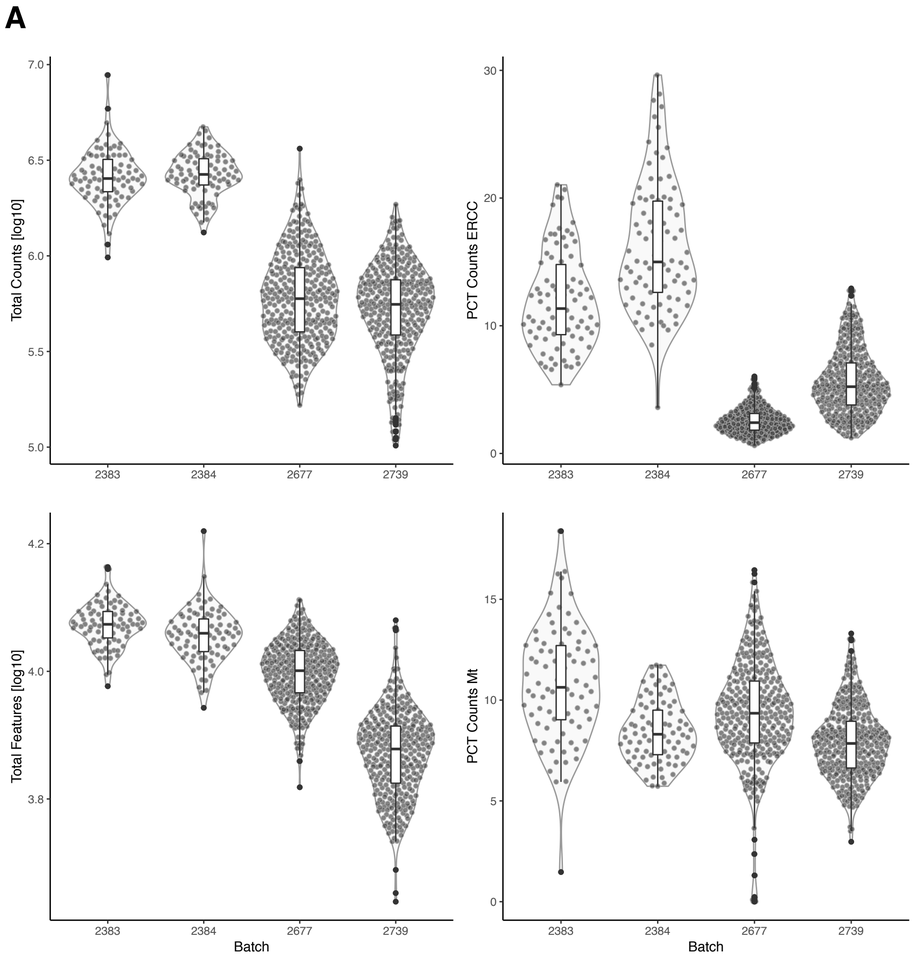
Cell Reports, Volume *26*

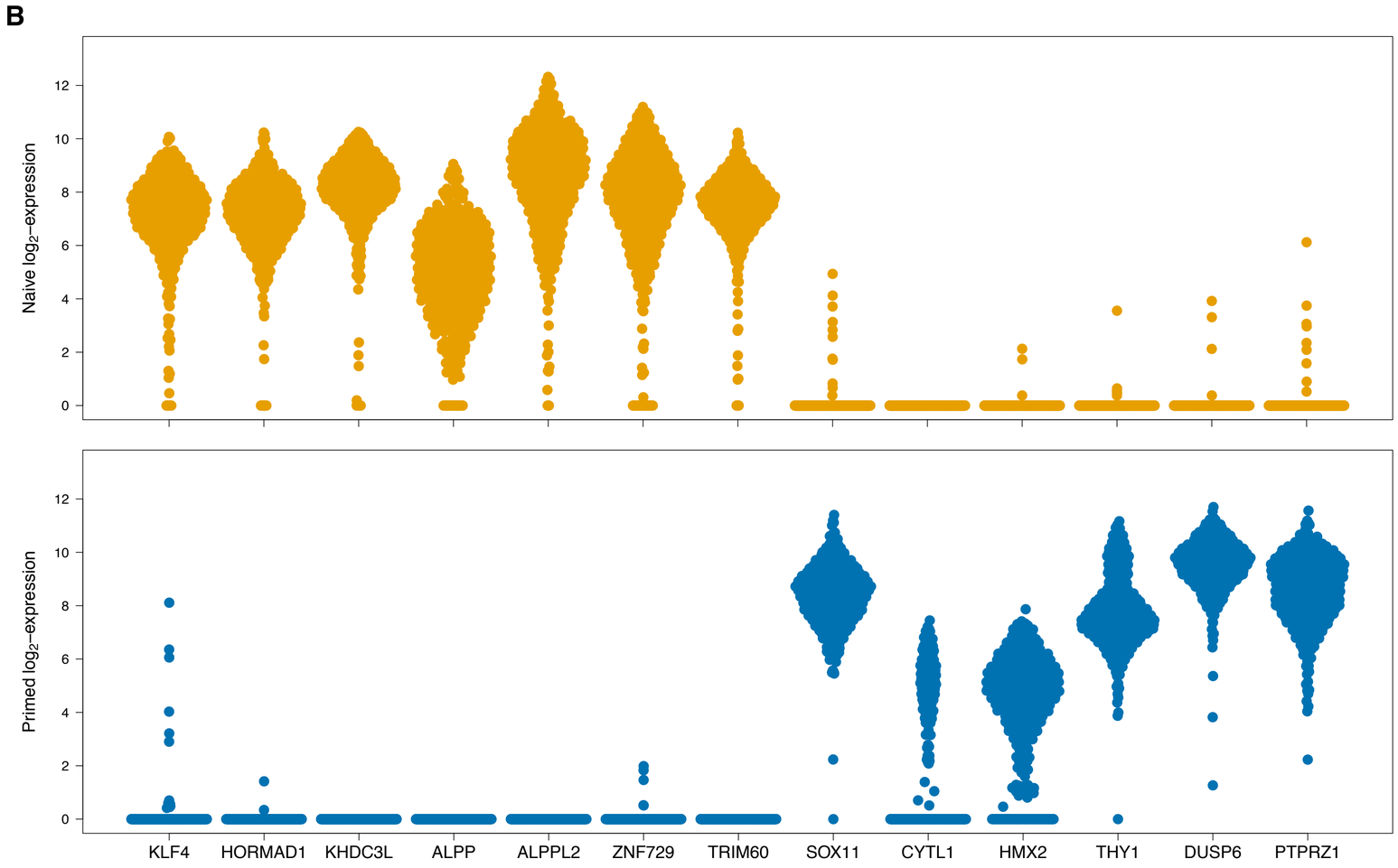
Supplemental Information

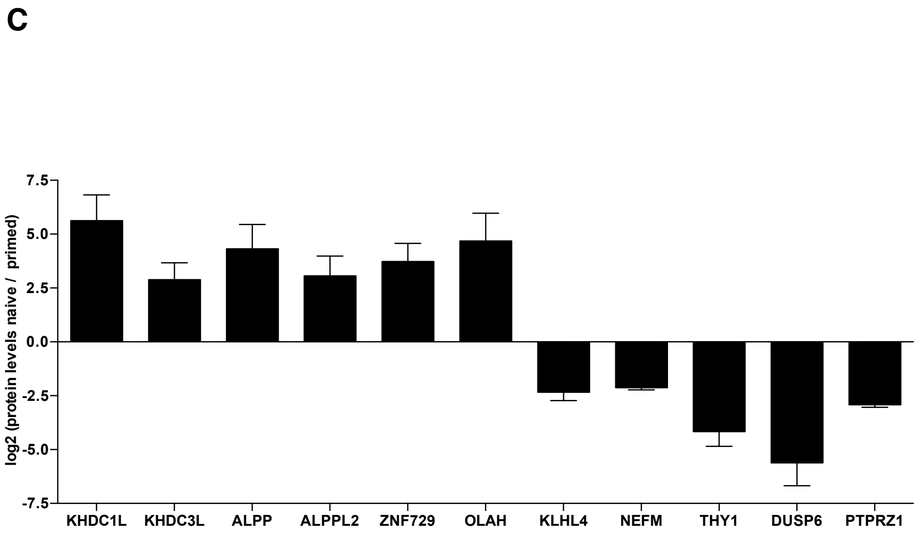
Transcriptional Heterogeneity in Naive and Primed Human Pluripotent Stem Cells at Single-Cell Resolution

Tobias Messmer, Ferdinand von Meyenn, Aurora Savino, Fátima Santos, Hisham Mohammed, Aaron Tin Long Lun, John C. Marioni, and Wolf Reik

Figure S1


A
Total Counts [log10]
PCT Counts ERCC
Total Features [log10]
PCT Counts Mt
2383
2384
2677
2739
Batch


B
Naive log2−expression
Primed log2−expression
KLF4
HORMAD1
KHDC3L
ALPP
ALPPL2
ZNF729
TRIM60
SOX11
CYTL1
HMX2
THY1
DUSP6
PTPRZ1


C
log2 (protein levels naive / primed)
KHDC1L
KHDC3L
ALPP
ALPPL2
ZNF729
OLAH
KLHL4
NEFM
THY1
DUSP6
PTPRZ1

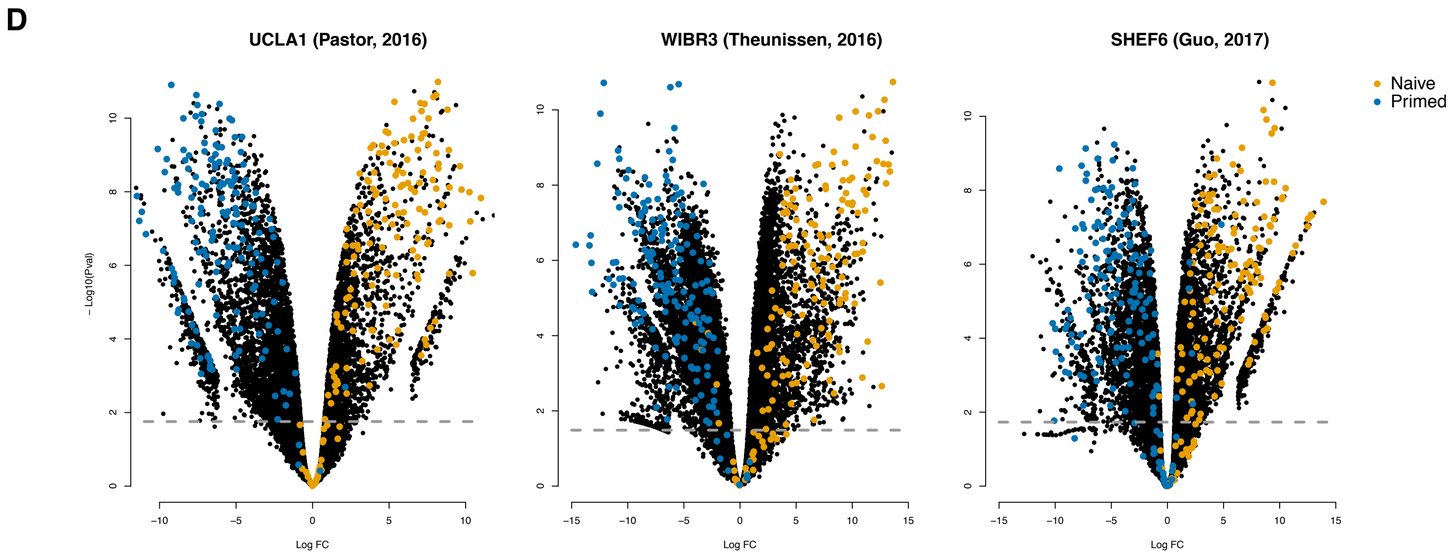

D
UCLA1 (Pastor, 2016)
WIBR3 (Theunissen, 2016)
SHEF6 (Guo, 2017)
− Log10(Pval)
Log FC
Naive
Primed

Figure S1. Statistics for quality control of single cell RNA-seq data, related to Figure 1

(A) Violin plots for each quality control metric including the log10-total count per cell, log10-total features per cell, percentages of spike-ins and percentages of mitochondrial genes. One violin plot is shown per condition in each batch, for naïve (2383) and primed (2384) in batch 1, and naïve (2677) and primed (2678) in batch 2.

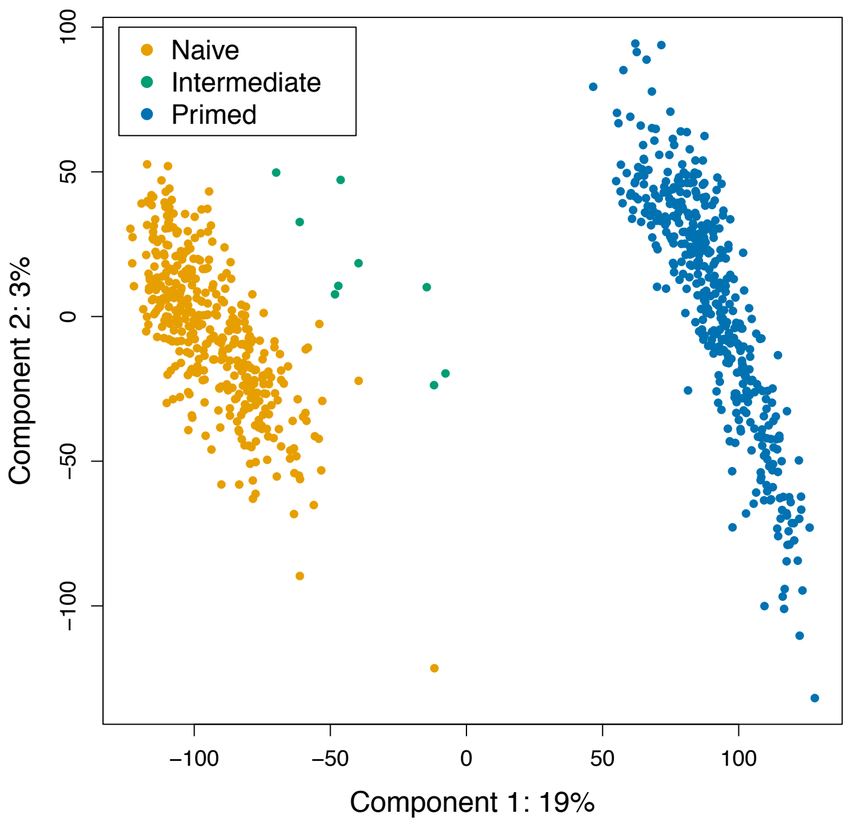
(B) Beeswarm plot of the normalized and batch-corrected log-expression values of marker genes for the naïve and primed population. Each point represents a cell in the naïve (top) or primed (bottom) population.

(C) Log2 protein expression level differences between primed and naïve hESCs measured by Mass Spectrometry (in biological triplicates) for a number of selected markers identified to be highly specific for each condition. Error bars indicate the standard deviation.

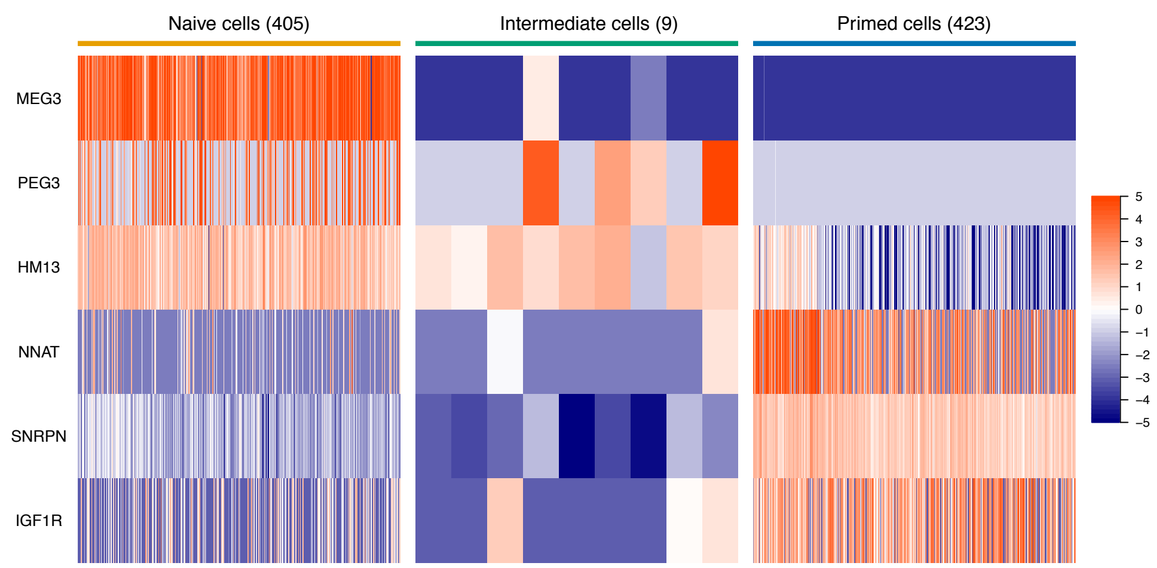
(D) Volcano plots of differentially expressed genes between naïve and primed conditions of different human ESC lines by Pastor et al., 2016 (left), Theunissen et al., 2016 (center) and Guo et al., 2017 (right). The previously identified top 200 markers are coloured in orange (naïve markers) or blue (primed markers). The grey dashed line marks a FDR of 0.05.

Figure S2

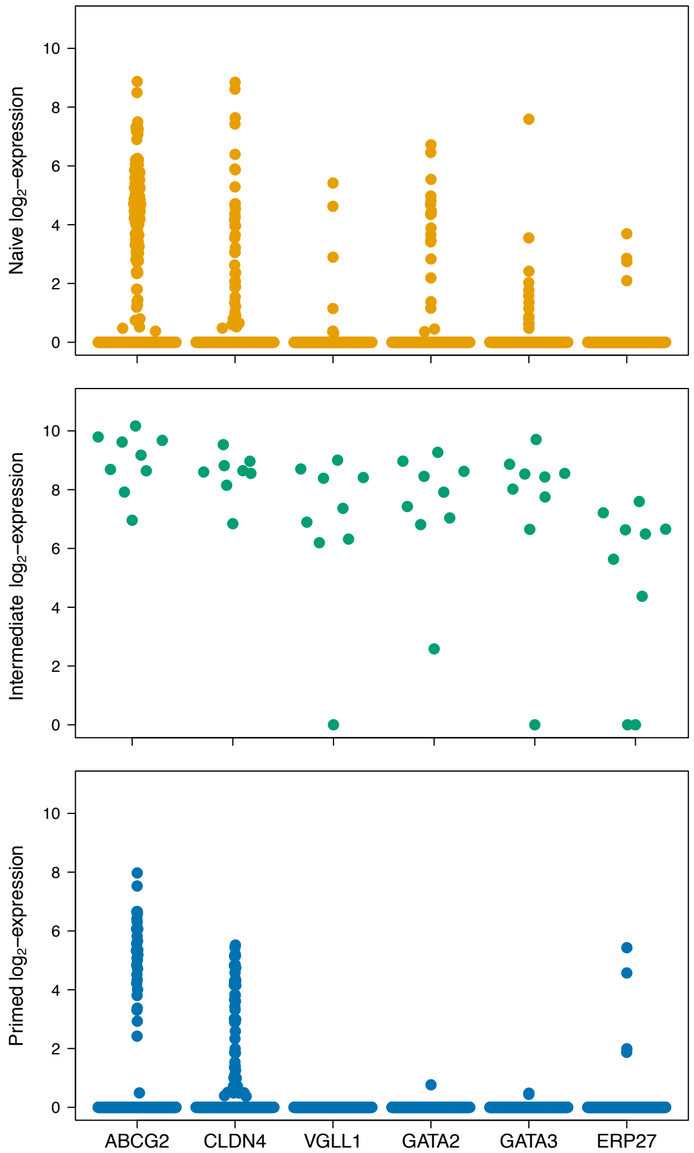
A



B



C



D

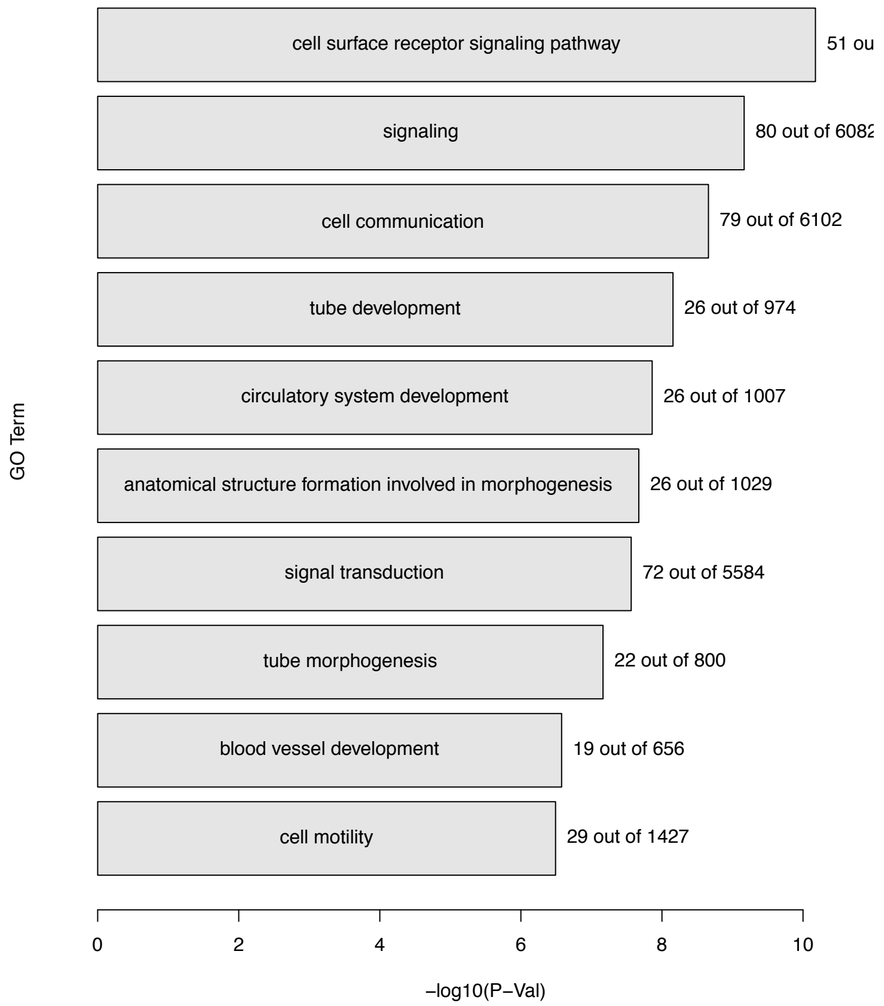


Figure S2. Cells in the naïve subpopulation state lie between the naïve and primed populations, related to Figure 2

(A) Cells in the intermediate population were identified after clustering on HVGs identified in the naïve condition. These cells are highlighted in green, using the same PCA plot in Figure 1B.

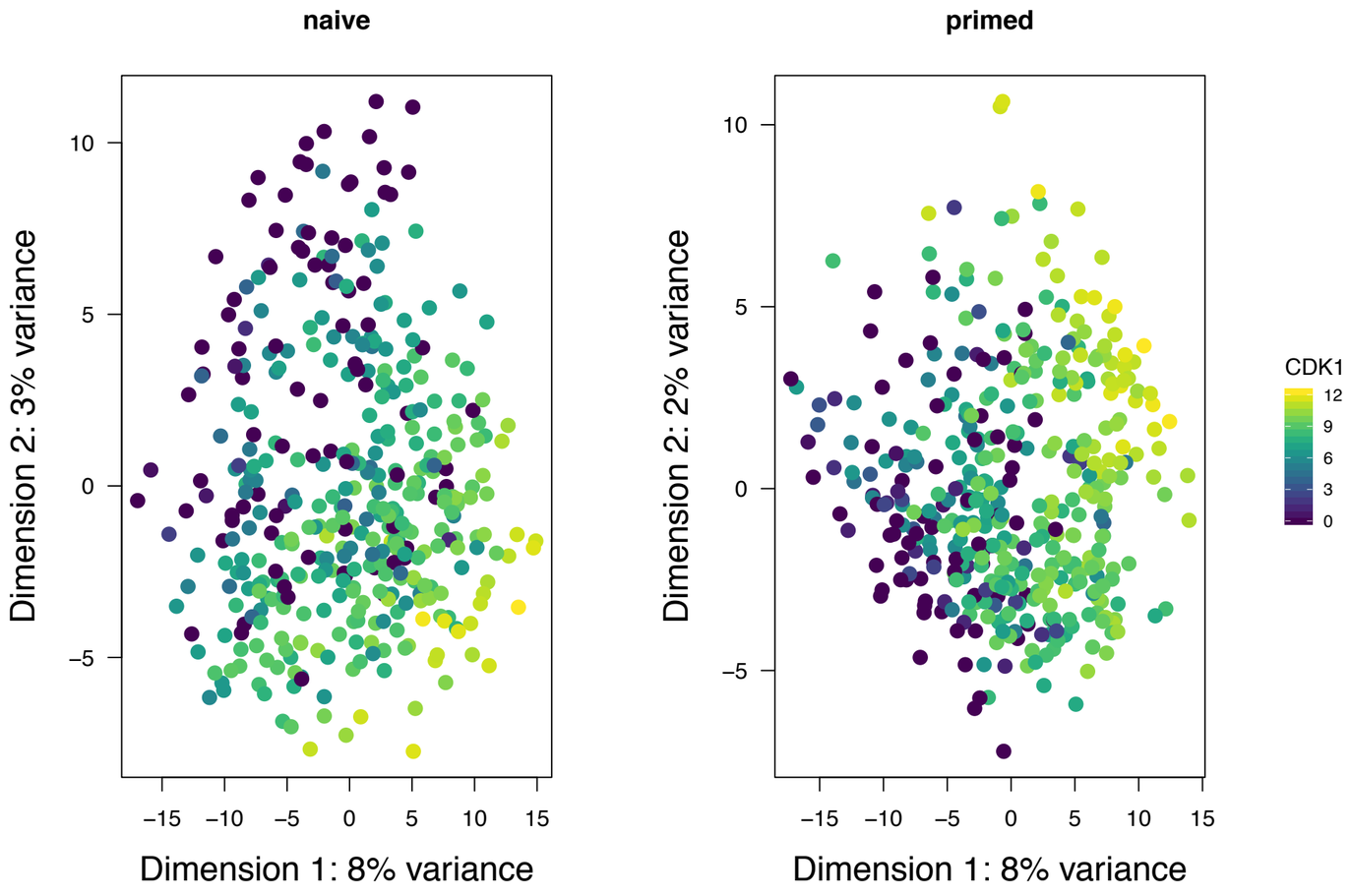
(B) Heat map of a selection of imprinted genes. Each box represents the $\log_2$-fold expression change of genes (rows) for naïve, intermediate and primed cells (columns).

(C) Beeswarm plots of genes that are uniquely expressed in the intermediate population compared to other naïve cells (top) and the primed population (bottom). Each point represents a cell.

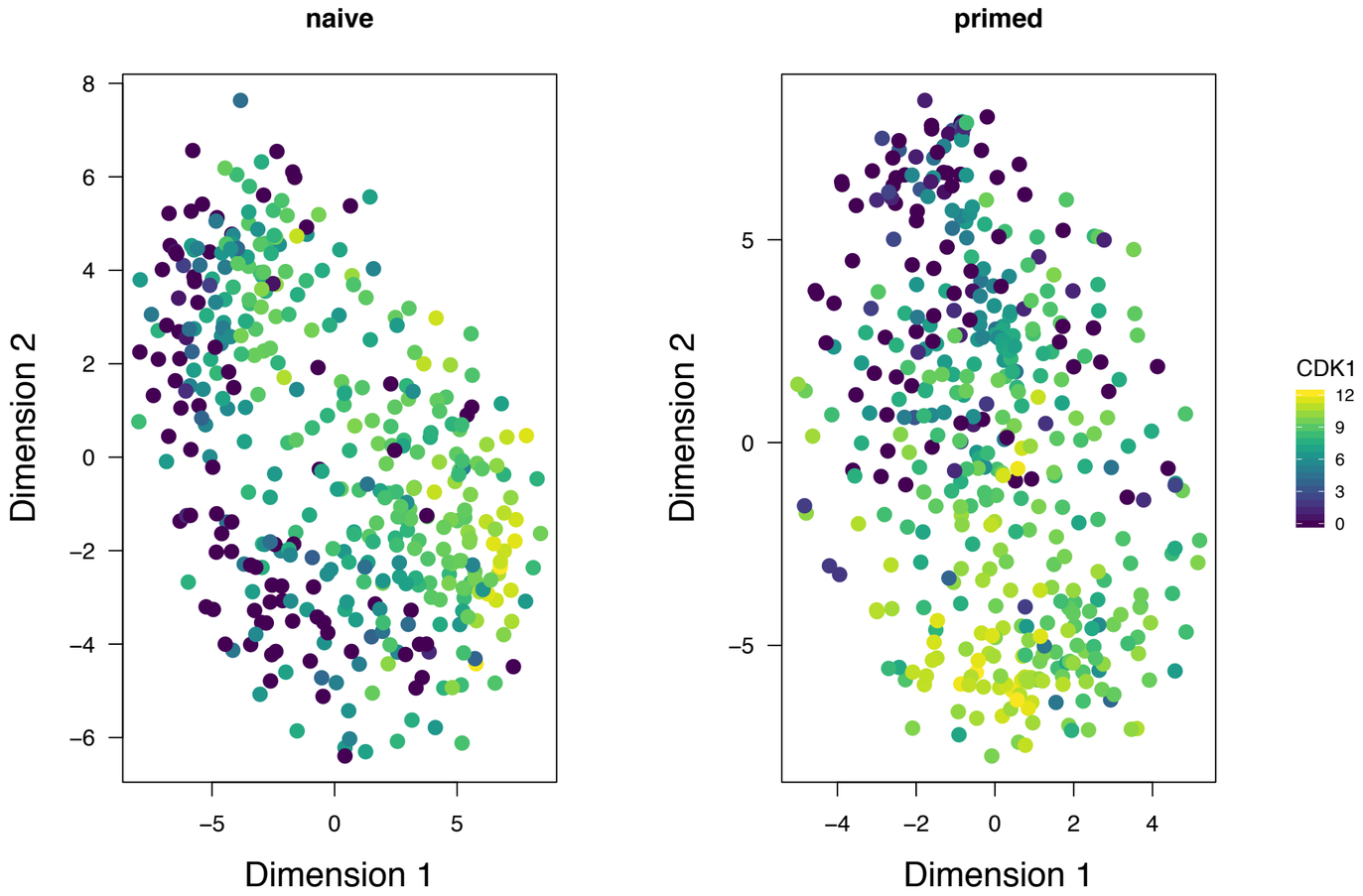
(D) The top 10 most strongly over-represented gene ontology terms in the set of genes that are uniquely expressed in the intermediate population, based on p-value. The values next to the bars indicate the number of DE genes in each gene set.

Figure S3

A



B



C

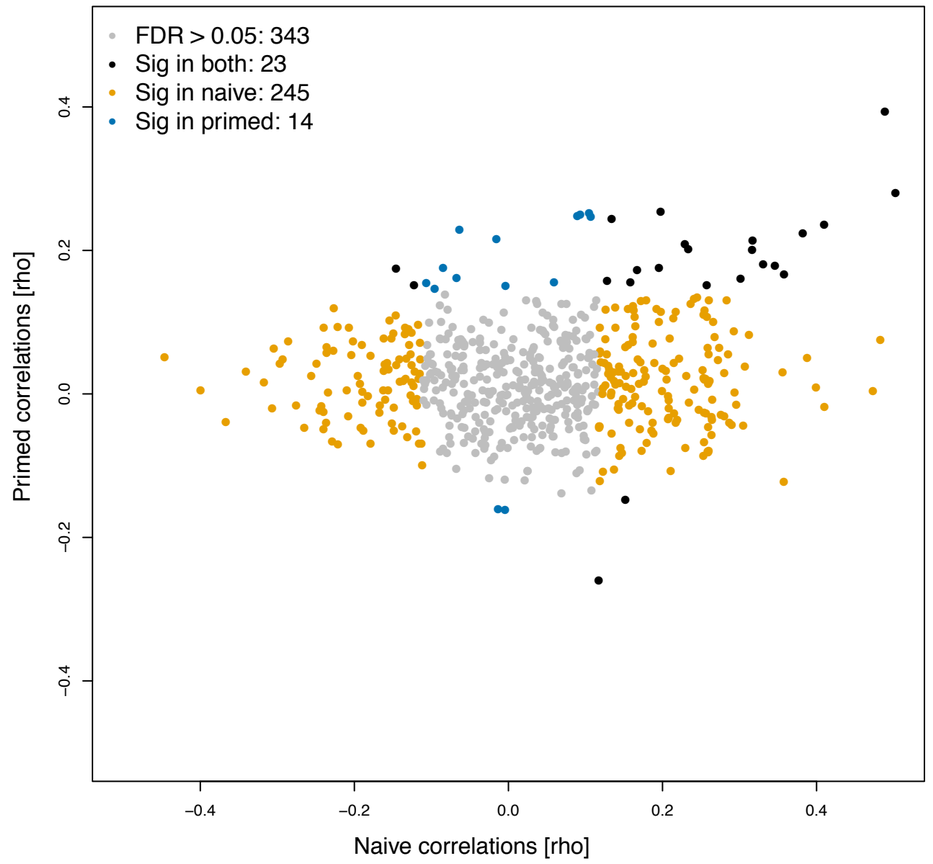


Figure S3. Heterogeneity is driven by cell cycle in both naïve and primed conditions but correlations between lineage markers and epigenetic modulators are more pronounced in naïve cells, related to Figure 3

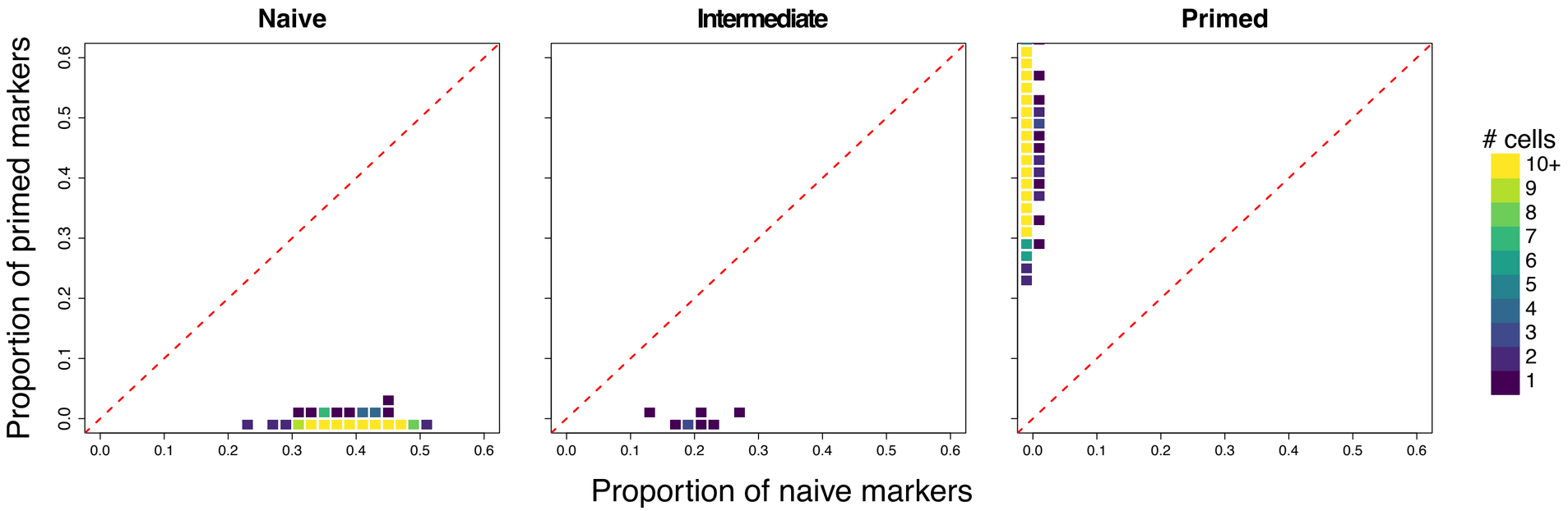
(A) PCA plots of the naïve and primed cells using an unbiased selection of genes. Each point represents a cell and is coloured according to the expression of CDK1.

(B) t-SNE plot based on the batch-corrected normalized log-expression of lineage-specific markers. Each point represents a cell that is coloured for the cell-cycle marker CDK1.

(C) Correlations between lineage markers and epigenetic regulators in the naïve and primed conditions, coloured according to whether they are significantly non-zero in either or both conditions, at a FDR threshold of 5%.

Figure S4

A


Naive
Intermediate
Primed
Proportion of primed markers
Proportion of naive markers
cells
10+
9
8
7
6
5
4
3
2
1

B

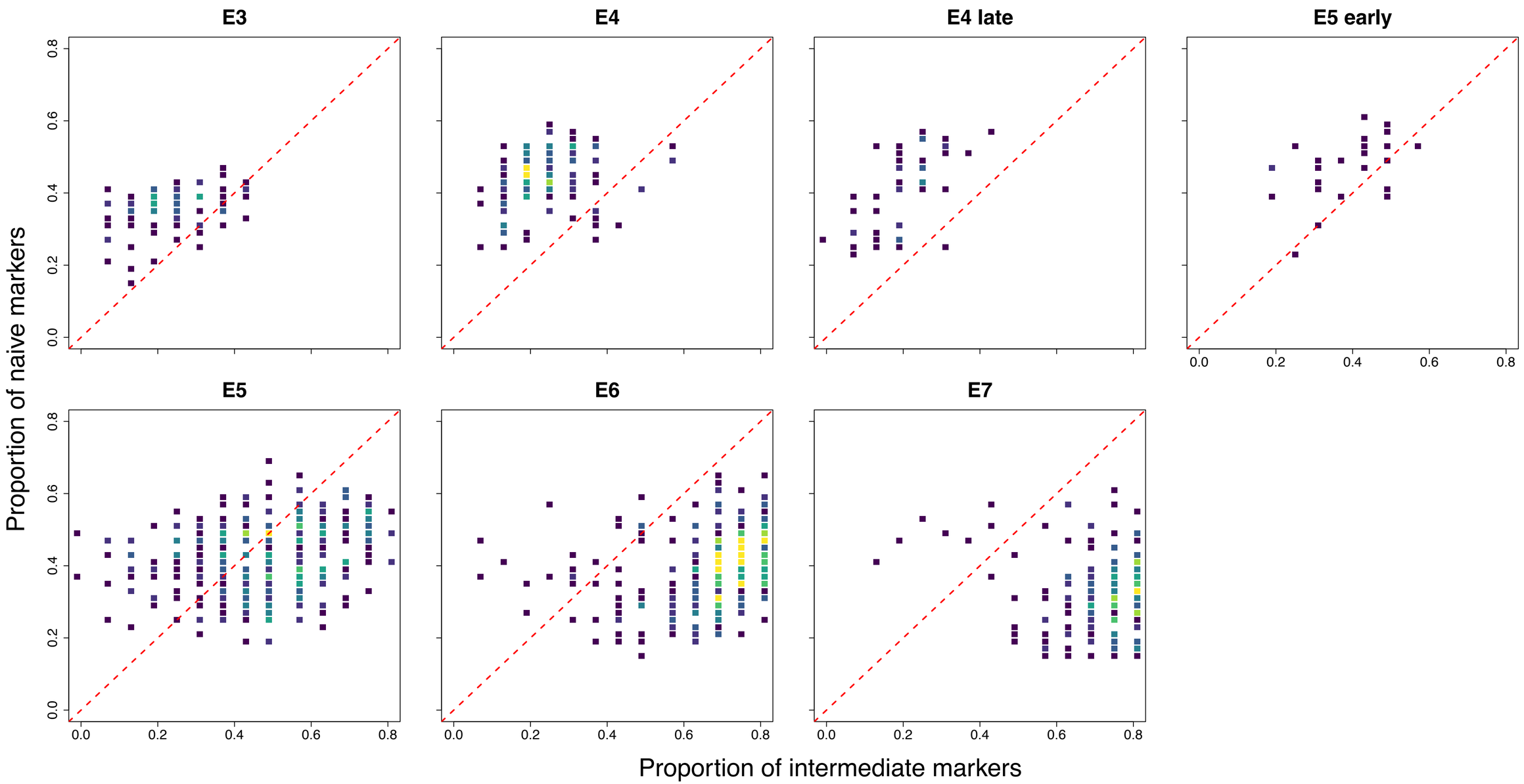

E3
E4
E4 late
E5 early
E5
E6
E7
Proportion of naive markers
Proportion of intermediate markers
cells
8+
7
6
5
4
3
2
1

Figure S4. Mapping of hESCs on the naïve/primed and a naïve/intermediate axis, related to Figure 4
(A) hESCs from the naïve, intermediate and primed populations were mapped onto the naïve/primed axis based on the expression of genes that were strongly DE between conditions to provide a proof-of-concept.
(B) Cells from human pre-implanation embryos are mapped on a naïve/intermediate axis analogously to the naïve/primed axis according to the expressed fraction of naïve or intermediate population markers.

Table S1. Differentially expressed genes between the naïve and the primed population, related to Figure 1

Table S2. Differential expression profile of the intermediate population, related to Figure 2

Table S3. Used germ-layer markers, related to Figure 3

Table S4. Used lineage, pluripotency and epigenetic markers, related to Figure 3