V9: Epigenetic landscape during early development

Embryonic development is a complex process that remains to be understood despite knowledge of the complete genome sequences of many species and rapid advances in genomic technologies.

A fundamental question is how the unique gene expression pattern in each cell type is established and maintained during embryogenesis.

It is well accepted that the gene expression program encoded in the genome is executed by transcription factors that bind to cis-regulatory sequences and modulate gene expression in response to environmental cues.

Xie et al., Cell 153, 1134-1148 (2013)

Epigenetic marks control cellular memory

Growing evidence now shows that maintenance of such **cellular memory** depends on **epigenetic marks** such as DNA methylation and chromatin modifications

DNA methylation at promoters has been shown to silence gene expression and thus has been proposed to be necessary for lineage-specific expression of developmental regulatory genes, genomic imprinting, and X chromosome inactivation.

Indeed, the DNA methyltransferase DNMT1 or DNMT3a/3b **double-knockout** mice exhibit severe defects in embryogenesis and die before midgestation, supporting an essential role for DNA methylation in embryonic development

Xie et al., Cell 153, 1134-1148 (2013)

Survival without DNMTs?

On the other hand, mouse embryonic stem cells (mESCs) lacking all 3 DNMTs can survive and self-renew and can even begin to differentiate to some germ layers

This raises the possibility that DNA methylation is dispensable for at least initial lineage specification in early embryos.

Thus, the role of DNA methylation in animal development needs to be more precisely defined.

Xie et al., Cell 153, 1134-1148 (2013)

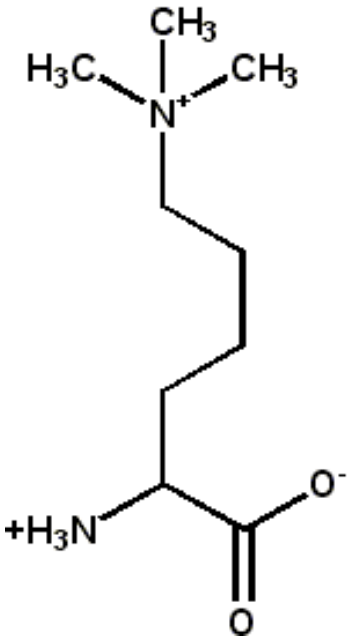
Epigenetic landscape during early development

Like DNA methylation, **chromatin modifications** have also been shown to play a key role in animal development.

In particular, enzymes responsible for **methylating histone H3 at lysines 4, 9, and 27** are essential for embryogenesis.

Nomenclature:
"H3K4me3" stands for histone 3, lysine 4, tri-methylation

Although both DNA methylation and chromatin modifications are critical for mammalian development, the exact role of each epigenetic mark in the maintenance of lineage-specific gene expression patterns remains to be defined



Tri-methylated lysine

Xie et al., Cell 153, 1134-1148 (2013)

www.wikipedia.org

Epigenetic landscape during early development

Human embryonic stem cells (hESCs) can nowadays be differentiated into a variety of precursor cell types, providing an *in vitro* model system for studying early human developmental decisions.

There exist protocols for differentiation of hESCs to various cell states, including

- trophoblast-like cells (TBL),
- mesendoderm (ME),
- neural progenitor cells (NPCs), and
- mesenchymal stem cells (MSCs).

The first 3 states (TBL, ME, NPC) represent developmental events that mirror critical developmental decisions in the embryo: the decision to become embryonic or extraembryonic, the decision to become mesendoderm or ectoderm, and the decision to become surface ectoderm or neuroectoderm, respectively.

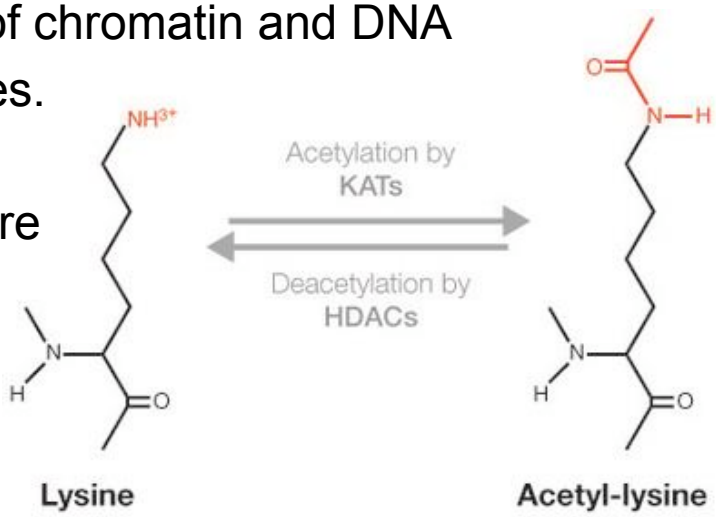
MSCs are fibroblastoid cells that are capable of multilineage differentiation to bone, cartilage, adipose, muscle, and connective tissues

Xie et al., Cell 153, 1134-1148 (2013)

Epigenetic landscape during early development

Several groups have reported genome-wide maps of chromatin and DNA methylation in pluripotent and differentiated cell types.

From these efforts, a global picture of the architecture and regulatory dynamics is beginning to emerge.



Active promoters contain modifications such as **H3K4me3** and **H3K27ac**.
Active enhancers are enriched for **H3K4me1** and **H3K27ac**.

Repressed loci exhibit enrichment for **H3K27me3**, **H3K9me2/3**, **DNAme**, or a combination of the latter two modifications.

The enrichment of repressive histone modifications, such as H3K27me3, which is initiated at **CpG islands (CGI),** is considered a facultative state of repression.

DNAme is generally considered a more stable form of epigenetic silencing.

Gifford et al., Cell 153, 1149-1163 (2013)

Epigenetic landscape during early development

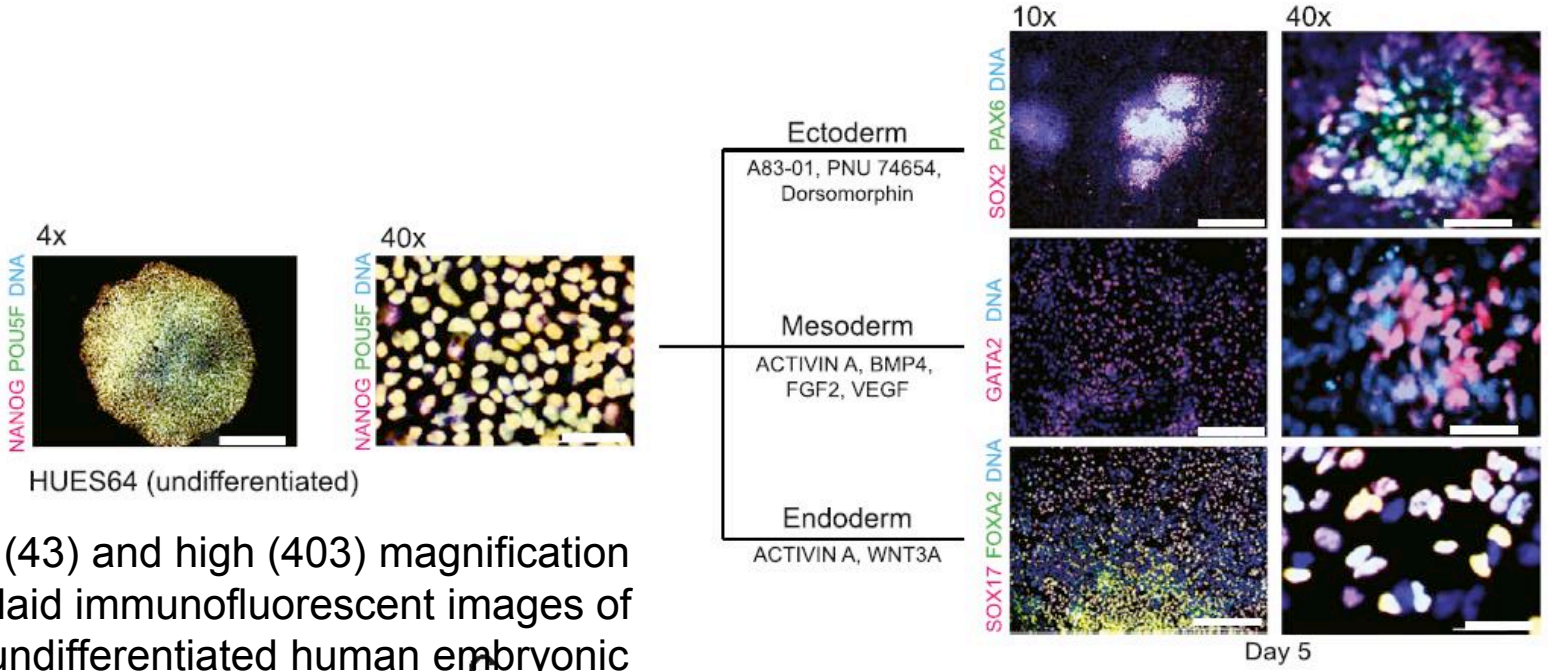
To dissect the early transcriptional and epigenetic events during hESC specification, Gifford et al. used directed differentiation of hESCs to produce early representative populations from the 3 germ layers, namely **ectoderm**, **mesoderm**, and **endoderm**, followed by fluorescence-activated cell sorting (FACS) to enrich for the desired differentiated populations.

These 3 cell types, in addition to **undifferentiated hESCs** (HUES64), were then subjected to
- ChIP-seq for six histone marks (H3K4me1, H3K4me3, H3K27me3, H3K27ac, H3K36me3, and H3K9me3),
- whole-genome bisulfite sequencing (to determine DNA methylation status), and
- RNA sequencing (RNAseq).

We also performed ChIP-seq for the TFs OCT4, SOX2, and NANOG in the undifferentiated hESCs, as well as ChIP bisulfite sequencing (ChIP-BS-seq) for FOXA2 in the endoderm population.

Gifford et al., Cell 153, 1149-1163 (2013)

Generation of hESCs and hESC-derived cell types



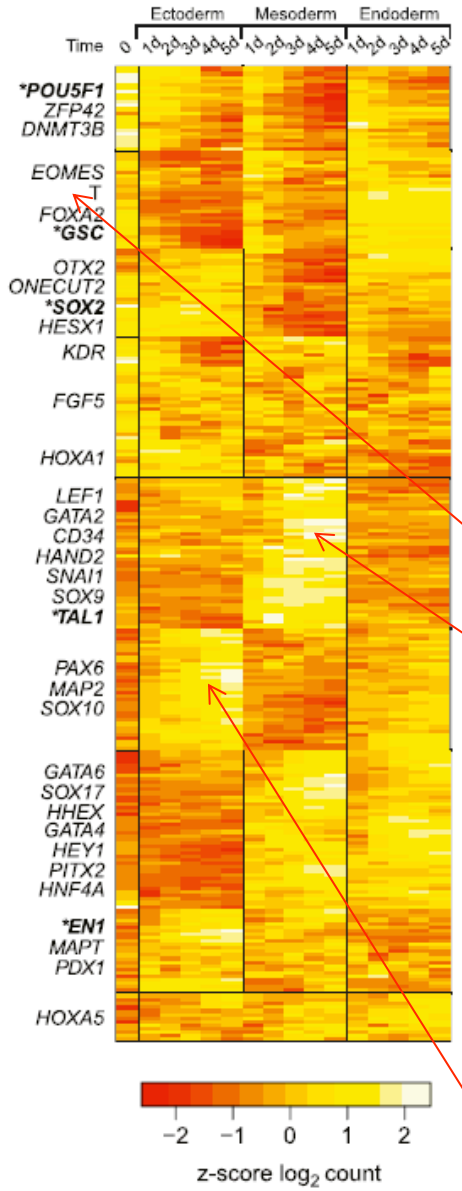
Low (43) and high (403) magnification overlaid immunofluorescent images of the undifferentiated human embryonic stem cell (hESC) line HUES64 stained with OCT4 (POU5F1) and NANOG antibodies.

Formation of ectoderm is induced by inhibition of TGFb, Wingless/ integrase1 (WNT), and bone morphogenetic protein (BMP) signaling

Established directed differentiation conditions were used to generate representative populations of the 3 embryonic germ layers: hESC-derived ectoderm, hESC-derived mesoderm, and hESC-derived endoderm. Cells were fixed and stained after 5 days of differentiation with the indicated antibodies. Representative overlaid images at low and high magnification are shown. DNA was stained with Hoechst 33342 in all images.

Gifford et al., Cell 153, 1149-1163 (2013)

Gene expression in 3 cell lineages



Z score $\log_2$ expression values during 5 days of in vitro differentiation. 268 out of 541 profiled genes changed by more than 0.5.

Z-score $$z = \frac{x - \mu}{\sigma}$$

μ : mean of population;

σ : standard deviation of population.

Selected lineage-specific genes are shown for each category that was identified based on hierarchical clustering.

Genes such as EOMES, T, FOXA2, and GSC are upregulated at 24 hr of mesoderm and endoderm induction, but not ectoderm differentiation.

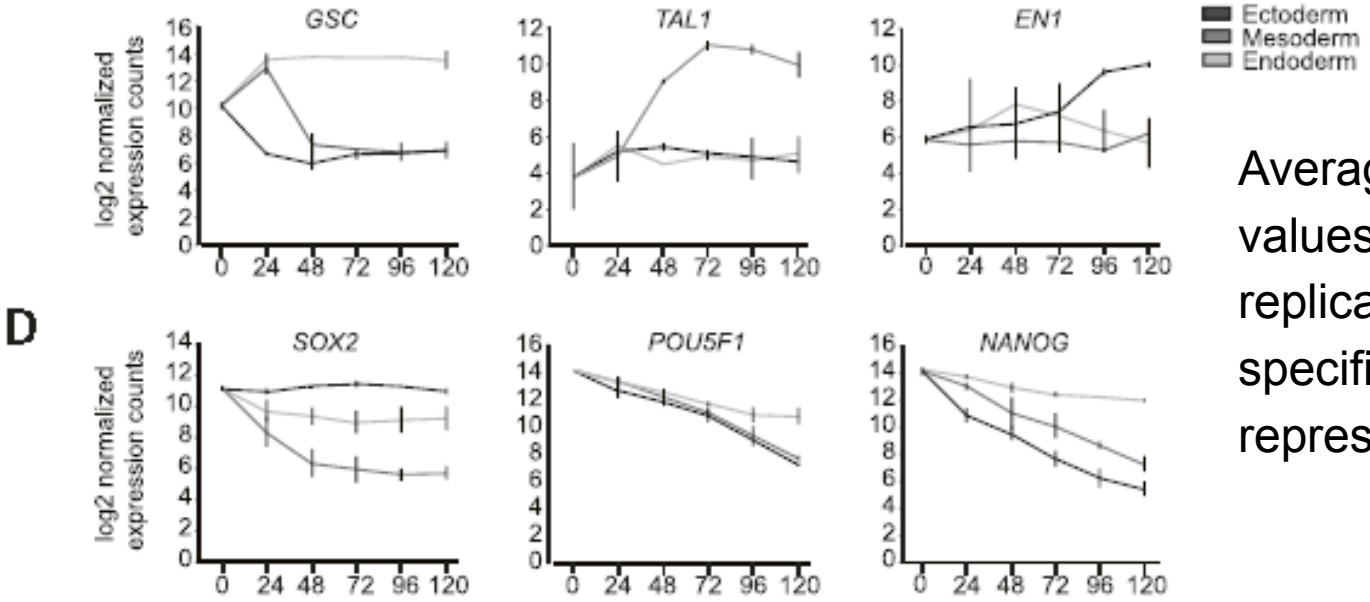
GSC expression decreases within 48 hr of differentiation in the mesoderm-like population, whereas the expression level is maintained in the **endoderm population**. EOMES and FOXA2 expression is also maintained in the endoderm population accompanied by upregulation of GATA6, SOX17, and HHEX.

After transient upregulation of mesendodermal markers, activation of mesodermal markers such as GATA2, HAND2, SOX9, and TAL1 is detected specifically in the **mesoderm conditions**.

None of these markers are detected during early **ectoderm differentiation**, which instead upregulates neural markers such as PAX6, SOX10, and EN1

Gifford et al., Cell 153, 1149-1163 (2013)

Gene expression of known pluripotency markers



Average $\log_2$ expression values of two biological replicates of lineage-specific genes. Error bars represent 1 SD.

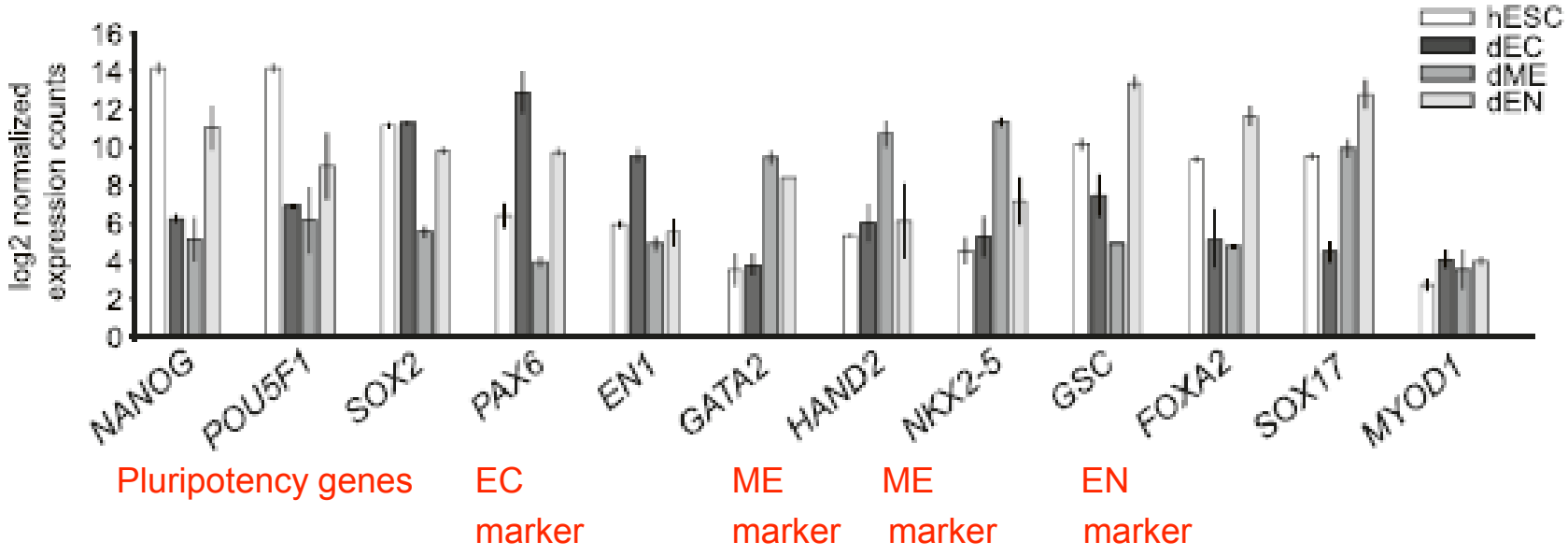
Yamanaka factors (for cell reprogramming): Oct4 (Pou5f1), Sox2, cMyc, and Klf4

POU5F1 (OCT4), NANOG, and, to some extent, SOX2 expression is maintained in the endoderm population.

SOX2 expression is downregulated in mesoderm and— to a lesser degree—in endoderm but is maintained at high levels in the ectoderm population.

Gifford et al., Cell 153, 1149-1163 (2013)

Gene expression in 3 cell lineages



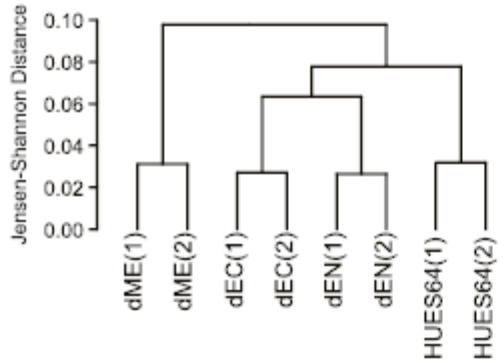
profiling of FACS-isolated ectoderm (dEC), mesoderm (dME), and endoderm (dEN).

Expression levels for MYOD1 (right) are included as a negative control.

Day 5 was selected as the optimal time point to capture early regulatory events in well-differentiated populations representing all three germ layers.

Gifford et al., Cell 153, 1149-1163 (2013)

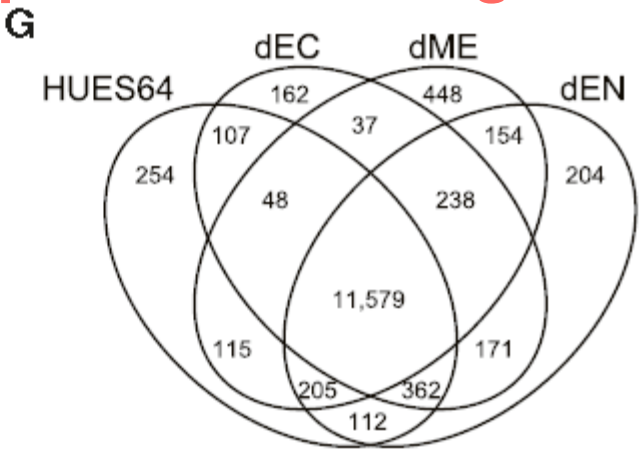
Relationship between lineages



Hierarchical clustering of global gene expression profiles for HUES64 and dEC, dME, and dEN shown as a **dendrogram**.

The **dME** population is the **most distantly** related cell type.

dEN and dEC are more similar to each other than to dME or hESCs



Venn diagram illustrating unique and overlapping genes with expression.

dME population expresses the largest number of unique genes (n = 448), such as RUNX1 and HAND2.

dEC and dME have the least transcripts in common (n = 37), whereas dEC and dEN have most transcripts in common (n = 171),

Gifford et al., Cell 153, 1149-1163 (2013)

Chromatin states

Analyze previously identified **informative chromatin states**

- H3K4me3+H3K27me3 (bivalent/poised promoter);
  „Poised“ genes: RNA-Polymerase II is located at their promoters in the absence of ongoing transcription, the genes are loaded to be transcribed soon
- H3K4me3+H3K27ac (active promoter); gene is actively transcribed
- H3K4me3 (initiating promoter);
- H3K27me3+H3K4me1 (poised developmental enhancer);
- H3K4me1 (poised enhancer);
- H3K27ac+H3K4me1 (active enhancer); and
- H3K27me3 (Polycomb repressed); and
- H3K9me3 (heterochromatin).

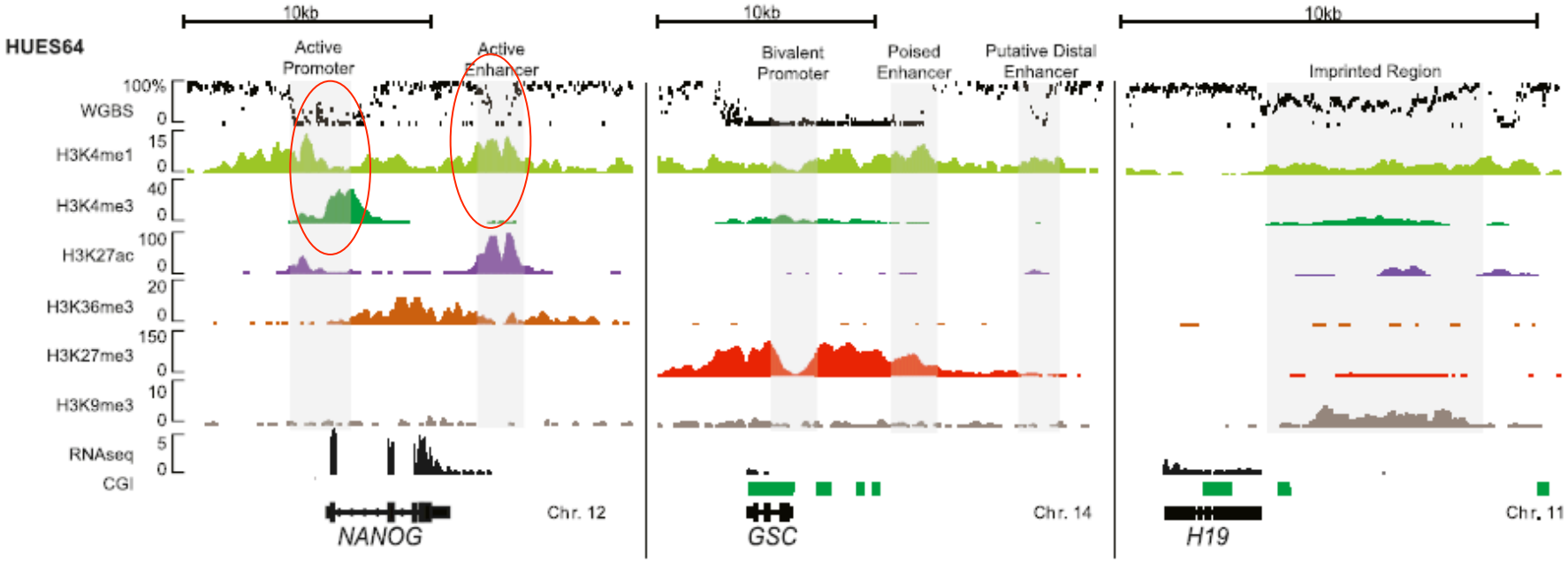
The WGBS data was segmented into three DNAme states:

- highly methylated regions (HMRs: > 60%),
- intermediately methylated regions (IMRs: 11%– 60%), and
- unmethylated regions (UMRs: 0%–10%).

Gifford et al., Cell 153, 1149-1163 (2013)

Epigenetic Data for hESC

One allele fully methylated,
other allele unmethylated
-> gene appears half methylated



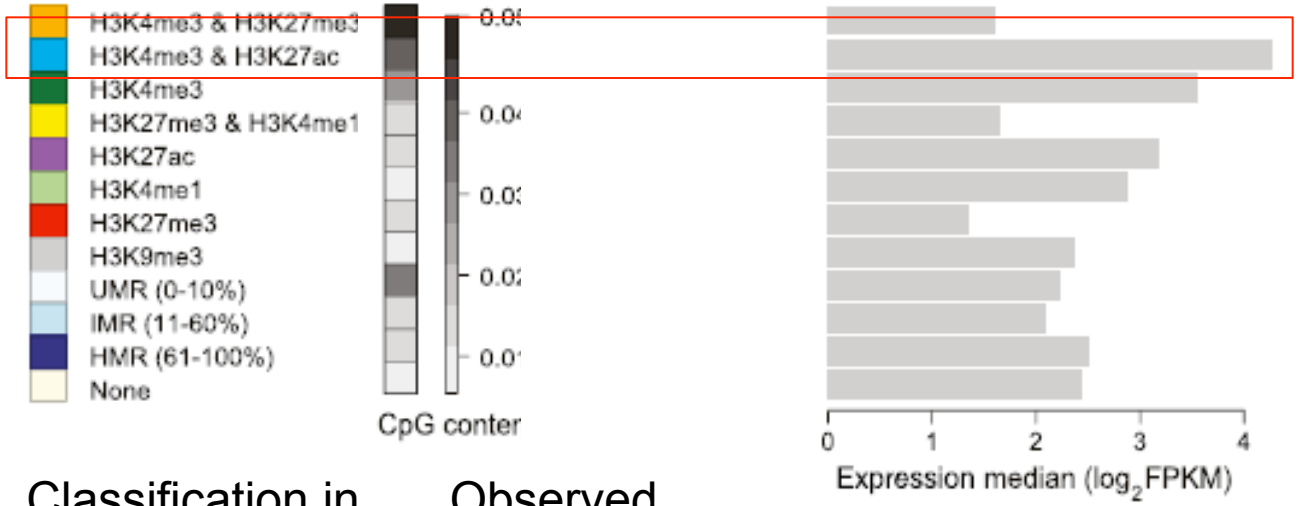
Data for the undifferentiated hESC line HUES64 at 3 loci: NANOG, GSC, and H19

WGBS (% methylation), ChIP-seq (read count normalized to 10 million reads), and RNA-seq (FPKM = fragments per kilobase of exon per million fragments mapped). **CpG islands** are indicated in **green**.

Same data was also collected for dEC, dME, and dEN cells (ca. 12 million cells each)

Gifford et al., Cell 153, 1149-1163 (2013)

Epigenetics linked to expression



Classification in distinct epige-netic states:

The combination of H3K4me3 and H3K27me3 exhibits the highest CpG content.

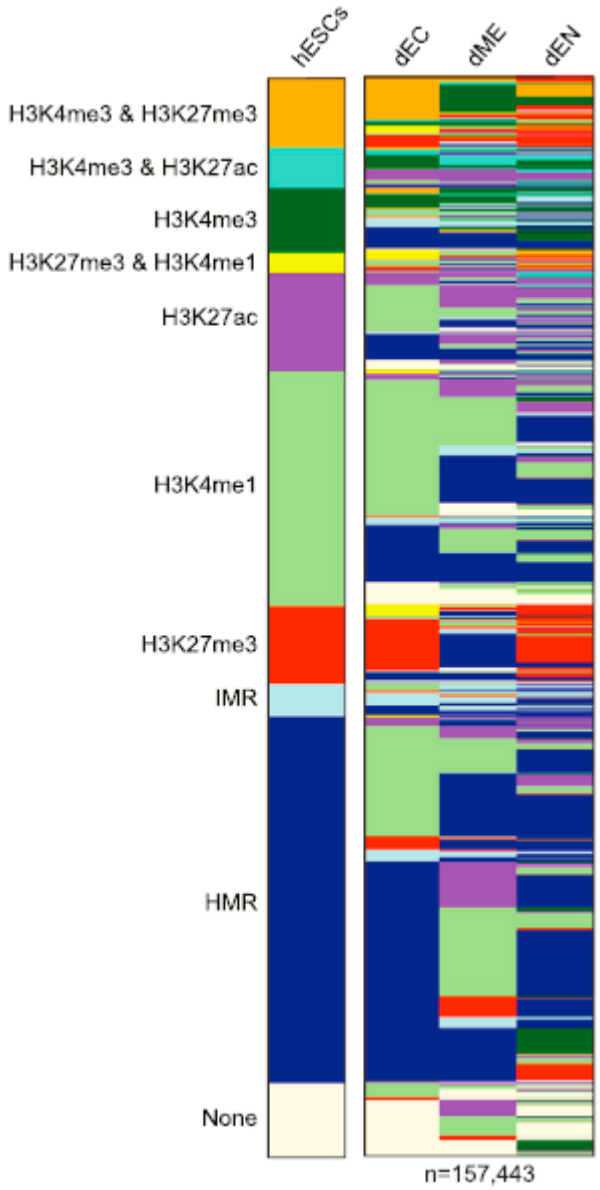
Observed median CpG content of genomic regions in states defined on the left

Right: Median expression level of epigenetic states based on assignment of each region to the nearest RefSeq gene. Regions of open chromatin (active promoter; H3K4me3 & H3K27ac) have highest expression.

Note that many (ca. 65%) epigenetic remodeling events are not directly linked to transcriptional changes based on the expression of the nearest gene.

Gifford et al., Cell 153, 1149-1163 (2013)

Regions changing their epigenetic state



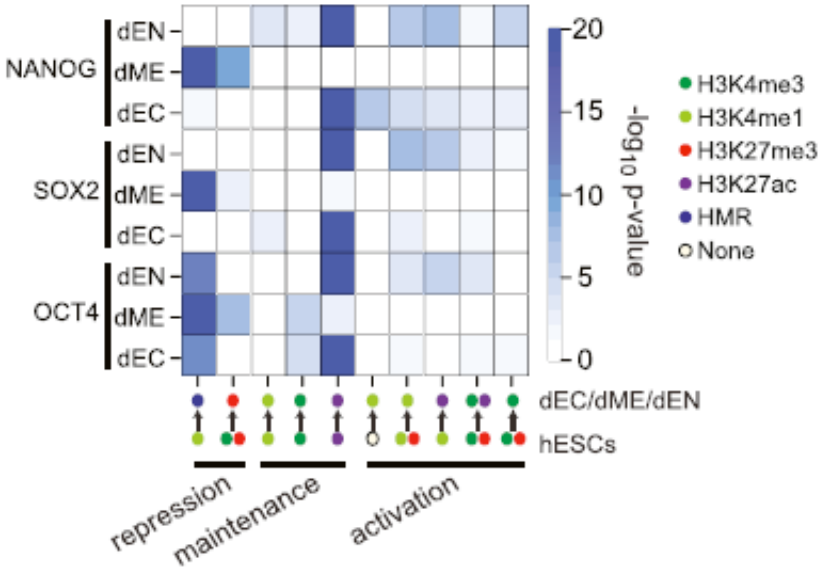
Epigenetic state map of regions enriched for one of 4 histone modifications in at least one cell type or classified as UMR/IMR in at least one cell type and changing its epigenetic state upon differentiation in at least one cell type.

Loss of H3K4 methylation (me1 and me3) is commonly associated with a transition to high DNAme, which is most prominent in the dEN population and genes involved in neural development.

We identified 4,639 proximal bivalent domains in hESCs and observed that 3,951 (85.1%) of these domains resolve their bivalent state in at least one hESC-derived cell type.

Gifford et al., Cell 153, 1149-1163 (2013)

Pluripotent TF binding linked to chromatin dynamics



H3K4me1 regions enriched for OCT4 binding sites frequently become HMRs in all three differentiated cell types, whereas NANOG and SOX2 sites are more prone to change to an HMR state in dME. In general, many regions associated with open chromatin that are bound by NANOG are more likely to retain this state in dEN compared to dME and dEC. We also found that regions enriched for H3K27ac in hESCs that maintain this state in dEN or dEC are likely to be bound by SOX2 and NANOG.

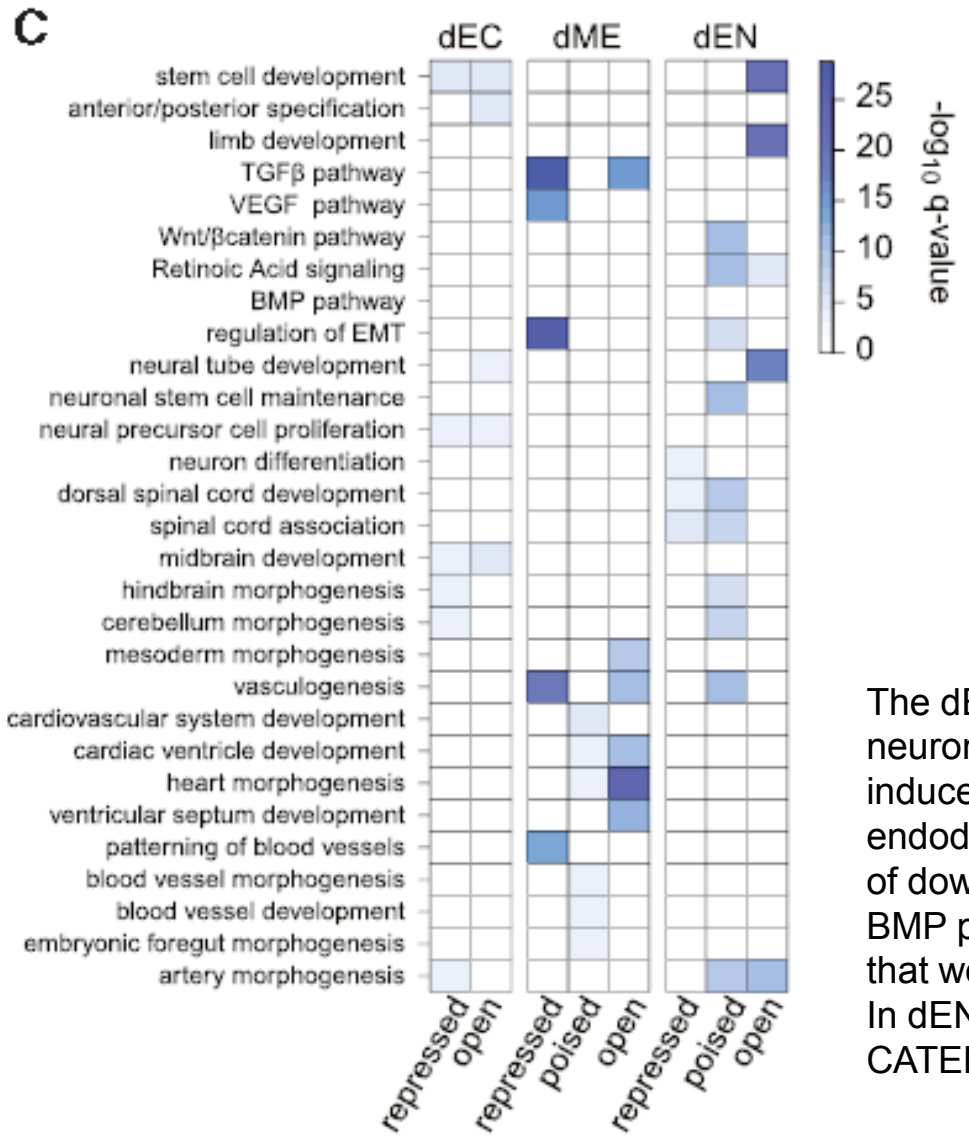
Enrichment of OCT4, SOX2, and NANOG within various classes of dynamic genomic regions that change upon differentiation of hESC.

Values are computed relative to all regions exhibiting the particular epigenetic state change in other cell types.

Epigenetic dynamics are categorized into three major classes: repression (loss of H3K4me3 or H3K4me1 and acquisition of H3K27me3 or DNAme), maintenance of open chromatin marks (H3K4me3, H3K4me1, and H3K27ac), and activation of previously repressed states.

Gifford et al., Cell 153, 1149-1163 (2013)

GO categories in regions gaining H3K27ac



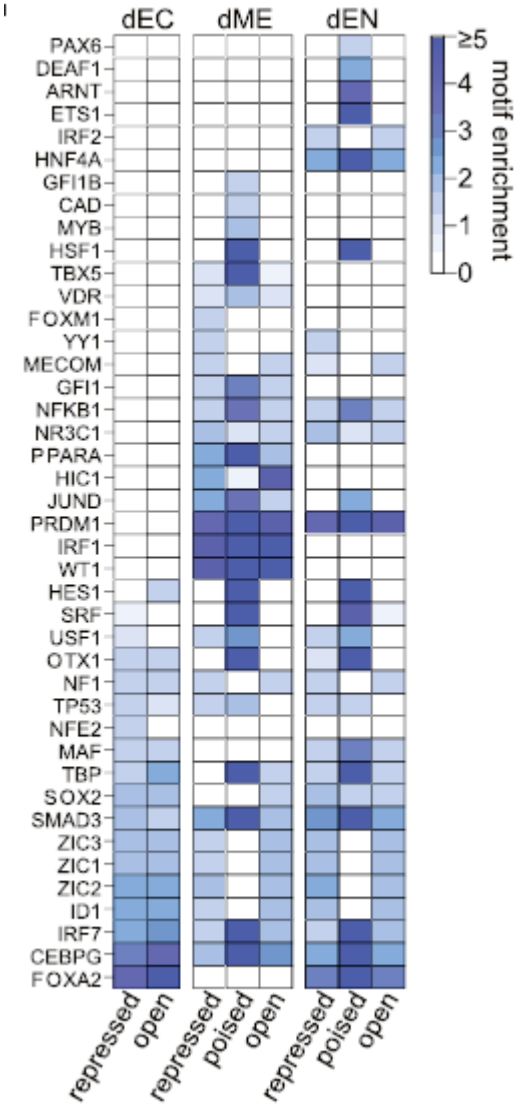
Regions gaining H3K27ac were split up by state of origin in hESC into repressed (none, IMR, HMR, and HK27me3), poised (H3K4me1/ H3K27me3), and Open (H3K4me3/ H3K27me3, H3K4me3, and H3K4me1).

Color code indicates multiple testing adjusted q value of category enrichment.

The dEN population shows an enrichment for early neuronal genes. This suggests that similar networks are induced in the early stages of both our ectoderm and endoderm specification. In dME, We find strong enrichment of downstream effector genes of the TGFb, VEGF, and BMP pathways, directly reflecting the signaling cascades that were stimulated to induce the respective differentiation. In dEN, we find enrichment of genes involved in WNT/b-CATENIN and retinoic acid (RA) signaling.

Gifford et al., Cell 153, 1149-1163 (2013)

TF motifs enriched in regions changing to H3K27ac



Color code indicates motif enrichment score .

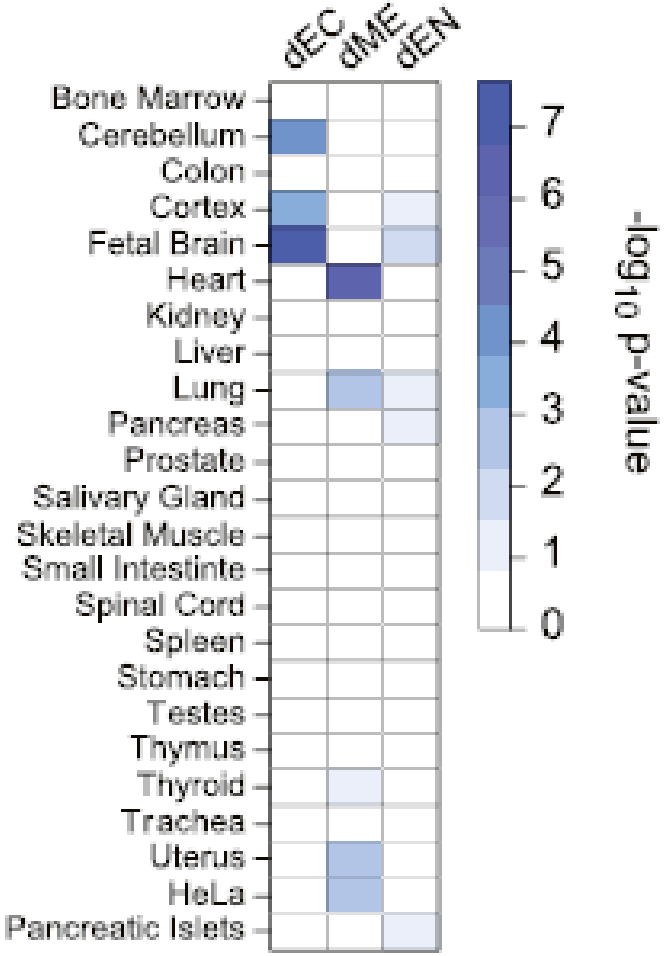
For each region class, the 8 highest-ranking motifs are shown.

We detected high levels of SMAD3 motif enrichment in the repressed dME and dEN, particularly in the poised putative enhancer populations. Similarly, we observe enrichment of key lineage-specific TF motifs such as the ZIC family proteins in dEC, TBX5 in dME, and SRF in dEN.

Interestingly, we also find the FOXA2 motif highly over-represented in dEN—in which the factor is active, and also dEC, in which the factor is inactive but becomes expressed at a later stage of neural differentiation, but not in dME.

Gifford et al., Cell 153, 1149-1163 (2013)

Tissue signature enrichment levels



Tissue signature enrichment levels of genes assigned to regions specifically gaining H3K4me1.

Regions that gain H3K4me1 in dEC are associated with fetal brain and specific cell types found within the adult brain.

The dME H3K4me1 pattern was associated with avrange of interrogated tissues, such as heart, spinal cord, andvstomach, which may be due to heterogeneity of the tissues collected.

The dEN associations were interesting given that, as with the RNA-seq and H3K27ac trends, H3K4me1 was again associated with brain-related categories.

Gifford et al., Cell 153, 1149-1163 (2013)

Transcriptional and Epigenetic Dynamics during Specification of Human Embryonic Stem Cells

Casey A. Gifford,[1,2,3,9] Michael J. Ziller,[1,2,3,9] Hongcang Gu,[1] Cole Trapnell,[1,3] Julie Donaghey,[1,2,3] Alexander Tsankov,[1,2,3] Alex K. Shalek,[4] David R. Kelley,[1,3] Alexander A. Shishkin,[1] Robbyn Issner,[1] Xiaolan Zhang,[1] Michael Coyne,[1] Jennifer L. Fostel,[1] Laurie Holmes,[1] Jim Meldrim,[1] Mitchell Guttman,[1] Charles Epstein,[1] Hongkun Park,[4] Oliver Kohlbacher,[5] John Rinn,[1,3,6] Andreas Gnirke,[1] Eric S. Lander,[1,7] Bradley E. Bernstein,[1,8] and Alexander Meissner[1,2,3,*]

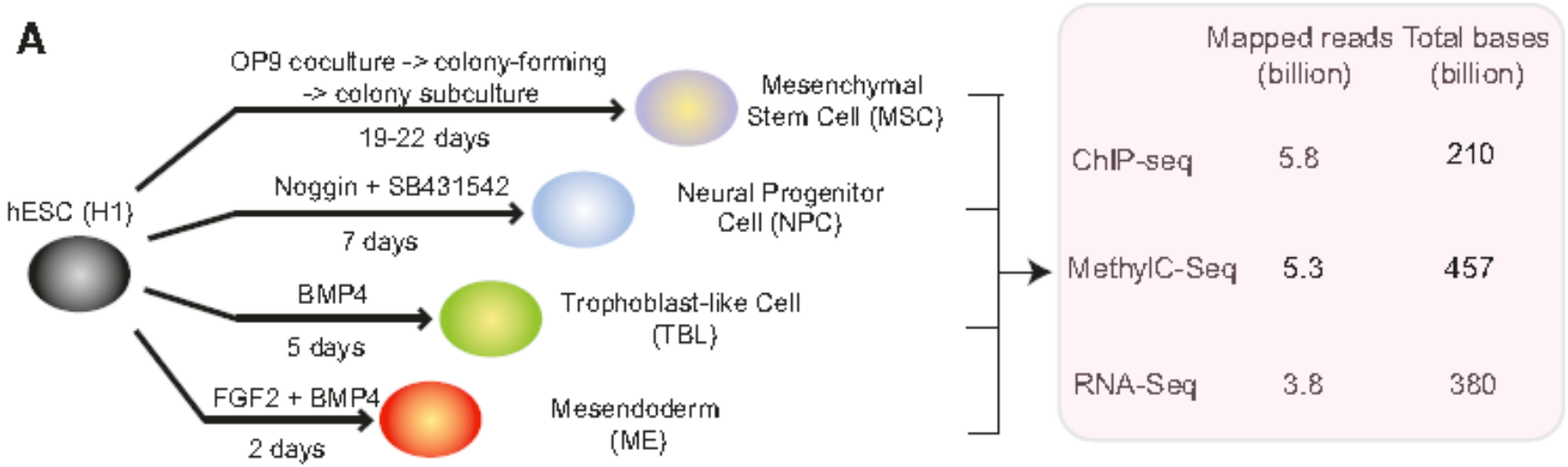
Epigenomic Analysis of Multilineage Differentiation of Human Embryonic Stem Cells

Wei Xie,[1] Matthew D. Schultz,[2] Ryan Lister,[2,14] Zhonggang Hou,[3] Nisha Rajagopal,[1] Pradipta Ray,[11] John W. Whitaker,[4] Shulan Tian,[3] R. David Hawkins,[1,15] Danny Leung,[1] Hongbo Yang,[7] Tao Wang,[4] Ah Young Lee,[1] Scott A. Swanson,[3] Jiuchun Zhang,[3,8] Yun Zhu,[4] Audrey Kim,[1] Joseph R. Nery,[2] Mark A. Urich,[2] Samantha Kuan,[1] Chia-an Yen,[1] Sarit Klugman,[1] Pengzhi Yu,[3] Kran Suknuntha,[12] Nicholas E. Propson,[3] Huaming Chen,[2] Lee E. Edsall,[1] Ulrich Wagner,[1] Yan Li,[1] Zhen Ye,[1] Ashwinikumar Kulkarni,[11] Zhenyu Xuan,[11] Wen-Yu Chung,[11,16] Neil C. Chi,[7] Jessica E. Antosiewicz-Bourget,[3] Igor Slukvin,[8,9,12] Ron Stewart,[3] Michael Q. Zhang,[11,13] Wei Wang,[4,6] James A. Thomson,[3,9,10,*] Joseph R. Ecker,[2,*] and Bing Ren[1,5,*]

[1]Ludwig Institute for Cancer Research, La Jolla, CA 92093, USA

[2]Genomic Analysis Laboratory, Howard Hughes Medical Institute, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA

Xie et al. did practically “the same thing”

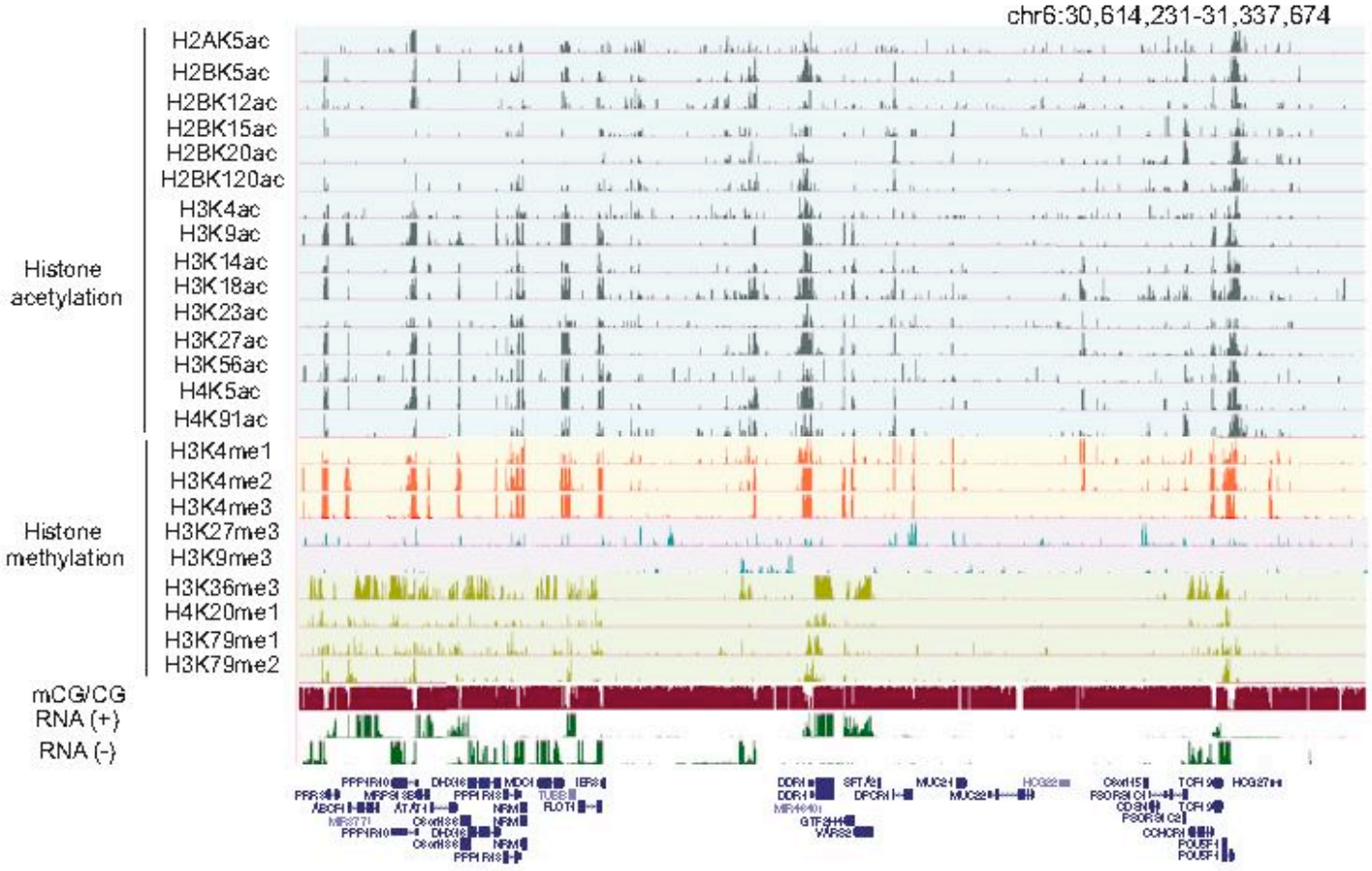


The hESC line H1 was differentiated to ME, TBL, NPCs, and MSCs.
ME, TBL, and NPC differentiation occurred quickly (2 days, 5 days, and 7 days, respectively) compared to that of MSC (19–22 days).

For each cell type, DNA methylation was mapped at base resolution using MethylC-seq (20–353 total genome coverage or 10–17.53 coverage per strand). We also mapped the genomic locations of 13–24 chromatin modifications by chromatin immunoprecipitation sequencing (ChIP-seq). Additionally, we performed paired-end (100 bp 3 2) RNA-seq experiments, generating more than 150 million uniquely mapped reads for every cell type.

Xie et al., Cell 153, 1134-1148 (2013)

Epigenetic marks of H1 cells



A snapshot of the UCSC genome browser showing the DNA methylation level (mCG/CG), RNAseq reads (+, Watson strand; , Crick strand), and ChIP-seq reads (RPKM) of 24 chromatin marks in H1.

Xie et al., Cell 153, 1134-1148 (2013)

Identify lineage-restricted genes

How is the genome differentially transcribed when hESCs are differentiated into each cell type?

→ Examine the expression of 19,056 RefSeq coding genes (33,797 isoforms). 76.6% (14,595) were expressed in at least one cell type.

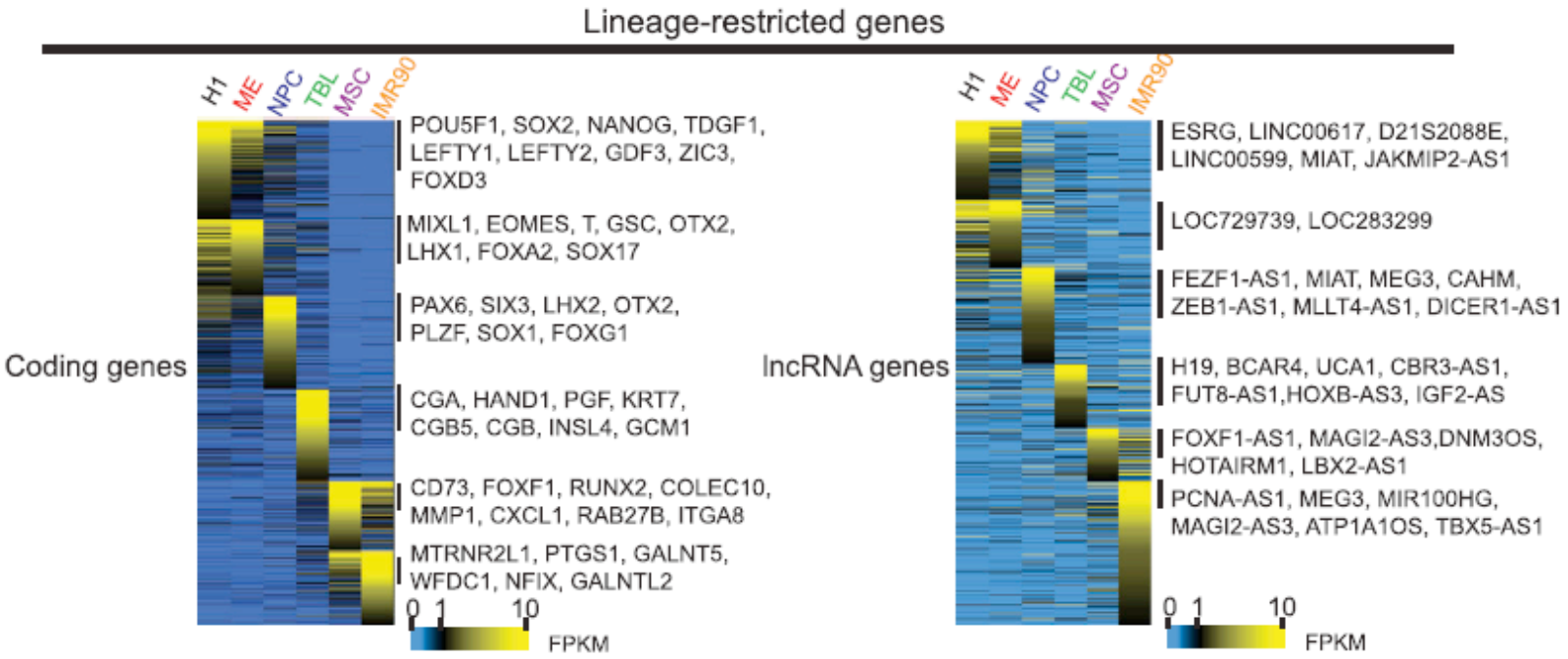
Using an entropy-based method, we identified 2,408 genes that showed cell-type-specific expression.

For convenience, we use "lineage-restricted genes" to reflect both H1-specific and differentiated cell-specific genes.

As expected, known lineage markers were highly expressed in the corresponding cell types

Xie et al., Cell 153, 1134-1148 (2013)

Lineage-restricted transcripts

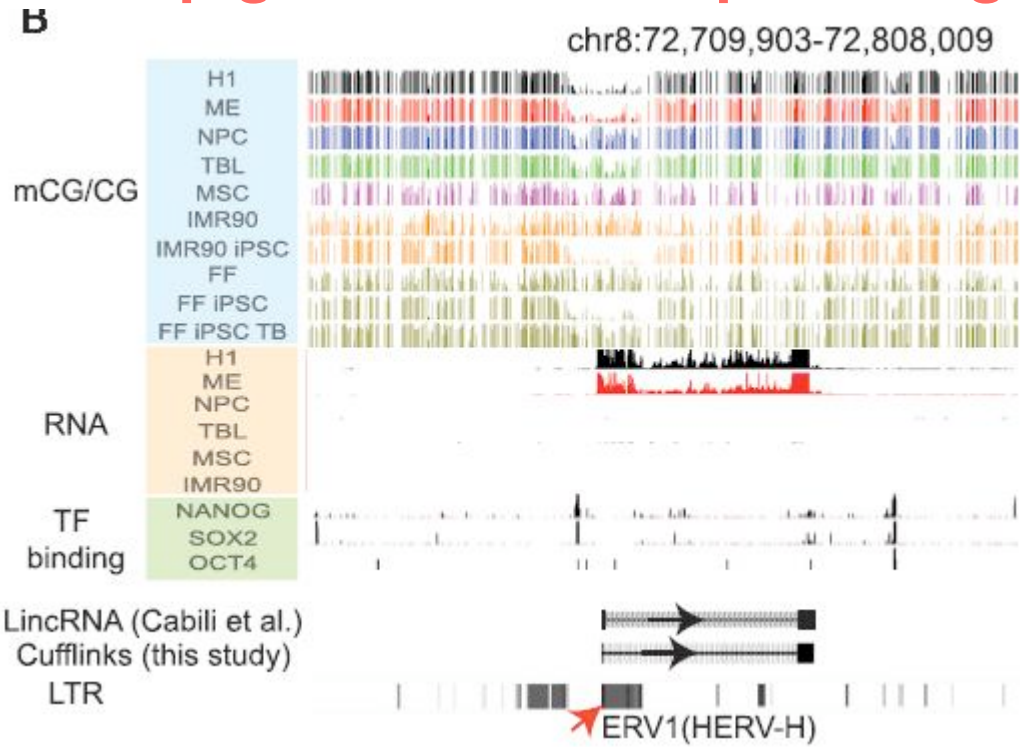


(A) Heatmaps showing the expression levels of lineage-restricted coding genes (left) and lncRNA genes (right). Genes are organized by the lineage in which their expression is enriched.

Certain genes (such as SOX2) can be expressed in more than one cell type.

Xie et al., Cell 153, 1134-1148 (2013)

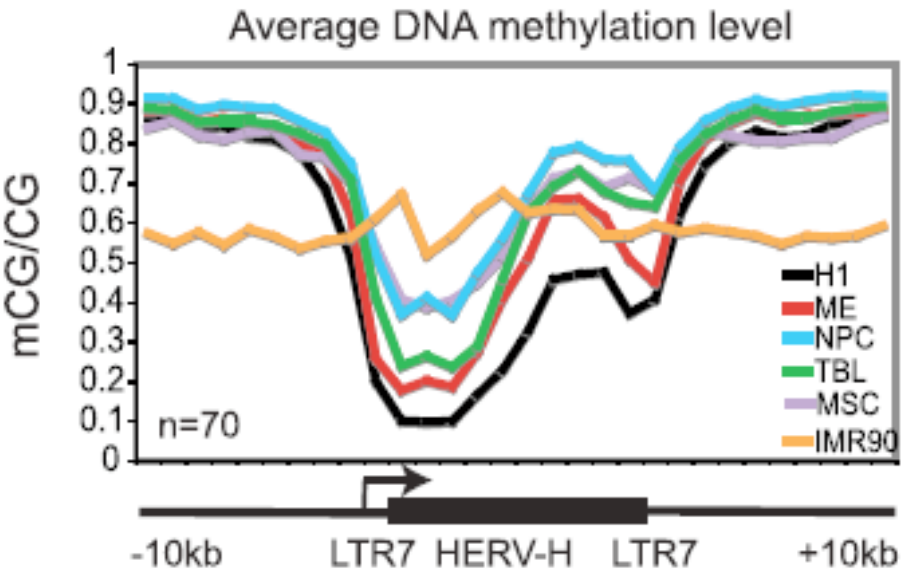
Epigenetic landscape during early development



The levels of DNA methylation and RNA, as well as the binding of NANOG, SOX2, and POU5F1, are shown around an annotated lincRNA gene with the promoter overlapping a HERV-H element.

Xie et al., Cell 153, 1134-1148 (2013)

Role of endoviral insertions



The average DNA methylation level in each cell type is shown for a subset (n=70) of H1-specific HERV-H elements.

Human endogeneous retrovirus (HERV) sequences were inserted into the human germline about 30 million years ago. They cover ca. 8% of the human genome.

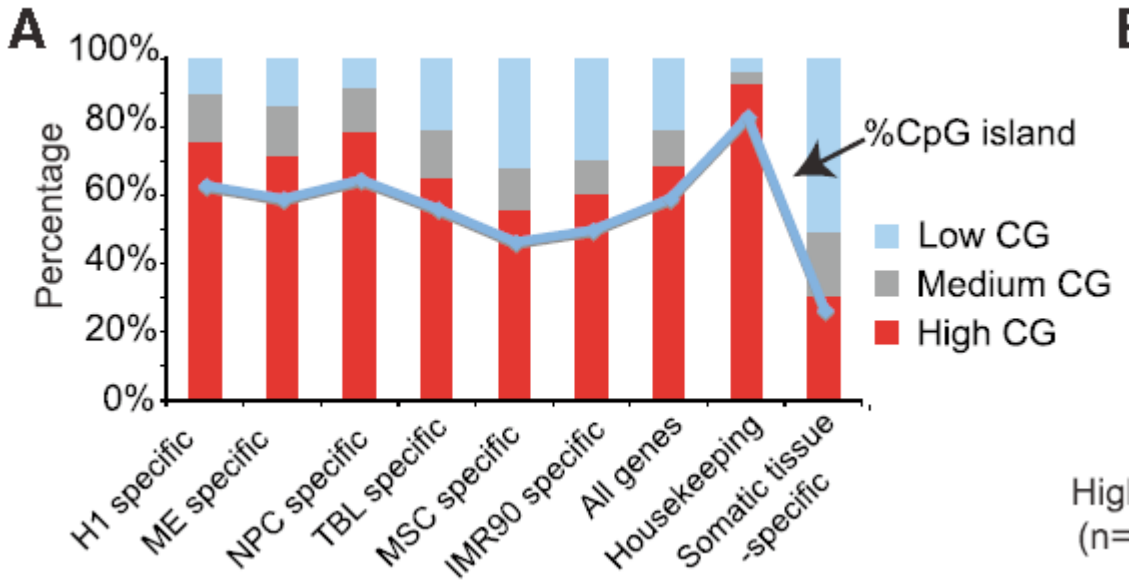
HERV sequences are usually silenced by DNA methylation.

These HERV-H elements show hypomethylation in H1 and ME but gain DNA methylation in other H1-derived cells.

These data suggest that many noncoding RNA genes may be transcriptionally regulated by endogenous retroviral sequences.

Xie et al., Cell 153, 1134-1148 (2013)

Epigenetic regulation of promoters for lineage-restricted genes



Percentages of promoters in the high, medium, and low CG classes for genes that are enriched in each cell type, all RefSeq genes, housekeeping genes, and somatic-tissue-specific genes.

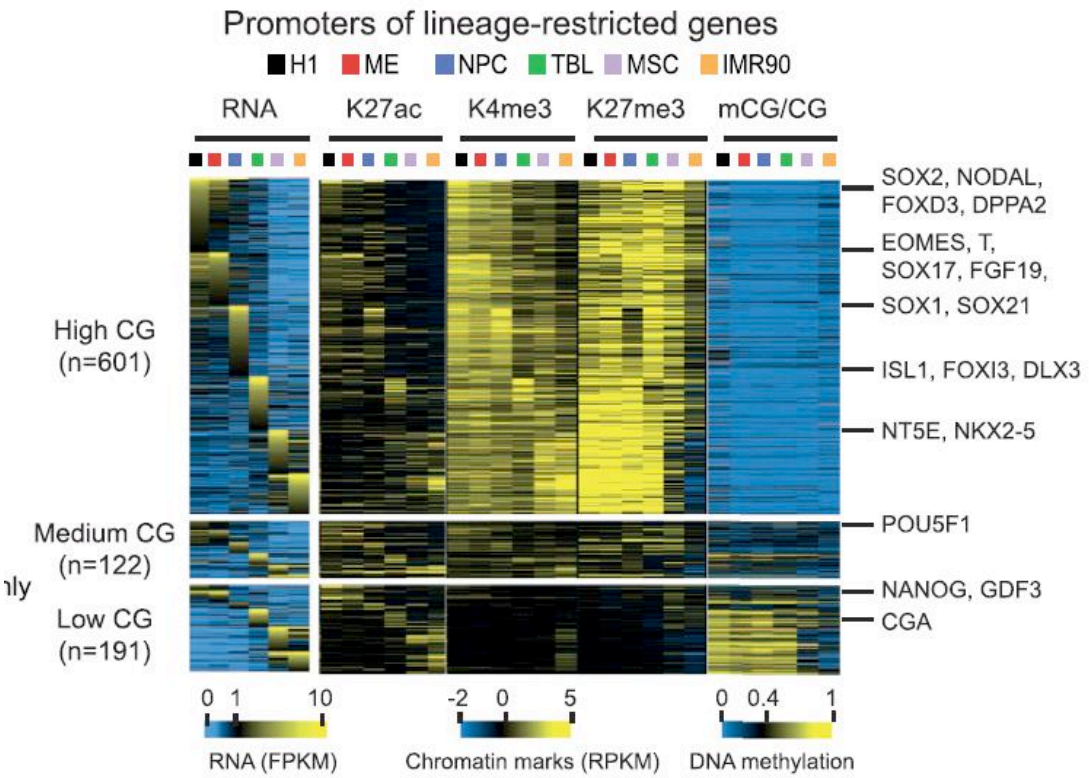
Blue line: percentages of promoters that contain CGIs.

Genes preferentially expressed in early embryonic lineages H1, ME, and NPC tend to be CG rich and contain CGIs. The percentages of CGI-containing promoters decreased for genes enriched in MSCs and IMR90, which are at relatively late development stages.

By contrast, a much lower percentage of promoters (23%) contain CGIs for somatic-tissue-specific genes identified from 18 human tissues.

Xie et al., Cell 153, 1134-1148 (2013)

Epigenetic landscape during early development



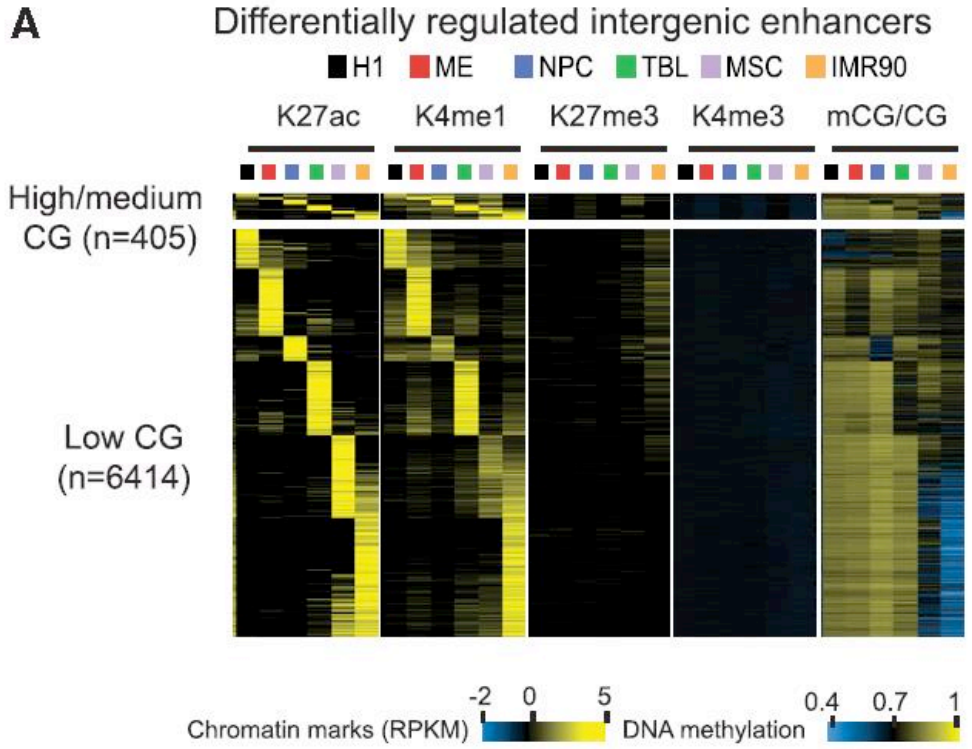
Average levels of RNA, H3K27ac, H3K4me3, H3K27me3, and DNA methylation for promoters of lineage-restricted genes.

Histone modifications, TSS ± 2 kb; DNA methylation, TSS ± 200 bp; promoter CG density, TSS ± 500 bp.

The DNA methylation machinery has been shown to be a mechanism of gene silencing during cell differentiation. In addition, the Polycomb protein complex, which deposits H3K27me3 at target genes, can also repress developmental genes. We set to determine which promoters are subject to regulation by DNA methylation, H3K27me3, or both.
A detailed analysis showed that promoters with high CG density tend to be enriched for H3K27me3, whereas those with low CG density are preferentially marked by DNA methylation

Xie et al., Cell 153, 1134-1148 (2013)

Epigenetic regulation of lineage-restricted enhancers



Heatmaps showing the average levels of H3K27ac, H3K4me1, H3K4me3, H3K27me3, and DNA methylation around the centers of lineage-restricted enhancers.

Histone modifications, enhancer center ± 2 kb; DNA methylation, enhancer center ± 500 bp; CG density, enhancer center ± 500 bp.

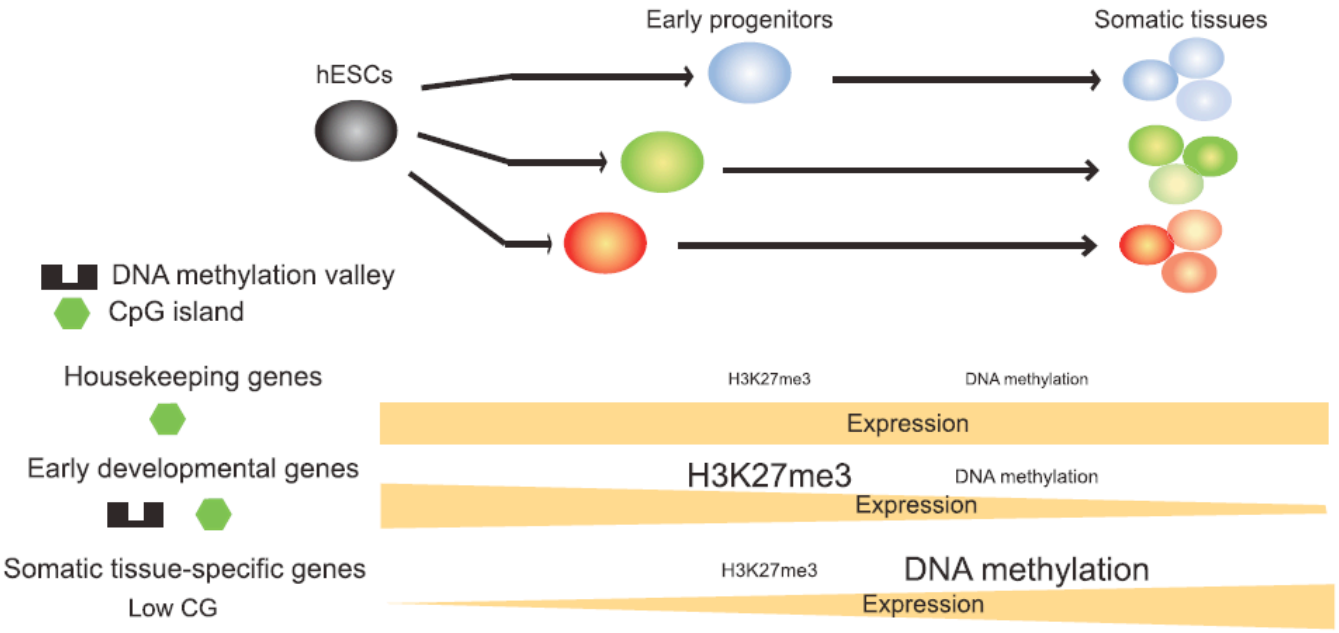
Most enhancers are CG poor (94%) and appear to be depleted of H3K27me3.

(However, weak enrichment of H3K27me3 is observed at a subset of enhancers in MSCs and IMR90.)

These enhancers are largely active in H1, ME, NPCs, and TBL, but not in MSCs and IMR90, as indicated by the levels of H3K27ac.

Xie et al., Cell 153, 1134-1148 (2013)

Model for early development



A model for 3 classes of promoters with distinct sequence features and epigenetic regulation mechanisms in cell differentiation.

The majority of genes differentially expressed in early progenitors are CG rich and appear to employ H3K27me3-mediated repression in nonexpressing cells.

Conversely, genes differentially expressed in later stages are largely CG poor and preferentially show DNA methylation-mediated gene silencing

Xie et al., Cell 153, 1134-1148 (2013)