DAZL regulates TET1-mediated reprogramming to a naive pluripotent state

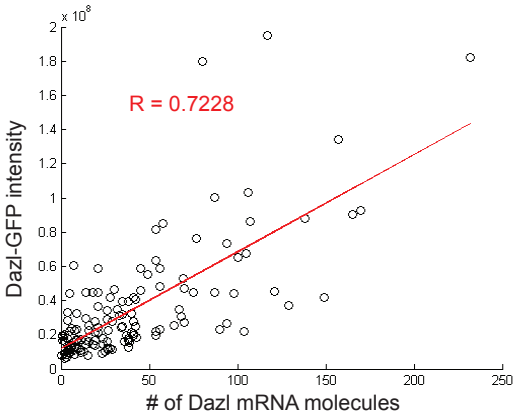
M Welling *et al*

APPENDIX

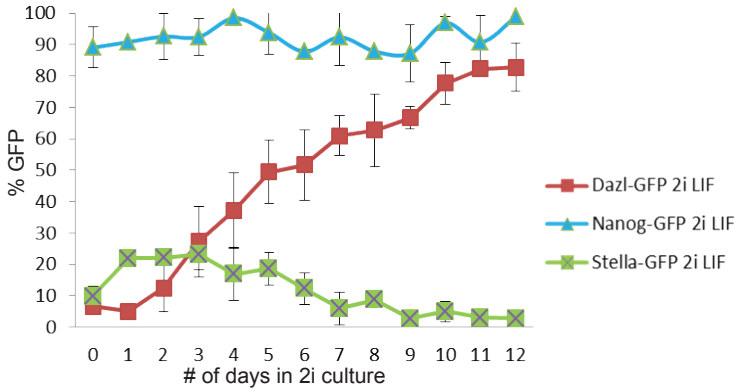
CONTENTS

Figure S1
Figure S2
Figure S3
Figure S4
Supplementary Figure Legends
Supplementary Methods

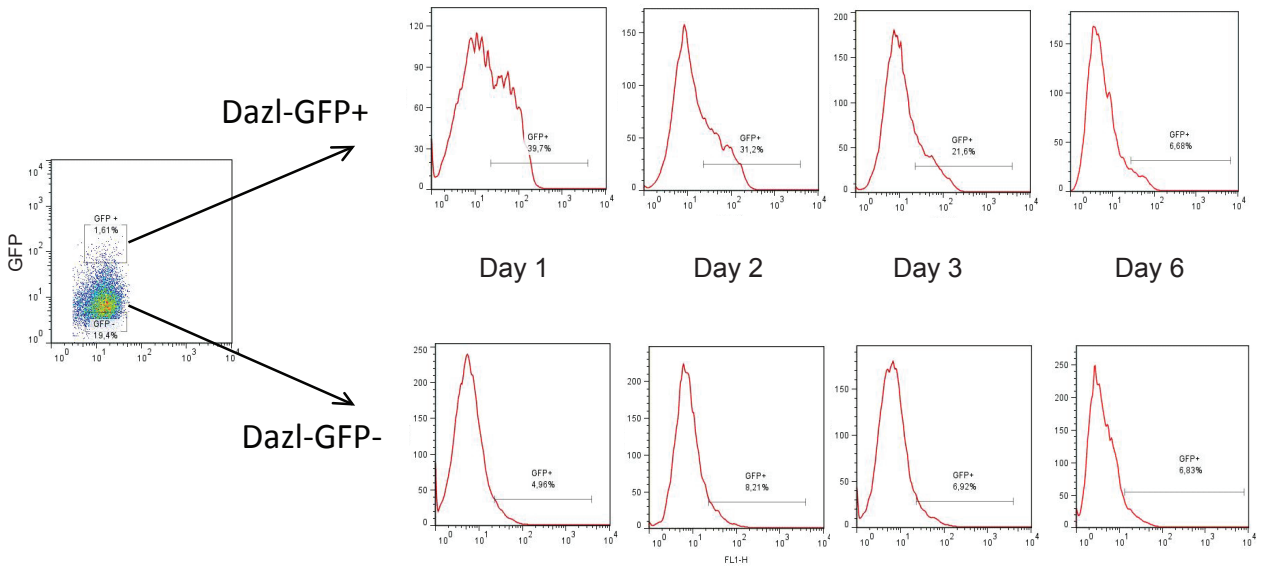
A



C



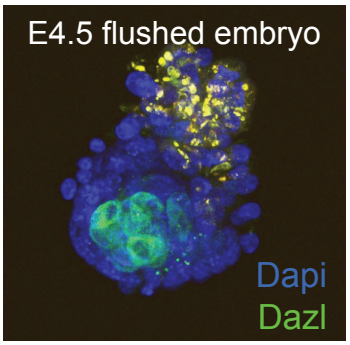
B



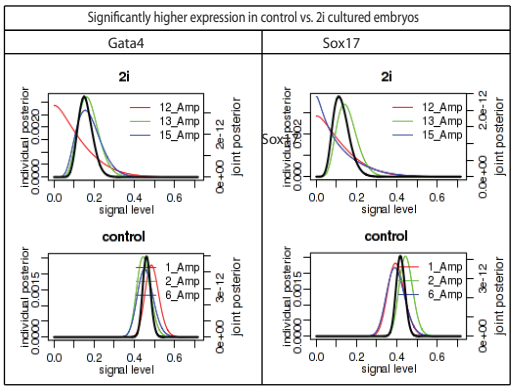
D

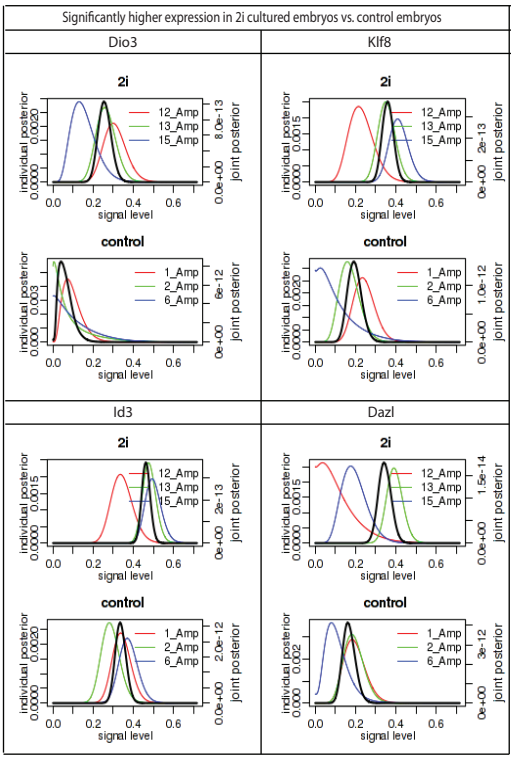
GO Term	1	2	3	4	5	6
TGFBR pathway	0.0011					
negative regulation of cell proliferation	0.0011					
heart development	0.0010					
extracellular matrix organization	0.0000					
collagen fibril organization	0.0000					
fatty acid beta-oxidation		0.0046				
regulation of ATPase activity		0.0046				
iron ion homeostasis		0.0039				
cellular iron ion homeostasis		0.0034				
protein metabolic process		0.0004				
negative regulation of programmed cell death			0.0012			
negative regulation of apoptosis			0.0011			
cellular alcohol metabolic process			0.0009			
cellular amino acid and derivative metabolic process			0.0003			
carbohydrate metabolic process			0.0000			
DNA repair				0.0022		
regulation of cell cycle				0.0015		
DNA damage checkpoint				0.0010		
induction of apoptosis by intracellular signals				0.0010		
DNA damage response, signal transduction				0.0004		
hexose biosynthetic process					0.0068	
endosome organization					0.0057	
DNA methylation						0.0093
positive regulation of foam cell differentiation						0.0093
fructose metabolic process						0.0004

A

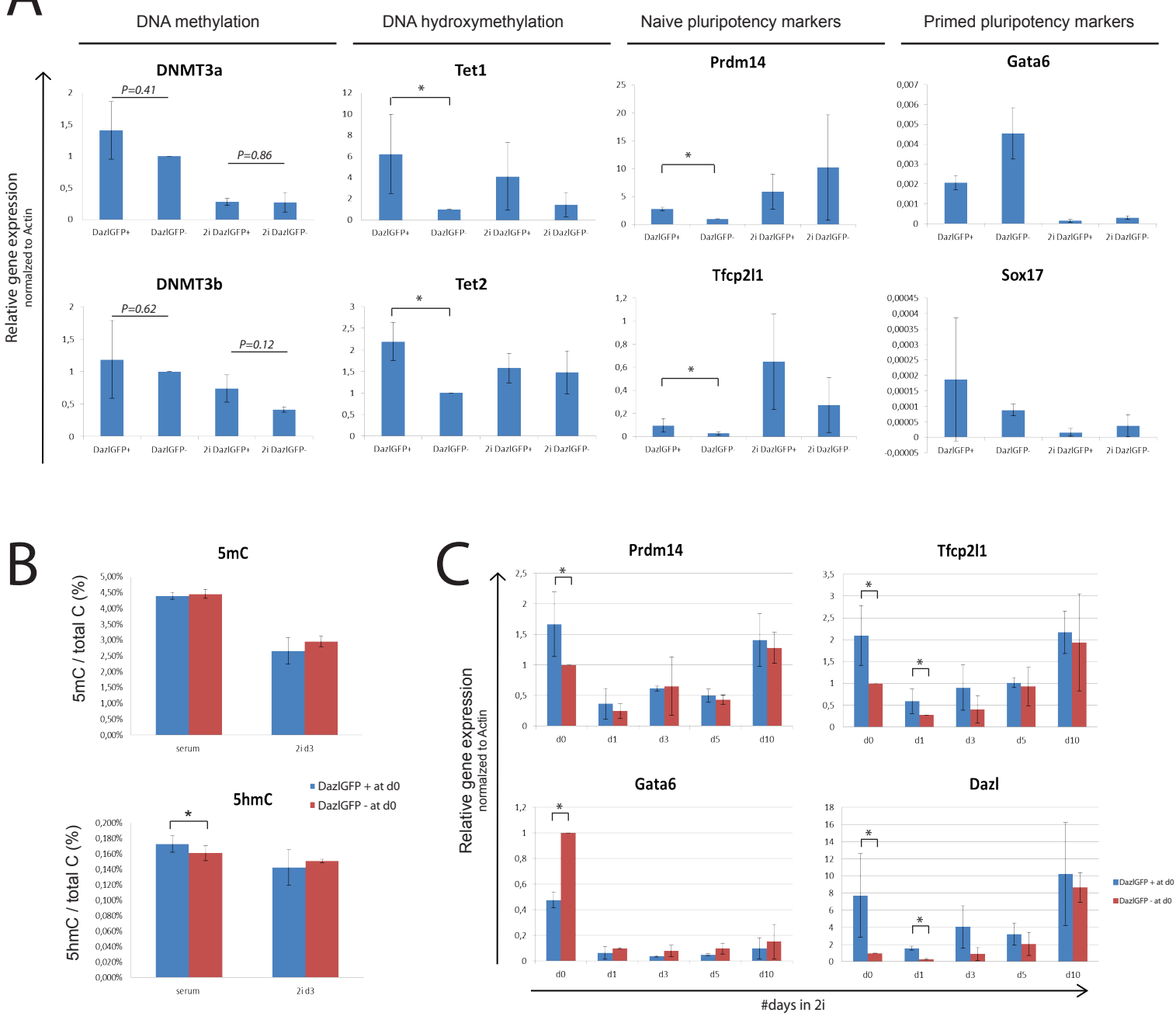

E4.5 flushed embryo
Dapi
Dazl

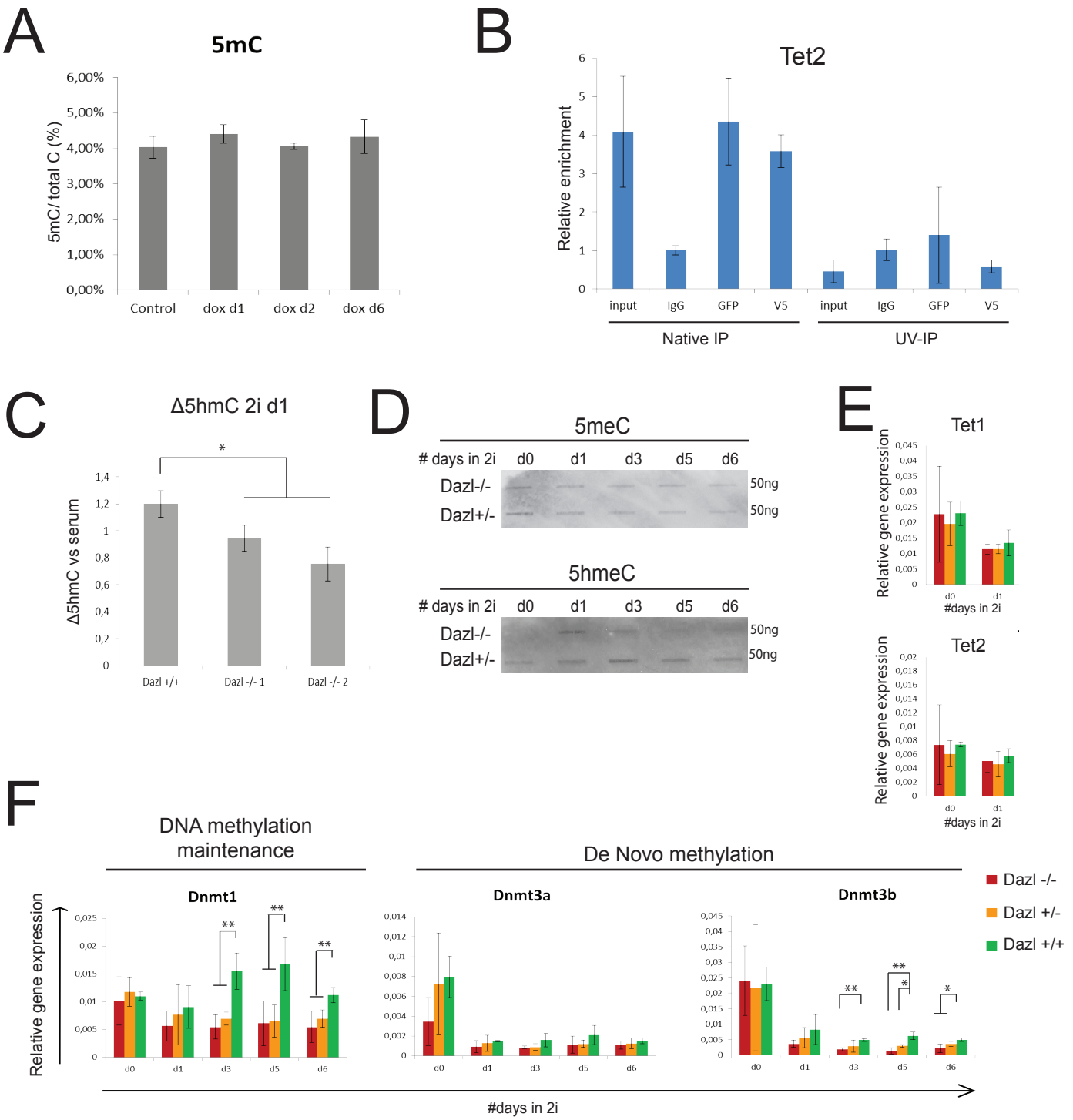
B


Significantly higher expression in control vs. 2i cultured embryos
Gata4
Sox17
2i
control
12_Amp
13_Amp
15_Amp
1_Amp
2_Amp
6_Amp
individual posterior
joint posterior
signal level


Significantly higher expression in 2i cultured embryos vs. control embryos
Dio3
Klf8
Id3
Dazl
2i
control
12_Amp
13_Amp
15_Amp
1_Amp
2_Amp
6_Amp
individual posterior
joint posterior
signal level

Welling *et al.* _ Supplemental figure S3




A
5mC
5mC/ total C (%)
6,00%
5,00%
4,00%
3,00%
2,00%
1,00%
0,00%
Control
dox d1
dox d2
dox d6
B
Tet2
Relative enrichment
input
IgG
GFP
V5
Native IP
UV-IP
C
Δ5hmC 2i d1
Δ5hmC vs serum
Dazl +/+
Dazl -/- 1
Dazl -/- 2
D
5meC
days in 2i
d0
d1
d3
d5
d6
Dazl-/-
Dazl+/-
50ng
5hmeC
E
Tet1
Relative gene expression
#days in 2i
Tet2
F
DNA methylation maintenance
De Novo methylation
Dnmt1
Dnmt3a
Dnmt3b
Dazl -/-
Dazl +/-
Dazl +/+
Relative gene expression
#days in 2i

Supplementary Figure S1:

A) Graph showing the correlation of the number of single Dazl mRNAs visualized by single molecule FISH and GFP intensity in individual Dazl-GFP ESCs.

B) FACS plots of Dazl-GFP+ and Dazl-GFP- sorted populations showing that Dazl-GFP ESCs quickly restore their original equilibrium after a few days.

C) Expression of Dazl-GFP, Nanog-GFP and Stella-GFP between day 0 and day 12 after 2i induction.

D) GO analysis of 6 clusters of differentially expressed genes at day and day 12 of ESCs in 2i culture conditions. Cluster 1: genes unregulated at day 3, Cluster 2: genes unregulated at day 12, Cluster 3: genes sustainably unregulated, Cluster 4: genes downregulated at day 3, Cluster 5: genes downregulated at day 12, Cluster 6: genes sustainably downregulated.

Supplementary Figure S2:

A) Dazl-GFP expression in an E4.5 flushed late blastocyst embryo.

B) Six examples of 377 differentially expressed genes (Z-score >3) from RNA sequencing data from control and 2i cultured blastocyst embryos. Left panel: Genes with a significantly higher expression in control than in 2i cultured embryos. Right panel: Genes with a significantly higher expression in 2i cultured embryos than control embryos.

Supplementary Figure S3:

A) qRT-PCR analysis for genes implicated in DNA methylation, DNA hydroxymethylation, naïve pluripotency and primed pluripotency in Dazl-GFP+ and Dazl-GFP- cells sorted from serum cultured ESCs or 2i-cultured ESCs. *$P < 0.05$, $n \geq 3$ biological replicates.

B) Mass spectrometric analysis of global 5mC and 5hmC levels in DazlGFP+ and DazlGFP- cells in serum and cells that were FACS sorted for DazlGFP+ and DazlGFP- followed by 2i culture for 3 days. *$P < 0.05$, Error bars indicate s.e.m. values of ≥3 biological replicates.

C) qRT-PCR analysis for the indicated genes at different time point in DazlGFP+ and DazlGFP- cells, FACS sorted at d0 and cultured in 2i conditions. *$P < 0.05$, n ≥ 3 biological replicates.

D) Representative smFISH images showing single Oct4 and Sox2 mRNA molecules in ESCs that were sorted for Dazl-GFP+ and Dazl-GFP- followed by 3 day culture in 2i conditions.

Supplementary Figure S4:

A) Mass spectrometric quantification of global 5mC levels upon overexpression of Dazl in ESCs at d0, d1, d2 and d6. N=3 biological replicates.

B) qRT-PCR results for Tet2 in both native as well as UV-crosslinked RNA-IP experiments in Dazl-GFP mESCs. Antibodies against GFP and V5 were used to immunoprecipitate Dazl and normal mouse IgG was used as a negative control. Error bars indicate s.d. values of 3 biological replicates.

C) Percentage change of global 5hmC levels in Dazl +/+ and two different Dazl -/- ESCs in serum and after 1 day in 2i conditions vs serum. N=3 biological replicates.

D) Dot blot analysis for 5meC and 5hmeC in Dazl -/- and Dazl +/- ESCs cultured in 2i conditions for several days.

E) qRT-PCR results for hydroxymethylation genes Tet1 and Tet2 in Dazl -/-, Dazl +/- and Dazl +/+ ESCs cultured in serum and in 2i conditions for 1 day.

F) qRT-PCR results for genes involved in DNA methylation in Dazl -/-, Dazl +/- and Dazl +/+ ESCs cultured in 2i conditions for several days. The average results were shown for 2 different Dazl -/- ESC-lines as well as 2 different Dazl +/- ESC-lines. P-values were calculated using student's *t*-test. **P*<0.05, ***P*<0.001. Error bars indicate s.d. values of ≥3 biological replicates per cell line.

Supplementary Methods

Generation of inducible Dazl ESCs

Dazl was cloned by PCR from mESC cDNA and inserted in the p2Lox-V5 vector as described previously [1]. Recipient ESCs were treated with doxycycline to induce Cre recombinase expression prior to the electroporation of p2Lox-DazlV5 construct. Selection was initiated the next day with 250 ng/ml of G418, resistant clones were picked, characterized and expanded.

Microarray analysis

For genome-wide expression analysis, total RNA was extracted using Trizol reagent (Invitrogen) and labeled and hybridized to Agilent Whole Mouse Genome Oligo 4X44K Microarrays (onecolor platform) according to the manufacturer's protocols. The gene expression results were analyzed using GeneSifter microarray analysis software, the DAVID bioinformatics tools and MATLAB programs developed in house.

Immunofluorescence microscopy

ESCs on coverslips were fixed in 4% paraformaldehyde (PFA), permeabilized in 0.1% Tween20/PBS or methanol and blocked in 20% donkey serum in 0.1% Tween20/PBS for 10 min at 4°C. The cells were then incubated in primary antibodies overnight at 4°C, washed and incubated with secondary antibodies for 1 hour at 4°C, washed and mounted in Vectashield with 1ng/ml DAPI. Cells and embryos were imaged with the Leica SPE or Nikon A1 confocal microscope. Primary antibodies used were: Stella (Abcam, ab19878), Dazl (Abcam , Nanog (Abcam, ab80892) and secondary antibodies were: Alexa-Fluor-488 (Life Technologies, A20181), Alexa-Fluor-568 (Life Technologies, A20184).

Q-PCR gene expression analysis

Total RNA was extracted with Trizol reagent (Invitrogen), treated with DNase I (Promega) and first strand cDNA was synthesized with SuperScript III reverse transcriptase using random primers (Invitrogen). Q-PCR was performed using SYBR Green master mix in a Biorad CFX thermocycler (Bio-Rad).

Western blotting

Cells were lysed with RIPA buffer containing proteinase inhibitors (Sigma). Equal amounts of lysed denatured protein were run on polyacrylamide-SDS gels (Biorad) and transferred to PVDF membranes (Millipore). Primary antibodies used were: TET1 (Milipore, 09-872), Dazl (Abcam, ab34139), Actin-HRP (Cell signaling, 5125), Tubulin (Sigma).

RNA immunoprecipitation

Two different methods were used to co-immunoprecipitate RNA complexes interacting with Dazl-GFP-V5 in mESCs. For the UV-crosslinking IP (CLIP), 10cm dishes of Dazl-GFP mESCs were washed with cold PBS and UV irradiated at 400 mJ/cm^2. Cells were lysed in CLIP lysis buffer (120mM NaCl, 50mM Tris pH7.5, 1% Igepal, 0,1% SDS, 0,5% sodium deoxycholate) supplemented with protease inhibitor cocktail (Sigma), 40U/ml RNasin (Promega), 2mM vanadyl ribonucleotide complex (Sigma) and 1mM DTT. Cell lysates were cleared by centrifugation, precleared and incubated with anti-V5 antibody (Invitrogen), or normal mouse IgG (Sigma) for 2-3 hours at 4°C. Dynabeads protein A (Invitrogen) were preblocked in 2% BSA and added to the cell lysates with antibodies for 1-2 hours at 4°C. The beads were washed 3 times in PXL buffer (1x PBS, 0,1% sodium dodecyl sulfate, o,5% Igepal, 0,5% sodium deoxycholate) with protease inhibitors and RNase inhibitors, followed by 2 times washing in high salt PXL buffer (0,5M NaCl) and 1 time in Wash buffer (150mM NaCl, 50mM Tris pH 7.5, 0,05% Igepal). Beads were resuspended in 100ul wash buffer and incubated with proteinase K (20ug per sample) for 20 min at 37°C. RNA was extracted from the beads using Trizol reagent (Invitrogen). For the native RNA-IP, mESCs were lysed in RIP lysis buffer (150mM NaCl, 50mM Tris pH 7.5, 5mM $MgCl_2$, 0,5% Igepal) with all supplements. IP was performed similarly and beads were washed 5 times in Wash buffer.

Mass spectrometry based proteomics analysis

Sample preparation was performed as in Munoz *et al.* [2]. Peptides were chemically labeled with stable isotope dimethyl labeling [3]. Briefly, peptides were labeled with a mixture of formaldehyde-H2 and sodium cyanoborohydride ("light" reagent for "Day 0"), formaldehyde-D2 with cyanoborohydride ("intermediate" reagent for "Day 12") and ^{13}C-D2-formaldehyde with cyanoborodeuteride ("heavy" reagent for "Day 3"). In a second biological replica experiment, these reagents were swapped. Prior to the MS analysis, both replicates were fractionated using strong cation exchange (SCX) systems according to (Pinkse *et al*, 2008) [4]. Peptides were further separated and analyzed by nanoflow LC-MS/MS using an LTQ-Orbitrap Velos mass spectrometer (Thermo Electron, Bremen, Germany) coupled to an Agilent 1200 HPLC system (Agilent Technologies, Waldbronn, Germany). MS data was acquired, processed and

quantified with Proteome Discoverer (version 1.3, Thermo) using identical workflows as described in Munoz *et al.* [2].

RNA-sequencing of blastocyst embryos

This procedure is based on the protocol of Tang *et al*. [5]. Single blastocysts were transferred into a lysis buffer, followed by reverse transcription carried on the whole blastocyst lysate. cDNAs are then amplified by 18 + 9 cycles of PCR and the resulting amplified cDNAs are used to construct a sequencing library, which were used for deep sequencing using the SOLiD system. Q-PCRs for spike RNAs were added to the lysis buffer in known quantities differing by the factor of 10 (40 copies, 400 copies etc) to make sure that despite increasing the amount of starting material the spike RNAs/cDNAs were proportionally amplified [6]. Differential expression was analyzed using the Bayesian method according to the protocol of Kharchenko and collegues [7].

Single molecule FISH

E3.5 embryos were fixed in 4% formaldehyde and permeabilized with 0.5% Triton X-100. ESCs were fixed in 4% paraformaldehyde and permeabilized in 100% ethanol. Probe libraries were designed and fluorescently labeled as previously described [8]. The Dazl, Oct4 and Sox2 probe libraries consist of 48 oligonucleotides of 20 bp length (Table EV1) complementary to the coding sequence of the genes. Embryos and ESCs were hybridized overnight with Dazl-TMR, Oct4-Cy5 or Sox2-A594 probes at 30^{o}C, as previously described [8]. DAPI and Phalloidin-Alexa488 (Life-Technologies) were added during washes. Images were acquired on a Perkin-Elmer Spinning Disc confocal microscope with a 100x oil-immersion objective (numerical aperture 1.4) using Perking Elmer Volocity software. Images were recorded as stacks with a z spacing of 0.3 µm. Diffraction-limited dots corresponding to single mRNA molecules were automatically detected using custom Matlab software, based on previously described algorithms [8]. Briefly, the images were first filtered using a three-dimensional Laplacian of Gaussian filter, followed by selection of the intensity threshold at which the number of connected components was least sensitive to the threshold. All embryo isolation experiments were performed according to Dutch guidelines for care and use of laboratory animals and were approved by the animal welfare committee of Utrecht University.

Dot blot analysis

Genomic DNA (50ng per sample) was denatured in 0.4M NaOH at 95°C and neutralized with 1M NH_4OAc on ice. DNA was spotted on a Hybond-N+ nitrocellulose membrane (GE Healthcare) using a Bio-Dot SF apparatus (Bio-Rad). The membrane was washed in 2× SSC, dried on Whatman 3MM paper UV cross-linked 120,000 µJ $cm^{-2.}$ The membrane was blocked in 5% BSA, incubated with primary mouse anti-5-methylcytosine antibody (Abcam 1:2000) or rabbit anti-5-hydroxymethylcytosine antibody (Active motif 1:10000) overnight at 4°C. The membrane was washed in PBS-T and incubated with secondary HRP-conjugated goat anti-mouse antibody (BD Biosciences 1:5000) or HRP-conjugated goat anti-rabbit (Santa Cruz 1:5000) for 1 hour at room temperature. The membrane was visualed by chemiluminescence with Pierce ECL (Fisher).

Accession numbers

All RNA-IP microarray data have been deposited at the NCBI Gene Expression Omnibus (GEO) under accession numbers GSE69356 and GSE69357.

Supplementary references

1. Mazzoni EO *et al* (2011) Embryonic stem cell-based mapping of developmental transcriptional programs. *Nat Methods* **8:** 1056-1058
2. Munoz J, Low TY, Kok YJ, Chin A, Frese CK, Ding V, Choo A, Heck AJ The quantitative proteomes of human-induced pluripotent stem cells and embryonic stem cells. *Mol Syst Biol* **7:** 550
3. Boersema PJ, Raijmakers R, Lemeer S, Mohammed S, Heck AJ (2009) Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat Protoc* **4:** 484-494
4. Pinkse MW, Mohammed S, Gouw JW, van Breukelen B, Vos HR, Heck AJ (2008) Highly robust, automated, and sensitive online TiO2-based phosphoproteomics applied to study endogenous phosphorylation in Drosophila melanogaster. *J Proteome Res* **7:** 687-697
5. Tang F, Barbacioru C, Nordman E, Li B, Xu N, Bashkirov VI, Lao K, Surani MA (2010) RNA-Seq analysis to capture the transcriptome landscape of a single cell. *Nat Protoc* **5:** 516-535
6. Hackett JA, Sengupta R, Zylicz JJ, Murakami K, Lee C, Down TA, Surani MA (2013) Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science* **339:** 448-452
7. Kharchenko PV, Silberstein L, Scadden DT (2014) Bayesian approach to single-cell differential expression analysis. *Nat Methods* **11:** 740-742
8. Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S (2008) Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods* **5:** 877-879