

ATR expands embryonic stem cell fate potential in response to replication stress

Sina Atashpaz^{1,*,#}, Sara Samadi Shams^{1,*}, Javier Martin Gonzalez², Endre Sebestyén^{1,±},
Negar Arghavanifard^{1,3}, Andrea Gnocchi^{1,3}, Eliene Albers⁴, Simone Minardi^{1,5}, Giovanni Faga⁶,
Paolo Soffientini¹, Elisa Allievi⁵, Valeria Cancila⁷, Angela Bachi¹, Oscar Fernandez-
Capetillo^{8,9}, Claudio Tripodo⁷, Francesco Ferrari¹, Andrés Joaquin López-Contreras⁴ and
Vincenzo Costanzo^{1,3,#}

¹IFOM-The FIRC Institute of Molecular Oncology, Milan, Italy

²Transgenic Core Facility, University of Copenhagen, Copenhagen, Denmark

³Department of Oncology and Hemato-oncology, University of Milan, Milan, Italy

⁴Center for Chromosome Stability and Center for Healthy Aging, Department of Cellular
and Molecular Medicine, University of Copenhagen, Copenhagen, Denmark

⁵Cogentech, IFOM-The FIRC Institute of Molecular Oncology Milan, Italy

⁶Experimental Therapeutics Program, IFOM-The FIRC Institute of Molecular Oncology
Milan, Italy

⁷Tumor Immunology Unit, Department of Health Sciences, Human Pathology Section,
University of Palermo School of Medicine Palermo, Italy

⁸Spanish National Cancer Research Center, Madrid, Spain

⁹Science for Life Laboratory, Division of Genome Biology, Department of Medical
Biochemistry and Biophysics, Karolinska Institute, Stockholm, Sweden

* These authors contributed equally to this work and are listed in alphabetical order.

± Present Address: 1st Department of Pathology and Experimental Cancer Research, Semmelweis
University, Budapest, Hungary

Correspondence should be addressed to: satashpaz@gmail.com and vincenzo.costanzo@ifom.eu

Summary

Unrepaired DNA damage during embryonic development can be potentially inherited by a large population of cells. However, the quality control mechanisms that minimize the contribution of damaged cells to developing embryos remain poorly understood.

Here, we uncovered an ATR- and CHK1-mediated transcriptional response to replication stress (RS) in ESCs that induces genes expressed in totipotent two-cell (2C) stage embryos and 2C-like cells. This response is mediated by *Dux*, a multicopy retrogene defining the cleavage-specific transcriptional program in placental mammals. In response to RS, DUX triggers the transcription of 2C-like markers such as murine endogenous retrovirus-like elements (MERVL) and *Zscan4*. This response can also be elicited by ETAA1-mediated ATR activation in the absence of RS. ATR-mediated activation of DUX requires GSRF1 dependent post-transcriptional regulation of *Dux* mRNA. Strikingly, activation of ATR expands ESCs fate potential by extending their contribution to both embryonic and extra-embryonic tissues. These findings define a novel ATR dependent pathway involved in maintaining genome stability in developing embryos by controlling ESCs fate in response to RS.

Keywords: ATR, Embryonic stem cell, Replication stress, 2C-like cells, DNA damage response pathways, ETAA1, Embryonic development, *Dux*

Background

ESCs are characterized by self-renewal and the ability to propagate for several cycles *in vitro* and *in vivo*¹. Even if ESCs exhibit several markers of RS², they are able to maintain genome integrity more efficiently than differentiated cells¹. The mechanisms underlying such distinctive feature are largely unknown.

ESC colonies harbor a small transient subpopulation of cells (2C-like cells) with functional and transcriptional features of totipotent 2C-stage embryos³⁻⁵. Transition to 2C-like cells has been shown to promote maintenance of genome integrity and survival of ESCs in long-term culture⁶⁻⁸. In addition, several studies have demonstrated that transition to the 2C-like state confers expanded developmental potential to ESCs, making them capable of contributing to both embryonic and extra-embryonic tissues (also referred to as bidirectional cell fate potential)³⁻⁵. However, the molecular players underlying the transition to the 2C-like state in ESC culture and its possible physiological relevance *in vivo* in maintaining genome integrity and expanding cell fate potential in a developing embryo are not fully understood.

Here we provide several lines of evidence that ATR and CHK1-mediated response to RS triggers the activation of 2C-specific genes in ESCs and mouse embryos. This transition is hampered in ESCs derived from ATR-deficient Seckel and CHK1 haploinsufficient mouse models and following ATR or CHK1 inhibition. Significantly, we show that ETAA1-mediated ATR activation is sufficient to trigger the formation of 2C-like cells in the absence of RS.

Mechanistically, ATR-induced transition to 2C-like state is mediated by post-transcriptional regulation of the *Dux* gene, which shapes the transcriptional signature of 2C-like cells and totipotent 2C-stage embryos in placental mammals⁹⁻¹¹. ATR-dependent regulation of *Dux* requires the GSRF1 protein, which directly binds to *Dux* mRNA promoting its stability. Importantly, activation of ATR promotes DUX-dependent formation of placental giant trophoblast-like cells (TGCs), which is

hampered in ATR-deficient Seckel ESCs. Consistent with this, unlike *Dux* KO ESCs, ATR activation in WT ESCs lead to expanded cell fate potential *in vivo*, as shown by their ability to contribute to both inner cell mass and extra-embryonic compartments.

Results

RS increases the number of 2C-like cells in ESCs culture and activates the expression of 2C-like genes in mouse embryos

Maintenance of genome stability along with unlimited self-renewal is a unique feature of ESCs¹. To understand how ESCs coordinate these functions, first we asked how ESCs transcriptionally respond to RS at the single cell level. To this end, we performed single cell transcriptional profiling¹² of E14 mouse ESCs cultivated in Leukemia Inhibitory Factor (LIF) plus MEK and GSK inhibitors (2i) upon treatment with aphidicolin (APH), a reversible inhibitor of DNA polymerases that activates ATR by stalling replication forks progression¹³. Unsupervised clustering analysis of CNTL and APH-treated cells¹² identified a distinct subset of cells (Fig.1a, Cluster 4), that was also clearly separated by Principal Component (PC) 1 from the rest of the population (Extended Data Fig. 1a,b). The analysis of differentially expressed genes (DEGs) between cluster 4 and the rest of the population identified a significant enrichment of 2C-specific genes in this cluster, including *Eif1a-like* genes (*Gm5662*, *Gm2022*, *Gm4027*, *Gm2016*, and *Gm8300*), *Tcstv3*, and *Zscan4* genes (*Zscan4a–Zscan4d*), (Fig. 1b,c; Extended Data Fig. 1c,d; Supplementary Table 1,2), the transcription of which has been shown to play a critical role in maintaining ESCs genome stability⁶⁻⁸. Remarkably, we found a statistically significant increase in the number of cells expressing 2C-specific markers upon RS (Fig. 1d). To understand whether the increase in the number of 2C-like cells was a response to RS or it was limited to APH treatment, we exposed *pZscan4*-Emerald ESCs generated by the stable introduction of the Emerald-GFP reporter under the *Zscan4c* promoter⁸ to a range of RS-inducing agents, including APH, hydroxyurea (HU) and ultraviolet light (UV)¹⁴. Fluorescence Activated Cell Sorting (FACS) analysis confirmed a significant increase in the number of Emerald positive (Em⁺) ESCs across all treated conditions in a dose-dependent manner (Fig. 1e). Of note, the increase in the number of Em⁺ ESCs upon short exposure to UV revealed that the continuous presence of the RS-inducing agent was not

necessary for the activation of 2C-like cells (Fig. 1e), suggesting that replication fork stalling induced by UV-mediated DNA lesions was sufficient to trigger this pathway. Next, to uncover the minimum timing required for the activation of ZSCAN4, we performed a time-course experiment in which ESCs were treated with APH and subsequently harvested at various time points. Immunoblot on ESCs revealed a rapid phosphorylation of CHK1 while a significant increase in ZSCAN4 protein took place after 4-8 hours of treatment with APH (Extended Data Fig. 1e). Importantly, although APH, mostly at a high concentration, induced mild cell apoptosis, the majority of Em⁺ cells were not co-stained with the apoptosis marker CASPASE-3, suggesting that the emergence of Em⁺ cells upon APH treatment was not due to the activation of apoptosis (Extended Data Fig. 1f). Moreover, no sign of cellular senescence was detected by β -galactosidase staining upon APH treatment in ESCs (Extended Data Fig. 1g).

Due to the key role of DUX and MERVL-derived long terminal repeat (LTR) elements in shaping the transcriptional signature of 2C-stage embryos and 2C-like cells^{3-5,9-11}, we next asked whether RS-induced expression of the *Zscan4* gene was accompanied by the activation of these retroelements. The RT-qPCR results revealed a linear correlation between concentration of RS inducing agents and the expression of key 2C-like markers, such as *Dux*, MERVL, *Zscan4d*, *Tcstv3*, *Gm12794*, *Gm4340* (Extended Data Fig. 2a-f). These results were further validated using pMERVL-GFP ESCs (ESCs expressing a GFP reporter under the control of a MERVL promoter^{4,9}) (Extended Data Fig. 2g,h). However, induction of RS did not affect the expression of main 2C-specific markers such as MERVL and *Zscan4* (no expression was detected by qPCR) in two mouse embryonic fibroblast (MEF) lines with distinct genetic backgrounds (C57BL/6J and 129P2/OlaHsd) (Extended Data Fig. 2i), suggesting the involvement of alternative repressive mechanisms that suppress the activation of 2C-like transcriptional program in more differentiated cells.

2C-like cells reduce the expression of pluripotency markers at protein but not transcriptional level³⁻⁵. Thus, to understand whether RS-induced cells share such features with 2C-like cells, we

monitored the expression of canonical pluripotency markers upon APH treatment. In agreement with previous findings^{3,4}, we detected downregulation of SOX2, and POU5F1 (also known as OCT4) proteins in ZSCAN4⁺/Em⁺ cells by immunostaining and immunoblot (Fig. 1f; Extended Data Fig. 2j,k). However, no significant alteration in the expression of pluripotency-related genes (e.g., *Nanog*, *Oct4* and *Rex1*) was observed at the transcriptional level (Extended Data Fig. 2l-n). As previously reported⁴, 2C- like cells that were identified by expression of 2C markers, MERVL-GAG and ZSCAN4, and the absence of OCT4 protein, were found to lack chromocenters (Extended Data Fig. 2j).

The ZSCAN4⁺/MERVL⁺ cells were reported to be present in all phases of the cell cycle albeit with higher percentage in G2/M phase¹⁵. Of note, the cell cycle analysis on p*Zscan4*-Emerald ESCs confirmed that APH treatment neither at low (0.3 μ M) nor at high (6 μ M) concentration could increase the G2/M population in culture (Extended Data Fig. 3a). Similar results were obtained upon exposure of ESCs to UV (Extended Data Fig. 3b), indicating that the increase in the Em⁺ population upon APH and UV treatment was not due to cell cycle arrest in G2/M. This is consistent with a previous work showing that G2/M arrest by nocodazole is not sufficient to trigger *Zscan4* expression¹⁶. Finally, to evaluate the physiological relevance of these findings *in vivo*, we asked whether transient induction of RS could activate the expression of key 2C-embryo specific genes at the later stages of mouse embryonic development. To this end, morula-stage embryos were treated with APH and subsequently, the synchronized embryos were selected and subjected to RT-qPCR. Strikingly, induction of RS activated the expression of several key 2C-embryo specific markers, including *Zscan4c*, *Dux*, MERVL and *Gm4981* in mouse embryos (Fig. 1g-j; Extended Data Fig. 3c-h). Importantly, APH treatment did not disrupt the blastocyst formation or expression of ICM and TE markers (Extended Data Fig. 3i-k).

Overall, these findings indicate that RS leads to the activation of 2C-embryo specific genes in ESCs and morula-stage mouse embryos.

ATR and CHK1-mediated RS response triggers the activation of key 2C-like genes in ESCs

Next, to gain further insight into the mechanisms through which RS response (RSR) could contribute to the emergence of 2C-like cells in ESCs culture, we asked whether activation of the endogenous DNA damage response pathways (DDR) is responsible for the emergence of 2C-like cells under normal culture condition. To this end, we inhibited DDR pathways by treating p*Zscan4*-Emerald ESCs with specific ATM and ATR inhibitors (ATMi, KU-55933 and ATRi, VE 822, respectively). FACS analysis revealed a slight reduction in the number of Em⁺ cells upon ATR inhibition; however, treatment with ATMi did not significantly reduce the fraction of Em⁺ cells in culture (Fig. 2a; Extended Data Fig. 4a). In addition, FACS analysis revealed a significant enrichment of γ H2AX⁺ and p-CHK1⁺ (indicators of RS) populations within Em⁺ ESCs (Extended Data Fig. 4c,d), suggesting that the transient activation of ZSCAN4 is linked to the presence of endogenous RS in 2C-like cells. Consistent with these results, inhibition of ATR but not ATM activity could robustly revert the expression of ZSCAN4 and *Dux*, which was induced in response to various RS-inducing agents (e.g., UV, HU and APH) (Fig. 2a, b; Extended Data Fig. 4a,b,e,f). These results were further validated on E14 and R1 ESC lines through RT-qPCR assay for *Zscan4d* and *MERV1* (Fig. 2c; Extended Data Fig. 4g). Similar results were obtained in ATM KO ESCs confirming that activation of APH-induced 2C-like genes in ESC culture is not ATM dependent (Extended Data Fig. 4h-j). Of note, inhibition of p38 MAPK signaling pathway did not lead to any decrease in the level of APH-induced 2C-gene expression. This result suggests that unlike the role of p38 inhibitors in suppressing DUX4 in cellular and animal models of facioscapulohumeral muscular dystrophy¹⁷, this pathway does not play a major role in the regulation of *Dux* upon RS induction in mouse ESC (Extended Data Fig. 4k).

In agreement with our findings, single cell transcriptional profiling revealed that suppression of ATR activity by ATRi led to a statistically significant reduction in the number of cells that express main 2C-genes in response to APH (e.g., *Zscan4d*⁺, *Gm8300*⁺, *Gm022*⁺, and *Usp14ib*⁺) (Fig. 2d).

However, we did not find any significant change in the expression level of 2C-specific genes at the single cell level across different conditions (i.e., CNTL, APH and APH+ATRi) (Extended Data Fig. 4l-n). Moreover, FACS analysis of *pMERV*L-GFP ESCs did not show any significant increase in the mean intensity of MERVL-GFP signal within MERVL⁺ population between CNTL and APH-treated conditions (Extended Data Fig. 2g,h). These results suggest that the increase in the expression of 2C-genes upon APH treatment is mainly due to the increase in the number of newly generated 2C-like cells and not to the overexpression of 2C-genes in the pre-existing 2C-like cell population.

To further validate these findings, we derived ATR-deficient Seckel (*Atr*^{Sec/Sec}) and haploinsufficient (*Chk1*^{+/-}) ESCs from previously reported mouse models^{18,19} as the complete ablation of ATR or CHK1 causes embryonic lethality in mice²⁰. To this end, ESC lines were established in culture using pre-implantation embryos obtained from crosses between heterozygous mice either for *Atr* Seckel or for *Chk1* KO alleles. On the basis of genotyping results, homozygous *Atr* Seckel and heterozygous *Chk1* ESC lines were characterized for further investigations (Fig. 2e, f; Extended Data Fig. 5a, b). Although immunoblot results confirmed that the levels of ATR and CHK1 proteins were severely reduced in comparison to the wild type (WT) ESCs (Fig. 2e, f), no significant alteration in the expression of key pluripotency genes was observed in *Atr*^{Sec/Sec} or *Chk1*^{+/-} ESCs compared to WT ESCs (Extended Data Fig. 5a, b). Noticeably, similar to the inhibitory impact of ATRi on the expression of 2C-genes, APH treatment on *Atr*^{Sec/Sec} ESCs led to only a mild increase in the expression of 2C-related genes unlike WT ESCs (Fig. 2e, g-k; Extended data Fig. 5c-f). As expected, similar results were obtained with *Chk1*^{+/-} ESCs and upon treatment with CHK1 inhibitor (CHK1i) (Fig. 2f,l; Extended Data Fig. 5g-n). Of note, we did not find any alteration in the level of APH-induced *Dux* upon *Trp53* knockdown (KD) or its complete ablation in *Trp53* KO ESCs, suggesting that activation of 2C-like pathway does not require the known mediator of DDR, P53 (Extended Data Fig. 5o-r).

Overall, these data indicate that the ATR and CHK1-stimulated response to RS regulates the activation of 2C-specific genes.

ATR induces the transcriptional signature of 2C-like cells in ESCs

Next, to understand whether activation of ATR-dependent response could result in a global transcriptional activation of 2C-specific genes, we performed high-throughput transcriptional profiling on three ESC lines, namely E14, R1 and MC1, upon APH treatment. Analysis of differentially expressed genes (DEGs) (FDR < 0.05 and $|\log_2$ fold change| > 1) identified 3074 upregulated genes with more than two-fold change in gene expression upon APH treatment, and only 640 downregulated genes (Extended Data Fig. 6a; Supplementary Table 3), which is in agreement with the general openness of chromatin in 2C-like cells⁴. To understand how many of the identified DEGs overlap with those specifically expressed in 2C-like cells, we compared our list of APH-induced genes with a recently published dataset¹⁵. Through such comparison, we found that a significant fraction of APH-induced retroviral elements and genes overlap with those expressed in 2C-like cells, including MERVL, MT2_Mm, *Dux*, *Eif1a*-like genes (*Gm5662*, *Gm2022*, *Gm4027*, and *Gm8300*), *Zscan4* genes (*Zscan4b* and *Zscan4d*), *Zfp352*, *Zfp750*, *Tdpoz* genes (*Tdpoz1* and *Tdpoz3*), and *Tmem92* (Fig. 3a, b; Extended Data Fig. 6b,l; Supplementary Table 4,5). Considerably, RT-qPCR results confirmed the significant upregulation of main retroviral elements and 2C-specific genes upon APH treatment (Extended Data Fig. 6c-k). Interestingly, the DEG analysis identified a large portion (48%) of APH-induced 2C-specific genes to be transcriptionally repressed upon ATR inhibition (Fig. 3c; Supplementary Table 4).

Recent reports demonstrated that 2C-like cells can be generated through genetic modulation of several factors, including the KD of chromatin assembly factor-1 (CAF-1), KD of KRAB (Krüppel-Associated Box Domain)-Associated Protein 1 (KAP1), KO of microRNA-34 (miR-34) and OE of

DUX^{3,4,10,21}. Thus, to test whether these factors were involved in ATR-induced expression of 2C-related genes, we compared the transcriptome of APH-induced ESCs with those from published datasets. While we found a significant overlap with all datasets, the highest number of overlapping genes was found with the transcriptome of CAF-1 KD ESCs (Extended Data Fig. 6l; Supplementary Table 4), possibly due to the previously reported role of CAF-1 in preventing RS²².

Next, to uncover the specific role of ATR in controlling 2C-genes regulating factors, we focused our analysis on genes whose expression was reverted by ATRi. Remarkably, through such analysis we found that 71% of the genes shared between APH and DUX-induced conditions were rescued by ATRi (Fig. 3d, Supplementary Table 4), suggesting a possible role of DUX in activation of 2C-specific genes through ATR. To validate this finding, we checked the expression of key 2C-related genes in *Dux* KO ESCs after induction of RS. Importantly, although APH treatment, through ATR activation, increased the expression of *Dux* and its downstream targets, such as *Zscan4* and *MERVL* in WT ESCs, it could not induce the expression of key 2C-genes in *Dux* KO ESCs (Extended Data Fig. 7a-c). Next, to understand whether the inability of APH in activating *MERVL* and *Zscan4* is due to a general suppression of these genes in *Dux* KO ESCs or not, we performed a siRNA-mediated KD experiment to partially silence *Dux* in WT ESCs. As expected, upon partial *Dux* KD in CNTL condition, the expression level of *Zscan4* and *MERVL* decreased but was not fully abolished (Extended Data Fig. 7d-f). Consistent with our previous findings, RT-qPCR results confirmed that the activation of ATR through APH treatment increases the expression of 2C-genes in WT ESCs, while APH could only modestly activate *Zscan4* and *MERVL* in *Dux* KD ESCs (Fig. 3e-g). These results were further validated by immunoblot for ZSCAN4 (Fig. 3h) and confirmed the role of ATR in regulating 2C-genes through *Dux* activation.

These findings overall suggest ATR as a potent upstream driver that can induce the expression of 2C-genes in ESCs culture.

ETAA1-mediated activation of ATR elevates the number of 2C-like cells in a RS-free context

Recent evidence showed that ATR kinase can be directly activated by the RPA-binding protein, Ewing's tumor-associated antigen 1 (ETAA1) in the absence of RS^{23,24}. Hence, to confirm that the activation of 2C-specific genes is due to ATR-mediated response and not the physical damage to DNA, we aimed to activate ATR in a DNA damage-free context through overexpression of ETAA1-ATR activating domain (ETAA1-AAD) in ESCs. To this end, p*Zscan4*-Emerald ESCs were infected with two independent lentiviruses generated from two different clones of ETAA1-expressing lentivectors in which ETAA1-AAD was expressed under the control of a doxycycline (Dox)-inducible promoter, and subsequently the cells were selected against puromycin and neomycin for two weeks. As expected, FACS analysis confirmed the activation of the ATR pathway, as shown by an increase in the number of γ H2AX⁺ cells upon Dox administration to ETAA1-AAD-inducible ESCs (Dox-iETAA1 ESCs) compared to ESCs infected with an empty vector (EV) (Extended Data Fig. 8a). However, ETAA1-AAD overexpression did not induce cell apoptosis as shown by CASPASE-3 FACS analysis (Extended Data Fig. 8b). Strikingly, ETAA1-AAD-stimulated activation of ATR was accompanied by a significant increase of the Em⁺ population in ESCs culture and transcriptional activation of *Zscan4*, *Dux* and *MERVL* (Fig. 4a-d; Extended data Fig. 8a, c-f). As expected, this transcriptional activation was fully abolished upon ATR and CHK1 inhibition (Fig. 4a-c; Extended Data Fig. 8f-j).

Importantly, several lines of evidence demonstrated that the γ H2AX positivity caused by the overexpression of ETAA1 is not a consequence of DNA breakage^{23,24}. Consistent with this, ATR activation through ETAA1-AAD expression led to the mild phosphorylation of H2AX and CHK1 but not CHK2 in ESCs (Fig. 4a, c). Moreover, while inhibition of ATR activity reduced the level of γ H2AX upon ETAA1-AAD-mediated activation of ATR, inhibition of ATM activity did not exert any noticeable effect (Fig. 4a, c). Finally, we demonstrated that the activation of canonical 2C-genes upon

ETAA1-AAD-mediated ATR activation is also DUX dependent, as *Dux* KD significantly reduced the ETAA1-AAD-mediated expression of *MERVL* and *ZSCAN4* (Fig. 4d-g; Extended Data Fig. 9e-h). Similar results were obtained upon ETAA1-AAD-mediated ATR activation in *Dux* KO ESCs (Extended Data Fig. 8k-m).

Finally, to further dissect the physiological role of ETAA1 in the transition to 2C-like state, we asked whether its KD through specific siRNA (Extended Data Fig. 9a) could lead to a reduction in the level of 2C-specific genes expressed under basal condition. Of note, the KD of ETAA1 did not lead to a reduction of 2C-genes expression (Extended Data Fig. 9b,c) in line with recent works where ETAA1 was found to function in parallel to TOPBP1 in regulating ATR and maintaining genome stability^{23,24}. Moreover, we found a higher level of ETAA1 in 2C-like cells further confirming that activation of RSR pathways could lead to the expression of 2C genes (Extended Data Fig. 9d).

These results collectively indicate that the DNA damage-independent, ETAA1-stimulated ATR activation in ESCs is sufficient to activate 2C-related genes through DUX.

Candidate-based screening identifies the involvement of the GRSF1 in the post-transcriptional regulation of the *Dux* mRNA downstream of ATR

Next, to identify the mediating factors regulating *Dux* expression in response to ATR activation, we performed a siRNA-based RT-qPCR screening using a library targeting 148 genes (Fig. 5a) and quantified the expression level of *Zscan4*, a *bona fide* DUX downstream target, as the final readout⁹⁻¹¹. To this aim, we took advantage of the MISSION esiRNA technology that provides a heterogeneous mixture of siRNAs targeting the same mRNA sequence, and thus offers a highly specific and effective gene-silencing approach at lower concentration, ensuring a minimal risk of off-target effects²⁵. We refined the list of siRNAs based on the overlap of the following categories: potential substrate of

ATR/CHK1^{26,27}, *Dux* promoter-bound factors based on our *in silico* analysis and previous findings²⁸, and potential activators or suppressors of *Zscan4* (Fig. 5b)²⁹. We included siRNAs targeting *Dux*, *Atr*, and *Zscan4* as positive controls and a siRNA targeting Renilla luciferase as a negative control.

To identify *Dux* activators linked to ATR, we focused on genes whose KD could decrease the APH-induced *Zscan4* expression, while genes whose KD was sufficient to induce *Zscan4* expression were considered as *Dux* suppressors. Through analyzing RT-qPCR results, we identified 49 hits whose downregulation altered *Zscan4* expression at least two-fold. Interestingly, these genes were found to be enriched in 2 main categories: i) genes involved in DNA replication and RSR such as *Vcp*, *Smc1a*, *Rfwd3*, *Rfc2* that further validated our finding on the involvement of RSR pathway in the regulation of *Dux*; and ii) mRNA processing factors (Fig. 5c; Supplementary Table 6), suggesting the possible regulation of *Dux* at the post-transcriptional level.

The significant enrichment of mRNA processing factors in our screening analysis, led us to test whether ATR could directly affect *Dux* mRNA level. To this end, we first checked whether ATRi could alter the level of *Dux* mRNA produced exogenously by a vector that did not contain *Dux* natural promoter. To verify this, we treated *Dux* overexpressing ESCs, generated by the stable integration of a Dox-inducible *Dux* construct into ESCs (Dox-*iDux* ESCs), with ATRi and probed the level of exogenous *Dux* transcript using oligos specifically recognizing this construct. While Dox administration led to a remarkable increase in the level of exogenous *Dux* mRNA, ATR inhibition significantly reduced the Dox-induced *Dux* transcript (Fig. 5d). Similar results were obtained upon ATR KD, however, neither ATRi nor ATR KD could dramatically alter the levels of exogenous Luciferase mRNA induced by Dox administration (Fig. 5e). Finally, to exclude any possible effect of ATR inhibition on endogenous *Dux* gene expression, we generated a Dox-inducible *Dux* OE system using *Dux* KO ESCs in which the endogenous *Dux* genomic region was completely deleted⁹, and obtained a similar result on the expression of endogenous *Dux* mRNA upon ATR inhibition (Extended

Data Fig. 9i). Importantly, *Dux* overexpression did not lead to a significant increase in the level of p-CHK1, ruling out the possibility of DUX activating RSR pathway through the ATR-CHK1 axis (Extended Data Fig. 9j). Overall these results suggest that ATR affects the level of *Dux* mRNA through its post-transcriptional regulation.

Next, to identify the potential *Dux* mRNA-binding factor(s) through which ATR could alter *Dux* mRNA level, we performed RNA-protein pull-down using a synthetic 3' untranslated region of *Dux* mRNA (*Dux* 3'UTR). Mass-spectrometry analysis of the pulled-down proteins revealed several *Dux* 3'UTR binding factors, including RNA binding proteins HNRNPA1, HNRNPA3, HNRNPA2B1, PABPN1, PABPC1, PCBP1, PCBP2 and GRSF1 (Supplementary Table 7). Strikingly, although both *Grsf1* and *Pcbp2* KD could reduce APH-induced *Dux* expression (Fig. 5f,g, Extended Data Fig. 9k), only GRSF1, which belongs to a group of heterogeneous nuclear RNPs bearing the RNA-recognizing domain RRM³⁰, was proven to bind *Dux* 3'UTR by immunoblot (Fig. 5h). These results suggest that GRSF1 is directly binding and likely stabilizing *Dux* mRNA. Importantly, we did not observe any alteration in the level of pCHK1 upon *Grsf1* KD (Fig. 5g), excluding the possibility of *Dux* modulation through induction of RS in GRSF1-deficient cells. Of note, GSRF1 has been found to be phosphorylated in putative ATR and CHK1 sites following cellular stress in mass spectrometry based phospho-proteomic screens³¹, suggesting that direct phosphorylation by ATR/CHK1 might affect its function in *Dux* regulation. Although further studies will be required to explain the specific role of GRSF1 in the post-transcriptional regulation of *Dux* through ATR, these findings shed light on the link between RSR and *Dux* regulation. Furthermore, they support recent observations reporting the involvement of mRNA processing factors in the regulation of 2C-genes²⁹.

ATR-activated ESCs gain expanded developmental potential

Our results so far indicated that ATR-activated ESCs exhibit similar transcriptional profile and characteristics of 2C-like cells. These findings led us to ask whether ATR-activated ESCs also gain expanded developmental potential similar to 2C-like cells, which are able to contribute to both embryonic and extra-embryonic tissues³⁻⁵.

To test this hypothesis, both *Atr*^{Sec/Sec} and *Atr*^{+/+} ESCs were cultivated in the absence or in the presence of APH and were subsequently differentiated *in vitro* toward trophoblast-like stem cells (TSCs) for three days, followed by terminal differentiation to TGCs upon withdrawal of fibroblast growth factor 4 (FGF4) and heparin for additional three days (Fig. 6a)³². The RT-qPCR results for the TGC specific marker, *Prl2c2*, revealed the highest expression in ATR-activated conditions, while *Atr*^{Sec/Sec} cells could not significantly up-regulate this gene in the presence of APH (Extended Data Fig. 10a). Moreover, *Prl2c2* basal level was significantly higher in WT cells compared to non-treated *Atr*^{Sec/Sec} cells (Extended Data Fig. 10a). These results were also consistent with the expression of Placenta-Expressed Transcript 1 protein (PLET1) and the number of TGCs generated in each condition (Fig. 6a,b). Next, to understand whether ATR-induced differentiation to TGCs is also mediated by DUX, we differentiated *Dux* KO and WT ESCs toward trophoblast cells upon ATR activation (Extended Data Fig. 10b). Significantly, ATR activation in *Dux* KO cells could not induce the formation of TGCs unlike WT cells, confirming the critical role of DUX in trophoblast-directed differentiation program upon ATR activation (Extended Data Fig. 10c). These experiments indicate that the transition to 2C-like state is required for ATR-induced TGC differentiation.

Next, to define the cell fate potential of ATR-activated ESCs during embryonic development, we traced the fate of their progenies in chimeric blastocysts. To this end, ten mCherry fluorescent protein-labeled (mCherry-labeled) Dox-iETAA1 ESCs were treated with Dox and subsequently microinjected into each C57BL/6N recipient mouse morulae. Untreated Dox-iETAA1 ESCs were injected in parallel as control. The contribution of ESCs to the inner cell mass (ICM) and the

trophectoderm (TE) layer of the blastocysts was then monitored 48 hours post-injection (Fig. 6c). While both Dox-treated and untreated ESCs contributed to the ICM, strikingly, upon ATR activation, we found a statistically significant (Fisher's exact test p-value: 0.0005) increase in the number of chimeric embryos in which ESC progenies localized to both ICM and TE (Fig. 6d; Extended Data Fig. 10d). Although the basal contribution level of untreated ESCs to the TE could be due to the previously reported effect of $2i^{33}$, overexpression of ETAA1-AAD significantly increased the number of embryos in which mCherry-labelled cells could contribute to the TE. Next, we asked whether activation of ATR through ETAA1-AAD expression is also able to impact on the number of cells contributing to the TE layer of the blastocyst. To this aim, we repeated this experiment and compared the number of cells in the TE layer of embryos injected upon Dox treatment to the one observed in CNTL. Significantly, ATR activation was able to increase about two-fold the average number of cells contributing to the TE layer (Extended Data Fig. 10e). Finally, to understand whether ATR-dependent expanded developmental potential is mediated by DUX, we generated mCherry-labeled ETAA1-AAD Dox-inducible ESCs in a *Dux* KO background. As shown in Extended Data Fig. 8k-m, unlike WT cells that increased the expression of 2C-genes upon Dox treatment, *Dux* KO ESC were not able to activate these genes. Importantly, through the mouse chimera assay we found that activation of ATR was not able to significantly increase the number of mCherry-labeled Dox-iETAA1 *Dux* KO ESCs contributing to TE layer upon Dox treatment, consistent with our *in vitro* results (Extended Data Fig. 10f).

The expanded cell fate potential of RS-induced ESCs could originate from cells with bidirectional potential or a heterogeneous population containing cells that preferentially differentiate into embryonic or extraembryonic lineages³. To distinguish between these two possibilities, we microinjected a single mCherry-labeled Dox-iETAA1 ESC upon Dox treatment into each C57BL/6N recipient morulae to generate chimeric blastocysts. Consistent with our previous findings, a considerable number of ATR-activated single cells were able to contribute to ICM+TE or TE, albeit

with lower efficiency (Fisher's exact test p-value: 0.08) (Fig. 6e). To further validate these observations, we generated post-implantation chimeric embryos by microinjecting ten mCherry-labeled Dox-iETAA1 ESCs into each C57BL/6N morulae that were subsequently transferred to foster mothers. While WT ESCs contributed exclusively to the embryonic tissue (epiblast), ATR-activated cells contributed to both embryonic and extra-embryonic cell lineages in 50% (3 out of 6) of chimeric embryos at embryonic day (E) 7.5 (Fig. 6f,g; Extended Data Fig. 10g; Extended Data Fig. 11a,b). These results indicate that ATR-activated ESCs gain expanded developmental potential, in contrast to ESCs, which are mostly restricted to generate embryonic cell types (Fig. 6h).

These findings overall indicate that ATR-activated ESCs gain expanded developmental potential *in vitro* and *in vivo*, and that this effect is mediated by DUX, whose physiological role in the activation of cleavage-stage transcription program has been reported in previous works⁹⁻¹¹.

Discussion

The physiological relevance of the 2C-like transcriptional program has been limited to 2C-stage embryos, and thus the molecular regulation involved in reactivation of 2C-like cells in ESCs culture and its possible significance *in vivo* at later stages of embryonic development have remained elusive.

Here, we provide evidence that transition to bipotent 2C-like cells in ESCs culture is triggered by activation of ATR, a developmentally essential DDR gene, which plays a crucial role in the maintenance of stem cells both in embryonic and adult tissues^{20,34}. First, we found a significant enrichment of γ H2AX⁺ and p-CHK1⁺ cells within Em⁺ cell population. Second, inhibition of RSR by specific ATR or CHK1 inhibitors, but not p38 inhibitors, could markedly reduce the number of 2C-like cells in ESCs culture. Third, ATR-deficient Seckel and haploinsufficient CHK1 ESCs exposed to RS could not significantly induce 2C-specific genes. Fourth, activation of ATR through ETAA1-AAD overexpression could further increase the population of bipotent 2C-like cells in ESCs culture in the absence of RS. Finally, through a candidate-based screen approach, we unraveled the mechanistic basis of ATR-stimulated transition to bipotent 2C-like cells by showing that this response induces a 2C-like transcriptional program through post-transcriptional regulation of *Dux*, the key inducer of zygotic genome activation (ZGA) in placental mammals⁹⁻¹¹.

Importantly, although DNA damage-induced cellular differentiation has been frequently shown in stem cells^{35,36}, here we report for the first time to our knowledge that RS in ESCs leads to the transition to a more developmentally potent state that is required for the subsequent differentiation toward TGCs. As far as the molecular mechanism of ATR dependent *Dux* level increase is concerned, our results show that ATR activity is required to promote *Dux* mRNA accumulation at the post-transcriptional level. This finding is consistent with previous observations showing that the mRNA processing machinery can regulate DUX4 mRNA level by promoting its degradation³⁷. Significantly, we discovered that GRSF1, a RNA binding protein involved in several aspects of mRNA metabolism³⁰,

directly and specifically binds to *Dux* mRNA and it is required for *Dux* mRNA accumulation in response to ATR activation. The links between GRSF1 and *Dux* are unexpected and will require further work to be fully dissected.

Elevated DUX level in turn increases bipotent 2C-like cells in ESC culture through global transcriptional activation of 2C-specific genes, including genome caretaker genes, *Zscan4* and *Tcstv1/3*^{8,38}. In addition to their role in telomere elongation through activation of telomere sister chromatid exchange^{7,8}, *Zscan4* genes can also promote DNA repair by facilitating heterochromatin decondensation and DNA demethylation^{6,15,39}. Therefore, ATR-mediated activation of *Zscan4* genes likely contributes to ESCs genomic integrity in response to RS. Intriguingly, the DUX4 genes were originally discovered in an attempt to identify the target genes of Helicase-like Transcription Factor (HLTF)⁴⁰, that was later found to act at replication forks upon RS⁴¹. Unexpectedly, our findings also show that ATR activation in ESCs can trigger the generation of cells with bidirectional cell fate potential (Fig. 6h). Activation of this pathway could act as a safeguard mechanism to ensure genome integrity of the developing embryo in response to RS by promoting DNA repair, on the one hand, or by diverting unrepaired ATR-activated ESCs to extra-embryonic tissues, on the other hand, thus limiting the incorporation of cells with unrepaired DNA in embryo proper tissues. These findings could in part explain the phenotype of *Atr*^{sec/sec} mouse embryo, in which failure in proper activation of ATR in response to RS leads to the accumulation of γ H2AX positive cells in embryo proper¹⁹.

Overall our findings revealed that 1) ATR is a potent upstream driver of 2C-specific genes in ESCs culture; 2) 2C-like transcriptional program can also be activated at later stages of embryonic development (rather than being activated only at 2C-embryo stage) in response to exogenous RS; 3) 2C-like state induced by RS and ATR influences ESC fate and impacts on its plasticity and likely genome integrity. These findings suggest that the physiological role of 2C-like transcriptional program might not only be restricted to ZGA. Importantly, *ATR*^{-/-} mice have been reported to develop up to

blastocyst stage^{20,42}, suggesting that presence of ATR may not play an essential role in ZGA at 2C-stage embryo. This would be consistent with a recent finding showing that *Dppa2* and *Dppa4*, which are expressed until the onset of gastrulation, were found to be the direct regulators of DUX-driven zygotic transcriptional program⁴³. This behavior would be compatible with the reactivation of 2C-like transcriptional program at later stages of development, concomitantly with a weakening of the suppressive mechanisms that keep 2C-genes repressed in response to endogenous or exogenous stimuli. However, although our findings indicate that ATR activation impacts on the cell fate even at later stages of embryonic development, we also clearly show that RS *per se* is not able to trigger main 2C-genes in more committed embryonic cells such as MEFs, suggesting an involvement of robust repressive mechanisms in suppressing this pathway in more differentiated cells.

It is tempting to speculate that reactivation of such ATR-dependent transcriptional program in more committed adult stem cells in which these suppressive mechanisms fail, could be responsible for the expression of genes that are involved in invasiveness, angiogenesis and immunosuppression shared by placenta and cancer cells⁴⁴⁻⁴⁶. This would be consistent with the recently reported *Dux/DUX4* reactivation in a diverse range of tumors^{47,48,49}. In line with this, a recent report has shown that DUX4 level increases specifically in cancer cells, promoting their ability to downregulate major histocompatibility complex I antigen presenting molecules⁵⁰, a feature shared between trophoblast and cancer cells that allows escaping from immune surveillance mechanisms⁴⁶. Activation of extraembryonic features in response to ATR in early stages of cellular transformation could be a major outcome of RS induced by oncogene activity in cancer cells and would provide an additional rationale for the potent tumor suppressive effect of ATR inhibitors^{44,51-55}. Intriguingly, extraembryonic tissues have been recently shown to have similar epigenetic methylation profiles to cancer cells⁵⁶, highlighting an additional common feature between these two tissues.

In summary, our findings shed light on the endogenous and exogenous stimuli that could contribute to the cellular plasticity of ESCs and also provide a fundamental insight into the alternative mechanisms these cells exploit to respond to RS and maintain genome integrity.

Figure Legend

Figure 1 | Induction of RS increases the number of 2C-like cells in ESCs culture and activates the expression of 2C-specific genes in mouse embryos

a, Clustering of 1399 Drop-seq single-cell expression profiles into four cell populations. The plot shows a two-dimensional representation (t-SNE) of global gene expression relationship; cells are colored according to their cluster. Clusters were identified by shared nearest neighbor algorithm (see methods). **b**, t-SNE plot showing *Zscan4d* expression level across all CNTL and APH-treated cells. **c**, Heatmap showing the list of top 30 genes that are differentially expressed between cluster 4 cells (orange cluster in Fig. 1a) and the rest of the population (cluster 1,2, and 3). **d**, Plot showing the scaled expression of 2C-specific markers and the percentage of cells expressing 2C-related genes in CNTL and APH-treated condition. Fisher's exact test was used to determine p-values. **e**, FACS analysis on p*Zscan4*-Emerald ESCs upon treatment with various RS-inducing agents. **f**, Immunostaining of ESCs for ZSCAN4-Emerald, MERVL-GAG and canonical pluripotency marker POU5F1 upon treatment with APH (bar= 25µm). **g-j**, RT-qPCR analysis on blastocyst-stage embryos treated with APH for key 2C-like markers. Statistical significance compared to CNTL unless otherwise indicated. All bar-plots show mean with \pm SD (*P \leq 0.05, **P \leq 0.01, ***P \leq 0.001, ****P \leq 0.0001, one-way ANOVA).

Figure 2 | ATR and CHK1-mediated RSR triggers activation of key 2C-specific genes in ESCs

a, FACS analysis on p*Zscan4*-Emerald ESCs showing the number of Em⁺ cells upon treatment with APH and specific ATR and ATM inhibitors. **b**, Immunoblot for the phosphorylation status of key DDR kinases (CHK1 and CHK2) and the ZSCAN4 protein level upon treatment with APH and ATM/ATR inhibitor in ESCs. **c**, RT-qPCR analysis of two ESCs lines for the expression of MERVL upon treatment with APH and ATRi. **d**, Plot showing the scaled expression of 2C-specific markers and the percentage of cells expressing 2C-related genes in CNTL, APH-treated and APH+ATRi conditions. Fisher's exact test was used to determine p-values. **e**, Immunoblot for ZSCAN4, ATR and the phosphorylation status of CHK1 upon APH treatment of *Atr*^{Sec/Sec} and *Atr*^{+/+} ESCs. **f**, immunoblot showing the expression of ZSCAN4 and p-CHK1 in *Chk1*^{+/-} and *Chk1*^{+/+} ESCs upon treatment with APH. **g-k**, RT-qPCR for 2C-specific genes in *Atr*^{Sec/Sec} and *Atr*^{+/+} ESCs treated with APH. **l**, FACS analysis of p*Zscan4*-Emerald ESCs showing the number of Em⁺ cells upon treatment with APH and a specific CHK1 inhibitor. Statistical significance compared to CNTL unless otherwise indicated. All bar plots show mean with \pm SD (*P \leq 0.05, **P \leq 0.01, ***P \leq 0.001, ****P \leq 0.0001, one-way ANOVA).

Figure 3 | ATR induces transcriptional signature of 2C-like cells in ESCs

a, MA plot showing gene expression in ESCs treated with APH in comparison to control. Key 2C-specific genes are highlighted **b**, MA plot showing retrotransposons expression upon APH treatment. **c**, Heatmap showing the robust z-scores for 2C-specific genes in the indicated samples. 2C-related genes were identified by performing a differential expression analysis on ZSCAN4⁺/MERV1⁺.s. ZSCAN4⁻/MERV1⁻ ESCs from Eckersley *et. al.*¹⁵ **d**, Bar plot displaying the percentage of ATR-dependent differentially expressed genes among the ones shared between APH-treated ESCs and each dataset. **e-g**, RT-qPCR analysis of *Dux* KD ESCs for *Dux* and *Zscan4d* genes, and MERV1 upon treatment with APH. **h**, Immunoblot for p-CHK1 and ZSCAN4 proteins upon treatment with APH in *Dux* KD ESCs in comparison with control ESCs. Statistical significance compared to CNTL unless otherwise indicated. All bar plots show mean with \pm SD (*P \leq 0.05, **P \leq 0.01, ***P \leq 0.001, ****P \leq 0.0001, one-way ANOVA).

Figure 4 | ETAA1-mediated activation of ATR induces 2C-like cells in a RS-free context

a, FACS analysis for γ H2AX and Emerald-GFP in Dox-iETAA1 ESCs in the presence or absence of Dox and upon treatment with ATRi or ATMi. **b**, RT-qPCR results for *Dux* expression in Dox-iETAA1 ESCs upon Dox induction in the presence or absence of ATRi or ATMi. **c**, Immunoblot showing the expression of ETAA1-AAD, ZSCAN4, MERV1-GAG and the phosphorylation status of CHK1, CHK2 and H2AX in Dox-iETAA1 ESCs upon treatment with Dox, ATRi or ATMi. **d**, Immunoblot showing the expression of ETAA1-AAD, ZSCAN4, MERV1-GAG in Dox-iETAA1 ESCs upon treatment with Dox and *Dux* knock down. **e-g**, RT-qPCR analysis of Dox-iETAA1 ESCs for expression of 2C-related genes (*Dux*, MERV1 and *Zscan4*) upon treatment with Dox and *Dux* knock down. Statistical significance compared to CNTL unless otherwise indicated. All bar plots show mean with \pm SD (*P \leq 0.05, **P \leq 0.01, ***P \leq 0.001, ****P \leq 0.0001, one-way ANOVA).

Figure 5 | Identification of RSR downstream molecular players regulating the 2C-like state

a,b, Schematic design of screening experiment and siRNA library selection. **c**, Protein interaction network for the hits identified through the siRNA screen. Activators and suppressors are highlighted in red and green, respectively. The interactions are made based on the STRING database. **c**, RT-qPCR analysis of iDox-*Dux* ESCs for exogenous *Dux* mRNA upon treatment with Dox. **d**, RT-qPCR analysis of iDox-Luciferase ESCs for Luciferase mRNA upon treatment with Dox. **e**, RT-qPCR analysis of *Dux* mRNA upon APH treatment and *Grsf1* KD. **f**, Immunoblot showing the expression of GRSF1,

ZSCAN4 and pCHK1 upon *Gsrfl* KD. **g**, Immunoblot showing the binding of GRSF1 protein to the *Dux* mRNA. Statistical significance compared to CNTL unless otherwise indicated. All bar plots show mean with \pm SD (* $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$, one-way ANOVA).

Figure 6 | ATR-activated 2C-like cells gain expanded developmental potential *in vitro* and *in vivo*

a, Phase contrast and immunofluorescence images of TGCs formed by *in vitro* differentiation of *Atr*^{+/+} or *Atr*^{sec/sec} ESCs treated with or without APH (bar = 25 μ m). **b**, Quantification of the number of TGCs detected in the conditions represented in Fig. 6a. **c**, Images of blastocysts displaying the contribution of mCherry-labeled Dox-iETAA1 ESCs to the ICM and TE layers in the presence and absence of Dox (bar = 20 μ m). **d**, Bar plot showing the percentage of chimeric embryos in which injected mCherry-labeled Dox-iETAA1 ESCs could contribute to either ICM or ICM+TE with or without Dox. The ratios on top of each bar show the actual number of embryos analyzed. **e**, Bar plot showing the percentage of chimeric embryos in which single-cell injection of mCherry-labeled Dox-iETAA1 ESCs could contribute to either ICM or ICM+TE/TE with or without Dox. The ratios on top of each bar show the actual number of embryos analyzed. **f**, Images showing the contribution of injected mCherry-labeled Dox-iETAA1 ESCs to the epiblast (EPI) or extra-embryonic layers (EEL) of mouse embryos at E7.5 with or without Dox treatment (bar = 50 μ m). **g**, Immunostaining of mouse embryos at E7.5. Arrows in the left panel indicate the contribution of mCherry-labeled Dox-iETAA1 ESCs to the EPI (marked by POU5F1) in Dox-treated and untreated conditions. Arrows in the right panel indicate the contribution of injected mCherry-labeled Dox-iETAA1 ESCs to the EEL (marked by CDX2) only in Dox-treated condition. (lower magnification, bar = 100 μ m, higher magnification, bar = 25 μ m). **h**, Schematic model defining a novel ATR-dependent transcriptional response to maintain the genomic integrity of developing embryos in response to RS. I) ATR and CHK1-mediated RSR triggers the *Dux* mRNA accumulation through GRSF1 that in turn increases bipotent 2C-like cells by global transcriptional activation of 2C-specific genes including *Zscan4*. II) ATR-induced bipotent ESCs extend their contribution to placental compartment. In the dot plot the mean is represented by a red line. Statistical significance compared to CNTL unless otherwise indicated. (* $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$, one-way ANOVA).

Extended Data Figure 1 | Induction of RS increases the number of 2C-like cells in ESCs culture

a, PCA plot on 1399 cells treated with or without APH. Cells are colored based on clusters. **b**, Heatmap showing scaled expression values of top 30 genes and 100 ranked cells according to PC1. **c**, Venn

diagram showing the overlap of the top 100 DEGs of cluster 4 with those expressed in 2C-like cells (according to Eckersley *et al.*¹⁵). **d**, Plot showing the scaled expression of 2C-specific markers and the percentage of cells expressing 2C-related genes in the four clusters. **e**, Immunoblot showing the expression of ZSCAN4 and p-CHK1 proteins upon treatment with APH at different time points. **f**, FACS analysis for Emerald and CASPASE 3-Alexa 647 on pZscan4-Emerald ESCs upon treatment with increasing concentrations of APH. **g**, β -galactosidase staining of H₂O₂-treated BJ fibroblast cells, untreated and APH-treated ESCs.

Extended Data Figure 2 | Characterization of APH-induced 2C-like cells

a-f, RT-qPCR analysis for 2C-like genes upon treatment with increasing concentrations of APH **g**, FACS analysis on pMERVL-GFP ESCs upon treatment with APH. **h**, FACS analysis showing the mean GFP signal intensity within pMERVL-GFP positive population in CNTL and APH-treated pMERVL-GFP ESCs. **i**, RT-qPCR analysis for MERVL upon treatment with APH in two different MEF lines. **j**, Immunostaining for 2C markers, MERVL-GAG and ZSCAN4 and ESC marker, POU5F1 in untreated and APH-treated ESCs, DAPI staining shows the lack of chromocenters in 2C-like cells (bar =20 μ m), **k-n**, Immunoblot and RT-qPCR analysis of ESCs for ZSCAN4 and canonical pluripotency markers upon treatment with APH. Statistical significance compared to CNTL unless otherwise indicated. All bar plots show mean with \pm SD (*P \leq 0.05, **P \leq 0.01, ***P \leq 0.001, ****P \leq 0.0001, one-way ANOVA).

Extended Data Figure 3 | Characterization of APH-induced 2C-like cells

a, Cell cycle analysis of E14 and R1 ESCs upon treatment with low (0.3 μ M) and high (6 μ M) concentrations of APH. **b**, Cell cycle analysis of E14 and R1 ESCs upon UV. **c-e**, **f-h**, Two replicates of RT-qPCR analysis on blastocyst-stage embryos for key 2C-specific genes *Dux*, *Gm4981* and MERVL. **i**, Wide field microscope images of mouse embryos at blastocyst stage after APH treatment (bar= 25 μ m) **j**, Immunostaining of blastocyst-stage embryos for key ESCs marker, NANOG and the key TSC marker, CDX2 after APH treatment (bar= 50 μ m). **k**, Percentage of embryos developed to the blastocyst stage after APH treatment in 3 replicates. Statistical significance compared to CNTL unless otherwise indicated. All bar plots show mean with \pm SD (*P \leq 0.05, **P \leq 0.01, ***P \leq 0.001, ****P \leq 0.0001, one-way ANOVA).

Extended Data Figure 4 | ATR-mediated RSR triggers activation of key 2C-specific genes in ESCs

a, FACS analysis of p*Zscan4*-Emerald ESCs upon treatment with ATM and ATR inhibitors after induction of RS upon HU. **b**, RT-qPCR analysis of *Dux* mRNA upon APH treatment and ATR KD **c,d**, FACS analysis of p*Zscan4*-Emerald ESCs for DNA damage markers γ H2AX and p-CHK1 (Em^+ and Em^- correspond to Emerald-GFP positive and negative populations, respectively). **e, f**, Immunoblot showing the expression of ZSCAN4, p-CHK1 and p-CHK2 proteins in ESCs upon treatment with UV or APH for 8 hours. **g**, RT-qPCR analysis for *Zscan4d* gene upon treatment with a specific ATRi in two distinct ESC lines (E14 and R1). **h**, Immunoblot showing the absence of ATM kinase in ATM KO cells. **i,j**, RT-qPCR analysis for *Dux* and *Zscan4* mRNA upon APH treatment in ATM WT and KO ESCs. **k**, RT-qPCR analysis for *Dux* mRNA upon APH and two different concentrations of p38 inhibitors (1 and 2). **l-n**, Box plot showing the expression level of 2C-related genes (within the subpopulation of cells expressing the specific marker) at the single-cell level in CNTL, APH and APH+ATRi condition. Statistical significance compared to CNTL unless otherwise indicated. All bar plots show mean with \pm SD (* $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$, one-way ANOVA).

Extended Data Figure 5 | ATR and CHK1-mediated RSR triggers activation of key 2C-specific genes in ESCs

a, b, Immunostaining of *Atr*^{Sec/Sec}, *Atr*^{+/+} ESCs and *Chk1*^{-/-}, *Chk1*^{+/+} ESCs for the canonical pluripotency markers POU5F1 and NANOG (bar= 25 μ m). **c-f**, RT-qPCR analysis for 2C-related genes in *Atr*^{sec/sec} and *Atr*^{+/+} ESCs upon treatment with APH. **g**, FACS analysis of *Chk1*^{-/-} and *Chk1*^{+/+} ESCs for the basal expression level of ZSCAN4 protein **h-n**, RT-qPCR analysis for 2C-related genes in *Chk1*^{-/-} and *Chk1*^{+/+} ESCs upon treatment with APH. **o**, RT-qPCR analysis for *Trp53* expression upon siRNA-mediated KD. CNTL sample was transfected with sc siRNA. **p**, Immunoblot showing ZSCAN4 expression and phosphorylation status of P53 and CHK1 upon APH treatment in CNTL and Trp53 KD ESCs. **q,r**, RT-qPCR analysis for *Trp53* and *Dux* expression in *Trp53* WT and KO ESCs. Statistical significance compared to CNTL unless otherwise indicated. All bar plots show mean with \pm SD (* $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$, one-way ANOVA).

Extended Data Figure 6 | ATR induces transcriptional signature of 2C-like cells in ESCs.

a, Heatmap showing the robust z-scores for all the DEGs in three ESCs lines upon treatment with APH or APH+ATRi. **b**, Plot showing the repeat subfamilies that are significantly differentially expressed

both in our data (CNTL *v.s.* APH) and the ZSCAN4⁺/MERVL⁺ *v.s.* ZSCAN4⁻/MERVL⁻ comparison¹⁵. The Pearson correlation (0.59) between the log₂ fold changes from the two comparisons is also shown on the plot. **c-h**, Validation of RNA-Seq results by RT-qPCR analysis. **i-k**, RT-qPCR analysis of ESCs for expression of MERVL elements upon treatment with APH and ATRi **l**, Plot shows the number of genes expressed in ESCs upon treatment with APH overlapping with previously published datasets. The -log₁₀ (Fisher p-value) shows the significance of overlaps obtained from Fisher test. Statistical significance compared to CNTL unless otherwise indicated. All bar plots show mean with ±SD (*P≤0.05, **P≤0.01, ***P≤0.001, ****P≤0.0001, one-way ANOVA).

Extended Data Figure 7 | ATR induces transcriptional signature of 2C-like cells in ESCs.

a-c, RT-qPCR analysis for *Dux*, *Zscan4*, and MERVL expression upon treatment with APH and ATRi in *Dux* KO and WT ESCs. **d-f**, RT-qPCR analysis of ESCs for expression of 2C-like genes in siRNA-mediated *Dux* KD ESCs compared to CNTL ESCs treated with sc siRNA. Statistical significance compared to CNTL unless otherwise indicated. All bar plots show mean with ±SD (*P≤0.05, **P≤0.01, ***P≤0.001, ****P≤0.0001, one-way ANOVA).

Extended Data Figure 8 | Characterization of ETAA1-AAD inducible ESCs

a, FACS analysis for γH2AX and Emerald upon overexpression of ETAA1-AAD by Dox in two different lentivirus clones with respect to empty vector (EV)-infected ESCs. **b**, FACS analysis of CASPASE-3 in Dox-iETAA1 ESCs. **c**, RT-qPCR analysis for the expression of ETAA1-AAD in Dox-iETAA1 ESCs (E14 and R1) upon Dox treatment. **d, e**, RT-qPCR results for *Dux* and *Zscan4* expression in Dox-iETAA1 ESCs after 24, 48 and 72 hours of Dox administration **f**, RT-qPCR results for MERVL element in Dox-iETAA1 ESCs upon Dox administration in the presence or absence of ATRi or ATMi. **g-h**, RT-qPCR results for 2C-related genes, *Zscan4* and *Dux* in Dox-iETAA1 ESCs upon Dox and CHK1i treatment. **i-j**, RT-qPCR analysis for *Zscan4* and *Dux* expression in Dox-iETAA1 ESCs in the absence or presence of Dox and CHK1i (LY2603618). **k-m**, RT-qPCR analysis for the expression of 2C-related genes in Dox-inducible *Dux* KO ESCs in comparison with WT ESCs upon treatment with Dox. Statistical significance compared to CNTL unless otherwise indicated. (*P≤0.05, **P≤0.01, ***P≤0.001, ****P≤0.0001, one-way ANOVA).

Extended Data Figure 9 | ATR-activated 2C-like cells gain expanded developmental potential.

a,b, RT-qPCR analysis for ETAA1 and *Dux* genes expression in siRNA-mediated EtAA1 KD ESCs upon treatment with APH. CNTL ESCs were treated with sc siRNA. **c**, Immunoblot for expression of ETAA1, ZSCAN4 and p-CHK1 in siRNA-mediated ETAA1 KD ESCs upon treatment with APH. **d**, Immunostaining showing the ETAA1 expression in ZSCAN4 positive cells. **e,h**, RT-qPCR analysis of 2C-related genes expression in Dox-iETAA1 ESCs (R1 and E14) upon Dox treatment and/or siRNA mediated *Dux* KD. **i**, RT-qPCR analysis for exogenous *Dux* mRNA in *iDux-Dux* KO cells upon treatment with Dox, and ATR inhibitor. **j**, Immunoblot showing the level of p-CHK1 upon DUX OE. **k**, RT-qPCR analysis for *Dux* mRNA level upon siRNA-based KD of *Dux* 3'UTR binding factors. Statistical significance compared to CNTL unless otherwise indicated. (* $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$, one-way ANOVA).

Extended Data Figure 10 | ATR-activated 2C-like cells gain expanded developmental potential.

a, RT-qPCR assay for the TGCs specific marker *Prl2c2* in *Atr*^{+/+} and *Atr*^{Sec/Sec} ESC-derived TGCs upon APH treatment. **b**, Immunostaining of *Dux* KO and WT-derived TGCs upon APH treatment (bar=25 μ m). **c**, Plot showing the number of TGCs generated from WT and *Dux* KO ESCs upon APH treatment. **d**, Immunostaining of blastocysts displaying the contribution of mCherry-labeled Dox-iETAA1 ESCs to the ICM and TE layers in the presence and absence of Dox (bar= 20 μ m). **e**, Bar plot showing the average number of injected mCherry-labeled Dox-iETAA1 ESCs that could contribute to TE with or without Dox. **f**, Beeswarm plot showing the average number of injected mCherry-labeled Dox-iETAA1 *Dux* KO ESCs that could contribute to TE with or without Dox. **g**, Images showing the contribution of injected mCherry-labeled Dox-iETAA1 ESCs to the epiblast (EPI) or extra-embryonic layers (EEL) of mouse embryos at E7.5 with or without Dox treatment (bar= 50 μ m). Statistical significance compared to CNTL unless otherwise indicated. (* $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$, one-way ANOVA).

Extended Data Figure 11 | ATR-activated 2C-like cells gain expanded developmental potential.

a, b Immunostaining of mouse embryos at E7.5. Arrows indicate the contribution of mCherry-labeled Dox-iETAA1 ESCs to the EPI (marked by POU5F1) or EEL (marked by CDX2) in Dox-treated and untreated conditions.

Materials & Methods

Animals

All mice used to generate ATR Seckel ESCs and MEF lines were bred and maintained under specific pathogen-free conditions. C57Bl/6J and 129P2/OlaHsd mice were purchased from Charles River Laboratories Harlan Italy (currently known as Envigo), respectively. Animals were kept in ventilated cages in standard 12 hours light-dark cycle. The procedure was approved by the FIRC Institute of Molecular Oncology Institutional Animal Care and Use Committee and performed in compliance with Italian law (D.lgs. 26/2014 and previously D.lgs. 116/92), which enforces Dir. 2010/63/EU (Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010, on the protection of animals used for scientific purposes).

Cell culture

ESCs were grown in feeder-free culture condition and incubated at 37 °C under 3% O₂ tension. ESC medium composed of KnockOut DMEM (ThermoFisher, 10829-018), 15% ESC qualified Fetal Bovine Serum (ThermoFisher, 16141-079), 2 mM L-glutamine, 1/500 home-made leukemia inhibitory factor (LIF), 0.1 mM non-essential amino acids, 0.1 mM 2-mercaptoethanol, 50 units/mL penicillin, 50 mg/mL streptomycin supplemented with the two inhibitors (2i); PD 0325901 (1 μM) and CHIR 99021 (3 μM) was used. MEFs were cultured in high glucose DMEM (Lonza, BE12-614F) supplemented with 10% North American FBS, 2 mM L-glutamine, 0.1 mM non-essential amino acids, 50 units/mL penicillin, 1 mM Sodium Pyruvate and 50 mg/mL streptomycin (MEF medium).

ESCs derivation from ATR Seckel and CHK1 haploinsufficient mice

To establish $Atr^{Sec/Sec}$ ESCs, $Atr^{+/Sec}$ heterozygous mice were crossed. Morulae were recovered and cultured overnight in KSOM medium (Millipore, MR-020P-5D) under mineral oil. Blastocysts were placed on MEFs feeder layer in ESC medium with 2i and LIF at 37 °C under 5% CO₂. Upon ICM

expansion, cells were passaged on feeder layer to obtain the first stock. The cells were then characterized by genotyping. To establish *Chk1*^{+/-} ESCs lines, embryos at morula stage were cultured overnight in KSOM supplemented with 2i. On the following day, blastocysts were plated individually in 96-well plate and cultured in N2B27 medium supplemented with 2i and LIF. After 6-7 days, the ICM outgrowth was disaggregated using Accutase. Clumps of cells were expanded every second day, or when the size of colonies reached to the proper expansion level.

MEFs generation

To produce MEFs, C57BL/6 and 129P2/OlaHsd mice were used. MEFs were prepared by mechanical disaggregation, trypsinization and seeding of embryos (E12-E13) in MEF medium after removal of the head, tail, limbs, and internal organs. Each trypsinized embryo was plated into a 10 cm dish. Once cells reached to 90% confluency (after 48 hours), they were frozen and stocked for the subsequent experiments.

Treatments

All treatments, including APH, HU, UV, ATRi, ATMi, p38i and CHK1i, were added into fresh ESC medium and the cells were kept at 37 °C under 3% O₂ tension overnight (16-17 hours) unless otherwise indicated. APH was used at concentrations ranging from 0.4 to 6 μM as indicated. Similarly, HU was used at 0.1 mM - 0.2 mM. ATMi (KU-55933) and ATRi (VE 822) were used at the concentrations of 10 μM and 1 μM respectively. The CHK1 inhibitors LY2603618 and UCN-01 were used at the concentrations of 200 nM and 100 nM, respectively. UV radiation exposure was performed at the dosage of 5 J/m² or 10 J/m². For induction of ETAA1-AAD expression, Dox was used at the concentration of 1 μg/mL for 48 hours. SB 203580 hydrochloride (Tocris Bioscience) and SB 239063 (Tocris Bioscience) were used at the concentration of 5 and 10 μM.

RNA extraction, cDNA synthesis and qPCR

RNA was extracted using RNeasy Mini Kit (QIAGEN, 74104) and quantified by spectrophotometer (NanoDrop, ThermoFisher) after removal of DNA contaminants by DNase I digestion (QIAGEN, 79254). cDNA was prepared from 2 µg of total RNA using SuperScriptTM III reverse transcriptase kit (ThermoFisher, 18080-044) with Oligo(dT)20 Primer (ThermoFisher, 18418-020) and dNTP Mix (ThermoFisher, 18427-013) following manufacturer's instructions. qPCR assay was performed based on standard protocol using final working concentration of 1X SsoFastTM EvaGreen® Supermix (Bio Rad, 1725201) or 1X LightCycler 480 SYBR Green I Master (Roche, 04707516001), 0.5 µM primer mix and 5 ng of cDNA. GAPDH, TBP or GUSB were used as internal controls to normalise the qPCR data following the $\Delta\Delta C_t$ method. For the list of oligos used in this study please refer to the Supplementary Table 8.

RNA Isolation from Embryos and cDNA preparation

Cryopreserved C57BL/6N morulae purchased from Janvier Labs (QUICKBLASTO®), were thawed according to the manufacturer's protocol. After recovery, a group of healthy morulae were treated with APH at final concentration of 1.5 µM in KSOM culture medium for 4 hours while the rest were kept as control. The day after, the same number of morphologically healthy and synchronized blastocysts (22 and 7 blastocysts in the 1st and 2nd round, respectively) were selected from both groups. The APH concentration and the treatment interval had no effect on the viability of the embryos. RNA isolation was performed based on the previously published works^{43,44}. Briefly, following the removal of zona pellucida in acidic tyrode's solution (Sigma-Aldrich, T1788), the pool of embryos was collected in 10 µl of lysis buffer containing 5X First Strand Buffer (from SuperScript III reverse transcriptase kit, ThermoFisher) supplemented with 0.1% Tween-20. Subsequently, embryos were mechanically ruptured by three cycles of freeze/thaw. As the quantity of isolated RNA is not detectable by NanoDrop spectrophotometer, the supernatant was directly used for cDNA preparation following centrifugation at

10621g for 2 minutes. cDNA preparation and qPCR were performed following the aforementioned protocols.

Flow cytometry (FACS)

ESCs were fixed and permeabilized using Cytoperm/Cytofix kit (BD Biosciences, 554714), and subsequently stained for 1 hr at room temperature with anti-H2AX-Phosphorylated (Ser139) antibody conjugated with Alexa Fluor 647 (Biolegend), or Anti-p-CHK1 antibody conjugated with PE (Cell Signaling, 12268) or Anti-ZSCAN4 (Merck Millipore, AB4340) antibody conjugated with Anti-Rabbit ALexa488. Cells were washed and acquired on a FACSCalibur instrument (BD Biosciences) or Attune NxT (Thermo Fisher Scientific). For Emerald, mCherry and GFP acquisition, cells were trypsinized, collected and subsequently acquired without fixation. Cell cycle profiles were acquired after incubation with DAPI (4',6-diamidino-2-phenylindole) overnight.

Immunocytochemistry

Briefly, cells were fixed in 4% formaldehyde (Sigma-Aldrich) for 20 minutes and subsequently blocked for 1 hr in 10% FBS and 0.1% Triton-X100. Then, cells were incubated with primary antibody at 4 °C overnight, followed by washes and incubation with secondary antibodies for 1 hr at room temperature. Next, samples were mounted and images were acquired with wide-field fluorescence microscope (Olympus AX70, Olympus). Acquired images were analyzed by ImageJ software (NIH). For staining mouse embryos, blastocysts were fixed in 4% formaldehyde, followed by permeabilization for 10 minutes in 0.5% Triton X-100 and blocking for 1 hr at room temperature in 3% BSA and 0.1% Tween-20 in PBS. Then blastocysts were incubated with primary antibodies at 4 °C overnight, before washes in 0.1% BSA in PBS and secondary antibodies incubation for 2 hours at room temperature, followed by a 15 minutes incubation with DAPI. Stained blastocysts were imaged with Leica TCS SP2 AOBS inverted confocal microscope.

Antibodies used in this study were Anti-Oct-3/4 (Santa Cruz, sc-5279), Anti-Nanog (Abcam, ab80892), Anti-GFP (Abcam, ab5450), Anti-Plet-1 (Nordic MUBio, MUB1512P), Anti-Cdx2 (Biocare Medical CM226A), Anti-mCherry (ChromoTek 5F8 a-RFP), Anti-ZSCAN4 (Merck Millipore, AB4340), Anti-MERVL-GAG (Huabio, ER50102), Anti-ETAA1 (Abcam, ab197017).

Immunoblotting

Cells were trypsinized and washed with PBS and then lysed for 30 minutes on a rotating wheel at +4 °C in RIPA buffer supplemented with protease/phosphatase inhibitor cocktail (Cell Signaling, 5872). Lysates were sonicated with a Bioruptor Sonication System (UCD200) at high power for 3 cycles of 30 seconds with one minute breaks. Lysates were centrifuged at 13000 rpm for 20-30 minutes and clear supernatants were transferred to new tubes. Protein content was quantified using Bio-Rad protein assay according to manufacturer's instructions. For the detection of each protein, 35 µg of total protein extracts were loaded. Standard western blot was performed using following antibodies: Anti-ATR (Cell Signaling, 2790), Anti-p-CHK1 (S317) (cell signaling, 12302), Anti-ZSCAN4 (Merck Millipore, AB4340), Anti-CHK1 (Santa Cruz, 8408), Anti-CHK2 (Millipore, 05-649), Anti-γH2AX (Merck Millipore, 05-636), Anti-P53 (Cell Signaling, 2524S), Anti-p-P53 (S15) (Cell Signaling, 12571S), Anti-GAPDH (Abcam, ab9484), Anti Oct-3/4 antibody (Santa Cruz, sc-5279), Anti-VINCULIN (Sigma, V9131), Anti-ATM (Sigma, A1106), Anti-GRSF1 (Abcam, AB246330), Anti-PCBP2/hnRNP E2 (Abcam, AB236137), Anti-MERVL-GAG (Huabio, ER50102), Anti-ETAA1 (Abcam, ab197017), Anti-SOX2 (Chemicon International, AB5603).

Drop-seq single cell mRNA sequencing

ESCs from CNTL, APH (6 µM) and APH-ATR_i experimental conditions were resuspended in PBS-BSA and processed with a microfluidic device according to the DropSeq Laboratory Protocol V3.1 from McCarroll's Lab website [<http://mccarrolllab.com/dropseq/>]. For each condition, 3 distinct

aliquots of collected emulsion, each containing 4000 beads underwent reverse transcription (RT) and fragment library preparation using Illumina Nextera XT Library Prep Kit. A distinct barcode was used for each library to allow subsequent demultiplexing of sequencing reads. mRNA sequencing was performed using an Illumina HiSeq2000 instrument, library fragments were sequenced at 50 base pairs (bp) in PE mode. Sequencing reads were aligned to UCSC Mouse Reference Genome version mm10 using STAR (version 2.5.3a), and processed according to DropSeq Alignment Cookbook V1.2 to generate a digital expression matrix of STAMPs (cells) for each experimental condition, which was then furtherly analyzed in the RStudio environment (R version 3.3.3) using Seurat package V2.0 from Satija Lab [<http://satijalab.org/seurat/>]. Seven-hundred highly expressing cells were selected in each condition; genes expressed in less than 3 cells and cells expressing less than 200 genes were pruned from the dataset. Data from CNTL and APH conditions (1399 cells) or CNTL, APH and APH-ATRI conditions (2096 cells) were merged to generate two datasets, which were log normalised and scaled. Expression mean and variance to mean ratio ($\text{LogVMR} > 0.5$) were used to estimate data dispersion and select variable genes across the dataset. Respectively 2614 and 2768 genes (in the 2-conditions and 3-conditions datasets) were used in principal component analysis (PCA) to identify the appropriate number of components to include in data modeling. Top 15 components were selected to explain dataset complexity. A shared nearest neighbor (SNN) graph and smart local moving algorithm was used to perform cells clustering; t-SNE (t-stochastic neighbor embedding) analysis was used to reduce (PCA) dimensionality and generate plots of cells distribution.

RNA sequencing library preparation and sequencing

RNA samples were quantified using Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA) and RNA integrity was checked with RNA Screen Tape on Agilent 2200 TapeStation (Agilent Technologies, Palo Alto, CA, USA). All RNA samples had a RIN score of 10. Ribo-Zero rRNA

Removal Kit and TruSeq Stranded Total RNA library Prep kit was used to generate barcoded fragment libraries for RNA sequencing, following manufacturer's protocol (Illumina, Cat# RS-122-2101). Briefly, rRNA depleted RNAs were fragmented for 8 minutes at 94 °C. First strand and second strand cDNA were subsequently synthesized. The second strand of cDNA was marked by incorporating dUTP during the synthesis. cDNA fragments were adenylated at 3' ends, and indexed adapters were ligated to cDNA fragments. Limited cycle PCR was used for library enrichment. The incorporated dUTP in second strand cDNA quenched the amplification of second strand, which helped to preserve the strand specificity. Sequencing libraries were validated using DNA Analysis Screen Tape on the Agilent 2200 TapeStation (Agilent Technologies, Palo Alto, CA, USA), and quantified by using Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA) as well as by quantitative PCR (Applied Biosystems, Carlsbad, CA, USA).

Barcoded libraries were normalized, pooled and loaded on a single flowcell's lane of Illumina HiSeq4000 sequencer for cluster generation and sequencing, according to manufacturer's instructions. The samples were sequenced using a 2x150 Pair-End (PE) High Output configuration. Image analysis and base calling were conducted by the HiSeq Control Software (HCS) on the HiSeq instrument. Raw sequence data (.bcl files) generated from Illumina HiSeq was converted into FASTQ files and demultiplexed using Illumina bcl2fastq program version 2.17. One mis-match was allowed for index sequence identification. On average ~29 million reads were generated for each sample and average insert size ranged between 204 – 231 for different libraries.

RNA sequencing data processing

Transcript and gene level quantification was done with Kallisto (version 0.43.0)⁵⁷. The RefSeq NM and NR sequence collection as of 2016-11-21, was used to build the transcriptome index. Kallisto was run with the --bias option to perform sequence based bias correction on each sample. Additionally, all reads were mapped with the STAR spliced aligner (version 2.5.2b)⁵⁸ to the GRCm38 (mm10)

genome after trimming the reads to 100 bp with Trimmomatic (version 0.36)⁵⁹. Repeat expression was quantified by counting fragments overlapping with the repeat annotation from the Choi et al. 2017 paper³, using FeatureCounts from the subread package (version 1.5.2)⁶⁰. FeatureCounts was run with the `-p -B -C --primary` options.

RNA sequencing: gene level differential expression

Read counts for each sample estimated by Kallisto were imported into the R statistical environment (version 3.2.2) using tximport (version 1.1.2)⁶¹. The limma (version 3.24.15) package⁶² was used to test for differential expression between CNTL and APH or CNTL and APH+ATRI treatments after TMM normalization and voom transformation⁶⁰. A linear model was fitted with the limma lmFit function and the moderated t-statistics was calculated with the eBayes function. Genes were defined as differentially expressed if they had FDR adjusted p-values < 0.05 and $|\log_2 \text{fold change}| > 1$.

The robust z-score for each gene in each sample was calculated using the median and mean absolute deviation (MAD) for each gene across samples, and standardizing expression with the following formula:

Robust Z - score = $(x - \text{median})/\text{MAD}$ where x is the expression in TPM for a single gene in a given sample.

RNA sequencing repeat analysis

Read counts for each sample based on the STAR alignment were imported into the R statistical environment (version 3.2.2) and summarized on the repeat subfamily level. edgeR (version 3.14.0)⁶³ was used to test for differential expression between CNTL and APH treatments, after TMM normalization. All repeat subfamilies were dropped where the count per million value did not reach 1 in at least 2 samples. A linear model was fitted with the edgeR glmFit function and a likelihood ratio p-

value was calculated with the glmLRT function. Repeat subfamilies were defined as differentially expressed if they had FDR adjusted p-values < 0.05 and $|\log_2 \text{fold change}| > 1$.

RNA sequencing gene set enrichment analysis

Gene set enrichment was calculated using a mouse version of MSigDB available from <http://bioinf.wehi.edu.au/software/MSigDB/> (Accessed September 12, 2016). The CAMERA method⁶⁴ available in the limma package was used to check for significant enrichment of specific gene sets between the CNTL and APH treatments, using the same TMM normalized read counts as in the gene differential expression analysis. Additionally, the GSVA⁶⁵ package was used to transform gene level read counts per sample to a gene set level enrichment score per sample and for each MSigDB gene set. After this, the same limma analysis was carried out on the gene set enrichment values as for the gene level data. Gene sets were defined as differentially expressed if they had FDR adjusted p-values < 0.05 and $|\log_2 \text{fold change}| > 0.5$.

Gene set enrichment for the Control and Dox treatment or Control and *Dux* KD+Dox treatment was calculated with the ROAST method available in the limma (version 3.24.15) package^{65,66}.

RNA sequencing literature comparison

For all of the literature comparisons, the same transcript and gene level quantification methods and transcriptome annotations were used as for our data using the same limma package and methods to define differentially expressed genes between conditions. The Fisher-test was used to check for the significance of the overlaps in differentially expressed genes between our data and the literature datasets.

Transcription factor binding site prediction

To predict transcription factor binding sites in the *Dux* promoter region, we extracted the promoter sequence around the *Dux* transcription start site (TSS), with 1000 nucleotides upstream and 100 nucleotides downstream of the TSS. We submitted this sequence to the JASPAR database, and

predicted putative binding sites using the JASPAR CORE set⁶⁷. We filtered out predicted binding sites with a score < 5, and ranked the putative transcription factors regulating *Dux* based on the total number of binding sites predicted.

Transfection

ESCs were plated in 2i plus LIF medium as described above with or without antibiotics according to the applied Lipofectamine (RNAiMAX or 2000, respectively). Lipofectamine (Sigma-Aldrich) was diluted in Opti-MEM (ThermoFisher) according to the manufacturer's protocol and incubated for 5 minutes at room temperature. siRNAs and MISSION esiRNAs (Sigma-Aldrich) were diluted in Opti-MEM to the opportune concentration and then added to the Lipofectamine emulsion at 1:1 ratio. Following incubation at room temperature (timing depends on the type of Lipofectamine), the mixture was added to the cells suspension in culture medium. The final concentration of *Dux* and *Trp53* siRNAs were 75nM and 40nM, respectively and the final concentration of MISSION esiRNAs were 60nM. For each experiment a sample was transfected with MISSION siRNA Fluorescent Universal Negative Control, Cyanine 5 (for siRNA transfection) or esiRNA targeting RLUC (for MISSION esiRNA transfection) as negative controls at the same concentration used for the siRNA or esiRNA of interest. Gene expression was assessed 48 hours post transfection.

Cloning, lentivirus production and induction

pLVX-EF1 α -IRES-mCherry vector was co-transfected with plasmids encoding POL, REV, TAT, and the vesicular stomatitis virus envelope glycoprotein (VSV-G) into 80% confluent 293T cells using calcium phosphate precipitation in the presence of 25 mM chloroquine. The supernatant of transfected cells was collected every 24 hours for two days and concentrated using PEG-itTM virus precipitation solution and viral particles were resuspended in DMEM and frozen in small aliquots at -80 °C.

The Lenti-XTM Tet-On® 3G Inducible Expression System was used to express ETAA1-AAD protein in ESCs. The cDNA of ETAA1-AAD (kind gift from Mailand lab, Novo Nordisk Foundation Center for Protein Research, Copenhagen, Denmark) was cloned in pLVX-TRE3G vector. Viral particles of pLVX-TRE3G-ETAA1-AAD and pLVX-EF1a-Tet3G vectors were separately produced (as mentioned above). ESCs were co-infected with both lentiviruses in the presence of 8 µg/mL polybrene, and subsequently infected cells were subjected to puromycin and neomycin selection for two weeks. To induce ETAA1-AAD expression, selected ESCs were treated with 1µg/mL Dox for 48 hours. To generate *Dux* and luciferase overexpressing ESCs, pCW57.1-Luciferase (Addgene, 99283) and pCW57.1-mDux-CA (Addgene, 99284) plasmids were used for lentivirus production.

siRNA-based screening

The esiRNA screening was performed on Tecan Freedom EVO 200 Automated workstation (Tecan Group Ltd., Switzerland), equipped with RoMa (Robotic Manipulator Arm) and LiHa (Liquid Handling Arm) with 1 ml dilutors. Briefly, ESCs were reversely transfected with esiRNAs and Lipofectamine RNAiMAX (Invitrogen 56532) in 96-well plates. Subsequently, cells were treated with APH or ATRi 36-hour post-transfection. Twelve hours after treatment, cell proliferation was assessed by CyQUANT Cell Proliferation Assay kit (ThermoFisher C35011), followed by cell lysis using Cells-to-CT Kit (ThermoFisher, AM1728). The Taqman qPCR assays were performed using Biomek FXP Automated Workstation (Beckman Coulter Life Sciences).

***In vitro* differentiation and analysis**

ESCs were differentiated toward TSC and TGCs as previously reported^{32,68,69}. Briefly, ATR Seckel or *Dux* KO ESCs along with WT ESCs were seeded on gelatin-coated 6-well plate in ESCs medium and 24 hours post treatment, medium was changed to trophoblast stem cell (TSC) differentiation medium, which contains: 30% RPMI 1640 (Lonza, BE12-167F) supplemented with

20% FBS, 1mM pyruvate, 2mM L-glutamine, 100mM β -mercaptoethanol), 70% of conditioned medium from mitomycin-C-inactivated fibroblasts, 25 ng/mL FGF4 (R&D Systems, 235-F4-025) and 1 μ g/mL heparin (Sigma-Aldrich, H3149). The medium was refreshed daily to maintain TSCs. To induce giant cell differentiation, TSCs were split at day 2 on gelatin-coated plate. After 24 hours, medium was changed to RPMI 1640 with 20% FBS, 1 mM pyruvate, 2 mM L-glutamine, 100 mM β -mercaptoethanol in the absence of heparin and FGF4. The medium was changed daily for 3 days.

Mouse chimera assay

Morulae were obtained from superovulated C57BL/6 prepubescent females from Janvier laboratories. One (single cell experiments,) five (E4.5 embryos) or ten (E7.5 embryos) ESCs were injected into each morula after piercing their zona pellucida with beveled microinjection needle. Embryos were kept in M2 medium during the injection and afterwards in KSOM medium at 37 °C under 5% CO₂. Morulae were either cultured for 48 hours to study ESC contribution at late blastocyst stage (E4.5), or transferred to pseudopregnant females and later dissected at E7.5. ESC contribution to the various embryonic layers were assessed by the localization of mCherry fluorescence signal.

For the APH-treated ESCs, frozen morula-stage embryos were purchased from Charles River and Janvier laboratories. Embryos were thawed 2 hours prior to injection and 4 to 8 ESCs were transferred into the perivitelline space by laser-assisted microinjection method. Afterwards, embryos were cultured for 24 hours at 37 °C, 5% CO₂ until blastocyst stage (E3.5).

RNA-pull down

RNA-pull down was performed using Pierce Magnetic RNA-Protein Pull-Down Kit (ThermoFisher, 20164) following provider's instruction using *Dux* 3'UTR synthetic RNA.

Liquid chromatography–tandem MS (LC–MS/MS) analysis

The entire gel lane was processed with STAGE-digging protocol as described by Soffientini *et al.*⁷⁰. The entire protocol occurs in a p1000 tip (Sarstedt 70.762.100) filled at the orifice with a double C18 Empore Disk (3M, Minneapolis, MN) plug, named STAGE-digging tip. Briefly, after Coomassie staining, the entire lane was carefully cut into ~ 1 mm³ cubes and transferred into the STAGE-digging tip. These gel cubes were dehydrated with 100% acetonitrile (CAN, Carlo Erba 412392000) and rehydrated in 100 mM NH₄HCO₃ (Sigma A-6141) twice before being dehydrated by the addition of ACN. To ensure that the gel pieces do not create a sticky surface on the C18, all the solutions were added with a gel-loader tip. The removal of solutions was accomplished by centrifugation at 1800 rpm using the commercial tip box as holder. Reduction of protein disulfide bonds was carried out with 10 mM dithiothreitol (DTT, Nzytech MB03101) in 100 mM NH₄HCO₃ and subsequent alkylation was performed with 55 mM iodoacetamide (IAA Sigma I1149-25G), in complete darkness, in 100 mM NH₄HCO₃, at room temperature for 30 minutes. Both DTT and IAA were removed by centrifugation or by syringe as previously described. The gel pieces were rehydrated and dehydrated with 100 mM NH₄HCO₃ and ACN respectively prior to digestion. Gel pieces were rehydrated with 40 µL of Trypsin (12.5 ng/µL in 100 mM NH₄HCO₃, after few minutes 60 µL of NH₄HCO₃ were added and samples were incubated at 37 °C o/n in a commercial tip box filled by water on the bottom to ensure that buffer will not evaporate. The digestion solution was then forced through the double plug with a syringe and the flow through was collected. Samples were acidified with 100 µL of formic acid (FA, Fluka 94318) 0.1%, forced with the syringe and collected as flow-through to desalt the peptides. Peptides were eluted twice by adding 100 µL of a solution composed of 80% ACN, 0.1% FA, an extra step of extraction with 100% ACN was performed and then all the eluates were dried in a Speed-Vac and resuspended in 12 µL of solvent A (2 % ACN, 0.1% formic acid). Four µL were injected for each technical replicate on the Q-Exactive –HF mass spectrometer.

Mass spectrometry analysis

Mass spectrometry analysis was carried out by LC–MS–MS on a quadrupole Orbitrap Q Exactive HF mass spectrometer (Thermo Scientific). Peptide separation was achieved on a linear gradient from 95% Solvent A to 50% Solvent B (80% acetonitrile, 0.1% formic acid) over 20 minutes and from 50% to 100% Solvent B in 2 minutes at a constant flow rate of 0.25 μl minutes⁻¹ on a UHPLC Easy-nLC 1000 (Thermo Scientific), where the LC system was connected to a 25 cm fused-silica emitter of 75 μm inner diameter (New Objective), packed in house with ReproSil-Pur C18-AQ 1.9 μm beads (Maisch) using a high-pressure bomb loader (Proxeon). MS data were acquired using a data-dependent top15 method for HCD fragmentation. Survey full scan MS spectra (300–1750 Th) were acquired in the Orbitrap with 60,000 resolution, AGC target 1e6, IT 120 ms. For HCD spectra the resolution was set to 15,000, AGC target 1e5, IT 120 ms; normalized collision energy 28 and isolation width 3.0 m/z . Two technical replicates of each sample were carried out.

Protein identification

For protein identification, the raw data were processed using Proteome Discoverer (version 1.4.0.288, Thermo Fischer Scientific). MS² spectra were searched with Mascot engine against uniprot_mouse_database (80894 entries), with the following parameters: enzyme Trypsin, maximum missed cleavage 2, fixed modification carbamidomethylation (C), variable modification oxidation (M) and protein N-terminal acetylation, peptide tolerance 10 ppm, MS/MS tolerance 20 mmu. Peptide Spectral Matches (PSM) were filtered using percolator based on q-values at a 0.01 FDR (high confidence). Proteins were considered identified with 2 unique high confident peptides⁷¹. Scaffold (version Scaffold_4.3.4, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability by the Peptide Prophet algorithm⁷² with Scaffold delta-mass correction.

Protein identifications were accepted if they could be established at greater than 99.0% probability and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm⁷³. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Proteins sharing significant peptide evidence were grouped into clusters. All the proteomic data as raw files, total proteins and peptides identified with relative intensities and search parameters were loaded on Peptide Atlas repository (accession number <http://www.peptideatlas.org/PASS/PASS01443>)

β-Galactosidase staining

β-Galactosidase staining was performed following provider's instruction (Cell Signaling Technology, 9860).

Data/material availability

Raw sequencing reads for the bulk and single cell RNA-seq have been deposited in the NCBI BioProject database under accession number PRJNA415135 and PRJNA415187. The materials/data that support the findings of this study will be available from the corresponding authors upon request.

Reference

- 1 Giachino, C., Orlando, L. & Turinetto, V. Maintenance of Genomic Stability in Mouse Embryonic Stem Cells: Relevance in Aging and Disease. *International Journal of Molecular Sciences* **14**, 2617-2636, doi:10.3390/ijms14022617 (2013).
- 2 Ahuja, A. K. *et al.* A short G1 phase imposes constitutive replication stress and fork remodelling in mouse embryonic stem cells. *Nat Commun* **7**, 10660, doi:10.1038/ncomms10660 (2016).
- 3 Choi, Y. J. *et al.* Deficiency of microRNA miR-34a expands cell fate potential in pluripotent stem cells. *Science* **355**, 596+, doi:10.1126/science.aag1927 (2017).
- 4 Ishiuchi, T. *et al.* Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. *Nature Structural & Molecular Biology* **22**, 662-U635, doi:10.1038/nsmb.3066 (2015).
- 5 Macfarlan, T. S. *et al.* Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57+, doi:10.1038/nature11244 (2012).
- 6 Akiyama, T. *et al.* Transient bursts of Zscan4 expression are accompanied by the rapid derepression of heterochromatin in mouse embryonic stem cells. *DNA Res* **22**, 307-318, doi:10.1093/dnares/dsv013 (2015).
- 7 Nakai-Futatsugi, Y. & Niwa, H. Zscan4 Is Activated after Telomere Shortening in Mouse Embryonic Stem Cells. *Stem Cell Reports* **6**, 483-495, doi:10.1016/j.stemcr.2016.02.010 (2016).
- 8 Zalzman, M. *et al.* Zscan4 regulates telomere elongation and genomic stability in ES cells. *Nature* **464**, 858-U866, doi:10.1038/nature08882 (2010).
- 9 De Iaco, A. *et al.* DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat Genet* **49**, 941+, doi:10.1038/ng.3858 (2017).
- 10 Hendrickson, P. G. *et al.* Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat Genet* **49**, 925+, doi:10.1038/ng.3844 (2017).
- 11 Whiddon, J. L., Langford, A. T., Wong, C. J., Zhong, J. W. & Tapscott, S. J. Conservation and innovation in the DUX4-family gene network. *Nat Genet* **49**, 935+, doi:10.1038/ng.3846 (2017).
- 12 Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214, doi:10.1016/j.cell.2015.05.002 (2015).
- 13 Aze, A., Sannino, V., Soffientini, P., Bachi, A. & Costanzo, V. Centromeric DNA replication reconstitution reveals DNA loops and ATR checkpoint suppression. *Nat Cell Biol* **18**, 684+, doi:10.1038/ncb3344 (2016).
- 14 Cimprich, K. A. & Cortez, D. ATR: an essential regulator of genome integrity. *Nat Rev Mol Cell Biol* **9**, 616-627, doi:10.1038/nrm2450 (2008).
- 15 Eckersley-Maslin, M. A. *et al.* MERVL/Zscan4 Network Activation Results in Transient Genome-wide DNA Demethylation of mESCs. *Cell Rep* **17**, 179-192, doi:10.1016/j.celrep.2016.08.087 (2016).

- 16 Storm, M. P. *et al.* Zscan4 Is Regulated by PI3-Kinase and DNA-Damaging Agents and Directly Interacts with the Transcriptional Repressors LSD1 and CtBP2 in Mouse Embryonic Stem Cells. *Plos One* **9**, doi:ARTN e8982110.1371/journal.pone.0089821 (2014).
- 17 Oliva, J. *et al.* Clinically Advanced p38 Inhibitors Suppress DUX4 Expression in Cellular and Animal Models of Facioscapulohumeral Muscular Dystrophy. *J Pharmacol Exp Ther* **370**, 219-230, doi:10.1124/jpet.119.259663 (2019).
- 18 Liu, Q. H. *et al.* Chk1 is an essential kinase that is regulated by Atr and required for the G(2)/M DNA damage checkpoint. *Gene Dev* **14**, 1448-1459 (2000).
- 19 Murga, M. *et al.* A mouse model of ATR-Seckel shows embryonic replicative stress and accelerated aging. *Nat Genet* **41**, 891-U846, doi:10.1038/ng.420 (2009).
- 20 Brown, E. J. & Baltimore, D. ATR disruption leads to chromosomal fragmentation and early embryonic lethality. *Gene Dev* **14**, 397-402 (2000).
- 21 Rowe, H. M. *et al.* KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**, 237-240, doi:10.1038/nature08674 (2010).
- 22 Hoek, M. & Stillman, B. Chromatin assembly factor 1 is essential and couples chromatin assembly to DNA replication in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 12183-12188, doi:10.1073/pnas.1635158100 (2003).
- 23 Bass, T. E. *et al.* ETAA1 acts at stalled replication forks to maintain genome integrity. *Nat Cell Biol* **18**, 1185-+, doi:10.1038/ncb3415 (2016).
- 24 Haahr, P. *et al.* Activation of the ATR kinase by the RPA-binding protein ETAA1. *Nat Cell Biol* **18**, 1196-1207, doi:10.1038/ncb3422 (2016).
- 25 Theis, M. & Buchholz, F. MISSION esiRNA for RNAi screening in mammalian cells. *J Vis Exp*, doi:10.3791/2008 (2010).
- 26 Matsuoka, S. *et al.* ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* **316**, 1160-1166, doi:10.1126/science.1140321 (2007).
- 27 Blasius, M. *et al.* A phospho-proteomic screen identifies substrates of the checkpoint kinase Chk1. *Genome Biol* **12**, doi:ARTN R7810.1186/gb-2011-12-8-r78 (2011).
- 28 Campbell, A. E. *et al.* NuRD and CAF-1-mediated silencing of the D4Z4 array is modulated by DUX4-induced MBD3L proteins. *Elife* **7**, doi:ARTN e310210.7554/eLife.31023 (2018).
- 29 Rodriguez-Terrones, D. *et al.* A molecular roadmap for the emergence of early-embryonic-like cells in culture. *Nat Genet* **50**, 106-+, doi:10.1038/s41588-017-0016-5 (2018).
- 30 Ufer, C. The biology of the RNA binding protein guanine-rich sequence binding factor 1. *Curr Protein Pept Sci* **13**, 347-357 (2012).
- 31 Mertins, P. *et al.* Ischemia in tumors induces early and sustained phosphorylation changes in stress kinase pathways but does not affect global protein levels. *Mol Cell Proteomics* **13**, 1690-1704, doi:10.1074/mcp.M113.036392 (2014).
- 32 Abad, M. *et al.* Reprogramming in vivo produces teratomas and iPS cells with totipotency features. *Nature* **502**, 340-+, doi:10.1038/nature12586 (2013).
- 33 Morgani, S. M. *et al.* Totipotent Embryonic Stem Cells Arise in Ground-State Culture Conditions. *Cell Rep* **3**, 1945-1957, doi:10.1016/j.celrep.2013.04.034 (2013).

- 34 Ruzankina, Y. *et al.* Deletion of the developmentally essential gene ATR in adult mice leads to age-related phenotypes and stem cell loss. *Cell Stem Cell* **1**, 113-126, doi:10.1016/j.stem.2007.03.002 (2007).
- 35 Santos, M. A. *et al.* DNA-damage-induced differentiation of leukaemic cells as an anti-cancer barrier. *Nature* **514**, 107-+, doi:10.1038/nature13483 (2014).
- 36 Schneider, L. *et al.* DNA Damage in Mammalian Neural Stem Cells Leads to Astrocytic Differentiation Mediated by BMP2 Signaling through JAK-STAT. *Stem Cell Reports* **1**, 123-138, doi:10.1016/j.stemcr.2013.06.004 (2013).
- 37 Feng, Q. *et al.* A feedback loop between nonsense-mediated decay and the retrogene DUX4 in facioscapulohumeral muscular dystrophy. *Elife* **4**, doi:ARTN e0499610.7554/eLife.04996 (2015).
- 38 Zhang, Q. *et al.* Tcstv1 and Tcstv3 elongate telomeres of mouse ES cells. *Sci Rep* **6**, 19852, doi:10.1038/srep19852 (2016).
- 39 Dan, J. M. *et al.* Zscan4 Inhibits Maintenance DNA Methylation to Facilitate Telomere Elongation in Mouse Embryonic Stem Cells. *Cell Rep* **20**, 1936-1949, doi:10.1016/j.celrep.2017.07.070 (2017).
- 40 Ding, H. *et al.* Characterization of a double homeodomain protein (DUX1) encoded by a cDNA homologous to 3.3 kb dispersed repeated elements. *Hum Mol Genet* **7**, 1681-1694, doi:DOI 10.1093/hmg/7.11.1681 (1998).
- 41 Peng, M. *et al.* Opposing Roles of FANCI and HLF Protect Forks and Restrain Replication during Stress. *Cell Rep* **24**, 3251-3261, doi:10.1016/j.celrep.2018.08.065 (2018).
- 42 de Klein, A. *et al.* Targeted disruption of the cell-cycle checkpoint gene ATR leads to early embryonic lethality in mice. *Curr Biol* **10**, 479-482 (2000).
- 43 Eckersley-Maslin, M. *et al.* Dppa2 and Dppa4 directly regulate the Dux driven zygotic transcriptional programme. *bioRxiv*, doi:citeulike-article-id:14642477 (2018).
- 44 Costanzo, V., Bardelli, A., Siena, S. & Abrignani, S. Exploring the links between cancer and placenta development. *Open Biol* **8**, doi:ARTN 18008110.1098/rsob.180081 (2018).
- 45 Ferretti, C., Bruni, L., Dangles-Marie, V., Pecking, A. P. & Bellet, D. Molecular circuits shared by placental and cancer cells, and their implications in the proliferative, invasive and migratory capacities of trophoblasts. *Human Reproduction Update* **13**, 121-141, doi:10.1093/humupd/dml048 (2007).
- 46 Costanzo, V., Bardelli, A., Siena, S. & Abrignani, S. Exploring the links between cancer and placenta development. *Open Biol* **8**, doi:10.1098/rsob.180081 (2018).
- 47 Yasuda, T. *et al.* Recurrent DUX4 fusions in B cell acute lymphoblastic leukemia of adolescents and young adults (vol 48, pg 569, 2016). *Nat Genet* **48**, 1591-1591 (2016).
- 48 Yoshimoto, T. *et al.* CIC-DUX4 Induces Small Round Cell Sarcomas Distinct from Ewing Sarcoma. *Cancer Research* **77**, 2927-2937, doi:10.1158/0008-5472.Can-16-3351 (2017).
- 49 Preussner, J., Zhong, J., Looso, M., Braun, T. & Kim, J. Connect-four: genomic analyses of regenerating stem cells identifies zygotic Dux factors as tumor initiators. *Mol Cell Oncol* **6**, 1565469, doi:10.1080/23723556.2019.1565469 (2019).

- 50 Chew, G. L. *et al.* DUX4 Suppresses MHC Class I to Promote Cancer Immune Evasion and Resistance to Checkpoint Blockade. *Developmental cell*, doi:10.1016/j.devcel.2019.06.011 (2019).
- 51 Bartek, J., Mistrik, M. & Bartkova, J. Thresholds of replication stress signaling in cancer development and treatment. *Nat Struct Mol Biol* **19**, 5-7, doi:10.1038/nsmb.2220 (2012).
- 52 Bartkova, J. *et al.* DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature* **434**, 864-870, doi:10.1038/nature03482 (2005).
- 53 Dobbstein, M. & Sorensen, C. S. Exploiting replicative stress to treat cancer. *Nature Reviews Drug Discovery* **14**, 405-423, doi:10.1038/nrd4553 (2015).
- 54 Macheret, M. & Halazonetis, T. D. DNA Replication Stress as a Hallmark of Cancer. *Annual Review of Pathology: Mechanisms of Disease, Vol 10* **10**, 425-448, doi:10.1146/annurev-pathol-012414-040424 (2015).
- 55 Nieto-Soler, M. *et al.* Efficacy of ATR inhibitors as single agents in Ewing sarcoma. *Oncotarget* **7**, 58759-58767, doi:10.18632/oncotarget.11643 (2016).
- 56 Smith, Z. D. *et al.* Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer. *Nature* **549**, 543-547, doi:10.1038/nature23891 (2017).
- 57 Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* **34**, 525-527, doi:10.1038/nbt.3519 (2016).
- 58 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
- 59 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).
- 60 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930, doi:10.1093/bioinformatics/btt656 (2014).
- 61 Sonesson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* **4**, 1521, doi:10.12688/f1000research.7563.2 (2015).
- 62 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47, doi:10.1093/nar/gkv007 (2015).
- 63 McCarthy, D. J., Chen, Y. S. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* **40**, 4288-4297, doi:10.1093/nar/gks042 (2012).
- 64 Wu, D. & Smyth, G. K. Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic Acids Res* **40**, e133, doi:10.1093/nar/gks461 (2012).
- 65 Hanzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-Seq data. *Bmc Bioinformatics* **14**, doi:Artn 710.1186/1471-2105-14-7 (2013).
- 66 Wu, D. *et al.* ROAST: rotation gene set tests for complex microarray experiments. *Bioinformatics* **26**, 2176-2182, doi:10.1093/bioinformatics/btq401 (2010).
- 67 Khan, A. *et al.* JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res* **46**, D1284, doi:10.1093/nar/gkx1188 (2018).

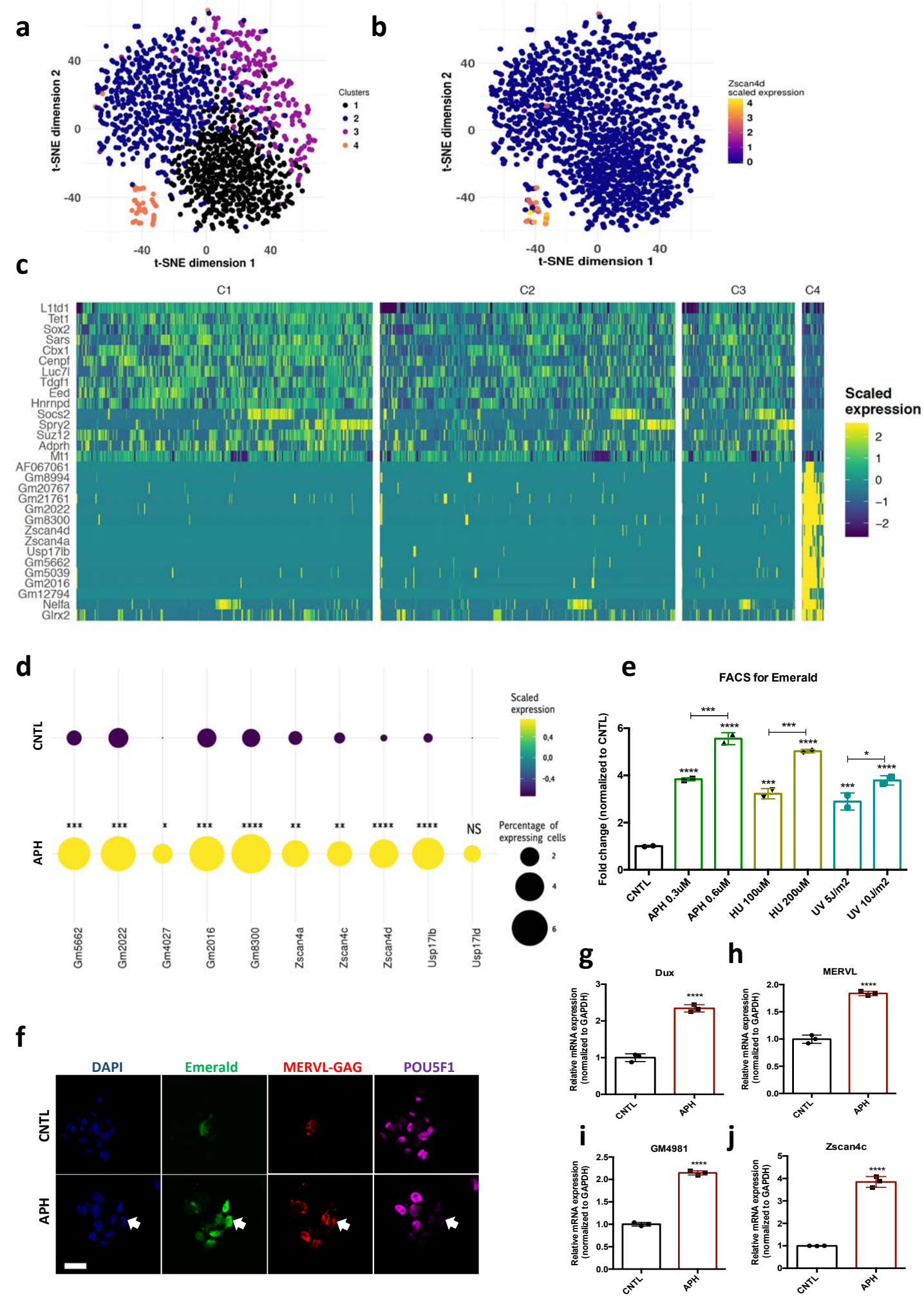
- 68 Tanaka, S., Kunath, T., Hadjantonakis, A. K., Nagy, A. & Rossant, J. Promotion of trophoblast stem cell proliferation by FGF4. *Science* **282**, 2072-2075, doi:10.1126/science.282.5396.2072 (1998).
- 69 Ng, R. K. *et al.* Epigenetic restriction of embryonic cell lineage fate by methylation of Elf5. *Nat Cell Biol* **10**, 1280-1290, doi:10.1038/ncb1786 (2008).
- 70 Soffientini, P. & Bachi, A. STAGE-diging: A novel in-gel digestion processing for proteomics samples. *J Proteomics* **140**, 48-54, doi:10.1016/j.jprot.2016.03.038 (2016).
- 71 Kall, L., Canterbury, J. D., Weston, J., Noble, W. S. & MacCoss, M. J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods* **4**, 923-925, doi:10.1038/Nmeth1113 (2007).
- 72 Keller, A., Nesvizhskii, A. I., Kolker, E. & Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* **74**, 5383-5392, doi:10.1021/ac025747h (2002).
- 73 Nesvizhskii, A. I., Keller, A., Kolker, E. & Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* **75**, 4646-4658, doi:10.1021/ac0341261 (2003).

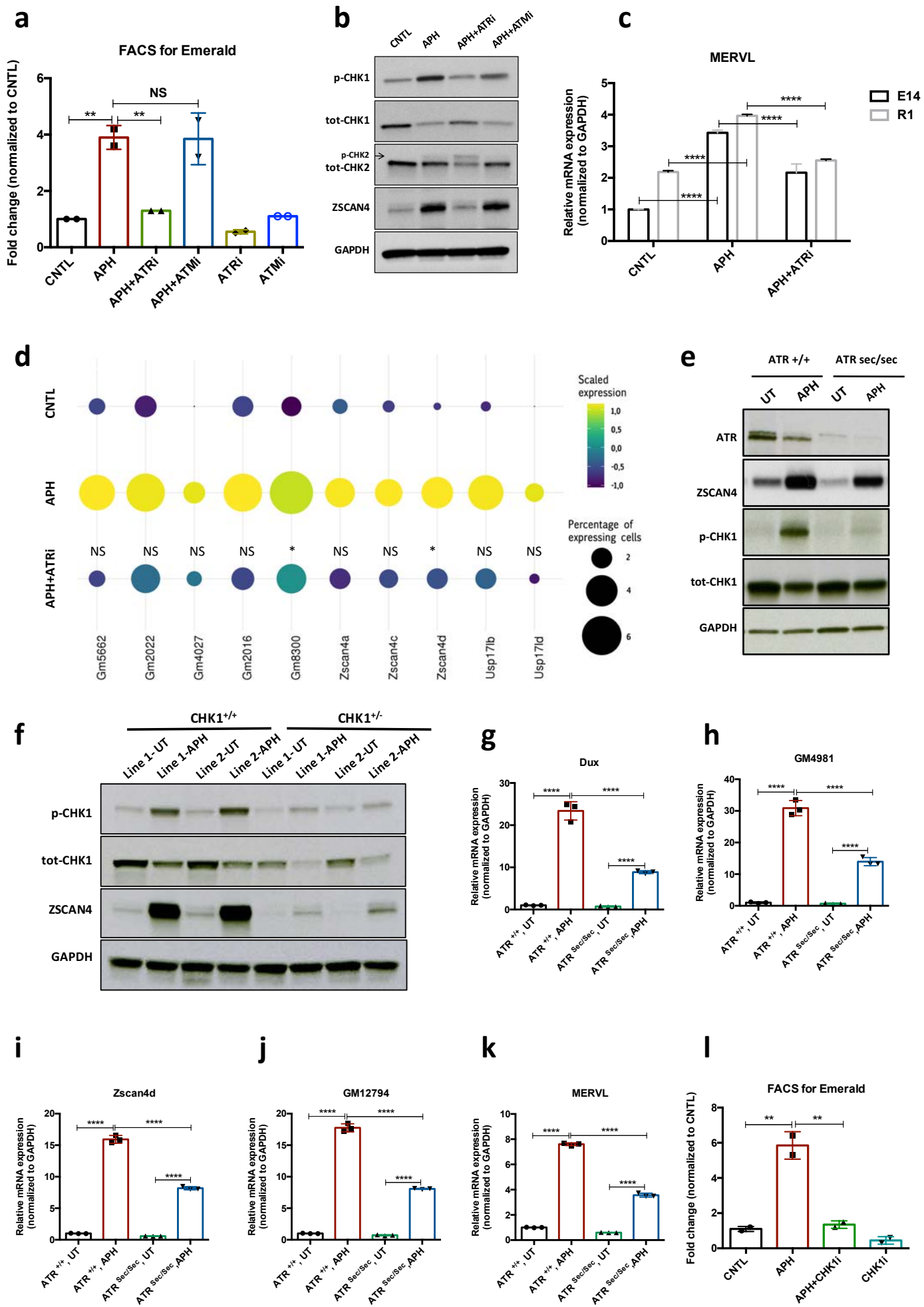
Acknowledgments

We thank Anna De Antoni, Ambra Belpietro, and Eleonora Verga for technical support, Christopher Bruhn for helping with ATR Seckel mice colony maintenance and Federica Zanardi and Fabio Iannelli for helping with RNA design. The Niels Mailand lab (The Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen, Denmark) for sharing ETAA1-AAD cDNA and advice. Alberto De Iaco and Didier Trono (School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland) for sharing the *Dux* KO ESCs. Geppino Falco (Istituto di Ricerche Genetiche Gaetano Salvatore Biogem scarl, Ariano Irpino, Italy) for sharing *pZscan4*-Emerald ESCs. Austin Smith (Department of Biochemistry, University of Cambridge, Cambridge, UK) for sharing E14 ESCs. Yang Xu (University of California San Diego for sharing ATMskO ESC. This work was funded by the Associazione Italiana per la Ricerca sul Cancro (AIRC), AIRC 5xmille special program, a European Research Council (ERC) consolidator grant (614541), the Giovanni Armenise-Harvard foundation award awarded to Vincenzo Costanzo and research fellowships awarded by the Umberto Veronesi Foundation (FUV) and by FIRC - Fondazione Italiana per la Ricerca sul Cancro (The triennial “Mario e Valeria Rindi” Post-doctoral research fellowship, (18112)) to Sina Atashpaz. Work in A.J.L.-C lab was supported by grants from the Danish Cancer Society (KBVU-2014), Danish Council for Independent Research (Sapere Aude, DFF Starting Grant 2014), the European Research Council (ERC-2015-STG-679068) and the Danish National Research Foundation (DNRF115). The authors declare no financial conflict of interest related to this work.

Author contributions

S.S.Sh. and S.A. performed the most of experiments including single-cell transcriptional profiling, high-throughput RNA sequencing, characterization of ATR Seckel and CHK1 haploinsufficient ESCs, cloning, candidate-based siRNA screening, generation of inducible ESCs, RNA pull-down and establishment of *in vitro* trophoctoderm differentiation. E. Albers, J. M., and A.J. L-C contributed to the generation of CHK1 haploinsufficient ESCs and the mice chimera experiments. E.S. and F.F. performed computational analysis on RNA-sequencing data. S.M. performed and analyzed the Drop-seq single cell gene profiling. P.S. and A.B. performed mass spectrometry analysis. E. Allievi established ATR Seckel ESCs, generated MEFs and helped with mice embryo experiments. N.A. performed RNA isolation and qPCR on mice embryos, helped with the characterization of ATR Seckel and CHK1 haploinsufficient ESCs and establishment of *in vitro* differentiation protocol. A.G. helped with characterization of CHK1 haploinsufficient ESCs, performed *in vitro* differentiation experiments and analyzed the results. G.F., helped with siRNA screening. O.F.C., shared ATR-Seckel mice and p53 KO cells and helped with technical analysis. C.T. and V. Cancila, performed immunostaining of mouse embryos. S.A., S.S.Sh., and V.C. conceived, designed and analyzed the study and wrote the manuscript. All authors commented and edited the manuscript.





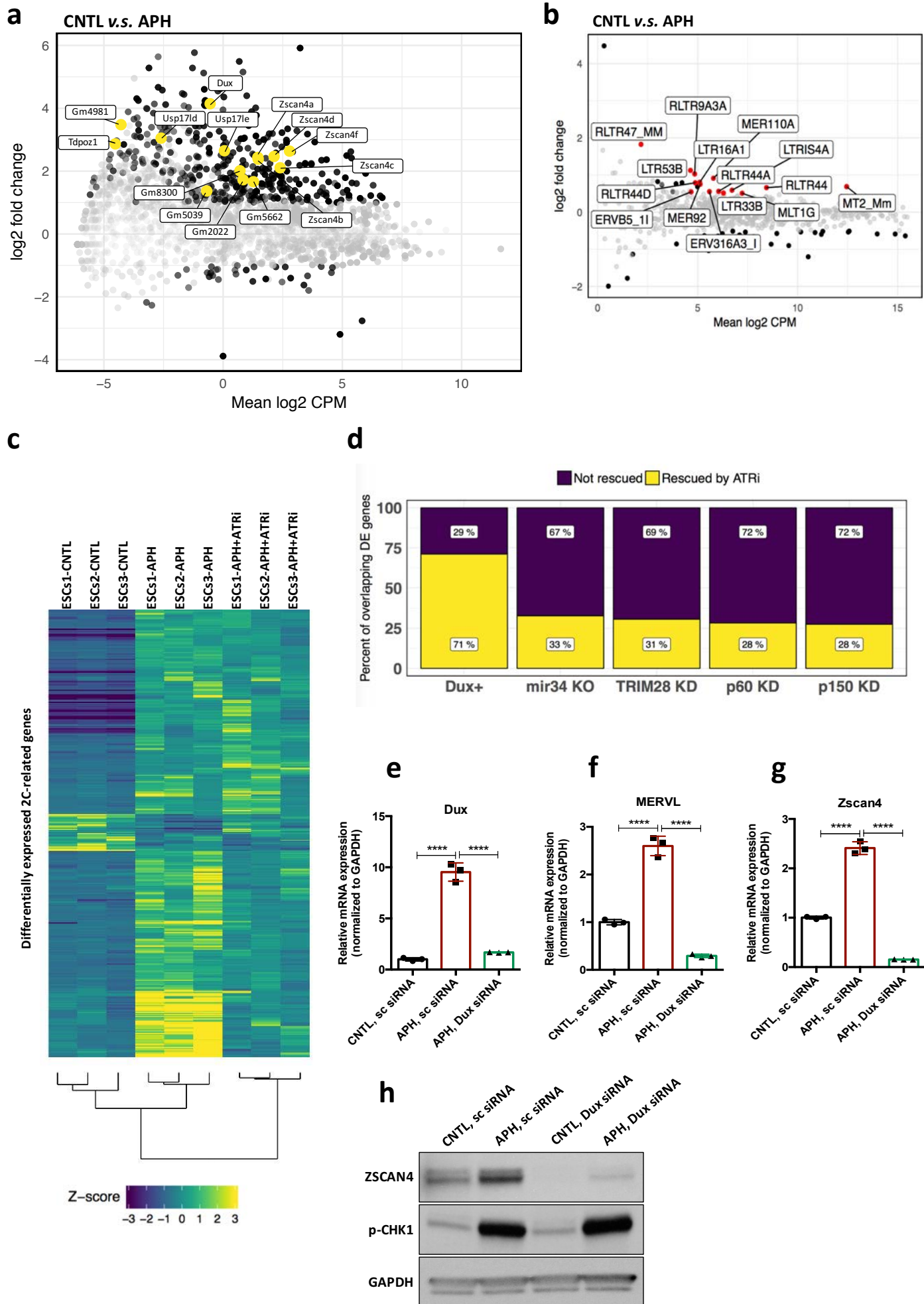
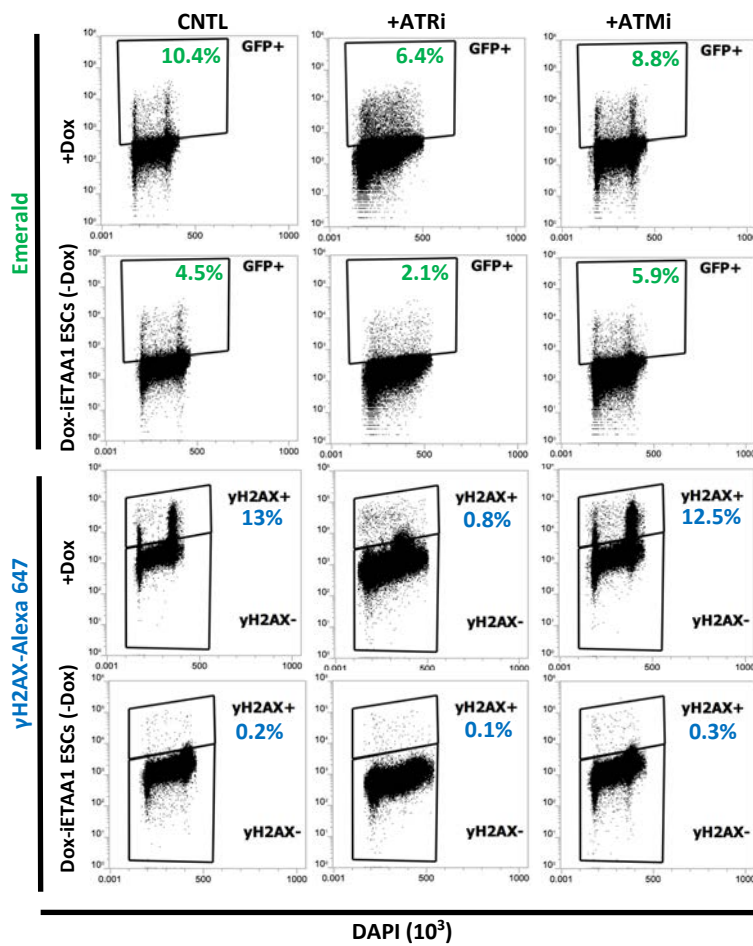
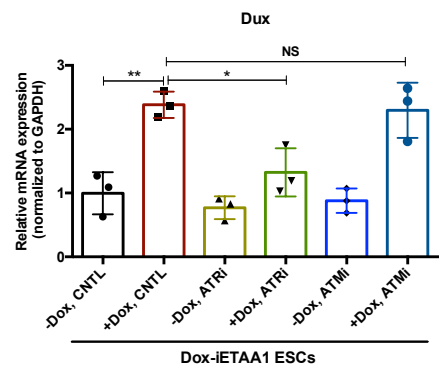


Figure 4

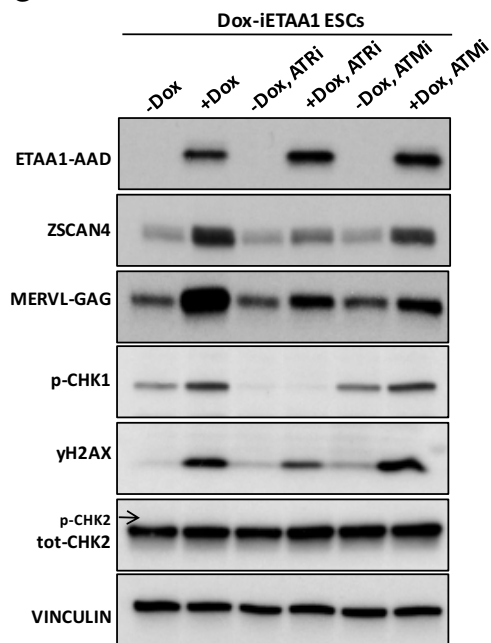
a



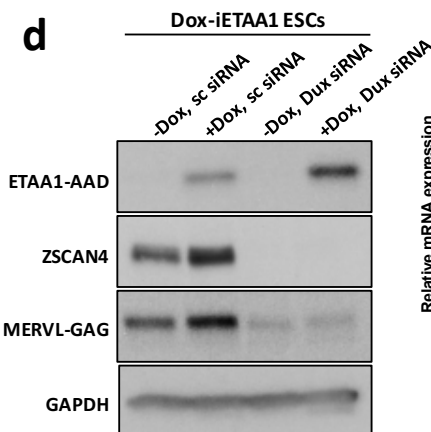
b



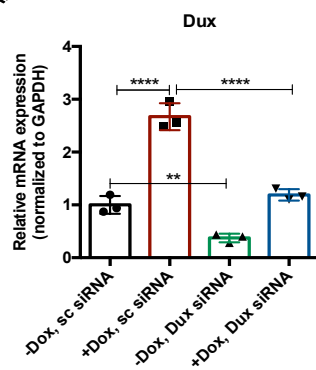
c



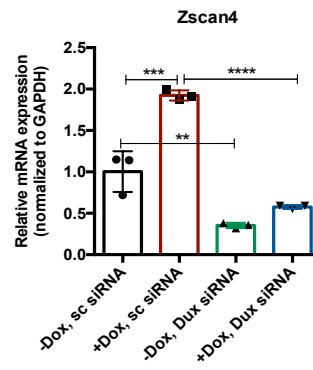
d



e



f



g

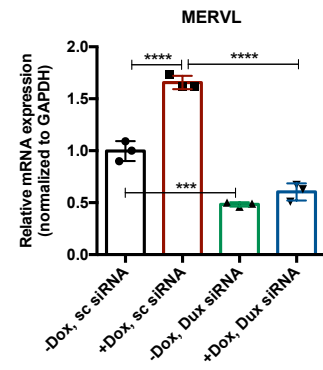


Figure 5

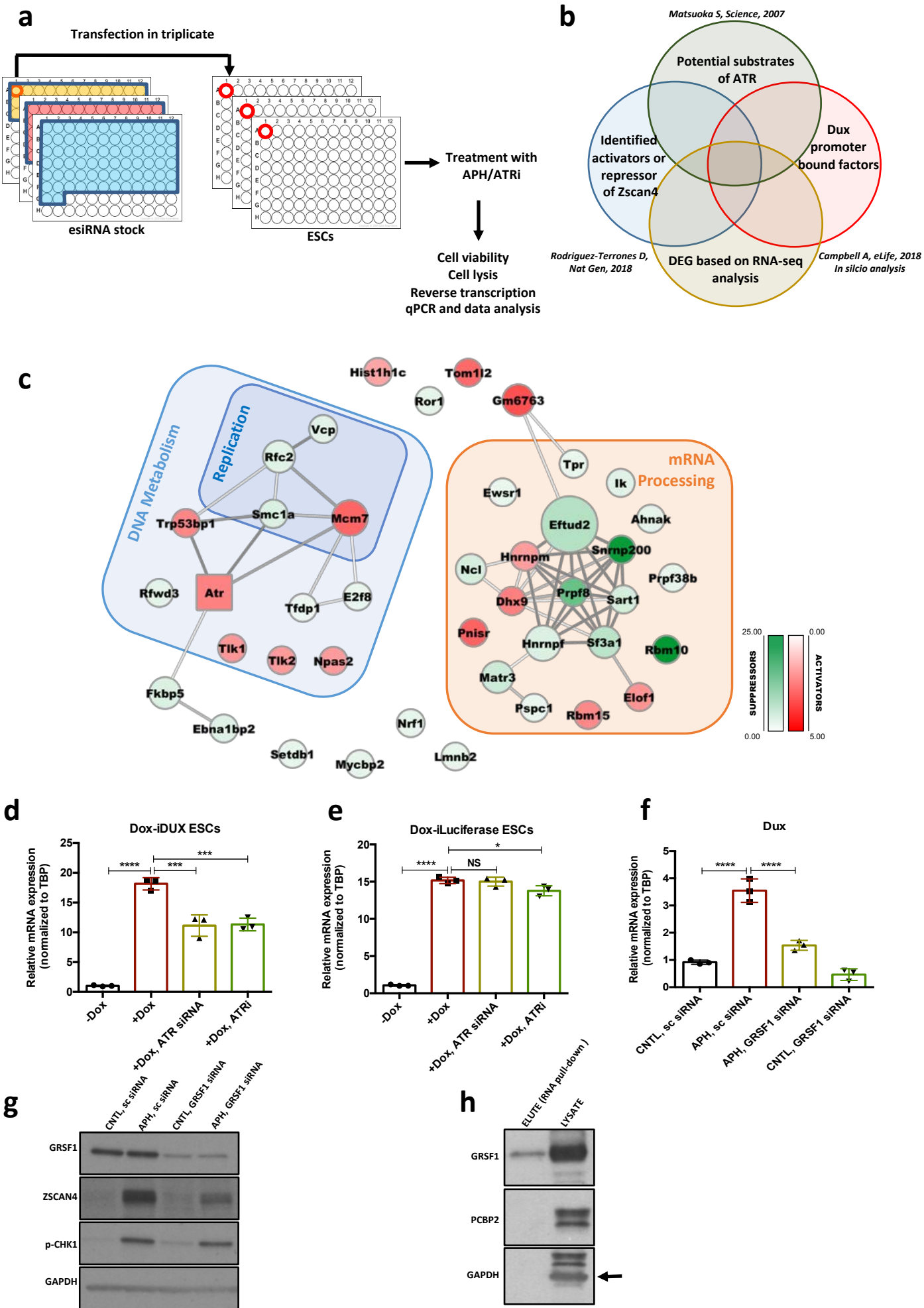


Figure 6

