V10: Epigenetics of stem cells

During development, epigenetic information is acquired in a progressive manner.

These changes regulate the transcriptional programme during **lineage commitment**.

Dynamic regulation of the epigenome underlies **cellular plasticity** and provides a **heritable** response to environmental and developmental cues.

The different layers of epigenetic information are closely **interconnected**.

Epigenetic deregulation is directly linked to a wide spectrum of **diseases** ranging from developmental disorders associated with aberrant genetic imprinting to various cancers that have defects in protein complexes involved in histone or DNA modifications.

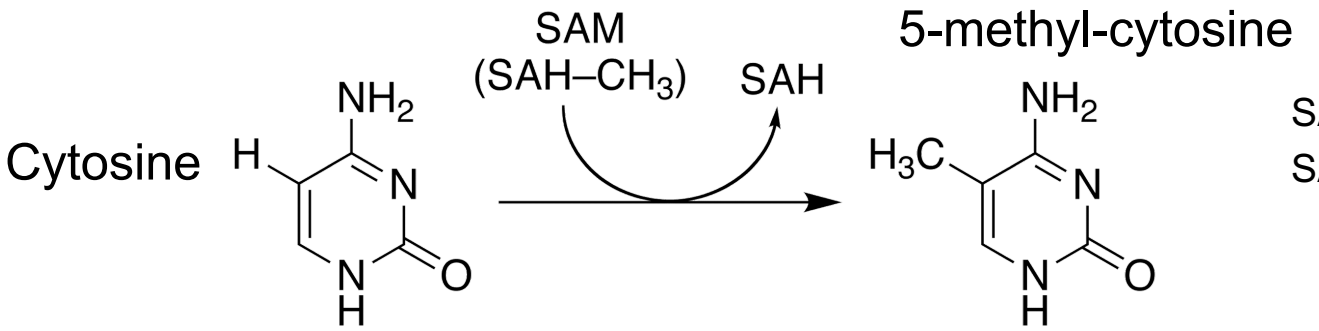
The fact that epigenetic modifications are, in principle, **reversible** renders epigenetic regulation amenable to **pharmacological intervention**.

Atlasi & Stunnenberg, *Nature Rev Genet* **18**, 643–658 (2017)

Review (V9): Cytosine methylation

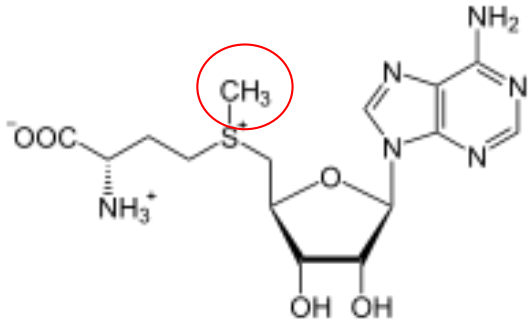
Observation: 3-6 % of all cytosines are methylated in human DNA.

This methylation occurs (almost) exclusively when cytosine is followed by a guanine base -> **CpG dinucleotide**.



SAM: S-adenosyl-methionine

SAH: S-adenosyl-homocysteine



Cytosines in **CpG islands** are usually not methylated.

Esteller, Nat. Rev. Gen. 8, 286 (2007)
www.wikipedia.org

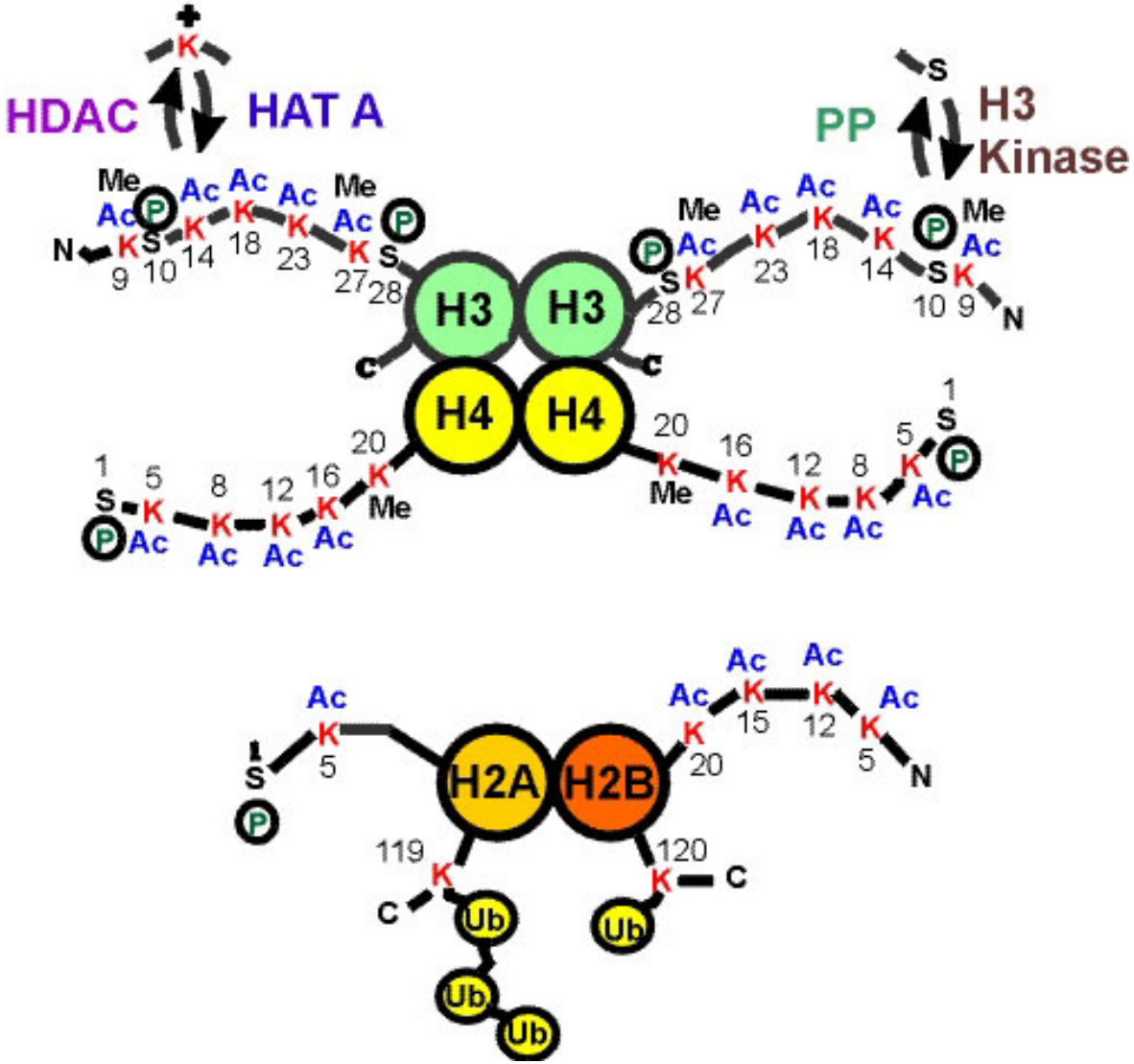
Review (V9): Post-translational modifications of histone tails

The disordered histone tails comprise 25-30% of the histone mass.

They extend from the compact histone multimer to provide a platform for various **post-translational modifications (PTMs)**.

These modifications affect the histones' ability to bind DNA and to other histones.

This, in turn, affects **gene expression.**



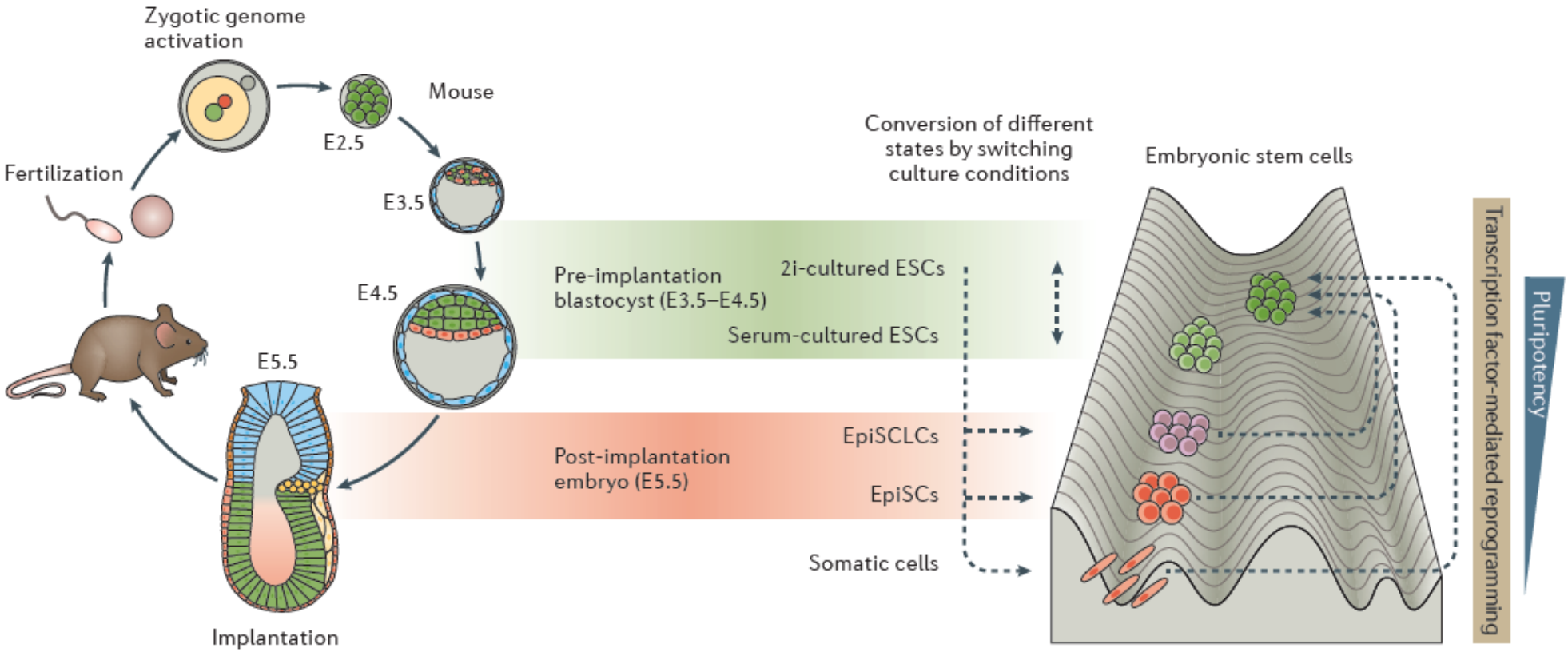
Strahl BD and Allis CD, 2000. Nature 403:41-45

Review (V9): Different states of pluripotency

E4.5 epiblast cells: represent ground-state pluripotency

Implantation: stage of pregnancy at which the blastocyst adheres to the wall of the **uterus**.

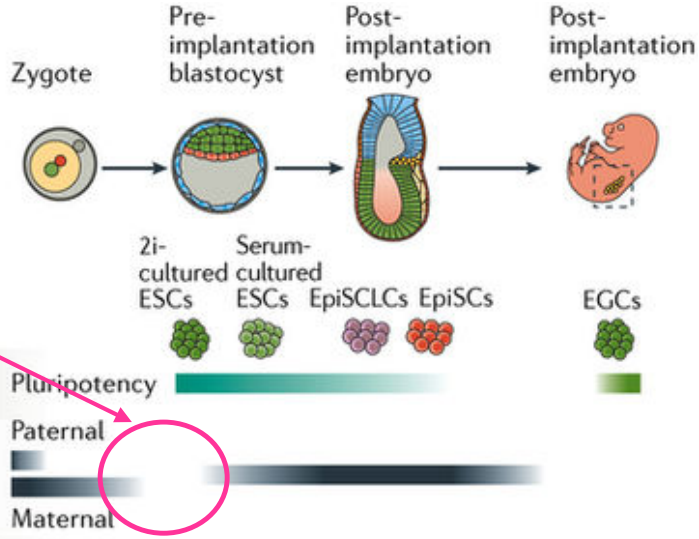
After implantation (E5.5): **epiblast cells** undergo a strong wave of epigenetic reprogramming. They are now „primed“.



Atlasi & Stunnenberg, *Nature Rev Genet* **18**, 643–658 (2017)

Review (V9): Dynamics of epigenetic modifications

DNA methylation is erased in the paternal and maternal genomes after fertilization and is put back on at later developmental stages.

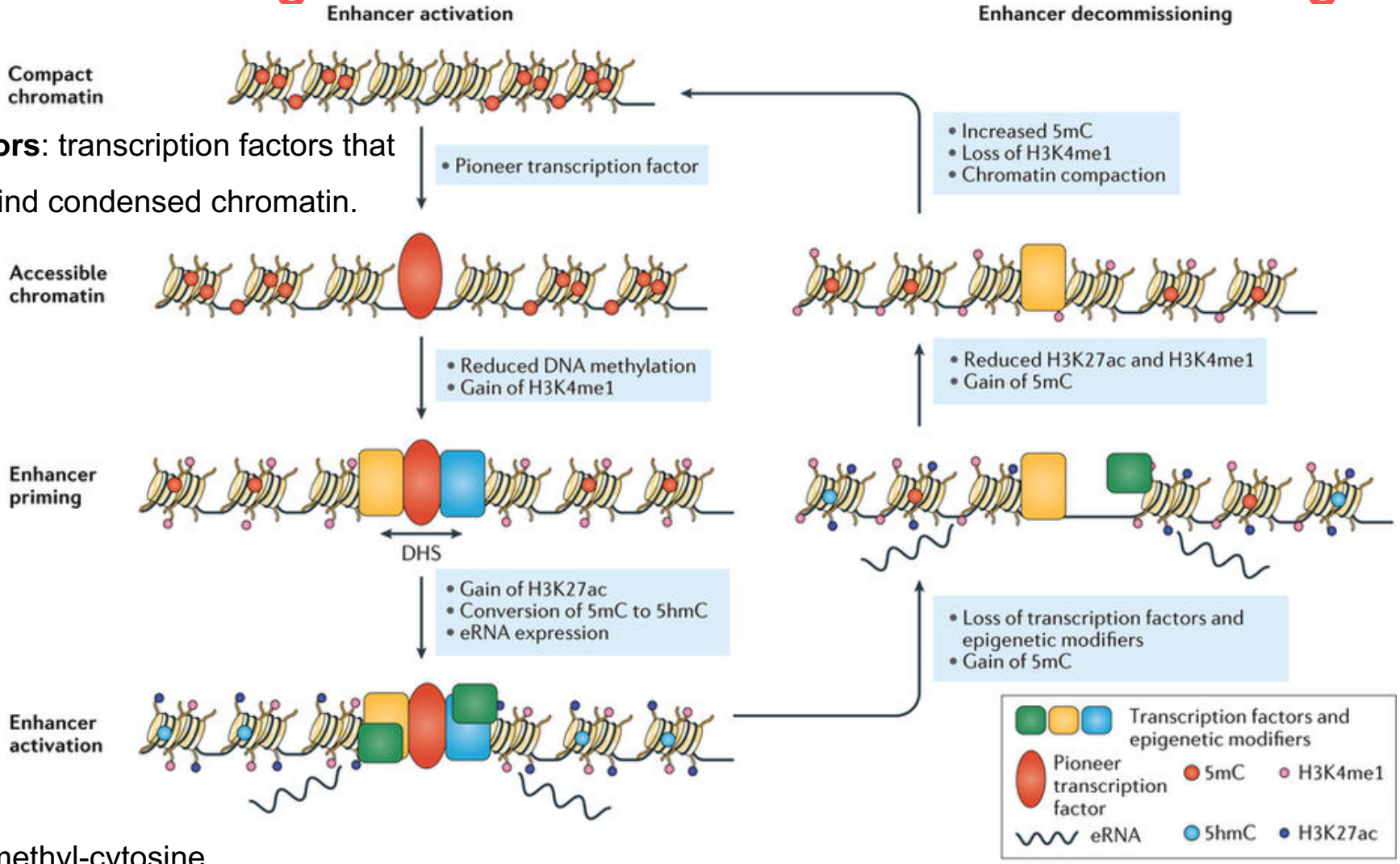


Chromatin modification	Writers	Erasers	Location	Function
DNA methylation	DNMT1, DNMT3A and DNMT3B	TET1, TET2 and TET3	CpG dinucleotides	Silencing and others
H3K27me3	PRC2	• UTX1 • JMJD3	CpG-rich promoters and intergenic regions	Silencing
H3K9me2	G9A and GLP	• JMJD2A, JMJD2B, JMJD2C and JMJD2D • JMJD1A, JMJD1B and JMJD1C	Gene bodies, intergenic regions and enhancers	Silencing
H3K4me3	COMPASS-like proteins (SET1, MLL1–MLL2)	• JARID1A, JARID1B, JARID1C and JARID1D • KDM2B	Mainly promoters	Possibly activating
H3K27ac	HATs (including CBP/p300, GNATs and MYSTs)	HDACs and sirtuins	Promoters and enhancers	Activating
H3K4me1	COMPASS-like proteins (MLL3–MLL4)	LSD1 and LSD2	Promoters, enhancers and intergenic regions	Priming and/or activating

Atlasi & Stunnenberg, *Nature Rev Genet* **18**, 643–658 (2017)

Review (V9): Events during enhancer activation / decommissioning

Pioneer factors: transcription factors that can directly bind condensed chromatin.



5mC: 5-methyl-cytosine

5hmC: 5-hydroxy-methyl-cytosine

Nature Reviews | Genetics

Atlasi & Stunnenberg, *Nature Rev Genet* **18**, 643–658 (2017)

Different pluripotent states

The different states that have been captured *in vitro* provide a **gradient of pluripotency** that resembles different stages of embryonic development:

naive ESCs cultured in serum-supplemented medium or in 2i medium resemble the **pre-implantation epiblast**

- 2i medium = serum-free medium supplemented with two inhibitors of MAP/ERK kinase (MEK)) and glycogen synthase kinase 3 (GSK3):
- (1) PD0325901 inhibits the autocrine stimulation of the mitogen-activated protein kinase (ERK1/2) pathway by fibroblast growth factor-4 (FGF4), which has been shown to be elemental for ES cell differentiation.
- (2) GSK3 inhibition impairs the endogenous repressor activity of Tcf3, a transcriptional repressor of the core pluripotency network

primed epiblast-derived stem cells (EpiSCs) resemble the **post-implantation embryo**.

Differentiation of embryonic stem cells

Human embryonic stem cells (hESCs) can be differentiated into a variety of precursor cell types.
This provides an *in vitro* model system to study early human developmental decisions.

There exist protocols for differentiation of hESCs to various cell states, including
- trophoblast-like cells (TBL),
- mesendoderm (ME), and
- neural progenitor cells (NPCs).

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

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

Differentiation of embryonic stem cells

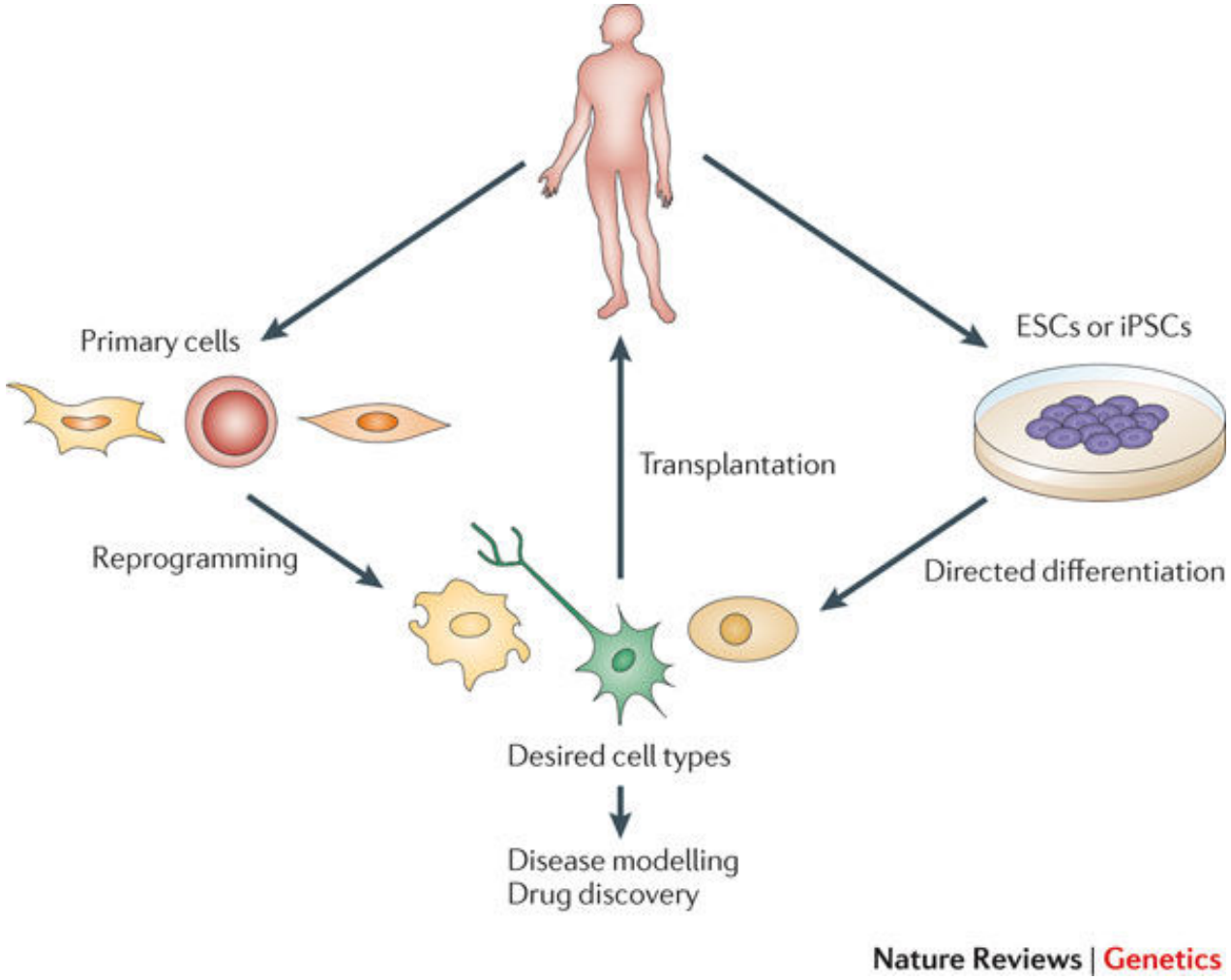
To dissect the early transcriptional and epigenetic events during hESC specification, Gifford *et al*. used **directed differentiation** of hESCs to produce early representative populations from the 3 germ layers, namely **ectoderm**, **mesoderm**, and **endoderm**.
This was followed by fluorescence-activated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

ChIP-seq was also performed for the TFs OCT4, SOX2, and NANOG in the undifferentiated hESCs (-> binding sites of these TFs).

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

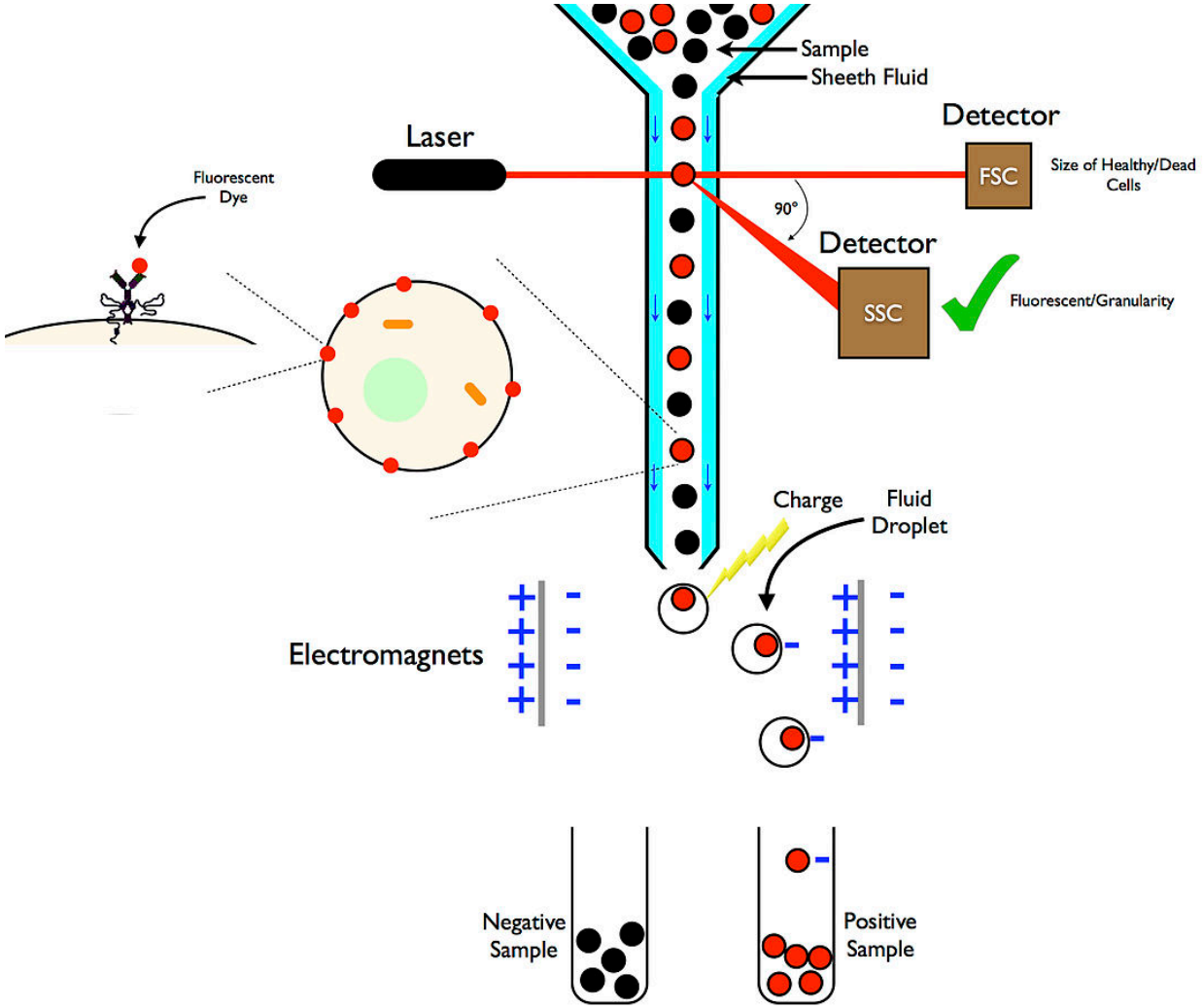
Directed differentiation



Pluripotent cells can be differentiated *in vitro* to a desired cell state (directed differentiation, right).

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

FACS

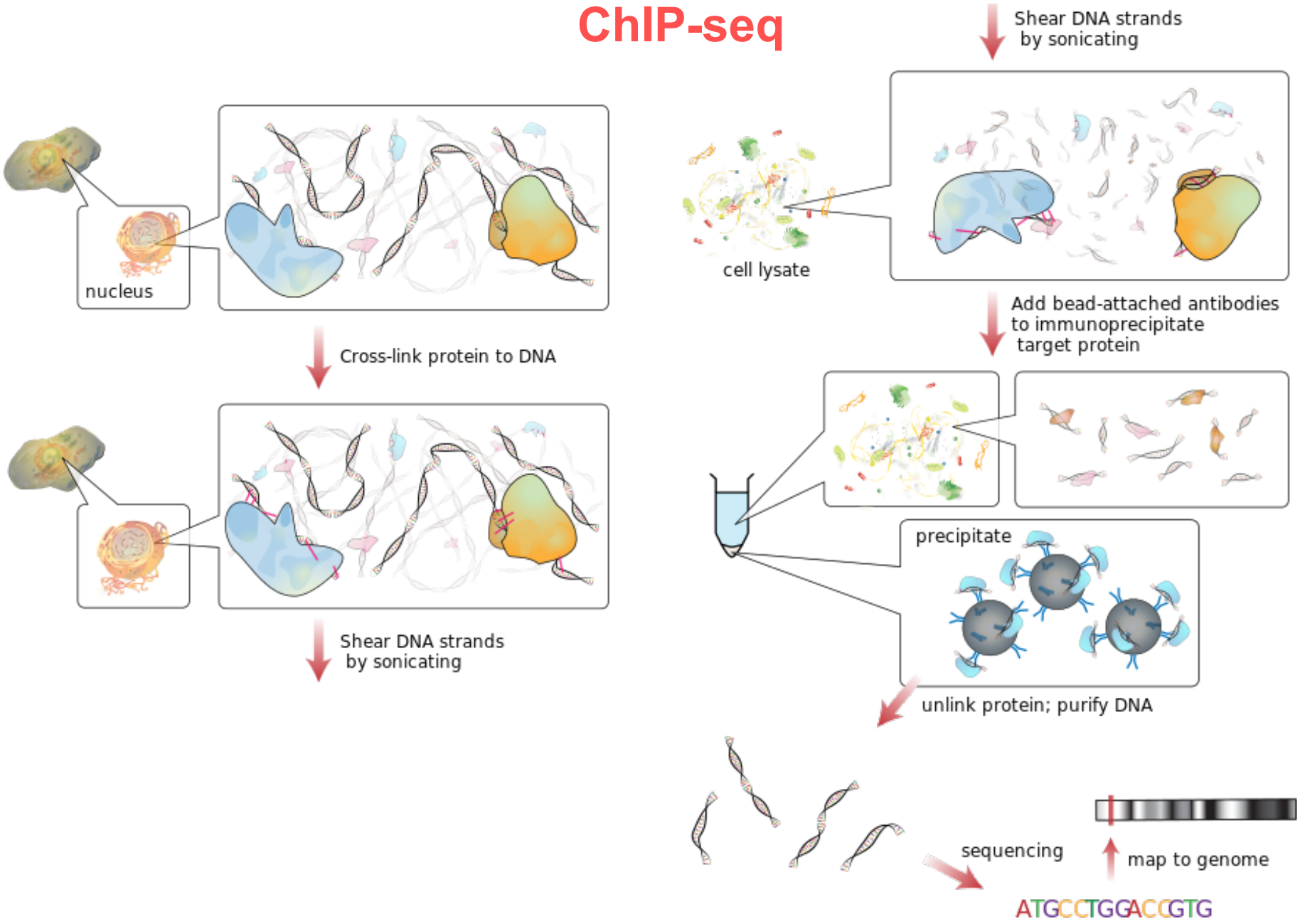


Fluorescence-activated cell sorting (FACS) is a specialized type of flow cytometry. It provides a method for sorting a heterogeneous mixture of biological cells into two or more containers, one cell at a time, based upon the specific light scattering and fluorescent characteristics of each cell.[

www.wikipedia.org

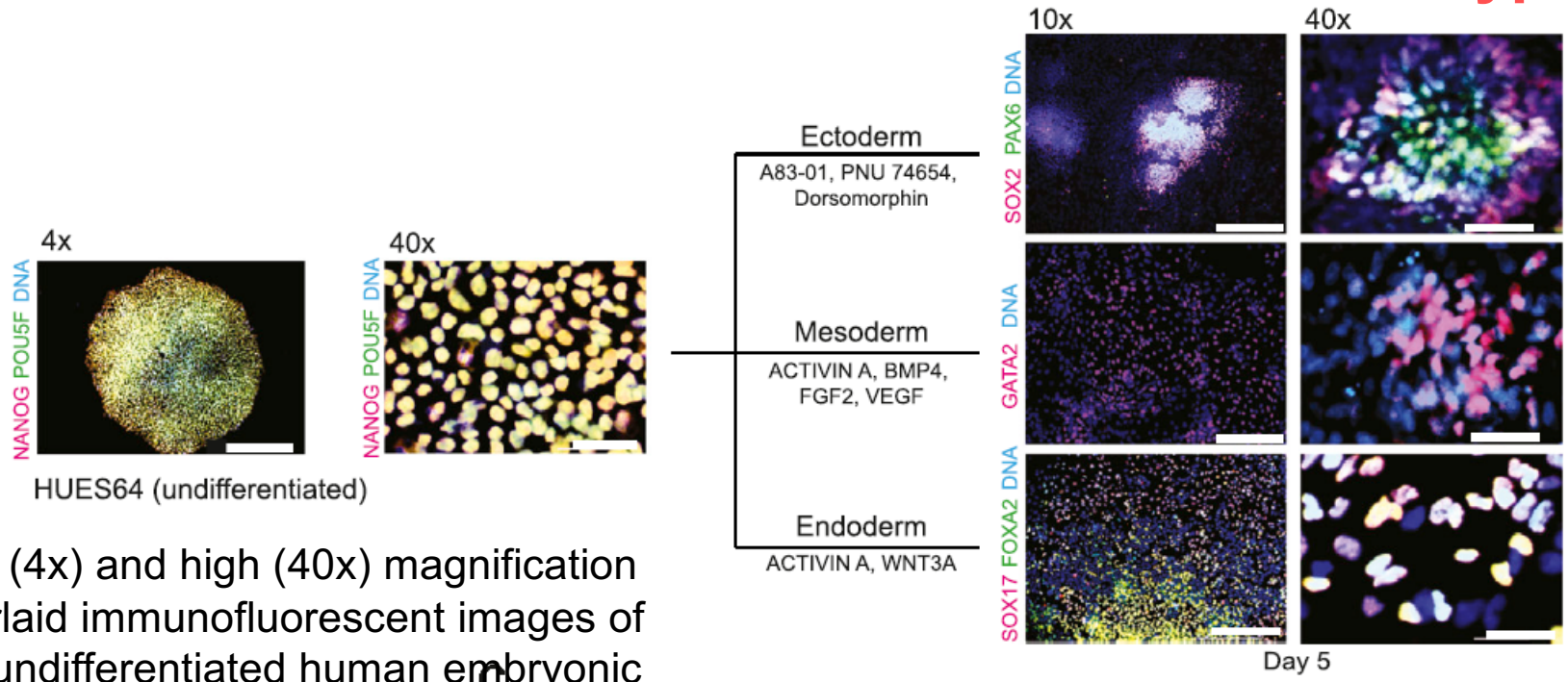
By SariSabban - Sabban, Sari (2011) https://commons.wikimedia.org/w/index.php?curid=18139883

ChIP-seq



www.wikipedia.org

Generation of hESCs and hESC-derived cell types



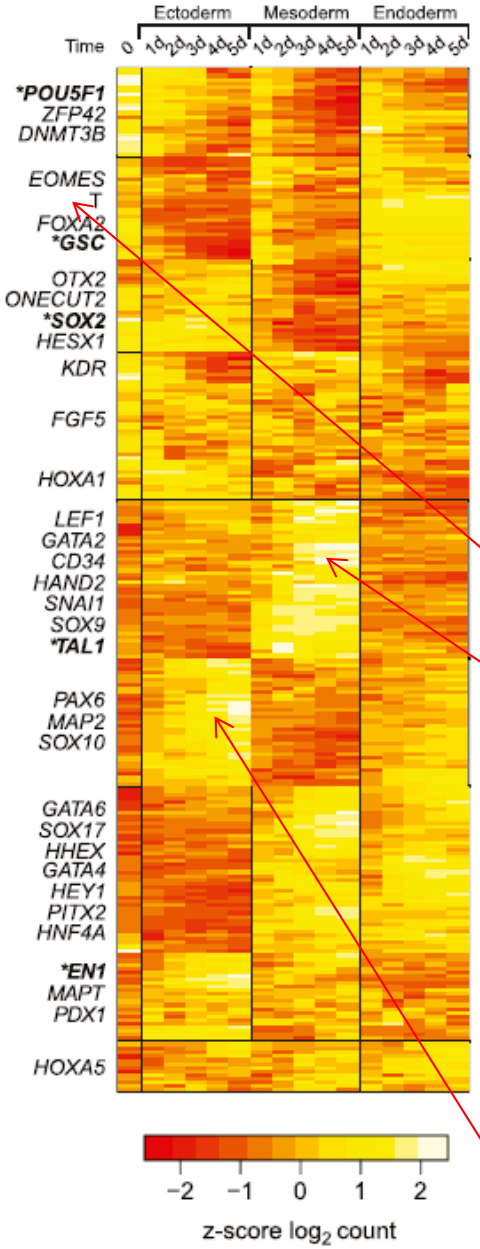
Low (4x) and high (40x) magnification overlaid immunofluorescent images of the undifferentiated human embryonic stem cell (hESC) line HUES64 stained with OCT4 (POU5F1) and NANOG antibodies.

E.g. 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. DNA was stained with Hoechst 33342 in all images.

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

Differential 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)

TFs in Core Pluripotency Network

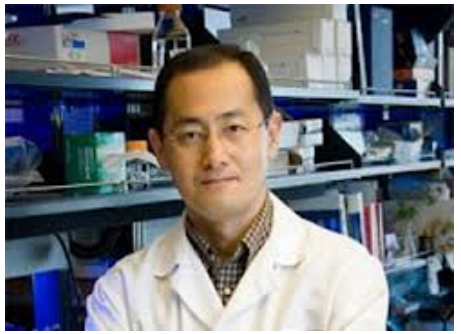
Oct4, encoded by ***Pou5f1***, is a POU domain-containing TF
that is essential to ES cells and early embryonic development.

Oct4 binds to **Sox2**, another TF.

Genome-wide mapping of OCT4 and SOX2 sites
in human ES cells shows that they **co-target** multiple genes.

Oct4 and Sox2, along with **c-Myc** and **Klf4**, appear to be sufficient for reprogramming fibroblasts to **induced pluripotent stem cells (iPS),** which are functionally similar to ES cells (→ **Yamanaka factors**).

Shinya Yamanaka
noble price for medicine 2012



Chen et al., Cell 133, 1106-1117 (2008)

Other TFs in Core Pluripotency Network

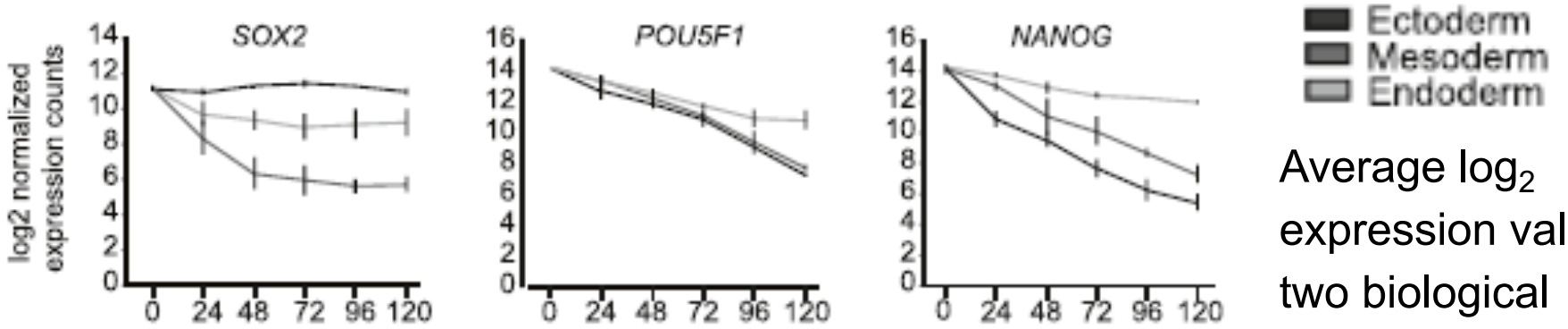
These 4 TFs can exert a dominant role in reconstructing
the transcriptional regulatory network of ES cells.

A third well-studied TF in ES cells is **Nanog**.
Nanog can sustain pluripotency in ES cells.

In addition to this, some further transcriptional regulators such as Esrrb and Zfx are required to maintain ES cells in the state of pluripotency. .

Chen et al., Cell 133, 1106-1117 (2008)

Gene expression of known pluripotency markers



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

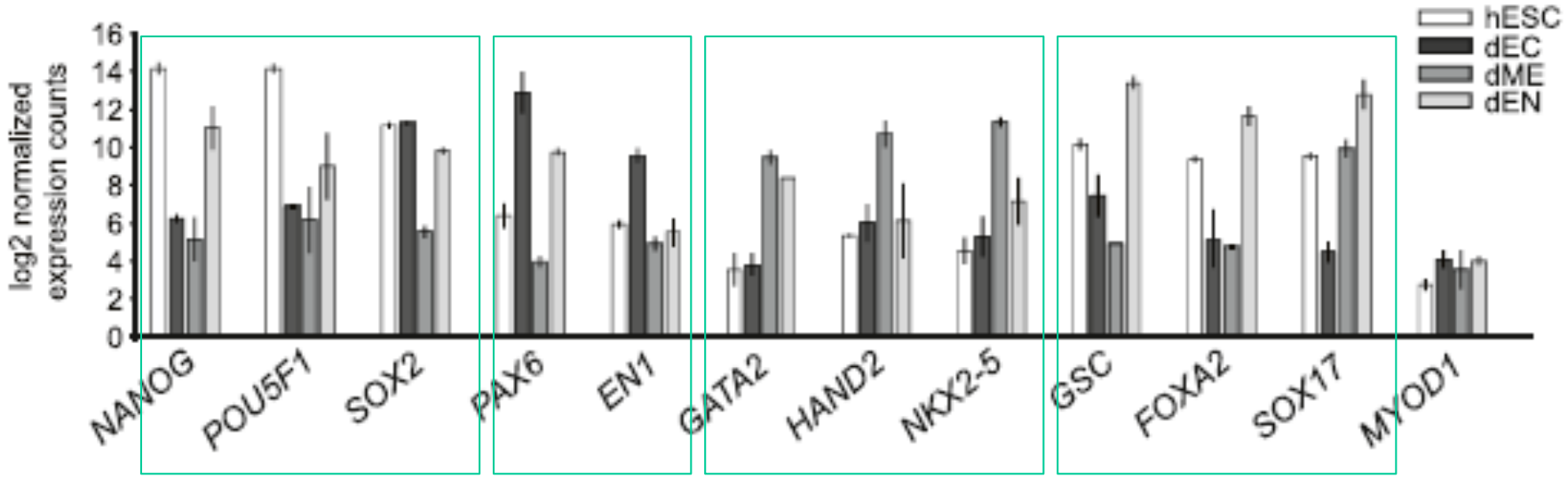
Yamanaka factors (for cell reprogramming): Oct4 (Pou5f1), Sox2, cMyc, and Klf4

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

In ectoderm, SOX2 expression is maintained at high levels.
In mesoderm, SOX2 expression is downregulated.

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

Gene expression in 3 cell lineages



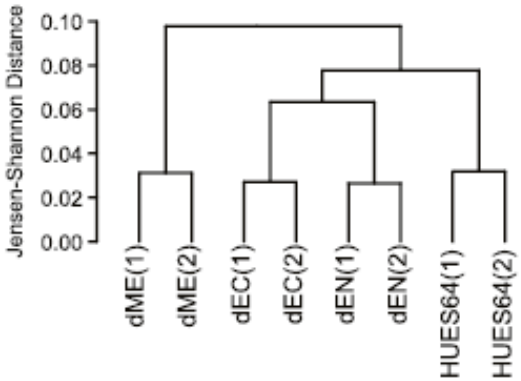
Pluripotency genes | EC markers | ME markers | EN markers

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

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

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

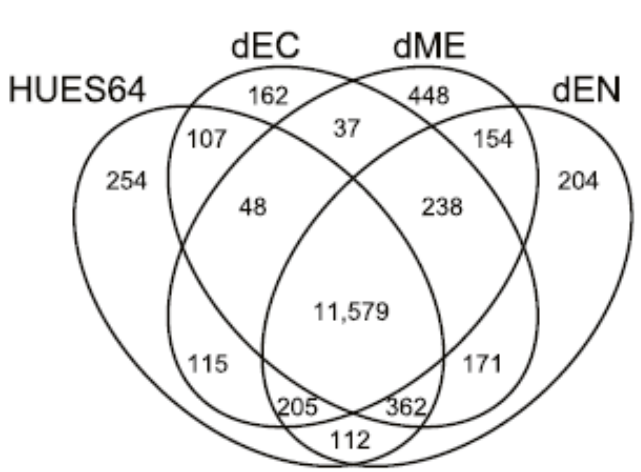
Transcriptional relationship between lineages



Hierarchical clustering of global gene expression profiles for HUES64 and dEC, dME, and dEN.

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)

Epigenetic marks control cellular memory

However, the expression levels of transcription factors are NOT everything!

For example, the maintenance of **cellular memory** also depends on **epigenetic marks** such as DNA methylation and chromatin modifications

DNA methylation at promoters has been shown to silence gene expression (weak correlation, ca. 0.15) 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)

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).

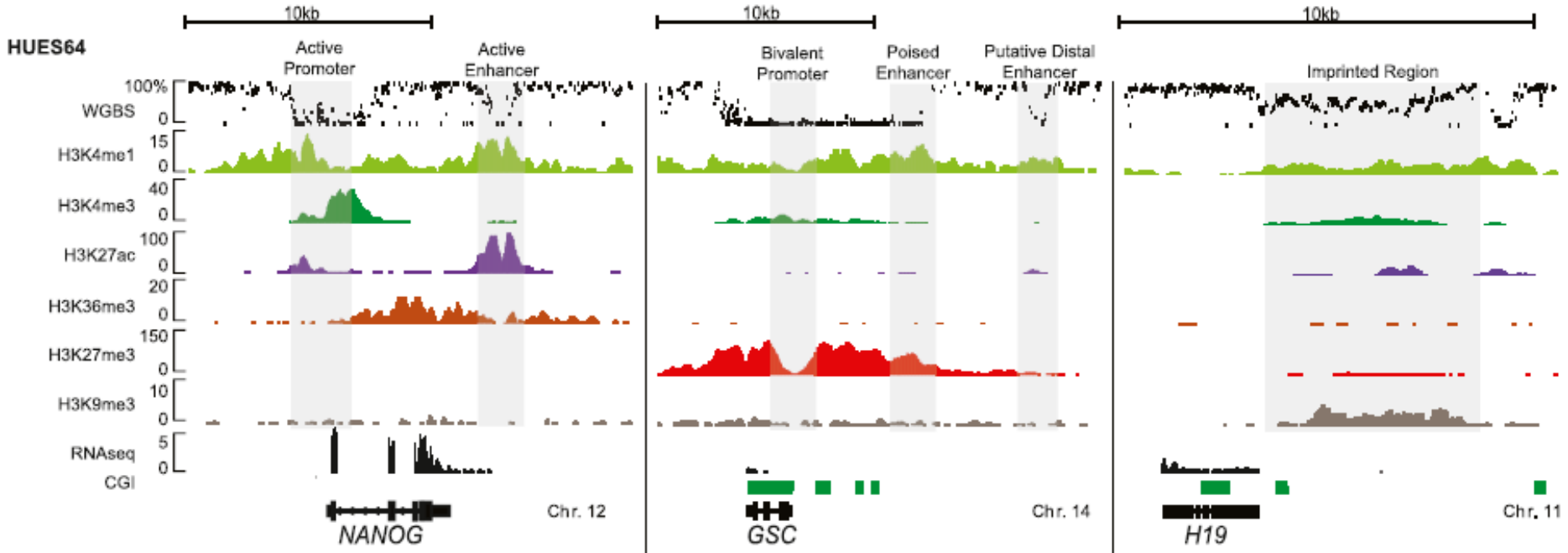
The WGBS data was segmented into three levels of DNA methylation:

- 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 WholeGenomeBisulfiteSequencing (% 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**.

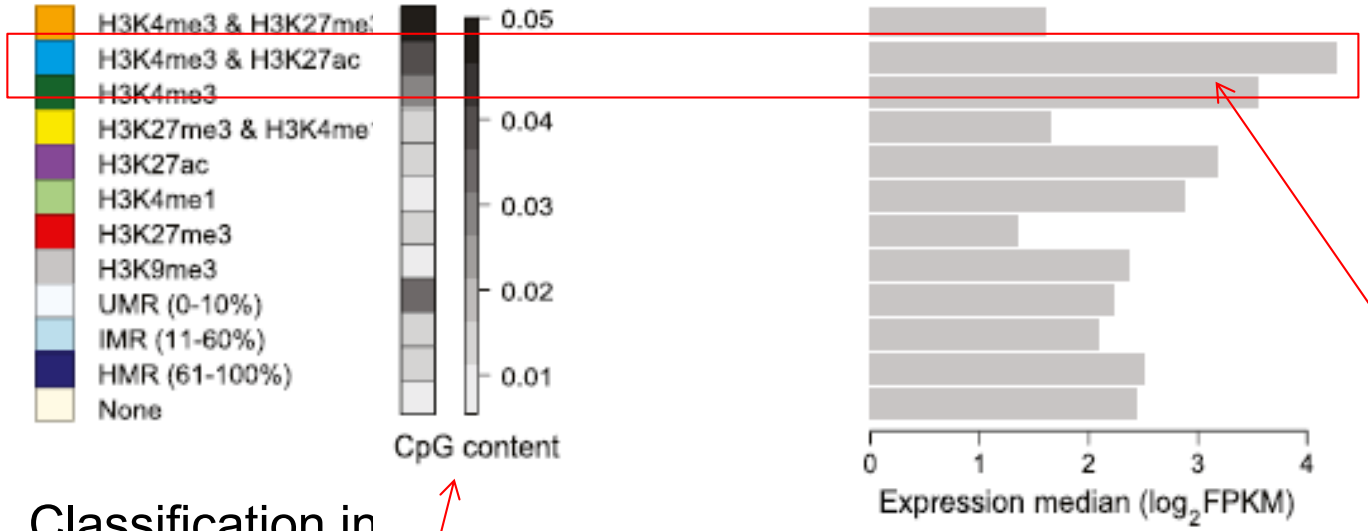
Same data was also collected for dEC, dME, and dEN cells (ca. 12 million cells each)

Bivalent promoter: carries activating (e.g. H3K4me3) and repressive (e.g. H3K27me3) histone marks

Poised enhancer: closed enhancer having H3K4me1 along with H3K27me3 and devoid of H3K27ac marks

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

35% of epigenetic marks are linked to expression levels



Classification in distinct epigenetic states:

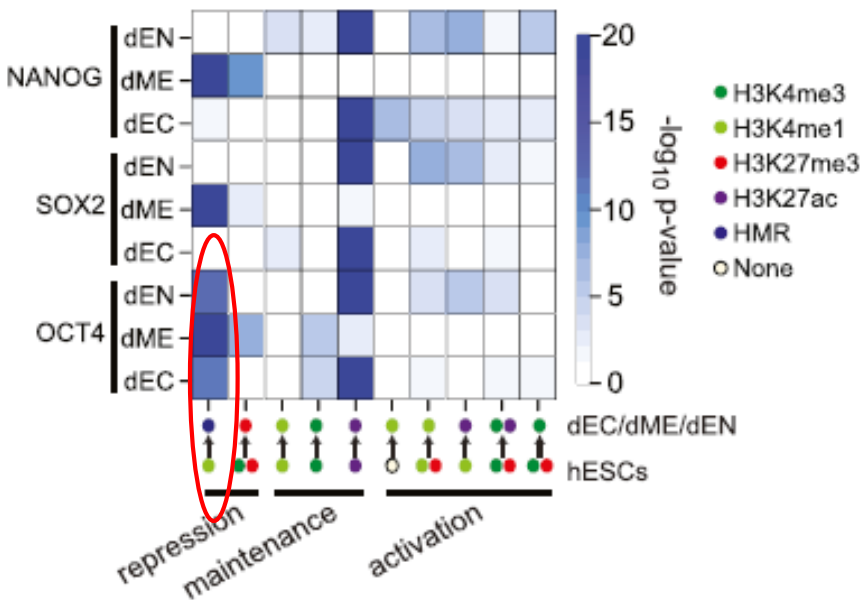
The combination of **H3K4me3** and **H3K27me3** exhibits the **highest CpG content**.

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)

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.

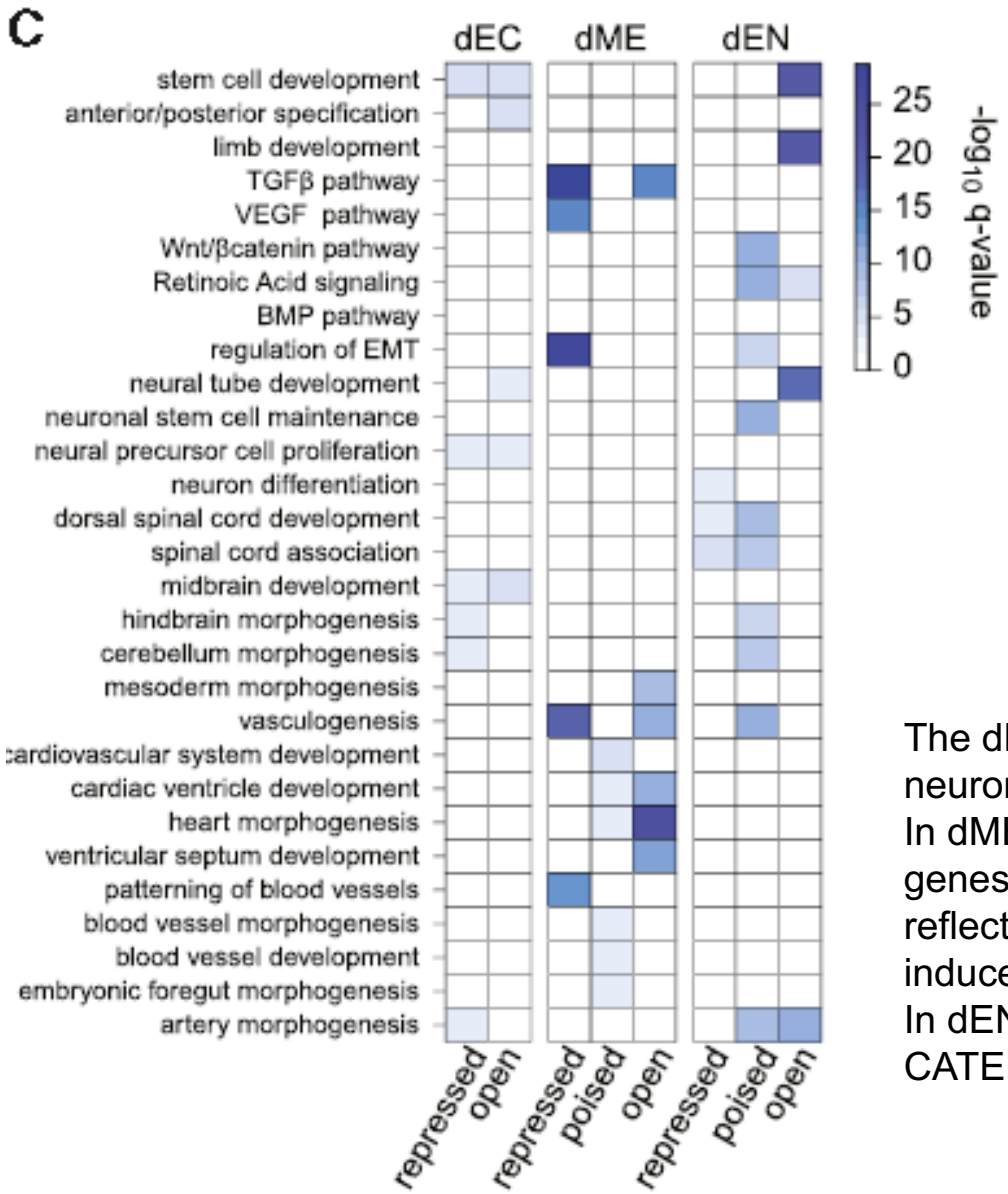
Enrichment of OCT4, SOX2, and NANOG within various classes of dynamic genomic regions that change upon differentiation of hESC.

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

Epigenetic dynamics are categorized into 3 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



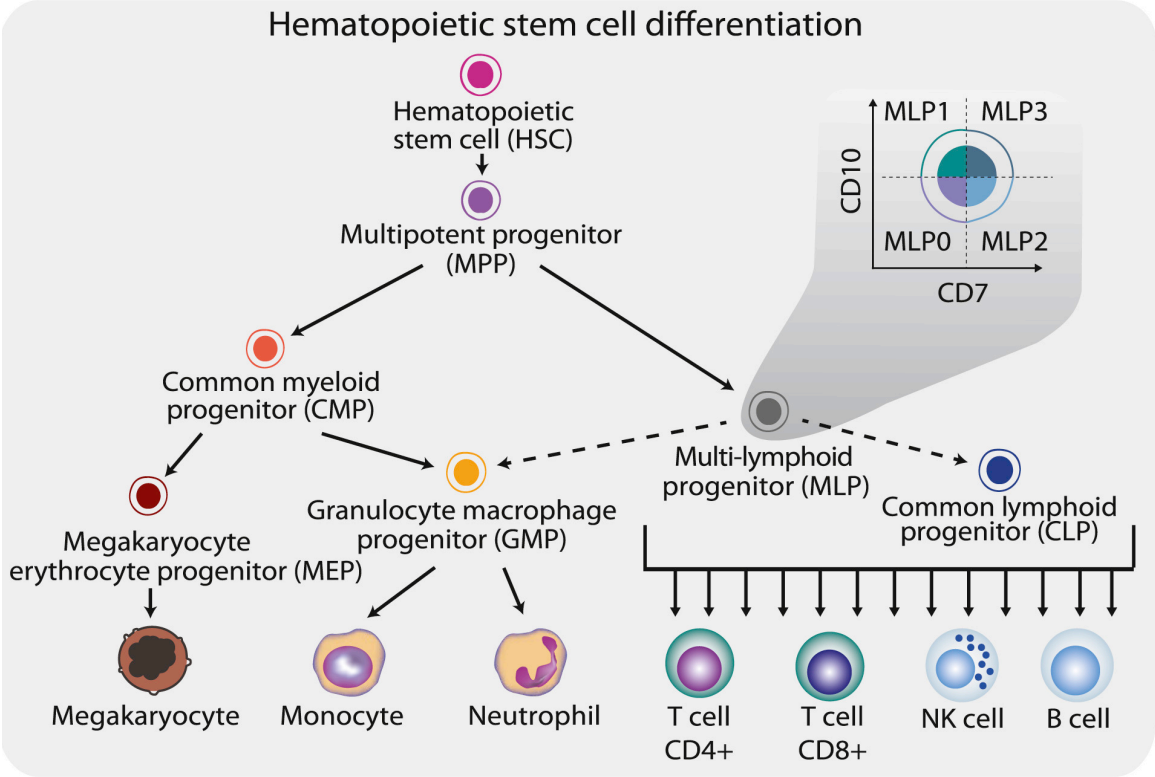
Regions gaining H3K27ac were split up by state of origin in hESC into repressed (none, IMR, HMR, and HK27me3), poised (H3K4me1/ H3K27me3), and Open (H3K4me3/ H3K27me3, H3K4me3, and H3K4me1).

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

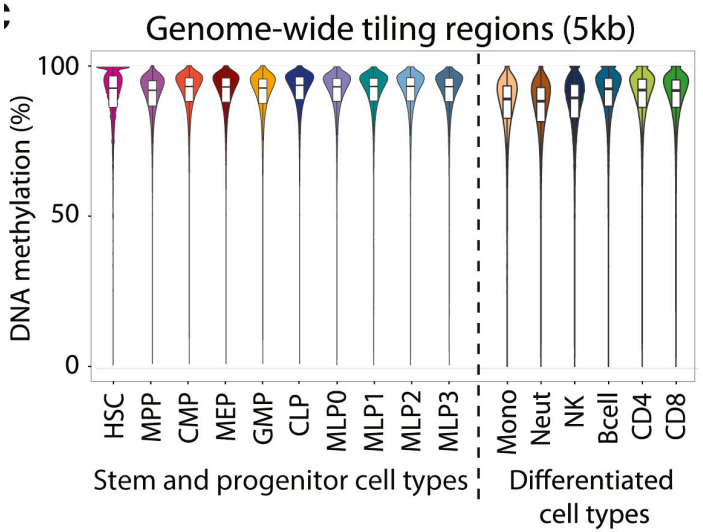
The dEN population shows an enrichment for early neuronal genes..
In dME, there is strong enrichment of downstream effector genes of the TGFb, VEGF, and EMT pathways, directly reflecting the signaling cascades that were stimulated to induce the respective differentiation.
In dEN, genes are enriched that are involved in WNT/b-CATENIN and retinoic acid (RA) signaling.

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

DNA methylation levels during hematopoiesis



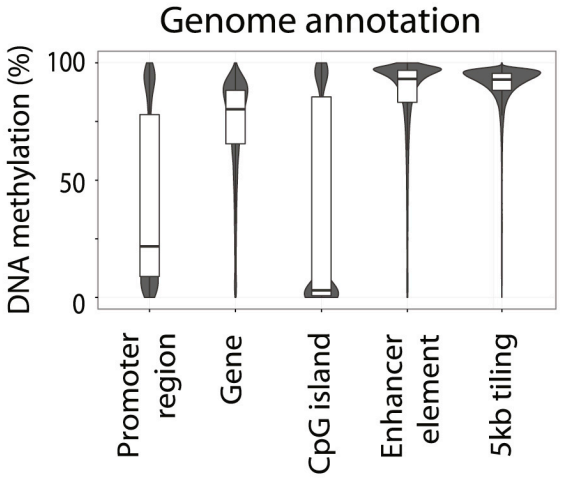
(Left) single-cell whole genome bisulfite sequencing for 17 hematopoietic cell types (multiple types of HSCs).



(right) The distribution of DNA methylation levels was similar across all stem and progenitor cell types.
Differentiated cell types are shifted to slightly lower values.

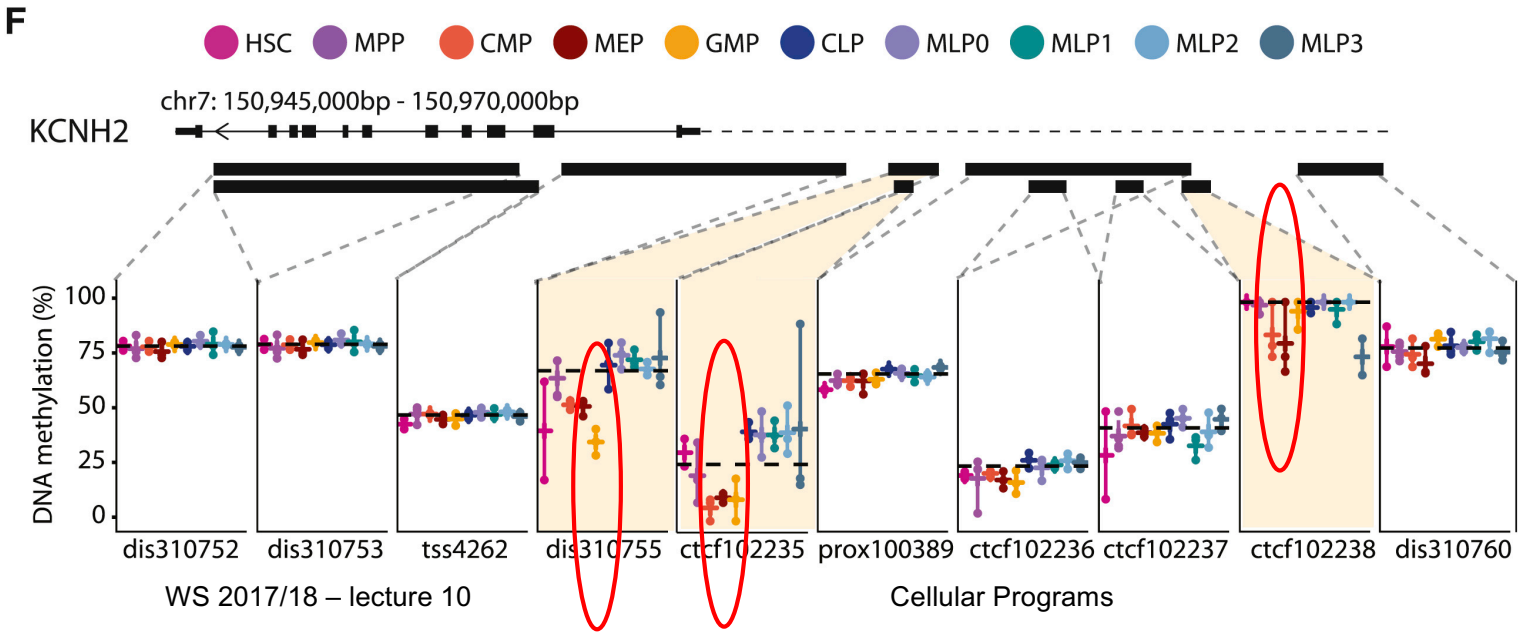
Farlik M et al. Cell Stem Cell (2016) 19:808-822

Local variation of DNA methylation levels



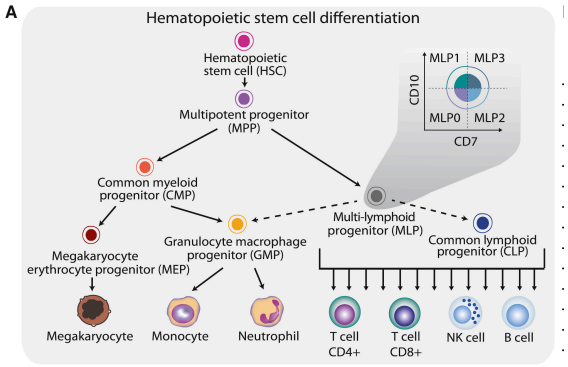
Typical behavior observed: high levels of DNA methylation in most parts of the genome; locally reduced levels at gene promoters and CpG islands

The *KCNH2* gene encodes a key factor for erythroid development. Here, two CTCF sites and a distal element inside the gene show decreased DNA methylation in the myeloid lineage, consistent with increased expression levels in **CMP** and **GMP** cells.

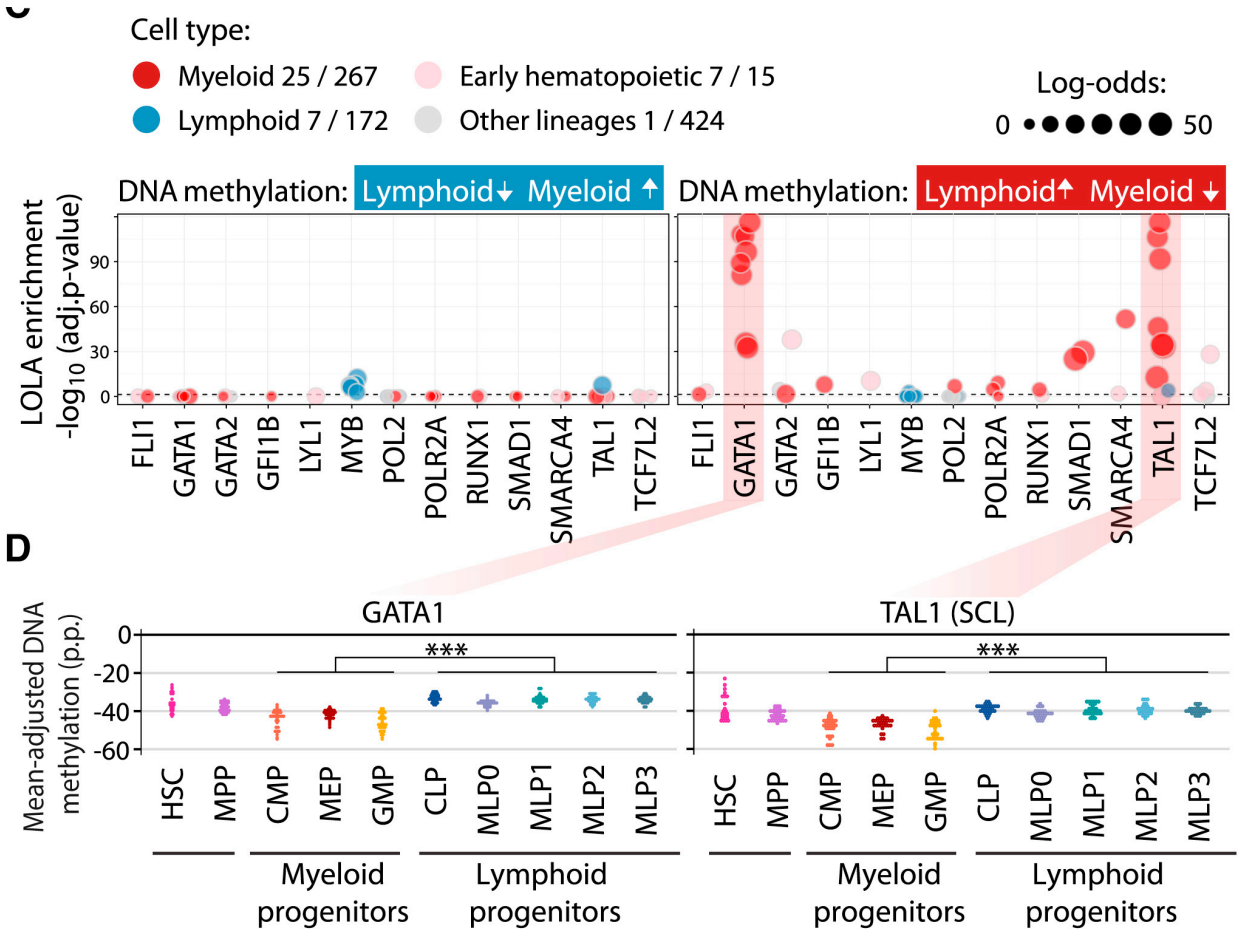


Farlik M et al. Cell Stem Cell (2016) 19:808-822

Myeloid-Lymphoid Lineage Choice



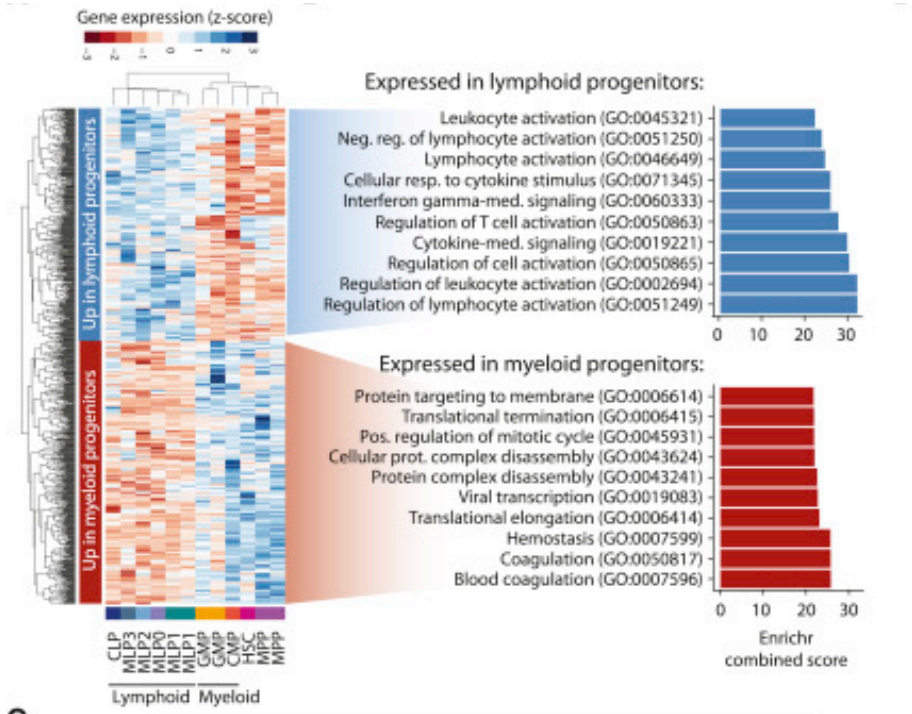
Differentially methylated regions between myeloid and lymphoid progenitors were enriched for binding sites of 11 transcription factors and for RNA polymerase II binding in hematopoietic cells

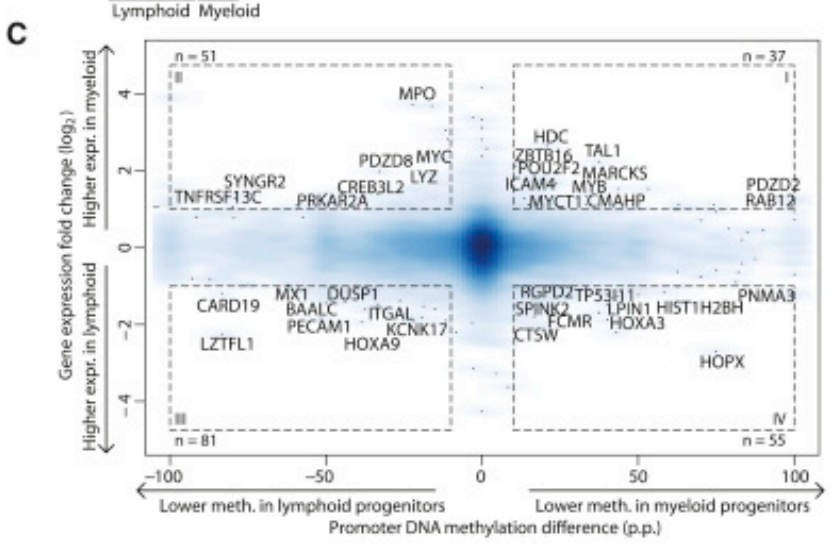


Strongest effects for GATA1 and TAL1.

Farlik M et al. Cell Stem Cell (2016) 19:808-822

Cell-type specific expression levels





656 genes were differentially expressed between myeloid and lymphoid progenitors.

Only few genes (left, bottom) showed concordant methylation and expression changes.

Farlik M et al. Cell Stem Cell (2016) 19:808-822

Tissue signature enrichment levels

DNA methyltransferase (DNMT) inhibitors and histone deacetylase (HDAC) inhibitors are in clinical trials.

A few molecules have already been approved as drugs.

Paper #8 (Fawaz, Salem, Hera): Moignard et al.
Decoding the regulatory network of early blood development from single-cell gene expression measurements
Nature Biotechnology 33, 269–276 (2015)
doi:10.1038/nbt.3154

Paper #9 (Fazaneh, Aditi, Jing Yu): Monika E. Hegi, et al.
MGMT Gene Silencing and Benefit from Temozolomide in Glioblastoma
New England Journal of Medicine 352, 997-1003 (2005)
doi: 10.1056/NEJMoa043331

Paper #10 (Samira, Aryan, Jeenu): Göke J, et al.
Combinatorial Binding in Human and Mouse Embryonic Stem Cells Identifies Conserved Enhancers Active in Early Embryonic Development.
PLoS Comput Biol 7(12): e1002304 (2011)
https://doi.org/10.1371/journal.pcbi.1002304