Alu retrotransposons promote differentiation of human carcinoma cells through the aryl hydrocarbon receptor

Citation for published version:

Digital Object Identifier (DOI):
10.1093/nar/gkw095

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Nucleic Acids Research

Publisher Rights Statement:
© The Author(s) 2016. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Alu retrotransposons promote differentiation of human carcinoma cells through the aryl hydrocarbon receptor

Antonio Morales-Hernández1, Francisco J. González-Rico1, Angel C. Román2, Eva Rico-Leo1, Alberto Alvarez-Barrientos3, Laura Sánchez4, Ángela Macía4, Sara R. Heras4, José L. García-Pérez4, Jaime M. Merino1 and Pedro M. Fernández-Salguero1,*

1Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias, Universidad de Extremadura, Avenida de Elvas s/n, 06071-Badajoz, Spain, 2Instituto Cajal, Consejo Superior de Investigaciones Científicas, Avenida Doctor Arce 37, 28002-Madrid, Spain, 3Servicio de Técnicas Aplicadas a las Biociencias, Universidad de Extremadura, Avenida de Elvas s/n 06071-Badajoz, Spain and 4GENYO. Centro de Genómica e Investigación Oncológica: Pfizer/Universidad de Granada/Junta de Andalucía, Avda. de la Ilustración 114, PTS Granada, 18016-Granada, Spain

Received November 13, 2015; Revised February 7, 2016; Accepted February 9, 2016

ABSTRACT

Cell differentiation is a central process in development and in cancer growth and dissemination. OCT4 (POU5F1) and NANOG are essential for cell stemness and pluripotency; yet, the mechanisms that regulate their expression remain largely unknown. Repetitive elements account for almost half of the Human Genome; still, their role in gene regulation is poorly understood. Here, we show that the dioxin receptor (AHR) leads to differentiation of human carcinoma cells through the transcriptional upregulation of Alu retrotransposons, whose RNA transcripts can repress pluripotency genes. Despite the genome-wide presence of Alu elements, we provide evidences that those located at the NANOG and OCT4 promoters bind AHR, are transcribed by RNA polymerase-III and repress NANOG and OCT4 in differentiated cells. OCT4 and NANOG repression likely involves processing of Alu-derived transcripts through the miRNA machinery involving the Microprocessor and RISC. Consistently, stable AHR knockdown led to basal undifferentiation, impaired Alu transcription and blockade of OCT4 and NANOG repression. We suggest that transcripts produced from AHR-regulated Alu retrotransposons may control the expression of stemness genes OCT4 and NANOG during differentiation of carcinoma cells. The control of discrete Alu elements by specific transcription factors may have a dynamic role in genome regulation under physiological and diseased conditions.

INTRODUCTION

Recent evidences suggest that active transposable elements (TEs) have an important role in defining Human Genome structure and function and, consequently, in controlling development and disease (1,2). Short interspersed nuclear Alu elements (SINE) are a class of TEs highly abundant in the Human Genome that account for nearly 10% of its size (3). Alu retrotransposons derive from the 7SL RNA and are highly abundant in non-coding genomic regions including upstream promoters and gene introns (4,5). Previous studies have shown that global transposon activity varies under diverse cellular conditions; yet, very little is known regarding the mechanisms through which TEs regulate the expression of specific genes (6). In this context, a recent study revealed that an Alu element inserted in human chromosome 9p21 within the long non-coding RNA (lncRNA) ANRIL was needed to trans-regulate genes presumably involved in atherosclerosis (7). In the mouse, Anril lncRNA regulated cell proliferation and differentiation through the Cdkn2A/B gene (8).

Notably, TEs are potential carriers of binding sites for transcription factors. Genome-wide analyses have found an enrichment of binding sites for ESR1, TP53, OCT4 (POU5F1), SOX2 and CTCF in human TEs (9–11). In fact, TEs provide up to 25% of the binding sites for the pluripotency regulators OCT4 (POU5F1) and NANOG and for the chromatin remodeler CTCF in both human and mouse embryonic stem (ES) cells (10). Consequently, it appears plausible that TEs assume an important role in the control
of transcriptional programs that regulate cell turnover and plasticity (10). Furthermore, certain classes of TEs were up-regulated whereas others were downmodulated during the reprogramming of differentiated cells into induced pluripotent stem (iPSc) cells, thus producing an expression profile reminiscent of that of ES cells (12,13). Overall, these former studies suggest that TEs could modulate specific transcriptional programs that drive pluripotency and cell reprogramming (12). Previous work from our laboratory identified a novel B1-SINE retrotransposon (B1-X35S) widely represented in upstream regulatory regions of the mouse genome that acts as a genomic insulator blocking target gene expression (14,15). B1-X35S-dependent insulation required the interaction of transcription factors dioxin receptor (AhR) and Slug (Sna12) with their consensus sequences present in B1-X35S and the transcriptional activity of RNA polymerases III and II (15,16).

It is becoming increasingly clear that some repetitive elements are relevant for cell functioning. Recent efforts have identified repetitive sequences with the potential to regulate gene expression and to participate in the control of specific cell processes under normal and pathological conditions (15,17–19). In this work, we have investigated the functional relevance of Alu retrotransposons regulated by the dioxin receptor AHR in the differentiation of human embryonic carcinoma cells. We have focused on individual Alu elements located in the upstream regulatory regions of pluripotency genes NANOG and OCT4. AHR exerted a pro-differentiation role that was responsible, at least in part, for the repression of OCT4 and NANOG. Interestingly, such repressive mechanism appears to be mediated by non-coding RNA transcripts produced by RNA polymerase III (15,16).

Human embryonic teratocarcinoma NTERA-wt and NTERA-sh cells were cultured in DMEM containing 10% FBS, 100 U/ml penicillin, 100 μg/ml streptomycin and 2 mM L-glutamine at 37°C and 5% CO2 atmosphere. NTERA-wt and NTERA-sh cell lines were authenticated by DNA profiling using 8 different and highly polymorphic short tandem repeat (STR) loci (German Biological Resource Centre DSMZ). The AhR agonist 6-formylindolo[3,2-b]carbazole (FICZ) was from Enzo.

Retroviral knockdown of human AHR

AHR was knocked-down in NTERA cells by retroviral transduction essentially as described (20,21) using the shRNA sequence 5'-TGCTGTGACAGTGAGCAGCAAGATGTACCTTCCTTGGAAATTCATTGCCTGCCT-3'. For rescue experiments, an shRNA was synthesized targeting the 3' UTR region of the human AHR: 5'-TGCTGTGACAGTGAGCAGCAAGATGTACCTTCCTTGGAAATTCATTGCCTGCCT-3'. shRNAs sequences were designed using the algorithm available at: http://www.stanford.edu/group/nolan/retroviral_systems/retsyst.html.

Bioinformatic analysis of Alu elements containing AHR binding sites

The Human Genome was analyzed for the presence of conserved elements containing an XRE site and an E-box using the algorithm previously described (14). Two of the most abundant groups of elements (e.g. X36S and X45S) were analyzed for their presence in the 5'promoter region of pluripotency and stemness genes within a 200-bp interval from the transcription start site. The background expression of Nanog and Oct4 Alus was estimated by referring the RNA Seq data under basal cell conditions (untreated NTERA-wt cells) to the total number of AluSq and AluSx elements identified in the Human Genome.

Chromatin immunoprecipitation (ChIP)

Chromatin immunoprecipitation (ChIP) and double-sequential ChIP for AHR and OCT4 binding to the Nanog and OCT4 Alus were performed in NTERA-wt cells essentially as described (15,22). NTERA-sh cells were used as negative controls for ChIP experiments.

MATERIALS AND METHODS

Antibodies

The following antibodies were used: βIII-tubulin (Santa Cruz Biotechnology sc-58888, clone TUJ-1), GAP43 (Millipore AB-5220), Tau (generous gift of Dr Lorenzo-Benayas, University of Extremadura), GAPDH (Cell Signaling 2118, clone 14C10), OCT4 (Santa Cruz Biotechnology sc-5279, clone C-10), NANOG (AbCam Ab-21624), AGO2 (Millipore 03–110), AHR (ENZO Life Sciences BML-SA210 and Immunostep 130605–1) and β-Actin (Sigma A2066).

Cells and reagents

The following reagents were used: 35S-succinylcysteine (Pierce, clone C-10), NANOG (AbCam Ab-21624), AGO2 (Millipore), OCT4 (Santa Cruz Biotechnology sc-58888, clone TUJ-1), GAP43 (Millipore), 6-formylindolo[3,2-b]carbazole (FICZ) was from Enzo.

Immunostep 130605–1) and 

SYBRR Green I was from Life Technologies. The iScript™ Reverse Transcription Supermix was from Bio-Rad and the SYBR Select Master Mix was from Life Technologies. The AhR agonist 6-formylindololo[3,2-b]carbazole (FICZ) was from Enzo.
RNA immunoprecipitation (RIP)

RNA immunoprecipitation (RIP) for **NANOG Alu-T1** and **Alu-T2** loading into AGO2 was done essentially as described (23). Briefly, NTERA-wt cells were lysed in polysome lysis buffer containing RNase and protease inhibitors. Pre-washed protein-A/G plus agarose was incubated overnight at 4°C with 5 μg anti-AGO2 antibody or the same amount of IgG. After washing, the mRNP lysate was added and the mix briefly centrifuged to obtain the input fraction in the supernatant. Reactions were then gently rotated at room temperature for 2 h, washed and incubated for 30 min in NT2 buffer containing sodium dodecyl sulphate and proteinase K. RNA was phenol-chloroform-isoamyl alcohol extracted and ethanol precipitated. RIP was analyzed by qPCR using the oligonucleotides indicated in Supplementary Table S1. Polymerase chain reaction (PCR) reactions were also analyzed in agarose gels to determine the presence of the expected product.

RNA sequencing (RNAseq)

RNA sequencing (RNAseq) was performed in NTERA-wt and NTERA-sh cells in presence or absence of 1 μM RA. Libraries were made using RNA molecules of 75–250 bases in length isolated from agarose gels. Libraries were sequenced in an Illumina MiSeq equipment, which produced 8.32, 5.92, 7.70 and 8.79 million lectures for NTERA-wt (UT), NTERA-wt (RA), NTERA-sh (UT) and NTERA-sh (RA), respectively. FASTQ raw reads were transformed into FASTA format using a custom Perl script available on request. These 75-bp FASTA reads were integrated into a database, and NCBI Blast was applied using **OCT4** and **NANOG Alu** full-length sequences as query. Search parameters were adjusted for the detection of sequences with the highest similarity, and normalized against the values obtained for standard **Alu** elements. When the analysis of identical sequences was required, ad hoc Perl scripts comparing specific sequences against the full Human Genome sequence were used.

Synthesis and cloning of **NANOG** and **OCT4 Alus** and 3′-UTR sequences

Wild-type **NANOG** and **OCT4 Alu** elements and **NANOG Alu** mutated in the XRE site and/or A + B boxes were synthesized by GenScript and cloned in the promoter-less pUC57 vector. **NANOG** mRNA 3′UTR was amplified using the primers indicated in Supplementary Table S1 and cloned in the pGL2-Basic-Luc vector. **NANOG** 3′UTR mutants lacking the homology sequences for **Alu-T1** and/or **Alu-T2** were constructed by PCR from the wild-type sequence and cloned in the pGL2-Basic-Luc vector. The pUC57 vector does not contain any promoter and it is therefore indicated to assay the endogenous transcription of **NANOG** and **OCT4 Alus** by RNA polIII.

Mouse teratoma induction

Swiss male immunodeficient mice at 8–10 weeks of age (Charles Rives) were used for teratoma induction. Mouse experimentation was approved and followed the regulation established by the Bioethics and Biosecurity Commission of the University of Granada. Teratoma assays were performed as described previously (24). Briefly, aliquots of 10⁶ exponentially growing NTERA-wt and NTERA-sh cells were injected subcutaneously in the flank of the neck of each mouse. Two experiments of two mice were done for each cell line. After 7–8 weeks, the mice were killed and the tumors extracted, fixed in buffered formalin, embedded in paraffin and sectioned. Sections were stained with antibodies for the three germinal layers: endoderm (pan-cytokeratins and α-fetoprotein), mesoderm (α-smooth muscle actin) and ectoderm (acidic fibrilar acidic protein). Immunohistochemistry sections were analyzed blinded by eight different investigators and the intensity of staining estimated on a 0–10 scale.

Transient transfection

**NANOG** and **OCT4 Alu** transposons, DROSHA and DGCR8 dominant negative mutants (25) and **NANOG** 3′UTR plasmids were purified and transiently transfected in NTERA-wt or NTERA-sh cells using FuGene 6 (Promega). Aliquots of 1 μg DNA were used for the transfections essentially as described (14). Transfection continued for 48 h in experiments analyzing the expression of **NANOG** and **OCT4 Alus** or for 36 h in luciferase activity assays. H9 cells were transiently transfected with the AHR sh-RNA sequence indicated above using the Human Stem Cell Nucleofector kit in an Amaxa equipment (A-23 program) following the recommendations of the manufacturer.

Luciferase activity assays

**NANOG** 3′UTR constructs were amplified by RT-PCR from cellular RNA using the oligonucleotides indicated in Supplementary Table S1. The amplified 3′UTR was cloned in the XhoI-KpnI sites of the pGL2-Basic vector containing the firefly luciferase gene. This 3′UTR sequence was further modified by PCR to delete the homology region for **NANOG Alu-T1**, **Alu-T2** or both. 3′UTR constructs were co-transfected with the pRLTK (renilla luciferase) vector to normalize luciferase activity. Firefly and renilla luciferase activities were measured after 36 h using the Dual Glow kit (Promega) in a Varioskan Flash luminometer (Thermo).

Nuclear and cytosolic cell extracts and immunoblotting

Sodium dodecyl sulphate-polyacrylamide gel electrophoresis and immunoblotting were performed using NTERA-wt and NTERA-sh cell extracts as indicated (21). Nuclear and cytosolic cellular extracts were prepared from both cell lines as described (26).

Immunofluorescence

Immunofluorescence was done in an Olympus FV1000 confocal microscopy on cultures fixed for 20 min at room temperature in 4% paraformaldehyde (Polysciences Inc.). Fluorochromes used were Alexa 488 (**NANOG**, **OCT4**, βIII-tubulin) and Alexa 633 (AHR, Tau and GAP43). DAPI was used to stain cell nuclei. Fluorescence distribution was analyzed using the FV10 software (Olympus).
Reverse transcription and real-time RT-qPCR

Total RNA was isolated from NTERA-wt and NTERA-sh cells using the RNeasy kit (Qiagen). Reverse transcription was performed with random hexamers and the iScript™ Reverse Transcription Supermix (Bio-Rad). qPCR was done using SYBR® Select Master Mix (Life Technologies) in a Step One equipment (Life Technologies). The oligonucleotides used for RT-qPCR are indicated in Supplementary Table S1. Conventional (no quantitative PCR) was done using Biotaq DNA polymerase in a Bio-Rad T100 equipment.

Cloneogenic assays

Clonogenic assays in 2D were done by plating 103 NTERA-wt and NTERA-sh cells in 35 mm plain tissue culture dishes basically as described (20). After 7 days, the medium was removed and clones washed in PBS, photographed and counted using the ImageJ software (version 1.45S).

Flow cytometry

The analysis and separation of cancer stem-like cells in NTERA-wt and NTERA-sh cultures was done in a MoFlo-XDP equipment (Beckman-Coulter) essentially as described (20). Both cell lines were stained for 10 min at 10°C with antibodies specific for CD133⁺ (CD133-PE), CD44⁺ (CD44-PerCP) and CD29⁺ (CD29-FITC). To-Pro (0.1 µM) was used to discriminate dead cells from the analyses. Cell cycle distribution was determined essentially as described (27).

Sphere formation assay

NTERA-wt and NTERA-sh cells were plated in Nunc ultra-low adherence 24-well plates (Thermo Fisher) at 3 × 10⁴ cells/well and spheres allowed to form for 24 h. Images were acquired with an Olympus Cell-R microscope and sphere number and size quantified using the ImageJ software (version 1.45S).

Transposition assays

For transposition assays, 2 × 10⁴ cells/35 mm² plate were co-transfected with the driver vector pCEPL1SM-NN (containing a LINE-1) and the pBSKS-NeoII vector harboring the NANOG or the OCT4 Alu elements. Transfectants were selected by culturing with 400 µg/ml G418 for 12-13 days.

DNA sequencing

NANOG Alu-T1 and Alu-T2 ncRNA transcripts were sequenced using the oligonucleotides indicated in Supplementary Table S1 in an Applied Biosystems 3130 automatic capillary sequencer.

Statistical analyses

Comparison between experimental conditions was done using GraphPad Prism 6.0 software (GraphPad). Unpaired two-sided Student’s t-test was used to analyze differences between two experimental groups. Differences between three or more experimental conditions were analyzed using ANOVA. Data are shown as mean ± SD. Differences were considered significant at **P < 0.01; ***P < 0.001. Data analyses are indicated in the figure legends.

RESULTS

AHR knockdown leads to an undifferentiated phenotype and blocks RA-induced differentiation of human carcinoma cells

To investigate the implication of AHR in cell differentiation, we have used human pluripotent embryonic teratocarcinoma NTERA2 cells (NTERA) because they can be differentiated in culture by the addition of the endogenous molecule retinoic acid (RA) and because they can be genetically engineered to modulate target gene expression (28–30). Consistent with their pluripotent phenotype, treatment of NTERA cells growing at low cell density with RA produced highly compact clones (Figure 1A) having increased expression of the neuronal markers Tau, βIII-tubulin and GAP43 (Figure 1B and C). This differentiation phenotype of NTERA cells is consistent with that reported in previous studies (29). Notably, total cellular levels of AHR significantly increased during RA-induced differentiation of NTERA cells (Figure 1D), suggesting a role for this receptor in the process. To test this hypothesis, we used retroviral transduction to stably knockdown AHR expression in NTERA cells by means of short hairpin RNAs (NTERA-sh). This experimental approach was very efficient in reducing AHR protein to very low levels in basal and RA-treated cells (Figure 1D). To avoid caveats derived from the use of retrovirally-transduced cells, we decided to work with a pool of NTERA cells after puromycin selection rather than using individual interfered clones. AHR depletion produced spread clones (Figure 1A) with a significant reduction in the expression of neuronal markers Tau, βIII-tubulin and GAP43 (Figure 1B and C). Moreover, AHR depleted cells had an almost negligible response to RA-induced differentiation (Figure 1A–C). The undifferentiated phenotype caused by AhR downmodulation in basal and RA-treated NTERA cells was not associated with significant changes in cell cycle distribution as determined by FACS analysis (Supplementary Figure S1A). Previous reports have shown that AHR depletion could be relevant for mouse ES cell maintenance (31) and for mouse melanoma cells to develop an undifferentiated and metastatic phenotype (20,32). In agreement, AHR-depleted NTERA-sh cells had increased levels of stem-like markers CD44⁺/CD133⁺/CD29⁺, more potential to form spheres under low attachment growing conditions and reduced levels of differentiation-related dehydrogenase activity than wild-type NTERA cells (Supplementary Figure S1B–G). Altogether, these data suggest that AHR is involved in NTERA differentiation.

Differentiation promotes AHR nuclear accumulation and represses pluripotency genes OCT4 and NANOG

We next explored if the differentiation-dependent increase in AHR levels could lead to changes in its intracellular distribution. RA treatment promoted a robust nuclear accumulation of AHR in NTERA-wt cells (Figure 2A and B)

Figure 1. AHR knockdown impairs RA-induced differentiation. (A) NTERA-wt and AHR knockdown NTERA-sh cells were grown in 2D cultures and clone formation analyzed 48 h later in absence (UT) or presence of 1 μM RA. (B) Both cell lines were left untreated (UT) or treated with 1 μM RA for 48 h. Total cell extracts were analyzed for the expression of neuronal markers Tau, βIII-tubulin and GAP43 by immunoblotting. GAPDH was used to normalize protein expression. (C) Tau, βIII-tubulin and GAP43 were also analyzed by immunofluorescence in untreated (UT) or RA-treated NTERA-wt and NTERA-sh cells using a Fluoview FV1000 confocal microscope. (D) NTERA-wt and NTERA-sh cells were left untreated or treated with 1 μM RA for 72 h and total AHR protein levels determined by immunoblotting. β-Actin was used as normalization control. Panels A–D: n = 4 biological replicates. *P < 0.05 and **P < 0.01. Data are shown as mean ± SD.
Figure 2. RA-induced differentiation translocates AHR to the cell nucleus and represses OCT4 and NANO.G. (A) NTERA-wt and NTERA-sh cells were left untreated or treated with 1 μM RA for 72 h. Cytosolic and nuclear extracts were prepared and analyzed for the presence of AHR in each cellular compartment by immunoblotting. GAPDH and RNA polIII were used as cytosolic and nuclear controls, respectively. (B) The intracellular distribution of AHR was determined by immunofluorescence in NTERA-wt and NTERA-sh cells left untreated or treated with 1 μM RA for 72 h in absence (UT) and presence of 1 μM RA. (C and D) NANO and OCT4 mRNAs were quantified by RT-qPCR in both cell lines left untreated (UT) or treated with 1 μM RA for 48 or 72 h. GAPDH mRNA was used to normalize gene expression (ΔΔCt) and 2−ΔΔCt to calculate variations with respect to control or untreated conditions. (E) OCT4 protein levels were analyzed by immunoblotting and normalized by GAPDH. (F) The effects of RA on OCT4 and NANO protein expression were also determined in NTERA-wt and NTERA-sh cells by immunofluorescence. Panels A–D: n = 4 biological replicates; panels E–F: n = 6, two technical replicates in three biological replicates. *P < 0.05 and **P < 0.01. Data are shown as mean ± SD.
but had a negligible effect in AHR-interfered cells, which maintained their very low basal levels of AHR (Figure 2A and B). To further analyze the implication of AHR in RA-induced differentiation, we performed long-term RA treatment (e.g. 4 weeks) of NTERA-sh cells. Under persistent differentiating conditions, AHR protein levels gradually increased and NTERA-sh cells acquired wild-type-like clone morphology and the expression of neuronal markers Tau and βIII-tubulin (Supplementary Figure S2A and B). Restoration of a differentiation response in NTERA-sh cells by AHR re-expression supports a role for AHR in preserving teratocarcinoma cell differentiation.

To address that hypothesis, we first analyzed whether the effects of AHR on cell differentiation were mediated by changes in the expression of pluripotency genes. RA treatment of NTERA-wt cells produced a large reduction in mRNA levels of the stemness-related genes NANOG and OCT4 (Figure 2C and D, and Supplementary Figure S3A and B) without a significant effect on SOX2 and NOTCH1 (Supplementary Figure S3C and D). In agreement, protein levels of OCT4 and NANOG were also reduced by RA treatment in NTERA-wt cells as determined by fluorescence confocal microscopy and western-blot analyses (Figure 2E and F). In contrast, NTERA-sh cells had increased levels of OCT4 and NANOG mRNA and protein under basal conditions and, importantly, they were less sensitive to RA treatment (Figure 2C–F). To further support the implication of AHR in OCT4 and NANOG repression, we did treatments with the AHR ligand 6-formylindolol[3,2-b]carbazole (FICZ) in absence of RA (33). FICZ decreased OCT4 and NANOG mRNA levels in untreated NTERA-wt cells (Supplementary Figure S3E and F) but not in NTERA-sh cells (Supplementary Figure S3G and H), suggesting that AHR is involved in the mechanism that regulates OCT4 and NANOG during the differentiation of human carcinoma cells.

Pluripotency genes contain Alu retrotransposons in their upstream regulatory regions that are transcribed upon differentiation

Bioinformatic analyses have identified thousands of potential AHR binding sites (XRE; Xenobiotic Responsive Element) dispersed throughout the Human Genome (14,16,34). However, their functional and biological significance remains to be determined. Using a previously developed Perl script (14), we have identified XRE sites for AHR binding that happen to be included within repetitive sequences that are located in the promoter of pluripotency genes NANOG, OCT4, SOX2 and KLF4. These repetitive elements have been assigned to the family of Alu-S retrotransposons having an XRE / E-box motif similar to that found in murine SINE-B1 repeats (14) (Figure 3A). Alu elements in the NANOG, SOX2 and KLF4 promoters belong to the 5x subfamily whereas the OCT4-Alu is included in the Sq subfamily of human retroelements (Figure 3A). Importantly, these Alus contain A and B boxes and a putative transcription start site for RNA polymerase III (RNA polIII)-dependent transcription and at least one XRE site for AHR binding (Figure 3B, NANOG and OCT4 Alus are shown). This sequence analysis suggested that these Alus could control the expression of their hosting genes through the transcription factor AHR.

We considered the possibility that transcription of the Alu elements present in the promoter of these pluripotency genes could repress their expression to induce differentiation of carcinoma cells. We focus our study on NANOG and OCT4 since they were the genes more significantly affected by RA treatment in NTERA cells in an AHR-dependent manner (see Supplementary Figure S3). Transcription of Alu retrotransposons was analyzed by deep sequencing of small RNA molecules (RNAseq) isolated from NTERA-wt and NTERA-sh cells under basal and RA-differentiated conditions. Raw RNAseq data have been deposited in the Sequence Read Archive (SRA, NCBI) with submission number PRJNA301447. Previous work has shown that RNAseq is a useful tool to identify transcripts produced from individual Alu elements by RNA polIII (35). To avoid sequencing mRNAs and long non-coding RNA transcripts, libraries were made using size-selected small RNA populations (Supplementary Figure S4A, ncRNA). Bioinformatic alignment of the RNAseq data against OCT4 and NANOG Alus identified transcripts produced from the OCT4-Alu (hereafter OCT4-Alu-T) and NANOG-Alu (hereafter NANOG-Alu-T1 and Alu-T2) under basal conditions in NTERA-wt cells (Figure 3C). Notably, the level of those transcripts was significantly increased by RA-induced differentiation in NTERA-wt cells but not in NTERA-sh cells (Figure 3C). Importantly, the level of N4N4O and OCT4 Alu-derived transcripts was significantly higher than the predicted background level between AluSq and AluSx elements (Figure 3C, black dotted line). We have previously found that transcription of an AhR-regulated murine SINE-B1 retrotransposon takes place mainly in the antisense orientation (15). Similarly, the analysis of RNAseq data indicated that both OCT4 and NANOG Alus were preferentially transcribed in the antisense direction, despite the presence of sense transcripts produced from the NANOG Alu (Figure 3D).

AHR and RNA polIII are involved in OCT4 and NANOG repression in differentiated carcinoma cells

Most human Alus are transcribed by RNA polIII, which can be inhibited by the molecules tagetitoxin (TGT) (36) and WYE132 (37). TGT treatment significantly blocked the repression of NANOG and OCT4 produced by RA at the mRNA level in NTERA-wt cells (Figure 4A). Control experiments showed that TGT was a potent RNA polIII inhibitor in NTERA cells able to efficiently inhibit Leu tRNA expression (Figure 4B). In addition, NTERA-wt cells had a similar response to WYE132, which could also efficiently block the repression of N4N4O and OCT4 by RA (Supplementary Figure S4B). WYE132 was an RNA polIII inhibitor in NTERA cells since it could repress Leu tRNA expression (Supplementary Figure S4C) without significantly affecting NANOG and OCT4 mRNAs in the absence of RA treatment (e.g. WYE132 did not inhibit RNA polII) (Supplementary Figure S4D). Thus, NANOG and OCT4 repression in differentiated NTERA cells appears to require AHR upregulation and the transcriptional activation of their Alu elements by RNA polIII. In agreement with such possibil-
Figure 3. Human OCT4 and NANOG genes contain Alu retrotransposons in their promoters harboring AHR and RNA polIII binding sites. (A) Sequence analysis was used to determine the distribution of Alu retrotransposons among the Alu-Sx (NANOG, SOX2 and KLF4) or Alu-Sq (OCT4) subfamilies. (B) The full-length sequences of NANOG-Alu-Sx and OCT4-Alu-Sq were analyzed for the presence of XRE sites for AHR binding and A and B boxes for RNA polIII-dependent transcription. A DR2 box was also found in the NANOG-Alu. OCT4 and NANOG Alus contained the putative +1 adenosine nucleotide for RNA polIII transcription. The algorithm used for the bioinformatics analysis has been described (14). (C) Non-coding RNA libraries were prepared and sequenced as indicated in Supplementary Figure S4A. The bulk of reads obtained from each library was aligned with the sequences of OCT4 and NANOG Alus. Transcripts frequently expressed in NTERA-wt (blue) or NTERA-sh (red) in absence (UT) or presence of 1 μM RA appeared as peaks overlapping the sequence of each Alu element. The baseline expression of human Alu-Sq and Alu-Sx transposons is indicated by a horizontal black dotted line. (D) The sense (red line) and antisense (black line) orientation of the transcripts produced from the NANOG and OCT4 Alus was determined from the RNAseq data. Transcripts showing complementary for either the sense or the antisense strand were plotted against the full-length Alus to produce peaks corresponding to the NANOG Alu-T1 and Alu-T2 and OCT4 Alu-T molecules.
Figure 4. **OCT4** and **NANOG** Alus bind AHR, are transcribed by RNA polIII in differentiated cells. (A) **NANOG** and **OCT4** mRNA levels were quantified in NTERA-wt cells left untreated (UT) or treated with 1 mM RA or with 1 mM RA + 10 mM tagetitoxin (TGT) for 48 h. (B) Leu tRNA levels were measured in NTERA-wt cells untreated (UT) or treated with 1 mM RA + 10 mM TGT for 48 h. (C) Chromatin immunoprecipitation (ChIP) for AHR and **OCT4** binding to the **NANOG** and **OCT4** Alus was done in NTERA-wt cells left untreated (UT) or treated with 1 μM RA for 48 h. For specificity, one primer for the PCR reaction to amplify each Alu was located in a unique genomic sequence flanking the transposon (see Supplementary Table S1). Double ChIPs involved a first immunoprecipitation with anti-**OCT4** followed by a second immunoprecipitation with anti-AHR antibody. NTERA-sh cells were used as negative controls. Input DNAs and immunoprecipitations done in absence of specific antibodies were also performed. (D and E) **NANOG** and **OCT4** Alu-derived transcripts were quantified by RT-quantitative PCR (qPCR) in absence (UT) or presence of 1 μM RA using the conditions described in Supplementary Figure S5A-C. (F) Basal levels of **NANOG** Alu-T1 and Alu-T2 and **OCT4** Alu-T were measured in NTERA-wt cells by RT-qPCR. **GAPDH** mRNA was used to normalize gene expression (ΔCt) and 2^−ΔΔCt to calculate variations with respect to control or untreated conditions. Panels A and B: n = 6, two technical replicates in three biological replicates; panel C: n = 4 biological replicates; panels D-F: n = 4, two technical replicates in two biological replicates. *P < 0.05, **P < 0.01 and ***P < 0.001. Data are shown as mean ± SD.

As AHR is a transcription factor, we next tested if the repression of **NANOG** and **OCT4** required AHR binding to the XRE sequences present in their upstream Alu elements (Figure 3B). ChIP experiments in NTERA-wt cells using genomic region-specific primers showed that, under basal conditions, AHR was present in the **OCT4** Alu but not in the **NANOG** Alu (Figure 4C). Upon RA-induced differentiation, AHR was recruited to the **NANOG-Alu** and released from the **OCT4 Alu** (Figure 4C). Interestingly, sequential double AHR + **OCT4** ChIP revealed that **OCT4** and AHR were recruited together to the **NANOG-Alu** in RA-differentiated NTERA-wt cells (Figure 4C). A negligible interaction of AHR with the **NANOG-Alu** (Figure 4C) or the **OCT4 Alu** (not shown) was found in basal or RA-treated NTERA-sh cells. These results indicate that AHR binds to the **NANOG Alu** in differentiated cells and that AHR could be part of a transcriptional complex that also includes OCT4.

The **NANOG** and **OCT4** Alu elements are transcribed during differentiation

Based on the apparent prevalence of defined transcripts in differentiated NTERA cells (Figure 3C), we next used the
RNAseq data to assess which amplicons may be used to detect the expression of *NANOG* and *OCT4* Alus. Considering the repetitive nature of *Alu* elements, we first interrogated the Human Genome for the presence of instances identical to either *NANOG* or *OCT4* Alus. At the current sequence resolution of the Human Genome, we did not find any other *Alu* elements fully homologous to those present in *NANOG* and *OCT4* promoters. Based on this finding, we then decided to map putative sites in the Human Genome having full sequence identity with the transcripts produced from the *NANOG* and *OCT4* Alus. This approach would give us an accurate estimation on to what extent those transcripts were produced from the *NANOG* and *OCT4* Alus (since both elements appear at unique locations in the Human Genome) and not from other related retrotransposons. Interestingly, the amplicons predicted by RNA-seq analysis (Figure 3C) were also unique in the Human Genome for the case of *NANOG* *Alu*-T1 and *Alu*-T2 in chromosome 12 (Figure 5, full length hit and asterisks) and *OCT4* *Alu*-T in chromosome 6 (Figure 5, full length hit and asterisks). Moreover, NCBI Blast analysis predicted partial sequence similarity with only 10 and 21 more *Alu* elements in the case of *OCT4*, respectively (Figure 5). Based on these results, transcript-specific RT-qPCR assays were designed to quantify the expression of *OCT4*- and *NANOG*- *Alu*-derived transcripts (Supplementary Figure S5A and B). Amplicons were selected to cover the T1 (yellow) and T2 (red) transcripts from the *NANOG* *Alu* and the T (cyan) transcript from the *OCT4* *Alu* (Supplementary Figure S5C). Capillary fluorescent dideoxi-sequencing was used to confirm the identity of the amplicons obtained by qPCR (Supplementary Figure S5D). The results obtained indicated that *NANOG* *Alu*-T1 and *Alu*-T2 (Figure 4D) and *OCT4* *Alu*-T (Figure 4E) were significantly upregulated by RA in NTERA-wt but not in NTERA-sh cells. In addition, the *OCT4* *Alu*-T was expressed at much lower levels than the *NANOG* *Alu*-T1 and *Alu*-T2 under basal cell conditions (Figure 4F). Moreover, treatment with TGT or WVE132 reduced the levels of *OCT4* *Alu*-T, *NANOG* *Alu*-T1 and *Alu*-T2 transcripts in untreated NTERA-wt cells, further supporting the implication of RNA polIII in the transcription of *OCT4* and *NANOG* *Alus* (Supplementary Figure S5E).

We next investigated the role of AHR in *OCT4* and *NANOG* transcription. As a first approach, we treated NTERA cells with the AHR non-toxic ligand FICZ. AHR activation by FICZ increased the levels of *NANOG* *Alu*-T1 and *Alu*-T2 and *OCT4* *Alu*-T transcripts in NTERA-wt cells (Supplementary Figure S6A–C), supporting a role for this receptor in the regulation of these non-coding RNA transcripts. The effect was transient and each *Alu*-derived transcript returned to its basal level 24 h after FICZ treatment (Supplementary Figure S6A–C). Consistently, ChIP experiments showed that FICZ was able to induce AHR binding to the *NANOG* *Alu* in NTERA-wt cells to a level similar to that produced by RA treatment (Supplementary Figure S6D). Surprisingly, FICZ was also a potent inducer of *OCT4* binding to the *NANOG* *Alu*, reaching levels significantly higher than those obtained by RA (Supplementary Figure S6D).

We then decided to rescue AHR expression in AHR knockdown cells to investigate if restoring AHR could increase the content in *Alu*-derived transcripts while reducing the levels of *NANOG* and *OCT4* mRNAs (Figure 6). To do that, NTERA-wt cells were retrovirally engineered to stably express an shRNA molecule targeting the 3′UTR of the human AHR (NTERA-sh 3′UTR); this approach allowed transfection of a human AHR coding sequence not targeted by the 3′UTR shRNA molecule (Supplementary Table S1). NTERA-sh 3′UTR cells had an almost complete loss of AHR expression, increased basal protein levels of *OCT4* and very low expression of neuronal markers βIII-tubulin and Tau, consistent with the results presented above (Figure 6A). Similar data were obtained using two different shRNA molecules and two different pools of NTERA cells, excluding potential off-target effects in the observed phenotype. Restoring AHR expression in NTERA-sh 3′UTR cells by transient transfection mimicked the effects induced by RA with respect to the expression of *OCT4* and the neuronal differentiation markers βIII-tubulin and Tau (Figure 6A). Control experiments showed that AHR produced from the transfected plasmid was transcriptionally active in NTERA-sh 3′UTR cells based on the induction of its canonical target gene *CYP1A1* (Figure 6B). In agreement with data shown above, rescuing AHR expression, in absence of RA treatment, increased the levels of *NANOG* *Alu*-T1 and *Alu*-T2 (Figure 6C and D) and repressed the mRNA expression of *OCT4* and *NANOG* (Figure 6E and F) in NTERA-sh 3′UTR cells. Moreover, transient transfection of AHR in wild-type NTERA cells further increased the levels of *NANOG* *Alu*-derived transcripts (Figure 6C and D) and strongly repressed *OCT4* and *NANOG* mRNAs (Figure 6E and F). Altogether, these results provide evidences that transcripts produced from *Alu* elements located upstream of *NANOG* and *OCT4* might be involved in controlling the expression of these genes during differentiation of carcinoma cells.

**Transcription of the *NANOG* *Alu* from its RNA polIII promoter represses *OCT4* and *NANOG* in absence of RA**

We next considered the possibility that the transcripts produced from the *NANOG* and *OCT4* *Alus* could regulate the expression of their respective genes in absence of RA. To do that, each *Alu* element was cloned in a promoter-less plasmid (pUC57) to allow transcription from its internal RNA polIII promoter (Figure 3B, A and B boxes). We first confirmed by RT-qPCR that the *OCT4* and *NANOG*- *Alu*-constructs specifically expressed the corresponding *Alu*-T and *Alu*-T1 and *Alu*-T2 transcripts, respectively (Figure 7A). Remarkably, whereas transfection of the *OCT4*- *Alu* did not repress *OCT4* nor *NANOG* in NTERA-wt cells, expression of the *NANOG*- *Alu* efficiently repressed the mRNA levels of both genes (Figure 7B and C), suggesting that transcripts produced from the *NANOG* *Alu* could potentially act in cis and trans to regulate *NANOG* and *OCT4* expression. To provide additional support for the role of *NANOG* *Alu*-derived transcripts in regulating *OCT4* and *NANOG*, we have introduced site-specific mutations in the XRE sites and in the A and B boxes for AHR and RNA polIII binding, respectively (Figure 7D). The selected mutations...
Figure 5. Full length NANOG and OCT4 Alu-derived transcripts map to their specific genomic locations. The complete sequence of the NANOG Alu-T1 and Alu-T2 and OCT4 Alu-T transcripts were aligned against the Human Genome to identify their fully homologous locations in each chromosome (100% sequence homology, full length hit, asterisk). Full length NANOG Alu-T1 and Alu-T2 map to chromosome 12 whereas full length OCT4 Alu-T map to chromosome 6. A similar analysis was done reducing the percentage of sequence homology of each Alu-derived transcript and the Human Genome (partial hit).
were highly effective in blocking \textit{NANOG} \textit{Alu}-derived transcripts from its internal RNA polIII promoter since the levels of \textit{NANOG} \textit{Alu}-T1 and \textit{Alu}-T2 transcripts were largely decreased in NTERA-wt cells transfected with the promoterless pUC57 vector harboring the XREmut, A + Bmut or XRE/A + Bmut sequences (Figure 7E). Notably, mutation of the XRE sites (XREmut), the A and B boxes (A + Bmut) or both (XREmut/A + B mut) significantly blocked the ability of the \textit{NANOG} \textit{Alu} to repress the mRNA expression of \textit{NANOG} and \textit{OCT4} (Figure 7F and G), confirming that, in the absence of transcription from the \textit{NANOG} \textit{Alu}, the repressive mechanism was markedly impaired. Control experiments excluded the possibility that the effects produced by \textit{NANOG} and \textit{OCT4} \textit{Alu} were due to changes in AHR protein levels (Supplementary Figure S6E). In addition, repression of \textit{NANOG} and \textit{OCT4} was AHR dependent since the \textit{NANOG} \textit{Alu} did not have a significant effect in NTERA-sh cells (Supplementary Figure S7A,B). Importantly, AHR- and RNA polIII-dependent transcription of the \textit{NANOG} \textit{Alu} may not have a general impact on pluripotency genes since it did not alter the mRNA levels of the reprogramming factor \textit{SOX2} (39) (Supplementary Figure S7C), the stem-cell factor and \textit{OCT4} regulator \textit{LIN28} (40) (Supplementary Figure S7D) or the pluripotency-inducing factor \textit{NR5A2} (41) (Supplementary Figure S7E). Moreover, even though the 3'UTR regions of \textit{LIN28} and \textit{NR5A2} have significant sequence homology with \textit{NANOG} \textit{Alu}-derived transcripts, their mRNA expression did not change by the \textit{NANOG} \textit{Alu} in NTERA-wt cells, further suggesting a specific effect on the \textit{NANOG} mRNA.

We then decided to analyze if AHR could also modulate \textit{OCT4} and \textit{NANOG} expression and the levels of \textit{NANOG} \textit{Alu}-derived transcripts in non-transformed human ES cells. To do that, we used H9 human ES cells that were either differentiated with RA (42,43) or transiently transfected with a sh-AHR against AHR. H9 cells treated with RA significantly increased their AHR protein levels while markedly decreased their \textit{OCT4} and \textit{NANOG} protein content (Supplementary Figure S8A). Downmodulation of pluripotency factors \textit{OCT4} and \textit{NANOG} likely arose from a significant reduction in their mRNA levels after RA treatment (Supplementary Figure S8B). Importantly, RA treatment largely increased the expression of the \textit{Alu}-T1 and \textit{Alu}-T2 transcripts derived from the \textit{NANOG Alu} while moderately enhanced the amount of the \textit{OCT4} \textit{Alu}-T transcript (Supplementary Figure S8C). Next, H9 cells were transiently transfected with the same sh-AHR construct used to generate the stable NTERA-sh cell line analyzed above. AHR knockdown produced a significant increase in the protein levels of both \textit{OCT4} and \textit{NANOG} as compared to wild-type H9 cells (Supplementary Figure S9A). Consistently, AHR downmodulation upregulated the mRNA levels of both genes in H9 cells (Supplementary Figure S9B). Furthermore, a transient reduction in AHR expression decreased the levels of \textit{NANOG} \textit{Alu}-T1 and \textit{Alu}-T2 transcripts without having a significant effect in the \textit{OCT4} \textit{Alu}-T transcript (Supplementary Figure S9C). Therefore, AHR appears to be able to regulate pluripotency factors \textit{OCT4} and \textit{NANOG} and the expression of transcripts derived from the \textit{NANOG} and \textit{OCT4} \textit{Alus} not only in tumorigenic NTERA cells but also in untransformed H9 human ES cells.

**\textit{NANOG} and \textit{OCT4} repression by \textit{Alu}-derived transcripts involves the miRNA machinery**

Previous studies have shown that the component of the miRNA pathway Microprocessor (DGC8R and DROSHA) can bind transcripts produced from \textit{Alu} and LINE-1 (long interspersed element-1) retrotransposons (25). It is also known that small non-coding RNAs produced by DICER/AGO3 from \textit{Alus} containing the DR2 motif can modulate the proliferation of human stem cells (44). A DR2 element is defined as a 6-bp direct repeat of...
Figure 7. *NANOG* Alu represses *NANOG* and *OCT4* expression in the absence of RA through AHR and RNA polIII. (A) *NANOG* and *OCT4* *Alu* were individually transfected in NTERA-wt cells and the levels of *NANOG* *Alu*-T1, *Alu*-T2 and *OCT4* *Alu*-T transcripts determined by RT-qPCR. (B and C) A promoter-less pUC57 expression vector containing the *NANOG* or *OCT4* *Alu* was transfected in NTERA-wt cells and the mRNA levels of *NANOG* and *OCT4* determined by RT-qPCR. RA was used as positive control. (D) Mutations introduced in the *NANOG* *Alu* to inactivate the XRE sites for AHR binding (cyan), the A (yellow) and B (green) boxes for RNA polIII or a combination of all of them. (E) *NANOG* *Alu* mutant constructs for AHR binding (XREmut), RNA polIII binding (A + Bmut) or both (XRE/A + Bmut) were transfected in NTERA-wt cells and the levels of *NANOG* *Alu*-T1 and *Alu*-T2 quantified by RT-qPCR. (F and G) Mutant constructs for the XRE sites and/or the A and B boxes in the *NANOG* *Alu* were cloned in the promoter-less pUC57 plasmid and transfected in NTERA-wt cells. Wild-type *NANOG* *Alu* (wt) and the empty plasmid (EV) were also transfected as controls. The mRNA levels of *NANOG* (F) and *OCT4* (G) were analyzed by RT-qPCR. Wild-type *NANOG* *Alu* (wt) and the empty plasmid (EV) were used as experimental controls. *GAPDH* mRNA was used to normalize gene expression (ΔΔCT) and and 2^−ΔΔCT to calculate variations with respect to control or untreated conditions. Panels A-C; E-G: n = 6, two technical replicates in three biological replicates. *P < 0.05, **P < 0.01 and ***P < 0.001. Data are shown as mean ± SD.
that such process requires miRNA pathway components RISC and Microprocessor.

As mentioned above, the DR2 element appears to be relevant for loading Alu-derived transcripts into the miRNA machinery via AGO2 (44). Since the *NANOG* Alu has a DR2 sequence (Figure 3B), we next tested whether AGO proteins could bind RNA transcripts produced from the *NANOG* and *OCT4* Alus. RIP followed by RT-qPCR showed that *NANOG* Alu-T1 and Alu-T2 and *OCT4* Alu-T transcripts were loaded into the RISC component AGO2, and that loading increased upon differentiation of NTERA-wt cells (Figure 8G). Interestingly, quantitative RNA analysis revealed a very low efficiency of *OCT4* Alu-T loading into AGO2 even under RA-induced differentiation. Loading of *NANOG* Alu-T1 and Alu-T2 transcripts into AGO2 increased with time for at least 48 h after RA treatment (Supplementary Figure S7G).

Certain TEs, including *Alu* retrotransposons, can regulate the generation of miRNAs (e.g. *Let*-7) known to be involved in cellular differentiation (46–48). The analysis of pre-miRNA expression revealed that differentiation reduced the levels of pre-*Let*-7 and increased pre-miR-21, pre-miR-424, pre-miR371 and pre-miR-141 in NTERA-wt but not in NTERA-sh cells (Figure 9A). Interestingly, time course experiments showed that *NANOG* Alu-T1 and Alu-T2 transcripts appeared early during differentiation (e.g. 24 h) whereas those pre-miR molecules reached their maximum levels only after 72 h of RA treatment (Figure 9B), suggesting that transcription of *NANOG* Alu-derived transcripts probably precedes the accumulation of miRNAs that could cooperate to control the differentiation of NTERA cells.

These results lead us to consider that *NANOG* Alu-T1 and Alu-T2 transcripts, loaded into AGO2, could target complementary sequences in the mRNAs of *NANOG* and *OCT4*. Sequence analysis revealed that *NANOG* Alu-T1 and Alu-T2 transcripts were highly homologous to the 3′UTR of *NANOG* and significantly homologous to the 3′UTR of *OCT4* (Supplementary Figure S10A and B). Conditional training analysis for the prediction of RNA secondary structures (CONTRAfold) revealed the optimal conformation of *NANOG* Alu-T1 and Alu-T2 transcripts in their homology regions with the 3′UTR of *NANOG* and *OCT4* (Supplementary Figure S10C and D). Importantly, sequence analyses by RepeatMasker (http://www.repeatmasker.org) revealed that the 3′UTR of *NANOG* harbors an *Alu*Y element homologous to *NANOG* Alu-T1 and Alu-T2 transcripts. However, the 3′UTR of *OCT4* did not
contain *Alu*-derived sequences despite its partial homology with *NANOG* *Alu*-T1 and *Alu*-T2 (Supplementary Figure S10A,B). Other genes involved in differentiation but not regulated by *Alu*-derived transcripts such as *SOX2* and *NOTCH1* did not contain *Alu*-derived sequences in their 3′UTR regions, nor homology with *NANOG* *Alu*-T1 and *Alu*-T2. These findings suggested that the *Alu* present in the 3′UTR region of *NANOG* could interact with *NANOG* *Alu*-derived transcripts to repress *NANOG* expression in differentiated cells.

To analyze if *NANOG* *Alu* could regulate *NANOG* 3′UTR, we cloned the wild-type *NANOG* 3′UTR and its mutant forms lacking the homology regions for *NANOG* *Alu*-T1 and *Alu*-T2 transcripts in a luciferase reporter vector (Supplementary Figure S7H). Transfection of NTERA-wt cells with *NANOG Alu* in absence of RA repressed the activity of the wild-type *NANOG 3′UTR*-Luc construct (Figure 9C). However, the *NANOG Alu* was unable to repress the activity of *NANOG 3′UTR*-Luc mutants lacking the homology regions for *Alu*-T1, *Alu*-T2 or both transcripts (Figure 9C). These data led us to speculate that *NANOG Alu* might regulate *NANOG* and *OCT4* at the mRNA level through the miRNA pathway. Consistent with this hypothesis, bioinformatic studies have suggested that *Alu* sequences located in the 3′UTR region of certain mRNAs can be recognized by miRNAs to regulate target gene expression (49).

**Figure 9.** *NANOG Alu*-derived transcripts could act through a miRNA mechanism to control *NANOG* and *OCT4* repression during differentiation. (A) NTERA-wt and NTERA-sh cells were left untreated (UT) or treated with 1 μM RA for 48 h and the levels of pre-miRs *pre*-Let7, *pre*-miR-21, *pre*-miR-424, *pre*-miR-21, *pre*-miR-141 quantified by RT-qPCR using the oligonucleotides indicated in Supplementary Table S1. (B) Both cell lines were treated with 1 μM RA for the indicated times and the levels of *NANOG Alu*-T1 and *Alu*-T2, *pre*-Let7, *pre*-miR-21 and *pre*-miR-141 determined by RT-qPCR. (C) A *NANOG 3′UTR*-Luc construct and their deletion mutants lacking the homology regions for *NANOG Alu*-T1 (T1 mut), *Alu*-T2 (T2 mut) or both (T1 + T2 mut) were co-transfected with the *NANOG Alu* in NTERA-wt cells. Firefly luciferase activity was measured and normalized by renilla luciferase. Control experiments were done in presence of RA (dashed). (D) NTERA-wt and NTERA-sh cells were injected subcutaneously in immunodeficient Swiss nude mice (Charles River) and the tumors formed collected and analyzed by immunohistochemistry using endodermal (pan-cytokeratins and α-fetoprotein), mesodermal (α-smooth muscle actin) and ectodermal (glial fibrilar acidic protein, GFAP) markers. Immunohistochemistry sections were analyzed blinded by eight independent observers and protein expression quantified in a 0–10 scale. Arrowheads indicate positivity areas used to estimate protein expression. (E) Scheme of the mechanism proposed for the *Alu*-dependent repression of *NANOG* and *OCT4* during carcinoma cell differentiation. *GAPDH* mRNA was used to normalize gene expression (ΔCt) and 2−ΔΔCt to calculate variations with respect to control or untreated conditions. Panels A and B: n = 6, three technical replicates in two biological replicates; panel C: n = 4, two technical replicates in two biological replicates; panel D: n = 4 mice for each cell line. *P < 0.05, **P < 0.01 and ***P < 0.001. Data are shown as mean ± SD.

**AHR knockdown increases the formation of undifferentiated teratomas**

We finally decided to validate AHR as a differentiation factor in human carcinoma cells using an *in vivo* teratoma assay. NTERA-wt and NTERA-sh cells were xenografted in immunocompromised Swiss mice and the teratomas formed analyzed for the presence of the three embryonic layers (39). Using this assay, we found that NTERA-wt cells generated teratomas containing endodermal (pan-cytokeratin and α-fetoprotein), mesodermal (α-smooth muscle actin) and ectodermal (glial fibrilar acidic protein, GFAP) embryonic layers (Figure 9D). Teratomas formed by NTERA-sh cells had significantly higher expression of endodermal, mesodermal and ectodermal markers than NTERA-wt cells (Figure 9D), suggesting that AHR depletion led to a more undifferentiated phenotype and an increased ability to differentiate *in vivo*. It is reasonable to assume that such properties of NTERA-sh cells might be due, at least in part, to their increased expression of pluripotency factors *NANOG* and *OCT4*. Notably, our hypothesis is in agree-
ment with recent studies showing that in vivo reprogramming induces undifferentiated teratomas with increased expression of NANOG (50), and that OCT4 has a major role in reprogramming and in the generation of iPS cells (51).

DISCUSSION

TEs, including LINEs, SINEs and human AluS, are widespread in mammalian genomes and, although their contribution to genome evolution and dynamics is now generally accepted, very limited knowledge is available regarding their contribution to gene regulation and cell function (2,52). Alu retrotransposons derive from the 7SL RNA and are the most abundant TEs in the Human Genome. Among the different Alu lineages known, some ancient genomic AluS contain an active core sequence in terms of retrotransposition ability (17). In addition, it has been suggested that they could be active in somatic tissues as the human brain (53,54). Furthermore, there is an increasing interest in identifying Alu elements whose activation could contribute to human disease, such as that inserted into the ANRIL long non-coding RNA (7).

In this study, we have presented data suggesting that the dioxin receptor AHR regulates the transcription of Alu elements located in the upstream promoter regions of the pluripotency genes NANOG and OCT4, and that Alu-derived transcripts probably repress the expression of both genes in differentiated carcinoma cells possibly through the miRNA pathway. Our study therefore integrates the activation of specific Alu elements in the control of important cellular functions such as differentiation and pluripotency.

Here, we demonstrate that AHR acts as a differentiation factor in human embryonic teratocarcinoma cells because its stable knockdown led to a basal undifferentiated phenotype that made cells unresponsive to RA-induced neuronal differentiation. Furthermore, differentiation increased the cellular levels of AHR and promoted its nuclear translocation, thus suggesting increased transcriptional activity of this receptor. The role of AHR in differentiation is also supported by in vivo experiments, as teratomas produced by AHR-depleted NTERA cells were more undifferentiated than those from AHR-expressing NTERA cells. These results are consistent with the activation of AHR during differentiation of mouse ES cells (31) and with the inhibitory role of mouse AHR in epithelial-to-mesenchymal transition (21) and in melanoma primary tumorigenesis and metastasis (20).

AHR has been extensively characterized as a transcription factor that binds the promoter region of xenobiotic-related and xenobiotic-unrelated genes (55). It is remarkable that AHR can also interact with specific repetitive retrotranspo-son sequences such as the murine B1-SINE to trigger the activation of a genomic insulator eventually able to modulate gene expression (14–16). Binding to TEs appears to be a conserved property of AHR, as we show here that it can also directly binds to human AluS retrotransposons located in the promoter region of the pluripotency genes OCT4 and NANOG. However, a complex interplay between AHR binding and AluS sequences seems to occur, since differentiation recruits AHR to the NANOG Alu while releasing it from the OCT4 Alu. One possible explanation for these results is that the increase in nuclear AHR that takes place during differentiation could shift the receptor from the OCT4 to the NANOG Alu, inducing at the same time the co-recruitment of OCT4 to the NANOG Alu. This would imply that AHR can cooperate with OCT4 in repressing NANOG expression and, thus, integrates this receptor into the proposed ‘pluripotency core regulatory circuitry’ of transcription factors that also includes SOX2 and that governs differentiation in human and mouse ES cells (56–58). In fact, AHR activation by its non-toxic ligand FICZ produced a robust binding of OCT4 to the NANOG Alu in absence of RA.

Notably, the repression of Alu-containing OCT4 and NANOG genes in differentiated NTERA cells is AHR dependent based on the lack of response shown by AHR-depleted cells, and appears to be specific since it does not take place in other examined Alu-containing pluripotency genes such as SOX2, NOTCH1, MYC and KLF4. Importantly, the repression process seemed to require the synthesis of non-coding RNA molecules since blockade of RNA polymerase III restored OCT4 and NANOG expression in an AHR-dependent manner. Notably, and as RNA polIII may transcribe 7SL RNAs and other non-coding sequences (59), we speculated that transcripts derived from the NANOG and OCT4 Alus could participate in the repression of both genes. Indeed, RNAseq and RT-qPCR experiments revealed that differentiation of NTERA cells increased the levels of transcripts derived from NANOG and OCT4 Alus, being predominant those synthesized from the NANOG Alu. In agreement with the gene expression data, transcription of these Alu elements is AHR dependent because their levels are upregulated by AHR knockdown in RA differentiated cells and increased by AHR rescue in RA-untransformed cells. Such AHR-dependent mechanism regulating the expression of OCT4, NANOG and NANOG Alu-derived transcripts may be operative in non-transformed human cells since RA treatment and AHR knockdown reproduced in human H9 ES cells the same regulatory pattern to that observed in embryonic carcinoma NTERA cells. One important observation is that the Alu retrotransposons located in NANOG and OCT4 promoters have a high degree of sequence specificity, not been detectable at additional locations in the Human Genome. This property most likely determines that NANOG- and OCT4 Alu-derived transcripts have full sequence identity and full homology only to the chromosomal regions where their respective Alus are located and from which they are transcribed. Another relevant finding from this study is that expression of these Alu elements, by themselves, could regulate NANOG and OCT4 during carcinoma cell differentiation. We consider this result relevant since a deep transcriptome profiling study has shown that non-coding transcripts derived from long terminal repeats can associate to enhancers of pluripotency related genes (18,60,61). Nevertheless, the effects of NANOG and OCT4 Alus in gene repression appear to have a significant degree of specificity regarding not only to the Alu element but also their target genes. Indeed, expression of NANOG Alu from its endogenous promoter repressed both NANOG and OCT4 mRNAs in absence of a differentiation stimulus. In contrast, expression of OCT4 Alu does not significantly affect either
gene under the same experimental conditions. Additionally, whereas \textit{NANOG}\textsubscript{Alu} most likely regulates the \textit{NANOG} gene in \textit{cis}, its effects on the \textit{OCT4} gene could be done in \textit{trans}. Such mechanism would resemble the one recently proposed for the long non-coding RNA ANRIL that regulates atherosclerosis genes in \textit{trans} by means of an \textit{Alu} element present in its sequence (7). Alternatively, the \textit{NANOG} \textit{Alu} could act on \textit{OCT4} by an indirect mechanism involving signaling through \textit{NANOG} itself. Nonetheless, this possible functional connection between \textit{NANOG} and \textit{OCT4} again supports the ‘pluripotency core regulatory circuitry’ mentioned above that would also include AHR (56–58). Finally, expression of \textit{NANOG} \textit{Alu} efficiently repressed \textit{NANOG} and \textit{OCT4} but not other genes containing similar \textit{Alu} elements such as \textit{SOX2}, \textit{LIN28} and \textit{NR5A2}. Although we cannot rule out at present time that other \textit{Alu} loci dispersed in the Human Genome might contribute to the generation of additional ncRNAs, when combined with the ChiP data, our results strongly support that \textit{NANOG} and \textit{OCT4} \textit{Alu}-derived transcripts may be involved in the regulation of pluripotency.

Interestingly, recent studies in rice have shown that the plant DICER homolog DICER-like 3 can use TE RNAs to generate small interfering RNAs (siRNA) with the potential to change some agricultural traits of the plant (62). Similarly, in human cells, the Microprocessor can interact and process transcripts derived from \textit{Alus} and \textit{LINE}-1 to regulate retrotransposition, although it is not clear whether processed RNAs give rise to functional miRNAs that might help to regulate the expression of cellular genes (25). Notably, we found that chemical inhibition of the miRNA pathway by ATA not only rescued \textit{OCT4} and \textit{NANOG} repression in differentiated cells but also increased the expression of \textit{NANOG} \textit{Alu}-T1 and \textit{Alu}-T2 and \textit{OCT4} \textit{Alu}-T transcripts. Furthermore, dominant negative forms of Microprocessor components DROSHA and DGCR8 also blocked the repression of \textit{NANOG} and \textit{OCT4} induced by the \textit{NANOG} \textit{Alu} in absence of RA. Intriguingly, the interaction between the miRNA pathway and transcripts derived from retrotransposons could be functionally conserved since inhibition of the Microprocessor also increased the levels of LINE1-derived transcripts and the expression of their target genes (25).

Although the mechanism by which \textit{NANOG} and \textit{OCT4} \textit{Alu}-derived transcripts may regulate gene expression require further experimentation, we speculate that it probably involves the interaction of \textit{Alu}-derived transcripts with the 3’UTR region of the \textit{NANOG} and \textit{OCT4} mRNAs. Several experiments support this model. First, \textit{NANOG} \textit{Alu}-T1 and \textit{Alu}-T2 and \textit{OCT4} \textit{Alu}-T transcripts are loaded into the RISC component AGO2 and such loading significantly increased with differentiation of NTERA cells. Second, a marked sequence similarity exists between \textit{NANOG} \textit{Alu}-T1 and \textit{Alu}-T2 transcripts and the 3’UTR region of \textit{NANOG} and \textit{OCT4} mRNAs. Third, the \textit{NANOG} \textit{Alu} was able to regulate the 3’UTR of \textit{NANOG} but only if the homology sequences for the \textit{Alu}-T1 and \textit{Alu}-T2 transcripts remained intact. Finally, the levels of miRNAs involved in differentiation such as \textit{Let7} and \textit{miR141} increased with time by RA treatment in NTERA-wt but not in NTERA-sh cells and, interestingly, the expression of \textit{NANOG} \textit{Alu}-T1 and \textit{Alu}-T2 preceded the accumulation of \textit{Let7} and \textit{miR141}. It will be interesting to address the processing and identity of these \textit{Alu}-derived transcripts and the mechanisms by which they probably interact with the 3’UTR regions of both genes. Our working model proposing that the interaction of \textit{Alu} derived-transcripts with 3’UTRs could regulate gene expression through the miRNA pathway has been also considered in previous studies. Indeed, bioinformatic analyses have shown that certain small RNAs have the potential to target \textit{Alu}-complementary sequences located in the 3’UTR of human mRNAs (49). Moreover, the FANTOM4 project has revealed that retrotransposons located in the 5’ end of mouse and human protein coding genes (as is the case for \textit{NANOG} and \textit{OCT4} \textit{Alus}) have the potential to be transcribed into non-coding RNAs. Importantly, this study also indicates that a large fraction of the expressed transcripts that contain retrotransposon complementary sequences in their 3’UTRs have lower expression levels than retrotransposon-free transcripts (63). Accordingly, we have found that the \textit{NANOG} mRNA contains an \textit{Alu}Y element in its 3’UTR that is highly homologous with ncRNAs generated from the \textit{Alu}X located in the \textit{NANOG} promoter. However, the 3’UTR of the \textit{OCT4} mRNA has only partial homology with \textit{NANOG} \textit{Alu}-derived transcripts, suggesting the existence of subtle differences in regulation between both genes.

In sum, we propose that AHR activates the transcription of \textit{Alu} retrotransposons located in the \textit{NANOG} and \textit{OCT4} upstream promoter regions in differentiated human carcinoma cells. Non-coding RNA transcripts produced from these \textit{Alus} could repress \textit{NANOG} and \textit{OCT4} expression probably through the miRNA pathway (Figure 9E). \textit{OCT4} and \textit{NANOG} repression in differentiated cells is AHR and RNA polIII dependent and, notably, the expression of the \textit{NANOG} \textit{Alu} is able to repress both genes in the absence of a differentiating stimuli. This regulatory mechanism could be relevant for normal and tumor cell growth and dissemination. Moreover, the results included in this study reveal how fixed TEs can participate in Human Genome regulation by a novel mechanism involving non-coding RNAs and the miRNA pathway (Figure 9E).

**ACCESSION NUMBERS**

Sequence Read Archive (SRA, NCBI). Submission number PRJNA301447.

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR Online.

**ACKNOWLEDGEMENTS**

Eva Barrasa and the Servicio de Técnicas Aplicadas a las Biociencias (STAB) of the Universidad de Extremadura are greatly acknowledged for their technical support. NTERA2 cells were a generous gift of Dr Esteban Ballestar (IDIBELL, Barcelona, Spain). The anti-Tau antibody was generously provided by Dr Maria J. Lorenzo-Benayas (University of Extremadura, Spain). Plasmids expressing dominant negative forms of DROSHA and DGCR8 were a kind gift of Dr Narry Kim (Seoul National University, Korea).


