESCs cells (Supplementary Fig. S2A–B). In contrast, *DGCR8* HET hESCs displayed a large reduction in their clonal expansion ability when plated at low cell density or as single cells (Fig. 1F and Supplementary Fig. S2B and C) suggesting poor maintenance of self-renewal capacity. Consistent with this finding, both H9 and PA-1 HET showed a decreased doubling time, confirming some proliferation defects (Supplementary Fig. S2D). Defective proliferation can result from a defect in cell cycle progression and/or increased apoptosis. *DGCR8* HET hESCs displayed delayed cell cycle progression, with a significant accumulation in G0/G1 (Supplementary Fig. S2E and F), in addition to a significant increase in early (7-AAD negative and PE Annexin V positive) and late (7-AAD positive and PE Annexin V positive) apoptosis (Fig. 1G and Supplementary Fig. S2G). Importantly, these cellular phenotypes, including defective proliferation, clonal expansion ability, increased apoptosis, and KLF4 misexpression, were reverted when DGCR8 expression was rescued in HET cells by lentiviral transduction (Fig. 1F and G and Supplementary Figs S1G and H and S2C, D, and G). All these together suggest that inactivation of one *DGCR8* allele in human pluripotent cells results in defective self-renewal capacity, which is characterized by increased apoptosis and cell cycle and proliferation defects.

DGCR8 heterozygous hESCs display differentiation defects

To evaluate if *DGCR8* heterozygosity results in defects during human embryonic development, WT and HET *DGCR8* hESCs were differentiated *in vitro* using spontaneous EB formation. Remarkably, EBs formed by the HET clones were of smaller size in comparison with WT cells, indicating differentiation and proliferation defects (Fig. 2A). To study the defects in differentiation, pluripotency and differentiation

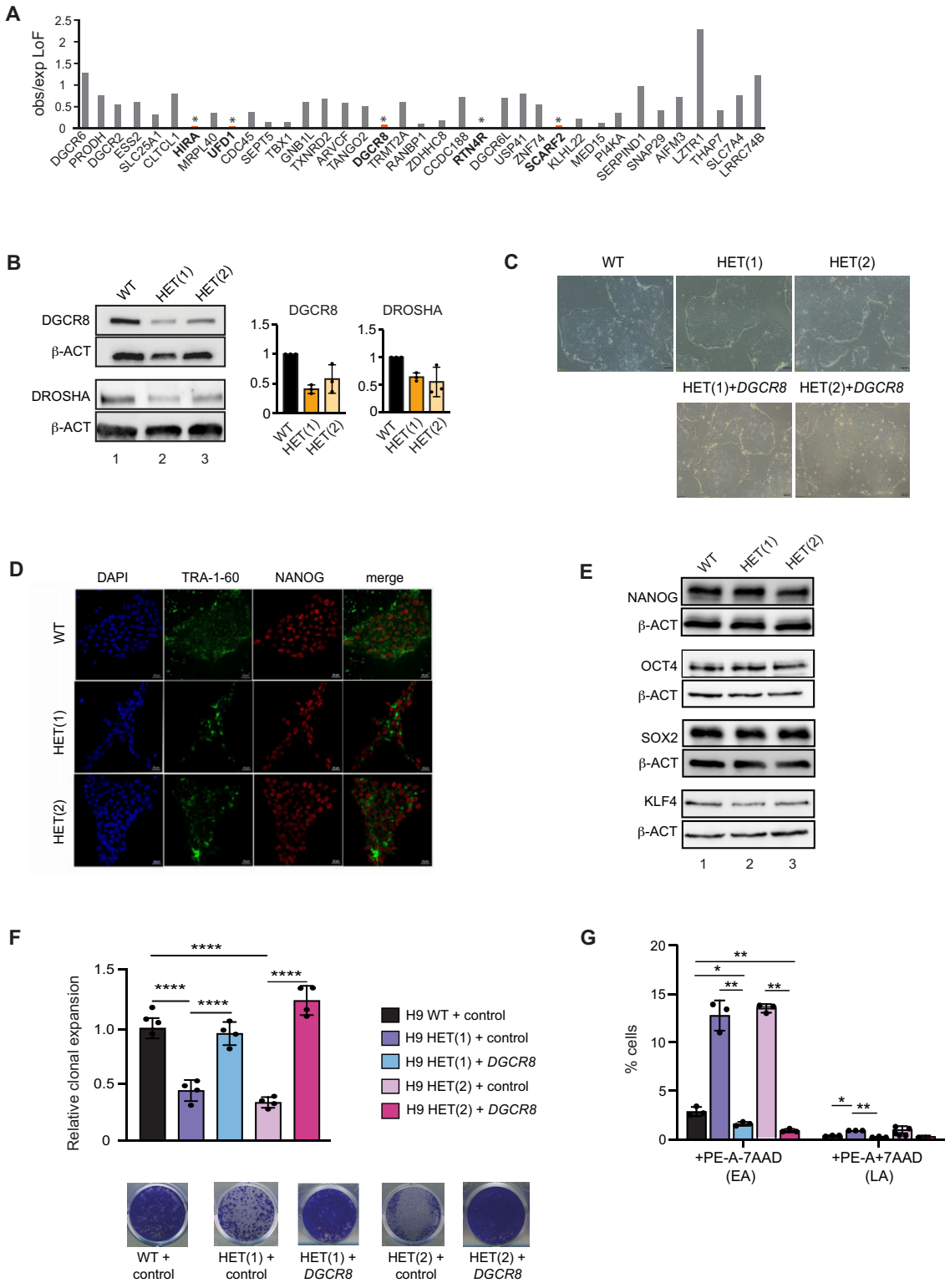


Figure 1. Generation and characterization of DGCR8 heterozygote cell lines. (**A**) Genes commonly deleted in 22qDS were searched for LoF mutations in the human population using the gnomAD. The obs/exp LoF for each gene was plotted. (*) obs/exp LoF ≤ 0.089, predicted as haploinsufficient. Only genes with at least 10 predicted LoF mutations were considered haploinsufficient (in bold). (**B**) *DGCR8* and DROSHA western blot analyses of parental cell line, WT, and two different HET *DGCR8* H9 hESC clones. Actin serves a loading control. Quantification of *DGCR8* and DROSHA protein levels in HET H9 hESCs. Data represent the average of three independent experiments ± standard deviation (SD). (**C**) Colony morphology in normal culturing conditions for WT, HET hESCs, and HET hESCs with rescued *DGCR8* expression by lentiviral transduction (scale bar = 100 μm). (**D**) Immunofluorescence images for pluripotency markers Tra-1–60 and NANOG (scale bar = 20 μm). (**E**) NANOG, OCT4, SOX2, and KLF4 western blot analyses of both WT and two different HET *DGCR8* hESC lines. Actin serves as a loading control. (**F**) Relative clonal expansion capacity, expressed as the fraction of stained area of HET hESCs lines versus WT transduced with a control-lentiviral vector or with lentivirus expressing human *DGCR8*. Data represent the average ± SD. of four biological replicates. (****) $P \leq .0001$, by one-way ANOVA followed by Tukey's multiple comparison test. (**G**) Percentage of cells in EA and LA. Data are the average of ($n = 3$) ± SD. (*) $P \leq .05$, (**) $P \leq .01$, by two-way ANOVA followed by Tukey's multiple comparison test.

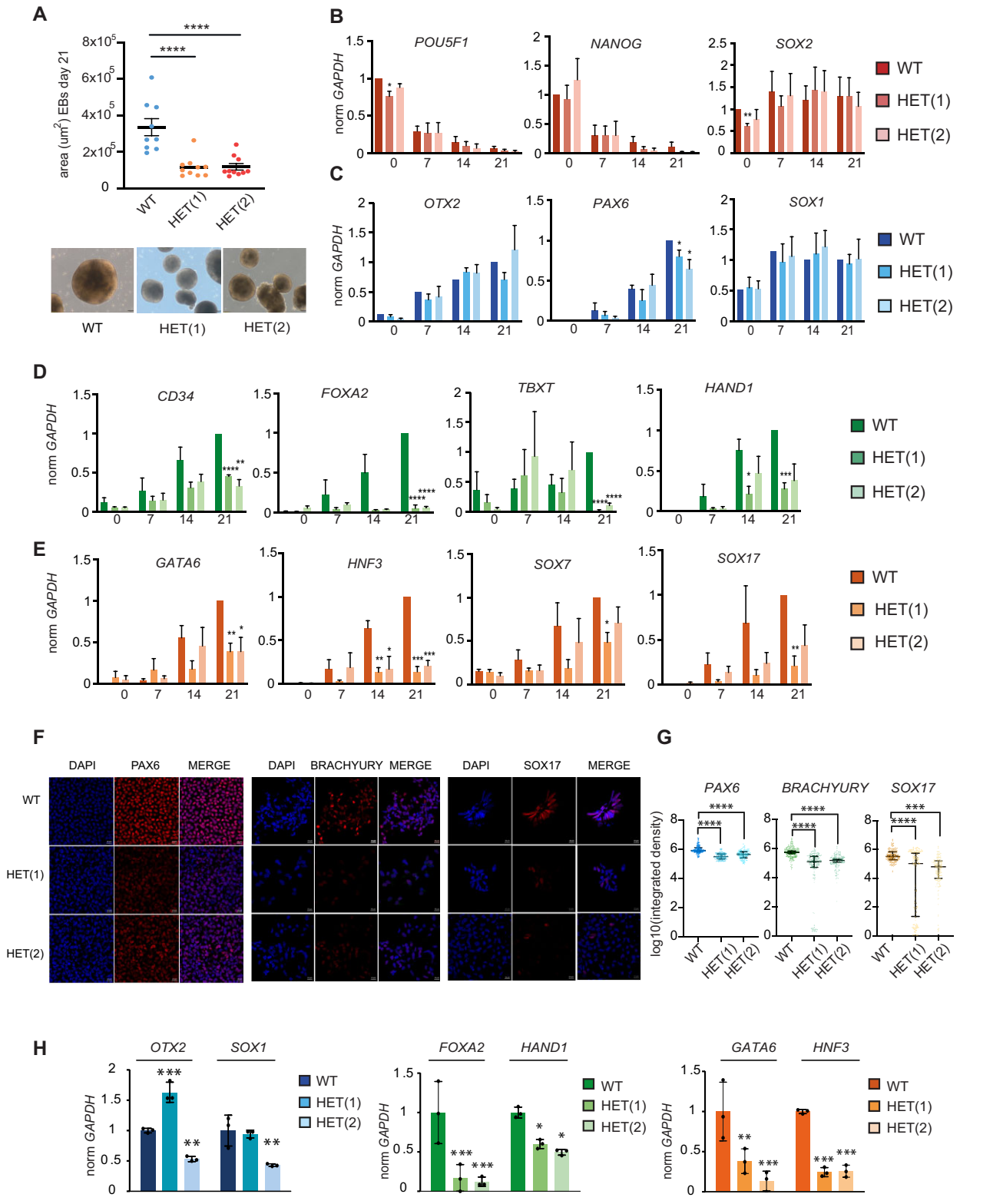


Figure 2. *DGCR8* heterozygote hESCs display differentiation defects. (**A**) Quantification of the area of embryoid bodies at day 21. Data are the average ± standard error of the mean (SEM); (****) $P \leq .0001$ by generalized linear models with negative binomial distribution, adjusted by Bonferroni *post-hoc* test (top panel). Representative images of EBs after differentiation at day 21 (bottom panel) (scale bar = 100 μm) (**B**) RT-qPCR analyses of pluripotency, (**C**) ectodermal, (**D**) mesodermal, and (**E**) endodermal markers after EB differentiation of *DGCR8* HET H9 hESCs compared with the parental cell line (WT). Data are the average ($n = 3$) ± SEM. (*) $P \leq 0.5$, (**) $P \leq .01$, (***) $P \leq .001$, (****), $P \leq .0001$, by two-tailed Student's *t*-test. (**F**) Representative immunofluorescence images from WT and HET cells after differentiation to ectoderm (PAX6), mesoderm (BRACHYURY), and endoderm (SOX17). Scale bar = 10 μm (**G**) Data represent the $\log_{10}$ integrated density for staining in WT and HET differentiated cells ($n = 150$). (***) $P \leq .001$, (****) $P \leq .0001$ by generalized linear models with negative binomial distribution, adjusted by Bonferroni post hoc test. (**H**) RT-qPCR analyses of ectodermal (OTX2, SOX1), mesodermal (FOXA2, HAND1), and endodermal (GATA6, HNF3) markers after guided differentiation. Data are the average ($n = 3$) ± SD. (*) $P \leq 0.5$, (**) $P \leq .01$, (***) $P \leq .001$, by two-way ANOVA followed by Sidak's multiple comparison test.

markers' expression from ectoderm, mesoderm, and endoderm, was compared by RT-qPCR at day 0, 7, 14, and 21 of differentiation. WT and HET EB differentiation resulted in a similar repression of the pluripotency markers, *NANOG* and *POU5F1*. The expression of *SOX2* was also similar between WT and HET EBs (Fig. 2B). Tested ectodermal markers (*OTX2*, *PAX6*, and *SOX1*) were also similarly increased during differentiation in both WT and HET hESCs. Only *PAX6* displayed a subtle reduction at day 21 of differentiation (Fig. 2C). In contrast, the expression of most of the tested mesodermal (*CD34*, *FOXA2*, *TBXT*, and *HAND1*) and endodermal (*GATA6*, *HNF3*, *SOX7*, and *SOX17*) markers was reduced upon differentiation of both clones of HET hESCs, in comparison with WT (Fig. 2D and E). These results suggest that *DGCR8* heterozygosity in pluripotent human cells resulted in differentiation defects most markedly for mesodermal and endodermal lineages.

To support these findings, directed differentiation protocols into ectoderm, mesoderm, and endoderm were performed, followed by immunofluorescence of well-stablished markers for these embryonic layers. This revealed that the average expression of the mesodermal marker BRACHYURY (encoded by the *TBXT*) and the endodermal marker SOX17 was significantly decreased, and nearly absent in a fraction of HET cells. A more homogeneous subtler reduction of the ectodermal marker PAX6 was observed (Fig. 2F and G). The consistent defects in mesoderm and endoderm differentiation of HET cells were confirmed using RT-qPCR against lineage-specific genes (Fig. 2H). All these results suggest that *DGCR8* heterozygosity leads to decreased pluripotency, affecting the establishment of the three major embryonic lineages, with a more pronounced defect in the mesodermal and endodermal germ layers.

DGCR8 heterozygous hESCs maintain cellular defects in a naïve-like state

Our findings demonstrate that the loss of a functional copy of *DGCR8* results in defects in the biology of hESCs. Conversely, ablation of a single *Dgcr8* allele in mESCs did not lead to significant or clear phenotypes [13], raising the possibility that the defects associated with human *DGCR8* heterozygosity could be species-specific. Alternatively, these findings could be attributed to the different pluripotency cellular states of hESCs, considered to be in a primed state compared with mESCs, which represent a naïve state [67]. To rule out this possibility, WT and HET hESCs were induced into a naïve-like state. As a result of this transition, colonies acquired the typical domed morphology of naïve hESCs (Fig. 3A), preserving the reduced protein levels of both DGCR8 and DROSHA (Fig. 3B). To confirm successful transition, upregulation of the naïve pluripotency markers *KLF17*, *DNMT3L*, *DPPA3*, and *DPPA5*, and silencing of the primed pluripotency marker, *DUSP6* was confirmed by RT-qPCR [68–70] (Fig. 3C). Despite the major differences in cellular phenotypes and gene expression profiles of naïve versus primed pluripotent states, the main cellular defects, including increased apoptosis and decreased colony formation capacity, were retained during the naïve stage (Fig. 3D–F). All these together suggest that the defects deriving from *DGCR8* heterozygosity may be species-specific rather than cell state-specific. These results prompted us to further characterise the impact of *DGCR8* heterozygosity at the molecular level.

Expression of PS miRNAs is altered in *DGCR8* heterozygous cells

Considering the defects in proliferation and differentiation of HET cells and the reduced levels of DGCR8 and DROSHA, we next investigated if these phenotypes were linked to abnormal expression of miRNAs. For this purpose, we performed miRNA expression analyses using small RNA-seq of two clones of *DGCR8* HET cells in H9 hESCs and one clone of PA-1 HET cells versus their WT counterparts. *DGCR8* HET hESCs displayed a remarkable reduction in mature miRNAs in a primed stage, with a similar proportion of miRNAs only showing a modest upregulation (Fig. 4A; for complete list of significant differentially expressed miRNAs see Supplementary Table S1). We also noted that a significant proportion of the common differentially expressed miRNAs in both hESC HET clones were PS (~30%) (Fig. 4B and Supplementary Fig. S3A). Dysregulated PS miRNAs mostly belonged to the big miRNA cluster C19MC [25] (Fig. 4C). The expression of the miR-371-3 cluster, homologous to the miR-291-295 cluster in mouse, was also markedly reduced (Fig. 4C). Interestingly, most members of both clusters share the seed sequence 'AAGUGC', which have been previously involved in self-renewal, proliferation, and apoptosis of hESCs [71–73]. The decreased expression of miRNAs belonging to these clusters was validated by RT-qPCR in both naïve and primed H9 hESC clones and was rescued upon reintroduction of DGCR8 (Fig. 4D and E). We next analysed the enrichment of Microprocessor-dependent motifs in the PS miRNAs that were sensitive to DGCR8 loss [74]. Our analyses revealed that these precursor miRNAs were depleted of the basal UG motif, which is usually found at position −14 upstream of the 5p miRNA (Supplementary Fig. S3B). The UG motif has been proposed to be recognized by Drosha [75]. We hypothesise that these pri-miRNAs may be less efficiently recognized by the Microprocessor due to the lack of the UG motif, rendering them, as a result, more susceptible to the reduction of DGCR8 and Drosha levels upon *DGCR8* haploinsufficiency. To experimentally test this prediction, we measured the C19MC unprocessed pri-miRNA levels. Although we observed accumulation of the unprocessed miR-522 precursor of C19MC, this effect was not consistent within other regions of the cluster (Supplementary Fig. S3C). Interestingly, the levels of precursor pri-miR-371-373 were highly reduced upon DGCR8 loss, suggesting that DGCR8 can regulate miRNA expression at multiple levels, both by driving efficient processing and by regulating their transcriptional control (Supplementary Fig. S3C).

To determine if alterations in miRNA levels could result in defective post-transcriptional gene silencing, we performed luciferase gene reporter assays. To this end, fully complementary binding sites for the miR-515-5p, miR-519-3p, and miR-372-3p, members of the C19MC and the miR-371-3 cluster, respectively, were cloned after the luciferase open reading frame. Introducing these binding sites led to a reduction in luciferase expression in WT cells, while the opposite was observed in HET cells (Supplementary Fig. S3D), suggesting that defective miRNA expression results in defective post-transcriptional gene silencing. To further investigate the relevance of the dysregulated PS miRNA expression, we performed pathway enrichment analyses with predicted mRNA targets of the PS miRNAs. Obtained pathways included 'signalling pathways regulating pluripotency of stem cells' and pathways involved in pluripotency maintenance and

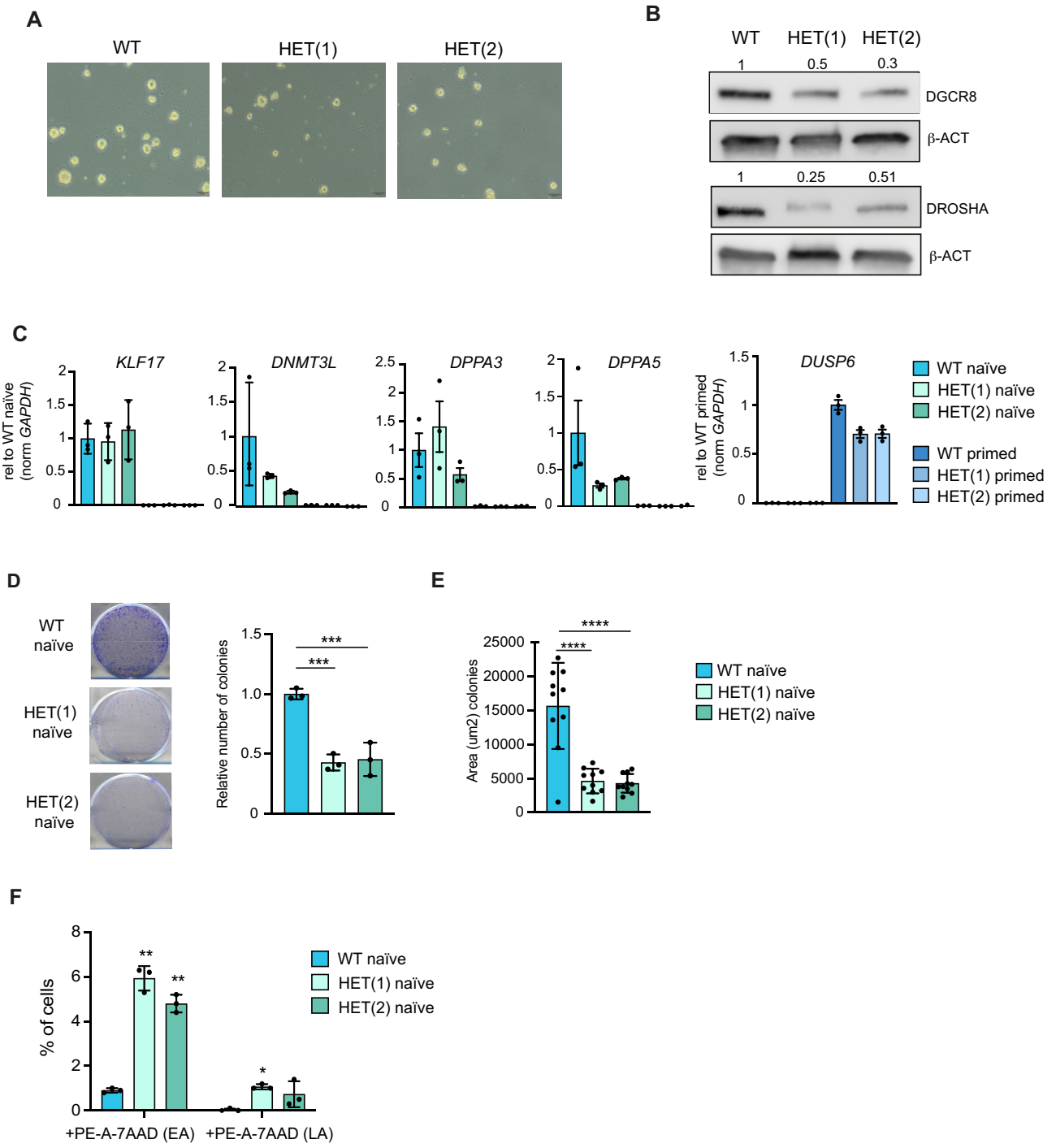


Figure 3. Defects in *DGCR8* heterozygous cells are conserved in a naïve pluripotent state. (**A**) Colony morphology for WT and HET hESCs in naïve culturing conditions (scale bar = 100 μm). (**B**) *DGCR8* and DROSHA western blot analyses of WT and HET *DGCR8* hESC lines. Actin serves as a loading control. (**C**) RT-qPCR analyses of naïve (*KLF17*, *DNMT3L*, *DPP3A*, and *DPP5A*) and primed (*DUSP6*) markers in both *DGCR8* HET and WT hESCs. Data are normalized to *GAPDH* relative to WT naïve or WT primed, respectively, and represent the average of three biological replicates ± SD. (**D**) Relative clonal expansion capacity, expressed as the number of colonies of HET hESCs lines in comparison to WT. Data represent the average ($n = 3$) ± SD. (***) $P \leq .001$, by one-way ANOVA followed by Dunnett's multiple comparison test. (**E**) Quantification of the area in naïve colonies at day 5. Data are the average ± ($n = 10$). (****) $P \leq .0001$, by one-way ANOVA followed by Dunnett's multiple comparison test. Colonies are visualized by crystal violet staining. (**F**) Percentage of cells in EA and LA. Data represent the average ± SD of three biological replicates. (*) $P \leq .05$, (**) $P \leq .01$, by two-way ANOVA followed by Tukey's multiple comparison test.

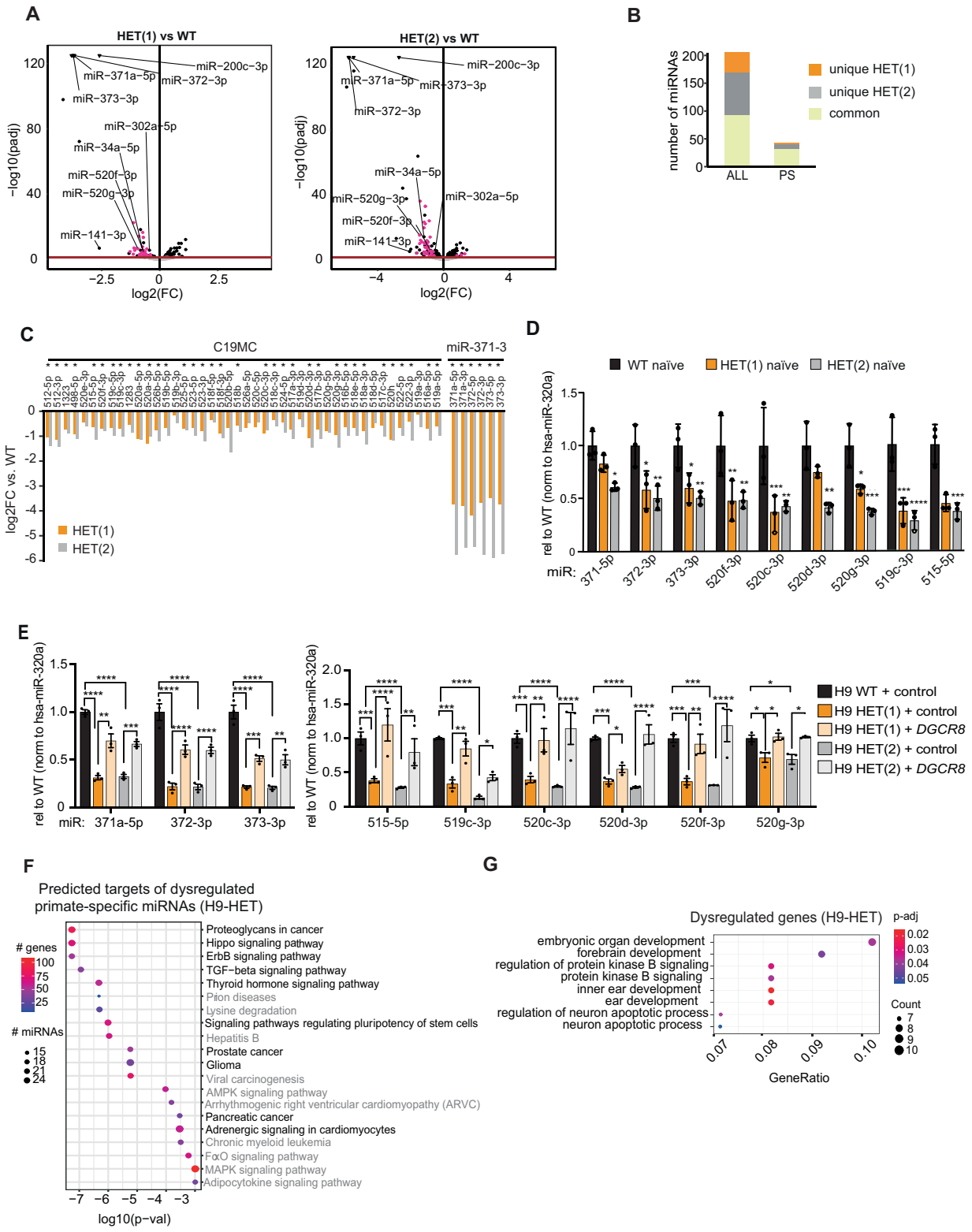


Figure 4. miRNA expression is affected in *DGCR8* heterozygote hESCs. (**A**) Differential expression (log_2FC) of miRNAs from two different clones (HET 1 and 2) of *DGCR8* HET hESCs compared with the parental cell line (WT). In brief, 69 and 93 miRNAs are downregulated whereas 59 and 76 miRNAs are upregulated in HET(1) and HET(2), respectively. MiRNAs in pink, indicate PS miRNAs. (**B**) Commonly dysregulated miRNAs in both HET H9 hESCs are significantly enriched in PS miRNAs (hypergeometric *P* 1.112e−5). (**C**) Log_2FC differential expression of miRNAs derived from the PS C19MC cluster and the miR 317–3 cluster in the two HET clones versus WT H9 hESC. Significant downregulated miRNAs are marked with an asterisk. (**D**) RT-qPCR of some mature C19MC and 371-3 miRNAs expression in H9 naïve hESCs and (**E**) in H9 WT and HET hESCs transduced with a control or a lentiviral vector expressing human *DGCR8*. Data represent the average ($n = 3$) ± SEM. Expression levels for each miRNAs are normalized to the levels of the *DGCR8*-independent miRNA, *hsa-miR-320-3p*, and expressed relative to WT levels. (*) $P \leq .05$, (**) $P \leq .01$, (***) $P \leq .001$, (****) $P \leq .0001$, by one-way ANOVA followed by Dunnett's or Tukey's multiple comparison test, respectively. (**F**) KEGG pathway analyses for the predicted targets (microT-CDS) of the commonly dysregulated PS miRNAs in both HET H9 clones. (**G**) GO pathway enrichment for differentially expressed genes in *DGCR8* HET hESC clones.

self-renewal, as well as embryonic development (e.g. 'TGF-β signalling', 'Hippo signalling', and 'ErbB signalling') [76–79] (Fig. 4F). Enriched pathways for PS miRNAs were similar to those obtained with all the dysregulated miRNAs (10 out of the top 20 predicted pathways) (compare Fig. 4F and Supplementary Fig. S3E, see common pathways in black). Despite the ontogenic differences between H9-hESCs and PA-1 cells and differences in their miRNA profile, we found that PA-1 HET cells also displayed a similar proportion of miRNAs being differentially expressed and a similar proportion of those being PS (Supplementary Fig. S3A and F). Dysregulated miRNAs were also predicted to regulate pathways involved in pluripotency maintenance and self-renewal in PA-1 cells (Supplementary Fig. S3G and H). These results highlight the potential importance of PS miRNAs, as a subgroup of dysregulated miRNAs during *DGCR8* haploinsufficiency.

To confirm the relevance of defective post-transcriptional gene silencing in HET cells, we performed total RNA high-throughput sequencing of HET H9 and PA-1 cells (Supplementary Fig. S4A and B and Supplementary Table S2). Functional enrichment analyses of differentially expressed genes ($P_{adj} \leq .05$) confirmed that affected pathways common to both cell lines were related to development, including 'embryonic organ development' (Fig. 4G and Supplementary Fig. S4C). To test if changes in gene expression were caused by defective miRNA levels, we investigated the expression of the predicted miRNA targets using the RNA-seq datasets. Changes in the expression of predicted targets for all dysregulated miRNAs (ALL), targets for the subset of PS dysregulated miRNAs, versus non-predicted targets (remaining genes) were compared. MiRNA-predicted targets significantly changed expression versus the non-targeted controls. This was true both for targets of all dysregulated miRNAs and only PS miRNAs in both PA-1 and H9 HET cells ($P < 2.22e{-}16$; Supplementary Fig. S4D and E). These findings suggest that, in part, alterations in the gene expression profile of HET cells can be attributed to altered post-transcriptional gene silencing. To better define the molecular mechanisms contributing to the defects of HET cells, we next investigated if the haploinsufficiency resulted in alterations in both well-defined canonical and non-canonical functions of *DGCR8*.

Heterozygous cells display defects in both the canonical and non-canonical functions of *DGCR8*

To investigate the impact of *DGCR8* heterozygosity on its canonical function, the biogenesis of miRNAs, we quantified Microprocessor cleavage efficiency both *in vitro* and in cells. For *in vitro* purposes, total cell extracts from the three PA-1 cell lines, WT, HET, and KO for *DGCR8*, were prepared. Extracts were incubated with radiolabelled pri-miRNAs to visualize precursor miRNA cleavage products as an indirect measurement of processing efficiency. We observed that extracts derived from HET cells only retained partial processing activity when compared with WTs, while KO extracts were not capable of processing pri-miRNAs (Fig. 5A). Next, to quantify pri-miRNA processing in cells, we measured the MPI using high-throughput sequencing data of chromatin-associated RNA for WT, HET, and KO PA-1 cells, as described in [33, 34]. For this purpose, cells were fractionated in cytoplasmic, nucleoplasmic, and chromatin fractions and confirmed that chromatin was enriched for pri-miRNA transcripts (Supplementary Fig. S5A and B). After sequencing, the MPI for each pri-miRNA was calculated as the negative $\log_2$ fraction of reads mapping to the hairpin versus reads mapping to the flanks of the pri-miRNA. The higher the value, the better processed the pri-miRNA, while values around 0 indicate absence of processing. Both Microprocessor-dependent and independent pri-miRNAs behaved as expected, with accumulation of reads over the hairpin of the Microprocessor-dependent pri-miRNA *pri-miR-374b*, in HET and KO cells, while no changes were observed for the DGCR8-independent miRNA *pri-miR-1234* (Fig. 5B; for more examples see Supplementary Fig. S5C). We next assessed how the MPI is affected by *DGCR8* heterozygosity and observed that, globally, PA-1 HET cells displayed an intermediate processing efficiency when compared with WT and KO cells (Fig. 5C; for a full list see Supplementary Table S3), suggesting that the biogenesis of miRNAs is affected in these cells.

To explore the relationship between changes in pri-miRNA processing efficiency and mature miRNAs, the levels of several pri-miRNA transcripts and mature miRNAs were compared by RT-qPCR in WT and HET cells. Despite pri-miRNAs accumulation in HET cells, no significant decrease in the mature miRNA levels was observed for some miRNAs, except for miR-135b and miR-767 (Fig. 5D and E). All these data together suggest that *DGCR8* heterozygosity results in defects in miRNA biogenesis, both *in vitro* and in cells. However, as we previously suggested for H9 hESCs, additional mechanisms may contribute to control the final mature miRNA levels.

Besides their canonical role in miRNA biogenesis, DGCR8 and Drosha have also been suggested to regulate gene expression at the transcriptional level, independently of miRNAs. These proteins have been shown to interact with promoter-proximal regions of human genes enhancing their transcription [80]. Furthermore, in an indirect manner, DGCR8 has been shown to alter gene transcription by regulating heterochromatin formation through physical association with KAP1 and HP1gamma [81]. Both functions of DGCR8 seem to be independent of the catalytical activity of Drosha. Thus, we next explored whether changes in chromatin structures or accessibility could be associated with the perturbation of gene expression observed in HET cells. For this purpose, we performed ATAC-seq in WT and HET PA-1 cells. As expected, an enrichment of ATAC-seq reads around the transcription start sites was observed for both cell lines (Supplementary Fig. S5D). Genome-wide differential peak analysis identified 52 347 high-confidence peaks and revealed that there was only a small proportion of peaks that were gained (0.36%, $n = 190$) and approximately twice as many were lost (0.6%, $n = 317$) in HET PA-1 cells (Supplementary Fig. S5E). Most differential peaks were located in introns and distal intergenic regions, followed by peaks annotated in proximal (≤1 kb) promoters (Supplementary Fig. S5F). In agreement with a role of chromatin accessibility in gene expression, 8% of genes containing a differential ATAC peak (±10 kb in distance) were also differentially expressed, according to the RNA-seq analysis. Although small, this enrichment was statistically significant ($P = 1.509e{-}9$), suggesting that changes in the chromatin accessibility of HET cells could also be influencing the gene expression profile. To further characterize the functional impact of changes in the accessibility of regulatory regions, we used the rGREAT package, which implements the Genomic Regions Enrichment of Annotations Tool [61]. This analysis revealed that some of the most significant terms associated with regions that lost accessibility in HET cells were

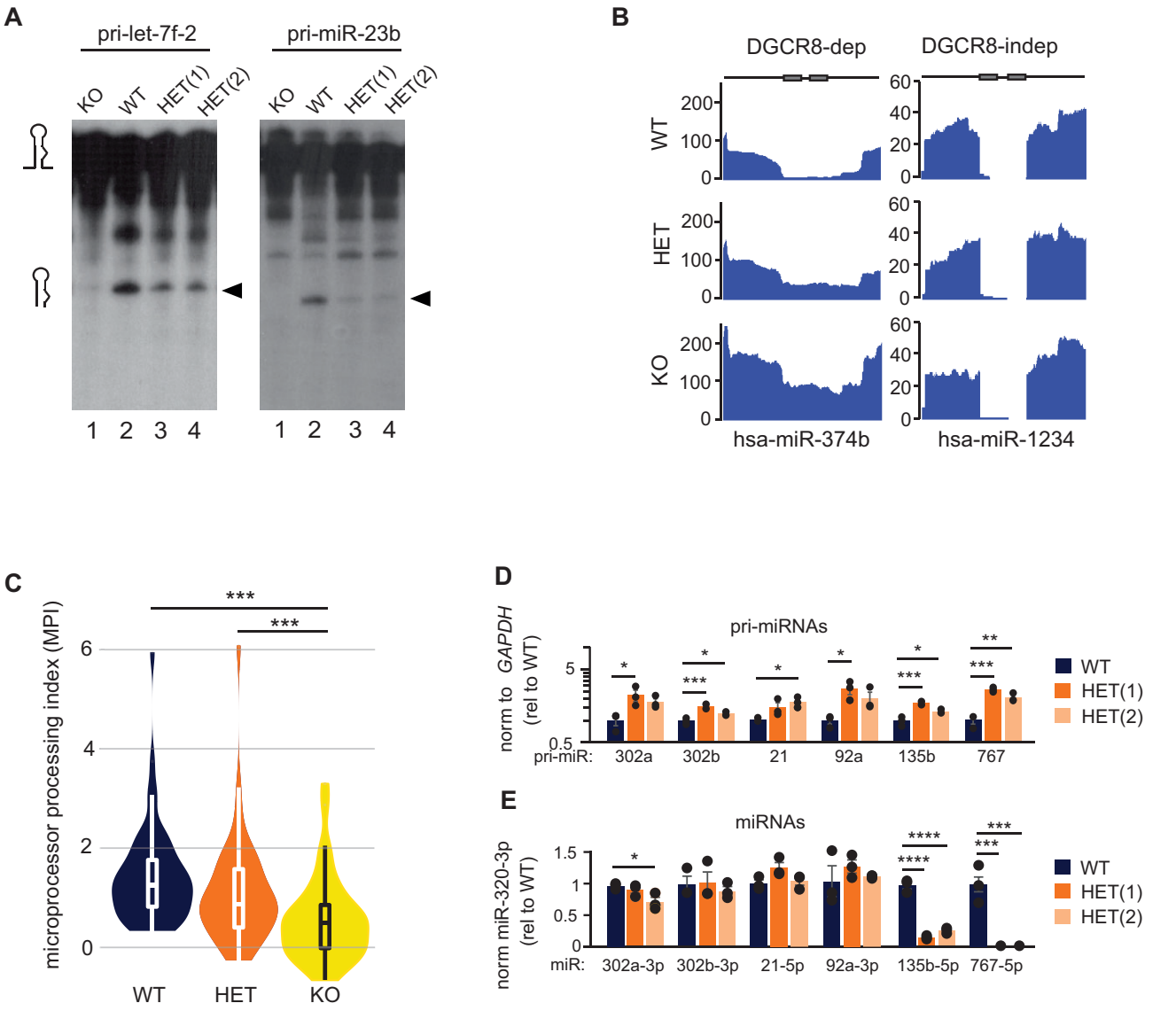


Figure 5. *DGCR8* heterozygosity in PA-1 cells results in impaired miRNA processing. (**A**) *In vitro* processing assays of radiolabelled pri-miRNAs, pri-let-7f-2, and pri-miR-23b, with WT, HET, and KO PA-1 derived extracts. Arrow is the resulting cleavage product, the pre-miRNA. (**B**) Read depth coverage across a *DGCR8*-dependent miRNA (*hsa-miR 374b*) and a *DGCR8*-independent mirtron (has-miR-1234) from chromatin-associated RNA high-throughput sequencing in WT, HET, and KO PA-1 cells. Grey boxes indicate mature miRNAs; black line represents surrounding genomic regions. (**C**) Violin plots for MPI values in WT, HET, and KO. Only pri-miRNAs with a minimum MPI > 0.3 in WT samples (processed in WT cells) were included in these analyses. (***) $P \leq .001$ by one-way ANOVA followed by Tukey's multiple comparison test. (**D**) Quantification of unprocessed pri-miRNA levels by RT-qPCR in WT and HET PA-1 cells. Data are normalized to *GAPDH* and represented relative to WT. Data are the average ($n = 3$) ± SEM. (*) $P \leq 0.5$, (***) $P \leq .001$ by one-way ANOVA followed by Dunnett's multiple comparison test (Y-axis, logarithmic) (**E**) Quantification of mature miRNA levels by RT-qPCR in WT and HET PA-1 cells. Data are normalized to the *DGCR8*-independent miRNA, *hsa-miR-320-3p*, and relative to WT levels. Data are the average ($n = 3$) ± SEM. (*) $P \leq 0.5$, (***) $P \leq .001$, (****) $P \leq .001$, by one-way ANOVA followed by Dunnett's multiple comparison test.

linked to 'development', 'differentiation', and ´morphogenesis" (Supplementary Fig. S5G). All these findings together suggest that the gene expression profile resulting from the loss of a single copy of *DGCR8* could be a combinatorial effect of both miRNA dysregulation and changes in chromatin accessibility.

DGCR8 haploinsufficiency reduces expression of primate-restricted HERVH and derived RNAs

Many TEs are transcribed during early human embryogenesis in a stage-specific manner and their expression is associated with stemness and pluripotency maintenance [28, 82]. For instance, knocking down the RNA derived from the HERVH or specific HERVH-derived RNAs (e.g. chimeric transcripts driven by their promoter activity) results in the loss of pluripotency and self-renewal capacity of hESCs [26–29]. To investigate if the stemness defects upon *DGCR8* haploinsufficiency could also originate from aberrant TE expression, we used a pipeline that allows analysing locus-specific TE expression using RNA-seq datasets (e.g. SQuIRE) [53]. This analysis revealed that a high number of genomic locations annotated by RepeatMasker as endogenous retroviruses type 1 (ERV1) were significantly downregulated in both H9 hESC HET clones compared with WT cells ($\log_2 FC < -1$; $P_{adj} \leq .05$; Fig. 6A). ERV1 retrotransposons have a structure resembling simple

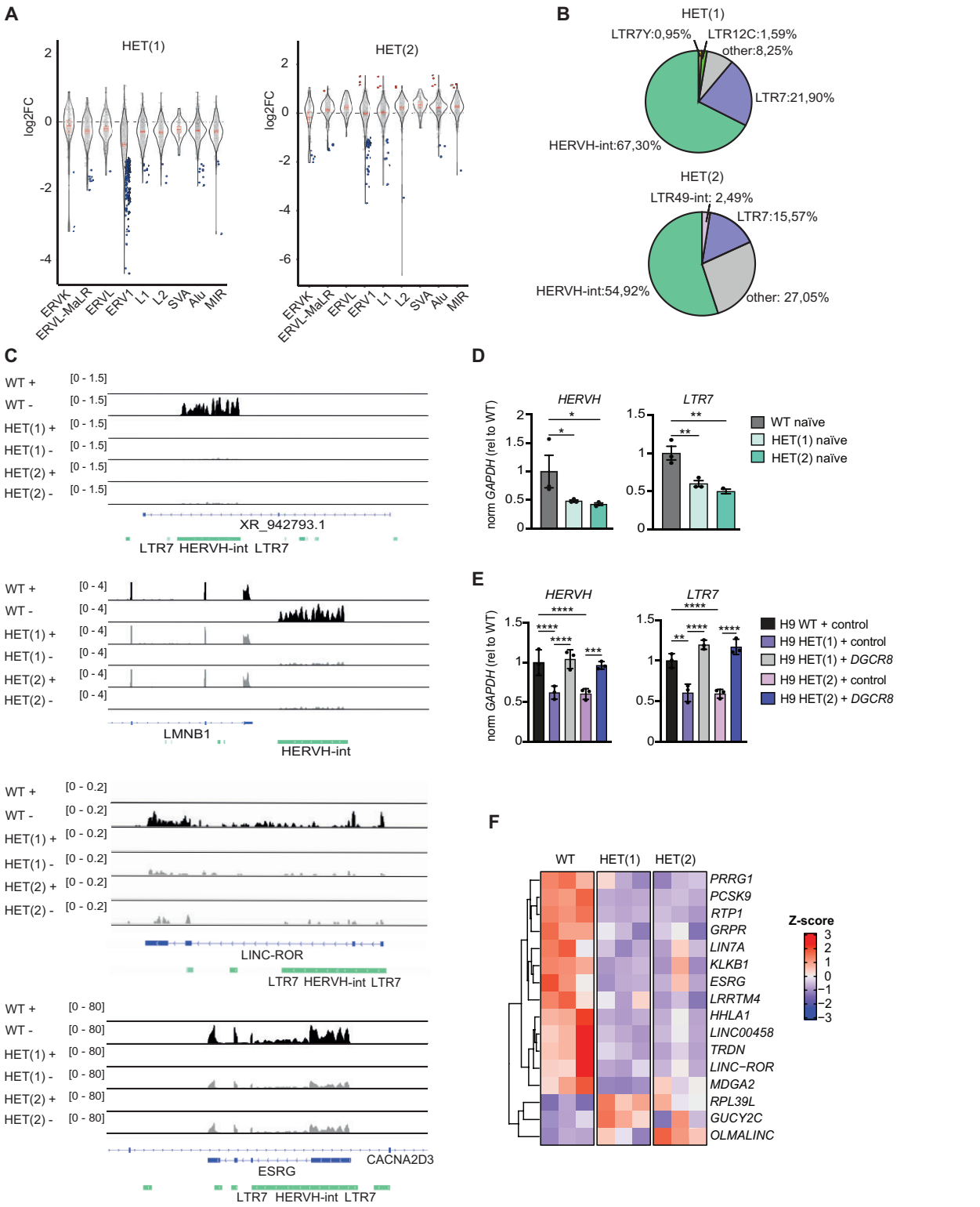


Figure 6. *DGCR8*-mediated control of HERVH expression. (**A**) Violin plots for locus-specific expression of retrotransposons in the two *DGCR8* hESC clones compared with WT cells. Significantly downregulated ($log_2FC < -1$; $P \leq .05$; baseMean > 100) and upregulated ($log_2FC > 1$; $P \leq .05$; baseMean > 100) elements are represented as blue and red dots, respectively. (**B**) Distribution of downregulated ERV1 subfamilies in both HET hESC clones. (**C**) Genome browser view of RNA-seq data from WT and HET *DGCR8* hESCs clones. Two representative regions containing differentially expressed LTR7-HERVH elements are shown, *XR_942793.1* and *LMNB1* (top panel) and two chimeric transcripts driven by LTR7 promoter activity are shown as representative examples, *LINC-ROR* and *ESRG* (bottom panel). Sense (+) and antisense (−) strands are represented. Genes and LTR7-HERVH are represented above and below, respectively. (**D**) RT-qPCR analyses for HERVH-int and LTR7 in WT and HET naïve hESCs (**E**) the same as (panel D), but in WT and HET-primed hESCs transduced with lentiviral control vector or with lentivirus expressing *DGCR8*. Data are normalized to *GAPDH* and relative to WT levels. Data are the average ($n = 3$) ± SD. (*) $P \leq .05$, (**) $P \leq .01$, (***) $P \leq .001$, (****) $P \leq .0001$, by one-way ANOVA, followed by Tukey's or Dunnett's multiple comparison test, respectively. (**F**) Heatmap for differentially expressed ($P \leq .05$) chimeric transcripts and lncRNAs driven by LTR7 promoter activity in HET hESCs.

retroviruses, as they encode for *gag* and *pol* genes, and are flanked by ∼450-bp long terminal repeats (LTRs) but they are not currently active in the human genome [28]. Downregulated ERV1 loci from HET H9 hESCs mostly belonged to the family members of the PS HERVH, with reads mapping to both their internal region (HERVH-int) and LTRs (known as LTR7) (89.2% and 70.49% of the ERV1 mapped reads in HET1 and HET2, respectively) (Fig. 6B, for locus-specific examples see Fig. 6C, upper panels).

HERVH elements are typically expressed in pluripotent human cells. Indeed, nearly half of all HERVH genomic copies (550 out of the 1225 full-length HERVH copies) are transcribed in hESC, although only a relatively small subset of loci ($n \sim 117$) is highly expressed [26]. High expression of HERVHs in hESCs is mostly driven by LTR7 rather than its counterparts, LTR7b, LTR7c, or LTR7y [26]. Remarkably, a significant proportion of the downregulated HERVH elements in both HET H9 hESCs clones belonged to the subgroup previously shown to be highly expressed in hESCs and mostly associated with LTR7 promoter activity (formerly known as Type I subfamily) (Supplementary Table S4). To validate if HERVH transcripts were reduced in H9 hESC HET cells, we measured the expression of both, total HERVH RNA levels and specific HERVH-derived transcripts. Using RT-qPCR, we observed a global reduction in LTR7 HERVH-int mRNAs in both naïve and primed *DGCR8* hESCs HET, which were rescued upon reintroduction of DGCR8 (Fig. 6D and E). Additionally, we observed a downregulation of specific HERVH-derived transcripts, as defined by [26], including the hESCs-specific lncRNAs and chimeric transcripts (lower panels in Fig. 6C and F). Downregulated HERVH-derived transcripts, including *linc-ROR*, *ESRG*, and *LINC00458*, have also been previously associated with pluripotency maintenance [29, 83]. All these data strongly suggest that *DGCR8* haploinsufficiency in pluripotent human cells leads to misregulation of HERVH expression. Furthermore, these data further support that the phenotype associated with *DGCR8* heterozygosity is species-specific, and that in human cells impacts the expression of PS miRNAs and HERVH retroelements.

C19MC and miR-371-373 miRNAs restore the molecular and cellular defects of *DGCR8* heterozygous cells

Our results suggest that *DGCR8* HET pluripotent human cells display two different PS molecular phenotypes. First, the downregulation of PS miRNAs and second, HERVH-derived transcripts, both of which are necessary for embryogenesis and pluripotency maintenance. We next wanted to test if HERVH downregulation was a consequence of the depletion of certain miRNAs, or if miRNAs and TEs were operating on separate pathways. To this end, HET cells were transfected with two different pools of miRNAs, belonging to the miR-371-3 cluster (hsa-miR-372-3p and hsa-miR-373-3p) and the C19MC cluster, to test their ability to rescue HERVH expression. Four miRNAs from the C19MC cluster were selected (hsa-miR-520g-3p, hsa-miR-520d-3p, hsa-miR-519c-3p, and hsa-miR-515-5p), as (i) these are significantly downregulated in HET hESCs, (ii) these display higher expression in hESCs than other members of the same cluster, (iii) these represent the diversity of seed sequences in the cluster, and (iv) these are predicted to silence important genes for pluripotency maintenance (Supplementary Fig. S6A and Supplementary Table S1). Full restoration of HERVH RNA levels was observed after reintroduction of miRNAs belonging to the miR-371-3 cluster and a partial rescue was observed after transfection of only the four miRNAs from the C19MC cluster (left panel in Fig. 7A). Similar results were obtained for HERVH-derived transcripts as *ESRG* and *LINC00458* (Fig. 7A, middle and right panel, respectively). Notably, KLF4 protein levels were also restored upon reintroduction of miRNAs from both clusters (Fig. 7B).

Next, to test if the rescue of HERVH levels upon miRNA reintroduction was dependent on KLF4, we measured HERVH after KLF4 depletion. Interestingly, the reduction of *KLF4* prevented HERVH increase upon miRNA reintroduction, demonstrating that the regulation of HERVH by these miRNAs is mediated by KLF4 (Fig. 7C and Supplementary Fig. S6B and C). Finally, to identify targets of these miRNAs that can explain the differences in KLF4 levels, we focused on the human-specific *ZNF398* and on *HEY2*, which have been previously implicated in pluripotency and embryonic development [84, 85]. These genes were selected as they were significantly upregulated in HET cells, in addition to containing predicted binding sites for the C19MC miRNAs at their 3′ untranslated regions. *ZNF398* contains four predicted miR-515-5p binding sites, while *HEY2* contains one miR-515-5p and two miR-519c-3p predicted binding sites (Supplementary Fig. S6A and Supplementary Table S2). Interestingly, we confirmed that reintroducing C19MC miRNAs resulted in reduced mRNA levels of *ZNF398* and *HEY2*, suggesting that they could be targets of these miRNAs (Fig. 7D). Interestingly, depletion of either *ZNF398* or *HEY2* resulted in upregulation of both KLF4 and HERVH. (Fig. 7E and F and Supplementary Fig. S6D). These suggest that the increased expression of *ZNF398*/*HEY2* upon *DGCR8* loss and subsequent alterations of C19MC miRNA levels could explain part of molecular phenotypes of HET cells, with reduced KLF4 and HERVH levels. Finally, we assessed whether overexpression of those miRNAs was also sufficient to restore some of the cellular phenotypes of HET cells. Transient transfection of miRNAs from the C19MC and miR-371/3 clusters re-established the clonal expansion capacity of *DGCR8* HET cells (Fig. 7G).

Importantly, most of these findings were confirmed when performing transient depletion of *DGCR8* in hESCs using siRNAs. These cells showed a reduction in DGCR8 and Drosha levels and a stable karyotype (Supplementary Fig. S7A and B). Consistent with our previous results, these cells retained the colony morphology and expression of pluripotency markers such as *NANOG* and *POU5F1*, but miRNAs belonging to the C19MC and miR-372-373 clusters as well as HERVH and KLF4 were reduced (Supplementary Fig. S7C–F). Interestingly, overexpression of these miRNAs was also sufficient to restore the molecular and cellular phenotypes during depletion (Supplementary Fig. S7G and H), confirming that these defects are caused primary by DGCR8 reduction and miRNA loss.

In sum, our findings indicate that the PS miRNA cluster C19MC, along with the miR-371-373 cluster, play crucial roles in hESC maintenance. Also, we showed that the loss of HERVH RNAs is a consequence of miRNA dysregulation and is mediated by KLF4 depletion (Fig. 7H). We predict that all these factors are acting in concert and contribute to the observed pluripotency defects of *DGCR8* HET cells.

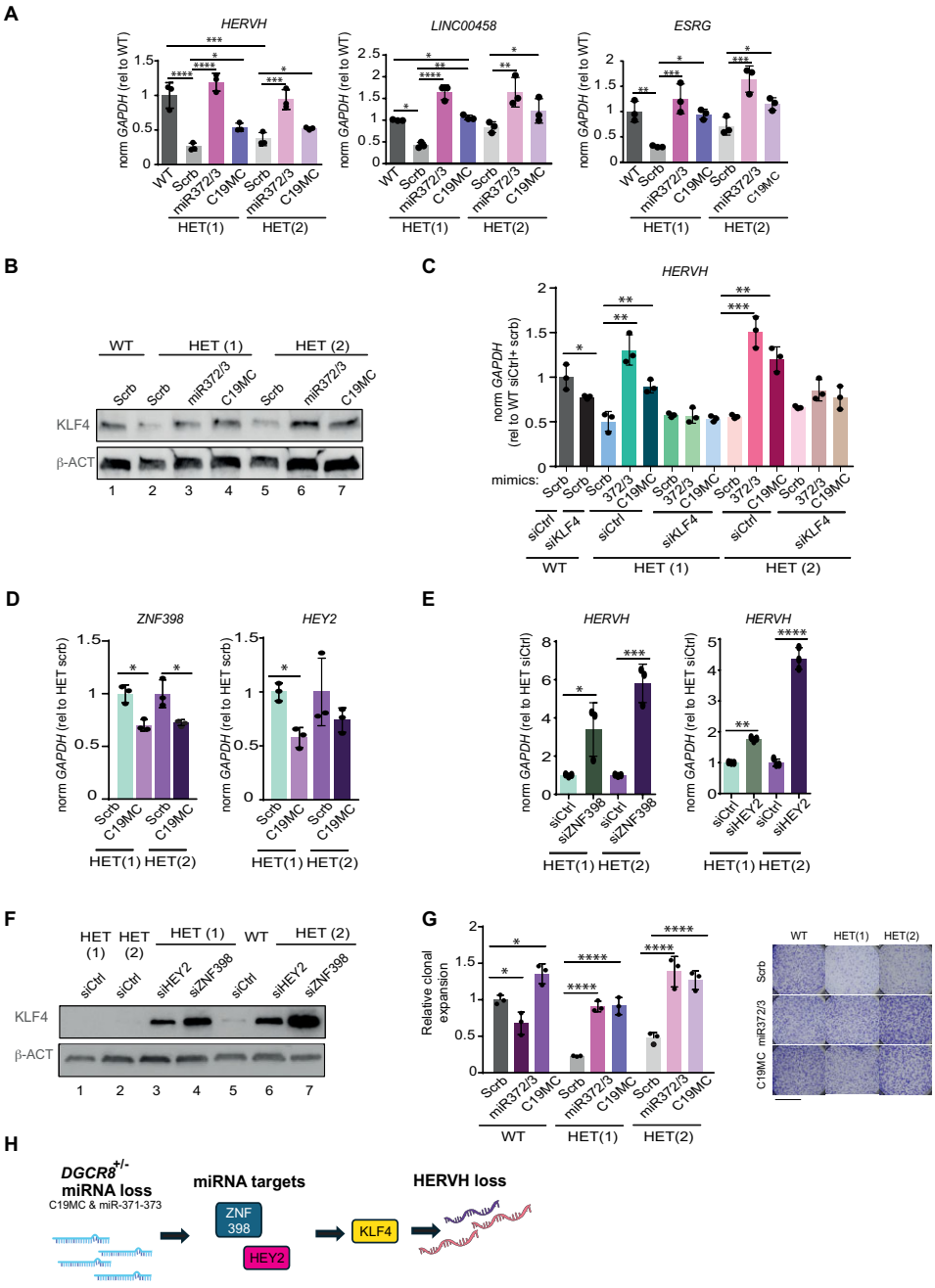


Figure 7. C19MC and miR-371-373 miRNAs rescue molecular and cellular defects in HET hESCs. (**A**) RT-qPCR analyses for HERVH-int and two HERVH chimeric transcripts, *LINC00458* and *ESRG*, in WT and HET hESCs transfected with mimic control (scrb) or with two miRNA mimics belonging to 371-3 cluster (372-3p and 373-3p) or four miRNA mimics from C19MC cluster (520g-3p, 520d-3p, 519c-3p, and 515-5p). Data are normalized to *GAPDH* and relative to WT levels. Data are the average ($n = 3$) ± SD. (*) $P \leq .05$, (**) $P \leq .01$, (***) $P \leq .001$, (****) $P \leq .0001$, by one-way ANOVA, followed by Tukey's multiple comparison test. (**B**) KLF4 western-blot analyses for WT and HET hESCs transfected with mimic control or with miRNA mimics belonging to 371-3 cluster and C19MC cluster. Actin serves as a loading control. (**C**) RT-qPCR for HERVH in WT and HET cells co-transfected with a siRNA (control or against *KLF4*) and miRNA mimics (scrb control or the pool of mimics belonging to 371-3 cluster or C19MC cluster). Data represent the average ($n = 3$) ± SD. Expression levels are normalized to *GAPDH* and expressed relative to the levels in WT cotransfected with siRNA and mimic control, (*) $P \leq .05$, (**) $P \leq .01$, (***) $P \leq .001$, by one-way ANOVA followed by Tukey's multiple comparison test. (**D**) RT-qPCR analyses for two predicted PS C19MC cluster target genes (*ZNF398* and *HEY2*) in WT and HET cells upon transfection with mimics from the C19MC cluster. Data represent the average ± SD of three biological replicates *GAPDH* and relative to HET control (scrb) levels. (*) $P \leq .05$, by one-way ANOVA followed by Tukey's multiple comparison test. (**E**) RT-qPCR analyses for HERVH in HET hESCs after depletion of *ZNF398* or *HEY2* with siRNAs. Data are normalized to *GAPDH* and expressed relative to HET siRNA control (siCtrl) levels. Data are the average ($n = 3$) ± SD. (*) $P \leq .05$, (**) $P \leq .01$, (***) $P \leq .001$, (****) $P \leq .0001$, by one-way ANOVA, followed by Tukey's multiple comparison test. (**F**) KLF4 western-blot analyses of WT and HET hESCs transfected with siRNA control (siCtrl) or siRNAs against *HEY2* or *ZNF398*. Actin serves as a loading control. (**G**) Relative clonal expansion capacity, expressed as the fraction of stained area with HET hESCs versus WT hESCs transfected with mimic control (scrb) or mimics of the 371-3 and C19MC cluster. Data represent the average ± SD of three biological replicates. (*) $P \leq .05$, (****) $P \leq .0001$, by two-way ANOVA followed by Tukey's multiple comparison test. Colonies are visualized by crystal violet staining (right panel). (**H**) Model for the molecular consequences of losing a single copy of *DGCR8* in hESCs. Upon *DGCR8* loss, the PS C19MC cluster and the miR-371-3 cluster are downregulated, and as result, their mRNA targets, *ZNF398* and *HEY2*, are upregulated. This upregulation leads to a reduction of the transcription factor KLF4, and as a consequence, of HERVH RNA levels.

Discussion

In this study, we have used two independent human pluripotent cellular models containing a single functional *DGCR8* allele to understand its relevance in the context of 22qDS. Our results indicate that *DGCR8* is haploinsufficient and that the most prominent defects are PS. Previous attempts to study the consequences of *DGCR8* haploinsufficiency were performed in mice. Mouse ESCs harbouring a single *Dgcr8* gene do not show significant defects in miRNA levels or differentiation [8, 13]. Despite the apparently negligible consequences, *Dgcr8* HET mice displayed behavioural and neuronal defects, which were attributed to altered miRNA expression [14]. In alternative to some of these findings in mice, our human models showed defects in pluripotency and dysregulation of miRNA expression. Inconsistencies between human and mouse models were also previously highlighted when comparing the transcriptome of 22qDS-derived neurons and those derived from the mouse model (*Df16A*$^{+/-}$), where no overlap was found [86, 87]. We hypothesize that this discrepancy is due to intrinsic differences between species. For instance, our results showed that *DGCR8* HET hESCs display increased apoptosis and defects in self-renewal, which demonstrates the importance of *DGCR8* for hESC survival. In agreement with this finding, *DGCR8* has also been identified as an essential gene for the survival of haploid hESCs [88]. In contrast, the total absence of *Dgcr8* in mESCs results in a reduction in proliferation but without changes in cell death or self-renewal [13]. Similarly, species-driven differences have been observed for DICER, another essential factor for miRNA biogenesis. While hESCs require *DICER1* for self-renewal, it seems to be dispensable for mESC survival [71]. These discrepancies could arise from differential sensitivities to defective miRNA levels within the different developmental stages that mouse and human ESCs represent; naïve versus primed states, respectively. However, our results showed that *DGCR8* HET hESCs maintain the main molecular and cellular defects after induction into a naïve-like stage, arguing against a cell-state-specific phenomenon and supporting species-specific differences. In agreement with these findings, our results suggest that PS miRNA dysregulation could be largely responsible for the stemness defects in *DGCR8* HET hESCs. We observed that a third of the miRNAs that are affected by *DGCR8* haploinsufficiency are PS and not present in rodents, some of which have been shown to be associated with pluripotency maintenance and self-renewal, including the C19MC cluster [73]. Our results show that the clonal expansion defects observed in *DGCR8* HET hESCs can be rescued by reintroducing four independent miRNAs belonging to this cluster, indicating its contribution to proliferation. Ree *et al.* have also suggested that *DGCR8* heterozygosity could result in processing defects of the C19MC cluster. However, their results showed some inconsistencies, probably due to lack of reduction in the DGCR8 protein levels for most of the heterozygous clones analysed [89].

Interestingly, only two of the miRNAs that rescue some of the phenotypes in *DGCR8* HET cells, miR-520d-3p and miR-519c-3p, share the seed sequence with miR-371-373 suggesting that the functions of these clusters in human pluripotent cells may be partially redundant. Similar to our findings, Teijero *et al.* showed that reintroduction of miR-372-3p and miR-373-3p rescues the apoptosis of *DICER* knockout hESCs [71]. Interestingly, a member of the mouse ortholog of this cluster (mmu-miR-294) has a role in mesoderm and endoderm EBs differentiation [90]. Considering the defects that we observed of mesoderm–endoderm differentiation in hESC HET cells, we proposed that the human miR-371-3, and mouse miR-290-295 may have conserved this function across species. Also in mice, the additive function of several miRNA families (miR-290-295 with miR-183-182 and miR-302) is required for optimal differentiation of mESCs [90, 91]. Interestingly, we also observed certain level of functional redundancy between the C19MC and miR371-3 on the preservation of critical stem cell functions.

The human and mouse early development display other remarkable differences, some of which seem to be driven by species-specific expression of transposable elements [28]. For instance, the murine endogenous retrovirus-L is transiently upregulated at the two-cell stage and is essential for mouse preimplantation development [92]. The endogenous retrovirus that colonized the common ancestor of apes, HERVH, is highly expressed in human pluripotent stem cells (hESCs and induced pluripotent stem cells, iPSCs) and epiblast, where it appears to play a role in promoting self-renewal and pluripotency [26, 93, 94]. Notably, almost all HERVH elements expressed in hESCs belong to a subfamily of elements transcribed from LTR7, which contains binding sites for specific transcription factors, including KLF4 [95, 96]. Haploinsufficiency of *DGCR8* led to a significant reduction of the HERVH/LTR7 subfamily transcripts and KLF4 levels, and these were both restored upon reintroduction of miRNAs. Our findings suggest that HERVH/LTR7 silencing is a consequence of miRNA downregulation and is mediated by KLF4 knockdown. The reduction of this PS endogenous retrovirus, in both naïve and primed stages, may also contribute to some of the cellular phenotypes of *DGCR8* HET hESCs.

Although both HET and WT hESCs clones display the chromosome 12 trisomy characteristic of hESC in culture, reintroduction of DGCR8 rescued both the cellular and molecular phenotypes, demonstrating that they are caused by DGCR8 loss. Also, most of our findings were recapitulated upon *DGCR8* transient depletion in hESCs with a normal karyotype (46,XX) (Supplementary Fig. S7). All these together confirm that our observations do not derive from clonal bias or are related to any indirect genomic alteration.

Importantly, we have shown that upregulation of two target transcripts of the C19MC cluster in HET cells leads to KLF4 depletion. These are (i) the *HEY2*, which plays an essential role in heart development [85] and (ii) the transcription factor *ZNF398*, which has a role in pluripotency as well as being specific to humans [84]. This supports the hypothesis that the phenotype associated with *DGCR8* haploinsufficiency in hESCs is largely related to PS pathways. It is also worth mentioning that while C19MC fully rescued KLF4 levels and the clonal expansion capacity, HERVH expression was only partially restored. This result suggests that KLF4 depletion in HET cells may contribute to pluripotency defects through additional mechanisms beyond reduced HERVH expression, as expected given the pleiotropic role of KLF4 in the stem cell transcriptional network [97].

Our previous results showed that RNAs derived from other types of TEs (LINEs and SINEs) are bound and processed by the Microprocessor [11]. However, HET hESCs showed no significant differences in the expression levels of these active retrotransposons, suggesting that inactivating a single allele

of *DGCR8* is not sufficient to abolish the post-transcriptional control of LINE/SINE in pluripotent cells. To identify which other functions of DGCR8 were altered in HET cells, we studied both the canonical and less-known functions. Similar to Stark *et al.* findings in mice [14], human *DGCR8* HET cells showed differential expression of a small proportion of miRNAs. However, miRNA biogenesis defects did not seem to be directly correlated with mature miRNA levels, indicating that additional factors may influence mature miRNA abundance, including differences in the transcription of precursor pri-miRNAs. Indeed, it has been shown that the recruitment of DGCR8 to superenhancers could boost both transcription and miRNA processing [98]. The retention of DGCR8–Drosha complex in the vicinity of C19MC genes [99] suggests that the reduction of DGCR8 could affect not only post-transcriptional processing but also transcription of C19MC and miR-371-373 miRNA clusters, which are located in close proximity in the genome. In addition to miRNA biogenesis, DGCR8 has been implicated in stimulating RNA-pol II transcription, but also promoting heterochromatin formation through interaction with KAP1 [80, 81]. ATAC-seq analysis of *DGCR8* HET cells only showed subtle changes in chromatin accessibility. Interestingly, affected regions were associated with development pathways, suggesting that changes at chromatin level could also be involved in some of the cellular phenotypes characterized in HET cells.

Although several studies have used iPSCs as model for 22qDS [87, 100, 101], there are also limitations associated with their use. Prenatal studies suggest that the incidence of the 22q11.2 microdeletion in foetuses is two or three times higher than the prevalence found in live births, suggesting that most 22q11.2 foetuses do not reach term (revised in [1]). Although the mechanisms leading to this fatal outcome are unknown, this highlights the extent of the heterogeneity in the presentation of the disease. Given the low frequency of DGCR8 LoF mutations in the general population, DGCR8 is one of the few genes predicted to be haploinsufficient in the entire deleted region of 22qDS, suggesting that its loss could have major impact on early human development and syndrome symptomatology. We hypothesize that certain levels of DGCR8 are required for survival and, consequently, our model allows studying the consequence of losing a single functional DGCR8 allele during early development without the requirement to be viable *in vivo*. Previously, Khan *et al.* showed that heterozygous loss of *DGCR8* in iPSCs-derived neurons recapitulated the calcium and excitability defects of 22qDS neurons [87]. Using heterozygous human pluripotent cell models, our analysis has focused on identifying the global effect of *DGCR8* loss in human embryonic cells and the molecular mechanism leading to this phenotype. Particularly, our results suggest that *DGCR8* could have a more profound role in the developmental issues present in 22qDS to what it had been previously suggested using other models and open new avenues to design novel targets for rescuing some of the developmental defects in 22qDS patients.

Collectively, we show that DGCR8 results in haploinsufficiency by altering the expression of PS miRNAs, as well as of PS transposable elements in human pluripotent cells independently of the developmental cellular stage. These findings stress the potential limitations of studying the function of human genes in evolutionarily distant animal models (e.g. rodents, zebrafish, etc.) where a proportion of the genome, especially the non-coding genome, is only partially conserved.

Acknowledgements

We are grateful to Javier F. Caceres for comments and critical reading of the manuscript. We thank Gregorz Kudla for initial bioinformatic analysis; Paul de Sousa and Rosa Montes for advice on hESC biology and techniques; and Marta García-Canadas, Jennifer Parra, Esther Prada, Meriem Benkaddour-Boumzaouad, and the GENYO's microscopy unit for technical support. We thank the Francisco Martin's group (GENYO, Granada) for help with the lentiviral transduction experiments. We would like to specially thank the Associations of 22q11.2 patients in Andalucia and Levante for their support and trust.

Author contributions: Conceptualization and supervision of this study: S.R.H. and S.M.; Methodology: A.C.B., L.K., S.M. and S.R.H.; Investigation with hESCs: A.C.B.; Investigation with PA-1 cells and formal analysis of transcriptomic approaches: L.K. supervised by A.I. and S.M.; Data curation and formal analysis of hESC and additional PA-1 datasets: G.P and G.B; Validation and technical support with hESCs: L.S.; Validation with PA-1 cells: P.C. and K.G; ATAC-seq experiment and analysis: S.P in the group of R.H.; HERVH formal analysis: G.P and A.G.G.; Resource generation (DGCR8 knockout PA1 cells and DGCR8 heterozygous hESCs): P.T.R.; Resources: J.L.G.P and F.J.S.L; Visualization: A.C.; Funding acquisition: S.M. and S.R.H; writing-original draft: A.C., S.M. and S.R.H.; Review and editing: All the authors. This article is part of the doctoral thesis of A.C.B., a student in the PhD programme in Pharmacy at the University of Granada (Spain).

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

None declared.

Funding

This work was supported by Ministerio de Ciencia e Innovación MCIN/AEI/10.13039/501100011033 [PID2020-115033RB-I00, PEJ2018-003280-A, RYC-2016-21395, and CNS2023-145402] ; ESF; Career Integration Grant—Marie Curie [FP7-PEOPLE-2011-CIG-303812]; FEDER/Consejería de Transformación Económica, Industria, Conocimiento y Universidades (Junta de Andalucía) [PY20_00619 and A-CTS-28_UGR20 grants]; donation to 'Aula de estudios 22qDS' (to S.R.H.); Wellcome Trust grants [221737/Z/20/Z and 107665/Z/15/Z]; the Royal Society grant [RGS\R1\191368]; the Wellcome Trust iTPA [PIII021] (to S.M.); MRC—Precision Medicine fellowship (to L.K.); Darwin Trust fellowship (to P.C.); Ministerio de Ciencia e Innovación, Agencia estatal de Investigación [PRE2021-098878] (to A.G.-G.); European Research Council (ERC) [ERC-Consolidator ERC-STG-2012-309433]; the Government of Spain [Ministerio de Ciencia e Innovación SAF2017-89745-R and PID2021-128934NB-I00]; Junta de Andalucia [P12-CTS-2256 and P18-RT-5067]; a private donation from Ms Francisca Serrano (Trading y Bolsa para Torpes, Granada, Spain) (to J.L.G.P.); and the Agencia Estatal de Investigación [RYC2021-031920-I, PID2022-143185NA-I00 and CNS2022-136033] (to F.J.S.L.). Funding to pay the

Open Access publication charges for this article was provided by Universidad de Granada/CBUA.

Data availability

All PA-1 sequencing data are deposited in GEO database, accession number GSE197474 (token: *qbulgmakldgrbap*). ATAC-seq datasets are deposited in GEO, accession number GSE205798 (token: *ivaduicidhwnlor*). H9 data are deposited in GEO, accession number GSE209843 (token: *obsnscqwbfshfuz*). Bioinformatic and software packages are described in the 'Materials and methods' section.

References

1. McDonald-McGinn DM, Sullivan KE, Marino B *et al.* 22q11.2 deletion syndrome. *Nat Rev Dis Primers* 2015;**1**:15071. https://doi.org/10.1038/nrdp.2015.71
2. Gebert LFR, Macrae IJ. Regulation of microRNA function in animals. *Nat Rev Mol Cell Biol* 2019;**20**:21–37. https://doi.org/10.1038/s41580-018-0045-7
3. Denli AM, Tops BBJ, Plasterk RHA *et al.* Processing of primary microRNAs by the Microprocessor complex. *Nature* 2004;**432**:231–5. https://doi.org/10.1038/nature03049
4. Gregory RI, Yan K-P, Amuthan G *et al.* The Microprocessor complex mediates the genesis of microRNAs. *Nature* 2004;**432**:235–40. https://doi.org/10.1038/nature03120
5. Han J, Lee Y, Yeom KH *et al.* The drosha-DGCR8 complex in primary microRNA processing. *Genes Dev* 2004;**18**:3016–27. https://doi.org/10.1101/gad.1262504
6. Landthaler M, Yalcin A, Tuschl T. The human DiGeorge syndrome critical region gene 8 and its *D. melanogaster* homolog are required for miRNA biogenesis. *Curr Biol* 2004;**14**:2162–7. https://doi.org/10.1016/j.cub.2004.11.001
7. Treiber T, Treiber N, Meister G. Regulation of microRNA biogenesis and its crosstalk with other cellular pathways. *Nat Rev Mol Cell Biol* 2019;**20**:5–20. https://doi.org/10.1038/s41580-018-0059-1
8. Han J, Pedersen JS, Kwon SC *et al.* Posttranscriptional crossregulation between Drosha and DGCR8. *Cell* 2009;**136**:75–84. https://doi.org/10.1016/j.cell.2008.10.053
9. Triboulet R, Chang H-M, Lapierre RJ *et al.* Post-transcriptional control of DGCR8 expression by the Microprocessor. *RNA* 2009;**15**:1005–11. https://doi.org/10.1261/rna.1591709
10. Knuckles P, Vogt MA, Lugert S *et al.* Drosha regulates neurogenesis by controlling neurogenin 2 expression independent of microRNAs. *Nat Neurosci* 2012;**15**:962–9. https://doi.org/10.1038/nn.3139
11. Heras SR, Macias S, Plass M *et al.* The Microprocessor controls the activity of mammalian retrotransposons. *Nat Struct Mol Biol* 2013;**20**:1173–81. https://doi.org/10.1038/nsmb.2658
12. Macias S, Plass M, Stajuda A *et al.* DGCR8 HITS-CLIP reveals novel functions for the Microprocessor. *Nat Struct Mol Biol* 2012;**19**:760–6. https://doi.org/10.1038/nsmb.2344
13. Wang Y, Medvid R, Melton C *et al.* DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nat Genet* 2007;**39**:380–5. https://doi.org/10.1038/ng1969
14. Stark KL, Xu B, Bagchi A *et al.* Altered brain microRNA biogenesis contributes to phenotypic deficits in a 22q11-deletion mouse model. *Nat Genet* 2008;**40**:751–60. https://doi.org/10.1038/ng.138
15. Earls LR, Fricke RG, Yu J *et al.* Age-dependent microRNA control of synaptic plasticity in 22q11 deletion syndrome and schizophrenia. *J Neurosci* 2012;**32**:14132–44. https://doi.org/10.1523/JNEUROSCI.1312-12.2012
16. Schofield CM, Hsu R, Barker AJ *et al.* Monoallelic deletion of the microRNA biogenesis gene Dgcr8 produces deficits in the development of excitatory synaptic transmission in the prefrontal cortex. *Neural Dev* 2011;**6**:11. https://doi.org/10.1186/1749-8104-6-11
17. Marinaro F, Marzi MJ, Hoffmann N *et al.* MicroRNA-independent functions of DGCR8 are essential for neocortical development and TBR1 expression. *EMBO Rep* 2017;**18**:603–18. https://doi.org/10.15252/embr.201642800
18. Fenelon K, Mukai J, Xu B *et al.* Deficiency of Dgcr8, a gene disrupted by the 22q11.2 microdeletion, results in altered short-term plasticity in the prefrontal cortex. *Proc Natl Acad Sci USA* 2011;**108**:4447–52. https://doi.org/10.1073/pnas.1101219108
19. Steiner DF, Thomas MF, Hu JK *et al.* MicroRNA-29 regulates T-box transcription factors and interferon-γ production in helper T cells. *Immunity* 2011;**35**:169–81.https://doi.org/10.1016/j.immuni.2011.07.009
20. Brandl A, Daum P, Brenner S *et al.* The microprocessor component, DGCR8, is essential for early B-cell development in mice. *Eur J Immunol* 2016;**46**:2710–8. https://doi.org/10.1002/eji.201646348
21. Khan IS, Taniguchi RT, Fasano KJ *et al.* Canonical micro RNA s in thymic epithelial cells promote central tolerance. *Eur J Immunol* 2014;**44**:1313–9. https://doi.org/10.1002/eji.201344079
22. Zimmermann C, Romero Y, Warnefors M *et al.* Germ cell-specific targeting of DICER or DGCR8 reveals a novel role for endo-siRNAs in the progression of mammalian spermatogenesis and male fertility. *PLoS One* 2014;**9**:e107023. https://doi.org/10.1371/journal.pone.0107023
23. Kim YS, Kim H-R, Kim H *et al.* Deficiency in DGCR8-dependent canonical microRNAs causes infertility due to multiple abnormalities during uterine development in mice. *Sci Rep* 2016;**6**:20242. https://doi.org/10.1038/srep20242
24. Bezman NA, Cedars E, Steiner DF *et al.* Distinct requirements of microRNAs in NK cell activation, survival, and function. *J Immunol* 2010;**185**:3835–46. https://doi.org/10.4049/jimmunol.1000980
25. Fromm B, Hoye E, Domanska D *et al.* MirGeneDB 2.1: toward a complete sampling of all major animal phyla. *Nucleic Acids Res* 2022;**50**:D204–10. https://doi.org/10.1093/nar/gkab1101
26. Wang J, Xie G, Singh M *et al.* Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature* 2014;**516**:405–9. https://doi.org/10.1038/nature13804
27. Lu X, Sachs F, Ramsay L *et al.* The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat Struct Mol Biol* 2014;**21**:423–5. https://doi.org/10.1038/nsmb.2799
28. Gerdes P, Richardson SR, Mager DL *et al.* Transposable elements in the mammalian embryo: pioneers surviving through stealth and service. *Genome Biol* 2016;**17**:100. https://doi.org/10.1186/s13059-016-0965-5
29. Römer C, Singh M, Hurst LD *et al.* How to tame an endogenous retrovirus: HERVH and the evolution of human pluripotency. *Curr Opin Virol* 2017;**25**:49–58. https://doi.org/10.1016/j.coviro.2017.07.001
30. Ran FA, Hsu PD, Wright J *et al.* Genome engineering using the CRISPR–Cas9 system. *Nat Protoc* 2013;**8**:2281–308. https://doi.org/10.1038/nprot.2013.143
31. Macia A, Widmann TJ, Heras SR *et al.* Engineered LINE-1 retrotransposition in nondividing human neurons. *Genome Res* 2017;**27**:335–48. https://doi.org/10.1101/gr.206805.116
32. Tristan-Manzano M, Maldonado-Perez N, Justicia-Lirio P *et al.* Lentiviral vectors for inducible, transactivator-free advanced therapy medicinal products: application to CAR-T cells. *Mol Ther Nucleic Acids* 2023;**32**:322–39. https://doi.org/10.1016/j.omtn.2023.03.018
33. Witteveldt J, Ivens A, Macias S. Inhibition of Microprocessor function during the activation of the type I interferon response. *Cell Rep* 2018;**23**:3275–85. https://doi.org/10.1016/j.celrep.2018.05.049

34. Conrad T, Marsico A, Gehre M *et al.* Microprocessor activity controls differential miRNA biogenesis *in vivo*. *Cell Rep* 2014;**9**:542–54. https://doi.org/10.1016/j.celrep.2014.09.007
35. Kim D, Paggi JM, Park C *et al.* Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 2019;**37**:907–15. https://doi.org/10.1038/s41587-019-0201-4
36. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie2. *Nat Methods* 2012;**9**:357–9. https://doi.org/10.1038/nmeth.1923
37. Clarke AW, Hoye E, Hembrom AA *et al.* MirGeneDB 3.0: improved taxonomic sampling, uniform nomenclature of novel conserved microRNA families and updated covariance models. *Nucleic Acids Res* 2025;**53**:D116–28. https://doi.org/10.1093/nar/gkae1094
38. Tristán-Ramos P, Rubio-Roldan A, Peris G *et al.* The tumor suppressor microRNA let-7 inhibits human LINE-1 retrotransposition. *Nat Commun* 2020;**11**:5712. https://doi.org/10.1038/s41467-020-19430-4
39. Li Q, Zhao X, Zhang W *et al.* Reliable multiplex sequencing with rare index mis-assignment on DNB-based NGS platform. *BMC Genomics* 2019;**20**:215.
40. Friedlander MR, Mackowiak SD, Li N *et al.* miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res* 2012;**40**:37–52. https://doi.org/10.1093/nar/gkr688
41. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucl Acids Res* 2014;**42**:D68–73. https://doi.org/10.1093/nar/gkt1181
42. Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. *Nucleic Acids Res* 2019;**47**:D155–62. https://doi.org/10.1093/nar/gky1141
43. Zhu A, Ibrahim JG, Love MI. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* 2019;**35**:2084–92. https://doi.org/10.1093/bioinformatics/bty895
44. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;**15**:550. https://doi.org/10.1186/s13059-014-0550-8
45. Vlachos IS, Zagganas K, Paraskevopoulou MD *et al.* DIANA-miRPath v3.0: deciphering microRNA function with experimental support. *Nucleic Acids Res* 2015;**43**:W460–6. https://doi.org/10.1093/nar/gkv403
46. Kanehisa M, Goto S, Sato Y *et al.* Data, information, knowledge and principle: back to metabolism in KEGG. *Nucl Acids Res* 2014;**42**:D199–205. https://doi.org/10.1093/nar/gkt1076
47. Yu G, Wang LG, Yan GR *et al.* DOSE: an R/bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics* 2015;**31**:608–9. https://doi.org/10.1093/bioinformatics/btu684
48. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014;**30**:923–30. https://doi.org/10.1093/bioinformatics/btt656
49. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010; https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
50. Dobin A, Davis CA, Schlesinger F *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;**29**:15–21. https://doi.org/10.1093/bioinformatics/bts635
51. Yu G, Wang L-G, Han Y *et al.* clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012;**16**:284–7. https://doi.org/10.1089/omi.2011.0118
52. Haunsberger SJ, Connolly NM, Prehn JH. miRNAmeConverter: an R/bioconductor package for translating mature miRNA names to different miRBase versions. *Bioinformatics* 2017;**33**:592–3. https://doi.org/10.1093/bioinformatics/btw660
53. Yang WR, Ardeljan D, Pacyna CN *et al.* SQuIRE reveals locus-specific regulation of interspersed repeat expression. *Nucleic Acids Res* 2019;**47**:e27. https://doi.org/10.1093/nar/gky1301
54. Tiedemann F. gghalves: compose half-half plots using your favourite geoms. 2020; https://erocoar.r-universe.dev/gghalves
55. Oh S, Abdelnabi J, Al-Dulaimi R *et al.* HGNChelper: identification and correction of invalid gene symbols for human and mouse. *F1000Res* 2020;**9**:1493. https://doi.org/10.12688/f1000research.28033.1
56. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 2016;**32**:2847–9. https://doi.org/10.1093/bioinformatics/btw313
57. Buenrostro JD, Wu B, Chang HY *et al.* ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol* 2015;**109**:21.29.1–9. https://doi.org/10.1002/0471142727.mb2129s109
58. Zhang Y, Liu T, Meyer CA *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 2008;**9**:R137. https://doi.org/10.1186/gb-2008-9-9-r137
59. Grandi FC, Modi H, Kampman L *et al.* Chromatin accessibility profiling by ATAC-seq. *Nat Protoc* 2022;**17**:1518–52. https://doi.org/10.1038/s41596-022-00692-9
60. Heinz S, Benner C, Spann N *et al.* Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol Cell* 2010;**38**:576–89. https://doi.org/10.1016/j.molcel.2010.05.004
61. Gu Z, Hubschmann D. rGREAT: an R/bioconductor package for functional enrichment on genomic regions. *Bioinformatics* 2023;**39**:btac745. https://doi.org/10.1093/bioinformatics/btac745
62. Zhu LJ, Gazin C, Lawson ND *et al.* ChIPpeakAnno: a bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinf* 2010;**11**:237. https://doi.org/10.1186/1471-2105-11-237
63. Karczewski KJ, Francioli LC, Tiao G *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020;**581**:434–43. https://doi.org/10.1038/s41586-020-2308-7
64. Karbarz M. Consequences of 22q11.2 microdeletion on the genome, individual and population levels. *Genes* 2020;**11**:977. https://doi.org/10.3390/genes11090977
65. Zeuthen J, Norgaard JO, Avner P *et al.* Characterization of a human ovarian teratocarcinoma-derived cell line. *Int J Cancer* 1980;**25**:19–32. https://doi.org/10.1002/ijc.2910250104
66. Garcia-Perez JL, Morell M, Scheys JO *et al.* Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. *Nature* 2010;**466**:769–73. https://doi.org/10.1038/nature09209
67. Nichols J, Smith A. Naive and primed pluripotent states. *Cell Stem Cell* 2009;**4**:487–92. https://doi.org/10.1016/j.stem.2009.05.015
68. Messmer T, von Meyenn F, Savino A *et al.* Transcriptional heterogeneity in naive and primed human pluripotent stem cells at single-cell resolution. *Cell Rep* 2019;**26**:815–24. https://doi.org/10.1016/j.celrep.2018.12.099
69. Guo G, von Meyenn F, Rostovskaya M *et al.* Epigenetic resetting of human pluripotency. *Development* 2017;**144**:2748–63. https://doi.org/10.1242/dev.146811
70. Theunissen TW, Powell BE, Wang H *et al.* Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell* 2014;**15**:524–6. https://doi.org/10.1016/j.stem.2014.09.003
71. Teijeiro V, Yang D, Majumdar S *et al.* DICER1 is essential for self-renewal of human embryonic stem cells. *Stem Cell Rep* 2018;**11**:616–25. https://doi.org/10.1016/j.stemcr.2018.07.013
72. Kobayashi N, Okae H, Hiura H *et al.* The microRNA cluster C19MC confers differentiation potential into trophoblast lineages upon human pluripotent stem cells. *Nat Commun* 2022;**13**:3071. https://doi.org/10.1038/s41467-022-30775-w

73. Mong EF, Yang Y, Akat KM *et al.* Chromosome 19 microRNA cluster enhances cell reprogramming by inhibiting epithelial-to-mesenchymal transition. *Sci Rep* 2020;**10**:3029. https://doi.org/10.1038/s41598-020-59812-8
74. Auyeung VC, Ulitsky I, McGeary SE *et al.* Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. *Cell* 2013;**152**:844–58. https://doi.org/10.1016/j.cell.2013.01.031
75. Nguyen TA, Jo MH, Choi YG *et al.* Functional anatomy of the human microprocessor. *Cell* 2015;**161**:1374–87. https://doi.org/10.1016/j.cell.2015.05.010
76. Chan R, Hardy WR, Laing MA *et al.* The catalytic activity of the ErbB-2 receptor tyrosine kinase is essential for embryonic development. *Mol Cell Biol* 2002;**22**:1073–8. https://doi.org/10.1128/MCB.22.4.1073-1078.2002
77. James D, Levine AJ, Besser D *et al.* TGFβ/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. *Development* 2005;**132**:1273–82. https://doi.org/10.1242/dev.01706
78. Ramos A, Camargo FD. The Hippo signaling pathway and stem cell biology. *Trends Cell Biol* 2012;**22**:339–46. https://doi.org/10.1016/j.tcb.2012.04.006
79. Xu Z, Robitaille AM, Berndt JD *et al.* Wnt/β-catenin signaling promotes self-renewal and inhibits the primed state transition in naïve human embryonic stem cells. *Proc Natl Acad Sci USA* 2016;**113**:E6382–90. https://doi.org/10.1073/pnas.1613849113
80. Gromak N, Dienstbier M, Macias S *et al.* Drosha regulates gene expression independently of RNA cleavage function. *Cell Rep* 2013;**5**:1499–510. https://doi.org/10.1016/j.celrep.2013.11.032
81. Deng L, Ren R, Liu Z *et al.* Stabilizing heterochromatin by DGCR8 alleviates senescence and osteoarthritis. *Nat Commun* 2019;**10**:3329. https://doi.org/10.1038/s41467-019-10831-8
82. Torres-Padilla ME. On transposons and totipotency. *Phil Trans R Soc B* 2020;**375**:20190339. https://doi.org/10.1098/rstb.2019.0339
83. Sexton CE, Tillett RL, Han MV. The essential but enigmatic regulatory role of HERVH in pluripotency. *Trends Genet* 2022;**38**:12–21. https://doi.org/10.1016/j.tig.2021.07.007
84. Zorzan I, Pellegrini M, Arboit M *et al.* The transcriptional regulator ZNF398 mediates pluripotency and epithelial character downstream of TGF-beta in human PSCs. *Nat Commun* 2020;**11**:2364. https://doi.org/10.1038/s41467-020-16205-9
85. Fischer A, Schumacher N, Maier M *et al.* The Notch target genes Hey1 and Hey2 are required for embryonic vascular development. *Genes Dev* 2004;**18**:901–11. https://doi.org/10.1101/gad.291004
86. Sun Z, Williams DJ, Xu B *et al.* Altered function and maturation of primary cortical neurons from a 22q11.2 deletion mouse model of schizophrenia. *Transl Psychiatry* 2018;**8**:85. https://doi.org/10.1038/s41398-018-0132-8
87. Khan TA, Revah O, Gordon A *et al.* Neuronal defects in a human cellular model of 22q11.2 deletion syndrome. *Nat Med* 2020;**26**:1888–98. https://doi.org/10.1038/s41591-020-1043-9
88. Yilmaz A, Peretz M, Aharony A *et al.* Defining essential genes for human pluripotent stem cells by CRISPR–Cas9 screening in haploid cells. *Nat Cell Biol* 2018;**20**:610–9. https://doi.org/10.1038/s41556-018-0088-1
89. Reé D, Fóthi Á, Varga N *et al.* Partial disturbance of microprocessor function in human stem cells carrying a heterozygous mutation in the DGCR8 gene. *Genes* 2022;**13**:1925. https://doi.org/10.3390/genes13111925
90. Wang XW, Hao J, Guo WT *et al.* A DGCR8-independent stable microRNA expression strategy reveals important functions of miR-290 and miR-183-182 families in mouse embryonic stem cells. *Stem Cell Rep* 2017;**9**:1618–29. https://doi.org/10.1016/j.stemcr.2017.08.027
91. Gu KL, Zhang Q, Yan Y *et al.* Pluripotency-associated miR-290/302 family of microRNAs promote the dismantling of naive pluripotency. *Cell Res* 2016;**26**:350–66. https://doi.org/10.1038/cr.2016.2
92. Sakashita A, Kitano T, Ishizu H *et al.* Transcription of MERVL retrotransposons is required for preimplantation embryo development. *Nat Genet* 2023;**55**:484–95. https://doi.org/10.1038/s41588-023-01324-y
93. Grow EJ, Flynn RA, Chavez SL *et al.* Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* 2015;**522**:221–5. https://doi.org/10.1038/nature14308
94. Goke J, Lu X, Chan YS *et al.* Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. *Cell Stem Cell* 2015;**16**:135–41. https://doi.org/10.1016/j.stem.2015.01.005
95. Carter TA, Singh M, Dumbović G *et al.* Mosaic *cis*-regulatory evolution drives transcriptional partitioning of HERVH endogenous retrovirus in the human embryo. *eLife* 2022;**11**:e76257. https://doi.org/10.7554/eLife.76257
96. Ohnuki M, Tanabe K, Sutou K *et al.* Dynamic regulation of human endogenous retroviruses mediates factor-induced reprogramming and differentiation potential. *Proc Natl Acad Sci USA* 2014;**111**:12426–31. https://doi.org/10.1073/pnas.1413299111
97. Di Giammartino DC, Kloetgen A, Polyzos A *et al.* KLF4 is involved in the organization and regulation of pluripotency-associated three-dimensional enhancer networks. *Nat Cell Biol* 2019;**21**:1179–90. https://doi.org/10.1038/s41556-019-0390-6
98. Suzuki HI, Young RA, Sharp PA. Super-enhancer-mediated RNA processing revealed by integrative microRNA network analysis. *Cell* 2017;**168**:1000–14. https://doi.org/10.1016/j.cell.2017.02.015
99. Bellemer C, Bortolin-Cavaille ML, Schmidt U *et al.* Microprocessor dynamics and interactions at endogenous imprinted C19MC microRNA genes. *J Cell Sci* 2012;**125**:2709–20. https://doi.org/10.1242/jcs.100354
100. Zhao D, Lin M, Chen J *et al.* MicroRNA profiling of neurons generated using induced pluripotent stem cells derived from patients with schizophrenia and schizoaffective disorder, and 22q11.2 del. *PLoS One* 2015;**10**:e0132387. https://doi.org/10.1371/journal.pone.0132387
101. Nehme R, Pietilainen O, Artomov M *et al.* The 22q11.2 region regulates presynaptic gene-products linked to schizophrenia. *Nat Commun* 2022;**13**:3690. https://doi.org/10.1038/s41467-022-31436-8

Received: June 3, 2024. **Editorial Decision:** February 13, 2025. **Accepted:** March 3, 2025