Supporting Information

Tompkins et al. 10.1073/pnas.1209620109

SI Materials and Methods

Total RNA Preparation, Gene Expression, and Ontology Analysis. Total RNA was isolated with RNeasy Mini Kit (Qiagen) according to manufacturers' specifications. Quality and purity was assessed by Nanodrop absorbance ratios and by Agilent Bioanalyzer (UCLA microarray core); 1 μg of total RNA for each sample was submitted for microarray analysis and was carried out at the UCLA DNA Microarray Core on NimbleGen 12 × 135 k human gene-expression arrays (Roche NimbleGen). Data were extracted from array images as log2 ratios and normalized using quantile normalization. Gene calls are made by Robust Multi-array Average (RMA) algorithm. Technical replicates were all highly correlated (R2 >0.94 all samples; Table S2) and combined (mean) for all subsequent gene expression analyses. Data analysis was conducted with ArrayStar4 software (DNAStar). Noise-filtered genes were considered differentially expressed between samples if possessing a P value of <0.01 (moderated t test) and a greater than twofold change. Irreversible expression changes maintained these cutoffs post–reverse adaptation. For validation of gene-expression data real-time RT–qPCR was carried out with iScript Reverse Transcription Supermix (BioRad) on the DNA Engine Opticon system (Bio-Rad). Primers and optimal PCR conditions are provided in Table S6. Relative enrichment normalized to an internal reference control (GAPDH) was calculated using the standard curve method. Ontology analysis was conducted using DAVID v6.7 using functional annotation (1). Significance thresholds required an Expression Analysis Systematic Explorer (EASE) score of $P <$ 0.01, a fold change of >1.3, and a minimum of four genes identified for each term.

MIRA-Chip Analysis. Genomic DNA from Tra-1–60+ sorted hESCs was isolated using the DNeasy Blood and Tissue Kit (Qiagen). DNA was subsequently fragmented by sonication to ~300–900 bp. Fragmentation was confirmed by gel electrophoresis on a 2% agarose gel stained with ethidium bromide. In accordance to manufacturer protocol, 500 ng of sonicated DNA was enriched for methylated DNA using the MethylCollector Ultra Kit (Active Motif), which is based on MIRA (2, 3). Eluted DNA was purified with MinElute reaction cleanup kit (Qiagen). MIRA-enriched and 10 ng of sonicated input DNA was subsequently amplified using the Whole Genome Amplification kit2 (Sigma). Then 2 μg of amplified samples were labeled and cohybridized with input DNA onto NimbleGen CpG Island Plus RefSeq Promoter Arrays. Microarray analysis was performed at UCLA DNA Microarray Core. Note that MIRA does not enrich for hydroxymethylated DNA and therefore 5hMC is not within the scope of these experiments (4).

Immunostaining, Karyotyping, and Cell Sorting. Cells were fixed in 4% (vol/vol) paraformaldehyde for 10 min, washed twice with PBS, and then blocked for 1 h in PBS 5% (vol/vol) goat serum/0.1% triton for 1 h. Cells were incubated overnight at 4 °C in either 40 μg/mL anti-Oct-4 (sc-5279; Santa Cruz Biotechnology), 40 μg/mL anti-Nanog (sc-33759; Santa Cruz Biotechnology) in PBS 8% (vol/vol) goat serum/0.1% triton, or buffer alone. Cells were washed twice and then incubated for 1 h at room temperature in Cy2 goat antimouse IgG (115–226-003; Jackson Immunoresearch), Rhodamine red-X goat antirabbit IgG (111–295-003; Jackson Immunoresearch). Cells were washed twice with PBS, Prolong gold antifade with DAPI (P36935; Invitrogen) was added, and the cells were visualized using an Olympus IX50 microscope.

Karyotyping of cells was conducted by GTG-banding by the City of Hope Cytogenetics Core Laboratory. At least 20 cells were studied for each sample and results are summarized in Table S1.

To prevent MEF contamination from altering methylation profiling, HES-2 cells were first enzymatically dissociated from MEFs using 1× collagenase solution and separated into single cells with accutase (Invitrogen). Cultures were then immunostained with α-Tra-1–60 (Millipore) followed by flow-cytometry-based isolation of Tra-1–60-expressing cells. For consistency, Mat/mTR and CS/SP cells were also enzymatically separated with accutase and sorted for Tra-1–60 (+) cells. High Tra-1–60 (+) staining was noted in all cultures (Table S1).

Identification and Annotation of Methylated Peaks and Methylated Regions. Data were extracted from microarray images as scaled log2 ratios of MIRA-enriched samples versus input. Log2 ratio data were converted into P value scores using the Kolmogorov–Smirnov test with a 750-bp window by using NimbleScan software [$S_p = -\log10(p)$]. Probes were selected as positive for peaks if their P value scores were above 2 ($P < 0.01$). For our analysis, we defined a methylated peak as a region spanning at least four positive probes (no gaps) covering a minimum length of 350 bp. This stringent definition yields few false-positive peaks. Peak scores were averaged from technical replicates. Identified methylation peaks were mapped relative to known transcripts defined in the University of Southern California genome browser HG18 RefSeq database (http://hgdownload.cse.ucsc.edu/goldenPath/hg18/database/). Methylation peaks falling −500 to +610 bp relative to TSS were defined as promoter-proximal peaks; methylation peaks falling within −2,440 to −501 bp relative to TSS were considered promoter-distal peaks; those falling +610 bp from TSS to the RefSeq transcript end sites were defined as intragenic peaks (Fig. 4*A*). Intergenic peaks were defined as any falling outside the constraints of the preceding definitions. To identify sample-specific DMRs, all identified methylated peaks were evaluated for significant differences between samples. As such, the average P value scores of probes falling within methylated peaks in one sample were compared with the average P value scores of corresponding probes in another sample and vice versa. The difference in scores was termed DMR score. Peaks possessing a >2.5 DMR score in pairwise comparison were considered sample-specific and no false-positive DMRs were identified with this cutoff. Individual DMRs were classified as promoter-proximal, promoter-distal, intragenic, or intergenic if >50% of the DMR fell within a single location. A Mat/mTR- and CS/SP-specific DMR must have been present both in the initial adaptation cells and those after additional passaging. A DMR was considered reversible if Mat/mTR- or CS/SP-specific and indistinguishable to MEF/DF12 original start cultures after reverse adaptation (DMR score <1.0 between reverse-adapted sample and original MEF/DF12 sample). Irreversible Mat/mTR or CS/SP DMRs maintained a DMR score >2.5 post–reverse adaptation compared with original MEF/DF12 cultures. There were no instances of "partial reversion." Last, for genomewide assessments of methylation, whole promoter and intragenic regions were considered methylated if possessing a methyl score (mean P value score of all probes) of >1.5 across the entire defined regions. Whole promoter region was defined as −2,440 to +610 bp, which corresponds to total tiling coverage of promoters by this array design.

Bisulfite Sequencing. Bisulfite conversion of genomic DNA was performed with EpiTect Bisulfite Kit (Qiagen). Primers for bisulfite PCR as well as optimal PCR conditions are described in Table S6. Bisulfite PCR-amplified DNA was electrophoretically separated on 2% agarose gel and bands extracted with a Gel Extraction Kit (Qiagen). Purified target DNA was cloned into pDrive vector and EZ competent cells were transformed with plasmid DNA (PCR Cloning plus Kit; Qiagen). DNA was isolated from transformed bacteria (Qiaprep Spin Miniprep; Qiagen) and sequenced at City of Hope DNA sequencing core.

Knockdown of GPSM3. HES-2 cells grown in Mat/mTR were subjected to siRNA transfections. siRNAs were selected from *Silencer* Select siRNAs (Life Technologies) to target GPSM3 transcripts. A scramble control was included. Transfections were conducted with RNAiMAX lipofectamine reagent according to manufacturers' protocol (Invitrogen). Knockdown was confirmed by microarray analysis comparing scramble control to untransfected cells and to GPSM3 knockdown samples.

SI Results

Despite five or seven passages on MEFD12 from Mat/mTR, and five or ten passages from CS/SP, many transcripts did not revert to levels seen in original MEF/DF12 cultures (Fig. 2*D*). In the case of Mat-MEF R.A., irreversible transcripts were insufficient in numbers for gene ontology analysis (28 transcripts). However, examination of irreversible gene lists demonstrates many adhesion factors escape reversion to original MEF/DF12 transcript levels. For example, gene ontology analysis had originally identified 35 CS/SP-specific adhesion factors that were up-regulated (GO:0007155~cell adhesion, Dataset S1). Twelve of these were irreversibly up-regulated.

1. Huang W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4:44–57.
2. Rauch TA, Wu X, Zhong X, Riggs AD, Pfeifer GP (2009) A human B cell methylome at 100-base pair resolution. *Proc Natl Acad Sci USA* 106:671–678.
3. Rauch T, Li H, Wu X, Pfeifer GP (2006) MIRA-assisted microarray analysis, a new technology for the determination of DNA methylation patterns, identifies frequent methylation of homeodomain-containing genes in lung cancer cells. *Cancer Res* 66: 7939–7947.
4. Jin SG, Kadam S, Pfeifer GP (2010) Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. *Nucleic Acids Res* 38:e125.

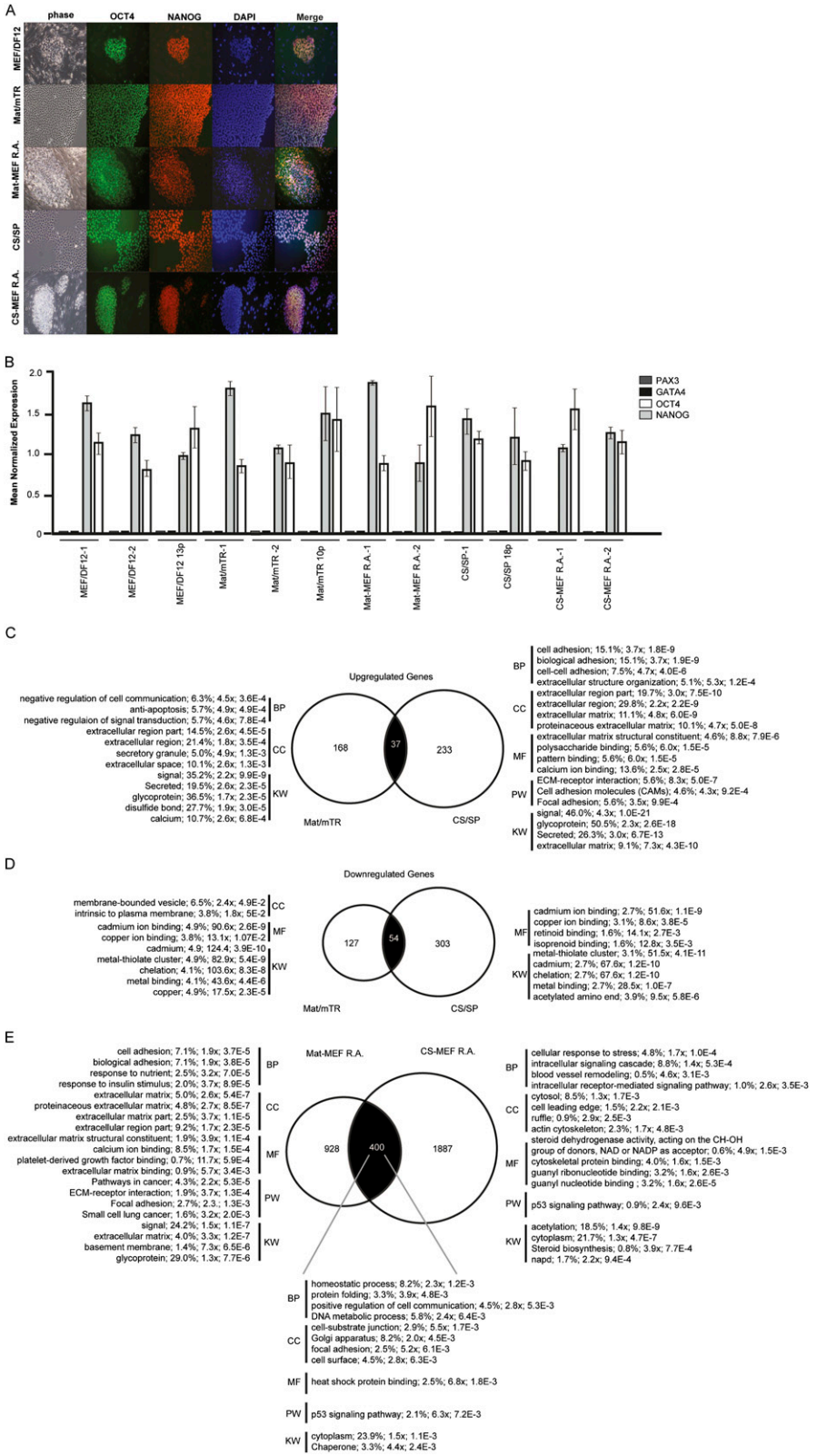

Fig. S1. (*A*) Representative images of hESCs grown in all tested conditions. Nanog and OCT4 protein expression was validated by immunofluorescence. Phase contrast images are provided. Merged images illustrates overlay of OCT4 and Nanog with nuclear marker DAPI. Feeder cells serve as inherent Nanog and OCT4 negative controls. (*B*) Validation of microarray expression results by real-time qRT–PCR. Expression values for OCT4 and Nanog pluripotency markers, as well as PAX3 and GATA4 lineage–specific markers were normalized to an internal reference control (GAPDH). Mean normalized expression ± SE are shown for each culture conditions tested. Reactions were conducted in triplicate. (*C*) Venn diagram and ontology analysis of culture-specific up-regulation. (*D*) Venn diagram and ontology analysis of culture-specific gene down-regulation. Overlapping regions indicate genes that are commonly up- or down-regulated in both

Legend continued on following page

Mat/mTR and CS/SP cultures. (*E*) Venn diagram and gene ontology analysis illustrating overlap of the unique differential gene expression between cells reverse adapted from either Mat/mTR or CS/SP. Ontology analysis was conducted with DAVID v6.7. The top four enriched ontology terms, when identified, arranged by EASE scores (*P* value) are shown for each database used. BP: biological process; CC: cellular component; MF: molecular function; PW: pathway; KW: keywords. Values from left to right following enriched terms are as follows: percent of total genes identified, fold change, and EASE score. See *Materials and Methods* for gene ontology significance thresholds.

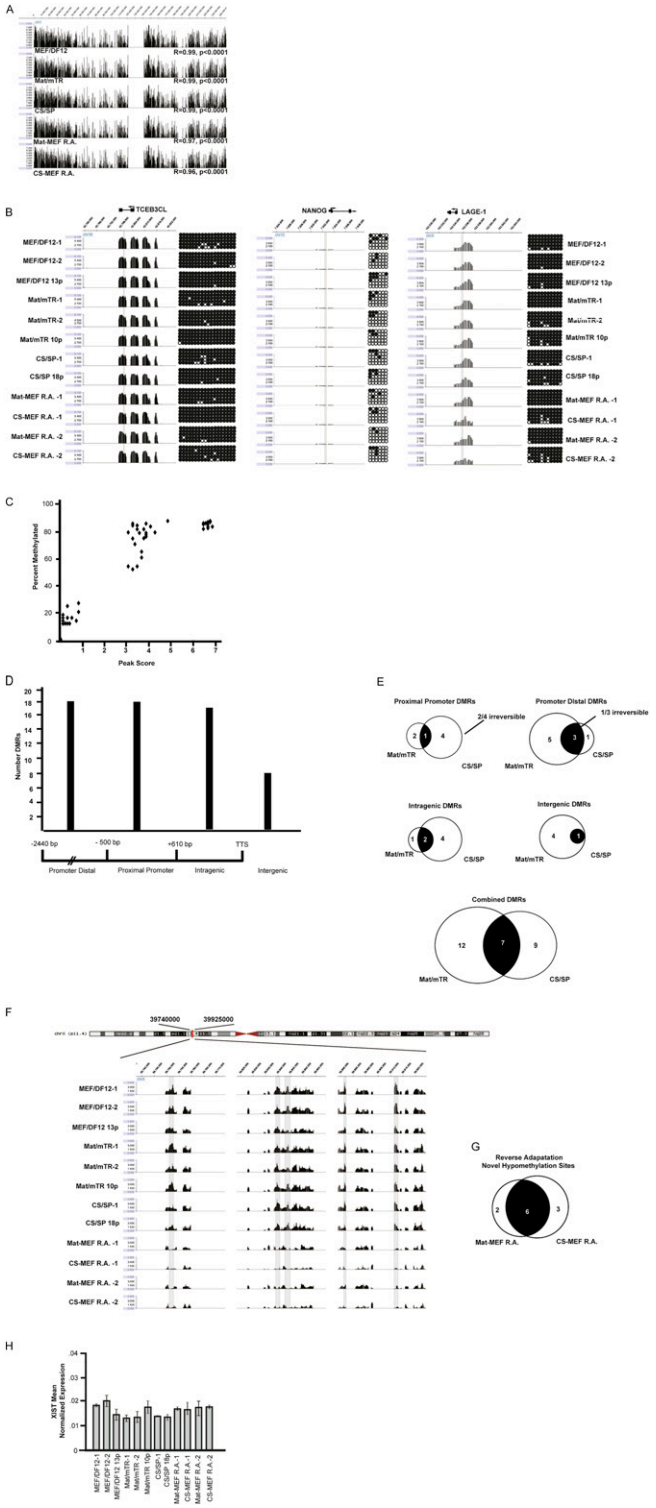

Fig. S2. (*A*–*C*) Visualization and validation of MIRA-chip results. (*A*) Representative chromosomewide view of DNA methylation patterns along chromosome 1 with Pearson's correlation values indicated right (comparison of *P* value data between indicated sample and start MEF/DF12 cultures). (*B*) In addition to validation of DMRs shown in Fig. 3, bisulfite sequencing was used to confirm microarray results at three distinct genomic loci for all 12 samples. Highlighted regions illustrate the targeted span of bisulfite PCR primer sets. Circles represent consecutive CpG dinucleotides. Dark circles: methylated CpG sites. Open circles: unmethylated CpG sites. (*C*) A direct correlation exists between methylation intensity as identified by MIRA-chip and percent methylation as identified by bisulfite sequencing. Bisulfite-sequencing data were pooled and percent methylation plotted versus corresponding methylation score. Methylation peak score is defined here as the mean probe intensity (*P* value) for all probes within the associated peak or in the case of unmethylated region, the mean probe intensity for four consecutive probes surrounding the bisulfite-sequencing target region. Bisulfite-sequencing details, including primer design and annealing temperature are provided in Table S6. (*D* and *E*) Genomic location of DMRs. (*D*) Total DMRs observed in each genomic context were tabulated. (*E*) DMRs observed in Mat/mTR and CS/SP were examined for overlap between the two culture systems and stratified further by gene context and reversibility. Over-

Legend continued on following page

lapping DMRs identified in both Mat/mTR and CS/SP are darkened on the Venn diagrams. (*F*) A region of apparent susceptibility to hypomethylation is shown over an ~180-kb region on chromosome x. Gray highlights show DMRs identified as hypomethylated in reverse-adapted cells only. Tiling between three expanded subregions was not available for this microarray design. (*G*) The majority of hypomethylation sites identified during reverse adaptation from Mat/mTR were shared with hESCs that were reverse adapted from CS/SP. Venn diagram displays overlap of these hypomethylated regions between cultures. (*H*) Expression values for X-inactive specific transcript (XIST) normalized to an internal reference control (GAPDH). Mean normalized expression ± SE are shown for each condition tested. Reactions conducted in triplicate.

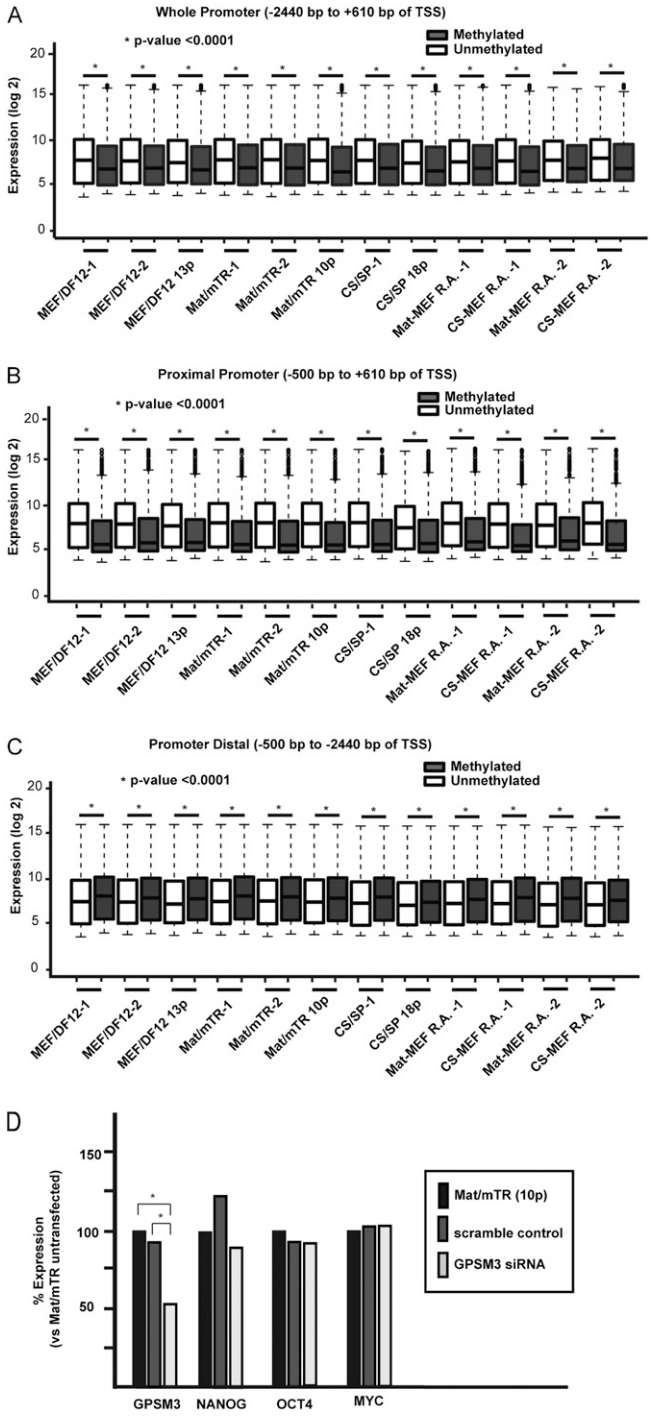

Fig. S3. Analysis of DNA methylation and corresponding gene expression by box plot. (A–C) Whole gene promoters (*A*) were broken down into promoter-proximal (*B*), and promoter-distal (*C*) regions as described in Fig. 4*A*. Methylation status and corresponding average gene-expression values were determined for each genomic region for all samples. Box plots, extending from 25th to 75th percentiles, represent the relative degree of gene expression for methylated and unmethylated genes in each sample. **P* value < 0.0001 (*t* test). (*D*) Examination of knockdown efficiency by microarray expression values. Percent expression was calculated by dividing linear expression values of scramble control and GPSM3, to Mat/mTR untransfected cells cultured for 10 passages. Pluripotency markers are included demonstrating that GPSM3 silencing does not alter expression levels for these genes. **P* value <0.01 (moderated *t* test).

Table S1. Summary of hESC samples collected for DNA methylation analysis and results of karyotype analysis

Table S1

Sample names and descriptions are provided. Tra 1–60 (+) values were identified by flow cytometry during cell sorting. Samples were intermittently karyotyped by GTG- banding (Giesma) and no abnormalities were detected. At least 20 cells were tested for each sample examined.

Table S2. Summary of correlation values between microarrays

Table S2

(*A*) Summary of correlations between technical and biological replicates. For MIRA-chip, Pearson correlation values were generated for *P* value data between samples as indicated. R^2 values from log2 expression data are displayed for replicates on the right. (*B*) Correlations between expression profiles for culture-adapted, prolonged-adaptation, and reverse-adapted samples. R^2 values from log2 expression data are displayed for pairwise comparison between samples as indicated.

Table S3. Distribution of DNA methylation across samples

Table S3

(*A*) Distribution of DNA methylation genomewide. All annotated promoter, intragenic CpG islands, and intergenic CpG islands were assessed for methylation content. (*B*) DNA methylation state at pluripotency genes. Pluripotency genes were examined for whole promoter and intragenic CpG island DNA methylation state and sample-specific DMRs.

Table S4. Medium composition and corresponding gene expression changes

Table S4

Medium supplements are shown *Left* and corresponding gene-expression results are displayed as either fold change for a particular gene or as enriched ontology term. Ontology terms are italicized and taken from ontology results shown in Dataset S1*B*. Several medium components target TGF-β signaling. Significant variations in TGF-β signaling is not described in detail within the table, but can be seen by examining lists of significantly changing genes by culture condition (Dataset S1*A*). These are organized by fold change. For example, the top five down-regulated genes in Matrigel/mTesR1 are directly associated with TGF-β signaling.

Table S5. Description of differential DNA methylation identified across cultures

Table S5

(*A*) Summaries of identified DNA methylation changes through culture adaptation. Chromosomal locations are provided. DMR score refers to the difference between mean *P* value scores between samples for probes falling within an identified methylation peak. Identification of DMRs and reversibility is described in *SI Materials and Methods*. Each DMR identified was examined for sample specificity as an indication of common methylation changes that may exist between cultures. Genes associated with DMRs and corresponding expression changes are described as well. NA indicates lack of expression microarray coverage for particular gene. (*B*) Descriptions of identified genes with DMRs. Full descriptions including DMR score, genomic coordinates, and corresponding expression change can be seen in *A*. DMR types: Rev Hyper/Hypo: Reversible Hypermethylation/Hypomethylation; Irrev Hyper/Hypo: Irreversible Hypermethylation/Hypomethylation; L.T. Hyper/Hypo: Long-term Hypermethylation/Hypomethylation; R.A.N. Hyper/Hypo: Reverse-Adapted Novel Hypermethylation/Hypomethylation. Functional classification is from Gene Ontology Annotation (UniProtKB-GOA). NA indicates no functional classification available. *AMPD3 showed progressive accumulation of DNA methylation in all long-term cultures and therefore was not considered irreversible in CS/SP cultures. **Multiple DMRs within gene region.

Table S6. Description of PCR targets, primers, and annealing temperature used to validate microarray results

Table S6

Top, RT-PCR; *Bottom*, bisulfite PCR.

Dataset S1. Differentially expressed genes and gene ontology lists

Dataset S1

(*A*) Differentially expressed transcripts by culture condition. Gene-expression changes were broken down by culture specificity and direction of transcriptional change. Fold change and *P* value (moderated *t* test) are shown for both initial and prolonged adaptation on Mat/mTR or CS/SP. Expression changes are sorted by fold change. Commonly up- or down-regulated genes are shown as well. (*B*) Complete ontology analysis of culture-specific expression changes. Genes identified as up- or down-regulated in a culture-specific manner were subjected to gene ontology analysis by DAVID. See Fig. S1*E* for gene ontology analysis description. Terms are ranked by *P* value for each ontology category. GenBank accessions are provided for each gene identified within a particular term. (*C*) Summary of irreversible gene-expression changes by culture environment. Differential gene expression identified during Mat/mTR or CS/SP acclimation was considered irreversible if still significantly up- or down-regulated (>twofold change, $P < 0.01$, *t* test) in final MEF/DF12 reverse-adapted samples compared with Mat/mTR- or CS/SP-intermediates cultures. Fold change and significance values are shown for both initial adaptation to Mat/mTR or CS/SP and reverse adaptation from these intermediates back to MEF/DF12. (*D*) Gene ontology analysis of differential gene expression unique to reverse-adapted cells. Transcripts were considered uniquely differentially expressed in reverse-adapted cells if they had significant expression changes (fold change >2, *P* value <0.01) compared with both their intermediate culture (Mat/mTR or CS/SP) and original MEF start culture. Therefore, reversible and irreversible transcripts were not considered here. See Fig. S1*E* for gene ontology analysis description.