iScience, Volume *25*

Supplemental information

Stabilization of hESCs in two distinct substates along the continuum of pluripotency

Chen Dekel, Robert Morey, Jacob Hanna, Louise C. Laurent, Dalit Ben-Yosef, and Hadar Amir

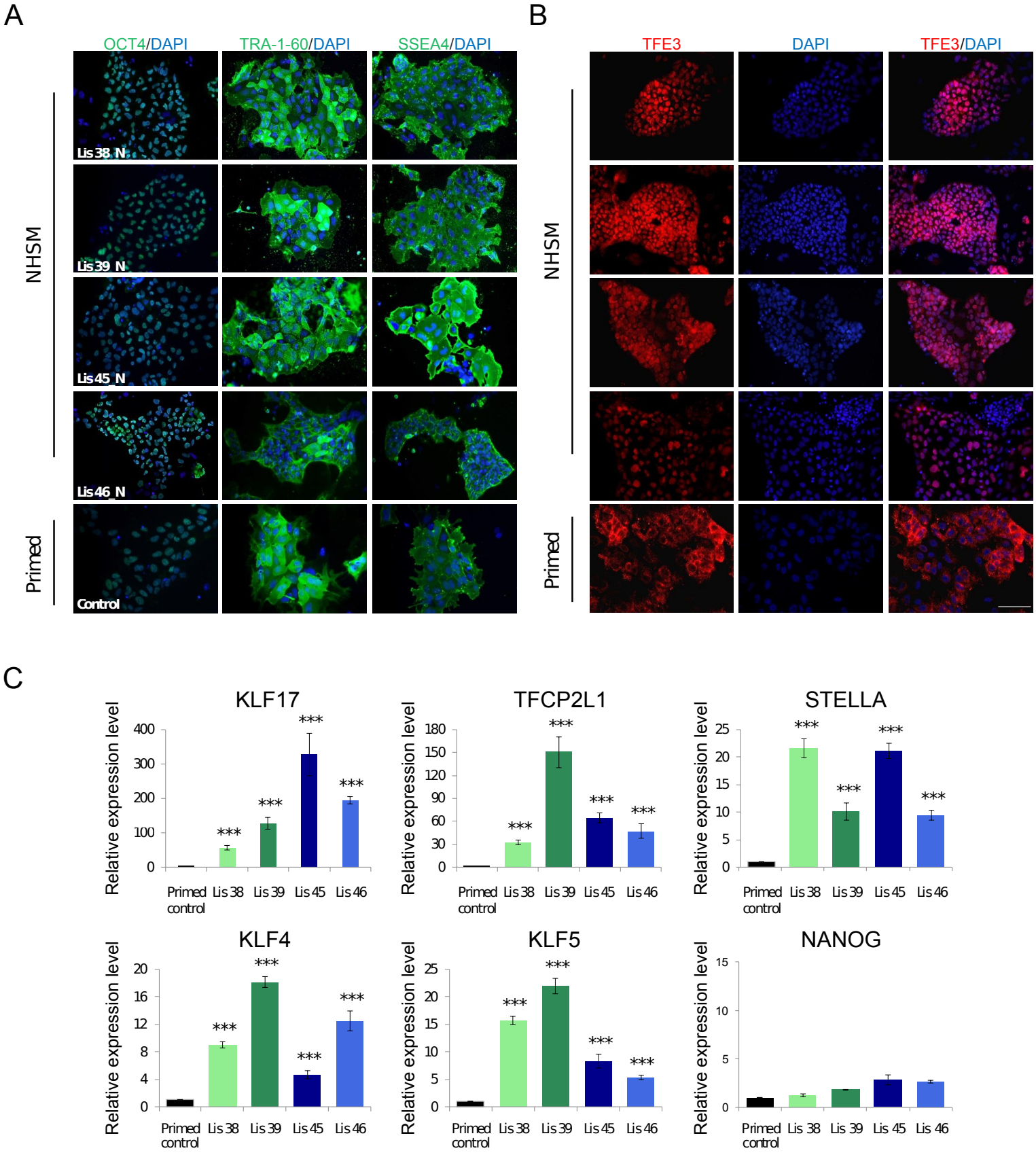


Figure S1: Characterization of NHSM derived hESC lines - related to Figure 1

(A) Immunofluorescence staining for pluripotency markers, OCT4, TRA-1-60 and SSEA4 (A) and the naïve marker TFE3 (B) in the NHSM derived and primed control hESC lines (scale bar-100 mm). (C) Relative mRNA expression levels of the naïve markers KLF17, TFCPL1, STELLA, KLF5, and KLF4 and a pluripotency marker (NANOG) as determined by qRT-PCR in NHSM (red columns) at p20 (derivation) and primed control (gray column) hESC cultures. All data are expressed as mean ± SEM, ***p < .005 by one sample t-test.

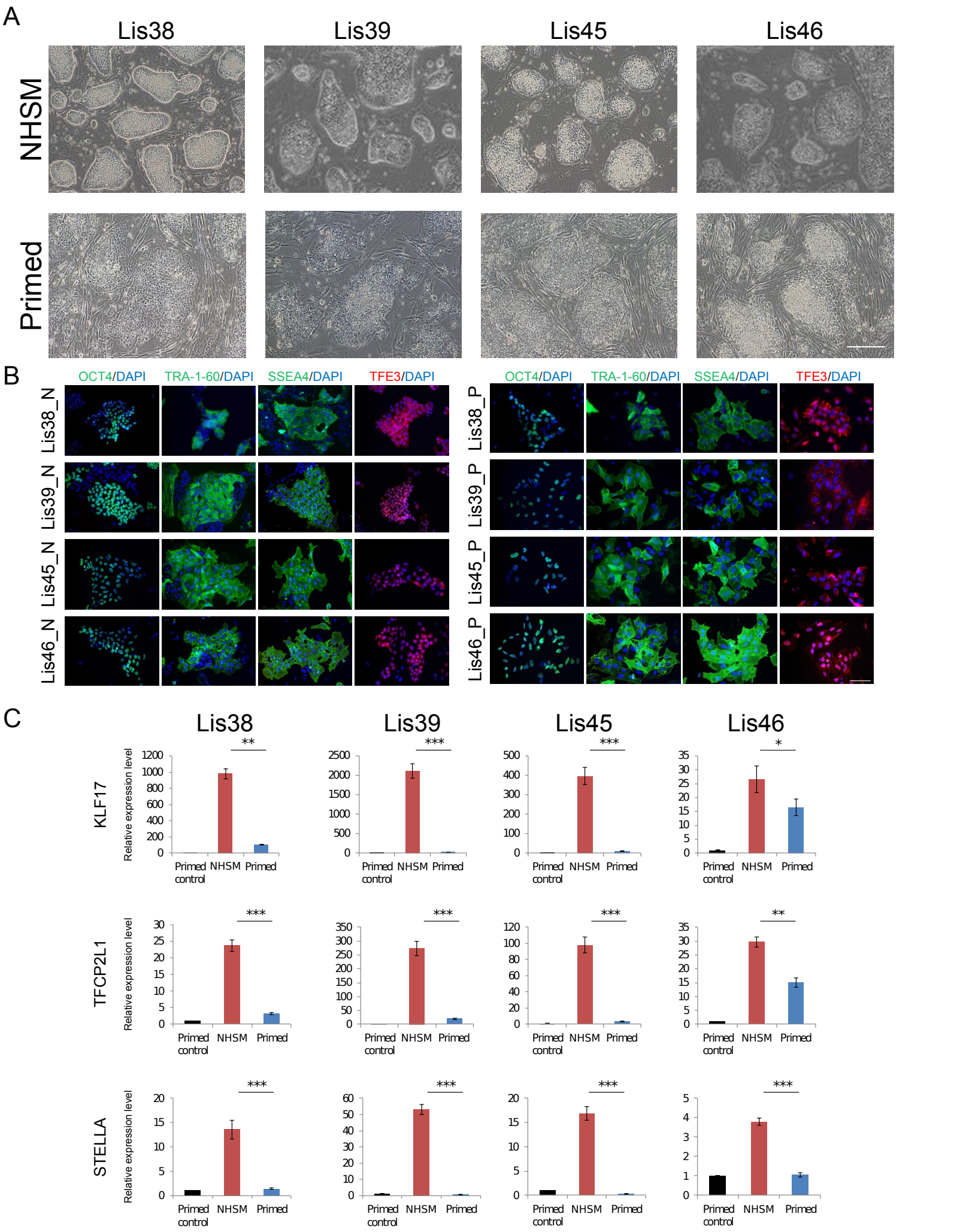

A
Lis38
Lis39
Lis45
Lis46
NHSM
Primed
B
OCT4/DAPI
TRA-1-60/DAPI
SSEA4/DAPI
TFE3/DAPI
Lis38_N
Lis39_N
Lis45_N
Lis46_N
Lis38_P
Lis39_P
Lis45_P
Lis46_P
C
Lis38
Lis39
Lis45
Lis46
KLF17
TFCP2L1
STELLA
Relative expression level
Primed control
NHSM
Primed
**
***
*

Figure S2: Validation of substates at early passage after conversion – related to Figure 1

(A) Colony morphology of early passage (p30, Early NHSM, top row) and their early passage isogenic primed counterparts (p30, Early primed, bottom row; scale bar-200 mm). (B) Immunofluorescence staining for pluripotent markers, OCT4, TRA-1-60 and SSEA4 and for the naïve marker TFE3 in Early NHSM (top 4 rows) and isogenic Early primed (bottom 4 rows) hESC cultures (scale bar-100 mm). (C) Relative mRNA expression levels of naïve markers KLF17, TFCP2L1 and STELLA as determined by qRT-PCR in Early NHSM (red column) and Early primed (blue column) cultures, relative to primed control (black column). All data are expressed as mean ± SEM, *p<.05, **p<.01, ***p < .005 by paired t-test.

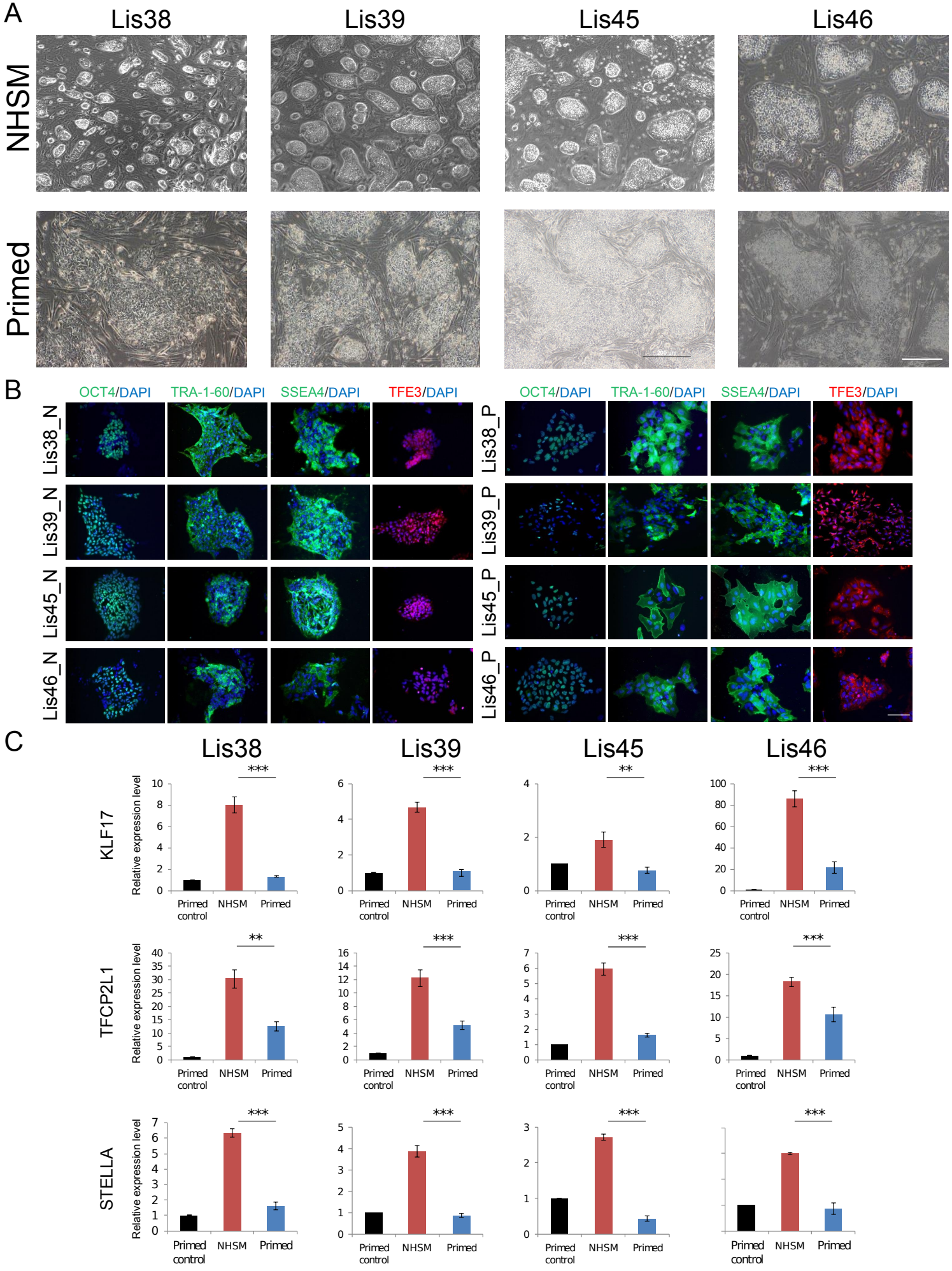

A
Lis38
Lis39
Lis45
Lis46
NHSM
Primed
B
OCT4/DAPI
TRA-1-60/DAPI
SSEA4/DAPI
TFE3/DAPI
Lis38_N
Lis39_N
Lis45_N
Lis46_N
Lis38_P
Lis39_P
Lis45_P
Lis46_P
C
Lis38
Lis39
Lis45
Lis46
KLF17
TFCP2L1
STELLA
Relative expression level
Primed control
NHSM
Primed
***
**

Figure S3: Analyzing the substates of hESC lines after prolonged culture – related to Figure 1

(A) Colony morphology of p50 Late NHSM (top row) and isogenic Late primed (bottom row) hESC cultures (scale bar-200 mm). (B) Immunofluorescence staining for pluripotency markers, OCT4, TRA-1-60 and SSEA4 and for the naïve marker TFE3 in NHSM and isogenic primed hESCs (scale bar-100 mm). (C) Relative mRNA expression levels of the naïve markers KLF17, TFCP2L1 and STELLA as determined by qRT-PCR in Late NHSM (red column) and Late primed (blue column) hESC cultures, relative to primed control cultures (black column). All data are expressed as mean ± SEM, *p<.05, **p<.01, ***p < .005 by paired t-test.

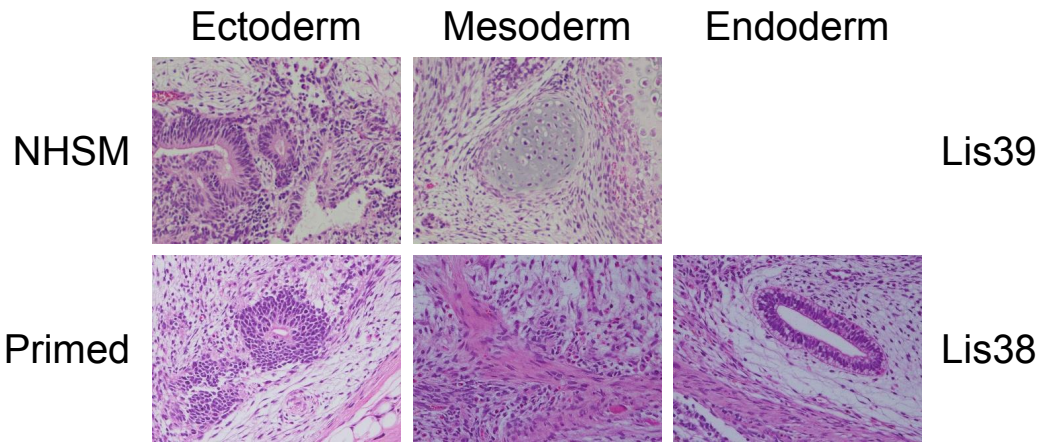


Figure S4: Differentiation potential of NHSM and isogenic primed hESC - related to Figure 1

In vivo differentiation was demonstrated by teratoma formation at p40. Tissues representing the three germ layers were identified using hematoxylin and eosin staining and marked by white arrows (scale bar 50 µm).

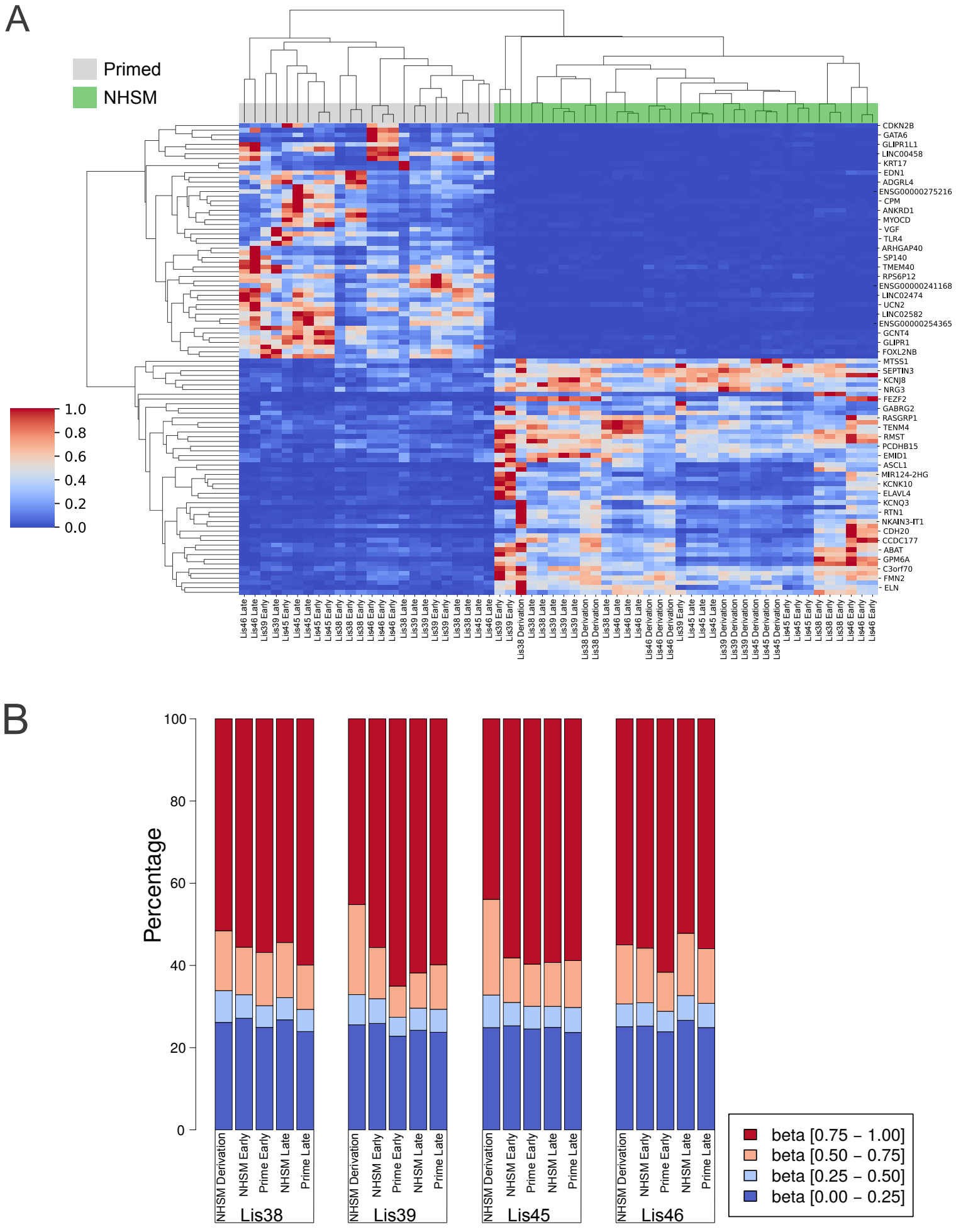

A
Primed
NHSM
1.0
0.8
0.6
0.4
0.2
0.0
CDKN2B
GATA6
GLIPR1L1
LINC00458
KRT17
EDN1
ADGRL4
ENSG00000275216
CPM
ANKRD1
MYOCD
VGF
TLR4
ARHGAP40
SP140
TMEM40
RPS6P12
ENSG00000241168
LINC02474
UCN2
LINC02582
ENSG00000254365
GCNT4
GLIPR1
FOXL2NB
MTSS1
SEPTIN3
KCNJ8
NRG3
FEZF2
GABRG2
RASGRP1
TENM4
RMST
PCDHB15
EMID1
ASCL1
MIR124-2HG
KCNK10
ELAVL4
KCNQ3
RTN1
NKAIN3-IT1
CDH20
CCDC177
ABAT
GPM6A
C3orf70
FMN2
ELN
Lis46 Late
Lis46 Late
Lis39 Early
Lis39 Late
Lis45 Early
Lis45 Late
Lis45 Early
Lis45 Early
Lis38 Early
Lis38 Early
Lis46 Early
Lis46 Early
Lis38 Late
Lis39 Late
Lis39 Early
Lis38 Late
Lis38 Late
Lis46 Late
Lis39 Early
Lis38 Derivation
Lis38 Late
Lis38 Late
Lis39 Late
Lis39 Late
Lis38 Derivation
Lis38 Derivation
Lis38 Late
Lis46 Late
Lis46 Late
Lis46 Derivation
Lis46 Derivation
Lis46 Derivation
Lis39 Early
Lis45 Late
Lis45 Late
Lis45 Derivation
Lis39 Derivation
Lis39 Derivation
Lis39 Derivation
Lis45 Derivation
Lis45 Derivation
Lis45 Derivation
Lis45 Early
Lis45 Early
Lis45 Early
Lis38 Early
Lis38 Early
Lis46 Early
Lis46 Early
Lis46 Early
B
Percentage
100
80
60
40
20
0
NHSM Derivation
NHSM Early
Prime Early
NHSM Late
Prime Late
Lis38
Lis39
Lis45
Lis46
beta [0.75 – 1.00]
beta [0.50 – 0.75]
beta [0.25 – 0.50]
beta [0.00 – 0.25]

Figure S5: Transcriptomic and epigenomic characterization of our NHSM and primed cultures – related to Figure 3

(A) Heatmap displaying the top 100 differentially expressed genes by fold change (50 upregulated in NHSM vs 50 upregulated in primed). (B) Global methylation levels from EPIC DNA methylation array. One sample from each line at each condition and timepoint. Methylation levels are displayed as a percentage of four beta value bins (Beta values range from 0 to 1 and are used to measure the percentage of methylation. They are the ratio of the methylated probe intensity and the overall intensity (sum of methylated and unmethylated probe intensities).

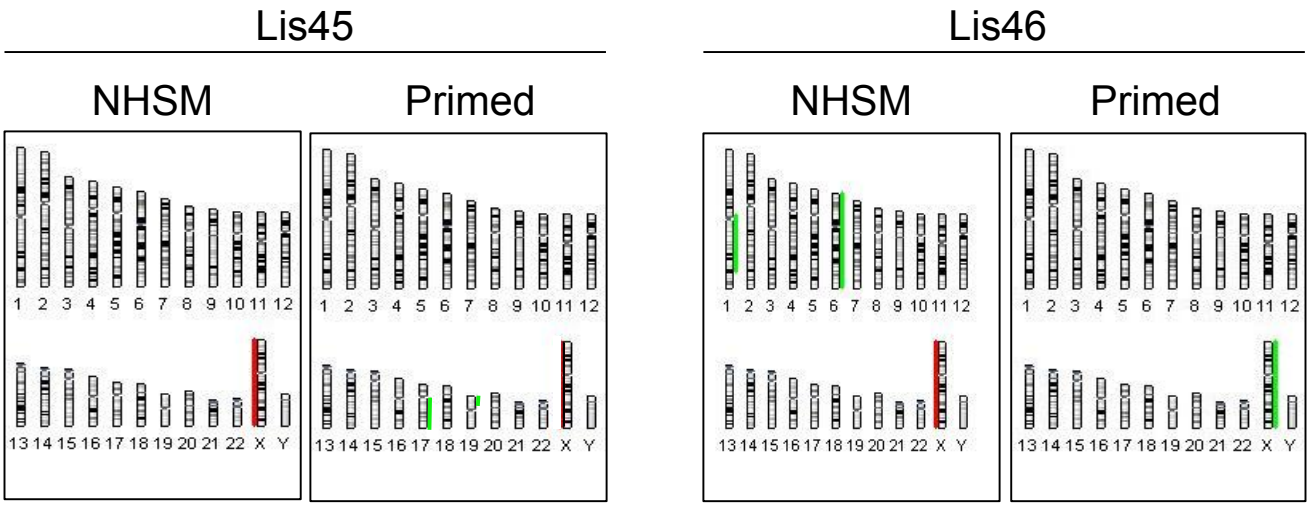
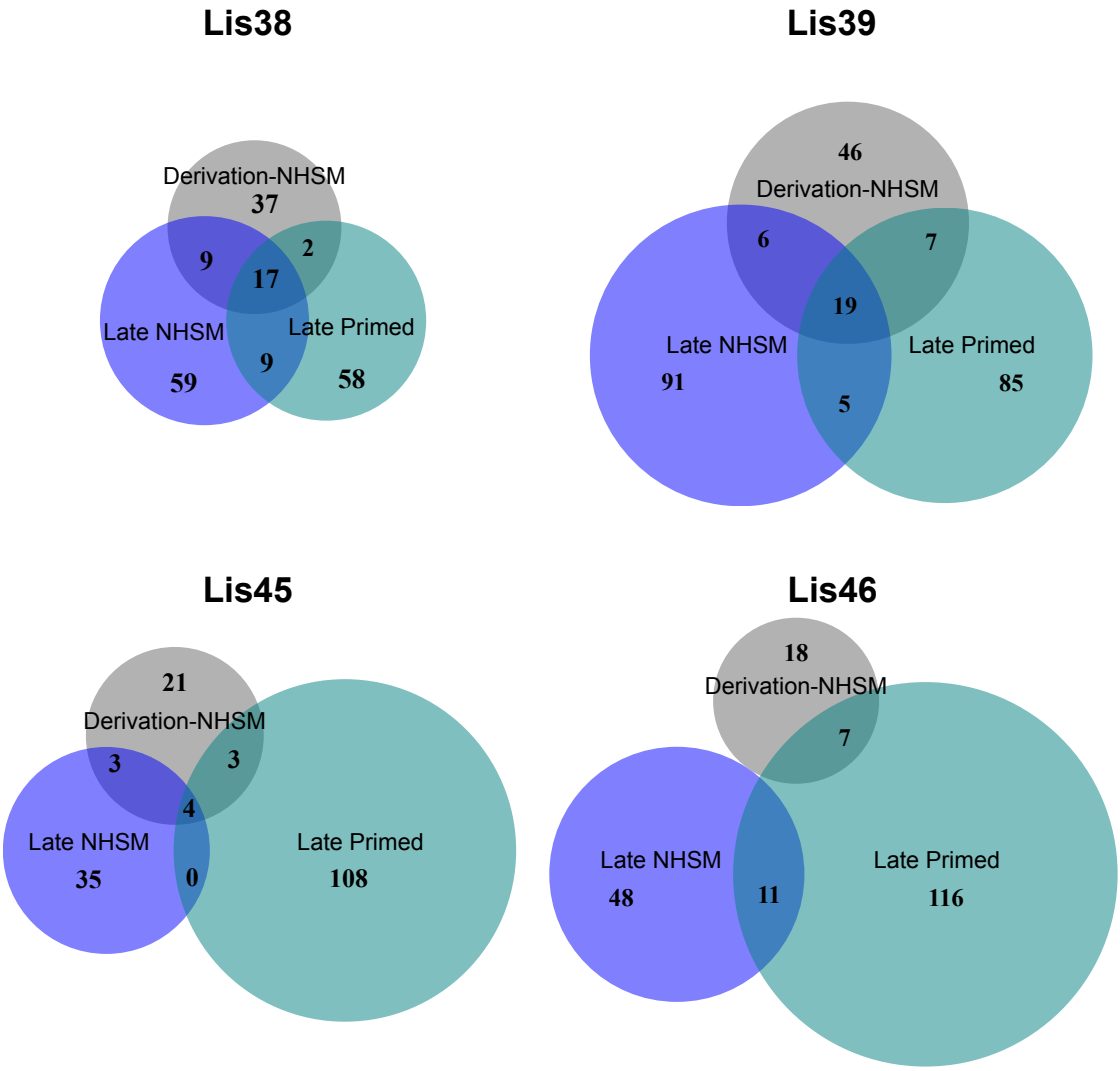


Figure S6: Chromosomal locations of copy number aberrations detected using CMA - related to Figure 2

Chromosomal charts of the CMA aberrations detected in late NHSM and isogenic primed cells. Duplications are shown in green and deletions in red.

A


Lis38
Derivation-NHSM
37
9
17
2
Late NHSM
59
9
Late Primed
58
Lis39
46
Derivation-NHSM
6
7
19
Late NHSM
91
Late Primed
85
5
Lis45
21
Derivation-NHSM
3
3
4
Late NHSM
35
0
Late Primed
108
Lis46
18
Derivation-NHSM
7
Late NHSM
48
11
Late Primed
116

B

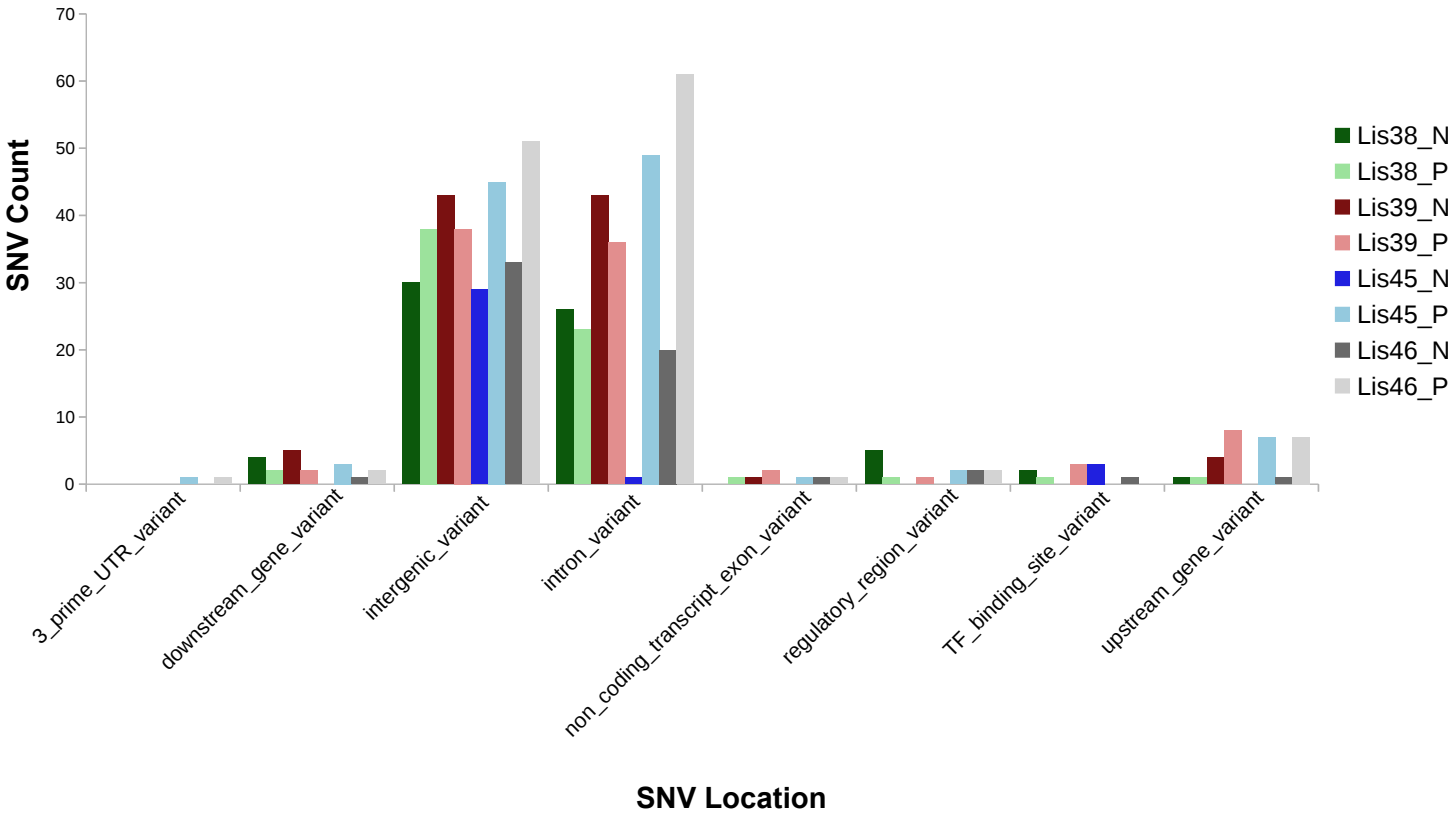

SNV Count
70
60
50
40
30
20
10
0
3_prime_UTR_variant
downstream_gene_variant
intergenic_variant
intron_variant
non_coding_transcript_exon_variant
regulatory_region_variant
TF_binding_site_variant
upstream_gene_variant
SNV Location
Lis38_N
Lis38_P
Lis39_N
Lis39_P
Lis45_N
Lis45_P
Lis46_N
Lis46_P

Figure S7: SNVs in NHSM derived compared to isogenic primed hESCs - related to Figure 4

(A) Venn-diagrams presenting the number of SNVs detected in NHSM derived hESCs, as well as in late NHSM and isogenic primed cells, including unique and overlapping SNVs. (B) Bar graph showing the distribution of the detected SNVs throughout the genome for each line in non-coding regions, including: SNVs within introns, up-stream or down-stream to genes, non-coding transcript exons, TF binding sites, regulatory regions, and intergenic regions.

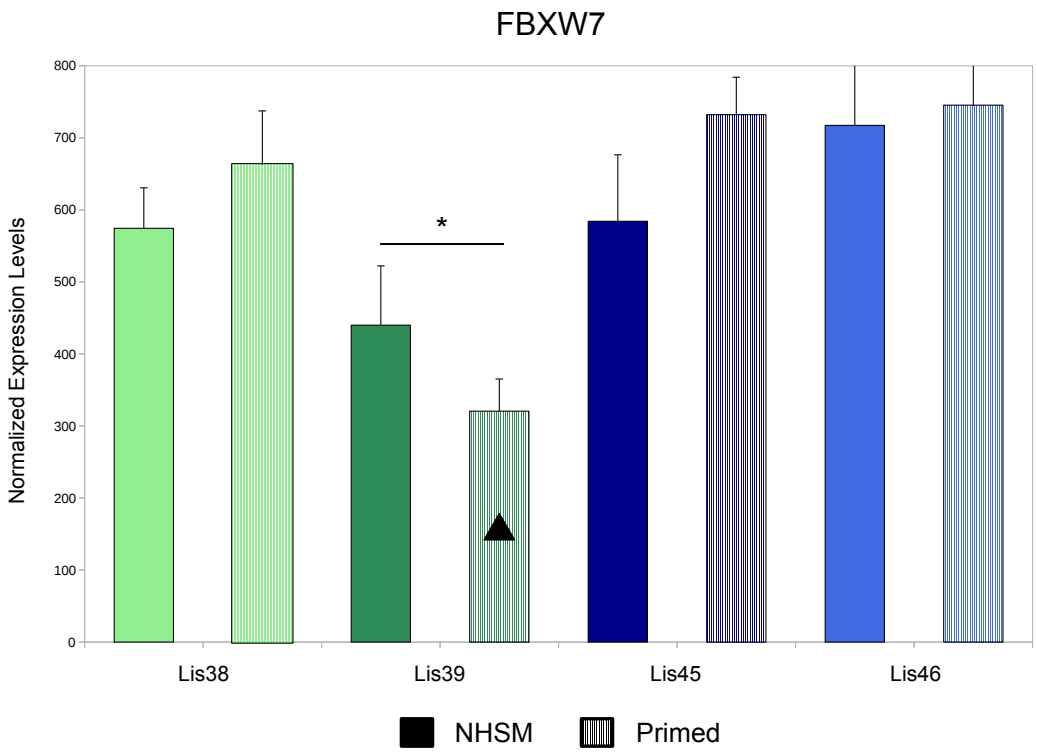


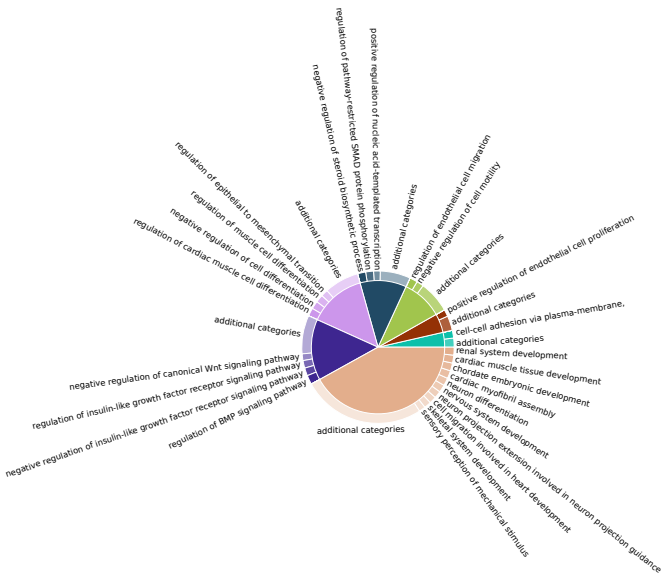
Figure S8: Aberrations in cancer-associated genes (Tier1 genes) - related to Figure 4

A deletion upstream to FBXW7 gene detected in Lis39_P (marked by a black triangle) is associated with decreased expression levels compared to its NHSM counterpart, in contrast to the expression pattern observed in the other lines. Data are expressed as mean ± SD, *p<.05 by paired t-test.

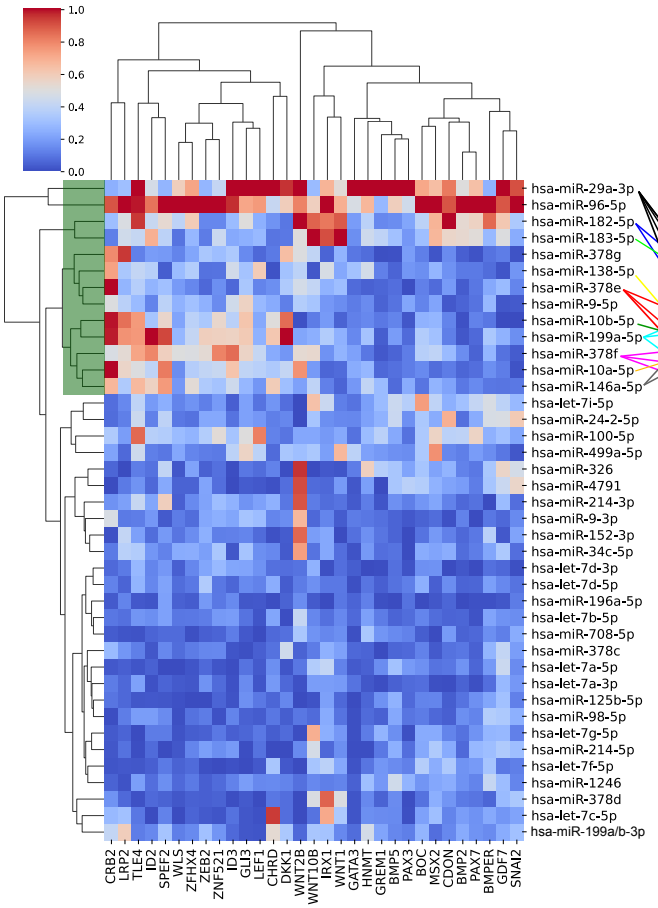
A

Name and Proportion of the Biological Process

- Various Development Terms, 41.9%
- Regulation of BMP Signaling Pathway, 14.8%
- Regulation Cell Differentiation, 13.9%
- Regulation of Steroid Biosynthetic Process, 11.2%
- Regulation of Endothelial Cell Migration, 9.9%
- Positive Regulation of Endothelial Cell Proliferation, 4.6%
- Cell-Cell Adhesion via Plasma-Membrane, 3.6%

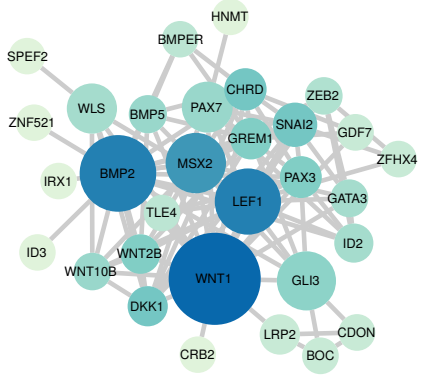


B

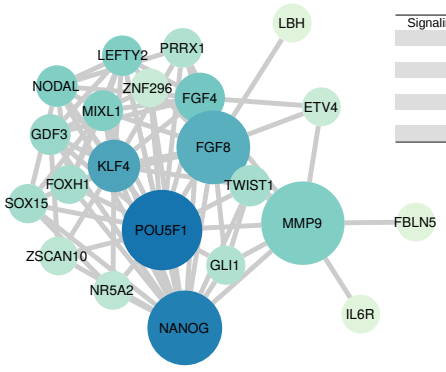


miRNA's analyzed for pathway enrichment

KEGG 2021 Term	Adjusted P-value
Hippo signaling pathway	1.53E-12
TGF-beta signaling pathway	1.98E-09
Basal cell carcinoma	7.24E-09
Wnt signaling pathway	1.90E-06
Hedgehog signaling pathway	1.34E-05
Signaling pathways regulating pluripotency of stem cells	1.74E-05

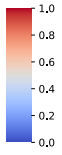


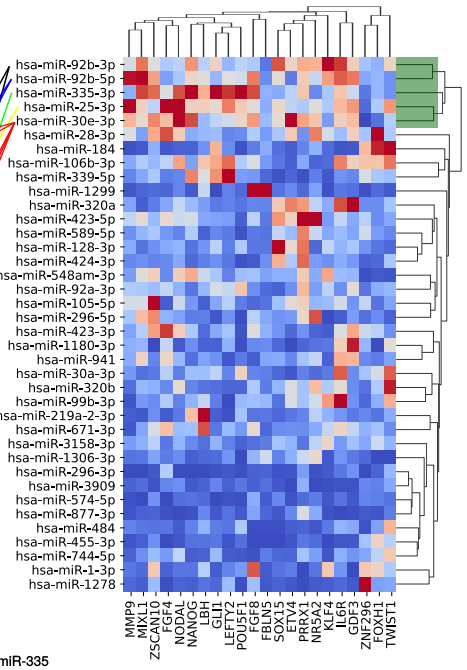
C



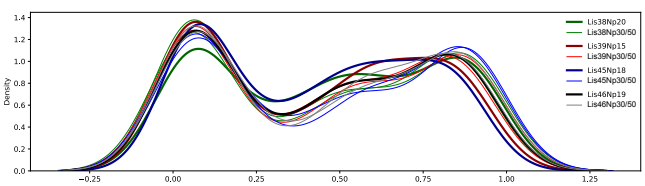
miRNA's analyzed for pathway enrichment

KEGG 2021 Term	Adjusted P-value
Signaling pathways regulating pluripotency of stem cells	1.66E-05
Pathways in cancer	4.69E-03
Proteoglycans in cancer	1.89E-02
Chemical carcinogenesis	2.19E-02
Melanoma	2.25E-02
Cytokine-cytokine receptor interaction	2.65E-02
TGF-beta signaling pathway	2.71E-02
PI3K-Akt signaling pathway	3.30E-02





D



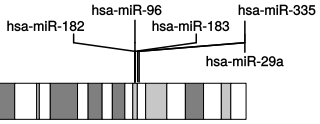
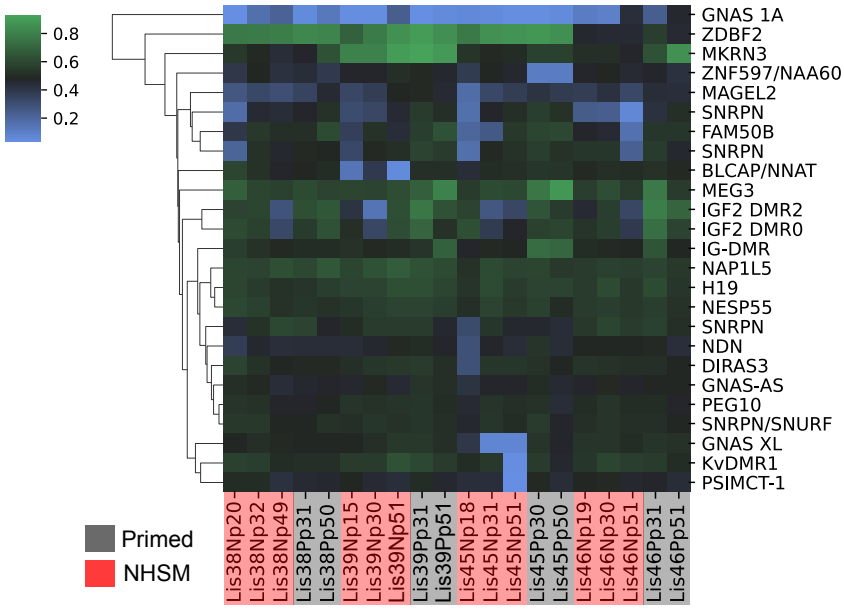


Figure S9: Effects of long-term culture on the transcriptome of NHSM hESCs – related to Figure 5

(A) Gene ontology (GO) enrichment analysis of uniquely downregulated genes in the NHSM derivation (p20) timepoint compared to the early (p30) and late (p50) timepoints among cultures. Genes that were downregulated in the early primed vs late primed timepoint were removed leaving 157 genes (**Table S7**). Revigo and CirGO were used to provide a non-redundant and biologically succinct visualization of GO terms with an adjusted p-value < 0.05. (B) Combined mRNA and miRNA expression analysis to identify regulatory changes occurring over time in NHSM conditions. The network (right) was constructed by creating STRING protein to protein interaction networks and then clustering the networks. The network shows the largest cluster of mRNAs downregulated in NHSM derived (p20) cultures compared to early (p30) and late (p50) NHSM cultures with the size relative to the Betweenness Centrality score and the color representing the node degree. The heatmap (left) was created using the importance scores generated by integrating the miRNA and mRNA using a gene regulatory inference algorithm. The colors represent the importance scores which reflect the degree to which each potential mRNA target is regulated by each miRNA in the dataset (see methods). The heatmap shows the mRNAs downregulated in the NHSM derived cultures that were also found in the largest clustered network (right). The table shows the significant (adj. p-value < 0.00005) KEGG 2021 pathways enriched in the mRNAs in the largest clustered network. The lines drawn from the miRNAs in the heatmap to the pathways in the table, indicate the pathways predicted to be regulated by the corresponding miRNAs by miRPathDB. Only the miRNAs highlighted in green were analyzed. (C) Combined mRNA and miRNA expression analysis to identify regulatory changes occurring over time in NHSM conditions. The network (left) was constructed by creating STRING protein to protein interaction networks and then clustering the networks. The network shows the largest cluster of mRNAs upregulated in NHSM derived (p20) cultures compared to early (p30) and late (p50) NHSM cultures with the size relative to the Betweenness Centrality score and the color representing the node degree. The heatmap (left) was created using the importance scores generated by integrating the miRNA and mRNA using a gene regulatory inference algorithm. The colors represent the importance scores which reflect the degree to which each potential mRNA target is regulated by each miRNA in the dataset (see methods). The heatmap shows the mRNAs upregulated in the NHSM derived cultures that were also found in the largest clustered network (right). The table shows the significant (adj. p-value < 0.05) KEGG 2021 pathways enriched in the mRNAs in the largest clustered network. The lines drawn from the miRNAs in the heatmap to the pathways in the table, indicate the pathways predicted to be regulated by the corresponding miRNAs by miRPathDB. Only the miRNAs highlighted in green were analyzed. (D) 5 of 18 miRNAs with the highest importance scores in **Fig. S9B** and **Fig. S9C** were found in the same miRNA cluster on chromosome 7. Density plot shows the methylation levels of derivation samples (thick lines) and later NHSM cultures (p30/50, thin lines), of CpG islands of miRNA cluster region on chromosome seven.

A


0.8
0.6
0.4
0.2
GNAS 1A
ZDBF2
MKRN3
ZNF597/NAA60
MAGEL2
SNRPN
FAM50B
SNRPN
BLCAP/NNAT
MEG3
IGF2 DMR2
IGF2 DMR0
IG-DMR
NAP1L5
H19
NESP55
SNRPN
NDN
DIRAS3
GNAS-AS
PEG10
SNRPN/SNURF
GNAS XL
KvDMR1
PSIMCT-1
Primed
NHSM
Lis38Np20
Lis38Np32
Lis38Np49
Lis38Pp31
Lis38Pp50
Lis39Np15
Lis39Np30
Lis39Np51
Lis39Pp31
Lis39Pp51
Lis45Np18
Lis45Np31
Lis45Np51
Lis45Pp30
Lis45Pp50
Lis46Np19
Lis46Np30
Lis46Np51
Lis46Pp31
Lis46Pp51

B

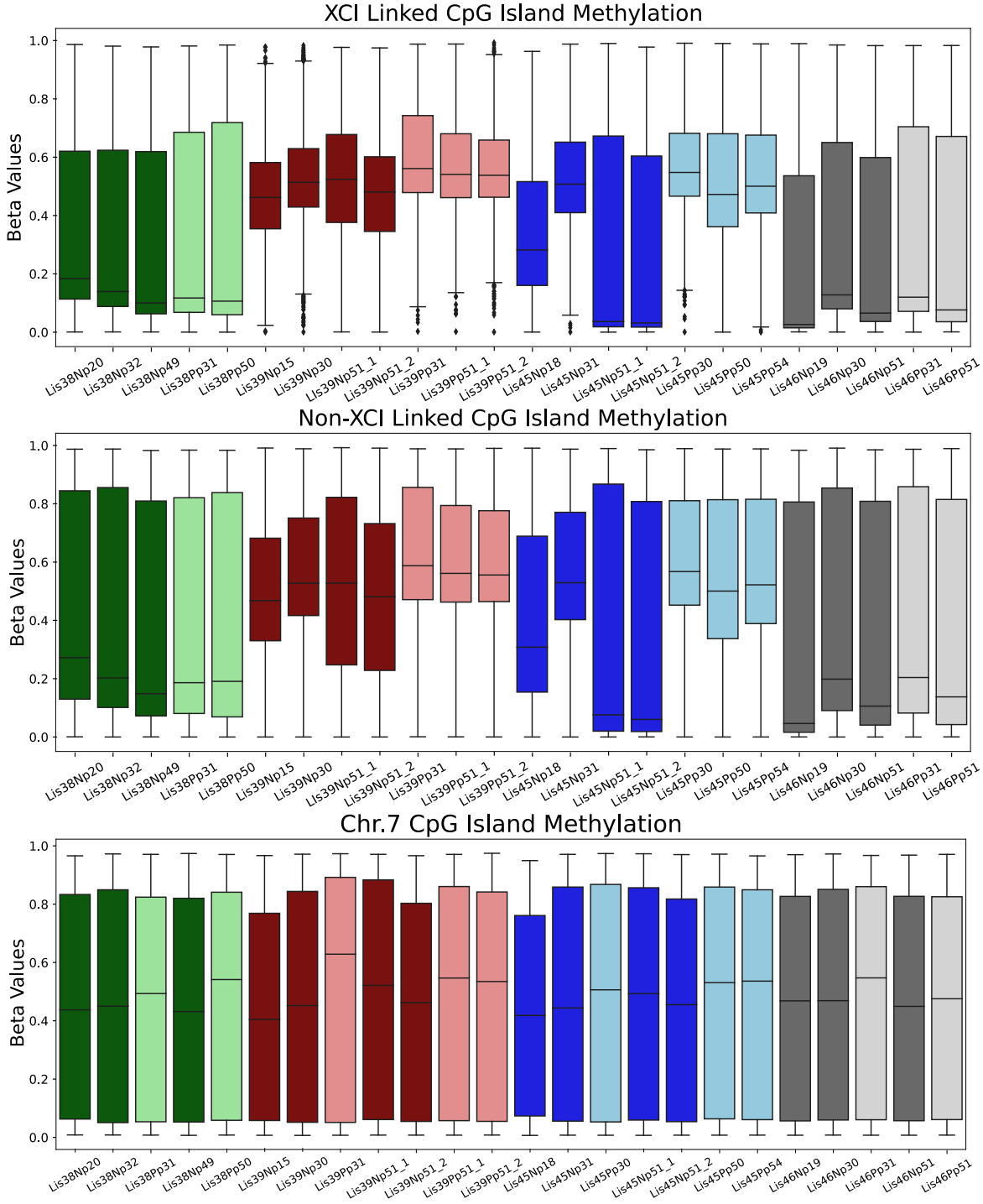

XCI Linked CpG Island Methylation
Non-XCI Linked CpG Island Methylation
Chr.7 CpG Island Methylation
Beta Values
1.0
0.8
0.6
0.4
0.2
0.0
Lis38Np20
Lis38Np32
Lis38Np49
Lis38Pp31
Lis38Pp50
Lis39Np15
Lis39Np30
Lis39Np51_1
Lis39Np51_2
Lis39Pp31
Lis39Pp51_1
Lis39Pp51_2
Lis45Np18
Lis45Np31
Lis45Np51_1
Lis45Np51_2
Lis45Pp30
Lis45Pp50
Lis45Pp54
Lis46Np19
Lis46Np30
Lis46Np51
Lis46Pp31
Lis46Pp51

Figure S10: Methylation at imprinted and X-Chromosome inactivation sites – related to Figure 6 & 7

(A) Heatmap displaying DNA methylation at regions associated with single-isoform imprinted genes. (B) Methylation levels of the CpG islands associated with X-Chromosome Inactivation (XCI) genes (top boxplot), non-XCI linked genes on the X-chromosome (middle), and with all genes found on chromosome 7 (bottom). Boxes span the top 75th percentile to bottom 25th percentile with the line in the middle of the boxes representing the median. Whiskers contain 95% of the data.

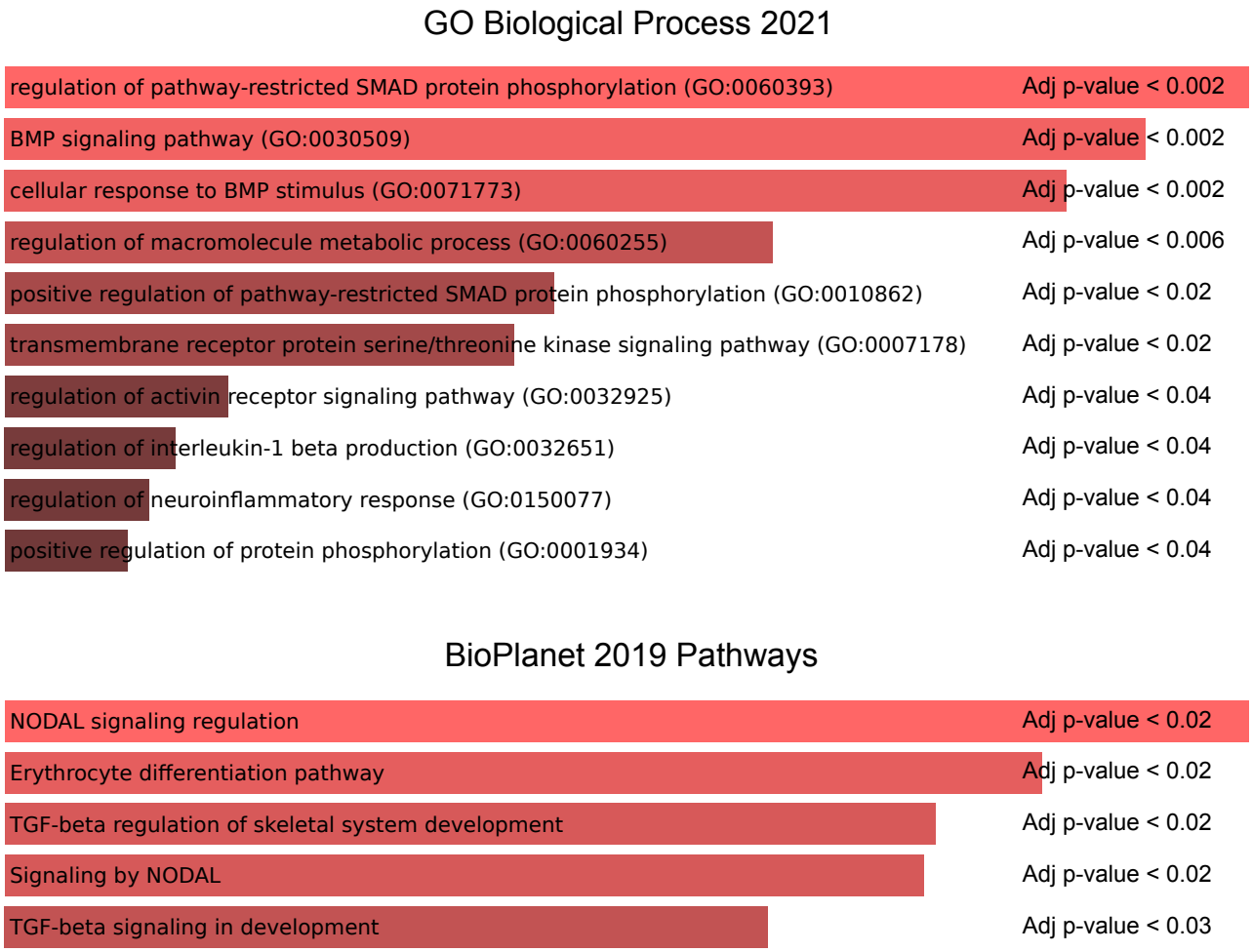


Figure S11: Gene ontology related to Figure 7

Gene ontology enrichment analysis of genes both hypomethylated and showing higher expression in the NHSM derived (p20) cultures compared to the early (p30) and late (p50) cultures.