**Supplementary Data**

## Index

## Supplementary Figure Legends

### Supplementary Fig.1

(a) Standard (left) and OxBiS treated Bisulfite Sequencing (BiS) profiles for a somatic male sample (46,XY SOM), female hESC line H9 (46,XX hESC) and male hESC line H1 (46,XY hESC). The region assessed by BiS analysis is the same as in Figure 1B. The methylation status of each CpG is indicated by black (methylated) and white (unmethylated), while each row is a single clone. (b) qChIP data for H3K9me3 at DXZ4 for female H9 hESC (46,XX hESC) and H9 derived EBOGs (46,XX EBOG). Each data set is graphed as a percent of input. Error bars indicate standard deviation from the mean of triplicate qPCR reactions. I-V are as in Figure 1C.

### Supplementary Fig.2

Differentiation-associated gene expression changes between hESCs and their corresponding EBOGs. (a) Quantitation of pluripotent marker gene expression in female (46,XX H9-left and H7-right) and male (46,XY H1) hESCs, female (46,XX RPE1-left and HDF-FET-right) and male (46,XY 1140) somatic cells (SOM), and female and male EBOG. Data is normalized to GAPDH expression and graphed relative to expression in H9. Error bars indicate standard deviation from the mean of triplicate qPCR reactions. (b) Indirect-immunofluorescence images showing the distribution of the indicated pluripotency markers in female (H9) hESCs (46,XX hESC), female (RPE1) somatic cells (46,XX SOM), as well as in female & male EBOG. For each sample the top row shows NANOG in green and SSEA4 in red. Nuclei are counter-stained with DAPI (white or blue). The white bars at the bottom of the merge images indicate 5μm. (c) Analysis of differentiation-associated gene expression in female (H9) hESCs (46,XX hESC) and their corresponding EBOGs by endpoint rtPCR analysis. Each is an inverted image of an ethidium bromide stained agarose gel.

### Supplementary Fig. 3

Characterization of transcripts around the distal EST cluster. (a) At the top is the RNA-seq profile for H1 taken from Fig. 3b and on this the area of focus is indicated, corresponding to 115,020,151-115,023,455 bp of the X chromosome (hg19). At the top is the zoomed-in RNA-seq profile for hESC H1 from the sense (+) and anti-sense strand (-) and the origins of the RNA-seq data from polyadenylated (polyA+) and nonpolyadenylated (polyA-) RNA. The y-axis is 0-300 reads. Beneath this is a schematic representation of adjacent DXZ4 monomers from the main array. The green box indicates the DXZ4 internal promoter region and the blue box the CTCF binding site. Regions of the distal EST interval showing sequence identity to DXZ4 are indicated along with the percent nucleotide identity at the top. Beneath this is the sequence verified DANT1 ATT and a DANT2 transcript inferred from EST entries BX642309 (5' read of clone DFKZp686H17267) and BX642310 (3' read of clone DFKZp686H17267). The dashed line in the open box indicates the portion of this clone that is not sequenced, whereas the solid black regions are sequence verified. Beneath this are independent unspliced EST annotations from the EST cluster. The

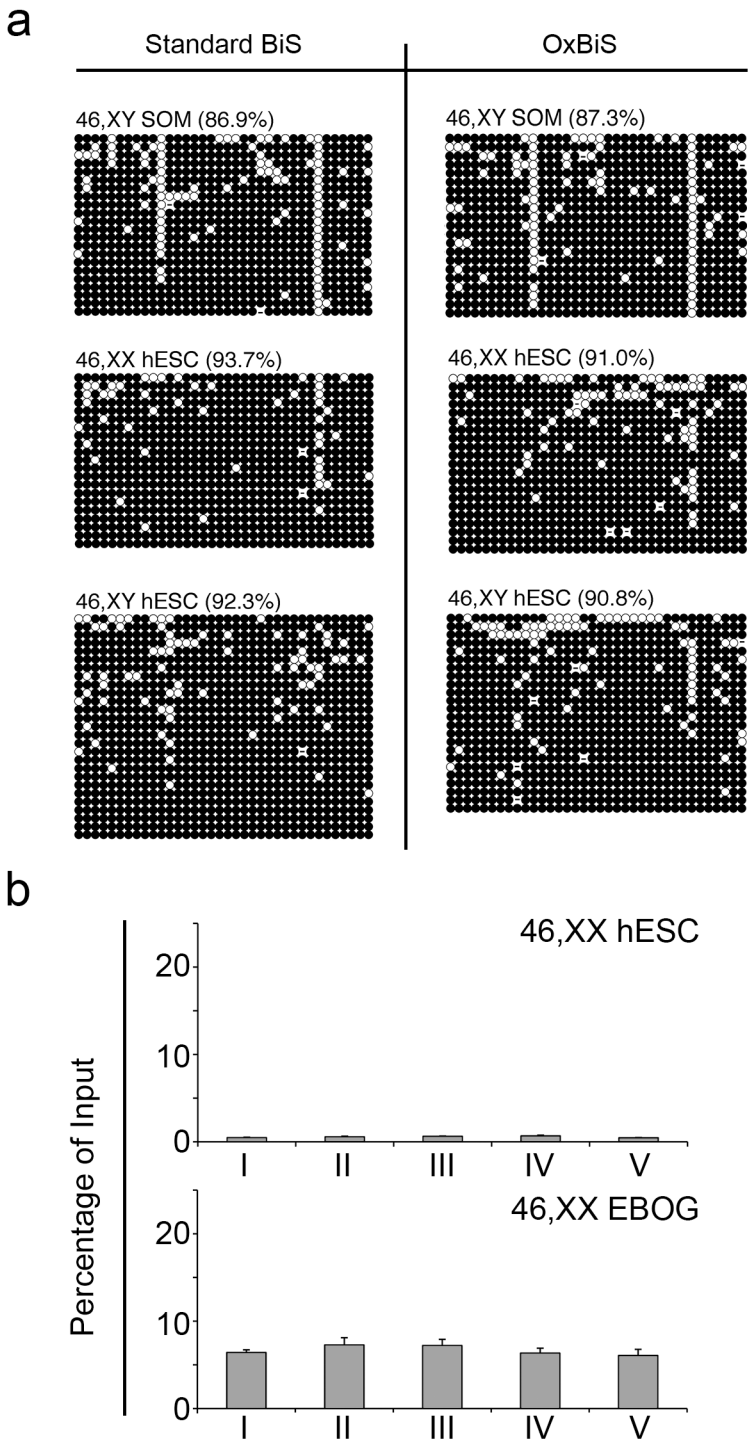
black arrow points to BX642310 from above. (b) Validation of the *DANT1* and *DANT2* array-traversing transcripts (ATT). A schematic representation of the DXZ4 interval is shown at the top, under which are novel ATTs identified in this study by RT-PCR, TA cloning, sequencing and mapping back to the genomic interval. Vertical lines represent exons and introns are horizontal lines with chevrons indicating the direction of transcription. The schematics correspond to the following accession numbers: (a) KM192213, (b) KM192212, (c) KM192214, (d) KM192215, (e) KM192216 and (f) KM192217. Both "a" and "b" were detected in H1 and H9 hESC, "c" was detected in H1 and H9 hESC (Figure 4) as well as CE, WB, FB, FL, HE, TR, UT, SC and ST, (d) was detected in FB, HE and UT, whereas (e) and (f) were detected in FB. Sample key, as in Fig. 4.

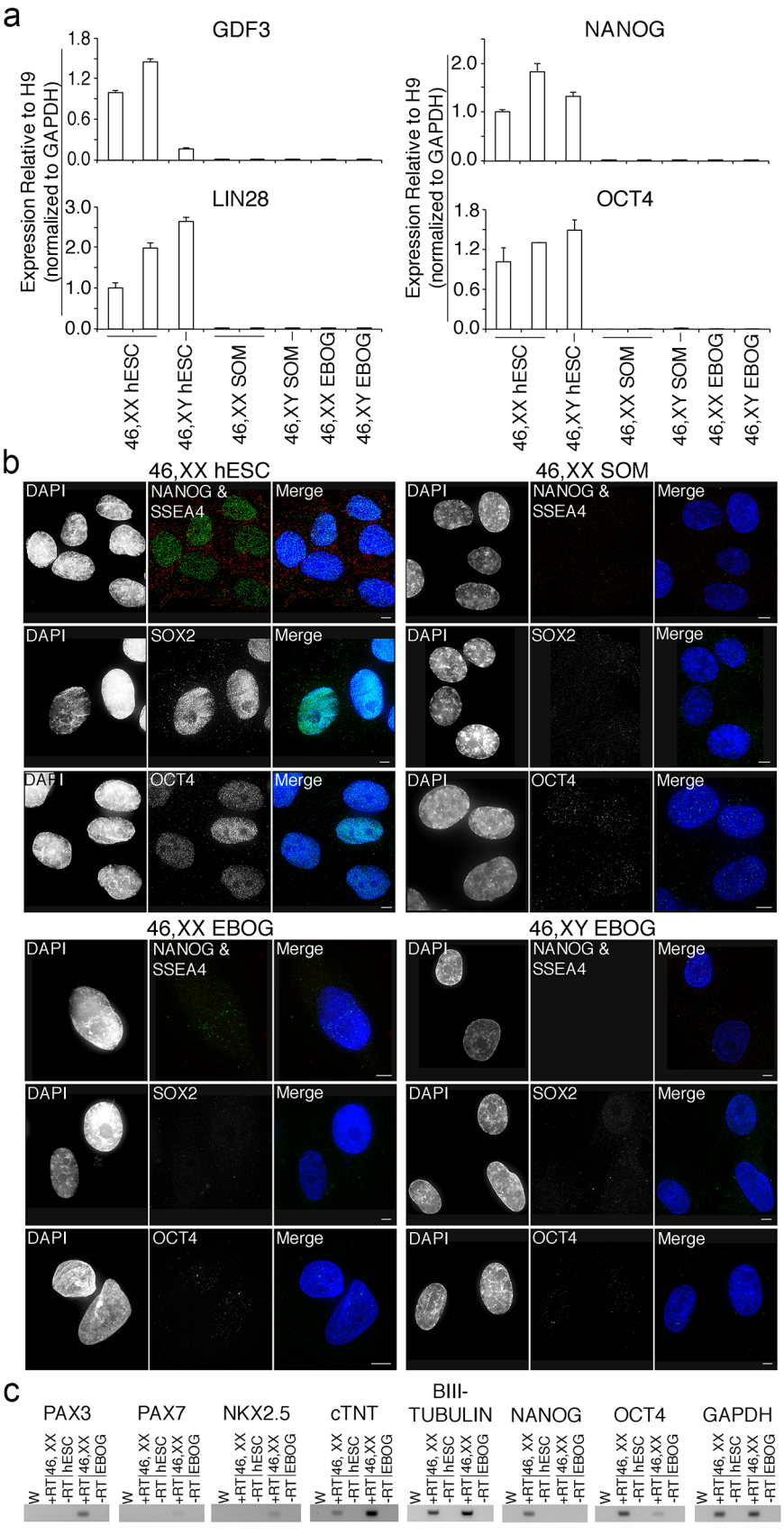
**Supplementary Fig. 4**
Expression of DXZ4 (Top) and total DANT2 (Bottom) in the 20 different human tissues assessed in Fig. 4b. Data represents qRT-PCR performed in triplicate and each is displayed as a percentage of GAPDH levels.

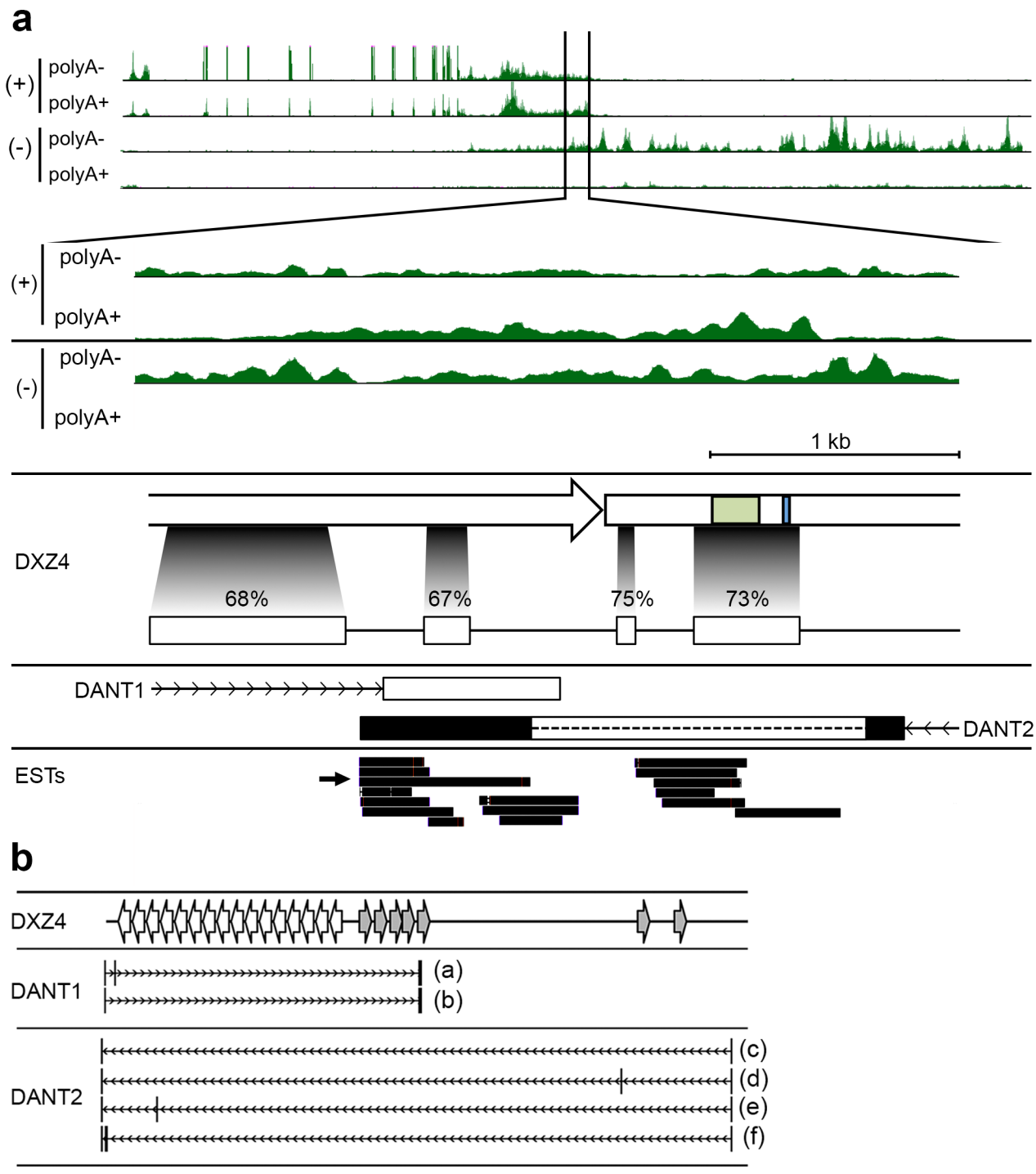
**Supplementary Fig. 5**
Detection of *DANT1*, *DANT2* and DXZ4 by direct-labeled RNA FISH probes in female H9 (top) and H7 (bottom) hESC. For each sample column 1 shows nuclei counterstained with DAPI (white), column 2 shows labeled signals indicated by arrowheads. Column 3 shows signals highlighted by arrows. Column 4 consists of a merge of the DAPI-staining (blue) with direct-RNA FISH in columns 2 (green) and 3 (red). The white bars at the bottom right of the merged images indicate 5μm.

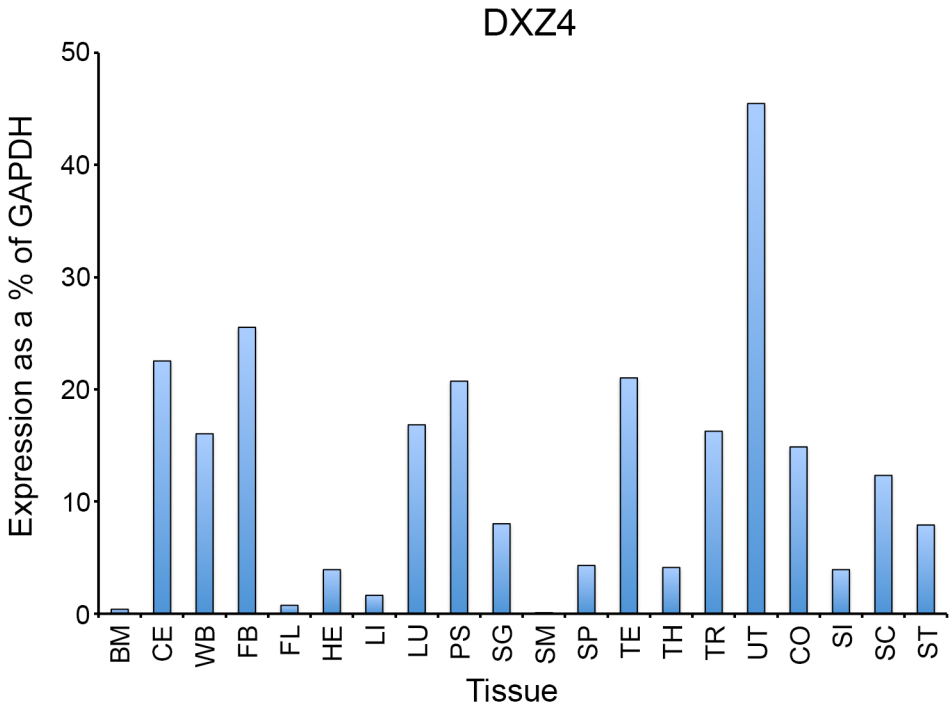
**Supplementary Fig. 6**
RNA and DNA FISH on RNAse-treated cells. (a) RNA FISH with direct-labeled probes (XIST, DANT1, DANT2 or DXZ4) in female (46,XX H9) and male hESCs (46,XY H1). Nuclei are counterstained with DAPI (Blue). Each panel shows data for the green or red probes to the transcripts indicated. The lack of signal is consistent with RNaseA treatment destroying the target RNA. The white bars at the bottom of the merge images indicate 5μm. (b) RNaseA treatment controls. Images are of 46,XX hTERT-RPE1 nuclei after RNA FISH (top row) or DNA FISH (bottom row). The probe is a spectrum-red direct-labeled DXZ4 BAC. Left images show nuclei not first treated with RNaseA, whereas right images show cells after pre-treatment with RNaseA. RNaseA treatment removes the RNA-FISH signal, but does not impact the DNA-FISH signal, confirming nucleic acid degradation is RNA specific.

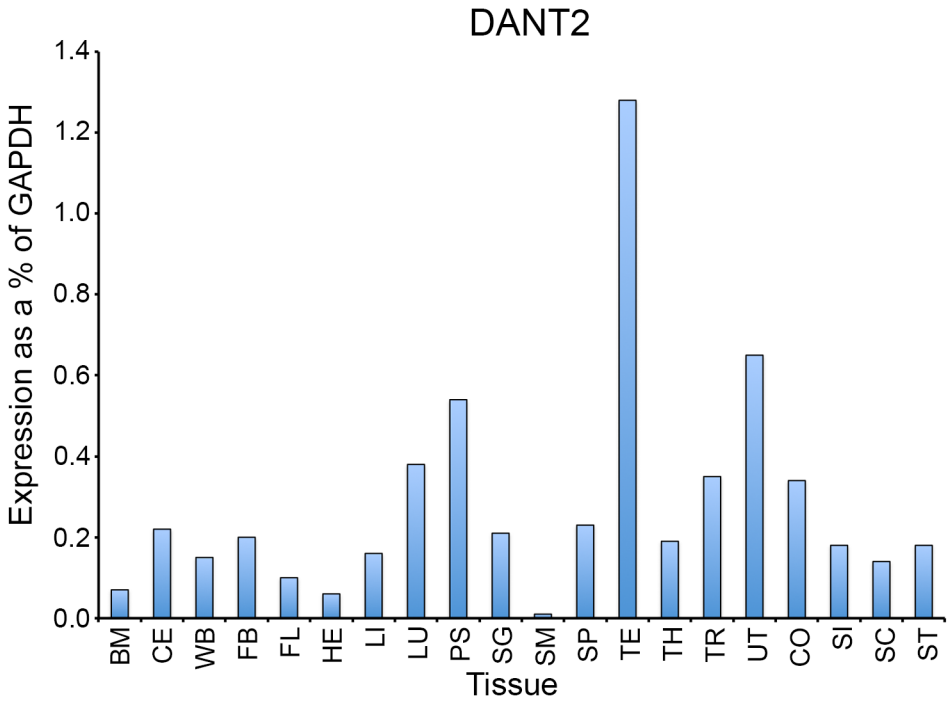
**Supplementary Fig. 7**
H3K9me3 qChIP data from the DANT1 promoter region assessed at locations I (F1.R1), II (F2.R2) and III (F3.R3) as shown in Figure 6. Each is graphed as a percent of input and is the mean of triplicate qPCR reactions. Error bars indicate standard deviation. Female (46,XX) data sets are on the left with male (46,XY) on the right. For each ChIP target, the top row shows data for female (H9) and male (H1) hESCs, middle row for female (RPE1) and male (BJ1) somatic cells and bottom row for EBOGs derived from the corresponding hESCs.

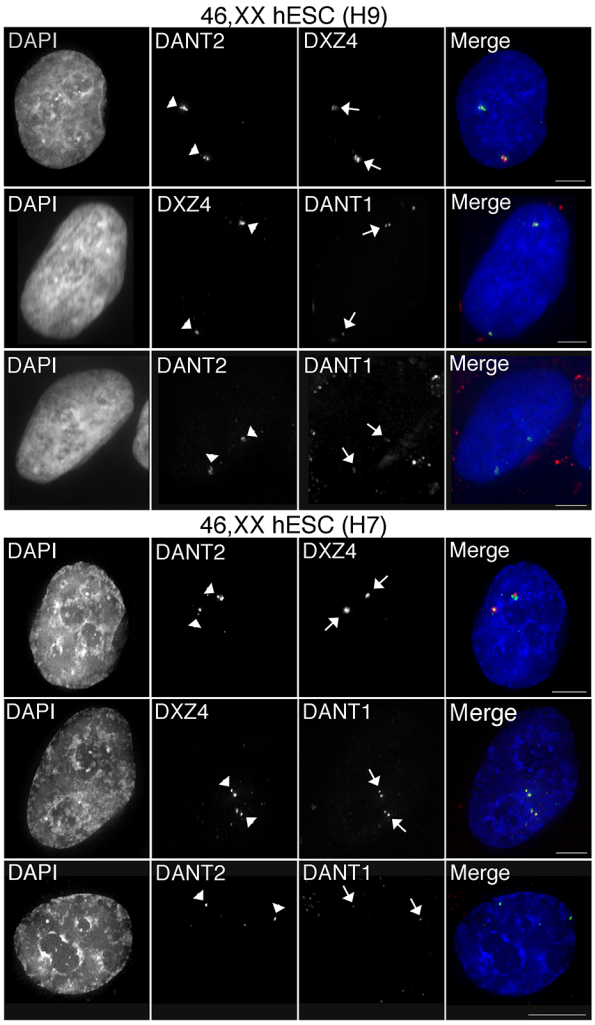

a
Standard BiS
OxBiS
46,XY SOM (86.9%)
46,XY SOM (87.3%)
46,XX hESC (93.7%)
46,XX hESC (91.0%)
46,XY hESC (92.3%)
46,XY hESC (90.8%)
b
46,XX hESC
Percentage of Input
20
10
0
I
II
III
IV
V
46,XX EBOG
20
10
0
I
II
III
IV
V

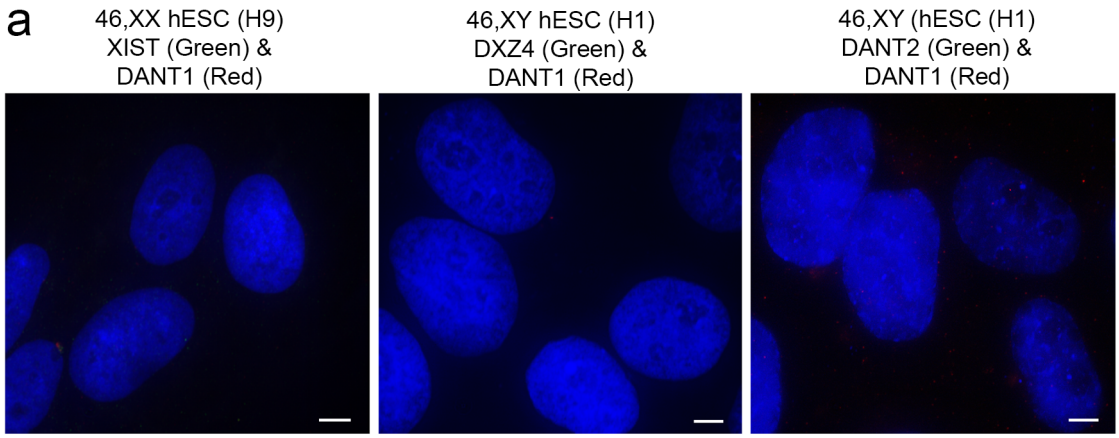

a
GDF3
NANOG
LIN28
OCT4
Expression Relative to H9 (normalized to GAPDH)
46,XX hESC
46,XY hESC
46,XX SOM
46,XY SOM
46,XX EBOG
46,XY EBOG
b
46,XX hESC
46,XX SOM
46,XX EBOG
46,XY EBOG
DAPI
NANOG & SSEA4
Merge
SOX2
OCT4
c
PAX3
PAX7
NKX2.5
cTNT
BIII-TUBULIN
NANOG
OCT4
GAPDH
W
+RT 46, XX
-RT hESC
+RT 46,XX
-RT EBOG

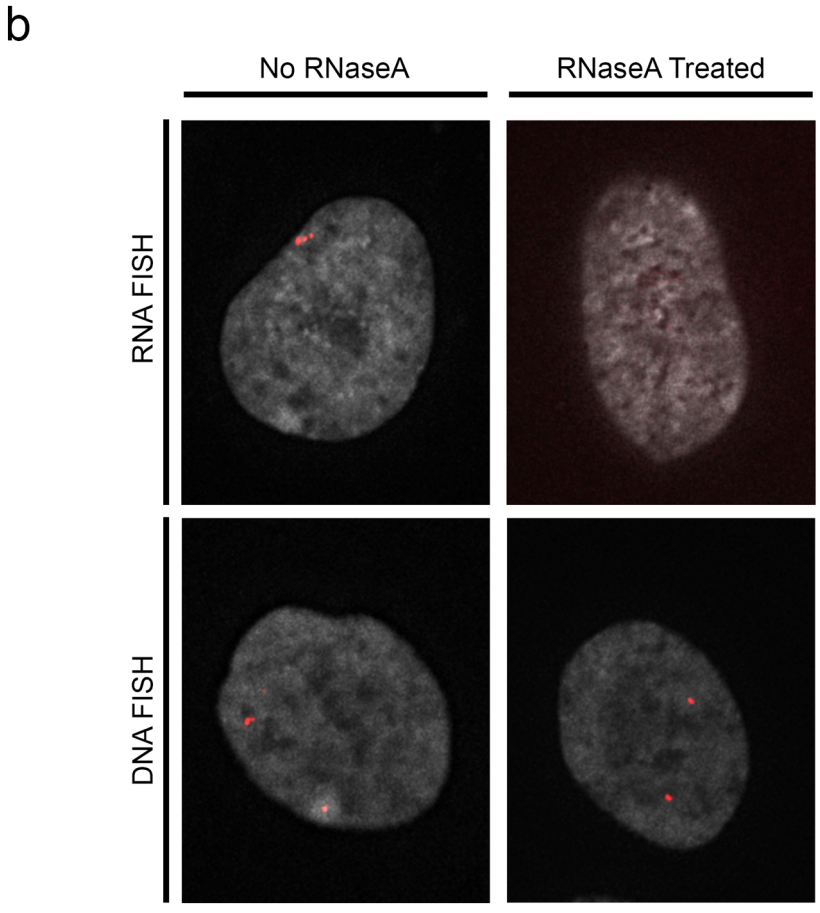

a
(+)
polyA-
polyA+
(-)
polyA-
polyA+
(+)
polyA-
polyA+
(-)
polyA-
polyA+
1 kb
DXZ4
68%
67%
75%
73%
DANT1
DANT2
ESTs
b
DXZ4
DANT1
(a)
(b)
DANT2
(c)
(d)
(e)
(f)

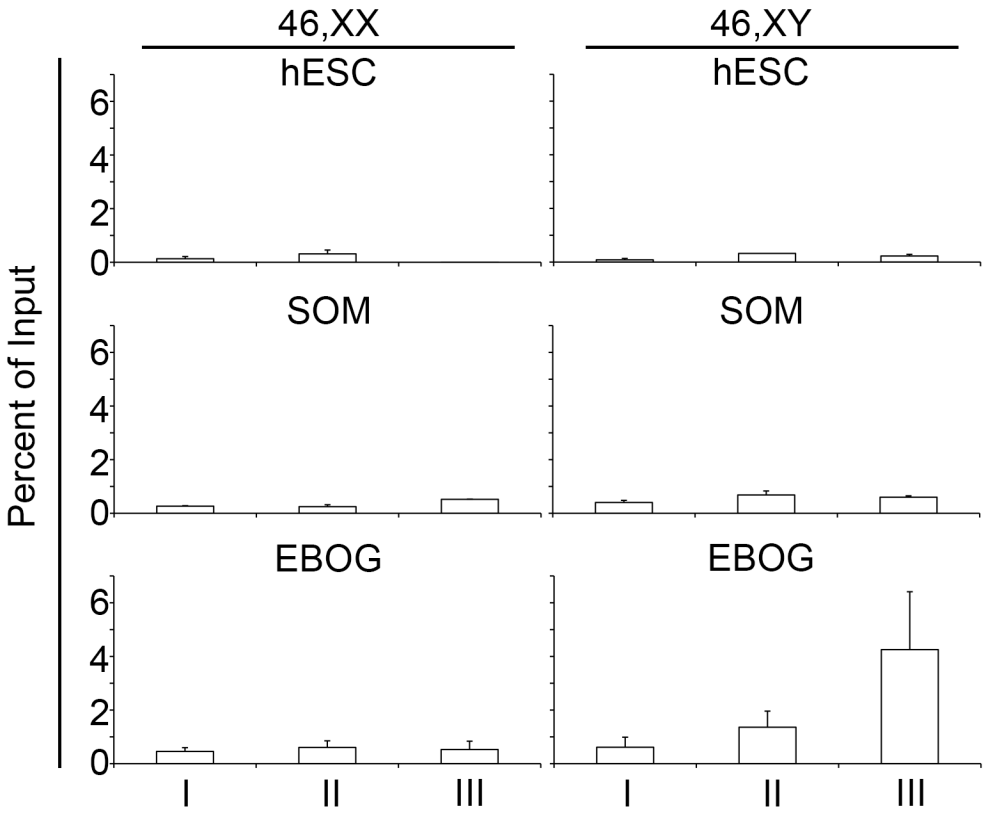

DXZ4
Expression as a % of GAPDH
50
40
30
20
10
0
BM
CE
WB
FB
FL
HE
LI
LU
PS
SG
SM
SP
TE
TH
TR
UT
CO
SI
SC
ST
Tissue


DANT2
Expression as a % of GAPDH
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
BM
CE
WB
FB
FL
HE
LI
LU
PS
SG
SM
SP
TE
TH
TR
UT
CO
SI
SC
ST
Tissue


46,XX hESC (H9)
DAPI
DANT2
DXZ4
Merge
DAPI
DXZ4
DANT1
Merge
DAPI
DANT2
DANT1
Merge
46,XX hESC (H7)
DAPI
DANT2
DXZ4
Merge
DAPI
DXZ4
DANT1
Merge
DAPI
DANT2
DANT1
Merge


a
46,XX hESC (H9)
XIST (Green) &
DANT1 (Red)
46,XY hESC (H1)
DXZ4 (Green) &
DANT1 (Red)
46,XY (hESC (H1)
DANT2 (Green) &
DANT1 (Red)


b
No RNaseA
RNaseA Treated
RNA FISH
DNA FISH


46,XX
46,XY
hESC
hESC
SOM
SOM
EBOG
EBOG
Percent of Input
0
2
4
6
I
II
III

| Sequence Name | Sequence 5' to 3' | Use |
|---|---|---|
| DXZ4-Fwd6 | CAGGCAGAAATGAGCACCAC | DXZ4 qChIP- Set I |
| DXZ4-Rev20 | TGTCCCCGAGGTTGTCTTG | DXZ4 qChIP- Set I |
| DXZ4-Fwd23 | GGACAGTCCCAAGCCACTC | DXZ4 qChIP- Set II |
| DXZ4-Rev14 | TCAGCAAAGGGCGGAGGG | DXZ4 qChIP- Set II |
| DXZ4-Fwd17 | TGACCAAGAGGTCAAAGGCG | DXZ4 qChIP- Set III |
| DXZ4-Rev8 | GTCCCGCTATTCCGGAATTG | DXZ4 qChIP- Set III |
| DXZ4-Fwd11 | CTACGCCAAGCTGCTAACTC | DXZ4 qChIP- Set IV |
| DXZ4-Rev22b | CCAGCGGAAAGTZCATGGG | DXZ4 qChIP- Set IV |
| DXZ4-Fwd4 | TTAAAGGTGCGTTGACGTGG | DXZ4 qChIP- Set V |
| DXZ4-Rev19 | GCCAGGGGGATAGGTGTG | DXZ4 qChIP- Set V |
| qDANT1-F1 | CAGTCTCCTTTGCCCTCTTG | DANT1 qChIP- Set I |
| qDANT1-R1 | AGTTGGGTGGGAAACCTCTT | DANT1 qChIP- Set I |
| qDANT1-F2 | TGGCCTCAATCAGGTAGGTC | DANT1 qChIP- Set II |
| qDANT1-R2 | AGGTCCTCTGCACCAGAAAG | DANT1 qChIP- Set II |
| qDANT1-F3 | AGAGCCTTTCTGGCCTTCC | DANT1 qChIP- Set III |
| qDANT1-R3 | CACACCTCCTCAGCCATGTAG | DANT1 qChIP- Set III |
| DXZ4-BiS-F1 | CCAAACAAACTACCCAAAACC | BiS |
| DXZ4-BiS-R2 | GAAGGTAGGTTAGTAAGAAGG | BiS |
| XIST-BiS-11F | TGGGTTAGAAAAATAAAAATTAAAGTAGG | BiS |
| XIST-BiS-11R | AATACCTACTAATTCCCTTCCCCTCTA | BiS |
| qrtDANT2-F1 | CCTAGCTGAGGCTGCGC | qrtPCR |
| qrtDANT2-R1 | GAGTCGCGTTGCCGCAG | qrtPCR |
| qrtDANT1cd-F3 | GGGCATCCTACCCCATATC | qrtPCR |
| qrtDANT1cd-R3 | AAGGCCCAGTCCGCTCTAC | qrtPCR |
| DANT1-Fwd | CATATCTGAGGTGCGACTGC | rtPCR; PCR-1; PCR-2 |
| DANT1-Rev | TCACCACAGAACTTATGTGAG | rtPCR; PCR-1 |
| Distal ESTs-Fwd1 | ATGCCCATGGCCTTGAGATG | rtPCR; PCR-4 |
| Distal-ESTs-Rev1 | GATCCAACATGGCATGTGAG | rtPCR; PCR-2 |
| DANT2-Fwd1 | AACGCAGATCAGAGGTAGTC | rtPCR; PCR-3; PCR-4; PCR-5 |
| DANT2-Rev1 | AAACCTGCGAGTCGCTGTC | rtPCR; PCR-3; PCR-6 |
| BE297956-Rev | AACCCTAGGATCTGGAGATG | rtPCR; PCR-5 |
| BM925596Ex2Fwd | CTTCCCAAGACCAGACACTG | rtPCR; PCR-6 |
| GAPDH-q-Fwd | CCCAATACGACCAAATCCGT | qrtPCR |
| GAPDH-q-Rev | TCTCTGCTCCTCCTGTTCGA | qrtPCR |
| GAPDH-Fwd | GAAGGTGAAGGTCGGAGTC | rtPCR |
| GAPDH-Rev | GAAGATGGTGATGGGATTTC | rtPCR |
| GDF3-qrt-F[a] | CGGGAATGTACTTCGCTTTC | qrtPCR |
| GDF3-qrt-R[a] | CCCTTTCTTTGATGGCAGAC | qrtPCR |
| Lin28-qrt-F[a] | AAGCGCAGATCAAAAGGAGA | qrtPCR |
| Lin28-qrt-R[a] | CTGATGCTCTGGCAGAAGTG | qrtPCR |
| totNANOG-F[b] | TACCTCAGCCTCCAGCAGAT | rtPCR, qrtPCR |
| totNANOG-R[b] | CCTTCTGCGTCACACCATT | rtPCR, qrtPCR |
| OCT4-F1[c] | CAGTGCCCGAAACCCACAC | rtPCR, qrtPCR |
| OCT4-R1[c] | GGAGACCCAGCAGCCTCAAA | rtPCR, qrtPCR |
| XISTqrt-F[d] | GAAGAGTCTCTGGCTCTTTAGAATACTGA | qrtPCR |
| XISTqrt-R[d] | CAGCGTGGTATCTTCAATGGG | qrtPCR |
| PAX3-F[e] | CTCACCTCAGGTAATGGGACT | rtPCR |
| PAX3-R[e] | CGTGGTGGTAGGTTCCAGAC | rtPCR |
| PAX7-F[e] | CCCCCGCACGGGATT | rtPCR |
| PAX7-R[e] | TATCTTGTGGCGGATGTGGTTA | rtPCR |
| NKX2.5-F[f] | TCTATCCACGTGCCTACAGC | rtPCR |
| NKX2.5-R[f] | AGATCTTGACCTGCGTGGAC | rtPCR |
| cTNT-F[f] | GGCAGCGGAAGAGGATGCTGAA | rtPCR |
| cTNT-R[f] | GAGGCACCAAGTTGGGCATGAACGA | rtPCR |
| βIII-TUBULIN-F[f] | CAACAGCACGGCCATCCAGG | rtPCR |
| βIII-TUBULIN-R[f] | CTTGGGGCCCTGGGCCTCCGA | rtPCR |
| DXZ4-DANT-F2 | CGTCAGGCATACAGATGATAC | promoter construct |
| DXZ4-DANT-R2 | CTCGGCTGCAGCCGCCATTG | promoter construct |
| Distal Fwd-6 | GAGGTTTAACCACCACTCAG | promoter construct |
| Distal Rev-6 | GATTCTACCGAGCAACTGGC | promoter construct |
| FMR1-Fwd[g] | TGGAAGGAGGTTAAATGTTTTG | SNP sequencing |
| FMR1-Rev[g] | GGCAGAACTTCAGTTTTTACGG | SNP sequencing |
| MSL3-Fwd[g] | TGAGCAATGCTAATGGGTTG | SNP sequencing |
| MSL3-Rev[g] | ACCTGATTCATGGCACCTG | SNP sequencing |
| TBL1X-Fwd[g] | TTGGGCTTGTTGTCTCTTCC | SNP sequencing |
| TBL1X-Rev[g] | CTTACTACGCGCCAAAGTGC | SNP sequencing |

[a]Tchieu et al. (2010) Cell Stem Cell 7:329-342
[b]Maherali et al. (2008) Cell Stem Cell 3:340-345
[c]Yu et al. (2007) Science 318:1917-1920
[d]Lengner et al (2010) Cell 141:872-883
[e]Ryan et al (2012) Stem Cell Rev and Rep 8:482-493
[f]Mahmood et al (2010) JMBR 25(6):1216-1233
[g]SNP sites from Teichroeb et al (2011) PLoS ONE 6(10):e23436