Stem Cell Reports, Volume *8*

Supplemental Information

Vitamin C and L-Proline Antagonistic Effects Capture Alternative States in the Pluripotency Continuum

Cristina D'Aniello, Ehsan Habibi, Federica Cermola, Debora Paris, Francesco Russo, Alessandro Fiorenzano, Gabriele Di Napoli, Dominique J. Melck, Gilda Cobellis, Claudia Angelini, Annalisa Fico, Robert Blelloch, Andrea Motta, Hendrik G. Stunnenberg, Dario De Cesare, Eduardo J. Patriarca, and Gabriella Minchiotti

SUPPLEMENTAL INFORMATION

Document S1. Figures S1-S3, Tables S1-S6 and Supplemental Experimental Procedures

Supplemental Figures

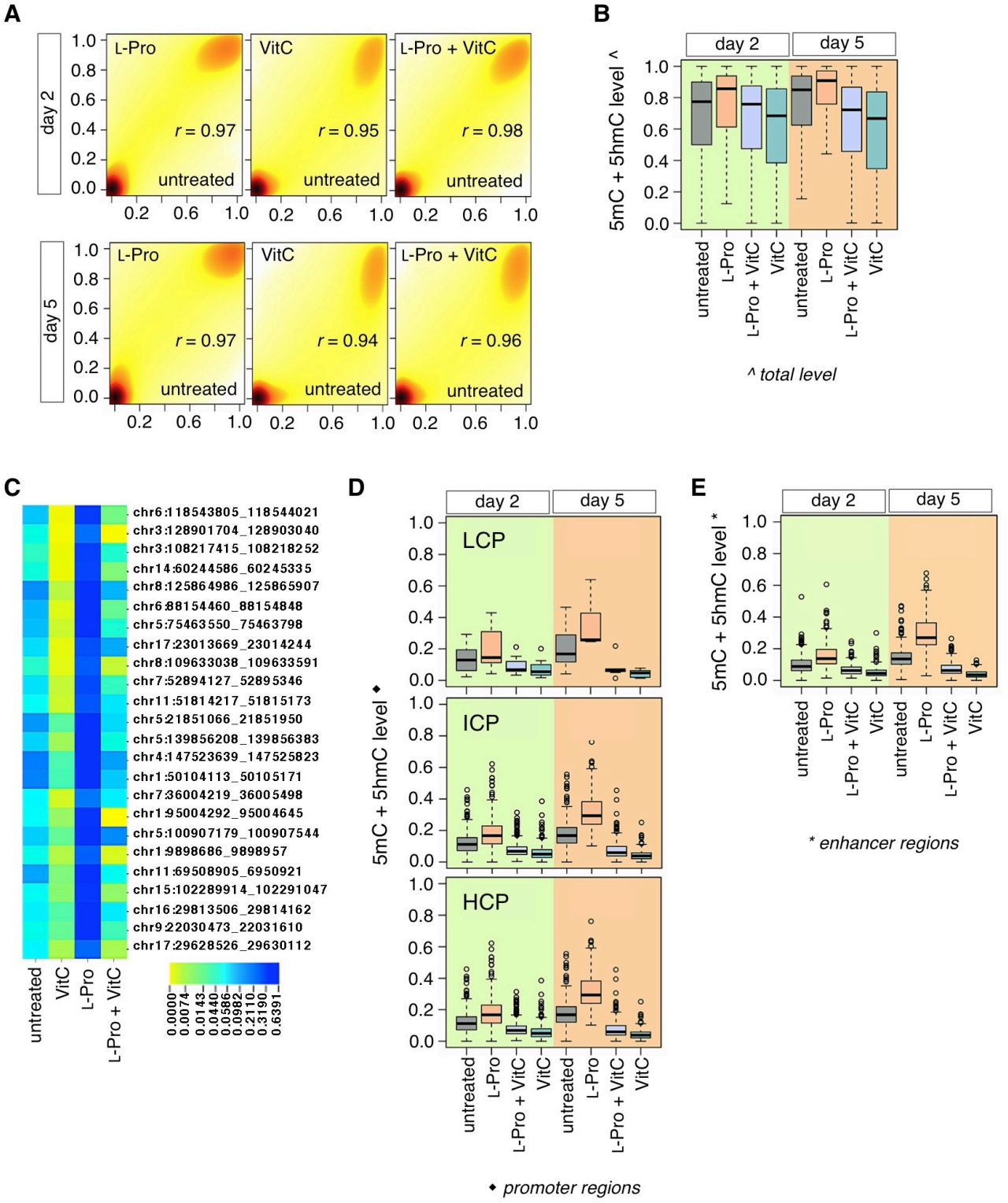

Figure S1. Effect of Vitamin C and L-Proline on DNA Methylation (related to Figure 1)

(A) Density scatterplot of mean methylation levels of the sliding windows. The Pearson correlation coefficients between the different conditions were high ($r > 0.94$).

(B) Global methylation levels in ESCs treated with VitC (500 µM) and L-Pro (150 µM), either alone or in combination, at day 2 and 5 or left untreated as control.

(C) Heatmap (CIMminer, http://discover.nci.nih.gov/cimminer/) showing the methylation levels at 24 selected loci in ESCs ± VitC (500 µM) and ± L-Pro (150 µM) at day 5.

(D) Average profile of methylation over low CpG density promoters (LCP), Intermediate (ICP) and High (HCP).

(E) Average profile of methylation over enhancers.

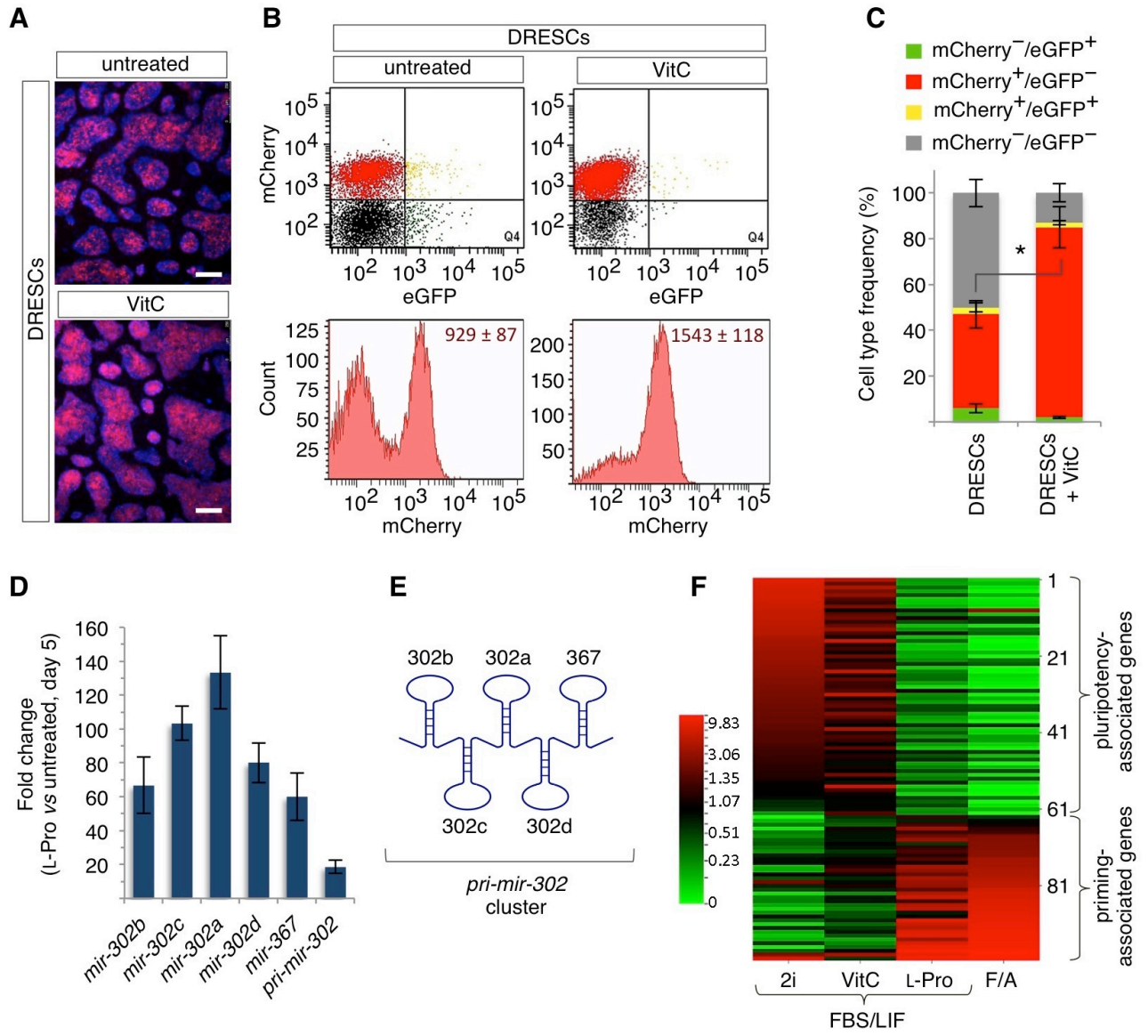

Figure S2. Effect of Vitamin C and L-Proline on *mir-290* and *mir-302* Expression (related to Figure 2)

(A) Representative fluorescence images of DRESCs grown on a feeder layer, in the presence or absence of VitC (100 µM) for three passages. Scale bar, 125 µm.

(B) Representative FACS plots showing the percentage of mCherry±/eGFP± cells (*upper panels*) and mCherry intensity mean (*lower panels*) in untreated and VitC- treated (100 µM) DRESCs.

(C) Quantification of mCherry±/eGFP± cells in DRESCs ± VitC (100 µM). Data are mean ± SEM, *p < 0.001 (n = 3 independent experiments).

(D) qPCR analysis of *mir-302* family members' expression in ESCs treated with L-Pro (1 mM) for 5 days. Data are fold change *versus* control (- L-Pro) after normalization to *Gapdh* and are mean ± SEM (n = 3 independent experiments).

(E) Schematic representation of *mir-302* cluster, including *mir-302b*, *c*, *a*, *d* and the *mir-367* members (*pri-mir-302/367*).

(F) Heatmap (CIMminer, http://discover.nci.nih.gov/cimminer/) showing pluripotency and priming associated genes in ESCs grown in FBS/LIF + 2i, VitC, −− (untreated), L-Pro or F/A.

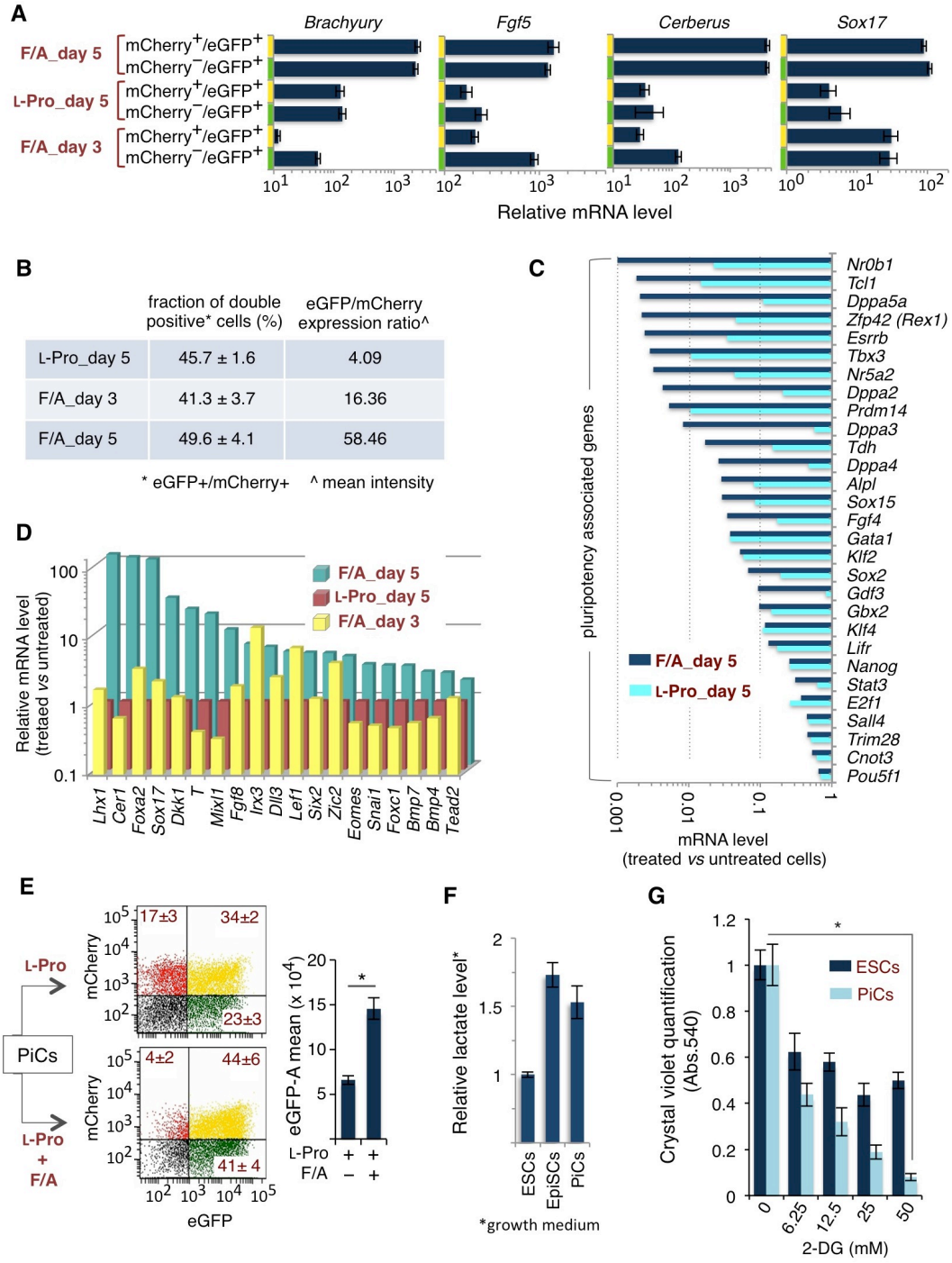

Figure S3. Effect of bFgf/Activin A and L-Proline on ESC Transcription Profile and on the Distribution of mCherry±/eGFP± Cells (related to Figure 3)

(A) qPCR gene expression analysis in mCherry±/eGFP+ cells derived from L-Pro- (day 5) and F/A- (day 3 and 5) treated DRESCs. mRNA levels were normalized to *Gapdh*. Data are mean ± SEM (n = 3 independent experiments).

(B) Fraction of mCherry+eGFP+ cells and eGFP/mCherry ratio in L-Pro- (day 5) and F/A- (day 3 and 5) treated cells. Data are mean ± SEM (n = 3 independent experiments).

(C) mRNA levels of pluripotency associated genes in L-Pro- and F/A- treated cells (day 5) *versus* untreated cells.

(D) mRNA levels of a set of mesendodermal genes in L-Pro- (day 5) and F/A- (day 3 and 5) treated cells.

(E) Representative FACS plots showing the percentage of mCherry±/eGFP± cells (*left panels*) and eGFP intensity mean (*right panels*) in L-Pro- induced cells (PiCs) treated with either L-Pro or L-Pro/F/A for further 3 days. Data are mean ± SEM, *p < 0.01 (n = 3 independent experiments).

(F) Lactate level measured in the conditioned medium derived from ESCs, F/A-EpiSCs and L-Pro -induced cells (PiCs). Data are shown as fold change *versus* ESCs and are mean ± SEM (n = 3 independent experiments).

(G) Dose-dependent effect of the glycolysis inhibitor 2-DG. Quantification of crystal violet staining of ESCs and L-Pro -induced cells (PiCs) at day 5 treated with the glycolysis inhibitor 2-DG (6.25-50 mM, for 16 hours) or left untreated as

control. Data are mean ± SEM (n = 3 independent experiments).

Supplemental Tables

Table S1. List of Differentially Methylated Regions (DMRs); Genes whose Promoters are Differentially Methylated; DMRs- Associated Genes in L-Pro- and VitC- Treated ESCs, Related to Figure 1.
Table S2. List of Differentially Expressed Genes and of Pluripotency-Associated and Priming Markers in 2i-, VitC-, Untreated, L-Pro- and F/A- Treated cells, Related to Figure 2.
Table S3. List of Differentially Expressed Genes; Genes Related to Pluripotency, Mesendodermal Specification, and ERK Superpathway Differentially Expressed in F/A- and L-Pro- Treated Cells, Related to Figure 3.
Table S4. List of Metabolites Deregulated by Supplemental L-Pro in ESCs, Related to Figure 3 and Discussion.
Table S5. Pluripotency-Associated Properties in ESCs, PiCs and EpiSCs, Related to Discussion.
Table S6. Primer Sequences for mRNA/miRNA and qPCR Analysis, Related to Experimental Procedures.

Table S4. List of Metabolites Deregulated by Supplemental L-Pro in ESCs. Relevant metabolites, as found by NMR-based metabolomics, (Variable Importance in Projection (VIP) > 1 and |pcorr| > 0.66) and the chemical shift of the buckets (the isolated variables selected for quantification) are shown. Values are fold-change ratio (L-Pro- treated *vs.* untreated cells) ± standard deviation of the NMR variables normalized to the total area of the spectrum, Related to Figure 3 and Discussion.

Metabolites	Chemical shift (ppm)	Fold change (PiCs vs ESCs)
Taurine	[3.39-3.41]	1.57±0.59
Proline	[3.33-3.35]	1.51±0.13
Lactate	[4.13-4.15]	1.47±0.14
Glycerol	[3.65-3.67]	1.37±0.13
GSH	[2.55-2.57]	1.37±0.24
Glutamate	[2.33-2.35]	1.34±0.21
Phosphocholine	[3.21-3.23]	1.24±0.13
Gaba	[2.25-2.27]	1.23±0.13
UDP-N-AcetylGlucosamine	[7.95-7.97]	-1.26±0.15
Serine	[3.95-3.97]	-1.28±0.23
Leucine	[0.95-0.97]	-1.34±0.35
AMP	[8.59-8.61]	-1.43±0.19
Valine	[0.99-1.01]	-1.29±0.31
Alanine	[1.47-1.49]	-1.73±0.34
Glycine	[3.55-3.57]	-1.99±0.28
Tyrosine	[6.87-6.89]	-1.98±0.76
Glutamine	[2.43-2.45]	-2.19±0.56

Table S5. Pluripotency-Associated Properties in ESCs, PiCs and EpiSCs, Related to Discussion.

Pluripotency-associated property	**ESCs (naïve)**	**PiCs (early-primed)**	**EpiSCs (primed)**
MEK_ERK activation	No	Yes	Yes
Tgfβ-ActivinA activation	No	Yes	Yes
Global DNA hypomethylation	Yes	No	No
Pluripotency markers	↑	↓	↓↓
Priming markers	↓	↑	↑↑
Adhesion molecules	E-Cadherin	N-Cadherin	N-Cadherin
Sensibility to Trypsin	No	Yes	Yes
Metabolism	OxPhos/Glycolytic	Glycolytic	Glycolytic
Contribution to chimeric embryos	High	High	Low

Table S6. Primer Sequences for mRNA/miRNA and qPCR Analysis, Related to Experimental Procedures.

Primer sequences for miRNA qPCR analysis

Primer	Target Sequence	Exiqon product no.
hsa-mir-302a-3p	UAAGUGCUUCCAUGUUUUGGUGA	206059
hsa-mir-302b-3p	UAAGUGCUUCCAUGUUUUAGUAG	204773
hsa-mir-302c-3p	UAAGUGCUUCCAUGUUUCAGUGG	204403
hsa-mir-302d-3p	UAAGUGCUUCCAUGUUUGAGUGU	204311
hsa-mir -367-3p	AAUUGCACUUUAGCAAUGGUGA	204784
mmu-mir -293-3p	AGUGCCGCAGAGUUUGUAGUGU	205003
mmu-mir -294-3p	AAAGUGCUUCCCUUUUGUGUGU	205166
mmu-mir -293-3p	AAAGUGCUACUACUUUUGAGUCU	205078

Primer sequences for qPCR analysis

Primer	Forward	Reverse
pri-mir-302	CTGTGGGTTTGCTCTTCTGTTTT	GAGACAGAAAGCATTCCCATGTT
Fgf5	CAAAGTCAATGGCTCCCACGAAG	CTACAATCCCCTGAGACACAGCAAATA
Pitx2	AGCTGTGCAAGAATGGCTTT	CACCATGCTGGACGACATAC
Inhbb	CGAGATCATCAGCTTTGCAG	TGGTTGCCTTCATTAGAGACG
Brachyury (T)	GAACCTCGGATTCACATCGTGAGA	ATCAAGGAAGGCTTTAGCAAATGGG
Cerberus	AGGAGGAAGCCAAGAGGTTC	CATTTGCCAAAGCAAAGGTT
Sox17	AGCTAAGCAAGATGCTAGGCAAG	TCTCTGCCAAGGTCAACGC
Gapdh	TGCACCACCAACTGCTTAGC	TCTTCTGGGTGGCAGTGATG

Supplemental Experimental Procedures

***In vitro* Generation of L-Proline- Induced Cells (PiCs) and Epiblast Stem Cells (EpiSCs) and Differentiation**

ESCs were cultured in high glucose Dulbecco's modified Eagle's medium (Invitrogen, Life Technologies) supplemented with 15% ES- screened fetal bovine serum (FBS, Euroclone) 0.1 mM β-mercaptoethanol (Sigma-Aldrich), 1 mM sodium pyruvate, 2 mM glutamine, 100 U/ml penicillin/streptomycin all from GIBCO and 1000 U/ml recombinant LIF (ESGRO, Millipore).

2i medium was supplemented with PD0325901 (1 μM) and CHIR99021 (3 μM).

To induce ESC to PiC transition, ESCs were seeded at low density (50–500 cells/cm^2) onto gelatin-coated plates and grown in ESC medium (described above) in the presence of L-Pro (from 500 μM to 2 mM) for 5 days. Medium was changed at day 3 with addition of fresh L-Pro. PiCs were harvested using accutase (Sigma-Aldrich) and cultured in the presence of L-Pro (500 μM).

For ESC to EpiSC transition, ESCs were plated at 1500 cells/cm^2 onto FBS- coated plates and grown in N2/B27 supplemented with 20 ng/ml Activin A (Invitrogen) and 12 ng/ml Fgf2 (Provitro) and cultured for 5 days. Cell colonies were washed twice with phosphate-buffered saline (PBS) and fixed/stained with a solution of 6% glutaraldehyde and crystal violet for 30 min at RT. Cells were carefully washed with tap water and dried for further analysis. For quantification, crystal violet was dissolved with 30% acetic acid for 15 min at RT and absorbance was read at 540 nm, using the Synergy H1 Microplate Reader (BioTek).

Alkaline phosphatase assay was performed using the AP staining kit (System Biosciences), following the manufacturer's instructions.

Neuronal differentiation was performed as previously described (Fico et al., 2008).

5mC and 5hmC HPLC-MS/MS Measurement

DNA degradase plus (ZymoResearch) was used to degrade 2 μg of DNA into individual nucleosides (dG, mdC and hmdC) that were measured using a high performance liquid chromatography-tandem mass spectrometry (HPLC-MS/MS) system. Calibration standards were used to obtain area-based linear regression curves for quantification. The 5mC and 5hmC levels are presented as a ratio of %mdC/dG and %hmdC/dG, respectively.

Flow Cytometry and Cell Sorting

ESCs, PiCs and EpiSCs were dissociated to obtain a single cell suspension, using trypsin-EDTA, and analysed/sorted on the basis of mCherry and eGFP signals, with a FACS Aria (Becton Dickinson). For evaluating oxidative phosphorylation, the TMRE Mitochondrial Membrane Potential Assay kit (Abcam) was used, using the manufacturer instructions.

RNA Extraction and Quantitative Real-Time PCR

Total RNAs were isolated using either the RNeasy kit or Trizol reagent (Invitrogen) and reverse transcribed using QuantiTect Reverse Transcription kit (Qiagen). For miRNA expression analysis, RNAs were reverse transcribed using miRCURY LNA Universal RT microRNA PCR kit (EXIQON), using the manufacturer instructions. qPCR was performed using SYBR Green PCR master mix (FluoCycle IITM SYBR, EuroClone). Primers for miRNAs were purchased from EXIQON and are listed in Table S6. Primers for the *pri-mir*-302 and the other genes were designed using Primer3 software and are listed in Table S6.

For RNA profiling 2 μg of extracted RNA was depleted from ribosomal RNA using Ribo-Zero Gold Kit (Epicentre Madison, Winsconsin, USA). Then, rRNA-depleted RNA was used for library preparation using the Kapa RNA Sample Prep Kit (Kapa Biosystems) following the manufacturer's instructions. Libraries were indexed using NEXTflex adapters (Bioo-Scientific Corporation, Austin, TX, USA) and 2 × 43 bp paired-end sequencing was performed on Illumina NextSeq 500 instruments using TruSeq reagents (Illumina, San Diego, CA, USA), according to manufacturer's instructions.

RNA Sequencing Analysis

For RNA-seq data analysis of FBS/LIF ± VitC, L-Pro, naïve/2i and F/A EpiSCs, raw data were aligned to the Mus musculus NCBIM37.67 (mm9) transcriptome using Bowtie 1 with the following setting: -a --best --strata -S -m 100 -X 500. Quantification of gene expression was performed using MMSEQ (Turro et al., 2011). The MMSEQ expression estimates are roughly in FPKM units (fragments per kilobase of transcript per million mapped reads or read pairs). To do the differential expression analysis, the raw counts were extracted from the MMSEQ outputs using the mmseq.R R script included in MMSEQ package. Next, the R package DESeq was used to normalize the data and then performed pair-wise comparisons (fold change 2, adjusted p-value <0.05 and FPKM>=1 in at least in any condition) to determine the differentially expressed genes per condition.

For RNA-seq data analysis of mCherry+/eGFP+ L-Pro- and F/A- induced cells, files were mapped against the reference genome (mm9, NCBI Build 37) http://hgdownload.cse.ucsc.edu/goldenPath/mm9/chromosomes/ and the transcript annotation Mus musculus.NCBIM37.67.gtf.gz (ftp://ftp.ensembl.org/pub/release-67/gtf/mus_musculus/ repository; version NCBIM37, release 67). For the mapping procedure, we run TopHat2 (Kim et al., 2013) by using the -G option to be guided by the annotation. We run htseq-count (version 0.6.1p1) (Anders et al., 2015) with the intersection-

nonempty mode in order to quantify the number of reads mapped on each gene. All the other analyses were performed by using RNASeqGUI (Russo and Angelini, 2014). Genes with low counts were filtered in all the samples by using the Filtering Proportional Test (Tarazona et al., 2011); this procedure reduced the number of genes from 37620 to 15441. 2. We then applied an Upper Quartile Normalization, and finally run NOISeq (version 2.10.0) (Tarazona et al., 2011) on the filtered and upperquartile- normalized count files. We considered a gene as differentially expressed across the samples if the posterior probability was greater or equal to 0.99 for NOISeq.

Reduced Representation Bisulfite Sequencing (RRBS)
FASTQ files were mapped against mm9 with BSMAP v2.88 using the parameters -R -v 0.1 -u -n 1 -s 12 -w 100 -D C-CGG -q 2. We extracted methylation levels per individual C as #C/(#C+#T) with methratio.py from BSMAP v2.88 using the parameters -z -u -g. For further analysis we filtered CGs to have a minimum sequencing depth of 10X.

Preparation of Cytospin Samples
Cells (1-1.5 x 10^4) were dissociated with accutase for 5 min at 37°C, resuspended in 15% FBS/1x PBS and centrifuged at 900 rpm for 15 min onto glass slides (2 spots of $1x10^5$ cells each) using a Thermo Shandon Cytocentrifuge (CytoSpin™ 4). Specimens were fixed for further analysis.

Immunofluorescence Analysis
Cells were fixed (4% PFA) and permeabilized (0.1% Triton X-100) for 10 minutes at RT and incubated with anti-Gfap (1:300, DAKO #Z0334) and βIII-Tubulin (1:400, SIGMA #T8660) antibodies overnight at 4°C. After washing in 0,5% Tween-1x PBS, cells were incubated with the appropriate secondary antibodies (Alexa Fluor DAR-594, 1:200, Molecular Probes #A21207; Alexa Fluor GAM-488, 1:200, Molecular Probes #A11001). Cell nuclei were counterstained with DAPI (Invitrogen). Images were obtained using the DMI6000B microscope and the DFC 350FX B/W digital camera (Leica). Leica FW4000 and AF6000 software were used for image acquisition/elaboration.

Western Blotting
Whole-cell lysates were prepared in 100 mM Tris pH 8, 140 mM NaCl, 20 mM EDTA, 0,2% SDS, 1% Nonidet P-40 lysis buffer, resolved on SDS-PAGE gels and transferred onto PVDF membranes using the iBlot dry Transfer System (Life Technologies). Blocked membranes were incubated with the following primary antibodies: phospho-Erk1/2 (1:1000, Cell Signaling #9101); Erk (1:1000, Cell Signaling #9102); phospho-Smad2 (1:200, Cell Signaling #3108); Smad2 (1:1000, Cell Signaling #3103); anti Gapdh (1:10000, Abcam #Ab8245), followed by the appropriate HRP-conjugated secondary antibodies (1:10000, GAM-HRP #P0447 and GAR-HRP #P0448 DAKO). Detection was performed with ECL reagents (Pierce, Thermo Scientific). For densitometric analysis the ImageJ software was used.

NMR-based Metabolomic Analysis
Cells extracts and NMR samples were prepared as described (Lindon et al., 2005). NMR spectra of polar extracts were recorded at 600.13 MHz on a Bruker Avance-600 spectrometer equipped with a TCI CryoProbe™ fitted with a gradient along the Z-axis, at a probe temperature of 300 K. Assignment of NMR signals to specific metabolites was obtained by homonuclear and heteronuclear 2D experiments (TOCSY to identify ^{1}H-^{1}H connectivities, and ^{1}H-^{13}C HSQC for directly bonded ^{1}H and ^{13}C nuclei). Assignments were also compared with literature data and/or online database (Wishart et al., 2007).

Multivariate Statistical Analysis
The 1D spectra, representing the metabolic profiles, were subjected to multivariate statistical analysis to identify trends and clusters (Eriksson et al., 2001). The statistical analysis projects NMR data into a lower dimensional space, where sample clustering based on similarities of biochemical profiles can be determined and easily visualized through scores plot. To that purpose, ^{1}H-NMR spectra were automatically data reduced to 440 integrated regions ("buckets") with an equal width of 0.02 ppm over the spectral region between 0.60 and 9.40 ppm by using AMIX 3.9.7 software package (Bruker Biospin, Germany). The region 5.42- 4.38 ppm containing the residual water resonance was excluded, and each integrated region was normalized to the total spectrum area. Hence, a data matrix **X** was created with samples in rows and NMR variables in columns, while class belonging for each sample was encoded into a **Y** matrix. Projection to Latent Structure-Discriminant Analysis (PLS-DA) was then applied to **X** and **Y** matrices, containing control and L-Pro-treated cells data, with SIMCA 14 software (Umetrics, Umeå, Sweden) in order to distinguish cell classes according to the induced metabolite distribution. The obtained statistical model showed R2Y=0.98 and Q2=0.96 (goodness of fit and goodness of prediction, respectively). Acceptable values must be ≥ 0.5 for both, with $|R2 - Q2| < 0.2–0.3$. The robustness of the PLS-DA classification was assessed by permutation test (800 times), which estimated $R2_{fit}$= 0.32 and $Q2_{fit}$=-0.35 suggesting a non-overfitted model. Moreover, CV-ANOVA testing of Cross-Validated predictive residuals was used to assess the reliability of the model obtaining $p=3\times10^{-8}$.

Lactate Activity Assay
Lactate was measured using the colorimetric L-Lactate Assay Kit (Abcam, Cambridge, MA, USA, ab65331), according to the manufacturer's instructions. Data were normalized to total cell number.

Supplemental References

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.

Eriksson L, Johansson E, AB. U, Academy U, Kettaneh-Wold N, Wold S. Multi- and Megavariate Data Analysis: Principles and Applications: Umetrics Academy; Umeå, Sweden, 2001.

Fico, A., Manganelli, G., Simeone, M., Guido, S., Minchiotti, G., and Filosa, S. (2008). High-throughput screening-compatible single-step protocol to differentiate embryonic stem cells in neurons. Stem cells and development *17*, 573-584.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome biology *14*, R36.

Lindon, J.C., Nicholson, J.K., Holmes, E., Keun, H.C., Craig, A., Pearce, J.T., Bruce, S.J., Hardy, N., Sansone, S.A., Antti, H.*, et al.* (2005). Summary recommendations for standardization and reporting of metabolic analyses. Nature biotechnology *23*, 833-838.

Russo, F., and Angelini, C. (2014). RNASeqGUI: a GUI for analysing RNA-Seq data. Bioinformatics *30*, 2514-2516.

Tarazona, S., Garcia-Alcalde, F., Dopazo, J., Ferrer, A., and Conesa, A. (2011). Differential expression in RNA-seq: a matter of depth. Genome research *21*, 2213-2223.

Turro, E., Su, S.Y., Goncalves, A., Coin, L.J., Richardson, S., and Lewin, A. (2011). Haplotype and isoform specific expression estimation using multi-mapping RNA-seq reads. Genome biology *12*, R13.

Wishart, D.S., Tzur, D., Knox, C., Eisner, R., Guo, A.C., Young, N., Cheng, D., Jewell, K., Arndt, D., Sawhney, S.*, et al.* (2007). HMDB: the Human Metabolome Database. Nucleic Acids Res *35*, D521-526.