Stem Cell Reports, Volume *10*

Supplemental Information

DNA Methylation and Regulatory Elements during Chicken Germline Stem Cell Differentiation

Yanghua He, Qisheng Zuo, John Edwards, Keji Zhao, Jinzhi Lei, Wentao Cai, Qing Nie, Bichun Li, and Jiuzhou Song

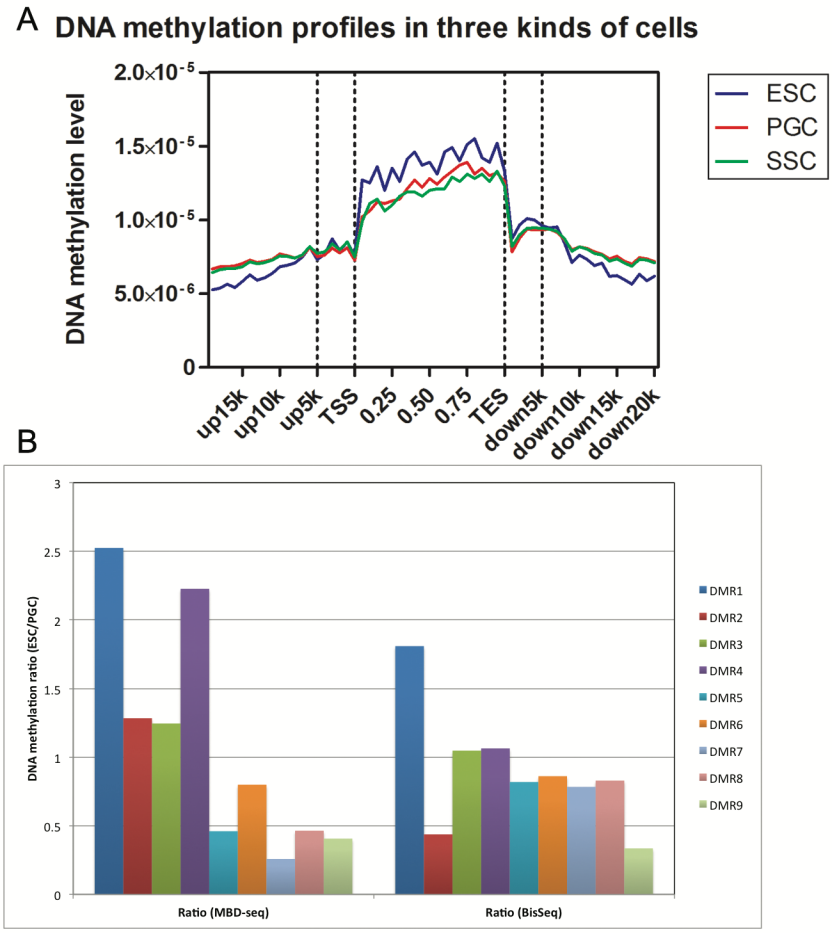

Figure S1. DNA methylation profiles in chick germ cells (related to Figure 1). (A) DNA methylation coverage among genic regions. For each gene, the read numbers detected in every 5% of the gene-body region and every 1 kb outside of the gene-body region were summed to obtain DNA methylation levels. These numbers were then normalized by the total number of base pairs in each region. (B) The validation of Differentially Methylated Regions (DMRs) by Bisulfite cloning sequencing (BisSeq) with ten positive clones for each DMR. Nine DMRs between ESCs and PGCs captured by MBD-seq assay were selected to validate their methylation levels by BisSeq. Y-axis is DNA methylation ratio of ESCs to PGCs.

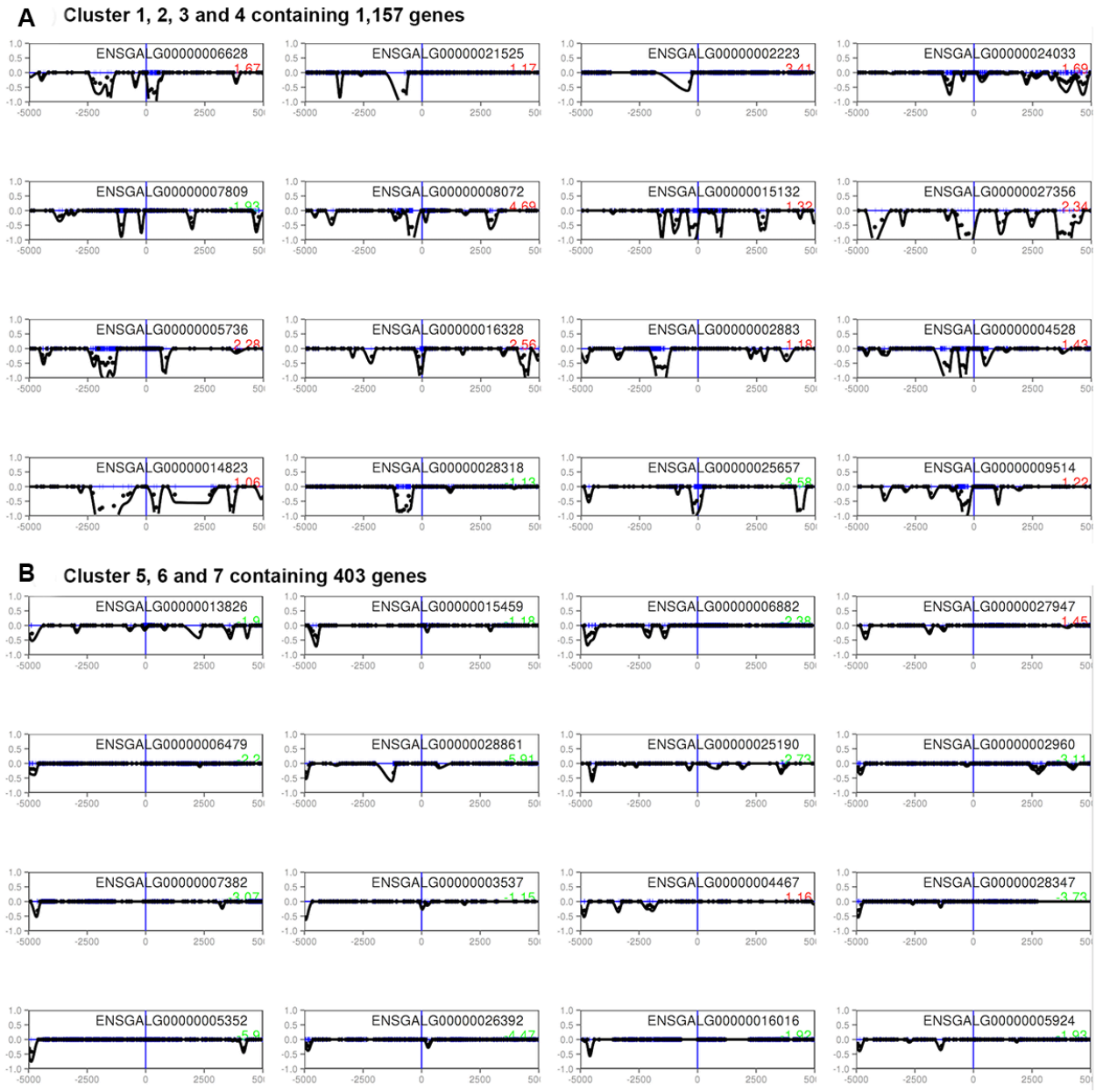

Figure S2. DNA methylation signatures in differentially expressed genes between PGCs and SSCs (related to Figure 2). Seven significant clusters of genes with same DNA methylation shape and gene expression change between PGCs and SSCs were identified by WIMSi. (A) Clusters 1, 2, 3, and 4 have similar methylation signatures with loss methylation through TSS and CGI shore. (B) Clusters 5, 6, and 7 have similar methylation signatures with loss methylation on distal regions of genes. Clustering was performed on 10 kb regions relative to TSS. The y-axis represents normalized methylation level and the x-axis represents genome position relative to the TSS (0). The number at the lower right corner denotes log2(gene expression fold change); green indicates downregulation, red indicates upregulation.

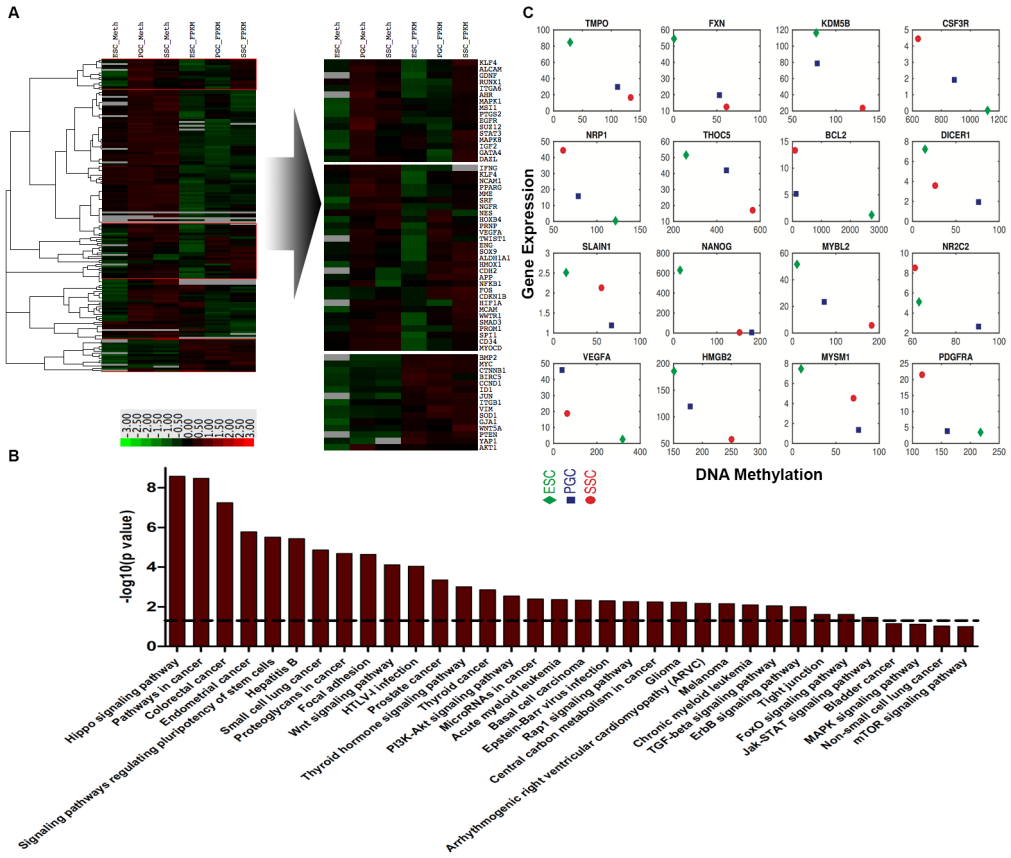

Figure S3. DNA methylation and gene expression profiles in genes associated with human stem cell differentiation (related to Figure 3). (A) The left panel: 162 genes with at least 5 hits to human stem cell differentiation were chosen to plot the DNA methylation profiles spanning upstream 2 kb relative to TSS and gene expression profiles. Normalized intensity values of genes (rows) and their methylation levels were ordered using Complete Spearman Rank Correlation and hierarchical clustering in Cluster3.0 software. The dendrogram showed the similarity (distance) of mRNA expression levels and methylation levels of genes and was divided into sub-trees as distinguished from different colors. Arrays (columns) were different cell-types for methylation level and gene expression level from left to right successively. Red and Green colors reflect the high and low intensities, respectively. Grey color represents no methylation or mRNA expression for corresponding genes. The right panel: DNA methylation and mRNA expression profiles of genes that are labeled in the three red boxes of the left panel. (B) KEGG Pathway analysis of genes in the bottom red box of (A). Dashed line: threshold line corresponds to P-value of 0.05. (C) The correlation between DNA methylation and gene expression at promoter regions (upstream 2kb relative to TSS). The x-axis represents normalized DNA methylation in promoter regions and the y-axis represents normalized gene expression.

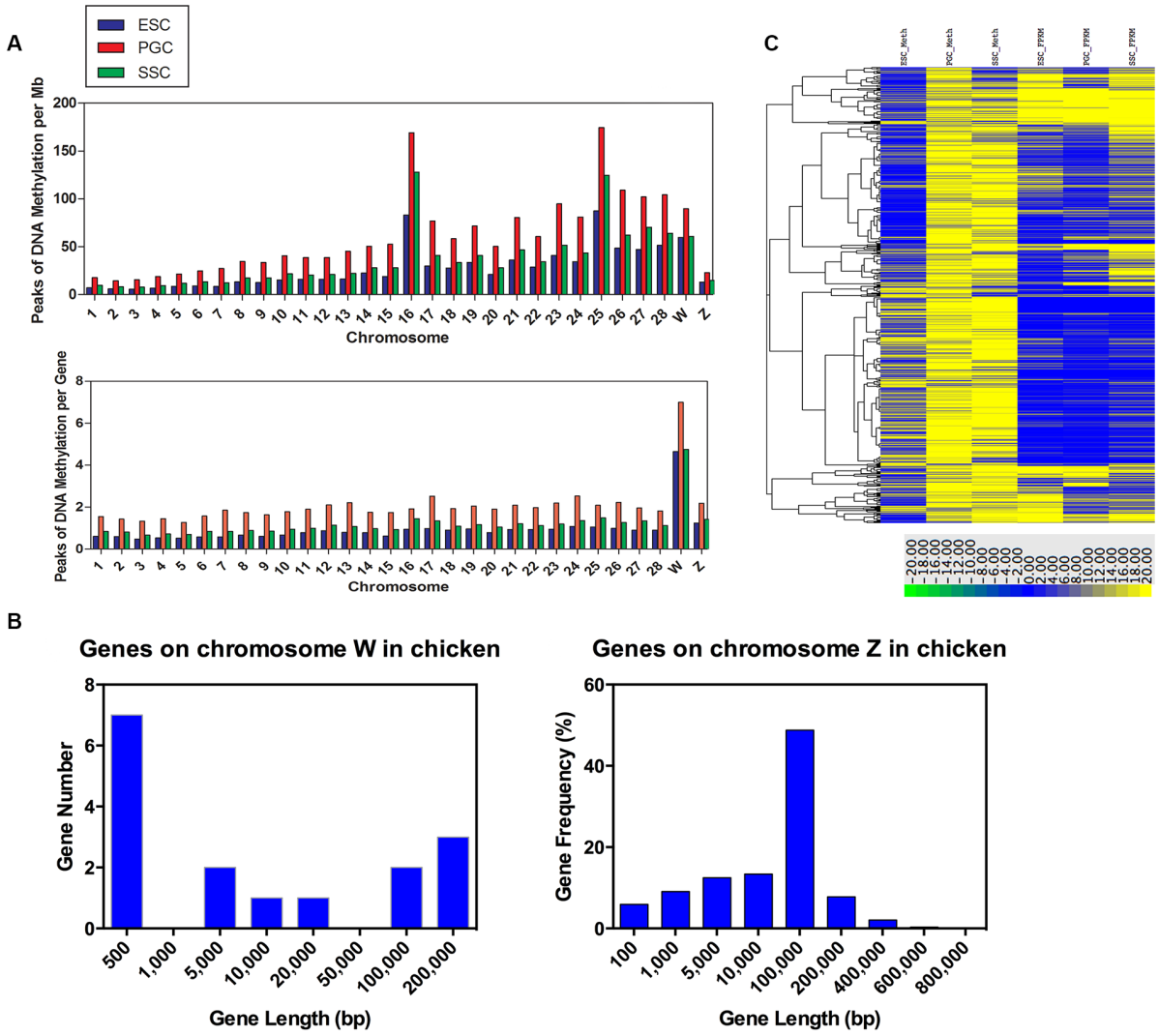

Figure S4. DNA methylation characteristics in chicken sex-chromosomes (related to Figure 3). (A) The distribution of DNA methylation on chicken chromosomes. The peaks of DNA methylation were mapped to each chromosome. The peak densities for per Mb (upper panel) and for per gene (lower panel) were calculated. (B) The statistics of genes on chicken sex chromosomes. The left panel: Gene numbers with different lengths for chromosome W. There are 16 genes in total on chromosome W in the chicken genome (galGal4). The right panel: The frequency distribution of gene size in chromosome Z. There are 859 genes in total on chromosome Z in the chicken genome (galGal4). (C) DNA methylation and gene expression profiles in all genes located in chromosome Z. Based on annotation for the chicken genome (galGal4), 859 genes were annotated on chicken chromosome Z. DNA methylation was profiling spanning upstream 2 kb relative to TSS for all chromosome Z genes. The heatmap was plotted as Figure S3A above. Yellow and blue colors reflect the high and low intensities, respectively.

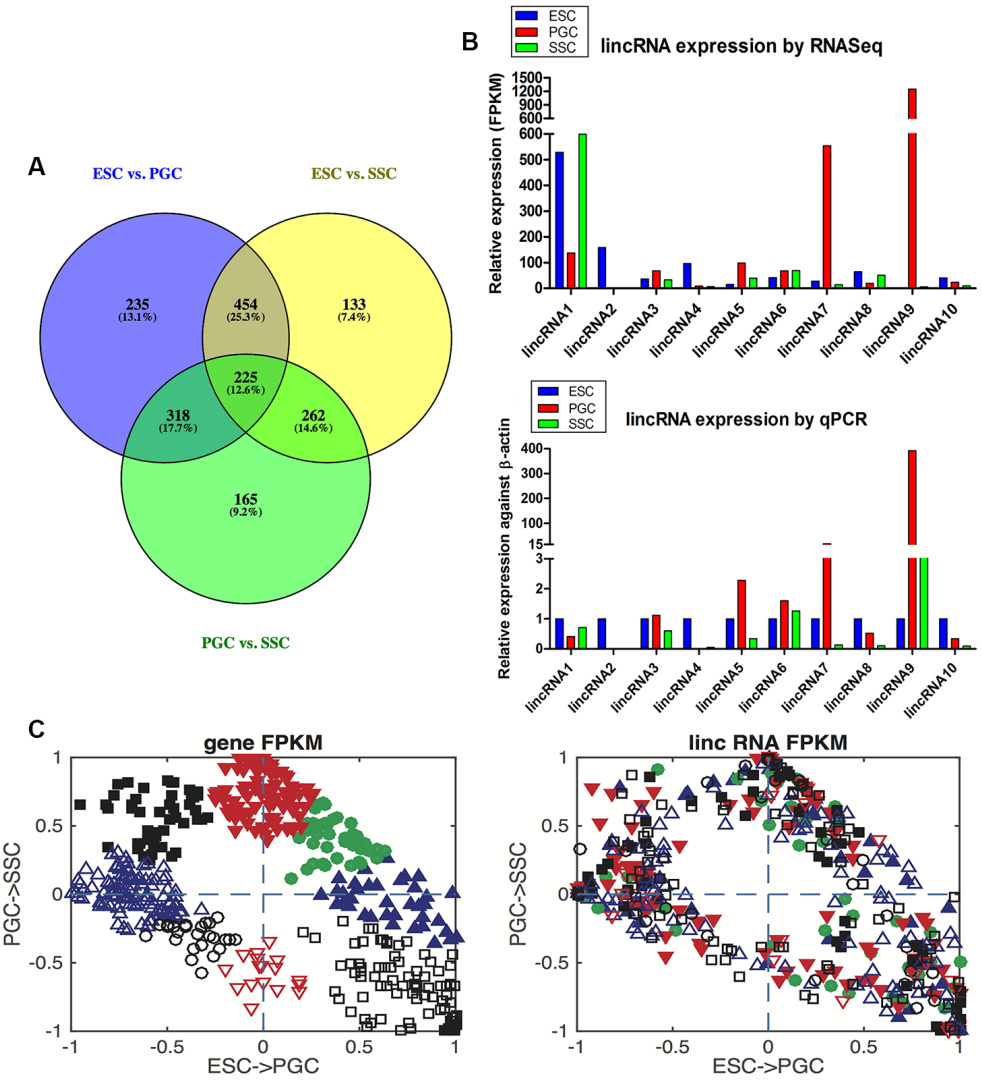

Figure S5. Differentially expressed lincRNAs and differentially expressed genes in chick germ cells (related to Figure 5). (A) Differentially expressed lincRNAs among three cell-types with the criteria of FDR <= 0.05. (B) The validation of differentially expressed lincRNAs by qPCR experiments in three chicken germ cell-types with three independent experiments. The upper panel: LincRNA expression analysis from RNA-seq data. The lower panel: LincRNA expression was validated by qPCR technique. The expression levels were normalized against *β-actin* cDNA in the corresponding samples. (C) Relative changes of gene expression and their neighboring lincRNA expression during the differentiation from ESC to PGC and PGC to SSC. The x-axis and the y-axis denote the relative expression changes of each gene (or lincRNA) at the stage of ESCs to PGCs and PGCs to SSCs, respectively. The left panel: Relative changes of mRNA expression for 451 differentially expressed genes between ESCs and PGCs, and between PGCs and SSCs. The right panel: Relative expression changes of the neighboring lincRNAs corresponding to the genes in the left panel. Here, 451 genes were separated into 8 groups according to their mRNA expression directions in two stages and each gene group was marked with different color shapes. The same color shape for the same gene was applied to its neighboring lincRNA (see Methods).

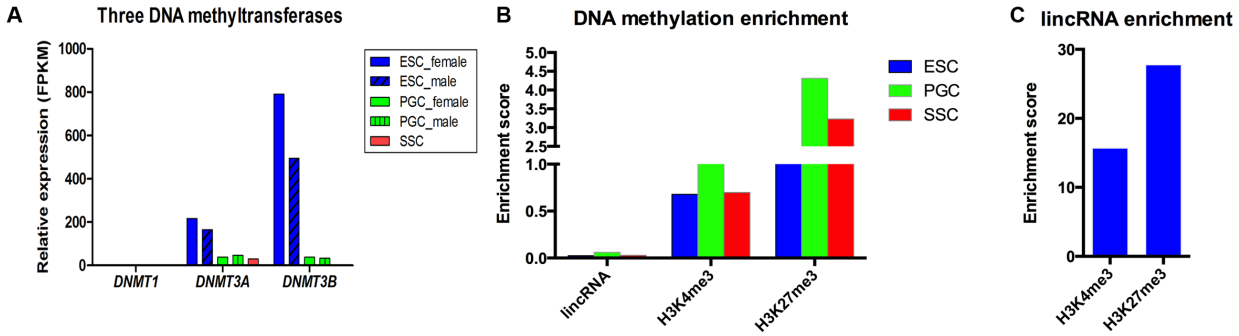

Figure S6. DNA methylation, histone modifications, and lincRNAs (related to Figure 6). (A) Expression levels of the three DNA methyltransferases were shown across three cell-types. (B) Enrichment score of DNA methylation in histone modification marks and lincRNA regions through three kinds of chicken germ cells. (C) Enrichment score of lincRNAs in histone modification marks.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Cell isolation and culture

All eggs were immediately collected after fertilization from the Experimental Poultry Farm at the National Poultry Institute of the Chinese Academy of Agricultural Sciences. A total of 18,340 fertile eggs were collected for isolation of three kinds of germline stem cells. All procedures involving the care and use of animals conformed to U.S. National Institute of Health guidelines (NIH Pub. No. 85-23, revised 1996) and were approved by the Laboratory Animal Management and Experimental Animal Ethics Committee of Yangzhou University.

The isolation of ESCs was carried out from in vitro culture of blastodermal cells taken from the area pellucida of 10,540 Stage X (EG&K) embryos. The isolated blastoderm cells were cultured and amplified on the primary chicken embryonic fibroblast (PCEF) cells as the feeder cells in the medium of Dulercco's Modified Eagle Medium (DMEM, Gibco) supplemented with 10% fetal bovine serum (FBS, Invitrogen), retinoic acid (5.5×10^{-5} mol/L, Sigma), murine leukemia inhibitory factor (LIF, 1 000 IU/ml, Sigma), basic fibroblast growth factor (bFGF, 10 ng/ml, Sigma), human insulin-like growth factor (hIGF, 10ng/ml, Sigma), and human stem cell factor (SCF, 5 ng/ml, Sigma) at 37°C in 5% CO2 with saturated humidity. The presence of alkaline phosphatase (AKP) was detected when single clones were produced.

To obtain PGCs, chicken embryos were isolated from 3,400 fertilized eggs on day 5 of incubation and rinsed with PBS. The genital ridge of embryos was cut up and digested with 0.25% trypsin and 0.05% EDTA for 5-10 minutes and then filtered and differentially cultured in a culture dish for 30 minutes. The cells were cultured in TCM-199 medium containing 10% fetal calf serum (FCS) with PCEF feeder cells. The medium was changed every few days and the cells were passaged to new wells every 4–7 days. A population of PGCs became visible after 7–10 days and was expanded under the same culture conditions. Periodic acid–Schiff (PAS) staining was performed on fixed PGCs and observed through an inverted microscope.

The testis of 4,400 fertile eggs at day 19 of incubation was cut up to obtain chicken SSCs and digested for 30 minutes by collagenase, and then digested with trypsin again. The culture medium contains DMEM supplemented with 10% FBS. These primary cultures were incubated at 37°C in 5% CO2 with saturated humidity and passaged every 2–3 days.

To obtain purified cells, a fluorescence-activated cell sorter (FACS) was used for cell sorting with two antibodies in combination to label and select cell types (see Figure 1A-B). Antibodies to OCT4 (Abcam, USA, no. ab181557) and NANOG (Abcam, USA, no. ab109250) were used to mark ESCs, antibodies to CVH (Abcam, USA, no. ab150390) and SSEA-1 (Abcam, USA, no. ab16285) were used to mark PGCs, and antibodies to INTEGRINS α-6 (Abcam, USA, ab20142) and INTEGRINS β-1 (Abcam, USA, ab3167) were used to mark SSCs. Goat anti-Rat IgG (H+L), FITC conjugate (Proteintech, USA, no. SA00003-11) and Goat anti-rabbit IgG (H+L), R-PE conjugate (Proteintech, USA, no. SA00008-2) were used as the secondary antibodies in this study.

Sorted positive cells were cultured and then determined with specific antibodies. SSEA-1 (Abcam, USA) and SOX2 (Abcam, USA, no. ab97959) antibodies were used to stain fixed ESCs, SSEA-1 and C-kit (Abcam, USA, no. ab5505) antibodies were used to stain fixed PGCs, and INTEGRINS α-6 and INTEGRINS β-1 antibodies were used to stain fixed SSCs. qRT-PCR was used to detect the expression of specific marker genes to only double-positive determined cells and the results were shown in the table below. The positive cells were collected for the next phase of the experiment. The primer pairs used for the chick sex determination at the early stage are F: GTTACTGATTCGTCTACGAGA and R: ATTGAAATGATCCAGTGCTTG with denaturing temperature of 49°C.

		ESC	PGC	SSC
Pluripotency-related genes	*NANOG*	+		
	SOX2	+	+	
	cPOUV/OCT4	+	+	+
Germline cell-related genes	*CVH*		+	+
	C-kit		+	+
Meiosis-related genes	*DAZL*		+	+
Pre-meiosis-related genes	*STRA8*			+
PGC-marker genes	*BLIMP-1*	+	+	+
SSC-marker genes	*INTEGRINS α6*			+
	INTEGRINS β1			+
Housekeeping gene	*β-actin*	+	+	+

MBD-seq and RNA-seq

Genomic DNA from ESCs, PGCs, and SSCs were extracted using the Wizard Genomic DNA purification kit (Promega, CA, USA) and DNA concentration was measured by the Qubit dsDNA Broad-Range Assay (Invitrogen, CA, USA). MethylCap kit (Diagenode, Denville, USA) was employed to obtain DNA containing methylated CpGs. Firstly, DNA was sheared into 300 - 500 bp of fragments using the Bioruptor Sonicator and checked on an agarose gel to visualize the size of the resultant segments. Secondly, the MethylCap protein was used to specifically isolate DNA containing methylated CpGs. Magnetic beads (coated with GSH) captured methylated DNA and unbound DNA was washed off. Captured DNA was eluted using sequentially Low Elution Buffer, Medium Elution Buffer and High Elution Buffer per capture reaction and collected together for sequencing library preparation. Total RNA from three cell-types was extracted using the standard TRIzol (Invitrogen, CA, USA) protocol and mRNA isolation was performed by Oligotex mRNA Mini Kit (QIAGEN, Hilden, Germany). Then mRNA was used to synthesize the first and the second strand cDNA by using SuperScriptTM III Reverse Transcriptase and oligo (dT) 12-18 primers (Invitrogen, CA, USA). After purification, the double-strand cDNA (dscDNA) was fragmented into ~300bp for sequencing library construction. The library for sequencing was constructed as follows. NEBNext End Repair Module (NEB, MA, USA) was used for the end repair of the fragmented methylated DNA and the fragmented dscDNA. Then a 3' poly-A was added using DNA Polymerase I, Large (Klenow) Fragment (NEB, MA, USA). Also, a pair of Solexa adaptors (Illumina, CA, USA) was ligated to the repaired ends by T4 ligase (Promega, CA, USA) and then 200 - 500 bp of fragments were selected on the Invitrogen® 2% E-Gel. Specific methylated fragments and dscDNA fragments were amplified by PCR using Phusion Hot Start High-Fidelity DNA Polymerase (NEB, MA, USA). After purification, DNA quality was examined by the Qubit assay (Life Technology, CA, USA). Finally, we performed DNA methylated fragment sequencing analyses with two biological replicates per cell-type on the Illumina HiSeq 2000 Analyzer following manufacturer protocols. For

RNA-seq, we had four biological replicates including two females and two male samples for ESCs, five biological replicates including two females and three male samples for PGCs, and three biological replicates for SSCs.

Analysis of DNA methylation data

The quality control of sequence files was examined by FastQC and the first 15 bases of all reads were trimmed off for high-quality assurance. The rest of 35 bases for each read were aligned to the galGal4 reference genome obtained from the UCSC browser (http://genome.ucsc.edu) by bowtie v1.1.1. For data manipulation, filtration, and format conversion, a combination of procedures available in SAMtools and BEDtools was applied. Mapped files of two biological replicates per cell-type were merged for peak calling. The peak-calling step was applied for each cell-type using MACS v1.4.2 with a *bandwidth* of 300 bp, *mfold* of 30, *p-value* of 1.00e-05 under an *FDR cutoff* of 1% to call peaks representing real methylated regions. Identification of the Differentially Methylated Regions (DMRs) was accomplished implementing the DiffBind R package (Ross-Innes *et al.* 2012) with an edgeR analysis. For normalization, the default method TMM (Trimmed Mean of M-values) that subtracts the controls reads and considers the effective library size (reads in peaks), was applied. The threshold utilized was 0.1 for False Discovery Rate (FDR).

For the general profiles of DNA methylation across gene bodies in each cell-type, the reads numbers detected in every 5% of the gene-body region and every 1 kb outside of the gene-body region for all annotated genes were summed and normalized by the total number of base pairs in each region to obtain DNA methylation level. Promoters were defined as the transcription start site (TSS) ± 2,000 bp. Over 60% of genes have a CpG-rich region, termed a CpG island, overlapping their promoter. Chicken CpG island information can be downloaded from UCSC Genome Database by the link: http://hgdownload.soe.ucsc.edu/goldenPath/galGal4/database/. Differential methylation in CpG island shores, defined as the regions up to 2kb away from a CpG island, has been linked to differential expression (Saxonov *et al.* 2006; Doi *et al.* 2009). The enrichment score of DNA methylation in annotated functional elements is essentially defined based on the proportion of genome matching one annotation (1bp resolution) and the proportion of DNA methylation matching the same annotation. For example, promoters cover about 4% of the genome and promoters in DNA methylation cover more than 4%, so the enrichment score of DNA methylation in promoters equals to the ratio of proportions and it is enriched (Pinello *et al.* 2014). Statistical significance of enrichment score was determined using Fisher's exact test (Fisher 1922). For hierarchical clustering of the DNA methylation profiles spanning upstream 2 kb relative to TSS and the correlation between DNA methylation and gene expression, DNA methylation level for each region was calculated and then normalized to TPM (number of transcript copies in per million clean reads).

Analysis of gene expression data

Sequence reads with 49 bp length were obtained from RNA-sequencing and the first 10 bases of all reads were trimmed off for high-quality assurance after quality control. The trimmed *fastq* files were mapped to the chicken reference genome (galGal4) by TopHat v2.0.9 software, and then they were assembled by cufflinks v2.1.1 with the ensemble transcriptome annotation *gtf* file that can be downloaded from UCSC - Table Browser, and the parameter of *'--GTF'* was used to obtain only known transcripts. The *cuffmerge* was then run to generate a merged *gtf* file for the use of *cuffdiff* that can compare differentially expressed genes between certain conditions. Finally, the expression levels of all genes were output and they were normalized by FPKM (Fragments Per Kilobase of transcript per Million mapped reads). Normalized gene expression levels were averaged with biological duplicates within each cell-type. The differentially expressed genes were filtered out by particular criteria.

For lincRNA identification, mapped transcripts were assembled individually with *Cufflinks*, which was run with *'--GTF-guide'* and then transcripts from all samples were then merged together with *cuffmerge* to build a consensus set of transcripts across samples. Transcripts were consequentially compared to genome annotations (ensGene and refGene) with *Cuffcompare* to exclude those that overlap with protein-coding genes, pseudogenes, and ncRNAs other than intergenic transcripts. Remaining transcripts located in the intergenic regions were then adapted to several stringent filtering steps to identify candidate lincRNAs (He *et al.* 2015). The ensGene annotation was used to identify nearest 3' and 5' neighbors of lincRNAs. To explore expression association between genes and their neighboring lincRNAs during the two stages of development - from ESC to PGC and from PGC to SSC, we calculated the relative expression changes of each gene at two stages, which were represented by

$\mathrm{ESC} \rightarrow \mathrm{PGC} = \frac{p_{i2}-p_{i1}}{\max(p_{i1},p_{i2},p_{i3})}$ and $\mathrm{PGC} \rightarrow \mathrm{SSC} = \frac{p_{i3}-p_{i2}}{\max(p_{i1},p_{i2},p_{i3})}$.

p_{ij} denotes normalized expression value (FPKM) of the i_{th} gene or lincRNA in the j_{th} cell-type that j = 1 means ESCs, 2 means PGCs and 3 means SSCs. Next, we refined the expression changes in eight directions as the table below. And genes were classified into eight groups according to different directions, and each gene group was marked with different color shapes.

Directions	Expression change during ESC->PGC	Expression change during PGC->SSC
East	Increase	Minor change
North-East	Increase	Increase
North	Minor change	Increase
North-West	Decrease	Increase
West	Decrease	Minor change
South-West	Decrease	Decrease
South	Minor change	Decrease
South-East	Increase	Decrease

The prediction of potential transcription factor (TF) motifs

In this study, HAYSTACK pipeline was used to identify cell-type-specific transcription factor motifs with DNA methylation data and quantify their activity on nearby genes (Pinello *et al.* 2014). Firstly, *haystack_hotspots* was used to find highly variable regions across different cell types also called hotspots, and then cell-type specific regions were captured from hotspots. To study what is controlling these specific regions, *haystack_motifs* was then conducted to identify enriched transcription factor motifs in each set of cell-type specific regions, and motif logos, motif hits on chicken genome, motif profiles, and motif nearby genes were obtained. Next, *haystack_tf_activity_plane* was performed to quantify motif specificity and activity on nearby genes via integrating RNA-seq data. The output is a set of Figures each containing the TF activity plane for a given motif. Average enrichment profile of motifs in cell-type specific regions: the center of each motif hits and their flanking ±2 Kb regions (x-axis) were divided into multiple bins to get a fold change value for each bin, and then all hits were piled up to get average and normalized fold change for each bin and finally plotted a motif enrichment profile. The regions above the horizontal black line and with a low q-value mean that this TF likely binds these sequences (see Figure 4B).

Integration analysis of multiple datasets

To uncover functional genes with differential methylation patterns associated with expression change of these genes, we used WIMSi based on MBD-seq and RNA-seq data to identify groups of genes with similarly shaped methylation signatures and corresponding expression changes (Vanderkraats *et al.* 2013). Firstly, only genes with greater than 2-fold expression change were filtered out for clustering methylation patterns by WIMSi. Secondly, genome-wide differential coverage analysis of MBD-seq data was conducted by the MEDIPS R package (Lienhard *et al.* 2014) with an edgeR analysis and a *windowsize* of 100 bp, and normalized *log2Foldchange* for each window was acquired for the screening of differentially methylated regions. And then a simple floor was applied to filter out large changes of DNA methylation in binding due to low signal counts, and then we re-scaled the data with a linear function and applied a ceiling to cap the values of DNA methylation on a bounded interval from [-1,1]. Finally, WIMSi pipeline was performed to get clustering genes with similarly shaped methylation signatures and more than 2-fold expression changes. The *p-value* is defined as the probability of clusters that were significantly correlated with differential expression. The purity of a set of genes is defined as the fraction of genes that have expression change in the same direction as the majority.

It was reported that epigenetic marks are at lower levels after birth in chickens. The levels of H3K9ac were found to reach the maximum value at E17, and then follow by a decrease at E20 until a minimum value at day of life 1 (D1), and then escalate until keep a stable low level after D14 through all three kinds of chicken tissues (liver, jejunum, and muscle) (Li *et al.* 2015). To investigate the interaction between DNA methylation and histone methylation, we combined our previous sequencing data of H3K4me3 and H3K27me3 (GEO Series accession number GSE65961) generated in somatic cells isolated from chicken bursa, spleen, and thymus at D5, D10, and D21 to check the enrichment of DNA methylation on histone methylation regions. Peak-calling was carried out using the WaveSeq algorithm (Mitra & Song 2012) with the mother function of *'morlet'* and *gap size* of 2 (400 bp) for H3K4me3 data and with the mother function of *'mexican hat'* and *gap size* of 10 for H3K27me3 data. The *p-value* threshold for H3K27me3 was lowered to 0.4. Peaks detected in the same genomic region of multiple biological replicates were merged to include all peaks and those appearing in only one sample were removed as possible false positives. The common peaks in all tissues and all time-points for H3K4me3 or H3K27me3 were extracted to consider as stable histone methylation marks of chickens for epigenetic integration analysis.

For TF motifs binding to lincRNAs, as long as there is 1 bp overlap between regions of a motif and a particular lincRNA we consider that this TF motif is associated with this lincRNA. For RNA-mediated interaction of TFs and histone modifications, as long as there are both of one motif sequence and one region of histone methylation peaks overlapped with a same lincRNA region, we consider that this TF motif and this histone methylation mark might form a molecular complex with lincRNA.

Bisulfite cloning sequencing for MBD-seq validation

MethylEdge™ Bisulfite Conversion System was used to treat 300 ng of DNA (Promega, CA, USA) following the standard protocol provided by the manufacturer. PCR primers were designed using MethPrimer (http://www.urogene.org/cgi-bin/methprimer/methprimer.cgi) and listed in below. Firstly, equal amounts of DNA from two samples of each cell-types were pooled together, serving as a template for the bisulfite conversion and the bisulfite PCR. Then, PCR products were purified using QIAquick Gel Extraction Kit (QIAGEN, Hilden, Germany). The purified PCR products were ligated to pGEM-T Vector (pGEM-T Vector System I, Promega, CA, USA), transformed to DH5α competent cells (Z-Competent E. Coli Cells—Strain Zymo 5α, ZYMO Research, CA, USA), and screened for successful insertions (blue-white clone selection) after incubation at 37°C overnight. In the next step, ten white colonies from each cell-type were cultured overnight in a 37°C shaker. Plasmid DNA was isolated using Zyppy Plasmid Miniprep Kit (ZYMO Research, CA, USA). M13 reverse primer and BigDye Terminator v3.1 Cycle Sequencing Kit (Life technologies, CA, USA) were employed for sequencing in the ABI

3730 machine. Bisulfite sequencing results were analyzed by QUMA (http://quma.cdb.riken.jp) and DNA methylation level for each region and group was obtained.

Primers used to validate DMRs and genes by bisulfite cloning sequencing

Bisulfite cloning sequencing	DMR Region (galGal4)	Sequence (5' → 3')	Product Length (bp)	Template	Annealing Temp. (°C)
DMR1	chr13:16771564-16772237	F: 5' AATATTAATTGGTATTTATGGGAAT 3'	350	Bisulfite treated DNA	60
		R: 5' TCAATACAAAACAAACTATCCC 3'			
DMR2	chrZ:77094055-77096685	F: 5' TGTTTGATTTTATTTGTTTTTTATA 3'	470	Bisulfite treated DNA	60
		R: 5' AAAATTTCCCTCTATTTTCTTC 3'			
DMR3	chr28:3815832-3816862	F: 5' GTGAGGAGGGTAGAGATTTGTT 3'	384	Bisulfite treated DNA	60
		R: 5' AACTAAACAAACCAACAACAATACA 3'			
DMR4	chr17:7560874-7561997	F: 5' AAGATTGTTTTGGAAATTAATA 3'	494	Bisulfite treated DNA	60
		R: 5' AAAAACACTTAACATCATCTAAC 3'			
DMR5	chr2:81631131-81631921	F: 5' GGTAGTTTTGGTTGGGGAGTAG 3'	367	Bisulfite treated DNA	60
		R: 5' AAACCTCATCCCCTAATTATCC 3'			
DMR6	chr13:2525658-2526202	F: 5' TTTTGGAGAAATTAATGGAGAT 3'	499	Bisulfite treated DNA	60
		R: 5' AACAATATAACACCAAAAAACAC 3'			
DMR7	chr2:37681914-37682389	F: 5' TTTGAGTTTTTTGTTAAAGAAGT 3'	304	Bisulfite treated DNA	60
		R: 5' ACCCTCAAATACCCTTCTCCTC 3'			
DMR8	chr9:12916536-12917412	F: 5' TGTATTTGGTAGTGGTTGGTAT 3'	441	Bisulfite treated DNA	60
		R: 5' CAAAAATAAAATTCAACCCATATAT 3'			
DMR9		F: 5' GTTTTGGGTGGAGTTTATTA 3'	422		60

	chr1:172441768-172443036	R: 5' AAAAACATTTCTTCCTACAACC 3'		Bisulfite treated DNA	
SMAD2	chrZ:1515315-1515759	F: 5' TGAAGGGACCAGACCCCA 3'	445	Bisulfite treated DNA	60
		R: 5' AGCGATGTGCCCTGCAC 3'			
SMAD5	chr13:14425700-14426034	F: 5' GCAGTGCTGAGGTTATCC 3'	335	Bisulfite treated DNA	60
		R: 5' TGACTTACATACAAGTATGC 3'			

Bisulfite Pyrosequencing for DNA methylation validation of genes

Sodium bisulfite conversion reagents were used to treat 500 ng of DNA extracted from three cell-types using the standard protocol provided by the manufacturer (EZ DNA Methylation Golden Kit). PCR primers for methylation validation are below. For pyrosequencing, we used a biotin labeled universal primer in the PCR reaction. The bisulfite PCR included 1 μl of 1:5 diluted bisulfite converted DNA, primers, and PCR reagents from Hotstar Taq polymerase kit (QIAGEN) with three biological replicates. The methylation level was detected individually by Pyro Q-CpG system (PyroMark ID, Biotage, Sweden) using 20 μl of PCR products.

Primer Name	Sequence (5'to3')	5'modifications
1.MAPKAPK5-F(290bp)	GAGATGGATAAAGAGATTAAGGTAAGGA	
1.MAPKAPK5-R	ATCCAAATAACCCTTTATCTCCT	5'Biotin
1.MAPKAPK5-S	GGTTATTTGGGAGGG	
2.GFRA2-F(128bp)	GGGAGAAGAATAGGATAGGAAGAG	
2.GFRA2-R	ACCACACAACTCACTATACTAAA	5'Biotin
2.GFRA2-S	GAGATGTTAAATAGGAAGTTT	
3.THYN1-F(126bp)	TGGTTTTAGGGAATAGAGAAATGG	
3.THYN1-R	ACCACTTTTTAAAATTCCTTCTACT	5'Biotin
3.THYN1-S	GGAATAGAGAAATGGGTT	

The validation of differentially expressed lincRNAs

The protocols of mRNA extraction and dsDNA synthesis were the same as those mentioned above. Real-time PCR using iQ™ SYBR® Green Supermix (Bio-Rad, CA, USA) was utilized to validate differentially expressed lincRNAs between cell-types. The primers were designed using Primer3 (http:// fokker.wi.mit.edu/primer3/input.htm) and confirmed by Oligo 6.0 (Supplementary Table 10). Three replicates were performed for RT-qPCR reactions. qPCR reaction was run using the program as follow: pre-incubation (95°C for 10min), 40 cycles of amplification (95°C for 10s, 60°C for 10s, and 72°C for 10s), melting curves using a heat ramp and cool down. Cycle threshold values (Ct values) were obtained from iCycler iQ PCR software. The expressions of lincRNAs were normalized against *β-actin* cDNA in the corresponding samples. The relative fold enrichment of each group was calculated by comparing the enrichment value for the given primer pair to *β-actin*.

The inhibition of TGF-β/BMP signaling pathway

To explore the function of TGF-β/BMP signaling pathway in the regulation of male germ cell formation, TGF-β signaling pathway specific inhibitors, LY-100 (100 nM of LY2109761 – antagonist to the TGF-β subgroup) and LDN-100 (100 nM of LDN193189 – antagonist to the BMP subgroup), were added to the BMP4 induction medium to inhibit SMAD2 and SMAD5 expression *in vitro*. LY2109761 is a novel selective TGF-β receptor type I/II (TβRI/II) dual inhibitor, and blocking TβRI/II kinase activity with LY2109761 completely suppresses TGF-β–induced SMAD2 phosphorylation (Melisi *et al.* 2008). LDN193189 is a highly potent small molecule BMP inhibitor that inhibits BMP4-mediated SMAD1, SMAD5, and SMAD8 phosphorylation and efficiently inhibits the transcriptional activity of the BMP type I receptors ALK2 and ALK3 (Yu *et al.* 2008). Cells at the different stages were collected to extract RNA and proteins. qRT-PCR was performed to evaluate the inhibition efficiency of TGF- β signaling. *SMAD2* and *SMAD5* mRNA expression in control and inhibition groups were measured at the differentiation day 4 (the time to form PGC-like cells) and 14 (the time to form SSC-like cells) of ESCs. Cell treatment groups were as follows: ① control group: with no treatment; ② BMP4 group: induction with 40 ng/mL of BMP4; ③ BMP4 + LY group: induction with BMP4 and 100 nM/L of LY2109761; ④ BMP4 + LDN group: induction with BMP4 and 100 nM/L of LDN193189; ⑤ BMP4 + DOUBLE group: induced with BMP4 and equal concentrations of both of the inhibitors. The expression of *SMAD2* and *SMAD5* in BMP4 group was regarded as a Positive Control, and the sample treat with H_2O was regarded as a Blank Control.

To further verify the regulatory mechanism of TGF-β/BMP signaling pathway during germ cell generation *in vivo*, LY-100 and LDN-100 were injected into chicken blastoderms to incubate. *SMAD2* and *SMAD5* mRNA expression in control and inhibition groups were measured at the embryo development day 5.5 (the time to form PGCs) and 18 (the time to form SSCs). Fertilized embryos were grouped as follows: ① blank group: without any treatment during the incubation process; ② control group (CON): injection with 100 μl of ddH2O; ③ LY-100 group: injection with 100 μl of LY2109761 (100 nM/L); ④ LDN-100 group: injection with 100 μl of LDN193189 (100 nM/L); ⑤ dual inhibitor group (DOUBLE): injection of more than 100 mol/L two inhibitors mixture. Injections were made into the egg white from the tip of the egg, and then sealed with paraffin and incubated at 38.5 ℃. The expression of *SMAD2* and *SMAD5* with the normal incubation process was regarded as a Control, and the sample treat with H_2O was regarded as a Blank Control.

Western blot analysis

Western blot analysis was performed in the cells collected from all stages before and after the inhibition of TGF-β/BMP signaling pathway *in vivo* and *in vitro*. The cell pellet was resuspended in ice-cold 1xRIPA buffer containing proteinase inhibitor cocktail (Sigma) and incubated on ice for a while. And then cells were lysed with 2X laemmli SDS buffer (Sigma, S3401), and protein lysate was loaded to NuPage@ 4-12% Bis-Tris Gel for electrophoresis (Novex, life technologies). After separation, proteins were transferred onto a polyvinylidene fluoride (PVDF) membrane overnight and then blocked with 5% nonfat milk in TBST containing 0.1% Tween 20. Primary antibodies against SMAD2 (Abcam, USA, no. ab40855, 58kDa) and SMAD5 (Abcam, USA, no. ab40771, 52kDa), and β-actin (Abcam, USA, no. ab8227, 42kDa) were prepared at 1:500 and 1:1000 dilutions, respectively. The membranes were incubated with primary antibodies for 2 hours at room temperature and then were washed quickly 3 times (less than 1 min) and 3 more times for 5 min. The membranes were then incubated with anti-rabbit IgG and anti-mouse IgG secondary antibodies (Santa Cruz) diluted in TBST (1:5000 or 1:10,000) for 1 hour at room temperature and then washed 3 more times for 5 min each. Finally, the membranes were developed with

ECL (Amersham) and measured using ChemiDoc XRS (Bio-Rad). Four independent experiments were done for each antibody.

Chromatin immunoprecipitation assay (ChIP) assay and qPCR

Cells were cross-linked with 37% formaldehyde solution (Sigma). Briefly, chromatin from fixed cells was fragmented to a size range of 200–700 bases and the pre-clear chromatin was added 200uL of protein A/G agarose bead slurry for rotating overnight in a cold room. The pre-bind antibody was added 50uL of protein A/G agarose bead slurry and 1mL of cold PBS on the ice and incubated with 10ug of HoxA5 antibody (Santa Cruz Biotechnology, USA, no. sc-365784) for rotating overnight in a cold room. And then 1mL of pre-cleared chromatin was added to each microcentrifuge tube of antibody: bead complex (IP samples) and rotated in a cold room for 2-4 hours. The antibody: chromatin: bead complex was then washed 6 times with buffers at 4C. The DNA: protein complexes from beads were eluted twice with 100uL of elution buffer. After cross-link reversal and proteinase K treatment, ChIP-DNA was extracted and treated with RNase and eventually quantified by the Qubit assay (Life Technology, CA, USA). qPCR was followed to measure HOXA5 affinity on the *GFRA2* promoter in three kinds of chick germ cells with HOXA5 ChIP-DNA as a template with a triple. The qPCR primer pair used is

F: CAACCTGAACGACAACTGCA and R: AAGAGGAGGCGGTAGGAGTA.

The Dual-Luciferase Reporter Assay

To avoid effects of endogenous genes on the detection efficiency of *MAPKAPK5* gene and its neighboring lincRNA binding in chick germ cells, human 293T cells were selected to perform the dual-luciferase reporter assay to detect the binding power of *MAPKAPK5* gene and its neighboring lincRNA. 293T cells were grown in DMEM supplemented with 10% FBS, streptomycin (100 mg/ml) and penicillin (100 U/ml) (Invitrogen). Cells were maintained at 37°C in 5% CO_2 with saturated humidity. *MAPKAPK5* promoter sequence was amplified by PCR and cloned into pGL3-Basic Luciferase reporter vector (Promega) to build the plasmid construct. The lincRNA sequence was also amplified by PCR and cloned into pCDNA3.1(+) expression vector (Invitrogen). The ligated products were then transformed into 5-alpha Competent E. coli cells (NEB). The positive clones were picked up to extract plasmid DNA and then sequenced, and the correct clones were eventually determined and extracted plasmid DNA with a large amount for transfection steps. For transfection, 293T cells were plated in 12-well plates and the medium was replaced with growth medium without antibiotics. The two kinds of vectors were individually transfected into 293T cells as well as their co-transfection. Cells were harvested after 24 hours, and Renilla and firefly luciferase activities were measured using Dual-Luciferase Reporter Assay System (Promega).

Public data sources

Chicken genome assembly galGal4, refGene annotation, and ensGene annotation were downloaded from UCSC Genome Browser (http://genome.ucsc.edu/index.html). Chromosome information (Genome) file and CpG island information were downloaded from UCSC at http://hgdownload.soe.ucsc.edu/goldenPath/galGal4/database. Human genes associated with stem cell differentiation were searched with *Search word(s)* of 'stem cell differentiation' (http://ci.smu.edu.cn/CooLGeN/index.php). Mouse genes associated with stem cell differentiation were searched by the keyword of 'mouse stem cell differentiation' in NCBI Gene Database (http://www.ncbi.nlm.nih.gov). As for gene functional annotation and functional clustering of genes, DAVID 6.8 (https://david.ncifcrf.gov) and GenCLiP 2.0 (Wang *et al.* 2014) (http://ci.smu.edu.cn/GenCLiP2.0/analysis.php) were applied.

Table S1. Pathway analysis of genes from cluster 1, 2, 3 and 4 with methylation change correlated with expression change between PGCs and SSCs with literature profiles in humans (related to Table 1 & Figure 2)

Pathway	Hit	Total	P-Value	Q-Value
BIOCARTA_VITCB PATHWAY	5	11	0.0002	0.0048
cluster1 Enrichment Score: 3.69				
REACTOME_AXON GUIDANCE	22	161	8.912e-06	0.0009
REACTOME_NCAM SIGNALING FOR NEURITE OUT GROWTH	11	69	0.0006	0.0072
REACTOME_SIGNALING BY PDGF	12	64	1.897e-05	0.0012
REACTOME_NCAM1 INTERACTIONS	9	44	8.525e-05	0.0033
REACTOME_SEMAPHORIN INTERACTIONS	8	66	0.0443	0.135
cluster2 Enrichment Score: 3.39				
BIOCARTA_TFF PATHWAY	6	21	4.621e-05	0.0022
BIOCARTA_GH PATHWAY	5	28	0.017	0.0743
BIOCARTA_AKT PATHWAY	6	22	8.555e-05	0.0027
cluster3 Enrichment Score: 3.35				
KEGG_ECM RECEPTOR INTERACTION	13	84	0.0002	0.0046
KEGG_FOCAL ADHESION	22	199	0.0008	0.0081
cluster4 Enrichment Score: 3.05				
BIOCARTA_GLEEVEC PATHWAY	7	23	2.674e-06	0.0005
BIOCARTA_FCER1 PATHWAY	6	39	0.0223	0.0874
BIOCARTA_PDGF PATHWAY	7	32	0.0003	0.0049
BIOCARTA_EGF PATHWAY	5	31	0.0327	0.1101
cluster5 Enrichment Score: 2.83				
KEGG_SMALL CELL LUNG CANCER	13	84	0.0002	0.0041
KEGG_COLORECTAL CANCER	8	62	0.0278	0.1008
BIOCARTA_BAD PATHWAY	6	26	0.0006	0.007

KEGG_PATHWAYS IN CANCER	32	325	0.0008	0.0085
KEGG_APOPTOSIS	12	87	0.0022	0.0167
cluster6 Enrichment Score: 2.77				
BIOCARTA_VEGF PATHWAY	7	29	9.273e-05	0.0025
KEGG_MTOR SIGNALING PATHWAY	7	52	0.0324	0.1131
KEGG_RENAL CELL CARCINOMA	11	70	0.0007	0.0073
REACTOME_COLLAGEN MEDIATED ACTIVATION CASCADE	5	23	0.0039	0.0248
cluster7 Enrichment Score: 2.68				
KEGG_TYPE II DIABETES MELLITUS	9	47	0.0002	0.0047
KEGG_INSULIN SIGNALING PATHWAY	14	137	0.0199	0.085
cluster8 Enrichment Score: 2.48				
SIG_IL4RECEPTOR IN B LYPHOCYTES	6	27	0.0009	0.0085
SIG_BCR SIGNALING PATHWAY	8	46	0.0017	0.0145
ST_B CELL ANTIGEN RECEPTOR	6	39	0.0223	0.0893
cluster9 Enrichment Score: 2.43				
BIOCARTA_NGF PATHWAY	5	18	0.0004	0.0052
BIOCARTA_INSULIN PATHWAY	5	22	0.0027	0.0196
BIOCARTA_CREB PATHWAY	6	27	0.0009	0.0081
BIOCARTA_IGF1 PATHWAY	5	21	0.0018	0.014
ST_DIFFERENTIATION PATHWAY IN PC12 CELLS	6	42	0.0367	0.1176
REACTOME_DOWNSTREAM SIGNALING OF ACTIVATED FGFR	6	43	0.0427	0.1322
BIOCARTA_P53HYPOXIA PATHWAY	5	23	0.0039	0.0233
REACTOME_INTEGRIN CELL SURFACE INTERACTIONS	11	81	0.0043	0.0249
cluster10 Enrichment Score: 2.35				
KEGG_PROSTATE CANCER	13	89	0.0002	0.0039

KEGG_CHRONIC MYELOID LEUKEMIA	9	73	0.0264	0.0993
KEGG_ACUTE MYELOID LEUKEMIA	8	57	0.014	0.0657
KEGG_DORSO VENTRAL AXIS FORMATION	5	24	0.0055	0.031
cluster11 Enrichment Score: 2.20				
BIOCARTA_PPARA PATHWAY	9	58	0.0029	0.0202
BIOCARTA_EDG1 PATHWAY	5	27	0.0133	0.0636
cluster12 Enrichment Score: 2.12				
BIOCARTA_CERAMIDE PATHWAY	5	22	0.0027	0.0189
BIOCARTA_HIVNEF PATHWAY	8	58	0.0163	0.0726
ST_GA13 PATHWAY	6	35	0.01	0.0503
cluster13 Enrichment Score: 2.07				
BIOCARTA_RAC1 PATHWAY	5	23	0.0039	0.024
BIOCARTA_MET PATHWAY	6	37	0.0152	0.0697
BIOCARTA_RACCYCD PATHWAY	5	26	0.0101	0.0498
cluster14 Enrichment Score: 2.02				
KEGG_NEUROTROPHIN SIGNALING PATHWAY	16	126	0.0006	0.0068
KEGG_ERBB SIGNALING PATHWAY	10	87	0.033	0.1091
KEGG_GLIOMA	9	65	0.0094	0.0489
KEGG_FC EPSILON RI SIGNALING PATHWAY	9	79	0.0489	0.1424
cluster15 Enrichment Score: 1.98				
KEGG_FC GAMMA R MEDIATED PHAGOCYTOSIS	12	96	0.0035	0.0235
KEGG_PHOSPHATIDYLINOSITOL SIGNALING SYSTEM	9	76	0.0364	0.1186
KEGG_INOSITOL PHOSPHATE METABOLISM	8	54	0.0087	0.0464
cluster16 Enrichment Score: 1.72				
SIG_PIP3 SIGNALING IN CARDIAC MYOCTES	9	63	0.007	0.0382

SIG_INSULIN RECEPTOR PATHWAY IN CARDIAC MYOCYTES	7	49	0.021	0.0875
ST_PHOSPHOINOSITIDE 3 KINASE PATHWAY	5	33	0.0472	0.1393
KEGG_ALZHEIMERS DISEASE	16	165	0.0234	0.0899
cluster17 Enrichment Score: 1.62				
BIOCARTA_TNFR1 PATHWAY	5	29	0.0215	0.0878
BIOCARTA_FAS PATHWAY	5	30	0.0267	0.0986
KEGG_ARRHYTHMOGENIC RIGHT VENTRICULAR CARDIOMYOPATHY ARVC	9	74	0.0295	0.1048
BIOCARTA_NO1 PATHWAY	5	31	0.0327	0.1121
REACTOME_SIGNALING BY ROBO RECEPTOR	5	32	0.0395	0.1243
KEGG_PEROXISOME	9	78	0.0445	0.1335

Enrichment score: The overall enrichment score for the group based on the P-value of each term members.

Hit: Genes involved in the keywords link to the related abstract.

Total: All genes involved in the keywords.

P-value: Chi-square Test P-value.

Q-value: corrected p-value.

Table S2. Pathway analysis of genes from cluster 5, 6 and 7 with methylation change correlated with expression change between PGCs and SSCs with literature profiles in humans (related to Figure 2)

Pathway	Hit	Total	P-Value	Q-Value
cluster1 Enrichment Score : 3.78				
REACTOME_PYRIMIDINE METABOLISM	5	22	6.954e-05	0.0024
REACTOME_METABLISM OF NUCLEOTIDES	6	71	0.0007	0.0081
KEGG_PYRIMIDINE METABOLISM	8	98	8.981e-05	0.0015
cluster2 Enrichment Score: 2.24				
REACTOME_AXON GUIDANCE	9	161	0.0037	0.0316
REACTOME_NCAM SIGNALING FOR NEURITE OUT GROWTH	5	69	0.0089	0.0606
KEGG_PPAR SIGNALING PATHWAY	5	69	0.0089	0.0505

Table S3. The genes related to human stem cell differentiation not only showed a linear correlation between DNA methylation and gene expression but also were found to be significant methylation signatures identified by WIMSi (related to Figure 2).

No.	Symbol	Ensembl ID	Description
1	*ALKBH4*	ENSGALG00000001844	alkB homolog 4, lysine demethylase
2	*BCL2*	ENSGALG00000012885	Apoptosis regulator Bcl-2
3	*BMPER*	ENSGALG00000012184	BMP-binding endothelial regulator protein precursor
4	*CHD7*	ENSGALG00000015472	Chromodomain-helicase-DNA-binding protein 7
5	*CSF3R*	ENSGALG00000002112	granulocyte colony-stimulating factor receptor
6	*DICER1*	ENSGALG00000010999	endoribonuclease Dicer
7	*FXN*	ENSGALG00000015108	frataxin
8	*HMGB2*	ENSGALG00000010745	high mobility group protein B2
9	*IGF2R*	ENSGALG00000011621	cation-independent mannose-6-phosphate receptor precursor
10	*KDM5B*	ENSGALG00000000427	lysine-specific demethylase 5B
11	*MYBL2*	ENSGALG00000003503	myb-related protein B
12	*MYSM1*	ENSGALG00000010869	histone H2A deubiquitinase MYSM1
13	*NANOG*	ENSGALG00000027772	nanog homeobox
14	*NF2*	ENSGALG00000008073	neurofibromin 2 (bilateral acoustic neuroma)
15	*NOS2*	ENSGALG00000005693	nitric oxide synthase, inducible
16	*NR2C2*	ENSGALG00000008519	nuclear receptor subfamily 2 group C member 2
17	*NRP1*	ENSGALG00000007140	neuropilin-1 precursor
18	*PDGFRA*	ENSGALG00000013929	platelet-derived growth factor receptor alpha precursor
19	*PDS5B*	ENSGALG00000017070	sister chromatid cohesion protein PDS5 homolog B
20	*PPARG*	ENSGALG00000004974	peroxisome proliferator-activated receptor gamma
21	*PPP1R1B*	ENSGALG00000028797	protein phosphatase 1 regulatory inhibitor subunit 1B
22	*PROM1*	ENSGALG00000014496	prominin 1
23	*SLAIN1*	ENSGALG00000016913	SLAIN motif family member 1
24	*SLC26A5*	ENSGALG00000008222	prestin
25	*SMAD2*	ENSGALG00000014697	mothers against decapentaplegic homolog 2
26	*SMAD5*	ENSGALG00000006309	mothers against decapentaplegic homolog 5
27	*SMAD9*	ENSGALG00000017050	mothers against decapentaplegic homolog 9
28	*SOX2OT_exon4*	ENSGALG00000026758	SOX2 overlapping transcript exon 4
29	*THOC5*	ENSGALG00000008082	THO complex subunit 5 homolog
30	*TMPO*	ENSGALG00000011504	lamina-associated polypeptide 2, isoform beta
31	*VEGFA*	ENSGALG00000010290	Vascular endothelial growth factor A

Table S4. List of transcriptional factor (TF) motifs predicted to be associated with DNA methylation (related to Figure 4)

No.	Cell-type	Motif Name	Motif ID	Presence in Target	Presence in BG	Ratio	p-value	q-value	Central Enrichment
1	ESC	FOXD1	MA0031.1	5.86%	0.00%	6.86	4.14E-04	1.98E-04	1.16
2	ESC	HOXA5	MA0158.1	5.33%	0.00%	6.33	1.08E-03	4.08E-04	1.19
3	ESC	CEBPB	MA0466.1	7.43%	0.67%	5.04	4.99E-04	2.27E-04	1.05
4	ESC	Hoxc9	MA0485.1	15.48%	2.68%	4.47	2.43E-06	4.01E-06	2.28
5	ESC	Gata4	MA0482.1	10.15%	2.01%	3.7	3.61E-04	1.87E-04	1.01
6	ESC	Hoxa9	MA0594.1	14.85%	3.36%	3.64	2.40E-05	2.21E-05	2.02
7	ESC	CREB1	MA0018.2	7.11%	1.34%	3.46	3.51E-03	1.06E-03	1.45
8	ESC	Meis1	MA0498.1	28.66%	8.72%	3.05	3.31E-08	1.20E-07	1.05
9	ESC	STAT1	MA0137.3	14.85%	4.70%	2.78	2.75E-04	1.59E-04	1
10	ESC	CDX2	MA0465.1	10.04%	3.36%	2.54	5.66E-03	1.54E-03	1.01
11	ESC	Pax4	MA0068.1	26.67%	10.74%	2.36	8.92E-06	1.17E-05	1.1
12	ESC	NR4A2	MA0160.1	15.48%	6.71%	2.14	3.54E-03	1.06E-03	1.16
13	ESC	MZF1_5-13	MA0057.1	43.51%	20.13%	2.11	2.90E-08	1.20E-07	1.1
14	ESC	EWSR1-FLI1	MA0149.1	57.43%	26.85%	2.1	2.39E-12	4.34E-11	1.05
15	ESC	RFX5	MA0510.1	32.85%	15.44%	2.06	8.98E-06	1.17E-05	1
16	ESC	Erg	MA0474.1	34.41%	16.78%	1.99	1.17E-05	1.42E-05	1.04
17	ESC	Rfx1	MA0509.1	22.59%	11.41%	1.9	1.62E-03	5.66E-04	1.22
18	ESC	Myod1	MA0499.1	42.36%	22.15%	1.87	1.54E-06	2.79E-06	1.05
19	ESC	Myog	MA0500.1	41.21%	22.15%	1.82	5.36E-06	8.12E-06	1.02
20	ESC	znf143	MA0088.1	50.94%	28.19%	1.78	2.34E-07	7.08E-07	1.04
21	ESC	KLF5	MA0599.1	57.95%	32.89%	1.74	1.62E-08	9.81E-08	1.05
22	ESC	Tcf12	MA0521.1	40.17%	22.82%	1.73	3.89E-05	3.08E-05	1.01
23	ESC	TFAP2C	MA0524.1	53.87%	31.54%	1.69	4.00E-07	9.08E-07	1.06
24	ESC	PAX5	MA0014.2	43.83%	26.85%	1.61	7.81E-05	5.07E-05	1.38
25	ESC	NR2C2	MA0504.1	45.92%	28.19%	1.61	4.08E-05	3.09E-05	1.06
26	ESC	TFAP2A	MA0003.2	50.00%	30.87%	1.6	1.36E-05	1.45E-05	1.12
27	ESC	RFX2	MA0600.1	27.20%	16.78%	1.59	6.44E-03	1.72E-03	1.36
28	ESC	Tcfcp2l1	MA0145.2	38.08%	26.17%	1.44	4.59E-03	1.30E-03	1.12
29	ESC	SP2	MA0516.1	71.44%	59.73%	1.19	5.22E-03	1.44E-03	1.22
1	PGC	TP53	MA0106.2	19.87%	8.01%	2.32	4.13E-04	2.27E-02	1.03
2	PGC	PPARG	MA0066.1	21.79%	10.58%	1.97	1.87E-03	2.58E-02	1.53
3	PGC	SREBF1	MA0595.1	19.87%	9.62%	1.97	3.27E-03	2.79E-02	1.77

4	PGC	SREBF2	MA0596.1	19.23%	9.94%	1.85	8.24E-03	5.04E-02	1.9
5	PGC	Nfe2l2	MA0150.2	23.72%	12.82%	1.79	3.55E-03	2.79E-02	1.16
6	PGC	ZEB1	MA0103.2	32.05%	18.91%	1.66	2.44E-03	2.68E-02	1.11
7	PGC	TP63	MA0525.1	34.62%	20.83%	1.63	1.59E-03	2.58E-02	1.01
8	PGC	HIF1A::ARNT	MA0259.1	42.95%	28.21%	1.5	1.75E-03	2.58E-02	2.12
9	PGC	Atoh1	MA0461.1	40.38%	26.92%	1.48	4.24E-03	2.92E-02	1.29
1	SSC	NFYB	MA0502.1	37.50%	6.68%	5.01	9.38E-38	1.44E-36	7.03
2	SSC	GABPA	MA0062.2	62.24%	18.38%	3.26	4.32E-50	1.99E-48	6.07
3	SSC	FOXP2	MA0593.1	25.00%	7.33%	3.12	4.63E-16	1.78E-15	3.67
4	SSC	NFYA	MA0060.2	36.73%	11.57%	3	5.25E-23	3.45E-22	5.78
5	SSC	T	MA0009.1	17.09%	5.14%	2.95	1.27E-10	4.50E-10	2.15
6	SSC	ELK4	MA0076.2	61.99%	21.34%	2.82	3.00E-42	6.91E-41	3.09
7	SSC	ELK1	MA0028.1	32.65%	11.70%	2.65	3.37E-17	1.55E-16	3.13
8	SSC	ELF1	MA0473.1	59.18%	22.37%	2.58	4.10E-35	4.72E-34	1.8
9	SSC	Mecom	MA0029.1	12.76%	4.37%	2.56	4.77E-07	1.22E-06	3.08
10	SSC	TEAD1	MA0090.1	38.52%	14.65%	2.52	2.80E-19	1.61E-18	2.34
11	SSC	TAL1::TCF3	MA0091.1	28.83%	13.11%	2.11	1.70E-10	5.58E-10	1.79
12	SSC	MAFF	MA0495.1	25.00%	11.31%	2.11	5.39E-09	1.55E-08	2.39
13	SSC	Pou5f1::Sox2	MA0142.1	14.80%	6.81%	2.02	1.98E-05	4.34E-05	1.27
14	SSC	FLI1	MA0475.1	63.27%	31.11%	2	9.68E-26	7.43E-25	1.48
15	SSC	FOXP1	MA0481.1	32.65%	17.48%	1.82	9.42E-09	2.55E-08	1.85
16	SSC	E2F1	MA0024.2	72.45%	40.23%	1.78	6.85E-26	6.30E-25	2.95
17	SSC	Ets1	MA0098.2	64.03%	38.05%	1.67	5.26E-17	2.20E-16	1.17
18	SSC	Gata1	MA0035.3	14.54%	8.61%	1.62	2.46E-03	4.35E-03	1.61
19	SSC	E2F4	MA0470.1	75.26%	48.71%	1.53	1.54E-18	7.88E-18	3.29
20	SSC	YY1	MA0095.2	25.00%	16.45%	1.49	7.14E-04	1.49E-03	1.68
21	SSC	Mafb	MA0117.1	44.13%	29.31%	1.49	6.58E-07	1.51E-06	2.91
22	SSC	REST	MA0138.2	71.17%	53.34%	1.33	3.99E-09	1.22E-08	3.77
23	SSC	THAP1	MA0597.1	64.80%	49.36%	1.31	6.11E-07	1.48E-06	1.68
24	SSC	RREB1	MA0073.1	63.01%	53.47%	1.18	2.18E-03	4.00E-03	1.19
25	SSC	EGR1	MA0162.2	77.81%	69.15%	1.12	1.88E-03	3.61E-03	3.37

Motif Name: a name that usually corresponds to the TF gene.

Motif ID: the id of the motif from the JASPAR database.

Presence in Target: the percentage of the specific regions contains a given motif.

Presence in BG: the percentage of the background contains a given motif.

Ratio: the ratio of Presence in Target to Presence in BG.

p-value: Fisher Exact Test P-value.

q-value: FDR.

Central Enrichment: the ratio of the signal intensity at the center of the regions to the flanking regions of the motif profile. The central enrichment is higher; the TF is more likely to sit at the center of these regions.

Table S5. List of TF motifs binding to lincRNAs in three cell-types (related to Figure 6)

ESC			PGC			SSC		
Motif ID	Number of motif hits lincRNAs	Motif Name	Motif ID	Number of motif hits lincRNAs	Motif Name	Motif ID	Number of motif hits lincRNAs	Motif Name
MA0516.1	93	SP2	MA0461.1	9	Atoh1	MA0162.2	3	EGR1
MA0014.2	72	PAX5	MA0103.2	5	ZEB1	MA0098.2	2	Ets1
MA0524.1	65	TFAP2C	MA0066.1	4	PPARG	MA0117.1	2	Mafb
MA0599.1	65	KLF5	MA0150.2	2	Nfe2l2	MA0470.1	2	E2F4
MA0500.1	60	Myog	MA0259.1	2	HIF1A::ARNT	MA0024.2	1	E2F1
MA0474.1	58	Erg	MA0595.1	2	SREBF1	MA0060.2	1	NFYA
MA0003.2	57	TFAP2A	MA0596.1	2	SREBF2	MA0062.2	1	GABPA
MA0521.1	57	Tcf12	MA0106.2	1	TP53	MA0073.1	1	RREB1
MA0499.1	56	Myod1	MA0525.1	1	TP63	MA0076.2	1	ELK4
MA0504.1	56	NR2C2				MA0095.2	1	YY1
MA0149.1	55	EWSR1-FLI1				MA0473.1	1	ELF1
MA0088.1	53	znf143				MA0475.1	1	FLI1
MA0510.1	53	RFX5				MA0495.1	1	MAFF
MA0498.1	51	Meis1				MA0597.1	1	THAP1
MA0057.1	44	MZF1_5-13						
MA0145.2	40	Tcfcp2l1						
MA0068.1	36	Pax4						
MA0600.1	34	RFX2						
MA0509.1	29	Rfx1						
MA0594.1	27	Hoxa9						
MA0485.1	26	Hoxc9						
MA0018.2	22	CREB1						
MA0137.3	18	STAT1						
MA0482.1	18	Gata4						
MA0466.1	12	CEBPB						
MA0158.1	11	HOXA5						
MA0160.1	11	NR4A2						
MA0031.1	7	FOXD1						

SUPPLEMENTAL REFERENCE

Doi A., Park I.H., Wen B., Murakami P., Aryee M.J., Irizarry R., Herb B., Ladd-Acosta C., Rho J., Loewer S., Miller J., Schlaeger T., Daley G.Q. & Feinberg A.P. (2009) Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat Genet* **41**, 1350-3.

Fisher R.A.C.F.p.d.J. (1922) On the Interpretation of χ² from Contingency Tables, and the Calculation of P. **85**, 87-94.

He Y., Ding Y., Zhan F., Zhang H., Han B., Hu G., Zhao K., Yang N., Yu Y., Mao L. & Song J. (2015) The conservation and signatures of lincRNAs in Marek's disease of chicken. *Sci Rep* **5**, 15184.

Li C., Guo S., Zhang M., Gao J. & Guo Y. (2015) DNA methylation and histone modification patterns during the late embryonic and early postnatal development of chickens. *Poult Sci* **94**, 706-21.

Lienhard M., Grimm C., Morkel M., Herwig R. & Chavez L. (2014) MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. *Bioinformatics* **30**, 284-6.

Melisi D., Ishiyama S., Sclabas G.M., Fleming J.B., Xia Q., Tortora G., Abbruzzese J.L. & Chiao P.J. (2008) LY2109761, a novel transforming growth factor beta receptor type I and type II dual inhibitor, as a therapeutic approach to suppressing pancreatic cancer metastasis. *Mol Cancer Ther* **7**, 829-40.

Mitra A. & Song J. (2012) WaveSeq: a novel data-driven method of detecting histone modification enrichments using wavelets. *PLoS One* **7**, e45486.

Pinello L., Xu J., Orkin S.H. & Yuan G.C. (2014) Analysis of chromatin-state plasticity identifies cell-type-specific regulators of H3K27me3 patterns. *Proc Natl Acad Sci U S A* **111**, E344-53.

Ross-Innes C.S., Stark R., Teschendorff A.E., Holmes K.A., Ali H.R., Dunning M.J., Brown G.D., Gojis O., Ellis I.O., Green A.R., Ali S., Chin S.F., Palmieri C., Caldas C. & Carroll J.S. (2012) Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389-93.

Saxonov S., Berg P. & Brutlag D.L. (2006) A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A* **103**, 1412-7.

Vanderkraats N.D., Hiken J.F., Decker K.F. & Edwards J.R. (2013) Discovering high-resolution patterns of differential DNA methylation that correlate with gene expression changes. *Nucleic Acids Res* **41**, 6816-27.

Wang J.H., Zhao L.F., Lin P., Su X.R., Chen S.J., Huang L.Q., Wang H.F., Zhang H., Hu Z.F., Yao K.T. & Huang Z.X. (2014) GenCLiP 2.0: a web server for functional clustering of genes and construction of molecular networks based on free terms. *Bioinformatics* **30**, 2534-6.

Yu P.B., Deng D.Y., Lai C.S., Hong C.C., Cuny G.D., Bouxsein M.L., Hong D.W., McManus P.M., Katagiri T., Sachidanandan C., Kamiya N., Fukuda T., Mishina Y., Peterson R.T. & Bloch K.D. (2008) BMP type I receptor inhibition reduces heterotopic [corrected] ossification. *Nat Med* **14**, 1363-9.