Cell Stem Cell, Volume *29*

Supplemental Information

8C-like cells capture the human zygotic genome activation program *in vitro*

Jasmin Taubenschmid-Stowers, Maria Rostovskaya, Fátima Santos, Sebastian Ljung, Ricard Argelaguet, Felix Krueger, Jennifer Nichols, and Wolf Reik

Supplemental Figures S1 – S4

Figure S1, related to Figure 1. Defining human ZGA-like gene expression in 8-cell embryos and stem cells.

Figure S2, related to Figure 2, 3. 8CLC signature and transposon expression in human embryos and stem cells.

Figure S3, related to Figure 3, 4. Identification of 8CLCs among naïve hESCs by marker protein expression *in vitro*.

Figure S4, related to Figure 4. Characterization of *DUX4* overexpressing naïve hESCs and 8CLCs.

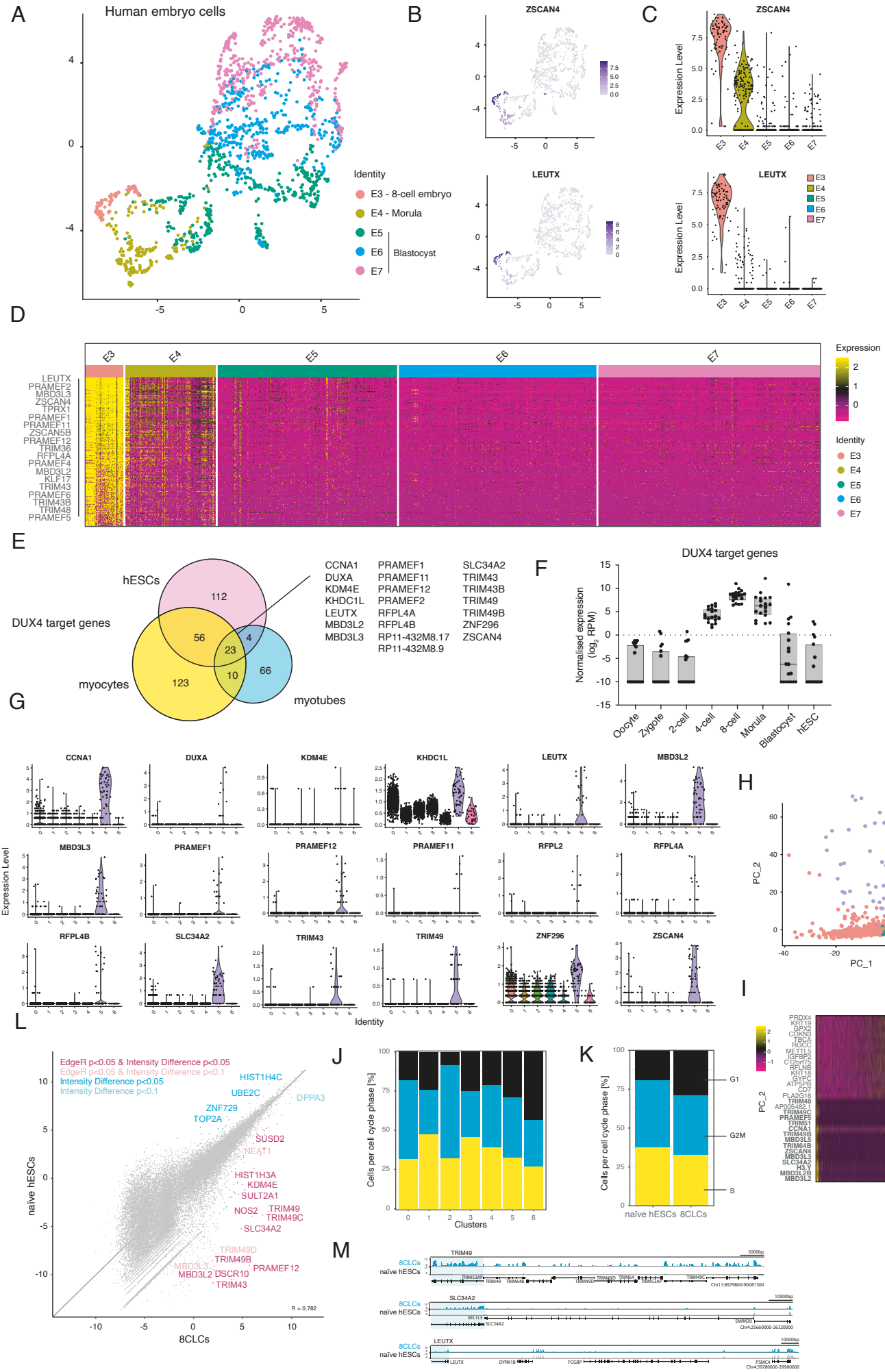
A
Human embryo cells
Identity
E3 - 8-cell embryo
E4 - Morula
E5
E6
E7
Blastocyst
B
ZSCAN4
LEUTX
C
ZSCAN4
LEUTX
Expression Level
E3
E4
E5
E6
E7
D
Expression
Identity
LEUTX
PRAMEF2
MBD3L3
ZSCAN4
TPRX1
PRAMEF1
PRAMEF11
ZSCAN5B
PRAMEF12
TRIM36
RFPL4A
PRAMEF4
MBD3L2
KLF17
TRIM43
PRAMEF6
TRIM43B
TRIM48
PRAMEF5
E
hESCs
DUX4 target genes
myocytes
myotubes
112
56
4
23
10
66
123
CCNA1
DUXA
KDM4E
KHDC1L
LEUTX
MBD3L2
MBD3L3
PRAMEF1
PRAMEF11
PRAMEF12
PRAMEF2
RFPL4A
RFPL4B
RP11-432M8.17
RP11-432M8.9
SLC34A2
TRIM43
TRIM43B
TRIM49
TRIM49B
ZNF296
ZSCAN4
F
DUX4 target genes
Normalised expression (log2 RPM)
Oocyte
Zygote
2-cell
4-cell
8-cell
Morula
Blastocyst
hESC
G
CCNA1
DUXA
KDM4E
KHDC1L
LEUTX
MBD3L2
MBD3L3
PRAMEF1
PRAMEF12
PRAMEF11
RFPL2
RFPL4A
RFPL4B
SLC34A2
TRIM43
TRIM49
ZNF296
ZSCAN4
Expression Level
Identity
H
Identity
PC_1
PC_2
I
PC_2
L
EdgeR p<0.05 & Intensity Difference p<0.05
EdgeR p<0.05 & Intensity Difference p<0.1
Intensity Difference p<0.05
Intensity Difference p<0.1
HIST1H4C
UBE2C
DPPA3
ZNF729
TOP2A
SUSD2
NEAT1
HIST1H3A
KDM4E
SULT2A1
NOS2
TRIM49
TRIM49C
SLC34A2
TRIM49D
TRIM49B
MBD3L5
PRAMEF12
MBD3L2
DSCR10
TRIM43
R = 0.782
naïve hESCs
8CLCs
J
Cells per cell cycle phase [%]
Clusters
K
Cells per cell cycle phase [%]
G1
G2M
S
naïve hESCs
8CLCs
M
TRIM49
SLC34A2
LEUTX
8CLCs
naïve hESCs

Figure S1

Figure S1, related to Figure 1. Defining human ZGA-like gene expression in 8-cell embryos and stem cells.

(A) UMAP of human 8-cell (E3), morula (E4) and blastocyst (E5 – E7) stage embryo cells based on single cell RNA expression data (Petropoulos et al., 2016). (B) UMAP and (C) Violin plots of normalized, scaled *ZSCAN4* and *LEUTX* expression in human pre-implantation embryo cells. (D) Heatmap of 8-cell embryo signature gene expression (rows) in human pre-implantation embryo cells (E3 – E7, columns). Markers of human ZGA are highlighted. (E) Differentially expressed genes that are upregulated upon *DUX4* overexpression in primed hPSCs (magenta), myocytes (yellow), and endogenous *DUX4* upregulation in human patient myotubes (blue) (Hendrickson et al., 2017; Jiang et al., 2020; Yao et al., 2014). Shared genes across all three datasets are highlighted. (F) Normalized expression of *DUX4* target genes during human pre-implantation development (oocyte to blastocyst) and in hESCs (Yan et al., 2013). (G) Violin plots of selected ZGA markers in clustered human naïve ESCs, based on normalized, scaled single cell RNA expression levels. Clusters were determined as in Figure 1 (B). (H) PC1 and PC2 of clustered naïve human HNES1 ESCs. (I) PC 2 loadings of naïve HNES1 are shown. ZGA marker genes are highlighted in bold. (J) Cell cycle analysis of single cell transcriptome data in naïve hESCs clusters (cluster 0 – 6) or (K) specifically 8CLCs as compared to naïve hESCs. (L) Differentially expressed genes, as defined by EdgeR or Intensity Difference analysis, between 8CLCs (n=5/480, based on the expression of ZGA markers) and naïve H9 PSCs cultured in t2iLGö (Messmer et al., 2019). (M) Genome browser tracks of pseudo-bulked ZGA marker genes (*TRIM49, SLC34A2, LEUTX*) in 8CLCs (as in G) and naïve H9 hPSCs (Messmer et al., 2019).

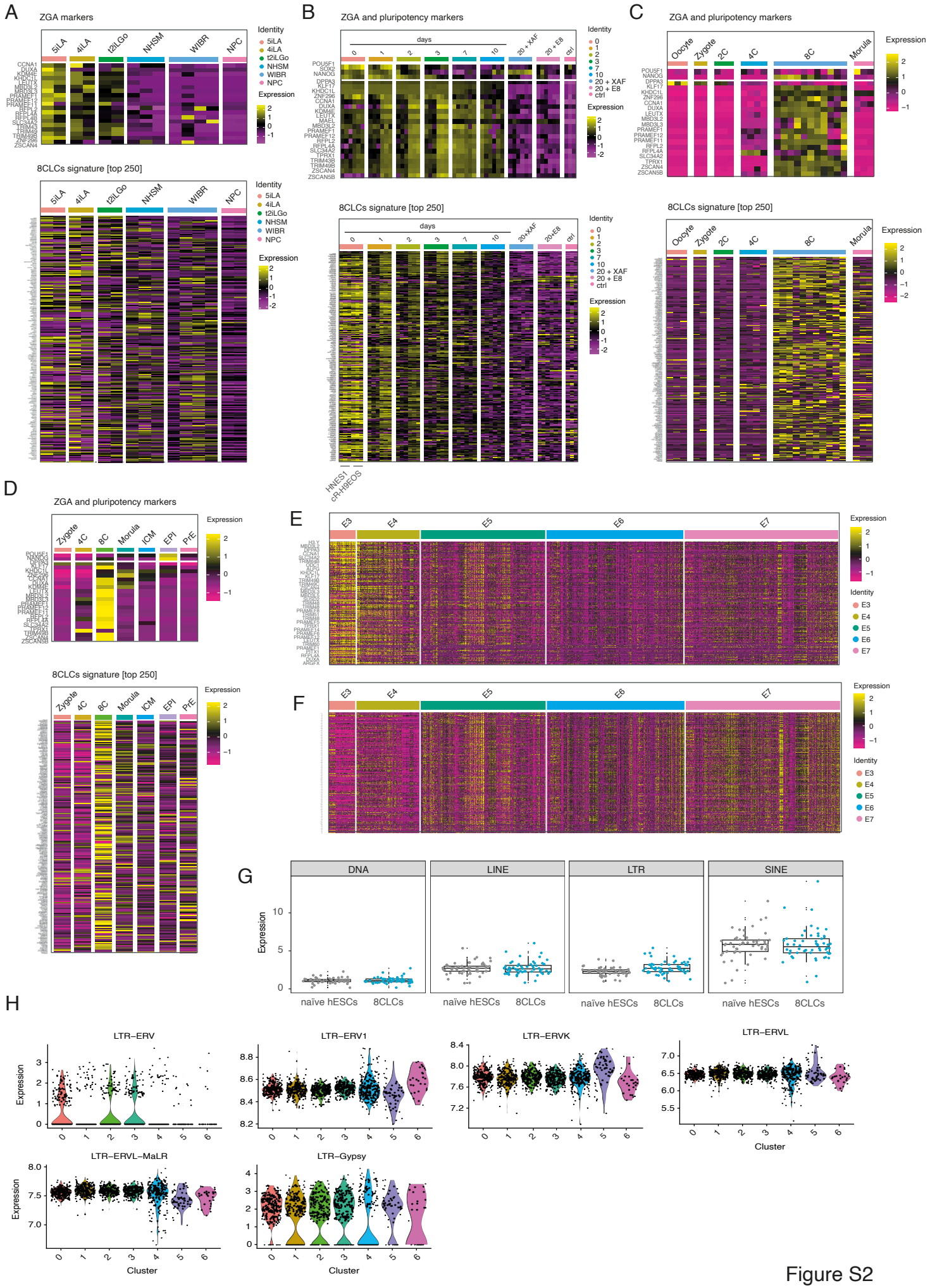
A
ZGA markers
5iLA
4iLA
t2iLGo
NHSM
WIBR
NPC
Identity
Expression
8CLCs signature [top 250]
B
ZGA and pluripotency markers
days
20 + XAF
20 + E8
ctrl
HNES1
cR-H9EOS
C
ZGA and pluripotency markers
Oocyte
Zygote
2C
4C
8C
Morula
D
ZGA and pluripotency markers
Zygote
4C
8C
Morula
ICM
EPI
PrE
8CLCs signature [top 250]
E
E3
E4
E5
E6
E7
F
G
DNA
LINE
LTR
SINE
Expression
naïve hESCs
8CLCs
H
LTR–ERV
LTR–ERV1
LTR–ERVK
LTR–ERVL
LTR–ERVL–MaLR
LTR–Gypsy
Cluster

Figure S2

Figure S2, related to Figure 2, 3. 8CLC signature and transposon expression in human embryos and stem cells.

(A) Heatmap of ZGA markers (top panel) and 8CLCs signature genes (bottom panel) of naïve (5iLA, 4iLA, t2iLGö), pseudo-naïve (NHSM), primed (WIBR) and differentiated cells (NPCs, neuronal progenitor cells). The analysis was done on bulk RNA expression data. (B) Heatmap of pluripotency genes (i.e. *POU5F1, SOX2, NANOG*), ZGA markers (top panel) and 8CLCs signature genes (bottom panel) in two naïve PSC lines (HNES1, cR-H9EOS in triplicate each). Cells were transitioned from the naïve into the primed state, in two different primed media (E8, XAF) (Rostovskaya et al., 2019). The days indicate the days of transition (day 0 – day 10, d20). Gene expression of control primed H9 cells is shown as well (ctrl). (C), (D) Heatmap of pluripotency genes (*POU5F1, NANOG*), ZGA markers (top panels) and 8CLCs signature genes (bottom panels) in human pre-implantation embryos from two different studies (C) (Xue et al., 2013) (D) (Stirparo et al., 2018). (E) 8CLCs signature and (F) naïve hPSCs marker gene expression that distinguishes them from 8CLCs, in human embryos (E3 – E7) (Petropoulos et al., 2016). (G) DNA, LTR, LINE, and SINE transposable element expression in naïve hESCs and 8CLCs. Reads are shown as percentage of total reads. (H) Normalized reads of LTR family transposons expressed in 8CLCs and naïve hESCs from single cell RNA sequencing data (for clusters see Figure 1B).

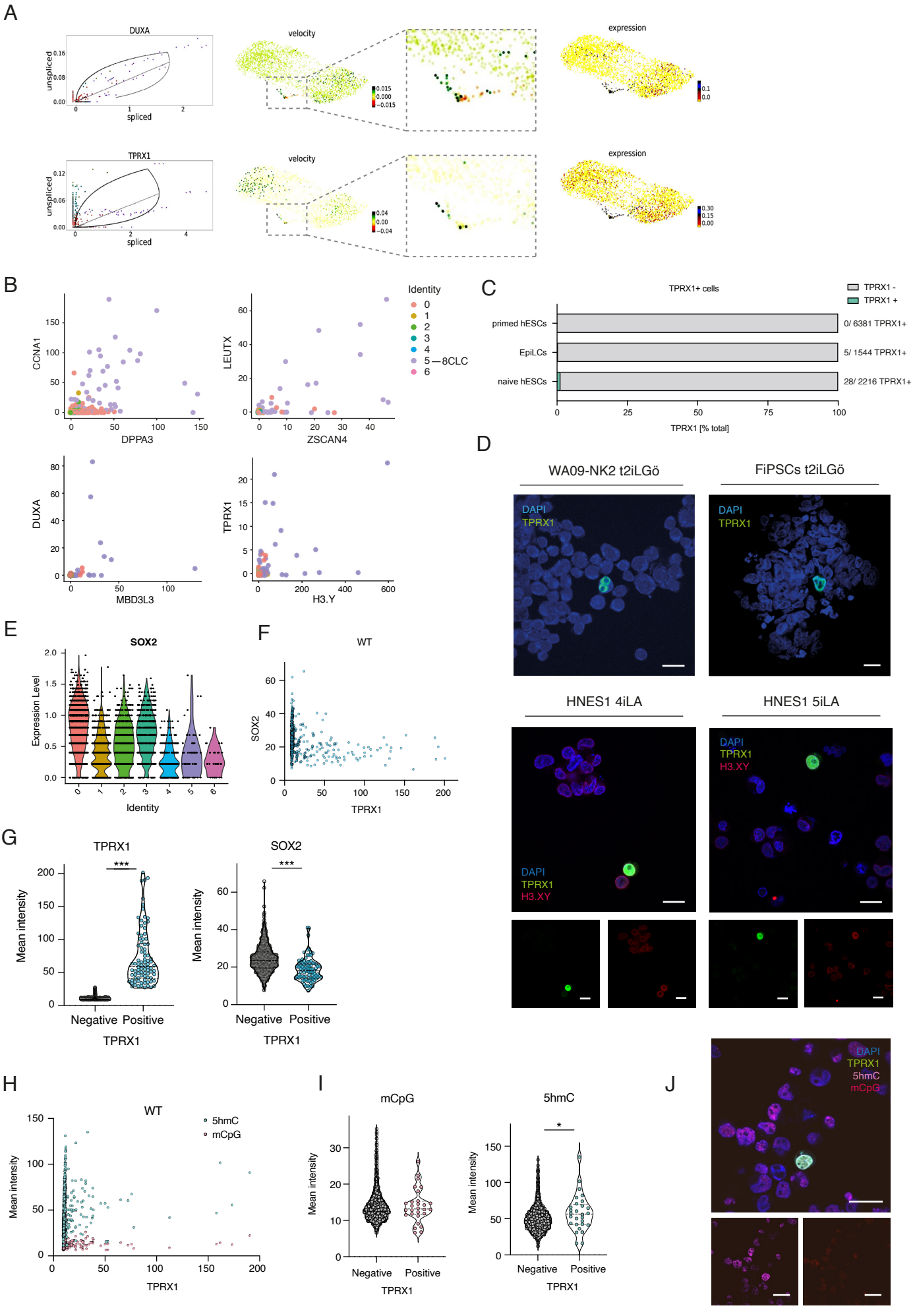
A
DUXA
velocity
expression
unspliced
spliced
TPRX1
B
Identity
0
1
2
3
4
5—8CLC
6
CCNA1
DPPA3
LEUTX
ZSCAN4
DUXA
MBD3L3
TPRX1
H3.Y
C
TPRX1+ cells
TPRX1 -
TPRX1 +
primed hESCs
EpiLCs
naive hESCs
0/ 6381 TPRX1+
5/ 1544 TPRX1+
28/ 2216 TPRX1+
TPRX1 [% total]
D
WA09-NK2 t2iLGö
FiPSCs t2iLGö
HNES1 4iLA
HNES1 5iLA
DAPI
TPRX1
H3.XY
E
SOX2
Expression Level
Identity
F
WT
SOX2
TPRX1
G
TPRX1
SOX2
Mean intensity
Negative
Positive
***
H
WT
5hmC
mCpG
Mean intensity
TPRX1
I
mCpG
5hmC
Mean intensity
Negative
Positive
TPRX1
*
J
DAPI
TPRX1
5hmC
mCpG

Figure S3

Figure S3, related to Figure 3, 4. Identification of 8CLCs among naïve hESCs by marker protein expression *in vitro*.

(A) Left panel: RNA velocity analysis of *DUXA* and *TPRX1* in individual 8CLCs and naïve hESCs (for clusters see Figure 1B); steady-state ratios (black lines), overall dynamics (black curve) and ratios of unspliced vs. spliced mRNA in single cells; middle panel: RNA velocity of *DUXA* and *TPRX1*; right panel: RNA expression levels of *DUXA* and *TPRX1*. (B) Scatter plots of 8CLC marker gene co-expression in naïve hESCs and 8CLC clusters. (C) Quantification of TPRX1-positive (TPRX1+) naïve hESCs, EpiLCs and primed ESCs as detected by immunofluorescence staining. Percentages are shown, total numbers are indicated on the right-hand side of the graph. (D) Upper panel: IF staining of TPRX1 in reprogrammed naïve H9 PSCs (left) and fibroblast derived naïve iPSCs (FiPSCs, right) cultured in t2iLGö. Bottom panel: IF staining of TRPX1 and H3.X/Y in naïve HNES1 cells cultured in 4iLA (left) or 5iLA (right). Scale bar, 20µm. (E) Violin plot of normalized, scaled *SOX2* expression in clustered naïve human ESCs (for clusters see Figure 1B). (F) TPRX1 and SOX2 mean fluorescent signal intensity per cell (mid optical slice) in antibody stained, cytospun naïve HNES1 cells cultured in PXGL. TPRX1-positive cells and control cells were selected to assess SOX2 protein levels. (G) Quantification of mean TPRX1 (left) and SOX2 (right) intensity levels in TPRX1-positive and TPRX1-negative selected HNES1 cells. Images were taken from four independent experiments and pooled. p-value *P<0.05, **P<0.01, ***P<0.001, absence of stars (non-significant, ns): p-value>0.05; two tailed Mann-Whitney U test. (H) TRPX1 and 5hmC/ mCpG mean fluorescent signal intensity per cell (mid optical slice) in antibody stained, cytospun naïve HNES1 cells cultured in PXGL. (I) Mean mCpG (left) and 5hmC (right) intensity levels in TPRX1-positive and TPRX1-negative HNES1 cells. Images were taken from at least two experiments and pooled. p-value *P<0.05, **P<0.01, ***P<0.001, absence of stars (non-significant, ns): p-value>0.05; unpaired, two tailed Mann-Whitney U test. (J) IF staining of TRPX1, 5hmC and mCpG in naïve HNES1 cells cultured in PXGL. Scale bar, 20µm.

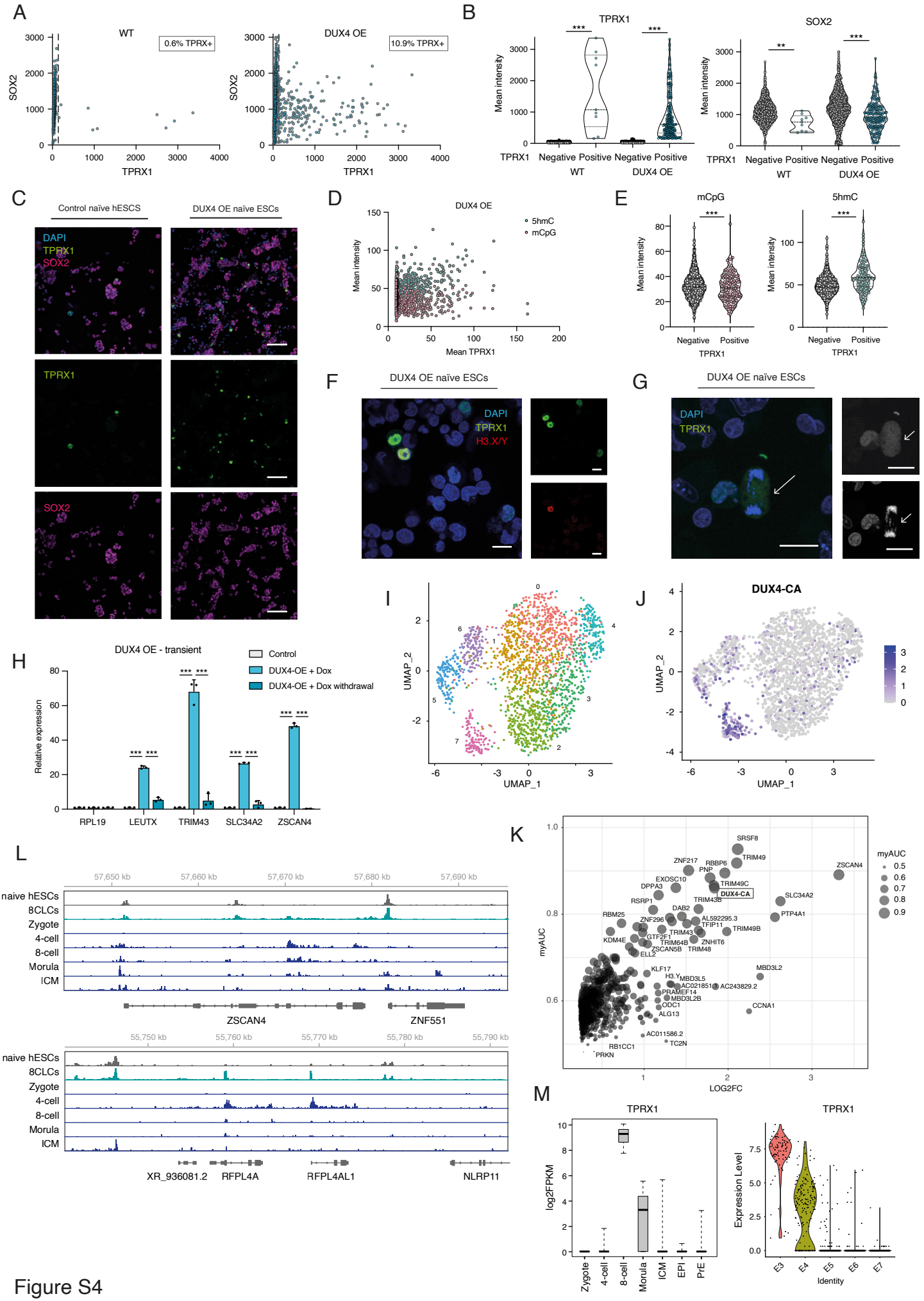
A
WT
0.6% TPRX+
DUX4 OE
10.9% TPRX+
SOX2
TPRX1
B
TPRX1
SOX2
Mean intensity
Negative Positive Negative Positive
WT
DUX4 OE
C
Control naïve hESCS
DUX4 OE naïve ESCs
DAPI
TPRX1
SOX2
D
DUX4 OE
5hmC
mCpG
Mean intensity
Mean TPRX1
E
mCpG
5hmC
Negative Positive
F
DUX4 OE naïve ESCs
DAPI
TPRX1
H3.X/Y
G
DUX4 OE naïve ESCs
H
DUX4 OE - transient
Control
DUX4-OE + Dox
DUX4-OE + Dox withdrawal
Relative expression
RPL19 LEUTX TRIM43 SLC34A2 ZSCAN4
I
UMAP_1
UMAP_2
J
DUX4-CA
K
myAUC
LOG2FC
SRSF8
TRIM49
ZNF217
RBBP6
PNP
ZSCAN4
EXOSC10
TRIM49C
DPPA3
DUX4-CA
TRIM43B
SLC34A2
RSRP1
DAB2
PTP4A1
RBM25
ZNF296
AL592295.3
TFIP11
TRIM43
TRIM49B
KDM4E
GTF2F1
TRIM64B
ZNHIT6
ZSCAN5B
TRIM48
ELL2
KLF17
MBD3L2
H3.Y
MBD3L5
AC021851.1
AC243829.2
PRAMEF14
MBD3L2B
ODC1
CCNA1
ALG13
AC011586.2
TC2N
RB1CC1
PRKN
L
57,650 kb 57,660 kb 57,670 kb 57,680 kb 57,690 kb
naive hESCs
8CLCs
Zygote
4-cell
8-cell
Morula
ICM
ZSCAN4
ZNF551
55,750 kb 55,760 kb 55,770 kb 55,780 kb 55,790 kb
XR_936081.2
RFPL4A
RFPL4AL1
NLRP11
M
TPRX1
log2FPKM
Zygote 4-cell 8-cell Morula ICM EPI PrE
Expression Level
E3 E4 E5 E6 E7
Identity

Figure S4

Figure S4, related to Figure 4. Characterization of *DUX4* overexpressing naïve hESCs and 8CLCs.

(A) TPRX1 and SOX2 mean fluorescent signal intensity per cell in antibody-stained control (WT) and *DUX4* overexpressing (*DUX4* OE) naïve HNES1 cells cultured in PXGL and plated on coverslips. The percentage of TPRX1-positive cells is indicated in the graphs, the cut-off is shown as a dashed line (x=150). (B) Mean TPRX1 (left) and SOX2 (right) intensity levels in TPRX1-positive and TPRX1-negative control (WT) and *DUX4* overexpressing (*DUX4* OE) HNES1 cells; two independent experiments. Cut-off as in (A). p-value *P<0.05, **P<0.01, ***P<0.001, absence of stars (non-significant, ns): p-value>0.05; two tailed Mann-Whitney U test. (C) IF staining of TPRX1 and SOX2 in control and *DUX4* overexpressing HNES1 cells plated on coverslips. Scale bar, 100µm. (D) TRPX1 and 5hmC/ mCpG mean fluorescent signal intensity per cell (mid optical slice) in antibody stained, cytospun *DUX4* overexpression HNES1 cells cultured in PXGL. (E) Mean mCpG (left) and 5hmC (right) intensity levels in TPRX1-positive and TPRX1-negative *DUX4* overexpressing HNES1 cells; two independent experiments. p-value *P<0.05, **P<0.01, ***P<0.001, absence of stars (non-significant, ns): p-value>0.05; nonparametric Mann-Whitney U test. (F) IF staining of TPRX1 and H3.X/Y in *DUX4* overexpressing HNES1 cells cytospun onto coverslips. Scale bar, 20µm (G) IF staining of TPRX1 in *DUX4* overexpressing cytospun HNES1 cells. Dividing TPRX1-positive cells (anaphase chromosomes) are indicated (see arrows). Scale bar, 20µm. (H) Expression levels of ZGA marker genes measured by RT-qPCR in *DUX4* overexpressing (DUX4-OE + Dox, 24h) cells as well as upon Dox withdrawal (DUX4-OE + Dox withdrawal, 24h + 72h). Data are shown as mean ± SD (n = 3 technical replicates) of fold-change compared to control HNES1, and are representative of three independent experiments. p-value *P<0.05, **P<0.01, ***P<0.001, absence of stars (non-significant, ns): p-value>0.05; unpaired, two tailed Student's t-test. (I) UMAP of clustered human naïve ESCs and 8CLCs (cluster 7). (J) Normalized scaled *DUX4-CA* expression in clustered hESCs and 8CLCs visualized on a UMAP. (K) Marker gene expression signature of 8CLCs (cluster 7 see S4I). AUC, area under curve; LOG2FC, log2 fold-change. (L) Genome browser views of accessibility at genomic loci of ZGA markers *ZSCAN4* (data range 0.0 – 0.2 stem cells, 0.0 – 1.0 embryo data) and *RFPL4A* (data range 0.0 – 0.1 stem cells, 0.0 – 1.0 embryo data) in naïve hESCs (grey) and 8CLCs (petrol) as compared to human embryos (blue) (Liu et al., 2019). (M) Normalized bulk (left) and single cell (right) *TPRX1* RNA expression levels in preimplantation human embryo cells (Petropoulos et al., 2016; Stirparo et al., 2018). Left: Figure adapted from https://app.stemcells.cam.ac.uk/human-embryo/.