TDG orchestrates ATF4-dependent gene transcription during retinoic acid-induced cell fate acquisition

Marion Turpin[1#], Thierry Madigou[1#], Maud Bizot[1], Rachael Acker[1,3], Erwan Watrin[1], Gérard Benoît[1], Katie Sawvell[1,4], Stephane Avner[1], Gaëlle Palierne[1], Cynthia Fourgeux[2], Martin Braud[2], Jeremie Poschmann[2], Christine Le Péron[1], Gilles Salbert[1,*]

[1]Univ Rennes, CNRS, Institut de Génétique et Développement de Rennes, UMR 6290, F-35000 Rennes, France

[2]CHU-Nantes, INSERM, Centre de Recherche en Transplantation et Immunologie UMR1064, Nantes Université, Nantes, France

[3]Present address: University of Pennsylvania, USA, rachael.acker@pennmedicine.upenn.edu

[4]Present address: Cincinnati Children's Hospital Medical Center, USA, Katie.sawvell@cchmc.org

[#] These authors contributed equally

*Corresponding author: gilles.salbert@univ-rennes1.fr, tel (+33) 223 236 625

Short title: TDG regulates ATF4-dependent transcription

Key words: TDG, ATF4, integrated stress response, mTORC1

ABSTRACT

During development, cell differentiation is associated to large-scale modifications in the methylome, which require the engagement of an active DNA demethylation machinery including Ten-Eleven-Translocation enzymes for oxidation of 5-methylcytosine and the T:G mismatch DNA glycosylase (TDG) for removal of the oxidized bases. Despite this well-defined molecular function, the biological output of TDG activity remains elusive. Here we combined transcriptomic and epigenomic approaches in TDG knock-out embryonal carcinoma cells, an epiblast stem-like cell model, to decipher TDG function in pluripotent cells and their retinoic acid-induced differentiated progeny. We determined that TDG activity is balancing differentiation in favor of a neural fate at the expense of a cardiac mesoderm fate. This process is associated with a sustained activity of a large set of ATF4-dependent genes in relation with a TDG-mediated nucleosome positioning at promoters and in conjunction with a TDG-dependent regulation of the mammalian target of rapamycin complex 1. These observations highlight the central role of TDG in cell differentiation and support a model linking metabolic reprogramming to cell fate acquisition.

INTRODUCTION

During differentiation, acquisition of a given cell identity relies on the establishment of specific gene expression profiles that result from changes in gene expression at the genome scale. These transcriptional changes involve both the remodeling and the covalent modifications of chromatin at various sites along the genome, among which stands DNA methylation. Steady-state genome-wide methylation patterns are a crucial layer of transcription regulation and result from the equilibrium between methylation reactions catalyzed by DNA methyltransferases (DNMTs) and demethylation ones, which rely on the Ten-Eleven-Translocation (TET) and

thymidine:guanidine mismatch thymine DNA glycosylase (TDG) enzymes (reviewed in Wu and Zhang, 2017). The transfer by DNMTs of a methyl group from S-adenosyl-L-methionine to cytosine at position 5 (5mC) occurs in the context of CpG dinucleotides and exerts a negative effect on gene expression. Methylated CpG (5mCpG) represses transcription by preventing or perturbing directly the binding of transcription factors to DNA through steric hindrance, and by its ability to recruit methyl-DNA binding proteins and associated repressive complexes (Nan et al., 1998; Kaluscha et al., 2022).

TDG was initially discovered as an enzyme catalysing removal of U and T mismatched with G in the genome, thereby participating in the maintenance of genome stability (Neddermann and Jiricny, 1993; 1994) and its role has expanded since then and the discovery of an active DNA demethylation pathway that involves TDG and TET enzymes (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). During this demethylation process, TETs iteratively oxidize 5mC into 5-hydroxy-mC (5hmC), 5-formyl-C (5fC) and 5-carboxyl-C (5caC), collectively referred to as oxi-mCs. The two more oxidised forms, 5fC and 5ca, are excised by TDG that hydrolyses the sugar-base bond, creating an abasic site that is then converted into an intact, unmodified cytosine by the base excision repair machinery. Newly incorporated base can then be targeted for *de novo* methylation by DNA methyltransferases (Maïti and Drohat, 2011). One crucial aspect of DNA methylation resides in the dynamic state that exists between modified and unmodified forms, which varies among different types of genomic regions. Of important note, the turnover of 5mC is notably high at enhancers, a particular class of gene regulatory elements that physically contact distant promoters for expression of cognate genes. As enhancer/promoter wiring plays an essential role in cell specific gene expression patterns, it is tempting to consider that such highly dynamic methylation of CpGs at enhancers as an indicator of its direct involvement in programming and acquisition of cell identity (Sérandour et al., 2012; Caron et al., 2015; Mahé et al., 2017; Ginno et al., 2020). As an actor of the demethylation pathway, TDG has been involved in this cyclic methylation/demethylation process in breast cancer cells and was later extended to embryonic stem cells (ESCs - Métivier et al., 2008; Rulands et al. 2018; Parry et al., 2021). Furthermore, TDG interacts with DNMT3A and DNMT3B, resulting in a decreased catalytic activity of both methyltransferases (Li et al., 2007; Boland and Christman, 2008; Sandoval et al., 2019). Hence, the opposing activities of DNMTs, TETs and TDG finely tune methylation dynamics that allow proper cell differentiation and identity acquisition. Besides this function in CpG methylation cycle, TDG has also been shown to physically interact with the histone acetyltransferase complex CBP () and with a number of transcription factors such as p53 and nuclear receptors, including retinoic acid (RA) receptors (RARs), and thereby participates in rendering chromatin accessible to DNA binding factors (Um et al., 1998; Tini et al., 2002; Onabote et al., 2022; Aranda et al., 2023), thereby. Hence, TDG is central to mechanisms regulating gene expression during cell-type specification processes, including those triggered by retinoic acid (RA). Nonetheless, the exact role of TDG needs to be further investigated.

The RA/RAR pathway exert pleiotropic effects during development in particular on cell fate commitment and identity acquisition processes, some of which can be recapitulated *in vitro*. Notably, RA treatment is able to trigger neural differentiation of ESCs as well as of embryonal carcinoma cells (ECCs - Jones-Villeneuve et al., 1982; Okada et al., 2004). The role of TDG in RA-induced neural differentiation has been previously investigated through *Tdg* inactivation approaches both in mESCs and in mouse primary neural stem cells (Cortazar et al., 2011;

Wheldon et al., 2014; Steinacher et al. 2019). Unfortunately, massive cell death occurred in both models early during neural differentiation induced by RA, precluding so far the study of TDG function in neural differentiation process. To circumvent these limitations, we took advantage of mouse ECCs called P19 (McBurney, 1993), because they present genic and morphological characteristics of epiblast stem cells (EpiSCs - Han et al., 2013) and are known as an appropriate alternative model to ESCs and EpiSCs when high cell death is observed (Magnúsdóttir et al., 2013). Using these P19 ECCs as a model, we succeeded in establishing a viable *Tdg*-KO model cell line that withstands RA treatment, providing us with a unique opportunity to study the role of TDG during cell differentiation. Using this model, we established the related impact of TDG inactivation on the transcriptome of cells treated with RA, demonstrating that TDG is required for balancing cell differentiation towards a neural fate at the expense of a cardiac mesoderm one. In addition, *Tdg* inactivation led to changes in nucleosome positioning and to a down-regulation of ATF4-target genes involved in adaptation to stress, which revealed that TDG is required to maintain ATF4-dependent gene transcription. ATF4 is a bZIP transcription factor that confers to cells the adaptability to various stress conditions like amino-acid (aa) deprivation and endoplasmic reticulum stress, through direct binding to promoters and co-activation by Lsd1 and p300/CBP of a large set of genes grouped into the so-called integrated stress response (ISR) pathway (Lassot et al., 2005; Pakos-Zebrucka et al., 2016; Faletti et al., 2021). We finally report that TDG maintains ATF4 target gene expression through regulation of the mTORC1 signaling complex. We propose that TDG regulates cell fate acquisition through diverse molecular mechanisms converging on ATF4 activity.

Materials and Methods

Cell Culture

P19 mouse embryonal carcinoma cells (ECCs) were grown and differentiated as described (Sérandour et al. 2012). Briefly, cells were exponentially grown in cell culture plastic dishes in 10% FCS DMEM containing penicillin and streptomycin. To obtain aggregates, cells were grown in bacterial grade plastic dishes. Neurons were observed 4 days after replating aggregates treated with 1 μM retinoic acid (RA, Sigma-Aldrich, reference R2625) for 48 hours. RA-induced aggregates in monolayer cultures were quantified by counting under phase-contrast microscopy. For mTOR inhibition, non-confluent adherent cultures of wt and TDG-null P19 ECCs seeded in 6-well plates (100,000 cells per well) were treated with 20 nM rapamycine (ApexBio, reference A8167) for 16 hours. halofuginone (8 nM, Supelco, reference 32481) or L-proline (0.5 mM, Sigma-Aldrich, reference P0380) were applied to monolayers of ECCs for 48 hours. MTT assays were run as previously described (Laurent et al., 2022).

Immunofluorescence microscopy

Wild type and TDG-null P19 ECCs were grown on glass coverslips and treated with 1 μM of RA for 72 hours, fixed with prechilled 100 % methanol at -20°C for 12 minutes. Fixed cells were then permeabilised with 0.02 % Triton-X100 in 1X PBS for ten minutes, blocked with 3 % Bovine Serum Albumin in 1X PBS containing 0.01% Triton-X100 for one hour and incubated with unconjugated primary antibodies diluted in blocking solution overnight at 20°C. Following three ten-minute washes in 0.01% Triton-X100 in 1X PBS, samples were incubated with 488- and 555-Alexa coupled secondary antibodies (1/500 dilution) for one hour, washed, DNA-counterstained with 0.2 µg/mL

of DAPI and mounted on glass slides using ProLong Gold antifading reagent. Rabbit anti-Smooth Actin Muscle antibodies were from Abcam (ab5694, 1/300), mouse anti-Pax6 from Invitrogen (13B10-1A10, 1/100). Images were acquired on a BX51 Olympus fluorescence microscope equipped with a Sola Light Engine light source, a Photometrics Prism BSI camera and Micromanager control and acquisition software.

CRISPR-Cas9-mediated *Tdg* knock-out

Knock-out was performed using the TDG mouse Gene Knock-Out kit (Origene, KN317363). This kit contained two gRNAs that targeted the first exon of TDG gene and a donor vector that contained a GFP-Puromycin cassette to facilitate the screening process. Briefly, wt P19 cells were seeded in P100 Petri dishes transfected with 1 µg of either gRNA vectors and 1µg of the donor vector using the JetPEI transfection reagent (Polyplus-transfection). Five passages after the transfection, Puromycin was added in the medium (0.5µg/ml). After the ninth passage, fluorescence acquired cell sorting was carried out to isolate GFP+ cells (LSR Fortessa X-20, Becton Dickinson). A limiting dilution was then performed to obtain clones from single cells. These clones were analyzed by RT-PCR and PCR on genomic DNA (primers are listed in Supplementary Table 1) to assess for the absence of TDG mRNA and the presence of the GFP-Puromycin cassette respectively.

CRISPR-Cas9-mediated *Cyp26a1* knock-out

Two guide RNAs (gRNAs) flanking the region of interest were designed using CRISPOR (http://crispor.tefor.net/) to be further introduced either into pX458 or pX459 vectors from Addgene (each vector contains one gRNA). pX458 and pX459 vectors were previously digested 2hours at 37°C by BbsI (NEB, R0539) and dephosphorylated 10 minutes at 37°C using Quick CIP (NEB, M0525). The vectors were then purified with PCR CleanUp kit (Macherey Nagel, 740609) and complete BbsI digestion was checked by agarose gel electrophoresis. Before ligation, gRNAs were phosphorylated with a T4 Polynucleotide Kinase (NEB, M0201) for 30 minutes at 37°C followed by inactivation at 95°C for 5 minutes. The ligation of vectors with gRNA inserts was performed overnight at 4°C using T4 DNA ligase (NEB, M0202). The ligation products were then transformed, amplified into NEB10 bacteria and plated on LB+Amp agar plates overnight at 37°C. The correct cloning was verified by PCR and agarose gel electrophoresis. For transfection, $7.5.10^5$ wt P19 cells were seeded into P100 Petri dishes and transfected for 24 hours with vectors containing the gRNAs, i.e., pX458, carrying a GFP coding sequence, and pX459 with a ratio 1:1.2. GFP positive cells were isolated using flow cytometry and sorted into a 96-well plate (LSR Fortessa X-20, Becton Dickinson). Clones were analyzed by PCR and KO was confirmed by Sanger sequencing.

MNase assay

For MNase-seq assays, $10\text{-}20.10^6$ cells were incubated for 10 min in 1% formaldehyde in PBS, washed twice and scraped in ice cold PBS and spun down at 100g for 10min at 4°C. Cells were re-suspended in digestion buffer (50 mM Tris-HCl pH7.6, 2 mM CaCl2, 0.2% NP40) containing protease inhibitors (complete EDTA-free, Roche) and warm-up to 37°C for 2min. 4U of MNase were added to wt ECC chromatin and the reaction mixture was incubated at 37°C for 5 min. For TDG-null ECCs, 2U of MNase were added, due to the limited amount of recovered mono-nucleosomal DNA when using 4U of MNase. The reaction was stopped by adding EDTA to a final concentration

of 10mM. Cells were then incubated for 10 min on ice in SDS lysis buffer (1% SDS, 50 mM Tris pH8, 10 mM EDTA pH8). An equal volume of TE/RNase (10 mM Tris pH8, 0.1 mM EDTA pH8, 0.4 µg/ml RNase A) was added before incubating at 37°C for 2 hours. A Proteinase K digestion (1 mg/ml) was then performed at 55°C overnight. The mixture was then extracted twice with Phenol/chloroform and the DNA was precipitated 1 hour at -20°C in the presence of glycogen (20µg). After centrifugation, DNA pellets were washed by ethanol, dissolved in TE buffer (10 mM Tris pH8, 1 mM EDTA pH8) and purified using NucleoSpin Gel and PCR clean-up (Macherey-Nagel). DNA was loaded on a 2% agarose gel. The mono-nucleosome band was excised and the corresponding DNA was purified using MinElute Gel Extraction kit (Qiagen). Mono-nucleosomal DNA was further analyzed by deep sequencing. MNase-qPCR assays followed the same procedure except that increasing concentrations of MNase were used. After PCR amplification of selected nucleosomes, values were normalized to the amplification values of undigested purified genomic DNA. Primers are listed in Supplementary Table 1.

Chromatin Immunoprecipitation (ChIP)

ChIP assays of H3K27ac were run as previously described (Sérandour et al., 2012). For TDG and ATF4 ChIP assays, chromatin from wt ECCs was cross-linked using 1% formaldehyde for 10min at room temperature. Cells were rinced with cold PBS, harvested and lyzed in Farnham lysis buffer (5mM PIPES pH8, 85mM KCl, 0.5% NP40, Protease inhibitor cocktail). Nuclei were pelleted and resuspended in RIPA buffer (1xPBS, 1% NP40, 0.5% Sodium deoxycholate, 0.1% SDS, protease inhibitor cocktail). Chromatin fragmentation was performed by sonicating samples 14 min (30sec on/off cycles) using a Bioruptor (Diagenode). The fragmented chromatin was incubated at 4°C overnight either with a rabbit polyclonal antibody against mTDG (Gallais et al., 2007), or a rabbit monoclonal antibody against hATF4 (Cell Signaling, D4B8), in RIPA buffer. Complexes were recovered after incubation with 50 µl protein A–conjugated Sepharose bead slurry at 4°C. Beads were then washed four times by LiCl washing buffer (100 mM Tris pH7.5, 1% Sodium deoxycholate, 1% NP40, 500mM LiCl, protease inhibitor cocktail) and twice by TE buffer (10mM Tris HCl pH8, 1mM EDTA). Fragments were eluted with extraction buffer (1% SDS, 100mM NaHCO3). Cross-linking was reversed by overnight incubation at 65°C and DNA fragments were purified using NucleoSpin Gel and PCR Clean-up columns (Macherey-Nagel).

RNA preparation, reverse transcription and real-time PCR (qPCR)

RNA was isolated using Trizol reagent (Ambion, Life Technology) according to the manufacturer's protocol. Reverse transcription was performed using 1µg of total RNA as template, 200 U of the M-MLV Reverse Transcriptase (Invitrogen) and 250 ng of Pd(N6) random hexamers (Jena Bioscience). Real-time PCR was performed using SYBR Green Master Mix (Bio-Rad) on Bio-Rad CFX96 machines. All primers were designed with Primer3 (Untergasser et al., 2012) and purchased from Sigma (Supplementary Table 1).

RNA Sequencing Methodology

Total RNA was extracted using Trizol and RNA Integrity Number (RIN) score were determined with a 2100 Bioanalyzer (Agilent). The protocol for 3' digital gene expression profiling was carried out as previously described (Chaumette et al., 2022). Library preparation was initiated from 10 ng of total RNA per sample. mRNA poly(A)

tails were tagged with universal adapters, well-specific barcodes, and unique molecular identifiers (UMIs) during the template-switching reverse transcription process. Barcoded complementary DNAs (cDNAs) from multiple samples were pooled, amplified, and subjected to transposon fragmentation, focusing on enriching the 3' ends of cDNA. The resulting library, with fragment sizes ranging from 350 to 800 bp, underwent sequencing in a 100-cycle S2 run on the NovaSeq 6000 system at the Genomics Atlantic platform (Nantes). Raw sequencing data were deposited in GEO under accession number GSE262587. On average, each sample yielded approximately 5 million 75 bp single-end reads. Post-sequencing, data demultiplexing was performed using Illumina bcl2fastq. The resulting FASTQ files were analyzed using the 3' Sequencing RNA Profiling (SRP) pipeline (Charpentier et al., 2021). This methodology aligns with previous DGEseq analyses in various studies (Chaumette et al., 2022; Letellier et al., 2022; Ménoret et al., 2023). The SRP pipeline incorporates cutadapt for read trimming and bwa aln (version 0.7.17) for alignment to the mm10 Mus musculus reference genome and transcriptome. A custom Python script was employed to parse and count UMIs, leading to the generation of a raw matrix containing unique transcript counts. Gene annotation was based on the RefSeq database. The dataset encompassed a total of 26,214 genes, where 3,453 genes had zero reads across all samples, while 22,761 genes had at least one read in at least one sample. The tests for differential expression between two conditions were performed using the R package DESeq2 (Love et al., 2014) version 1.42.0. A gene was declared modulated if it displayed a significant difference between the two indicated conditions (the cut-off was fixed at 5.10^{-2} for adjusted P-value (Benjamini-Hochberg). MA-plot shows the shrunken log2 fold changes between the indicated conditions, using the adaptive shrinkage estimator from the ashr package (Stephens, 2016).

Library preparation, sequencing, and bioinformatics

All MNase-seq libraries were prepared using the TrueSeq ChIP Sample Prep kit (Illumina) and single-end sequenced using a HiSeq 1500 sequencing system at the "Human and Environmental Genomics" platform (now EcoGeno) (Rennes, France). Raw sequencing data were deposited in GEO under accession number GSE261602. Raw sequencing data from publicly available datasets used in this study were downloaded from GEO at NCBI. Sequencing reads were mapped to mm8 using Bowtie (Langmead et al., 2009) and processed by SAMtools (Li et al., 2009) to generate bam files followed by wig file generation with MACS (Zhang et al., 2008). Peak calling followed a previously described procedure (Sérandour et al., 2012). Sequencing reads from publicly available datasets were mapped and treated following the same procedure. For comparison, sequencing data were normalized with respect to the number of reads. Heatmaps were generated online with Cistrome (Liu et al., 2011; http://cistrome.org/ap/root). Motif enrichment in TDG ChIP-seq peaks was run using the SeqPos tool from Cistrome. Bar graphs (mean +/- SD, n=3 to 6) were generated with GraphPad Prism 5.0 and analyzed by unpaired t-test (Prism 5.0).

Total cell extract and immunoblot

Total cell extracts were prepared and immunoblotting experiments performed as described previously (Watrin et al, 2014). Rabbit antibodies against U2AF65 (sc-53942) were purchased from Santa-Cruz Biotechnologies,

p70S6K (# 9202) and T389p70S6K-Ph (# 9234) from Cell Signalling Technology. Rabbit anti-TDG antibodies were already described (Gallais et al., 2007). All antibodies used in this study were affinity-purified.

Results

TDG KO skews retinoic acid-induced differentiation of ECCs towards cardiac mesoderm

To thoroughly characterize the dynamics of RA-induced ECC differentiation, we first analyzed by 3'RNA-seq the time-dependent gene expression pattern of P19 ECCs grown as monolayers and treated with 1 μM RA. Expression profiles of selected differentiation markers are shown in supplementary figure 1A. Results indicated that, upon RA treatment, P19 ECCs lost expression of pluripotency markers and acquired a posterior fate as revealed by the rapid inhibition of *Otx2* and induction of *Cdx1* and *Hoxa1* (Supplementary Fig. 1A). Posteriorization was rapidly followed by the expression of neuromesodermal progenitor (NMP) markers, all of which except *Sox2* dropped after 24 hours of RA exposure while NPC markers rose dramatically (Supplementary Fig. 1A). In parallel, mesoderm progenitor cell (MPC) and paraxial mesoderm markers showed no to low induction (except for *Meox1*) whereas several cardiomyocyte markers were induced (Supplementary Fig. 1A). These data indicate that wt ECCs are prone to undergo neural differentiation upon RA treatment but that a fraction of the cells can adopt a cardiac mesoderm fate. Consistent with this observation, signaling pathways favoring NMP establishment and maintenance as well as mesoderm induction (*i.e.* Fgf, Wnt and Nodal; Henrique et al., 2015; Edri et al., 2019) were turned down at the time when NPC markers started to be expressed (Supplementary Fig. 1A). Hence, upon RA addition, the P19 ECC system recapitulates key steps of EpiSC differentiation observed *in vivo* during gastrulation and gives rise to a mixed population of NPCs and cardiomyocytes.

To interrogate the role of TDG in RA-induced differentiation of ECCs, the *TDG* gene was next inactivated by CRISPR/Cas9 in wt P19 ECCs (Supplementary Fig 1B) and the transcriptomes of control and RA-treated wt and TDG-null ECCs were determined by 3'-RNA-seq. Differential gene expression analysis between vehicle-treated wt and TDG-null ECCs grown as monolayers revealed 977 differentially expressed genes (DEGs, adjusted P value $<$ 0.05, Supplementary Table 2). Gene ontology (GO) enrichment analysis revealed no specific annotation for up-regulated genes and "RNA processing" and "amino acid metabolism" annotations for down-regulated genes (Supplementary Fig. 1C). Cells in monolayers treated with RA for 48 hours showed 700 genes expressed significantly higher and 609 genes expressed significantly lower (adjusted P value $<$ 0.05) in TDG-null cells compared to TDG-proficient cells (Fig. 1A, Supplementary Table 3). GO enrichment analysis of up- and down-DEGs (Log2 fold change (FC) $>$ 1 and adjusted P value $<$ 0.05) indicated skewing from neural to mesodermal differentiation and revealed an upregulation of cardiac mesoderm genes in RA-treated TDG-null cells (Fig. 1A,B). In particular, a decreased expression of NPC-specific markers like *Pax6* and *Dbx1* and an increase in mesoderm markers such as *Acta2*, *Myl9* and *Meox1* were observed, whereas the mixed-lineage marker *Meis1* showed no change in expression (Fig. 1A). Consistent with the RNA-seq data, RT-qPCR analysis after long-term RA treatment indicated that the observed reduced expression of neural markers after 48 hours of RA treatment in TDG-null cells compared to wt ECCs was not due to a delayed activation but, rather, to a switch from neural to mesodermal fate as revealed by changes in mRNA levels of the neural markers *Pax6* and *Irx3*, the cardiogenic gene *Wnt11* and the cardiac mesoderm marker *Acta2* (Fig. 1C).

Further, analysis of the 17 most significantly down-regulated genes revealed a network of known ATF4-regulated genes (including *Atf4* itself) involved in one-carbon metabolism as well as genes relevant to amino-acid (aa) metabolism, including aa biosynthesis, transport and tRNA charging (Fig. 1D,E). Interestingly, expression of the 3 top genes (*i.e.*, *Asns*, *Mthfd2* and *Psat1*, in red Fig. 1D) has been shown to depend on the catalytic activity of Tet1 and Tet2 in ESCs (Mulholland et al., 2020), suggesting that TDG acts in concert with and downstream of 5mC oxidation at these genes. The down-regulation of *Asns* in the context of an inhibition of neural differentiation in TDG-null ECCs is reminiscent of the negative influence of low Asns levels on the differentiation of neural stem cells and the association of *Asns* mutations with a microcephaly phenotype in human (Ruzzo et al., 2013; Appiah et al., 2023). Collectively, these data suggest that, in a model recapitulating EpiSC differentiation into NMPs followed by a bifurcation towards neural or mesoderm fates, TDG favors the acquisition of a neural fate, possibly through the regulation of ATF4-dependent gene expression.

Tdg KO enhances sensitivity to RA but does not phenocopy _Cyp26a1_ KO

The observed differentiation skewing in RA-treated TDG-null cells could be due either to differential gene regulation upon TDG depletion, or to a disequilibrium in cell-type survival. To answer this question, we first investigated cell viability through MTT assays that measure mitochondrial reductase activity. Consistent with a higher cell death in RA-treated TDG-null cells, results showed lower reductase activity in monolayers of TDG-null compared to wt ECCs (Fig. 2A). To evaluate a possible cell-type bias in cell death, we next examined expression levels of the neural marker PAX6 and the cardiac mesoderm marker ACTA2 through immunofluorescence in wt and TDG-null ECC monolayers treated with RA for 72 hours. Upon RA treatment, PAX6-positive small aggregates could be observed in wt ECCs but were mainly absent from TDG-null cells (Fig. 2B). Conversely, cells positive for cardiac actin (ACTA2) and well-spread onto glass coverslips were readily observed in RA-treated TDG-null cells whereas they were rarely observed in RA-treated wt ECCs (Fig. 2B). Hence the skewing in cell fate upon TDG depletion could reflect high cell death of PAX6-positive neural progenitors. Since formation of embryoid bodies or cell aggregates has been shown to increase neural differentiation of wt P19 ECCs (Jones-Villeneuve et al., 1982), cells were then grown in suspension, to favor aggregate formation, for 48 hours with RA before being re-plated on cell-culture plastic dishes. In these conditions, neurons could be observed at day 6 after RA-addition (Fig. 2C,D), indicating that a fraction of neural progenitors escaped cell death in the absence of TDG. These data indicate that TDG is required for the correct balance between mesoderm and neurectoderm differentiation of ECCs, possibly through controlling cell survival of neural progenitors.

Previous studies demonstrated that RA induces apoptosis concomitantly with neuronal differentiation of P19 ECCs (Horn et al., 1996; Fujita et al., 1999), suggesting that the observed high cell death in RA-treated TDG-null cells could be due to an excess in RA signaling. Excess RA signaling could result from a lack of CYP26A1 activity. Indeed, CYP26A1 catabolizes the first step of RA elimination by oxidizing RA into 4-oxoRA and, by doing so, reduces the intracellular levels of RA (Langton and Gudas, 2008). Consistent with that hypothesis, *Cyp26a1* expression was turned off in TDG-null cells (Fig. 1A). Hence, to investigate whether alterations in RA-signaling could be at the basis of the observed differentiation skewing, we generated *Cyp26a1*-KO ECCs (Supplementary Fig. 2A,B). Supporting a higher sensitivity to RA of TDG-null cells, RA-induced Meis1 expression as well as growth

of neurogenic aggregates was observed at lower RA concentrations compared to wt ECCs (Fig. 2E,F). Quantification of these aggregates showed that, whereas aggregates were rarely observed at 1 μM RA in TDG-null cells, probably due to due to excessive cell death, *Cyp26a1*-KO cells and wt ECCs generated a high number of aggregates (Fig. 2G). In agreement with this observation, RT-qPCR data obtained with two different *Cyp26a1*-KO clones treated with 1 μM RA did not show differentiation skewing (Fig. 2H). Hence, excessive RA signaling alone cannot be accounted for the high cell death observed in TDG-null cells. However, induction by RA of the neural pro-survival gene *Bdnf* was impaired in TDG-null cells but not in *Cyp26a1*-KO cells, suggesting that high cell death of TDG-null cells could be due to the combinatorial effect of a lack of BDNF and excessive RA-signaling (Fig. 2I). Finally, whereas the *Cyp26a1* gene was turned down in TDG-KO cells, *Tdg* expression was reciprocally down-regulated in *Cyp26a1*-KO clones, underlining a complex and unexpected cross-talk between the *Tdg* and *Cyp26a1* genes (Fig. 2J,K). Consistent with a down-regulation of *Tdg* expression in *Cyp26a1*-KO ECCs, lower levels of mRNAs from ATF4-dependent genes were observed in RA-treated *Cyp26a1*-KO compared to wt ECCs (Fig. 2L; Supplementary Fig. 2C,D,E). These data indicate that high levels of *Tdg* expression are required for a correct regulation of ATF4-dependent genes whereas low levels are sufficient for survival of neural progenitors. Hence, the differentiation skewing observed in RA-treated TDG-null cells could be partly due to cell-type specific cell death in the complete absence of TDG.

TDG is required for sustained transcription of stress-response genes in differentiating ECCs

Since we observed a significant down-regulation of known ATF4 target genes in RA-treated TDG-null cells compared to RA-treated wt ECCs (Fig. 1D), we next addressed the dynamics of ATF4 target gene expression and the relationship between TDG and ATF4 networks. The time-course response to RA treatment was first analyzed by 3'RNA-seq of wt ECC monolayers. Data revealed a marked trend of ATF4-target gene expression after RA addition to wt ECC monolayers, with a decrease in mRNA levels during the first 24 hours followed by a recovery in gene expression at 48 hours, which was confirmed by RT-qPCR assays (Fig. 3A, Supplementary Fig. 3A). As an illustration, solute carrier (*Slc*) genes that showed a decrease in expression at 24 hours followed by a recovery at 48 hours of RA treatment were ATF4 targets, whereas other *Slc* genes were not (Supplementary Fig. 3B,C). Such oscillations in gene expression were accompanied by comparable oscillations in the active chromatin mark H3K27ac (GEO dataset GSE82314) at ATF4 target promoters but not at the unregulated gene *Nampt* (Fig. 3B), suggesting that the observed variations in these mRNA levels are correlated with transcriptional regulation of the corresponding genes. To identify common TDG and ATF4 target genes, we next reanalyzed publicly available TDG (GSM134311, in ESCs) and ATF4 (GSM5440974, in cardiomyocytes) ChIP-seq datasets and evidenced that selected gene promoters from the top down-regulated gene list (Fig. 1) are bound by TDG and by ATF4 (Fig. 3C). These observations indicated a potential involvement of TDG in the transcriptional control of ATF4 target genes. Consistent with a positive role of TDG in transcription of these ATF4-bound genes in RA-treated cells, mRNA levels were decreased in 48h RA-treated TDG-null cells compared to wt ECCs (Fig. 3D). In addition, H3K27ac was significantly reduced at the +1 nucleosome of ATF4 target genes in RA-treated TDG-null cells compared to wt ECCs (Fig. 3E), further consolidating the hypothesis that TDG positively regulates ATF4 target genes at the transcriptional level.

The observation that RA induces changes in the 3D organization of ECCs growing as monolayers (Fig. 2F) and the fact that RA down-regulates ATF4 target genes such as *Asns* (Fig. 2L), both processes being reinforced in TDG-null cells, led us to hypothesize that cell-cell interactions directly impact ATF4 activity. To test this hypothesis, wt and TDG-null ECCs were grown either as monolayers or in aggregates and their transcriptome were analyzed. A reduction in expression of ATF4-target genes was observed when growing wt ECCs in aggregates in the absence of RA, an effect that was amplified either by TDG depletion (Fig. 3F,G) or by RA treatment of wt ECCs (Supplementary Fig. 3D, Supplementary Table 4, Supplementary Table 5). Collectively, these data indicate that ATF4 target genes are positively regulated by TDG in aggregates and that the transient RA-mediated down-regulation of these genes observed in neurogenic aggregates obtained by treating monolayers with RA is probably initiated both through the 3D reorganization of cells in culture and through direct RA-dependent transcriptional regulations (Supplementary Fig. 3E). Hence, we hypothesized that RA treatment of ECCs induces a transient stress state that triggers a TDG-dependent ATF4-mediated adaptative response.

Stress adaptation relies on the stimulation of the *Atf4* mRNA translation by the Gcn2 pathway (Harding et al., 2000). To test whether the GCN2-ATF4 pathway is engaged in wt or TDG-null cells, we treated cell monolayers for 48 hours either with Halofugigone, a drug that blocks the prolyl-tRNA synthetase, thus mimicking aa starvation, or with L-Proline that lowers the GCN2 activation by preventing the accumulation of uncharged tRNA (D'Aniello et al., 2015). Contrary to what has been observed in ESCs (D'Aniello et al., 2015), no significant change in *Asns* or *Psat1* mRNA levels was observed in these conditions, either in wt or in TDG-null ECCs, suggesting that ATF4-target genes are not under a major regulation by the GCN2-ATF4 pathway in ECCs (Supplementary Fig. 3F). On the other hand, *Atf4* mRNA stability and translation have also been shown to be up-regulated by the mechanistic target of rapamycin complex 1 (mTORC1) activity in mouse embryonic fibroblasts and various cancer cell lines (Torrence et al., 2021). Thus, wt and TDG null ECCs were next treated with the mTORC1 inhibitor rapamycin. RT-qPCR assays evidenced that mTORC1 controls ATF4-target genes in ECCs and suggested that TDG-null ECCs have a lower mTORC1 activity compared to wt cells (Fig. 3H). This was further demonstrated by western blots showing the lower phosphorylation of p70s6k, a direct target of mTORC1, in TDG-null cells *versus* wt ECCs (Fig. 3I). With the aim of understanding the functional relationship between TDG and mTORC1 activity, transcriptomic data were then interrogated and revealed that a number of indirect activators of mTORC1 (*i.e*. *Mgst1*, *Gpx7*, *Rpl22l1* and *Slc7a5*) showed lower gene expression in TDG-null compared to wt ECCs (Supplementary Fig. 3G). In addition, the mTORC1 inhibitor gene *Sesn1* was down-regulated in RA-treated wt ECCs but not in TDG-null ECCs (Supplementary Fig. 3G). Hence, these observed complex changes in gene expression patterns might in part explain the differential activity of mTORC1 in wt compared to TDG-null cells.

In addition, mTORC1 activity has been linked to α-ketoglutarate (αKG) metabolism in colorectal cancer cells, with the ability of L-2-hydroxyglutarate (L2HG) to activate ATF4 through the mTOR pathway (Tabata et al., 2023). Nonetheless, mRNA levels of LDHA, the enzyme transforming αKG into L2HG were similarly down-regulated upon RA treatment of both wt and TDG-null ECCs, indicating that this specific pathway is not at the basis of the observed differential mTORC1 activity (Supplementary Fig. 3H). To evaluate the impact of mTORC1 activity on cell differentiation, monolayers of wt ECCs were next treated with rapamycin 24 hours after RA addition. Whereas expression of the mixed-lineage marker *Meis1* and the early neural marker *Pax6* were not

affected, late neural markers like *Irx3* and *Dbx1* were down-regulated, and the cardiac mesoderm marker *Acta2* increased (Fig. 3J). Collectively, these data reveal that TDG ensures a sufficient level of mTORC1 activity to sustain the transcriptional activity of ATF4-regulated genes and that the mTOR pathway is essential for proper cell fate choice in RA-treated ECCs.

TDG is engaged at ATF4 target promoters and regulates nucleosome positioning

To gain further insight on the role of TDG in the control of gene expression we next sought to obtain a comprehensive view of TDG engagement at promoters. In this aim, we compared the genome wide distribution of TDG with TSSs and CpG islands (CGIs) in ESCs. Results indicated that 78% (12,400 out of 15,908) of CGIs were bound by TDG, among which 10,202 were annotated as TSSs (Fig. 4A). The Sp1 motif (5'-CCCCGCCCC-3'), consistent with an engagement of TDG at CpG-rich sequences, and the CTCF motif, were highly enriched in TDG ChIP-seq peaks (Fig. 4B). When ranked according to the expression levels of the corresponding genes in untreated wt ECCs, TSSs of highly expressed genes were found to be associated with both TDG and ATF4 whereas TSSs of low expressed genes were not (Fig. 4C). On the contrary, CTCF (GSE131356 for ESCs, and GSE103198 for P19 ECCs) was mostly absent of TSSs (Fig. 4C). Consistent with this observation, except for *Mthfd2* in ESCs, CTCF was not found at TSSs of ATF4 target genes down-regulated by TDG depletion (Supplementary Fig. 4A), suggesting that CTCF is not involved in TDG-mediated regulations at TSSs of ATF4 target genes. This is likely to be the case both in ECCs and ESCs since CTCF showed a similar distribution in these two cell types (Supplementary Fig. 4B).

Importantly, we could confirm TDG engagement at selected ATF4 target promoters by ChIP-qPCR in wt ECCs, with a specially-high enrichment at the *Atf4* promoter itself that was tended to increase in RA-treated cells (Fig. 4D), suggesting that TDG regulates ATF4 target genes through direct binding to their promoter. We next interrogated ATF4 binding to the same selected ATF4 target promoters by ChIP-qPCR. Except for *Atf4* itself, data evidenced ATF4 engagement at all promoters and, consistent with a lower activity of the corresponding genes, a significant decrease in ATF4 binding in RA-treated TDG-null ECCs (Fig. 4E). To further qualify the relationship between ATF4 and TDG engagement genome-wide, we next clustered ATF4 binding sites according to their engagement by TDG in ESCs (Fig. 4F). Whereas ATF4 binding sites not bound by TDG were mostly intronic and intergenic (cluster C5), sites that were highly enriched in TDG (cluster C3) were mostly associated with promoters and 5'UTRs, reminiscent of CGI positions (Fig. 4G). Accordingly, C3 showed high TDG and ATF4 binding and high CpG density, whereas C5 showed high ATF4 binding but low CpG density and low TDG engagement, and an intermediate situation was observed for C4 (Fig. 4H). This was also true for 5fC/5caC accumulation as determined by methylation assisted bisulfite sequencing (MAB-seq) in *Tdg* knocked-down ESCs (GEO dataset GSE62631; Neri et al., 2015). These data indicate that TDG engagement and activity at CGI-TSSs could influence chromatin structure and facilitate ATF4-mediated gene transcription. To test this hypothesis, genome-wide nucleosome maps were obtained through MNase-seq experiments run on monolayers of wt and TDG-null ECCs treated or not with RA for 48 hours. First, TSSs were ranked in quartiles of expression in RA-untreated wt and TDG-null ECCs and, as previously observed in mESCs (Voong et al., 2016), average profiles of nucleosome occupancy showed a positive correlation with gene expression in both cell types (Fig. 4I). Next, nucleosome occupancy around TSSs of genes from cluster C3 (n=3,867), cluster C4 (n=1,090) or cluster C5 (n=441) was analyzed in wt and TDG-null ECCs

in the presence or absence of RA (Supplementary Fig. 4C). Data indicated a slight decrease in nucleosome density in both RA-treated wt and TDG-null ECCs at C3 and C4 TSSs that seemed to be more pronounced for C3 TSSs (Supplementary Fig. 4C). This was further confirmed by MNase-qPCR targeting the +1 nucleosome of selected ATF4 target genes (Supplementary Fig. 4D). Interestingly, although nucleosome density did not seem to vary between wt and TDG-null ECCs, nucleosome positioning was altered downstream of the TSSs of C3 genes with the +1 nucleosome being shifted towards the TSS by 24 bp in TDG-null cells compared to wt ECCs (Fig. 4J). Overall, these data indicate that TDG does not regulate nucleosome density around TSSs in ECCs and, instead, strongly argue for a role of TDG in nucleosome positioning at TSSs. Since CTCF sites are known to have highly positioned nucleosomes (Clarkson et al., 2019), we next interrogated the role of TDG on nucleosome positioning at non-promoter CTCF binding sites common between ESCs and ECCs (n=29,207). Consistent with what was observed at TSSs, the positioning of -1/+1 nucleosomes directly flanking CTCF binding sites was altered in TDG-null cells compared to wild-type cells (Fig. 4K). Indeed, whereas these nucleosomes were positioned 185 bp away from the CTCF binding sites in wt ECCs, as shown for most CTCF sites in ESCs (Clarkson et al., 2019), their distance to CTCF sites was lowered to 155 bp in TDG-null ECCs (Fig. 4K). In addition, this effect on the -1/+1 nucleosome position was observed for all quartiles of regions ranked by their level of CTCF binding in wt ECCs, suggesting that it is not related to the strength of CTCF engagement (Supplementary Fig. 4E). Hence, the observed 30 bp shift in nucleosome positioning is likely to reflect a change in a remodeler activity, indicating that TDG regulates one or several remodelers at CTCF-bound sites, which could also be involved in regulating nucleosome positioning at TSSs.

TDG expression is of bad prognosis in cancer and correlates with expression of ATF4 target genes

Given the importance of the ATF4-controlled integrated stress response (ISR) for tumor growth in certain cancer types (Tian et al., 2021; Lines et al., 2023), we next interrogated a putative correlation between *Tdg* and *Asns* (as a paradigmatic stress response gene) expression on the one hand, and between *Tdg* expression and survival prognosis on the other hand, in 37 cohorts of patients covering a wide range of cancer types, through the UCSC Xena browser (https://xenabrowser.net/). A handful of cancers showed association (Pearson's rho ≥ 0.5, $p \leq 0.05$) between *TDG* and *ASNS* expression (Fig 5A). In line with our data obtained in a neural differentiation system, the highest correlation was found for the Pediatric Brain Tumor Atlas cohort (Children's Brain Tumor Tissue Consortium - CBTTC, 854 tumors from 33 different types) where *TDG* expression was of adverse prognosis (Fig. 5A), as was that of *ASNS* (not shown). Further examination of *ASNS* and *TDG* expression in the most represented tumor types of the CBTTC cohort showed variable expression and correlation levels of these two genes, with the highest correlation being observed for Low grade glioma (Fig. 5B,C). Of note, somatic genome structural variant analysis highlighted *ASNS* as the gene mostly associated with enhancer highjacking and overexpression in the CBTTC cohort (Zhang et al., 2021). Indeed, 4.7% of all detected enhancer highjacking events involved *ASNS* and occurred primarily in medulloblastoma embryonal tumors (Fig 5D). Interestingly, activation of the *ASNS* gene by ATF4 in asparagine-depleted acute lymphoblastic leukemia cells was shown to be counteracted by DNA methylation (Jiang et al., 2019), suggesting that expression of ISR genes is repressed by DNA methylation and is activated by erasure of the methylation mark. Next, we tested the existence of an inverse

correlation between DNA methylation and *ASNS* expression in the TCGA lung cancer cohort, in which correlation between *TDG* and *ASNS* expression was also high (pval=1.285e-87, Fig. 5A). While most CpGs immediately upstream of the *ASNS* gene were unmethylated (not shown), cg25906151 methylation showed inverse correlation with both *ASNS* and *TDG* expression (Fig. 5E,F), suggesting that *ASNS* is negatively regulated by DNA methylation in lung cancer patients. Expression of other ATF4 target genes also showed a positive correlation with TDG expression in lung cancer patients (Fig. 5G). These included *MTHFD2* and the amino-acid transporters *SLC7A1* and *SLC7A5*, whereas the glucose transporter *SLC2A3*, which is not an ATF4 target in mouse, showed no correlation (Fig. 5G). Finally, *TDG* expression showed a negative correlation with expression of the mTOR pathway inhibitor *SESN1*, whereas a positive correlation was observed with *LDHA* and *RPL22L1* (Fig. 5H), together with SLC7A5 (Fig. 5G). Extending our findings to human, these data indicate that TDG is likely to be involved in ATF4-mediated gene regulation as well as the control of the mTOR pathway in cancer patients.

Discussion

Deciphering the role of TDG in cell lineage commitment stands of prime importance but has been hampered so far by cell lethality caused by RA in *Tdg* knocked-out mESCs as well as in *Tdg* knocked-down primary neural stem cells (Cortazar et al., 2011; Steinacher et al. 2019). Here, we describe the establishment of a *Tdg* knock-out cell model from wt ECCs. These cells proliferate *in vitro* and are viable upon RA-induced differentiation, and therefore represent a surrogate for EpiSCs and allowed us to investigate the molecular function of TDG in cell fate commitment.

In the course of this study, we could evidence that TDG is required for efficient neural differentiation, as TDG-null cells underwent differentiation skewing toward a cardiac mesoderm fate when treated with RA. Such a shift was recently observed using an elaborate system in which transient *Tdg* knock-down followed by a recovery period before embryoid body differentiation of mESCs led to an increase in cardiac mesoderm formation, possibly through a down-regulation of repressors of mesoderm differentiation (Aranda et al., 2023). Differentiation skewing towards cardiac mesoderm has also been observed *in vivo* in TET mutants (Li et al., 2016), suggesting a common requirement of various components of the DNA demethylation machinery for neural differentiation. Consistent with the idea that DNA demethylation associates with neural differentiation, the knock-out of the DNA methyltransferase gene *Dnmt3b* in mESCs inversely skews differentiation toward neurectoderm (Lauria et al., 2023). Although a number of differentially methylated regions (DMRs) observed in EpiSC-like cells derived from *Dnmt3b*-KO ESCs (Lauria et al., 2023) were found to overlap with the TSSs of genes repressed in TDG-null ECCs, they also overlapped in a similar proportion with TSSs of activated genes (data not shown). This argues against a simple model in which the co-engagement of TDG at DNMT3b-bound TSSs would prevent DNA methylation by inhibiting DNMT3B activity, thus favoring gene activity. Previous mapping of 5fC and 5caC through MAB-seq indicated that a fraction of these oxidized bases is found at promoters, including in CpG islands, but the role of TDG at promoters remained elusive in this case (Neri et al., 2015). Nonetheless, highly expressed gene promoters seem to undergo DNA methylation cycling that would be disrupted in TDG-null cells. Our nucleosome mapping data suggest that TDG could fine-tune nucleosome positioning at promoters as well as at other genomic sites such as CTCF binding sites. Indeed, the +1 nucleosome downstream of TSSs and the

nucleosomes immediately flanking CTCF binding sites were repositioned in the absence of TDG, suggesting that 5mC turnover could impact specific nucleosomes. DNA methylation has been proposed to favor nucleosome positioning and stability (Chodavarapu et al., 2010) but other studies found an opposite relationship (Li et al., 2022). In addition, 5fC has been shown to associate with well-positioned nucleosomes, likely through direct interaction with histones (Teif et al., 2014, Raiber et al., 2018), and sites that gain 5mC in Tet1/Tet2 KO ESCs also have a higher nucleosome occupancy (Wiehle et al., 2019). Hence, part of the transcriptional effect of TDG depletion could be related to alterations in nucleosome positioning. Accordingly, we observed that nucleosomes flanking CTCF sites are shifted 30 bp towards CTCF sites in TDG-null ECCs compared to wt ECCs, whereas +2, +3 and +4 nucleosome positions remain unchanged. This nucleosome-shift effect is very reminiscent of that observed upon alterations in the activity of chromatin remodelers that are associated to CTCF binding sites, such as SNF2H and CHD4 (Barisi et al., 2019; Clarkson et al., 2019). Hence, TDG could participate in coordinating nucleosome positioning and gene expression through functional interaction with chromatin remodelers. Alternatively, the observed shift could be due to variations in the strength of CTCF binding itself, i.e., *Tdg* inactivation would increase CTCF recruitment, which in turn would indirectly impact on the positioning of flanking nucleosomes. Indeed, the nucleosome repeat length flanking CTCF binding sites decreases when CTCF binding strength increases (Clarkson et al., 2019). However, our data indicate that the 30 bp-shift observed in TDG-null cells does not depend on the level of CTCF engagement, therefore supporting a model in which TDG is involved in regulating the activity of chromatin remodelers upon gene activation.

In the present study we evidenced a functional link between TDG and ATF4-dependent transcription of ISR genes. Indeed, TDG binds to the promoters of *Atf4* and ATF4 target genes in ECCs, and a large set of these ATF4-target genes is down-regulated in TDG-null cells. This global effect could result from the integration of several interdependent molecular mechanisms including direct regulation of ATF4 and its partner C/EBPb DNA binding by 5mC turnover at binding sites, nucleosome (re)positioning and acetylation of H3K27 at specific sites, and regulation of ATF4 levels by stress pathways that regulate both mRNA stability and translation of ATF4 mRNA (Fig. 5I - Wortel et al., 2017). DNA binding of ATF4 and C/EBPb occurs at 5'-TGACGTCA-3' sequences and ATF4 binding is inhibited by methylation of the central CpG (Kribelbauer et al., 2017; Yin et al, 2017). ATF4-C/EBPb heterodimers also bind to the non-canonical sequence 5'-CGATGCAA-3' with higher affinity when the CpG is methylated than when it is not (Mann et al., 2013). In addition, C/EBPb binds with a slightly higher affinity to the 5’-TTGCGTCA-3’ motif when the CpG is methylated, formylated or carboxylated (Sayeed et al., 2015; Yin et al., 2017). Given the documented impact of CpG methylation status on TF binding to DNA, a complete ablation of TDG activity is predicted to result in altered chromatin binding of these major stress integrators in various ways depending on the sequence of the binding site and the type of dimer engaged.

Upon stress conditions like amino-acid starvation, ATF4 is produced and activates transcription of its target genes, the products of which act in a concerted manner in the so-called ISR. Members of the ISR include Atf4 itself, aa synthetases, tRNA charging and solute carrier (Slc) genes (Wortel et al., 2017; Lines et al., 2023). Here we show that RA-induced 3D growth or embryoid formation of wt ECCs in the absence of RA is associated with a decrease in *Slc* mRNAs encoding amino-acid transporters, suggesting that aggregate formation might induce amino-acid shortage hence leading to ATF4 activation. In line with a possible regulation of ATF4 by cell

growth conditions, mesenchymal stem cells grown at high density produce high levels of ATF4 protein compared to cells grown at low density (Cohen et al., 2015). The down-regulation of *Slc* genes involved in aa transport in 3D-growing wt ECCs was found to be amplified in TDG-null cells, strongly suggesting a defect in stress response caused by TDG inactivation. Amino-acid shortage can be sensed by two major pathways involving GCN2- and mTOR-regulated translation of ATF4 mRNA (Harding et al., 2000; Park et al., 2017; Torrence et al., 2021). In addition, mTORC1 also induces a stabilization of ATF4 mRNA (Park et al., 2017). In the GCN2-ATF4 pathway, amino-acid starvation results in an increase in uncharged tRNAs which bind to and activate the kinase GCN2, leading to phosphorylation of eIF2A and selective translation of the ATF4 mRNA. In growth factor stimulated cells, uptake of essential amino-acids through SLC7A5 activates mTORC1, which allows for sustain growth and inhibition of TFE3-regulated autophagy (Nicklin et al., 2009). Our data unequivocally identified the mTORC1 pathway as being required for ATF4 activity in P19 ECCs and being affected in TDG-null cells. In particular, depletion of TDG selectively down-regulates *Slc7a5* mRNA levels in RA-treated cells. Although the Gcn2-Atf4 pathway has been suggested to control ESC identity (D'Aniello et al., 2015), the mTORC1 pathway is involved in pluripotency exit of ESCs through TFE3 inhibition (Betschinger et al., 2013; Villegas et al., 2019). Furthermore, consistent with the differentiation skewing observed in the absence of TDG, mTORC1 inhibition favors differentiation of ESCs into cardiomyocytes (Zheng et al., 2017). Thus, by controlling the activity of mTORC1 and ATF4, TDG is likely to be involved in metabolic reprogramming during cell differentiation and to sustain cell growth and survival in cancer.

Acknowledgements

We thank L. Deleurme and A. Aimé for conducting cell sorting experiments at the Biosit cell sorting plateform (https://biosit.univ-rennes.fr/cytometrie-en-flux-et-tri-cellulaire). We also thank V. Dupé for providing the Pax6 antibody, M. Pucéat for providing the Acta2 antibody, E. Chevet and E. Lafont for the gift of p70S6K and Phospho-p70S6K antibodies, PA Bidaud-Meynard for the gift of rapamycine, and A. Sérandour for providing sequencing reagents. We are grateful to R. Gibeaux for his help with microscopy experiments and to A. Laurent for MNase-seq library preparation. Appreciation is extended to the Genomics Core Facility GenoA and the Bioinformatics Core Facility BiRD, both members of Biogenouest and France Genomique, as well as the Institut Français de Bioinformatique for their resources and technical support.

Funding: This work was funded by the Centre National de la Recherche Scientifique and the University of Rennes. MT was a recipient of a Doctoral Fellowship from the University of Rennes. RA was a recipient of a Fullbright grant. KS was a recipient of a scholarship from the Northern Kentucky University STEM International Research and Scholarly Exchange Program.

Author contributions

MT, TM, EW, RA and MB: methodology and investigation. GB: methodology and formal analysis. SA: software and data curation. KS, GP, CF, MB, JP and CLP: investigation. GS: conceptualization, project administration,

funding acquisition, investigation, visualization and original draft preparation. All authors reviewed and edited the manuscript.

References

Appiah B, Fullio CL, Ossola C, Bertani I, Restelli E, Cheffer A, Polenghi M, Haffner C, Garcia-Miralles M, Zeis P, Treppner M, Bovio P, Schlichtholz L, Mas-Sanchez A, Zografidou L, Winter J, Binder H, Grün D, Kalebic N, Taverna E, Vogel T. (2023). DOT1L activity affects neural stem cell division mode and reduces differentiation and ASNS expression. *EMBO Rep*, 24(8):e56233.

Aranda S, Alcaine-Colet A, Ballaré C, Blanco E, Mocavini I, Sparavier A, Vizán P, Borràs E, Sabidó E, Di Croce L. (2023). Thymine DNA glycosylase regulates cell-cycle-driven p53 transcriptional control in pluripotent cells. *Mol Cell*, 83(15):2673-2691.e7.

Barisi D, Stadler MB, Iurlaro M, Schübeler D. (2019). Mammalian ISWI and SWI/SNF selectively mediate binding of distinct transcription factors. *Nature*, 569:136-140

Betschinger J, Nichols J, Dietmann S, Corrin PD, Paddison PJ, Smith A. (2013). Exit from pluripotency is gated by intracellular redistribution of the bHLH transcription factor Tfe3. *Cell*, 153(2):335-47.

Boland MJ, Christman JK. (2008). Characterization of Dnmt3b:thymine-DNA glycosylase interaction and stimulation of thymine glycosylase-mediated repair by DNA methyltransferase(s) and RNA. *J Mol Biol*, 379(3):492-504.

Caron G, Hussein M, Kulis M, Delaloy C, Chatonnet F, Pignarre A, Avner S, Lemarié M, Mahé EA, Verdaguer-Dot N, Queirós AC, Tarte K, Martín-Subero JI, Salbert G, Fest T. (2015). Cell-Cycle-Dependent Reconfiguration of the DNA Methylome during Terminal Differentiation of Human B Cells into Plasma Cells. *Cell Rep*, 13(5):1059-71.

Charpentier E, Cornec M, Dumont S, Meistermann D, Bordron P, David L, Redon R, Bonnaud S, Bihouée A. (2021). 3’ RNA sequencing for rapid and low cost gene expression profiling. *Protocolexchange*, 10.21203/rs.3.pex-1336/v1.

Chaumette T, Cinotti R, Mollé A, Solomon P, Castain L, Fourgeux C, McWilliam HEG, Misme-Aucouturier B, Broquet A, Jacqueline C, Vourc'h M, Fradin D, Bossard C, David L, Montassier E, Braudeau C, Josien R, Villadangos JA, Asehnoune K, Bressollette-Bodin C, Poschmann J, Roquilly A. (2022). Monocyte Signature Associated with Herpes Simplex Virus Reactivation and Neurological Recovery after Brain Injury. *Am J Respir Crit Care Med*, 206(3):295-310.

Chodavarapu RK, Feng S, Bernatavichute YV, Chen PY, Stroud H, Yu Y, Hetzel JA, Kuo F, Kim J, Cokus SJ, Casero D, Bernal M, Huijser P, Clark AT, Krämer U, Merchant SS, Zhang X, Jacobsen SE, Pellegrini M. (2010). Relationship between nucleosome positioning and DNA methylation. *Nature*, 466(7304):388-92.

Clarkson CT, Deeks EA, Samarista R, Mamayusupova H, Zhurkin VB, Teif VB. (2019). CTCF-dependent chromatin boundaries formed by asymmetric nucleosome arrays with decreased linker length. *Nucleic Acids Res*, 47(21):11181-11196.

Cohen DM, Won KJ, Nguyen N, Lazar MA, Chen CS, Steger DJ. (2015). ATF4 licenses C/EBPβ activity in human mesenchymal stem cells primed for adipogenesis. *Elife*, 4:e06821.

Cortázar D, Kunz C, Selfridge J, Lettieri T, Saito Y, MacDougall E, Wirz A, Schuermann D, Jacobs AL, Siegrist F, Steinacher R, Jiricny J, Bird A, Schär P. (2011). Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature*, 470(7334):419-23.

D'Aniello C, Fico A, Casalino L, Guardiola O, Di Napoli G, Cermola F, De Cesare D, Tatè R, Cobellis G, Patriarca EJ, Minchiotti G. (2015). A novel autoregulatory loop between the Gcn2-Atf4 pathway and (L)-Proline [corrected] metabolism controls stem cell identity. *Cell Death Differ*, 22(7):1094-105.

Edri S, Hayward P, Baillie-Johnson P, Steventon BJ, Martinez Arias A. (2019). An epiblast stem cell-derived multipotent progenitor population for axial extension. *Development*, 146(10):dev168187.

Faletti S, Osti D, Ceccacci E, Richichi C, Costanza B, Nicosia L, Noberini R, Marotta G, Furia L, Faretta MR, Brambillasca S, Quarto M, Bertero L, Boldorini R, Pollo B, Gandini S, Cora D, Minucci S, Mercurio C, Varasi M, Bonaldi T, Pelicci G. (2021). LSD1-directed therapy affects glioblastoma tumorigenicity by deregulating the protective ATF4-dependent integrated stress response. *Sci Transl Med*, 13(623):eabf7036.

Fujita E, Soyama A, Kawabata M, Momoi T. (1999). BMP-4 and retinoic acid synergistically induce activation of caspase-9 and cause apoptosis of P19 embryonal carcinoma cells cultured as a monolayer. *Cell Death Differ*, 6(11):1109-16.

Gallais R, Demay F, Barath P, Finot L, Jurkowska R, Le Guével R, Gay F, Jeltsch A, Métiver R, Salbert G. (2007) Deoxyribonucleic Acid Methyl Transferases 3a and 3b Associate with the Nuclear Orphan Receptor COUP-TFI during Gene Activation. *Mol Endocrinol*, 21:2085-2096.

Ginno PA, Gaidatzis D, Feldmann A, Hoerner L, Imanci D, Burger L, Zilbermann F, Peters AHFM, Edenhofer F, Smallwood SA, Krebs AR, Schübeler D. (2020). A genome-scale map of DNA methylation turnover identifies site-specific dependencies of DNMT and TET activity. *Nat Commun*, 11(1):2680.

Han DW, Tapia N, Araúzo-Bravo MJ, Lim KT, Kim KP, Ko K, Lee HT, Schöler HR. (2013). Sox2 Level Is a Determinant of Cellular Reprogramming Potential. *PLoS One*, 8(6):e67594.

Harding HP, Novoa I, Zhang Y, Zeng H, Wek R, Schapira M, Ron D. (2000). Regulated translation initiation controls stress-induced gene expression in mammalian cells. *Mol Cell*, 6(5):1099-108.

Henrique D, Abranches E, Verrier L, Storey KG. (2015). Neuromesodermal progenitors and the making of the spinal cord. *Development*, 142(17):2864-75.

Horn V, Minucci S, Ogryzko VV, Adamson ED, Howard BH, Levin AA, Ozato K. (1996). RAR and RXR selective ligands cooperatively induce apoptosis and neuronal differentiation in P19 embryonal carcinoma cells. *FASEB J*, 10(9):1071-7.

Jiang J, Srivastava S, Seim G, Pavlova NN, King B, Zou L, Zhang C, Zhong M, Feng H, Kapur R, Wek RC, Fan J, Zhang J. (2019). Promoter demethylation of the asparagine synthetase gene is required for ATF4-dependent adaptation to asparagine depletion. *J Biol Chem*, 294:18674-84

Jones-Villeneuve EM, McBurney MW, Rogers KA, Kalnins VI. (1982). Retinoic acid induces embryonal carcinoma cells to differentiate into neurons and glial cells. *J Cell Biol*, 94(2):253-62.

Kaluscha S, Domcke S, Wirbelauer C, Stadler MB, Durdu S, Burger L, Schübeler D. (2022). Evidence that direct inhibition of transcription factor binding is the prevailing mode of gene and repeat repression by DNA methylation. *Nat Genet*, 54(12):1895-1906.

Kriaucionis S, Heintz N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*, 324(5929):929-30.

Kribelbauer JF, Laptenko O, Chen S, Martini GD, Freed-Pastor WA, Prives C, Mann RS, Bussemaker HJ. (2017). Quantitative Analysis of the DNA Methylation Sensitivity of Transcription Factor Complexes. *Cell Rep*, 19(11):2383-2395.

Langmead B, Trapnell C, Pop M, Salzberg SL. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, 10(3):R25.

Langton S, Gudas LJ. (2008). CYP26A1 knockout embryonic stem cells exhibit reduced differentiation and growth arrest in response to retinoic acid. *Dev Biol*, 315(2):331-54.

Lassot I, Estrabaud E, Emiliani S, Benkirane M, Benarous R, Margottin-Goguet F. (2005). p300 modulates ATF4 stability and transcriptional activity independently of its acetyltransferase domain. *J Biol Chem*, 280(50):41537-45.

Laurent A, Madigou T, Bizot M, Turpin M, Palierne G, Mahé E, Guimard S, Métivier R, Avner S, Le Péron C, Salbert G. (2002). TET2-mediated epigenetic reprogramming of breast cancer cells impairs lysosome biogenesis. *Life Sci Alliance*, 5(7):e202101283.

Lauria A, Meng G, Proserpio V, Rapelli S, Maldotti M, Polignano IL, Anselmi F, Incarnato D, Krepelova A, Donna D, Levra Levron C, Donati G, Molineris I, Neri F, Oliviero S. (2023). DNMT3B supports meso-endoderm differentiation from mouse embryonic stem cells. *Nat Commun*, 14(1):367.

Letellier T, Kervella D, Sadek A, Masset C, Garandeau C, Fourgeux C, Gourain V, Poschmann J, Blancho G, Ville S, On Behalf of The Divat Consortium. (2022). Time-Limited Therapy with Belatacept in Kidney Transplant Recipients. *J Clin Med*, 11(11):3229.

Li YQ, Zhou PZ, Zheng XD, Walsh CP, Xu GL. (2007). Association of Dnmt3a and thymine DNA glycosylase links DNA methylation with base-excision repair. *Nucleic Acids Res*, 35(2):390-400.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078-9.

Li X, Yue X, Pastor WA, Lin L, Georges R, Chavez L, Evans SM, Rao A. (2016). Tet proteins influence the balance between neuroectodermal and mesodermal fate choice by inhibiting Wnt signaling. *Proc Natl Acad Sci U S A*, 113(51):E8267-E8276.

Li S, Peng Y, Panchenko AR. (2022). DNA methylation: Precise modulation of chromatin structure and dynamics. *Curr Opin Struct Biol*, 75:102430.

Lines CL, McGrath MJ, Dorwart T, Conn CS. (2023). The integrated stress response in cancer progression: a force for plasticity and resistance. *Front Oncol*, 13:1206561.

Liu T, Ortiz JA, Taing L, Meyer CA, Lee B, Zhang Y, Shin H, Wong SS, Ma J, Lei Y, Pape UJ, Poidinger M, Chen Y, Yeung K, Brown M, Turpaz Y, Liu XS. (2011). Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol*, 12(8):R83.

Love MI, Huber W, Anders S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15:550. 10.1186/s13059-014-0550-8

Magnúsdóttir E, Dietmann S, Murakami K, Günesdogan U, Tang F, Bao S, Diamanti E, Lao K, Gottgens B, Azim Surani M. (2013). A tripartite transcription factor network regulates primordial germ cell specification in mice. *Nat Cell Biol*, 15(8):905-15.

Mahé EA, Madigou T, Sérandour AA, Bizot M, Avner S, Chalmel F, Palierne G, Métivier R, Salbert G. (2017). Cytosine modifications modulate the chromatin architecture of transcriptional enhancers. *Genome Res*, 27(6):947-958.

Maiti A, Drohat AC. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J Biol Chem*, 286(41):35334–38.

Mann IK, Chatterjee R, Zhao J, He X, Weirauch MT, Hughes TR, Vinson C. (2013). CG methylated microarrays identify a novel methylated sequence bound by the CEBPB|ATF4 heterodimer that is active in vivo. *Genome Res*, 23(6):988-97.

McBurney MW. (1993). P19 embryonal carcinoma cells. *Int J Dev Biol*, 37(1):135-40.

Ménoret S, Tesson L, Remy S, Gourain V, Sérazin C, Usal C, Guiffes A, Chenouard V, Ouisse LH, Gantier M, Heslan JM, Fourgeux C, Poschmann J, Guillonneau C, Anegon I. (2023). $CD4^+$ and $CD8^+$ regulatory T cell characterization in the rat using a unique transgenic Foxp3-EGFP model. *BMC Biol*, 21(1):8.

Métivier R, Gallais R, Tiffoche C, Le Péron C, Jurkowska RZ, Carmouche RP, Ibberson D, Barath P, Demay F, Reid G, Benes V, Jeltsch A, Gannon F, Salbert G. (2008). Cyclical DNA methylation of a transcriptionally active promoter. *Nature*, 452(7183):45-50.

Mulholland CB, Nishiyama A, Ryan J, Nakamura R, Yiğit M, Glück IM, Trummer C, Qin W, Bartoschek MD, Traube FR, Parsa E, Ugur E, Modic M, Acharya A, Stolz P, Ziegenhain C, Wierer M, Enard W, Carell T, Lamb DC, Takeda H, Nakanishi M, Bultmann S, Leonhardt H. (2020). Recent evolution of a TET-controlled and DPPA3/STELLA-driven pathway of passive DNA demethylation in mammals. *Nat Commun*, 11(1):5972.

Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, Bird A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*, 393(6683):386-9.

Neddermann P, Jiricny J. (1993). The purification of a mismatch-specific thymine-DNA glycosylase from HeLa cells. *J Biol Chem*, 268(28):21218-24.

Neddermann P, Jiricny J. (1994). Efficient removal of uracil from G.U mispairs by the mismatch-specific thymine DNA glycosylase from HeLa cells. *Proc Natl Acad Sci U S A*, 91(5):1642-6.

Neri F, Incarnato D, Krepelova A, Rapelli S, Anselmi F, Parlato C, Medana C, Dal Bello F, Oliviero S. (2015). Single-Base Resolution Analysis of 5-Formyl and 5-Carboxyl Cytosine Reveals Promoter DNA Methylation Dynamics. *Cell Rep*, 10(5):674-683.

Nicklin P, Bergman P, Zhang B, Triantafellow E, Wang H, Nyfeler B, Yang H, Hild M, Kung C, Wilson C, Myer VE, MacKeigan JP, Porter JA, Wang YK, Cantley LC, Finan PM, Murphy LO. (2009). Bidirectional transport of amino acids regulates mTOR and autophagy. *Cell*, 136(3):521-34.

Okada Y, Shimazaki T, Sobue G, Okano H. (2004). Retinoic-acid-concentration-dependent acquisition of neural cell identity during in vitro differentiation of mouse embryonic stem cells. *Dev Biol*, 275(1):124-42.

Onabote O, Hassan HM, Isovic M, Torchia J. (2022). The Role of Thymine DNA Glycosylase in Transcription, Active DNA Demethylation, and Cancer. *Cancers (Basel*), 14(3):765.

Pakos-Zebrucka K, Koryga I, Mnich K, Ljujic M, Samali A, Gorman AM. (2016). The integrated stress response. *EMBO Rep*, 17(10):1374-1395.

Park Y, Reyna-Neyra A, Philippe L, Thoreen CC. (2017). mTORC1 Balances Cellular Amino Acid Supply with Demand for Protein Synthesis through Post-transcriptional Control of ATF4. *Cell Rep*, 19(6):1083-1090.

Parry A, Rulands S, Reik, W. (2021). Active turnover of DNA methylation during cell fate decisions. *Nature reviews. Genetics*, *22*(1), 59–66. https://doi-org.insb.bib.cnrs.fr/10.1038/s41576-020-00287-8

Raiber EA, Portella G, Martínez Cuesta S, Hardisty R, Murat P, Li Z, Iurlaro M, Dean W, Spindel J, Beraldi D, Liu Z, Dawson MA, Reik W, Balasubramanian S. (2018). 5-Formylcytosine organizes nucleosomes and forms Schiff base interactions with histones in mouse embryonic stem cells. *Nat Chem*, 10(12):1258-1266.

Rulands S, Lee HJ, Clark SJ, Angermueller C, Smallwood SA, Krueger F, Mohammed H, Dean W, Nichols J, Rugg-Gunn P, Kelsey G, Stegle O, Simons BD, Reik W. (2018). Genome-Scale Oscillations in DNA Methylation during Exit from Pluripotency. *Cell Syst*, 7(1):63-76.e12.

Ruzzo EK, Capo-Chichi JM, Ben-Zeev B, Chitayat D, Mao H, Pappas AL, Hitomi Y, Lu YF, Yao X, Hamdan FF, Pelak K, Reznik-Wolf H, Bar-Joseph I, Oz-Levi D, Lev D, Lerman-Sagie T, Leshinsky-Silver E, Anikster Y, Ben-Asher E, Olender T, Colleaux L, Décarie JC, Blaser S, Banwell B, Joshi RB, He XP, Patry L, Silver RJ, Dobrzeniecka S, Islam

MS, Hasnat A, Samuels ME, Aryal DK, Rodriguiz RM, Jiang YH, Wetsel WC, McNamara JO, Rouleau GA, Silver DL, Lancet D, Pras E, Mitchell GA, Michaud JL, Goldstein DB. (2013). Deficiency of asparagine synthetase causes congenital microcephaly and a progressive form of encephalopathy. *Neuron*, 80(2):429-41.

Sandoval JE, Huang YH, Muise A, Goodell MA, Reich NO. (2019). Mutations in the DNMT3A DNA methyltransferase in acute myeloid leukemia patients cause both loss and gain of function and differential regulation by protein partners. *J Biol Chem*, 294(13):4898-4910.

Sayeed SK, Zhao J, Sathyanarayana BK, Golla JP, Vinson C. (2015). C/EBPβ (CEBPB) protein binding to the C/EBP|CRE DNA 8-mer TTGC|GTCA is inhibited by 5hmC and enhanced by 5mC, 5fC, and 5caC in the CG dinucleotide. *Biochim Biophys Acta*, 1849(6):583-9.

Sérandour AA, Avner S, Oger F, Bizot M, Percevault F, Lucchetti-Miganeh C, Palierne G, Gheeraert C, Barloy-Hubler F, Péron CL, Madigou T, Durand E, Froguel P, Staels B, Lefebvre P, Métivier R, Eeckhoute J, Salbert G. (2012). Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. *Nucleic Acids Res*, 40(17):8255-65.

Steinacher R, Barekati Z, Botev P, Kuśnierczyk A, Slupphaug G, Schär P. (2019). SUMOylation coordinates BERosome assembly in active DNA demethylation during cell differentiation. *EMBO J*, 38(1):e99242.

Stephens M. (2016). False discovery rates: a new deal. Biostatistics, 18:2. 10.1093/biostatistics/kxw041

Tabata S, Kojima Y, Sakamoto T, Igarashi K, Umetsu K, Ishikawa T, Hirayama A, Kajino-Sakamoto R, Sakamoto N, Yasumoto KI, Okano K, Suzuki Y, Yachida S, Aoki M, Soga T. (2023). L-2hydroxyglutaric acid rewires amino acid metabolism in colorectal cancer via the mTOR-ATF4 axis. *Oncogene*, 42(16):1294-1307.

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, 324(5929):930-5.

Teif VB, Beshnova DA, Vainshtein Y, Marth C, Mallm JP, Höfer T, Rippe K. (2014). Nucleosome repositioning links DNA (de)methylation and differential CTCF binding during stem cell development. *Genome Res*, 24(8):1285-95.

Tian X, Zhang S, Zhou L, Seyhan AA, Hernandez Borrero L, Zhang Y, El-Deiry WS. (2021). Targeting the Integrated Stress Response in Cancer Therapy. *Front Pharmacol*, 12:747837.

Tini M, Benecke A, Um SJ, Torchia J, Evans RM, Chambon P. (2002). Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription. *Mol Cell*, 9(2):265-77.

Torrence ME, MacArthur MR, Hosios AM, Valvezan AJ, Asara JM, Mitchell JR, Manning BD. (2021). The mTORC1-mediated activation of ATF4 promotes protein and glutathione synthesis downstream of growth signals. *Elife*, 10:e63326.

Um S, Harbers M, Benecke A, Pierrat B, Losson R, Chambon P. (1998). Retinoic acid receptors interact physically and functionally with the T:G mismatch-specific thymine-DNA glycosylase. *J Biol Chem*, 273(33):20728-36.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M and Rozen SG. (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Res*, 40(15):e115.

Villegas F, Lehalle D, Mayer D, Rittirsch M, Stadler MB, Zinner M, Olivieri D, Vabres P, Duplomb-Jego L, De Bont ESJM, Duffourd Y, Duijkers F, Avila M, Geneviève D, Houcinat N, Jouan T, Kuentz P, Lichtenbelt KD, Thauvin-Robinet C, St-Onge J, Thevenon J, van Gassen KLI, van Haelst M, van Koningsbruggen S, Hess D, Smallwood SA, Rivière JB, Faivre L, Betschinger J. (2019). Lysosomal Signaling Licenses Embryonic Stem Cell Differentiation via Inactivation of Tfe3. *Cell Stem Cell*, 24(2):257-270.e8.

Voong LN, Xi L, Sebeson AC, Xiong B, Wang JP, Wang X. (2016). Insights into Nucleosome Organization in Mouse Embryonic Stem Cells through Chemical Mapping. *Cell*, 167(6):1555-1570.e15.

Watrin E, Demidova M, Watrin T, Hu Z, Prigent C. (2014). Sororin pre-mRNA splicing is required for proper sister chromatid cohesion in human cells. *EMBO Rep*, 15(9):948-55.

Wiehle L, Thorn GJ, Raddatz G, Clarkson CT, Rippe K, Lyko F, Breiling A, Teif VB. (2019). DNA (de)methylation in embryonic stem cells controls CTCF-dependent chromatin boundaries. *Genome Res*, 29(5):750-761.

Wheldon LM, Abakir A, Ferjentsik Z, Dudnakova T, Strohbuecker S, Christie D, Dai N, Guan S, Foster JM, Corrêa IR Jr, Loose M, Dixon JE, Sottile V, Johnson AD, Ruzov A. (2014). Transient accumulation of 5-carboxylcytosine indicates involvement of active demethylation in lineage specification of neural stem cells. *Cell Rep*, 7(5):1353-1361.

Wortel IMN, van der Meer LT, Kilberg MS, van Leeuwen FN. (2017). Surviving Stress: Modulation of ATF4-Mediated Stress Responses in Normal and Malignant Cells. *Trends Endocrinol Metab*, 28(11):794-806.

Wu X, Zhang Y. (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat Rev Genet*, 18(9):517-534.

Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, Das PK, Kivioja T, Dave K, Zhong F, Nitta KR, Taipale M, Popov A, Ginno PA, Domcke S, Yan J, Schübeler D, Vinson C, Taipale J. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*, 356(6337):eaaj2239.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol*, ;9(9):R137.

Zhang Y, Chen F, Donehower LA, Scheurer ME, Creighton CJ. (2021). A pediatric brain tumor atlas of genes deregulated by somatic genomic rearrangement. *Nat Commun*, 12:937.

Zheng B, Wang J, Tang L, Shi J, Zhu D. (2017). mTORC1 and mTORC2 play different roles in regulating cardiomyocyte differentiation from embryonic stem cells. *Int J Dev Biol*, 61(1-2):65-72.

Figure legends

Figure 1: Transcriptomic analysis of wt and TDG-null ECCs undergoing retinoic acid-induced differentiation. (**A**) MA plot showing differentially expressed genes in 1 μM RA-treated TDG-null ECCs compared to RA-treated wt ECCs grown as monolayers. Up-regulated genes are highlighted in red and down-regulated genes in blue. Specific marker genes are shown by a green dot and their names are indicated. (**B**) Gene ontology (GO) enrichment analysis of down- and up-regulated genes in 1 μM RA-treated TDG-null ECCs compared to RA-treated wt ECCs (BP: biological process). (**C**) Quantification of the mRNA levels (RT-qPCR) of the neural markers *Pax6* and *Irx3*, the mesoderm inducer *Wnt11*, and the cardiac mesoderm marker *Acta2* during long-term RA treatment of wt and TDG-null ECCS. For each gene analyzed, expression values were reported to the average value of expression in untreated wt ECCs and are shown as "fold change". (**D**) Log2 fold change (Log2FC) and adjusted *P* value (Adj *P* val) of the 17 most significantly down-regulated genes in 1 μM RA-treated monolayers of TDG-null ECCs compared to RA-treated wt ECCs. (**E**) Network showing functional interactions between the 17 genes shown in (D). The network was drawn using the online tool String (https://string-db.org/). The thickness of the network edges indicates confidence of the corresponding supporting data.

Figure 2: Reciprocal gene expression regulation between TDG and Cyp26a1. (**A**) Results of MTT assays in wt and TDG-null ECCs treated with increasing concentrations of RA for neurectodermal differentiation or with 1% DMSO for mesodermal differentiation. (**B**) Immunofluorescence detection of the neural marker PAX6 and the cardiomyocyte marker ACTA2 in untreated ECCs (top) and wt (middle) and TDG-null (bottom) cells treated for 72 hours with 1 μM RA. Cells were counterstained with DAPI. (**C**) Phase-contrast microscopy images of wt and TDG-null ECCs treated with 1 μM RA for 6 days. Cells were first grown as embryoid bodies for 2 days and then re-plated in cell-culture dishes for 4 days. (**D**) Quantification of neurite extension from embryoid bodies 4 days after replating. (**E**) RT-qPCR assay of *Meis1* mRNA levels in wt and TDG-null ECCs treated with increasing

concentrations of RA or 1% DMSO for 48 hours. (**F**) Phase-contrast microscopy images of wt and TDG-null ECCs grown as monolayers and treated with increasing concentrations of RA or with 1% DMSO. White triangles point to regions of RA-induced 3D growth. (**G**) Quantification of 3D growth regions in wt, TDG-null and two clones of Cyp26a1-null ECCs. (**H**) RT-qPCR assay of the pluripotency marker *Oct4*, the neural marker *Pax6*, and the cardiomyocyte marker *Acta2* mRNAs in the indicated cells treated or not with 1 μM RA for 48 hours. (**I**) RT-qPCR assay of *Bdnf* mRNA in the indicated cells treated or not with 1 μM RA for 48 hours. (**J**) RT-qPCR assay of *Tdg* mRNA in the indicated cells treated or not with 1 μM RA for 48 hours. (**K**) Western blot detection of TDG in wt, TDG-null, and Cyp26a1-null ECCs. Detection of U2AF65 serves as a loading control. Arrowheads indicate TDG bands and the asterisk indicates a non-specific band. (**L**) RT-qPCR assays for the indicated ATF4/ISR pathway genes in cells treated or not with 1 μM RA for 48 hours.

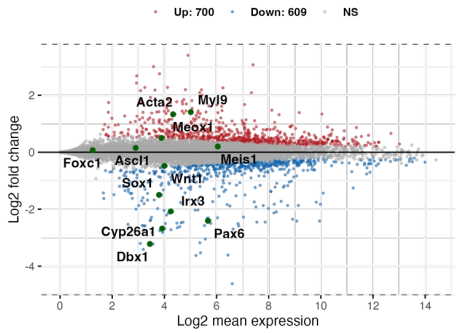
Figure 3: TDG and mTORC1 regulate ATF4-mediated gene transcription during cell differentiation. (**A**) RNA-seq data of a 1 μM RA treatment time-course experiment in wt ECCs for the indicated genes. Data are shown as Log2FC (LFC). (**B**) H3K27ac ChIP-seq signal of a 1 μM RA treatment time-course experiment in wt ECCs. (**C**) TDG and ATF4 ChIP-seq signal at selected loci. (**D**) Expression levels of the indicated genes in wt and TDG-null ECCs grown as monolayers and treated with 1 μM RA for 48 hours (extracted from RNA-seq data and expressed in CPM). (E) ChIP qPCR analysis of H3K27ac enrichment at the +1 nucleosome of ATF4 target genes. *Nampt*-nuc1 and -nuc2 indicates that two different nucleosome positions were targeted in the *Nampt* gene. For each gene, values were normalized to the H3K27ac signal in untreated wt ECCs and are expressed as fold change. (**F**) Expression levels of the indicated genes in untreated wt and TDG-null ECCs grown either as monolayers (Mon.) or embryoid bodies (Agg. - extracted from RNA-seq data and expressed in CPM). (**G**) Venn diagram of the genes down-regulated in untreated embryoid bodies compared to monolayers (aggregation-repressed genes) of wt ECC or TDG-null ECCs (left). Gene ontology biological process annotation of the 182 genes commonly down-regulated in embryoid bodies *versus* monolayers of wt and TDG-null ECCs and of the 323 genes specifically down-regulated in TDG-null ECCs. (**H**) RT-qPCR analysis of selected ATF4-target gene mRNAs in wt and TDG-null ECCs grown as monolayers and treated or not with 20 nM rapamycin for 16 hours (+ rap.). (**I**) Western blot analysis of the mTORC1 target p70s6k in wt and TDG-null ECC monolayers treated or not with 1 μM RA for 48 hours. Detection of p70s6k phosphorylated on T389 is shown on top, p70s6k in the middle, and the loading control Smc1a on the bottom. (**J**) RT-qPCR detection of mRNAs encoding differentiation markers in monolayers of wt ECCs treated or not with 1 μM RA for 48 hours. Twenty four hours after RA addition, rapamycin was added to the cell culture medium (20 nM) and cells were further cultured for 24 hours.

Figure 4: TDG binds active gene promoters and regulates nucleosome positioning. (**A**) Venn diagram analysis showing the overlap between TDG peaks in mESCs, CGIs and TSSs. (**B**) Enrichment in transcription factor motifs in mESC TDG peaks. (**C**) Heatmap of TDG, ATF4 and CTCF ChIP-seq signal centered on the TSSs of genes expressed in wt ECCs (cutoff of 4 reads per sample) and ranked by expression level. (**D**) ChIP-qPCR assay of TDG at promoters of ATF4-target genes in wt ECCs. (**E**) ChIP-qPCR assay of ATF4 at promoters of ATF4-target genes in wt ECCs. (**F**) Heatmap clustering of ATF4 binding sites according to their enrichment in TDG in mESCs. (**G**) Genomic

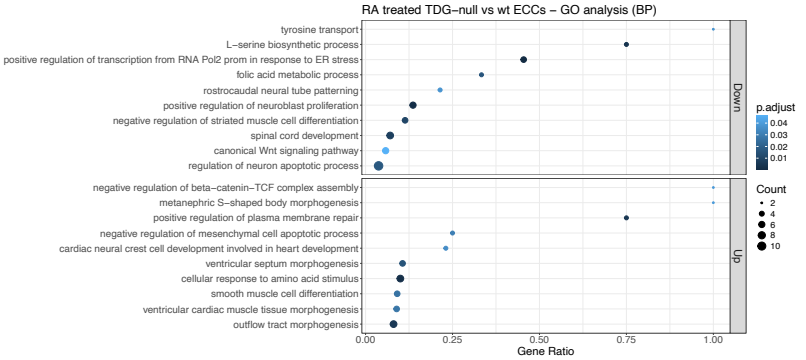
distribution of ATF4 binding sites from clusters C3, C4 and C5 as defined in (F). (**H**) Average profiles of CpG density, MAB-seq signal, and TDG and ATF4 ChIP-seq signal centered on the oriented TSS (arrow) of genes associated with ATF4 binding (ATF4 ChIP-seq peak within -/+ 500 bp of the TSS) and included in clusters C3, C4 and C5. (**I**) Average profiles of MNase-seq signal at the TSS of genes ranked by quartiles of expression in monolayers of untreated wt or TDG-null ECCs. (**J**) Average profiles of MNase-seq signal at C3, C4, and C5 TSSs in monolayers of wt or TDG-null ECCs treated with 1 μM RA. (**K**) Average profiles of MNase-seq signal at non-promoter CTCF binding sites common between ESCs and ECCs of wt and TDG-null ECCs grown as monolayers and either untreated (left panel) or treated with 1 μM RA (middle and right panels). The right panel shows the same data as in the middle panel but restricted to the -400/+400 bp distance.

Figure 5: TDG is of bad prognosis in cancer. (**A**) Graphical representation of Pearson's correlation between *TDG* and *ASNS* expression levels as a function of the bad prognosis *P* value of TDG expression in patient tumors from 37 TGAC cohorts. (**B**) Violin plots of *ASNS* and *TDG* expression levels in pediatric brain tumors. (**C**) Pearson's correlation between *TDG* and *ASNS* expression levels in pediatric brain tumors. (**D**) Pie chart representation of the distribution of *ASNS* association with enhancer highjacking events in pediatric brain tumors. (**E**) Heatmap representation of cg25906151 methylation levels and *ASNS* and *TDG* expression levels in lung tumors (high methylation/expression: red, low methylation/expression: blue). (**F**) Expression levels of *TDG* and *ASNS* in lung tumors classified by quartiles of cg25906151 methylation (Q1: low DNA methylation, Q4: high DNA methylation). (**G**) Heatmaps showing the expression levels of the indicated genes in lung tumors ranked according to TDG expression levels. (**H**) Heatmaps showing the expression levels of the indicated genes in lung tumors ranked according to TDG expression levels. (**I**) Working model depicting ISR gene regulation by TDG. ATF4 and TDG co-binding at ATF4 target genes is associated to H3K27 acetylation by CBP and +1 nucleosome positioning by a chromatin remodeler, leading to expression of the ISR genes which include the aa transporter gene *Slc7a5* (PM: plasma membrane). Amino-acid influx stimulates mTORC1 activity which in turn favors *ATF4* mRNA stability and translation, sustaining ISR gene activity. The mTORC1 inhibition by SESN1 is lowered in a TDG-dependent manner when cells are treated with RA, allowing for maintenance of the ISR gene transcriptional activity.

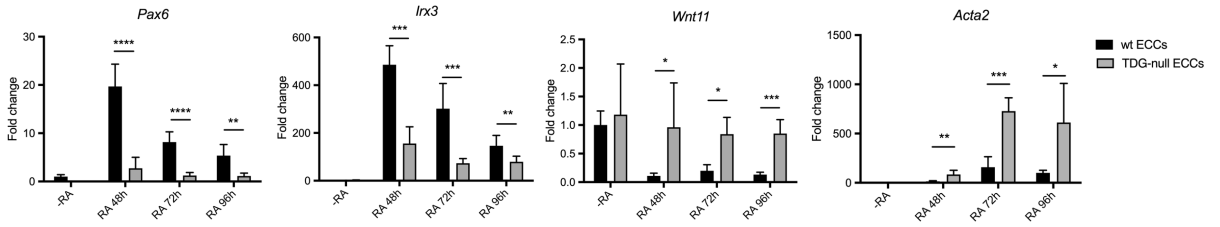
A



B



C



D

Gene name	log2FC	Adj *P* val
Asns	-2.43	1.61e-117
Mthfd2	-2.43	1.15e-99
Psat1	-1.76	2.75e-79
Nars	-1.84	1.81e-59
Bcat1	-2.11	2.36e-56
Atf4	-1.73	1.91e-52
Eif4ebp1	-1.88	3.49e-48
Yars	-1.99	5.08e-47
Eif2s2	-1.26	5.92e-47
Cth	-2.49	4.80e-46
Sars	-1.44	1.03e-45
Slc3a2	-1.88	1.038e-45
Chac1	-4.90	2.26e-44
Shmt2	-1.89	6.54e-42
Pck2	-2.49	1.70e-41
Herpud1	-2.31	9.60e-40
Psph	-1.84	1.97e-39

E

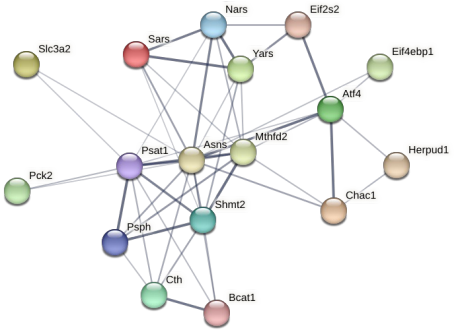


Fig. 1

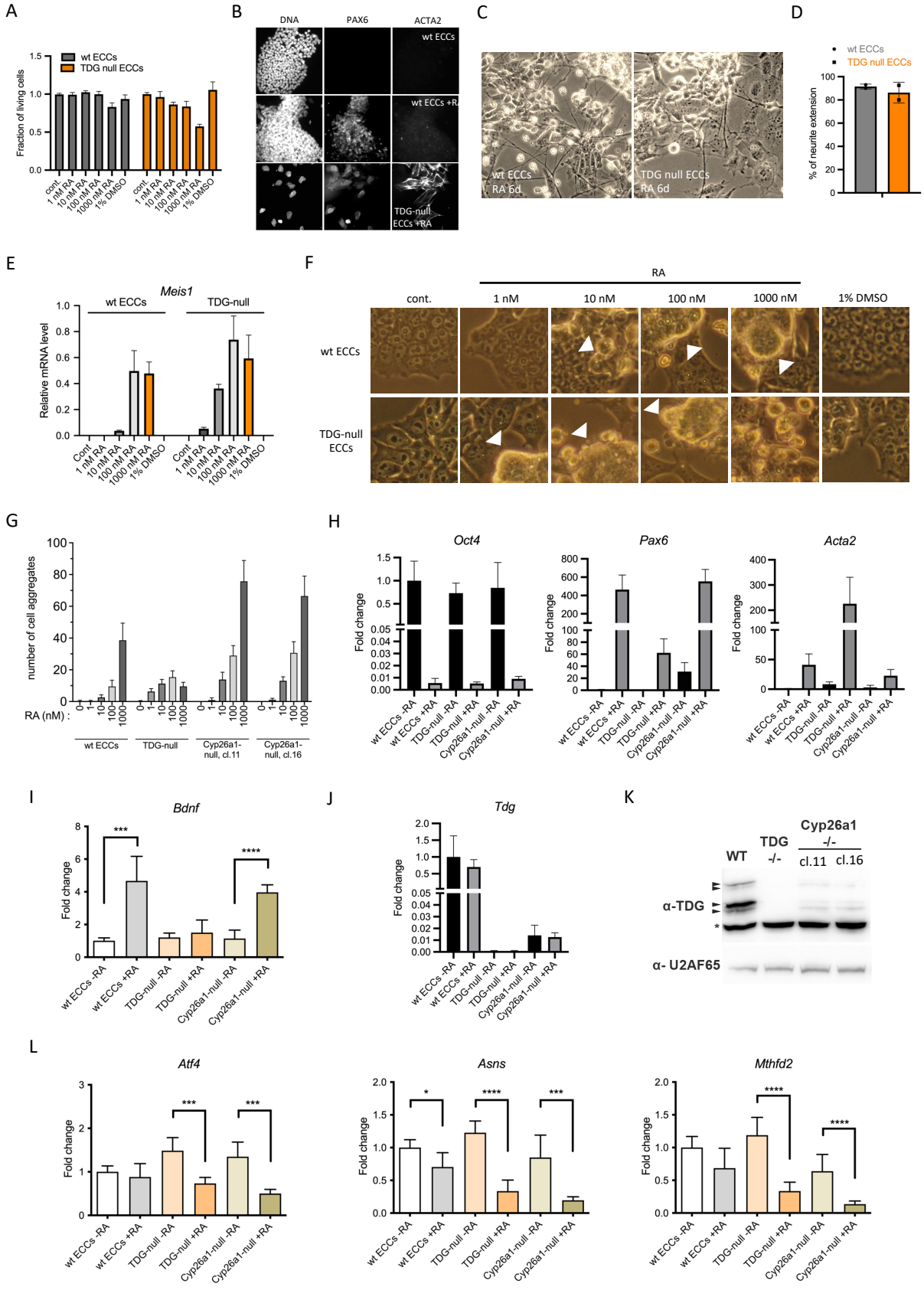

A
wt ECCs
TDG null ECCs
Fraction of living cells
cont.
1 nM RA
10 nM RA
100 nM RA
1000 nM RA
1% DMSO
B
DNA
PAX6
ACTA2
wt ECCs
wt ECCs +RA
TDG-null ECCs +RA
C
wt ECCs RA 6d
TDG null ECCs RA 6d
D
wt ECCs
TDG null ECCs
% of neurite extension
E
Meis1
wt ECCs
TDG-null
Relative mRNA level
Cont
F
RA
cont.
1 nM
10 nM
100 nM
1000 nM
1% DMSO
wt ECCs
TDG-null ECCs
G
number of cell aggregates
RA (nM) :
wt ECCs
TDG-null
Cyp26a1-null, cl.11
Cyp26a1-null, cl.16
H
Oct4
Pax6
Acta2
Fold change
wt ECCs -RA
wt ECCs +RA
TDG-null -RA
TDG-null +RA
Cyp26a1-null -RA
Cyp26a1-null +RA
I
Bdnf
Fold change
***
****
J
Tdg
Fold change
K
WT
TDG -/-
Cyp26a1 -/-
cl.11
cl.16
α-TDG
α- U2AF65
L
Atf4
Asns
Mthfd2
Fold change
*
***
****

Fig. 2

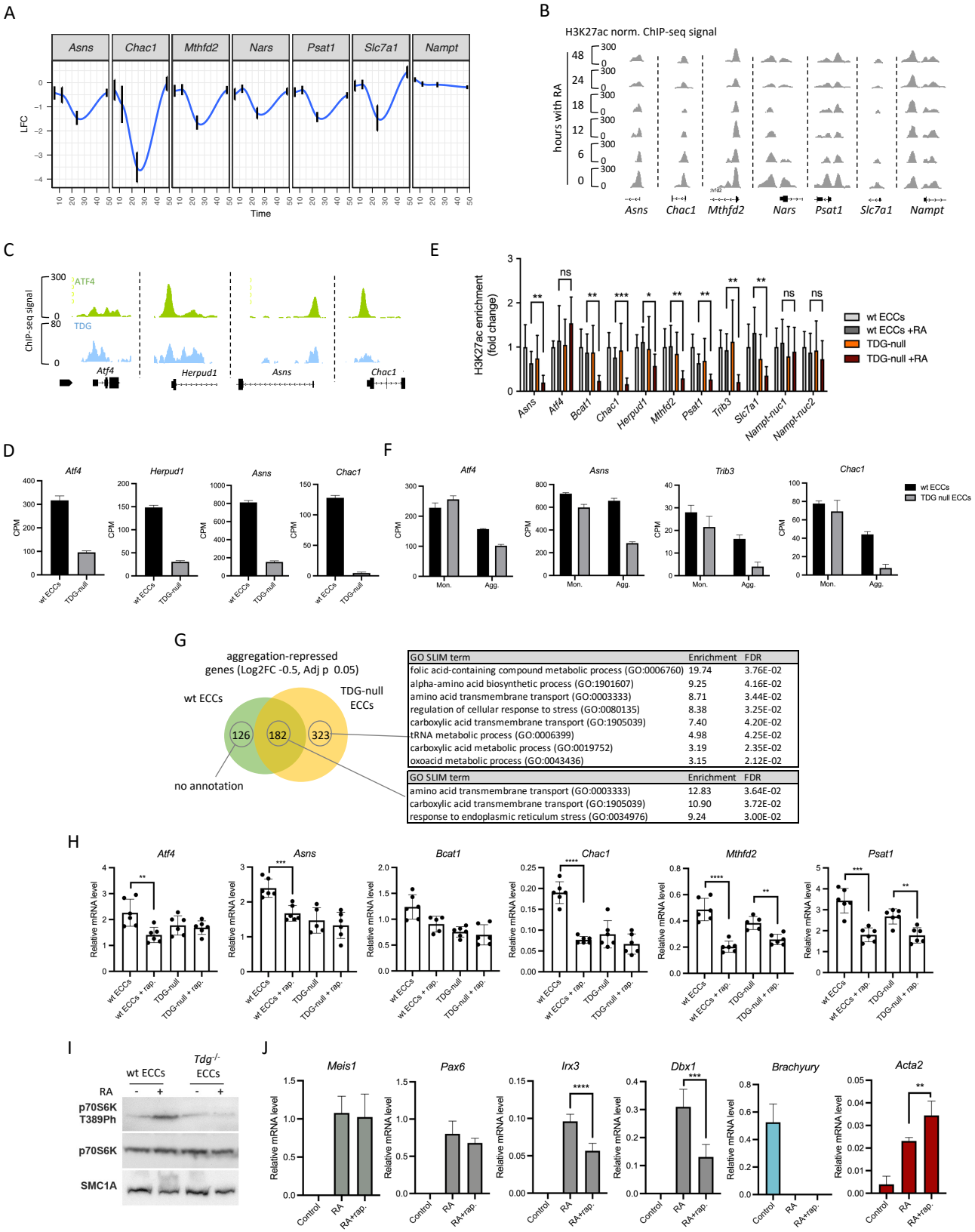

A
Asns
Chac1
Mthfd2
Nars
Psat1
Slc7a1
Nampt
LFC
Time
B
H3K27ac norm. ChIP-seq signal
hours with RA
48
24
18
12
6
0
Asns
Chac1
Mthfd2
Nars
Psat1
Slc7a1
Nampt
C
ChIP-seq signal
ATF4
TDG
Atf4
Herpud1
Asns
Chac1
E
H3K27ac enrichment (fold change)
wt ECCs
wt ECCs +RA
TDG-null
TDG-null +RA
Asns
Atf4
Bcat1
Chac1
Herpud1
Mthfd2
Psat1
Trib3
Slc7a1
Nampt-nuc1
Nampt-nuc2
D
Atf4
Herpud1
Asns
Chac1
CPM
wt ECCs
TDG-null
F
Atf4
Asns
Trib3
Chac1
CPM
Mon.
Agg.
wt ECCs
TDG null ECCs
G
aggregation-repressed genes (Log2FC -0.5, Adj p 0.05)
wt ECCs
TDG-null ECCs
126
182
323
no annotation
GO SLIM term | Enrichment | FDR
folic acid-containing compound metabolic process (GO:0006760) | 19.74 | 3.76E-02
alpha-amino acid biosynthetic process (GO:1901607) | 9.25 | 4.16E-02
amino acid transmembrane transport (GO:0003333) | 8.71 | 3.44E-02
regulation of cellular response to stress (GO:0080135) | 8.38 | 3.25E-02
carboxylic acid transmembrane transport (GO:1905039) | 7.40 | 4.20E-02
tRNA metabolic process (GO:0006399) | 4.98 | 4.25E-02
carboxylic acid metabolic process (GO:0019752) | 3.19 | 2.35E-02
oxoacid metabolic process (GO:0043436) | 3.15 | 2.12E-02
GO SLIM term | Enrichment | FDR
amino acid transmembrane transport (GO:0003333) | 12.83 | 3.64E-02
carboxylic acid transmembrane transport (GO:1905039) | 10.90 | 3.72E-02
response to endoplasmic reticulum stress (GO:0034976) | 9.24 | 3.00E-02
H
Atf4
Asns
Bcat1
Chac1
Mthfd2
Psat1
Relative mRNA level
wt ECCs
wt ECCs + rap.
TDG-null
TDG-null + rap.
I
wt ECCs
Tdg-/- ECCs
RA
p70S6K T389Ph
p70S6K
SMC1A
J
Meis1
Pax6
Irx3
Dbx1
Brachyury
Acta2
Relative mRNA level
Control
RA
RA+rap.

Fig. 3

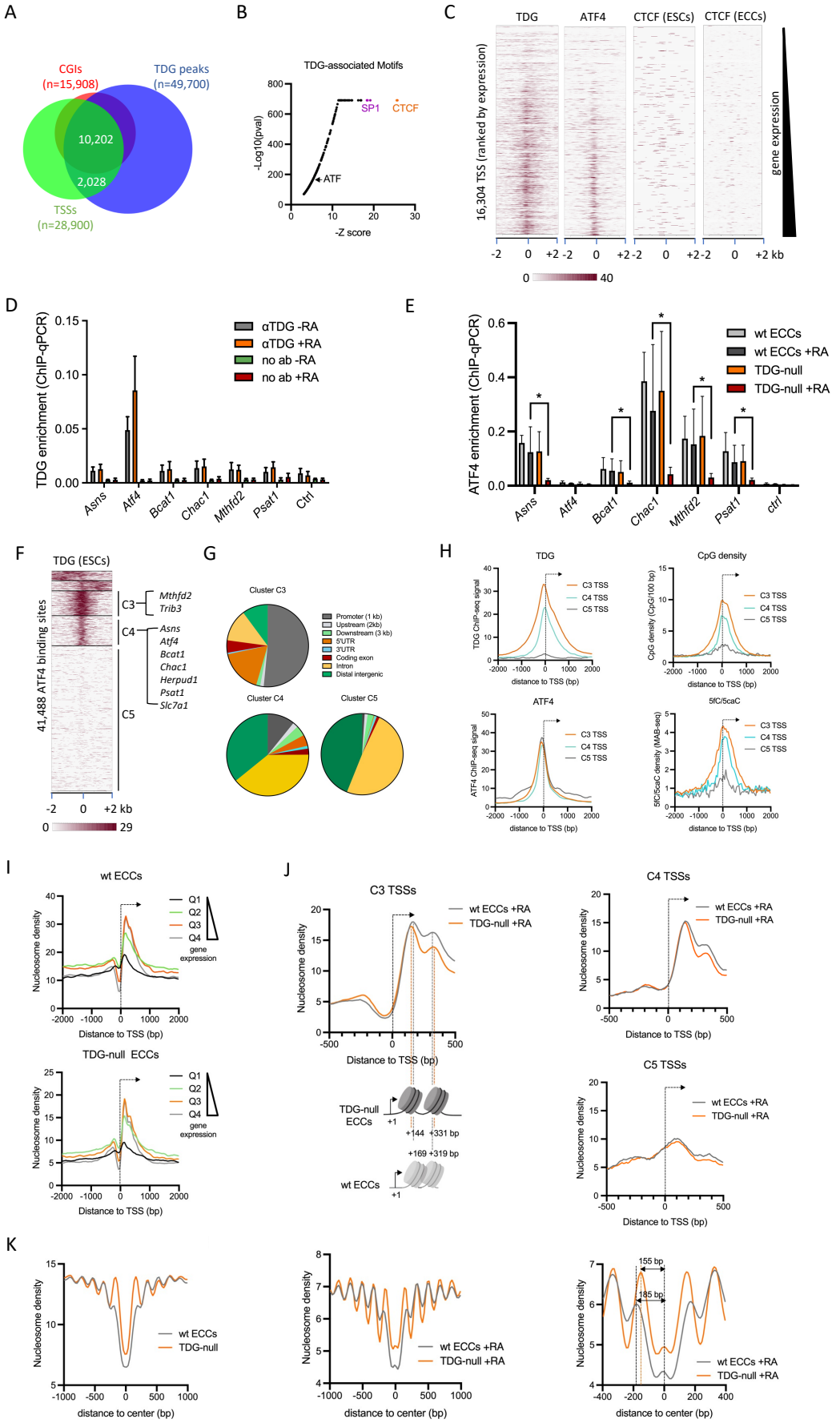

A
CGIs
(n=15,908)
TDG peaks
(n=49,700)
10,202
2,028
TSSs
(n=28,900)
B
TDG-associated Motifs
-Log10(pval)
-Z score
SP1
CTCF
ATF
C
TDG
ATF4
CTCF (ESCs)
CTCF (ECCs)
16,304 TSS (ranked by expression)
gene expression
-2 0 +2 kb
0 40
D
TDG enrichment (ChIP-qPCR)
αTDG -RA
αTDG +RA
no ab -RA
no ab +RA
Asns
Atf4
Bcat1
Chac1
Mthfd2
Psat1
Ctrl
E
ATF4 enrichment (ChIP-qPCR)
wt ECCs
wt ECCs +RA
TDG-null
TDG-null +RA
Asns
Atf4
Bcat1
Chac1
Mthfd2
Psat1
ctrl
F
TDG (ESCs)
41,488 ATF4 binding sites
C3
Mthfd2
Trib3
C4
Asns
Atf4
Bcat1
Chac1
Herpud1
Psat1
Slc7a1
C5
-2 0 +2 kb
0 29
G
Cluster C3
Cluster C4
Cluster C5
Promoter (1 kb)
Upstream (2kb)
Downstream (3 kb)
5'UTR
3'UTR
Coding exon
Intron
Distal intergenic
H
TDG
CpG density
ATF4
5fC/5caC
C3 TSS
C4 TSS
C5 TSS
TDG ChIP-seq signal
CpG density (CpG/100 bp)
ATF4 ChIP-seq signal
5fC/5caC density (MAB-seq)
distance to TSS (bp)
I
wt ECCs
TDG-null ECCs
Q1
Q2
Q3
Q4
gene expression
Nucleosome density
Distance to TSS (bp)
J
C3 TSSs
C4 TSSs
C5 TSSs
wt ECCs +RA
TDG-null +RA
Nucleosome density
Distance to TSS (bp)
TDG-null ECCs
wt ECCs
+1
+144
+331 bp
+169
+319 bp
K
wt ECCs
TDG-null
wt ECCs +RA
TDG-null +RA
Nucleosome density
distance to center (bp)
155 bp
185 bp

Fig. 4

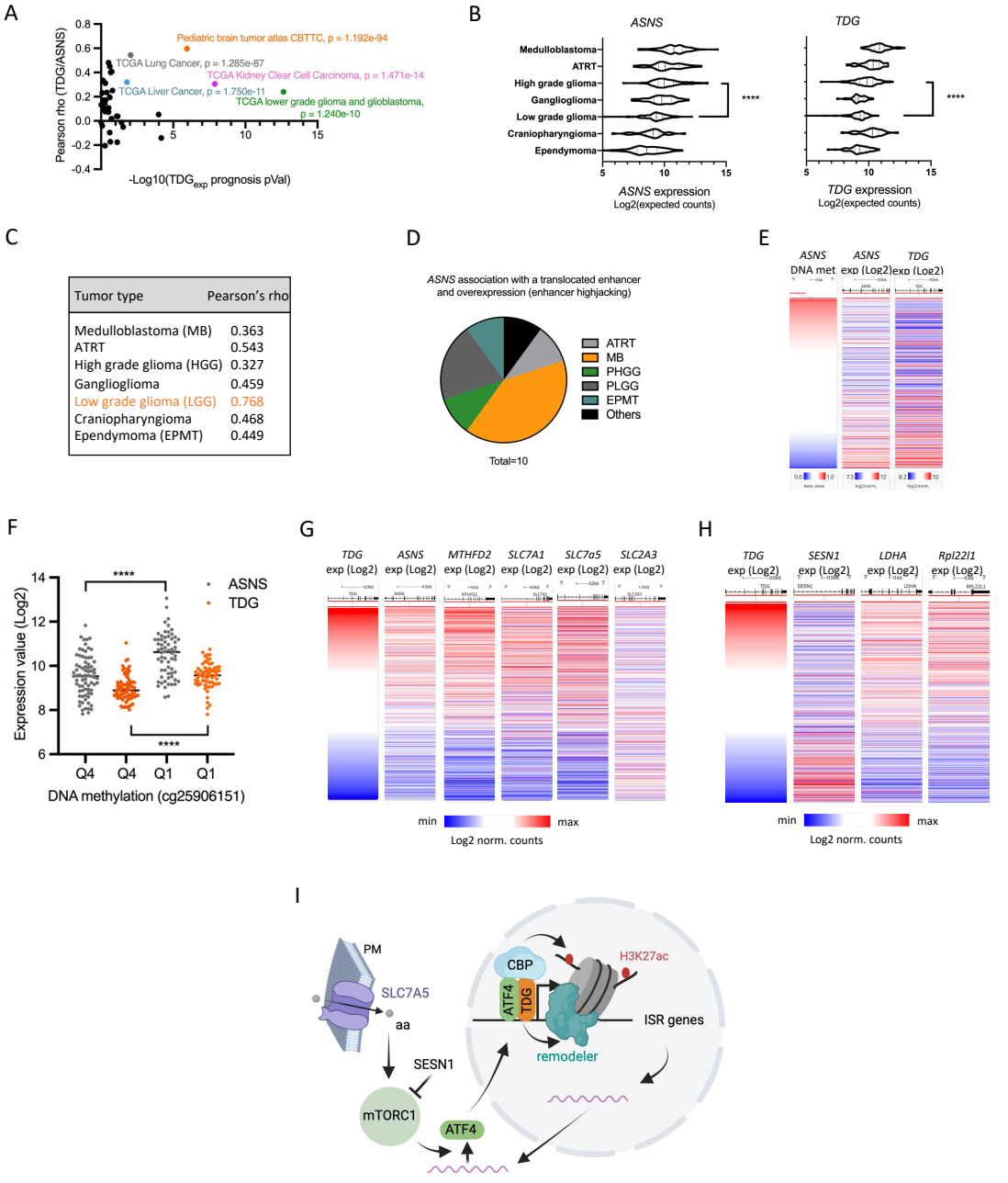

A
Pediatric brain tumor atlas CBTTC, p = 1.192e-94
TCGA Lung Cancer, p = 1.285e-87
TCGA Kidney Clear Cell Carcinoma, p = 1.471e-14
TCGA Liver Cancer, p = 1.750e-11
TCGA lower grade glioma and glioblastoma, p = 1.240e-10
Pearson rho (TDG/ASNS)
-Log10(TDGexp prognosis pVal)
B
ASNS
TDG
Medulloblastoma
ATRT
High grade glioma
Ganglioglioma
Low grade glioma
Craniopharyngioma
Ependymoma
****
ASNS expression
Log2(expected counts)
TDG expression
Log2(expected counts)
C
Tumor type | Pearson's rho
Medulloblastoma (MB) 0.363
ATRT 0.543
High grade glioma (HGG) 0.327
Ganglioglioma 0.459
Low grade glioma (LGG) 0.768
Craniopharyngioma 0.468
Ependymoma (EPMT) 0.449
D
ASNS association with a translocated enhancer and overexpression (enhancer highjacking)
ATRT
MB
PHGG
PLGG
EPMT
Others
Total=10
E
ASNS DNA met
ASNS exp (Log2)
TDG exp (Log2)
F
ASNS
TDG
Expression value (Log2)
****
Q4
Q4
Q1
Q1
DNA methylation (cg25906151)
G
TDG exp (Log2)
ASNS exp (Log2)
MTHFD2 exp (Log2)
SLC7A1 exp (Log2)
SLC7a5 exp (Log2)
SLC2A3 exp (Log2)
min
max
Log2 norm. counts
H
TDG exp (Log2)
SESN1 exp (Log2)
LDHA exp (Log2)
Rpl22l1 exp (Log2)
min
max
Log2 norm. counts
I
PM
SLC7A5
aa
SESN1
mTORC1
ATF4
CBP
ATF4
TDG
H3K27ac
ISR genes
remodeler

Fig. 5