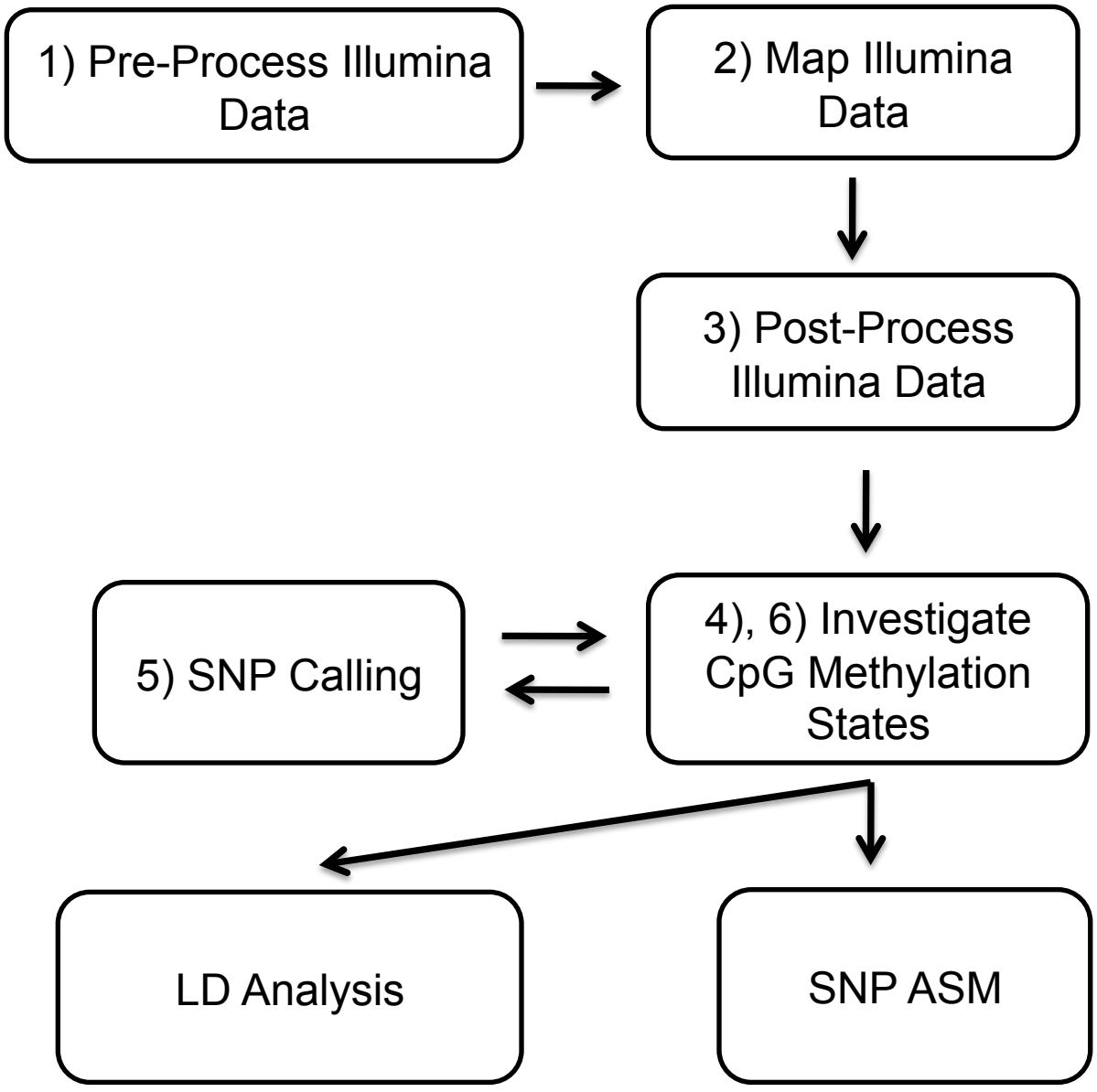
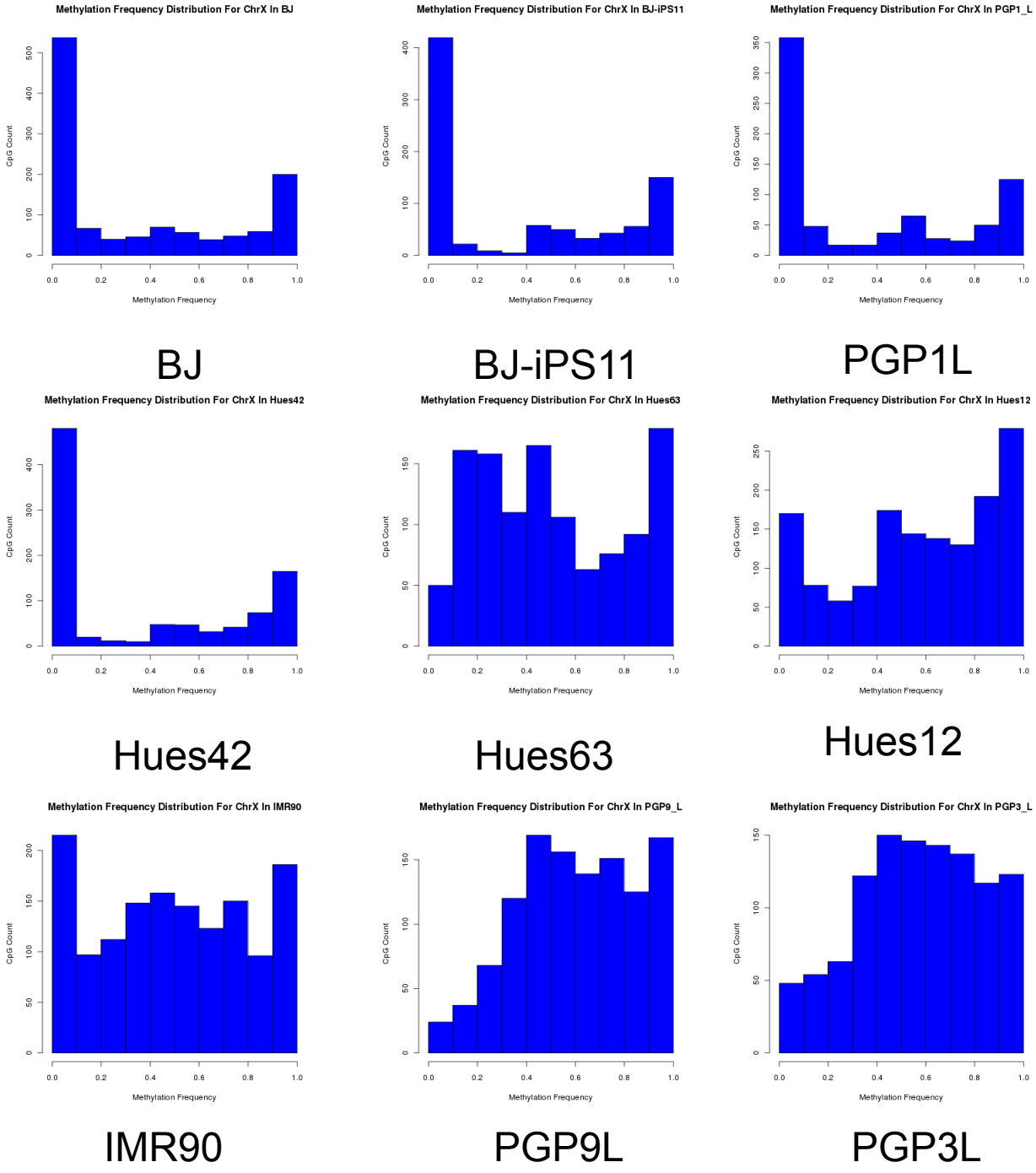
Supplementary Figure 1. The workflow of methylation haplotype analysis.



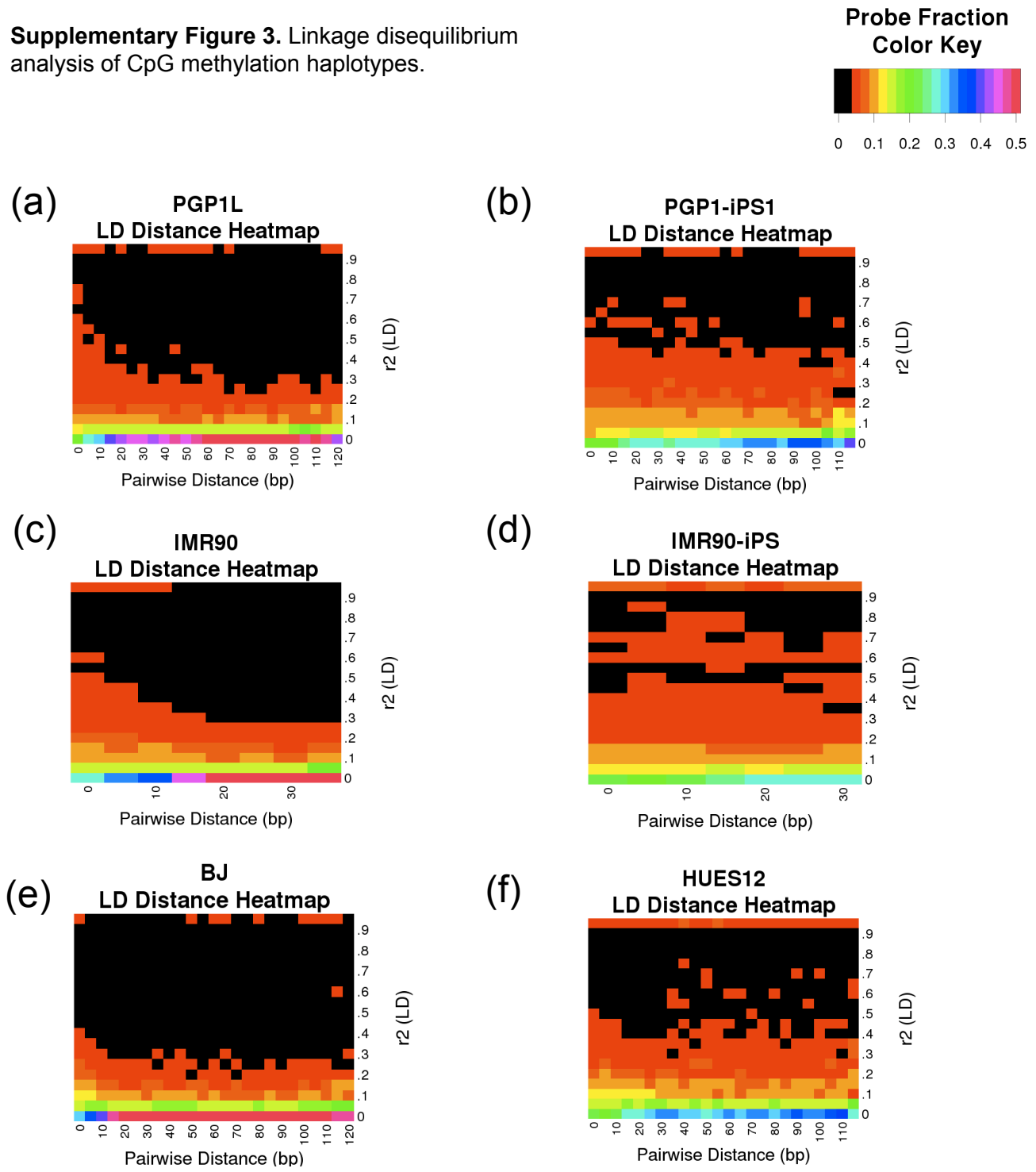
Supplementary Figure 1. The workflow of data analysis contains the following steps: 1) Pre-processing of Illumina data involved detecting whether the read was a PCR product or bisulfite converted sequence and taking the reverse complement of PCR product sequences. Then the read sequences were demethylated *in silico*. 2) Reads were mapped to the unmethylated bilsulfite converted hg18 reference genome. 3) Reads that did not uniquely map to the reference are filtered out. Paired end reads that mapped too closely together or too far away were treated as single end reads. 4) The read coordinates of the mapped *in silico* demethylated reads were assigned to the read's original sequence. The methylated state of any known CpGs covered by the reads were investigated. 5) SNPs were called based on the mapped read data using our own algorithm and SAM Tools. The intersection of SNP calls from these two algorithms was used in further anlayses. 6) The methylation states of new CpGs created by SNPs were investigated. 7) The methylation information was used for LD and SNP ASM analyses.

Supplementary Figure 2. Methylation frequency histograms for CpG dinucleotides located in chromosome X.



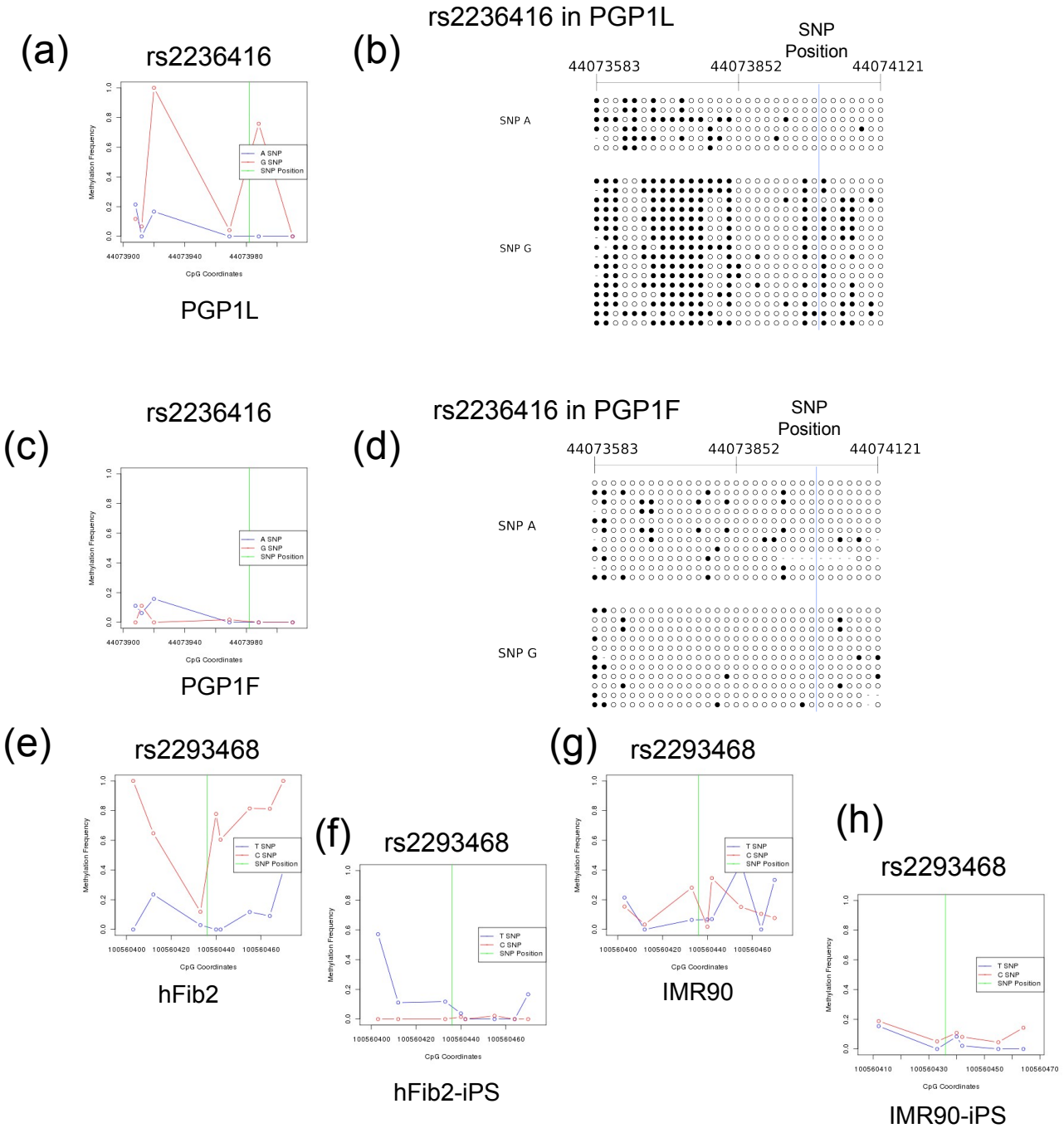
Supplementary Figure 2. Methylation frequency histograms for CpG dinucleotides located in chromosome X. The X-axis represents the methylation frequency bins and the y-axis represents the number of CpG dinucleotides within a certain methylation frequency bin. A methylation frequency value of one represents complete methylation at a CpG. Hues63, a male cell line, has a methylation frequency distribution in chromosome X that visibly contains more fuzzily methylated probes (.25 < methylation frequency <.75) than other male cell lines. Fuzzily methylated probes in chromosome X are indicative of female cell lines (Hues12, IMR90, PGP3L, PGP9L) rather than male cell lines (BJ family, Hues42, PGP1L), which is due to allele specific methylation.

Supplementary Figure 3. Linkage disequilibrium analysis of CpG methylation haplotypes.



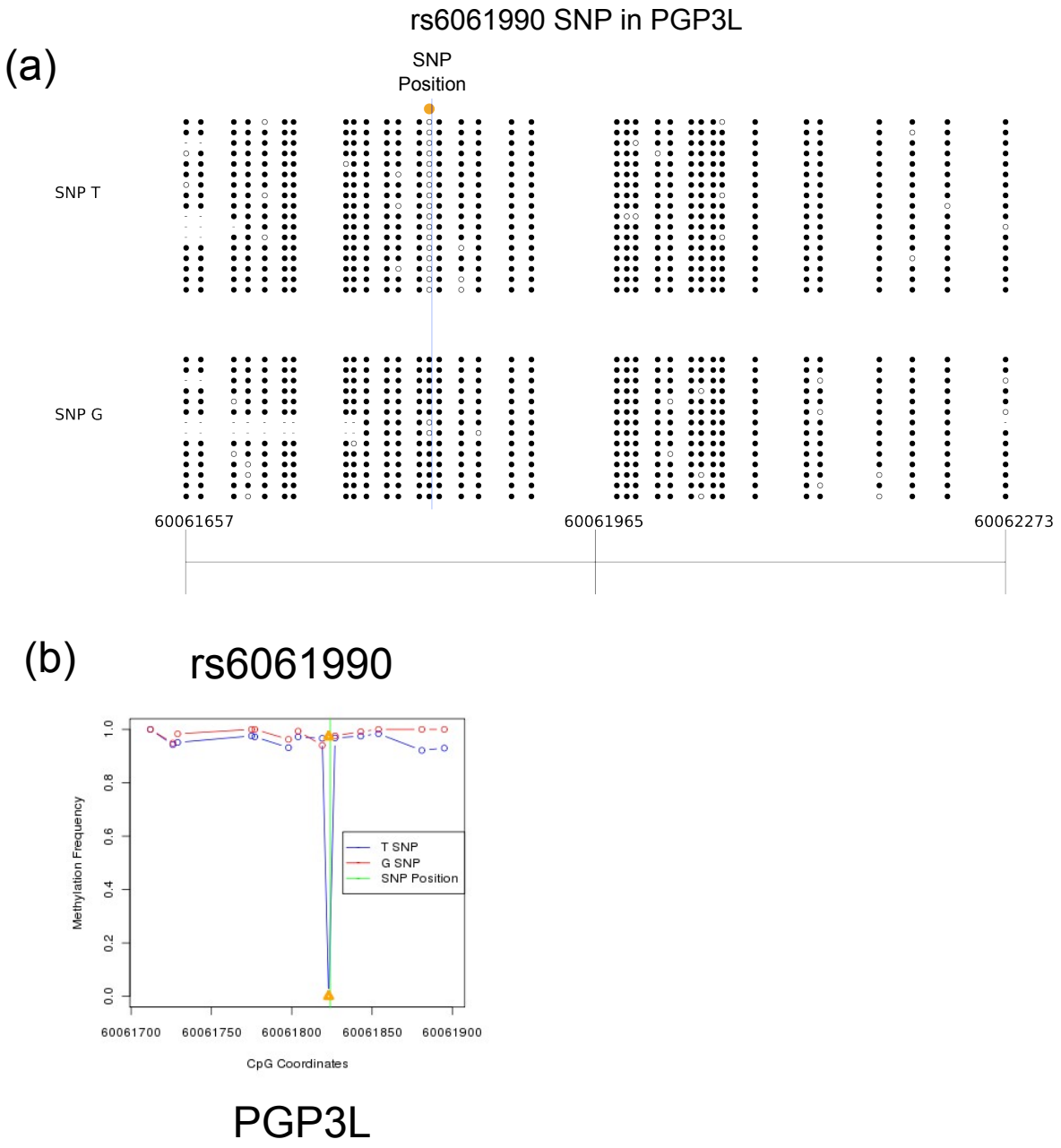
Supplementary Figure 4. Linkage disequilibrium analysis of CpG methylation haplotypes. (a), (c), and (e) are from differentiated cells and the heatmap shows that the r^2 value distribution shifts downwards as pairwise distance increases. (b), (d), and (f) represent undifferentiated cells and the distribution of r^2 values is maintained as pairwise distance increases. These figures show that undifferentiated cells contain organized methylation patterns that tend to span farther distances than in differentiated cells.

Supplementary Figure 4. Cell type specificity and individual dependent of ASM.



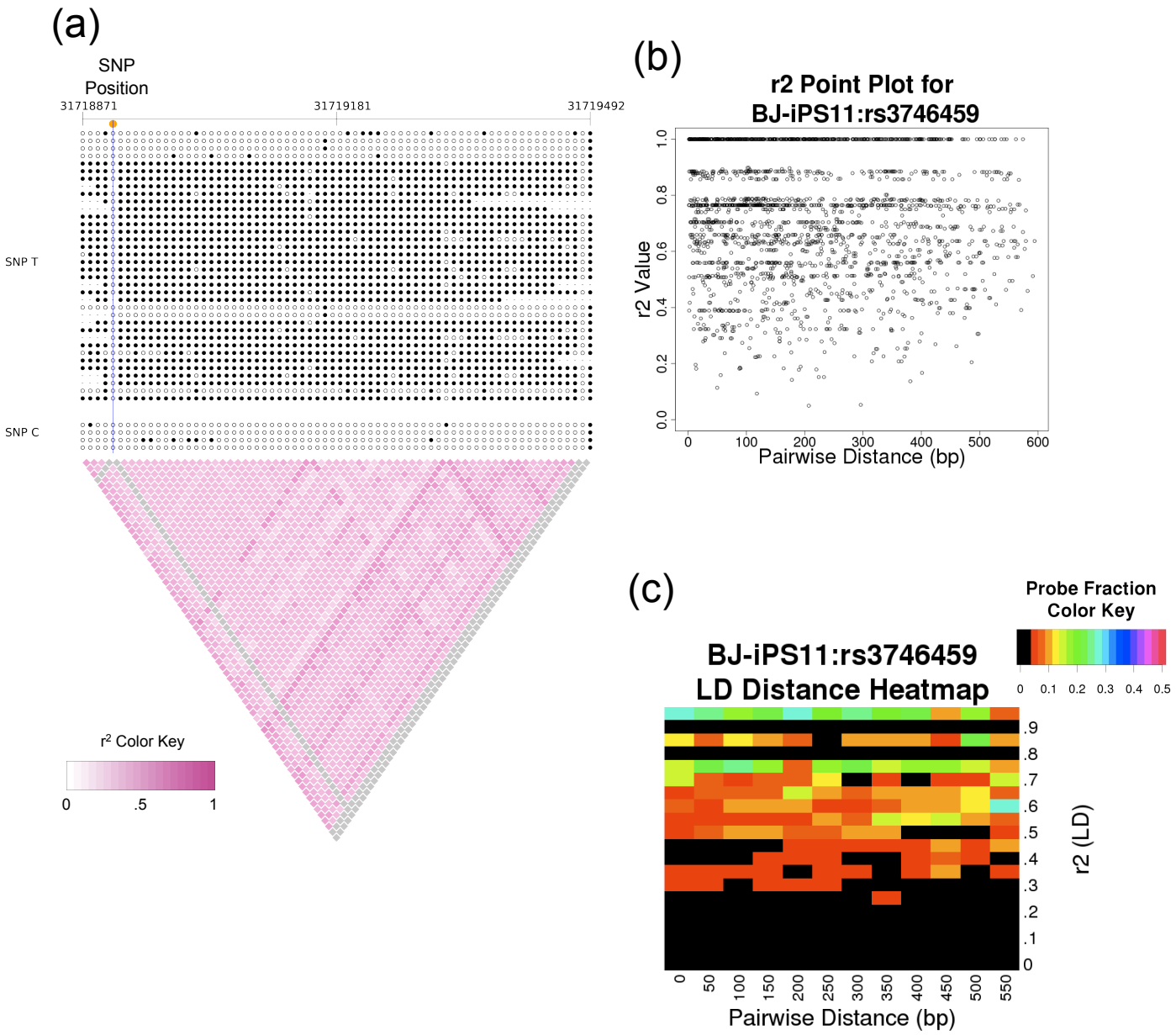
Supplementary Figure 6. Cell type specificity and individual dependence of ASM. (a)-(d) Illumina and Sanger sequences from PGP1L and PGP1F showing cell type specificity of ASM in rs2236416 indexed region in the intron of MMP9 gene (chr20:44073908-44074010 in Illumina and chr20:44073583-44074121 in Sanger data). ASM is seen in PGP1L near rs2236416 but not seen in PGP1F. (e)-(h) Illumina reads from hFib2, IMR90, hFib2-iPS (chr12:100560403-100560470) and IMR90-iPS (chr12:100560412-100560464) show the cell type specificity of ASM in rs2293468 indexed region. ASM around SNP rs2293468 is seen in hFib2, hFib2-iPS, and IMR90 but not IMR90-iPS.

Supplementary Figure 5. ASM example in PGP3L.



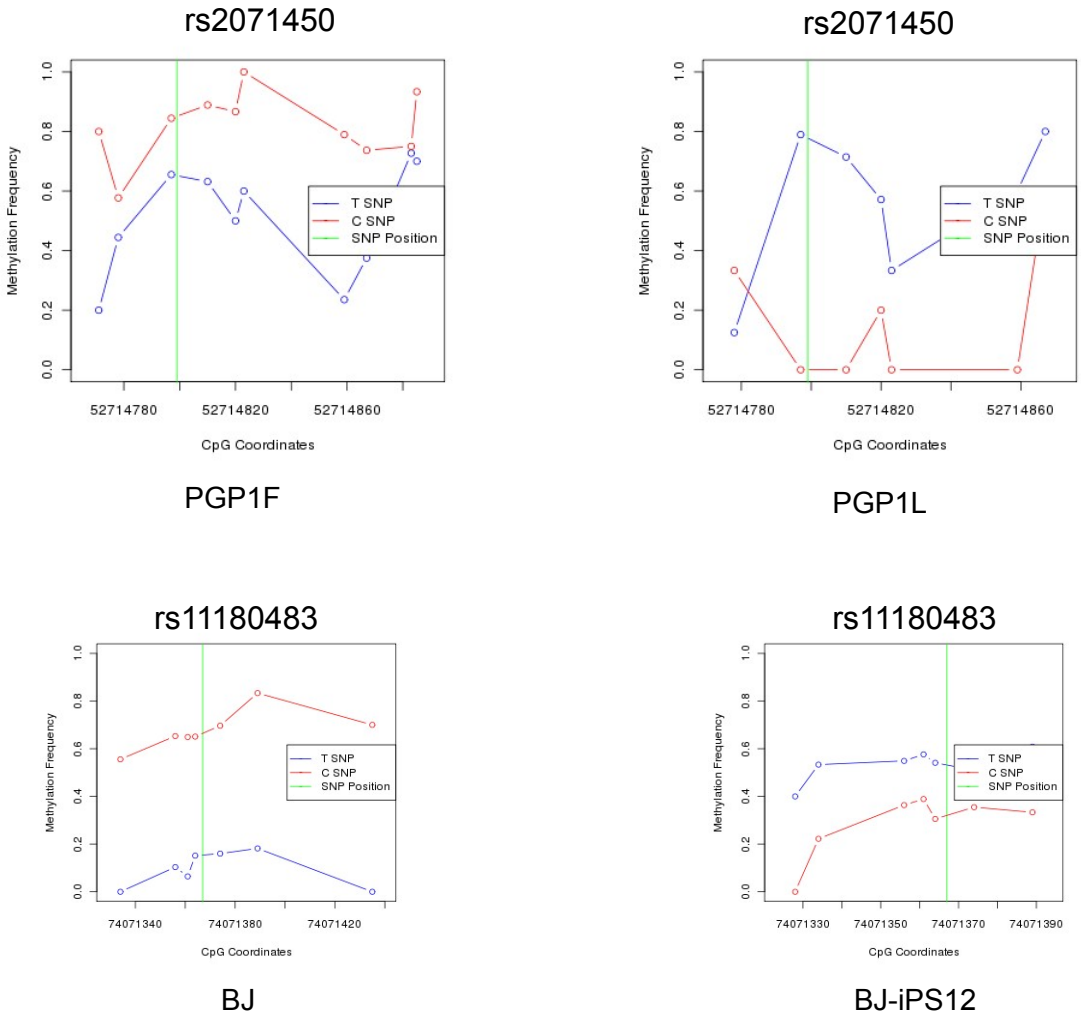
Supplementary Figure 5. ASM example in PGP3L. (a) Sanger reads from PGP3L show ASM at a CpG dinucleotide that overlaps with SNP rs6061990 (chr20:60061657-60062273). The other CpG dinucleotides do not show ASM. (b) Methylation frequency diagram of the SNP rs6061990 region produced by Illumina data (chr20:60061712-60061895). The diagram supports the ASM behavior observed in the Sanger sequences. The orange triangles represent a SNP site that overlaps with a CpG. There is a difference in methylation frequency at coordinate 60061798 but it is not considered significant due to low read depth; only 6 mapped Illumina reads of the G allele cover this CpG dinucleotide.

Supplementary Figure 6. Linkage disequilibrium analysis of CpG methylation haplotypes in SNP indexed region rs3746459 (NECAB3 Intron) in BJ-iPS11.



Supplementary Figure 3. Linkage disequilibrium analysis of CpG methylation haplotypes around rs3746459 in BJ-iPS12 (chr20:31718871-31719492). (a) Sanger shows sequences that are either mostly methylated or unmethylated. In this case, the T SNP allele destroys the CpG dinucleotide and CpGs at this location on the T allele are thusly unmethylated. The methylation organization shows characteristics of ASM but the SNP identity at rs3746459 does not clearly separate the methylated and unmethylated sequences. (b) and (c) CpG pairwise r2 value plots and heatmap show that the r^2 values in this region are significant and maintain significance as the relative distance between the two CpG dinucleotides increases.

Supplementary Figure 7. Examples of ASM flipping.



Supplementary Figure 7. Examples of ASM flipping. (a) and (b) Illumina reads from PGP1F and PGP1L show ASM in SNP rs2071450 region (chr12:52714771-52714885). The G allele is predominantly more methylated than A allele in PGP1F whereas A allele is more methylated than G allele in PGP1L. (c) and (d) Illumina reads from BJ (chr12:74071334-74071435) and BJ-iPS12 (chr12:74071328-74071389) show ASM around SNP rs11180483 indexed region. C allele is predominantly more methylated than T allele in BJ whereas T allele is more methylated than C allele in BJ-iPS12.

Supplementary Table 1. SNPs and ASM by Sanger sequencing.

Cell Line	SNP 129 ID	Associated Ref Seq Genes	Sequenced Strand	Minimum Allele Read Depth	Sanger ASM Class	Illumina ASM Class
BJ	rs3746459	NECAB3:intron	-	1	III	III
BJ	rs7266947		+	7	I	I
BJ-iPS11	rs10877897		+	46	I	I
BJ-iPS11	rs1061726		+	6	I	III
BJ-iPS11	rs3746459	NECAB3:intron	-	4	I	I
BJ-iPS12	rs10846023	FLJ22662:intron	-	9	I	I
hFib2	rs2277324	SLC26A10:5 prime	+	0	NA	I
Hues42	rs2277324	SLC26A10:5 prime	+	2	III	III
Hues63	rs220030	SNRPN:intron	-	0	NA	III
IMR90-iPS	rs2018002	DNMT3B:intron	+	2	III	III
IMR90-iPS	rs2089908	LSP1:5 prime	+	8	I	II
PGP1F	rs2236416	MMP9:intron	+	11	III	III
PGP1F	rs2089908	LSP1:5 prime	+	7	III	II
PGP1F	rs2072788	MATN4:exon	-	12	III	I
PGP1L	rs2236416	MMP9:intron	+	6	I	I
PGP1L	rs2072788	MATN4:exon	-	8	I	I
PGP1-iPS1	rs2072788	MATN4:exon	-	11	III	I
PGP3L	rs6061990	TAF4:intron	+	14	II	II
PGP3L	rs2018002	DNMT3B:intron	+	1	III	I
PGP9L	rs6061990	TAF4:intron	+	13	II	I
PGP9L	rs2018002	DNMT3B:intron	+	4	III	III

Supplementary Table 1. SNPs and ASM by Sanger sequencing. ASM statistics of the Sanger sequences compared to the Illumina read regions. Minimum allele read depth shows the minimum number of aligned Sanger sequences for either allele at each SNP region. The ASM categories were determined using the same criteria as for the Illumina sequences. Two called heterozygous SNPs in Hues63 and hFib2 from the Illumina data were not found in Sanger sequences. Illumina and Sanger ASM categories were considered consistent if both sequencing methods called ASM (i.e. category I or II). There were 7 cases where the ASM calls were not consistent between the Illumina and Sanger data. Rs2072788 in PGP1F and PGP1-iPS1 was a SNP that destroyed the cytosine of a CpG dinucleotide in one of the alleles in the Sanger sequences, which made this region ineligible for a category II ASM label. These two SNP regions were considered category I ASM in the Illumina data because of a neighboring CpG site that showed ASM. This corresponding CpG site in the Sanger data did not have sufficient read depth for a statistical call of ASM. Rs20889908 in PGP1F and rs2018002 in PGP3L also had insufficient Sanger read depth for ASM categorization. The discrepancy involving rs1061726 in BJ-iPS11 cannot be explained by read depth or single stranded coverage.

Supplementary table 2. Primer sequences.

Index SNP	Forward primer	Reverse primer	Amplicon size (bp)
rs1061726	GTAGAAATTTGGAAGTGGAAATTT	TAATCAATAATTTTTCCAAAAAAAA	645
rs10846023	TTTAGGGGTTGTTAGAGGGTTAGA	AAATTTTTAAAACCAACCCAAACTC	674
rs2018002	GTTTTTGTTTTGGGAAAAGTTAAG	CAAAAAACAACTCAAAATTCATACT	773
rs2072788	GTGAGGTTTTGTTGATTTAGGAGAG	ATCCAAACATTTAAATTAAAAATTC	671
rs2089908	AAGTTTTTGTTGGTTGGATTTTTTA	AAAAAAACCCATATTACCCCTAATC	746
rs2236416	TGGGTTAAAGAATAGGATATATTTGG	AAAAAACCCAAAACCTTAAATAAAC	643
rs3746459	GAAGTTAGGAAATAGTGTGGAGT	AATATACCCAAAACAATAACCC	640
rs6061990	AGGTTTGGGTTATTTTATTTGTTTG	ACTTTCCCAACTCTCAAAACTCTAC	753
rs2277324	GGTTAAGGATGTTTGTAGAAA	ATTAAAACCTCTCACCCTAAAACAC	557
rs220030	TAGGTTGTTTTTTGAGAGAAGTTAT	CTTTTAAAAAAATTTCAAATCTAAC	559
rs10877897	GAGATGATGGTTTGGATTTTTTAG	AAAATCTTTAACCACTACCTACCC	615

Supplementary Table 3. Summary of Illumina Sequencing Data.

Cell Line	Total Reads	Mapped Reads	% Mapped	Single End, Paired End, or Mixture.
BJ	18,516,100	9,319,863	50.33%	Mixture
BJ-iPS11	12,276,017	5,235,162	42.65%	Single
BJ-iPS12	10,539,126	4,587,661	43.53%	Single
hFib2	14,408,007	5,780,509	40.12%	Single
hFib2-iPS	14,805,000	5,652,624	38.18%	Single
Hues12	18,764,749	8,849,367	47.16%	Mixture
Hues42	11,185,368	6,378,978	57.03%	Single
Hues63	14,315,301	5,811,519	40.60%	Single
Hybrid1	16,522,597	9,093,421	55.04%	Mixture
IMR90	24,687,802	10,493,117	42.50%	Single
IMR90-iPS	10,985,469	5,445,173	49.57%	Single
PGP1F	8,441,376	5,021,880	59.49%	Paired
PGP1-iPS1	5,617,832	3,611,457	64.29%	Paired
PGP1L	5,258,188	3,704,072	70.44%	Paired
PGP3L	5,286,442	3,481,123	65.85%	Paired
PGP9L	5,689,552	4,121,764	72.44%	Paired

Supplementary Table 3. Summary of Illumina sequencing data. Single end data is taken from the previously published data set and the paired end data is from new experiments. Cell lines labeled as a mixture of reads contain single end and paired end Illumina sequences, which were merged together before alignment.

Supplementary Table 4. Genes with Conserved LD Blocks across 12 or more Cell Lines.

Gene	Cell Lines With LD Block	Cell Lines
FRG1B	16	P1F,P1iPS11,IMR90,IMR90iPS,H12,BJ,P9L,H63,H42,Hy,BJiPS12,P3L,hFib2,P1L,BJiPS11,hFiPS
HM13	16	P1F,P1iPS11,IMR90,IMR90iPS,H12,BJ,P9L,H63,H42,Hy,BJiPS12,P3L,hFib2,P1L,BJiPS11,hFiPS
NR_003579	16	P1F,P1iPS11,IMR90,IMR90iPS,H12,BJ,P9L,H63,H42,Hy,BJiPS12,P3L,hFib2,P1L,BJiPS11,hFiPS
GNAS	16	P1F,P1iPS11,IMR90,IMR90iPS,H12,BJ,P9L,H63,H42,Hy,BJiPS12,P3L,hFib2,P1L,BJiPS11,hFiPS
SNRPN	16	P1F,P1iPS11,IMR90,IMR90iPS,H12,BJ,P9L,H63,H42,Hy,BJiPS12,P3L,hFib2,P1L,BJiPS11,hFiPS
KCNS1	14	P1F,P1iPS11,IMR90iPS,H12,P9L,H63,H42,Hy,BJiPS12,P3L,hFib2,P1L,BJiPS11,hFiPS
KCNK15	13	P1F,P1iPS11,IMR90,IMR90iPS,H12,BJ,P9L,Hy,BJiPS12,P3L,hFib2,hFiPS,P1L
NR_003531	13	P1F,P1iPS11,IMR90,H12,BJ,P9L,H63,H42,Hy,P3L,hFib2,hFiPS,P1L
LOC100134868	13	P1F,P1iPS11,IMR90,IMR90iPS,H12,BJ,P9L,H63,H42,Hy,BJiPS12,P3L,BJiPS11
MEG3	13	P1F,P1iPS11,IMR90,H12,BJ,P9L,H63,H42,Hy,P3L,hFib2,hFiPS,P1L
NNAT	13	P1F,IMR90,IMR90iPS,H12,BJ,P9L,H63,H42,Hy,BJiPS12,P3L,hFib2,hFiPS
NR_004846	13	P1F,P1iPS11,IMR90,IMR90iPS,H12,BJ,P9L,H63,H42,Hy,BJiPS12,P3L,BJiPS11
JPH2	12	P1F,P1iPS11,IMR90,H12,BJ,P9L,H42,Hy,P3L,hFib2,hFiPS,P1L
C20orf96	12	P1F,P1iPS11,H12,BJ,P9L,H63,H42,BJiPS12,P3L,hFib2,BJiPS11,P1L
TXNRD1	12	P1F,P1iPS11,IMR90,IMR90iPS,H12,BJ,P9L,H63,BJiPS12,P3L,BJiPS11,P1L
THBD	12	P1F,P1iPS11,IMR90,IMR90iPS,H12,P9L,H63,Hy,hFib2,BJiPS11,hFiPS,P1L
KRT86	12	P1F,P1iPS11,IMR90iPS,H12,BJ,P9L,Hy,BJiPS12,hFib2,BJiPS11,hFiPS,P1L
KCNQ2	12	P1F,P1iPS11,IMR90,BJ,P9L,Hy,BJiPS12,P3L,hFib2,BJiPS11,hFiPS,P1L
CACNA2D4	12	P1iPS11,IMR90,IMR90iPS,H12,P9L,H63,H42,Hy,P3L,hFib2,hFiPS,P1L
TBX3	12	P1F,P1iPS11,H12,P9L,H63,H42,Hy,BJiPS12,P3L,hFib2,BJiPS11,P1L

Supplementary Table 4. Genes with Conserved LD Blocks across 12 or more cell lines. These genes contain CpG dinucleotides that are strong candidates for biological regulation via methylation. Cell line names are abbreviated so that they can be printed on a single line.

Supplementary Table 5. SNP calling statistics.

Cell Line	Total Candidate SNP Sites Examined	Called Heterozygous SNPs	Heterogzyous SNP Call Percentage	Called Homozygous SNPs	Homozygous SNP Call Percentage
BJ	1,525,138	457	0.030%	391	0.026%
BJ-iPS11	1,123,678	381	0.034%	260	0.023%
BJ-iPS12	1,192,417	402	0.034%	273	0.023%
hFib2	1,350,920	279	0.021%	307	0.023%
hFib2-iPS	1,144,060	283	0.025%	233	0.020%
Hues12	1,348,641	391	0.029%	344	0.026%
Hues42	1,128,168	308	0.027%	283	0.025%
Hues63	1,178,908	382	0.032%	197	0.017%
Hybrid1	1,390,831	395	0.028%	199	0.014%
IMR90	1,434,099	436	0.030%	334	0.023%
IMR90-iPS	1,139,705	430	0.038%	260	0.023%
PGP1F	1,020,601	292	0.029%	253	0.025%
PGP1-iPS1	835,010	240	0.029%	204	0.024%
PGP1L	913,151	257	0.028%	227	0.025%
PGP3L	951,308	254	0.027%	240	0.025%
PGP9L	982,170	272	0.028%	236	0.024%

Supplementary Table 5. SNP calling statistics. Candidate sites include all chromosomal locations with sufficient read coverage (minimum 10x coverage on analyzed strands) and sequence quality. Double stranded sites were checked for reverse complementarity. SNP calls were filtered so that only SNP calls made at rs129 sites that matched the reported SNP bases were recorded.

Supplementary Table 6. Pairwise common heterozygous SNPs between cell lines.

Cell Line	BJ	BJ-iPS11	BJ-iPS12	hFib2	hFib2-iPS	Hues12	Hues42	Hues63	Hybrid1	IMR90	IMR90-iPS	PGP1F	PGP1-iPS1	PGP1L	PGP3L	PGP9L
BJ	502	369	378	114	111	152	102	144	219	149	150	113	94	97	104	120
BJ-iPS11	369	421	366	115	116	140	95	136	229	128	137	99	84	85	97	117
BJ-iPS12	378	366	450	111	117	152	104	138	230	141	154	100	95	89	96	124
hFib2	114	115	111	305	241	116	81	98	127	121	122	89	74	70	85	87
hFib2-iPS	111	116	117	241	296	119	79	94	126	120	119	81	73	68	80	82
Hues12	152	140	152	116	119	427	110	119	137	138	123	105	90	96	99	98
Hues42	102	95	104	81	79	110	341	101	116	100	99	97	85	84	79	93
Hues63	144	136	138	98	94	119	101	417	129	134	134	107	86	94	92	85
Hybrid1	219	229	230	127	126	137	116	129	412	135	147	120	104	109	120	125
IMR90	149	128	141	121	120	138	100	134	135	490	360	111	101	102	101	101
IMR90-iPS	150	137	154	122	119	123	99	134	147	360	493	109	97	97	100	105
PGP1F	113	99	100	89	81	105	97	107	120	111	109	322	222	231	85	82
PGP1-iPS1	94	84	95	74	73	90	85	86	104	101	97	222	266	223	71	73
PGP1L	97	85	89	70	68	96	84	94	109	102	97	231	223	290	79	80
PGP3L	104	97	96	85	80	99	79	92	120	101	100	85	71	79	272	89
PGP9L	120	117	124	87	82	98	93	85	125	101	105	82	73	80	89	288

Supplementary Table 5. Pairwise common heterozygous SNPs between cell lines. Each matrix entry represents the number of heterozygous SNPs called in the row i cell line that match with those called in the column j cell line.

Supplementary Table 7. Pairwise common homozygous SNPs between cell lines.

Cell Line	BJ	BJ-iPS11	BJ-iPS12	hFib2	hFib2-iPS	Hues12	Hues42	Hues63	Hybrid_1	IMR90	IMR90-iPS	PGP1F	PGP1-iPS1	PGP1L	PGP3L	PGP9L
BJ	391	248	259	121	119	159	122	137	165	155	141	143	121	129	130	132
BJ-iPS11	248	260	229	106	107	129	105	118	140	114	123	124	103	108	114	114
BJ-iPS12	259	229	273	104	105	140	112	118	137	120	125	120	104	115	114	117
hFib2	121	106	104	307	170	127	114	95	101	114	92	120	108	110	125	115
hFib2-iPS	119	107	105	170	233	119	111	87	97	105	98	114	104	105	115	111
Hues12	159	129	140	127	119	344	155	111	102	137	129	145	122	139	140	143
Hues42	122	105	112	114	111	155	283	101	93	121	111	147	121	128	119	123
Hues63	137	118	118	95	87	111	101	197	106	117	116	108	88	92	104	99
Hybrid1	165	140	137	101	97	102	93	106	199	116	109	101	94	87	94	93
IMR90	155	114	120	114	105	137	121	117	116	334	208	128	117	113	116	129
IMR90-iPS	141	123	125	92	98	129	111	116	109	208	260	118	104	105	100	118
PGP1F	143	124	120	120	114	145	147	108	101	128	118	253	183	186	121	129
PGP1-iPS1	121	103	104	108	104	122	121	88	94	117	104	183	204	173	113	114
PGP1L	129	108	115	110	105	139	128	92	87	113	105	186	173	227	114	114
PGP3L	130	114	114	125	115	140	119	104	94	116	100	121	113	114	240	127
PGP9L	132	114	117	115	111	143	123	99	93	129	118	129	114	114	127	236

Supplementary Table 6. Pairwise common homozygous SNPs between cell lines. Each matrix entry represents the number of homozygous SNPs called in the row i cell line that match with those called in the column j cell line.