

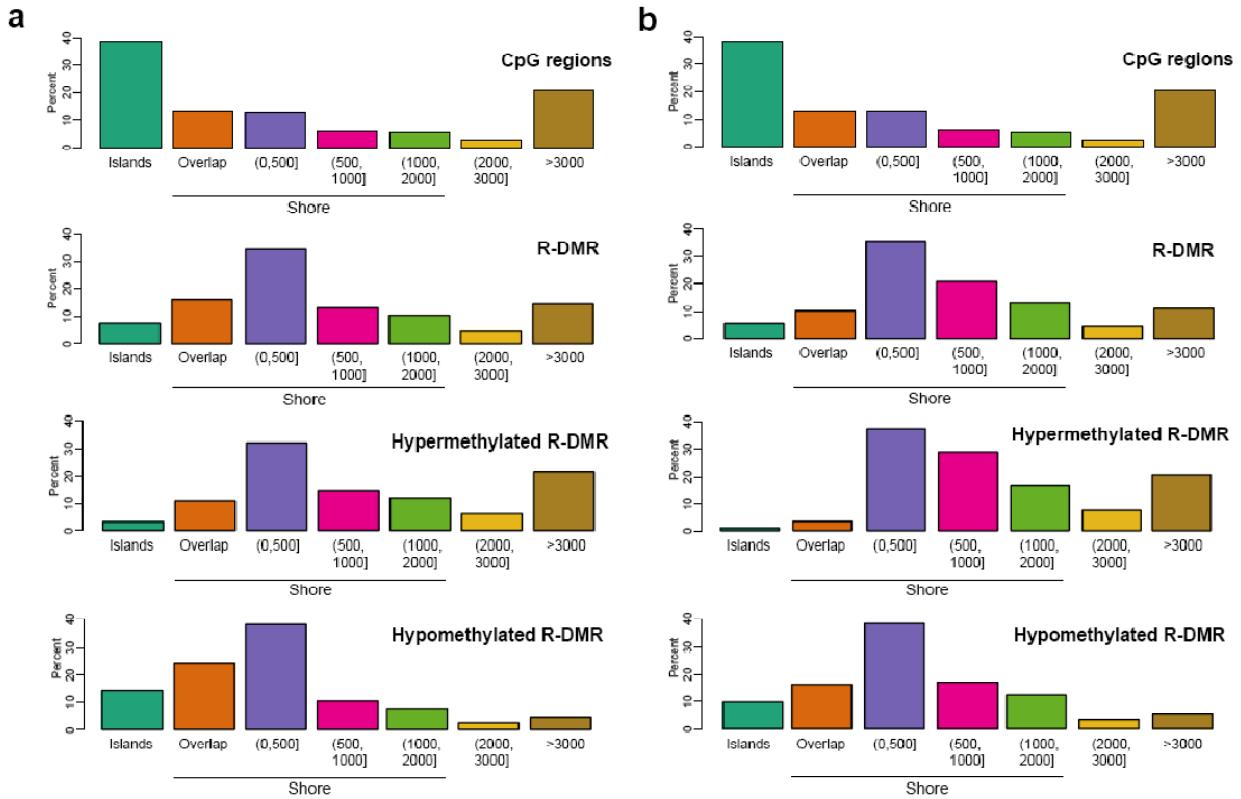
Differential Methylation of Tissue- and Cancer-specific CpG Island Shores Distinguishes Human Induced Pluripotent Stem Cells, Embryonic Stem Cells, and Fibroblasts

Akiko Doi^{1,*}, In-Hyun Park^{2,*}, Bo Wen^{1,*}, Peter Murakami¹, Martin J. Aryee^{3,4}, Rafael Irizarry³, Brian Herb¹, Christine Ladd-Acosta¹, Junsung Rho², Sabine Loewer², Justine Miller², Thorsten Schlaeger², George Q. Daley^{2,+}, Andrew P. Feinberg^{1,+}

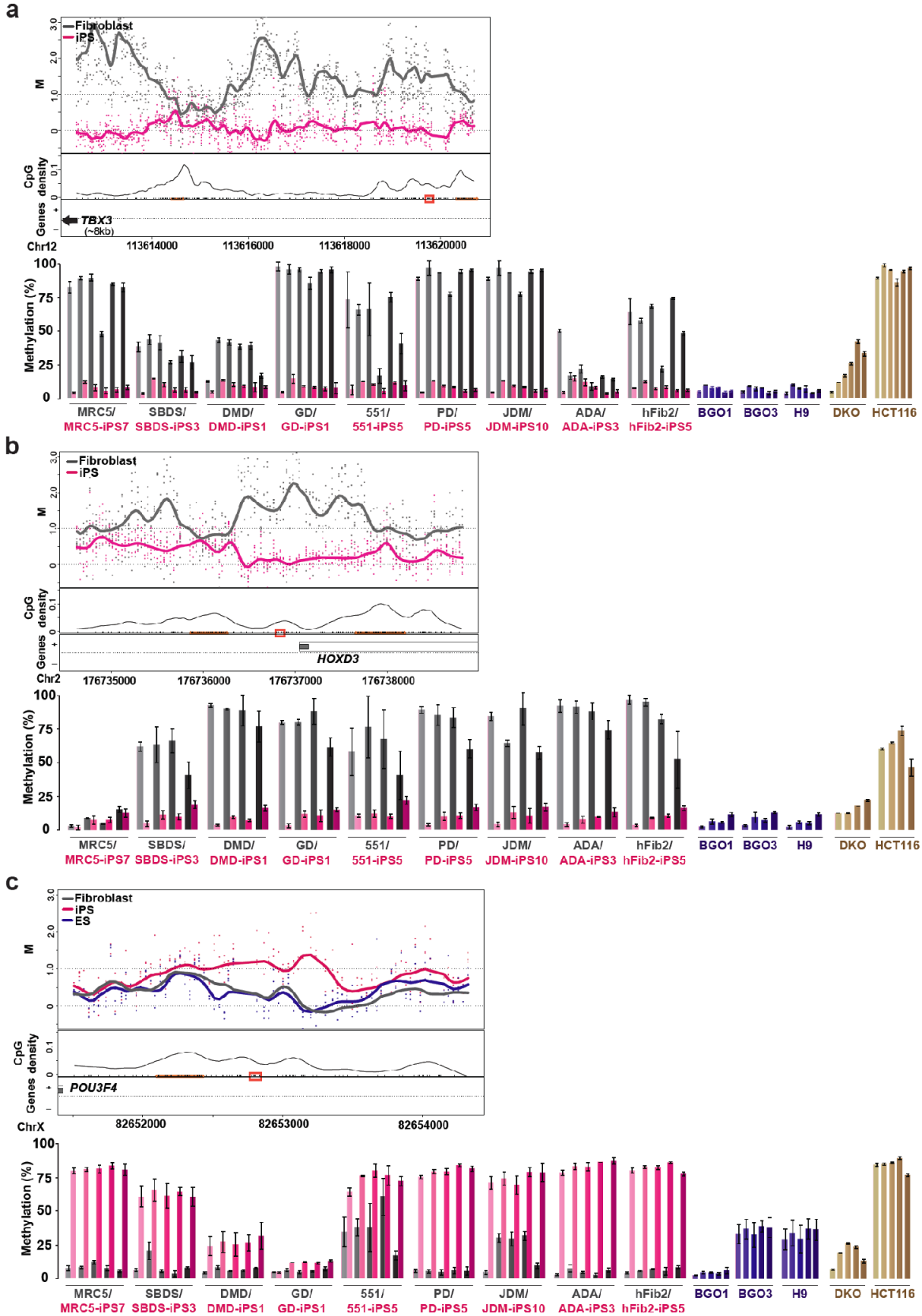
¹Center for Epigenetics and Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; ²Division of Pediatric Hematology/Oncology, Children's Hospital Boston and Howard Hughes Medical Institute, Boston, Massachusetts 02115, USA; ³Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA; ⁴Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, Maryland, USA.

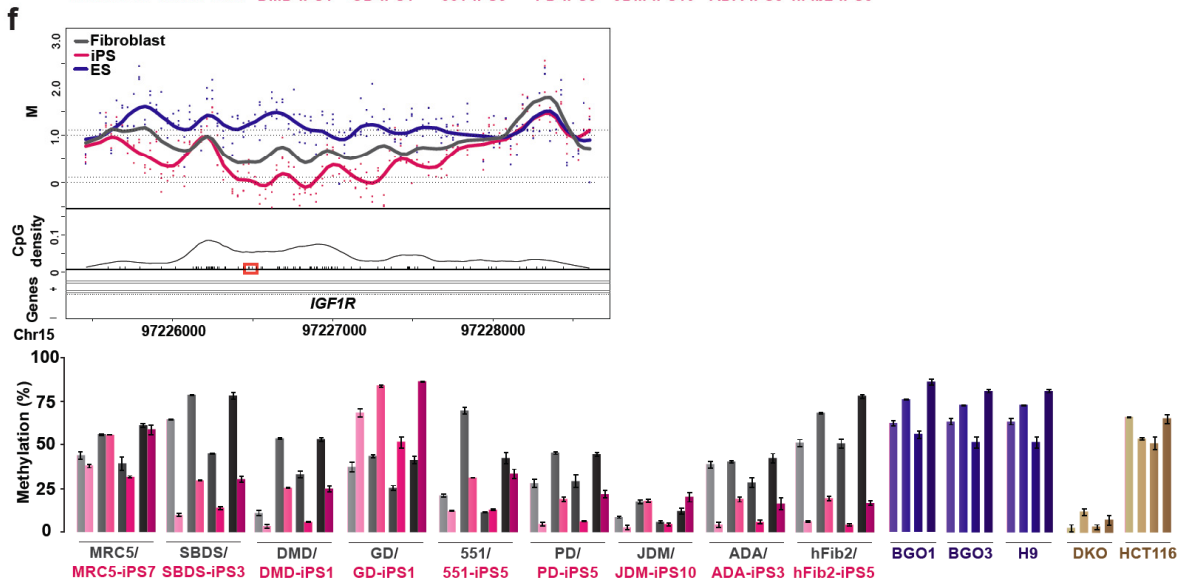
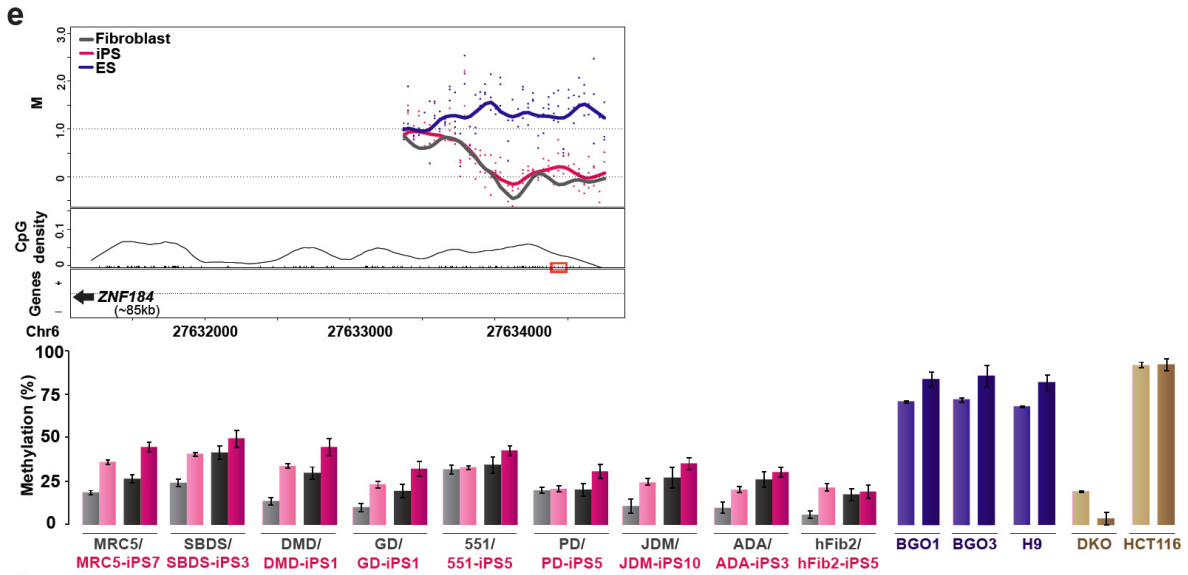
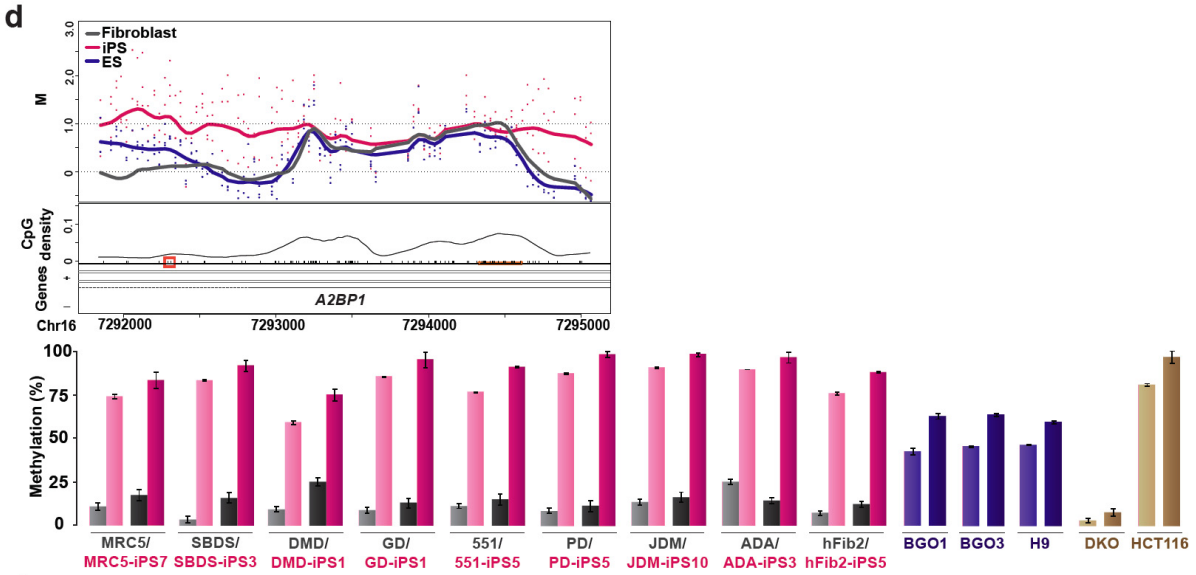
*Equal contribution from these authors

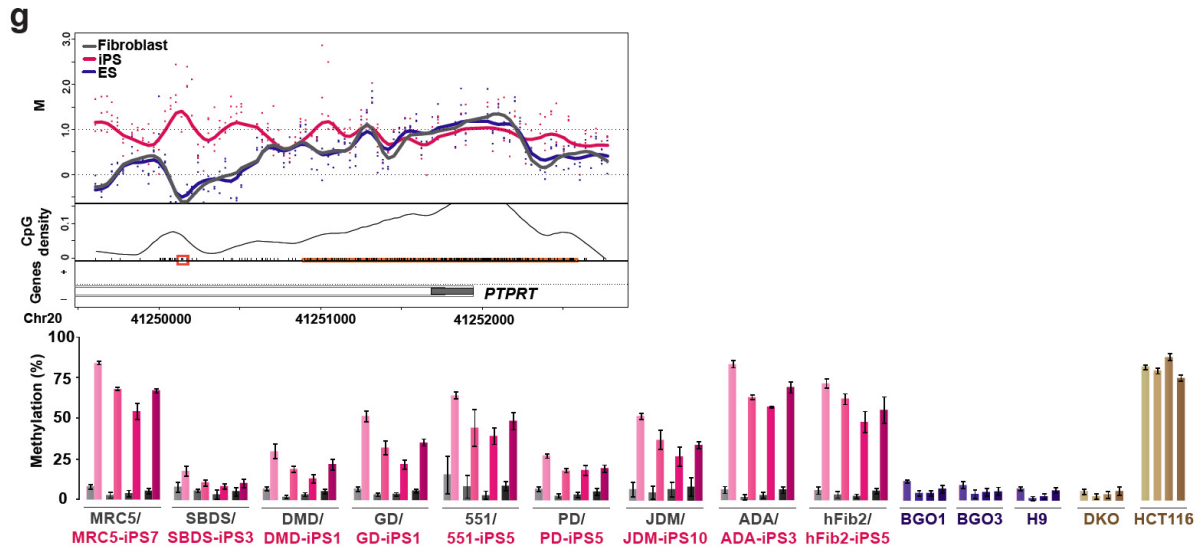
⁺These authors jointly supervised this work. Address correspondence to G.Q.D. (George.Daley@childrens.harvard.edu) and A.P.F. (afeinberg@jhu.edu)



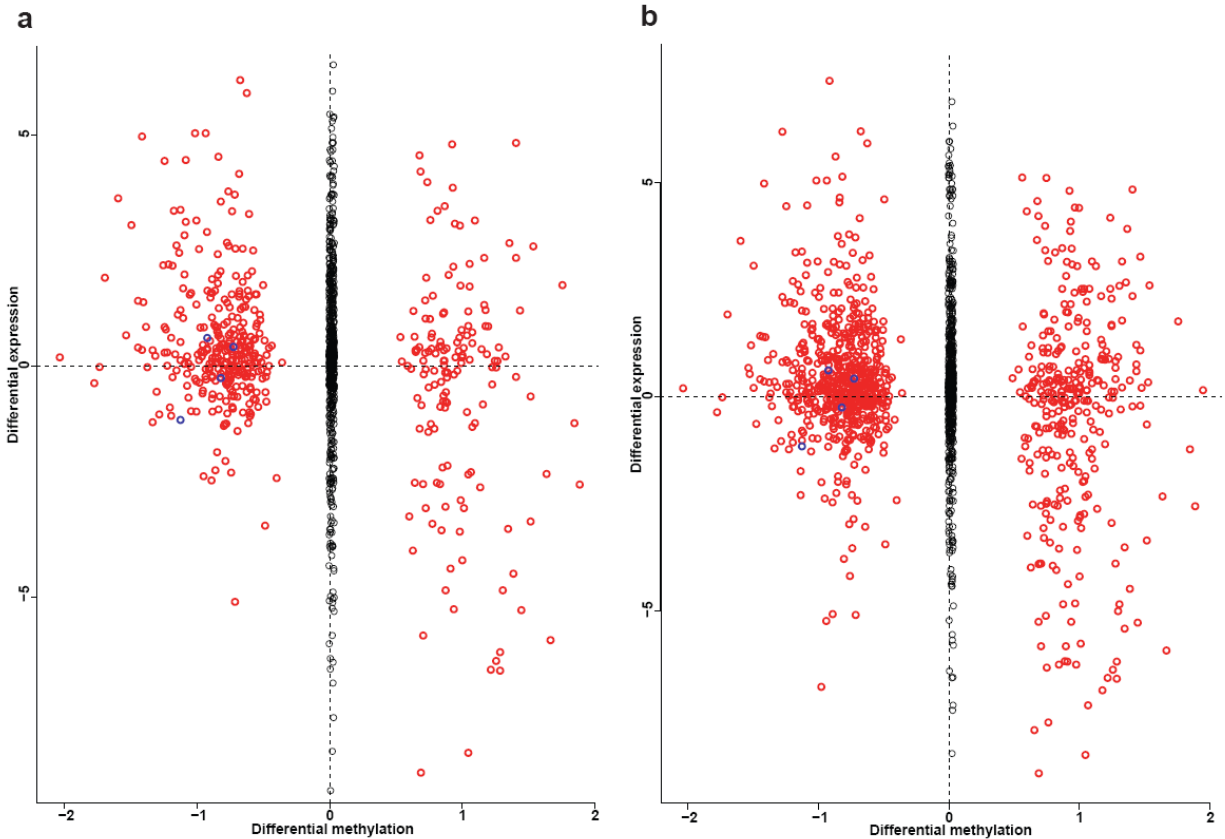
Supplementary Figure 1. Distribution of distance of reprogramming differentially methylated regions (R-DMRs) from CpG islands. Islands (teal) are regions that are inside, cover, or overlap more than 50% of a CpG island. Overlap (orange) are regions that overlap 0.1-50% of a CpG island. Regions denoted by (0, 500] (blue) are regions located ≤ 500 bp but do not overlap an island. Regions denoted by (500, 1000] (pink) are regions located > 500 bp and ≤ 1000 bp from an island. Regions denoted by (1000, 2000] (green) are regions located > 1000 bp and ≤ 2000 bp from an island. Regions denoted by (2000, 3000] (yellow) are regions located > 2000 bp and ≤ 3000 bp from an island. Regions denoted by > 3000 (brown) are > 3000 bp from an island. Percentage are given for the CpG regions (CHARM array, null hypothesis) and reprogramming differentially methylated regions (R-DMRs) as well as the R-DMRs subdivided into hypermethylation and hypomethylation in iPS relative to fibroblast. Percentages of each class is given for **(a)** R-DMRs from the first experiment ($n=6$ for each cell type) (R-DMR panel is duplicated from Fig. 1a) and **(b)** R-DMRs from second experiment ($n=3$ for each cell type).







Supplementary Figure 2. (a-g) Examples of differential DNA methylation (upper panel) and confirmation by bisulfite pyrosequencing (lower panel). The upper panel is a plot of M value versus genomic location, where the curve represents averaged smoothed M values. Also shown in the upper panel are the location of CpG dinucleotide (black tick marks on x axis), CpG density (smoothed black line) calculated across the region using a standard density estimator, location of CpG islands (orange line), as well as gene annotation indicating the transcript (thin outer gray line), coding region (thin inner gray line), exons (filled gray box) and gene transcription directionality on the y axis (sense marked as +, antisense as -). The lower panel represents the degree of DNA methylation as measured by bisulfite pyrosequencing. The red box indicated on the x axis of the CpG density plot in the upper panel indicates the CpG sites that were measured. Reactions were done in triplicate; bars represent the mean methylation \pm SD of iPS cells (pink), fibroblast (gray) and ES (blue) as well as DKO (*DNMT1* and *DNMT3B* Double KO cell line) and HCT116 (parental colon cancer cell line) for each individual CpG site measured. (a-b) are DMRs found by comparison between iPS cells and fibroblast (n=6), (c-g) is a DMR found by comparison between iPS cells and ES cells (n=3). (a) *TBX3* (T-box 3 protein), (b) *HOXD3* (Homeobox D3), (c) *POU3F4* (POU domain, class 3, transcription factor 4), (d) *A2BP1* (ataxin 2-binding protein 1), (e) *ZNF184* (zinc finger protein 184), (f) *IGF1R* (insulin-like growth factor 1 receptor), (g) *PTPRT* (protein tyrosine phosphatase, receptor type, T).



Supplementary Figure 3. Gene expression strongly correlates with reprogramming differentially methylated regions (R-DMRs) at CpG island shores. Red circles represent R-DMRs that are within 2kb from a CpG island, blue circles represent those that are more than 2kb away from a CpG island, and black circles represent log ratios for all genes not within (a) 500bp or (b) 1kb from the transcriptional start site (TSS) of an annotated gene. The \log_2 ratios of fibroblast to iPS expression were plotted against ΔM values (fibroblast minus iPS) for R-DMRs in which one of the two points had approximately no methylation. **(a)** DMRs that are within 500bp from a TSS of a gene. **(b)** DMRs that are within 1kb from a TSS of a gene.

Supplementary Table 3. Reprogramming differentially methylated regions (R-DMRs) overlap with bivalent domains. Overlap is significant in hypomethylated R-DMRs but not in hypermethylated R-DMRs.

Methylation	Total Number	Overlap with bivalent domain		
			Number	<i>P</i> Value*
R-DMRs	4401	Observed	1624 (36.9%)	<0.0001
		Random	833	
Hypermethylated R-DMRs (iPS>Fibroblast)	2663	Observed	495 (18.6%)	0.5699
		Random	498	
Hypomethylated R-DMRs (iPS<Fibroblast)	1738	Observed	1129 (65%)	<0.0001
		Random	335	

**P* values based on 10,000 permutations. Random values are averages over all 10,000 iterations.

Supplementary Table 4. Reprogramming differentially methylated regions (R-DMRs) overlap with *POU5F1*, *NANOG*, and *SOX2* binding sites. Overlap is significant in hypomethylated R-DMRs but not in hypermethylated R-DMRs.

Methylation	Total Number	Number of DMRs on promoter array*	Overlap with <i>POU5F1</i> -binding sites*		Overlap with <i>NANOG</i> -binding sites*		Overlap with <i>SOX2</i> -binding sites*		
			Number	<i>P</i> Value**	Number	<i>P</i> Value**	Number	<i>P</i> Value**	
R-DMRs	4401	1963	Observed	85	<0.0001	157	<0.0001	101	<0.0001
			Random	27		83		65	
Hypermethylated R-DMRs (iPS>Fibroblast)	2663	1202	Observed	2	1	20	1	9	1
			Random	16		49		38	
Hypomethylated R-DMRs (iPS<Fibroblast)	1738	761	Observed	83	<0.0001	137	<0.0001	92	<0.0001
			Random	11		34		27	

* Data from Boyer *et al. Cell.* **122**, 947-56 (2005).

***P* values based on 10,000 permutations. Random values are averages over all 10,000 iterations.

Supplementary Table 7. Gene ontology functional categories enriched in differentially methylated regions between iPS cells and ES cells.

Methylation	GO ID	Term	Count	%	Fold Enrichment	<i>P</i> Value
Hypermethylated in iPS compared to ES	GO:0007275	multicellular organismal development	14	31.82%	2.69251	6.81E-04
	GO:0032501	multicellular organismal process	18	40.91%	2.163279	7.23E-04
	GO:0048731	system development	11	25.00%	2.823529	0.00294
	GO:0065007	biological regulation	20	45.45%	1.724951	0.00477
	GO:0007399	nervous system development	7	15.91%	4.054299	0.00575
	GO:0045944	positive regulation of transcription from RNA polymerase II promoter	4	9.09%	10.5676	0.00579
Hypomethylated in iPS compared to ES	GO:0050794	regulation of cellular process	12	70.59%	2.778833	1.63E-04
	GO:0050789	regulation of biological process	12	70.59%	2.582055	3.36E-04
	GO:0065007	biological regulation	12	70.59%	2.345934	8.57E-04
	GO:0006355	regulation of transcription, DNA-dependent	8	47.06%	3.368421	0.00303
	GO:0006351	transcription, DNA-dependent	8	47.06%	3.293928	0.00345
	GO:0032774	RNA biosynthetic process	8	47.06%	3.28996	0.00348
	GO:0045449	regulation of transcription	8	47.06%	3.153195	0.00446
	GO:0019219	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	8	47.06%	3.082017	0.00509
	GO:0006350	transcription	8	47.06%	3.039703	0.00551
	GO:0043283	biopolymer metabolic process	11	64.71%	2.101101	0.00588
	GO:0010468	regulation of gene expression	8	47.06%	2.968116	0.00632
	GO:0010467	gene expression	9	52.94%	2.500271	0.00792
	GO:0031323	regulation of cellular metabolic process	8	47.06%	2.851375	0.00795
	GO:0019222	regulation of metabolic process	8	47.06%	2.753613	0.00968

Supplementary Table 9. Reprogramming differentially methylated regions (R-DMRs), and the DMRs that distinguish iPS cells from ES cells, overlap with cancer-specific differentially methylated regions (C-DMRs).

DMR	Total Number	Methylation	Overlap with					
			Hypermethylated C-DMRs (1508)			Hypomethylated C-DMRs (1199)		
			Number	Fold Enrichment	<i>P</i> Value	Number	Fold Enrichment	<i>P</i> Value
iPS-Fibroblast DMRs (R-DMRs)* (n=6)	4401	iPS>Fib (2663)	99	1.54	<0.0001	293	5.49	<0.0001
		Random	64			53		
		Fib>iPS (1738)	294	6.53	<0.0001	33	0.91	0.74
		Random	45			36		
iPS-ES DMRs** (n=3)	71	iPS>ES (51)	10	7.93	<0.0001	2	1.93	0.28
		Random	1.3			1.0		
		ES>iPS (20)	4	7.86	0.0015	1	2.48	0.33
		Random	0.51			0.40		
iPS-Fibroblast DMRs (R-DMRs)** (n=3)	2179	iPS>Fib (988)	56	2.72	<0.0001	83	4.42	<0.0001
		Random	21			18		
		Fib>iPS (1191)	230	8.84	<0.0001	20	0.86	0.78
		Random	26			23		

* FDR cutoff of 0.05.

**Absolute area cutoff of 10.0.

P values based on 10,000 permutations. Random values are averages over all 10,000 iterations.

Supplementary Table 10. Primer sequences and location of CpG sites examined using bisulfite pyrosequencing.

Gene	Primer	Sequence (5' → 3')	Chromosomal Coordinates						
			Chr	CG1	CG2	CG3	CG4	CG5	CG6
<i>BMP7</i>	Forward	TTTGGTTTGAAATGTATTAATATA							
	Reverse	TAACAATACCAAAAAATACTAAAACACTACTA							
	Nested forward	TTTTGGTTTTAAAATAATAAAGTAATTATT	20	55268092	55268094	55268113	55268115	55268135	
	Nested reverse	/5Biosg/AACTCAAACAAACATATACAATACC							
	Sequencing (F)	GTTGTTATTAATTTAATTTATT							
<i>GSC</i>	Forward	GATTTAAGTTATTATGTTTTAGGGTAGATA							
	Reverse	AAAACAATATTCCAAATAAAAAAAA							
	Nested forward	TTAGGTTTAAAGTTATAGGGTAGTTGATG	14	94308388	94308412	94308417	94308422	94308424	
	Nested reverse	/5Biosg/TTAACATCTTTACAAAAACAAAAC							
	Sequencing (F)	GTAATTTATTAGTGATTGTTT							
<i>PTPRT</i>	Forward	TTAAAGAGTAAATAAAGAAAAGGTGTT							
	Reverse	AATCCTAAAAATCCAAACATAATTC							
	Nested forward	TGAAAGTAATTAGATTTGTATTTTAATAGT	20	41250110	41250112	41250118	41250145		
	Nested reverse	/5Biosg/AATTTTATATCCTCTAAAACATAACC							
	Sequencing (F)	GATGGAATATTTTTGATTTTGT							
<i>TBX3</i>	Forward	TTAGGATTTAGGGTTTTGTTTTTT							
	Reverse	TATCATCTTCCTAAATATTTACAAAATATT							
	Nested forward	GTGGGTAGGAAGAAGTTTTAAGGTT	12	113619701	113619716	113619732	113619736	113619746	113619751
	Nested reverse	/5Biosg/AACTCATTCTCAAATAAAAAACCC							
	Sequencing (F)	TTATTAGAGTTTTTTAGTAGATT							
<i>HOXD3</i>	Forward	GTAGATTGGTTTTTTTGTATTTTTG							
	Reverse	TATAAACTCTCAAATTTCTTTTAATATCT							
	Nested forward	GATTTATTTGGTTAGAGGGTTTGG	2	176736800	176736811	176736815	176736852		
	Nested reverse	/5Biosg/AAAAAACTTTTCCCACTTAAAAAAC							
	Sequencing (F)	GATTTATTTGGTTAGAGGGTTTGG							
<i>POU3F4</i>	Forward	AAGGTTATAGGGATTTTGGTTTATT							
	Reverse	CCACAACAACACTACATATTTTTAAAA							
	Nested forward	ATTTTTGTGTGTATGTGTTTTTGTG	X	82652791	82652795	82652797	82652805	82652832	
	Nested reverse	/5Biosg/CTCTACACAACCTAACCAATTTTT							
	Sequencing (F)	ATTTTTGTGTGTATGTGTTTTTGTG							
<i>A2BP1</i>	Forward	TTTTTGATAAATTGATGGGATGTG							
	Reverse	AACCCTAAAACCTAACCAAAAAAC							
	Nested forward	TAAGATGAAAAGTGAAAGAAATAG	16	7292289	7292308				
	Nested reverse	/5Biosg/ATAAAAACTCTAAACCCAACCATCA							

	Sequencing (F)	GAAGATTTTATAGTTATTTTAAATAG					
<i>ZNF184</i>	Forward	AAAAGAAAATTTTAAAGTTATAAAATT					
	Reverse	AAATCAAAATCCATATCTCATTTAATCTAA					
	Nested forward	TTGGGAGAGTTTTAAAGTTATTTGG	6	27634436	27634468		
	Nested reverse	/5Biosg/TAActCCAATCCAAAATTTTCTCTC					
	Sequencing (F)	TGGGAGAGTTTTAAAGTTATTTGGA					
<i>IGF1R</i>	Forward	GTGGTTTGGGAAGATATGAATTTT					
	Reverse	AAAATAAAAACCCCTTTTCTTAC					
	Nested forward	AAGGTTTTTTATTGTTTTTGATTA	15	97226473	97226475	97226481	97226497
	Nested reverse	/5Biosg/AAAATCCTAAACCTCCACTTC					
	Sequencing (F)	AGGTTTTTTATTGTTTTTGATTA					

/5Biosg/ = 5' biotin added

F = forward