V11 Differentiation of embryonic stem cells

Human embryonic stem cells (hESCs) can be differentiated into a variety of precursor cell types.
With this *in vitro* model system, one can 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

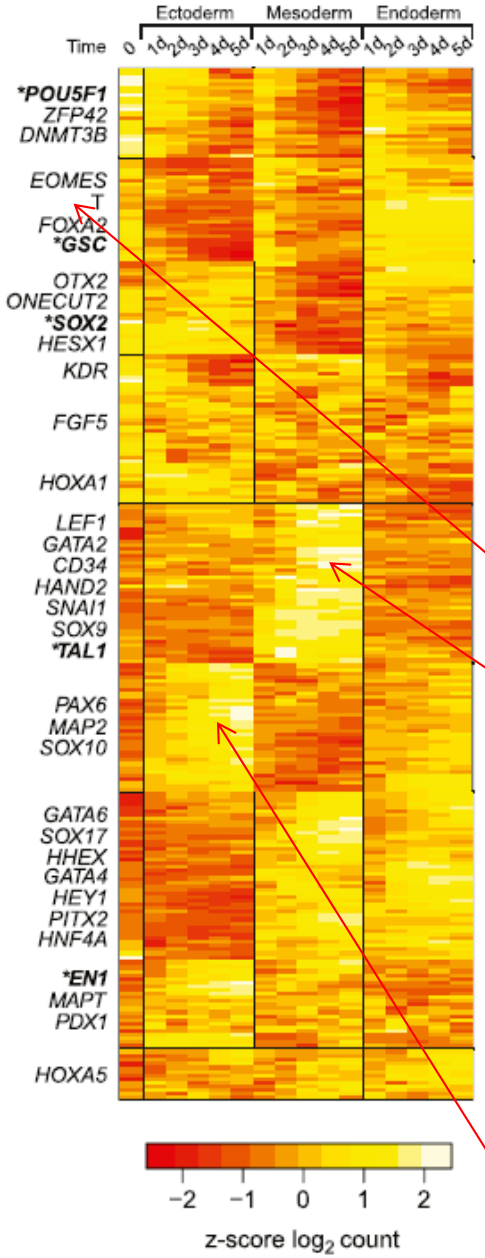
To dissect the early transcriptional and epigenetic events during hESC specification, Gifford *et al*. used **directed differentiation** of hESCs to produce early representative populations from the 3 germ layers, namely **ectoderm**, **mesoderm**, and **endoderm**.
Then they used fluorescence-activated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

These 3 cell types, in addition to **undifferentiated hESCs** (HUES64), were then subjected to
- ChIP-seq for 6 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)

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

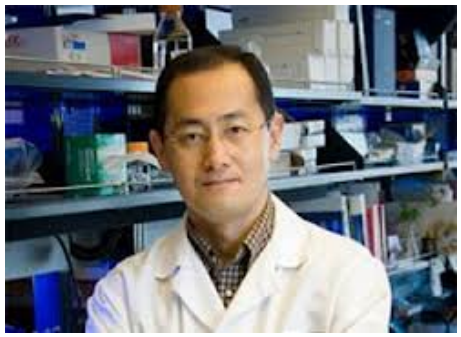
Oct4, encoded by ***Pou5f1***, is a POU domain-containing TF
that is essential to ES cells and early embryonic development.

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

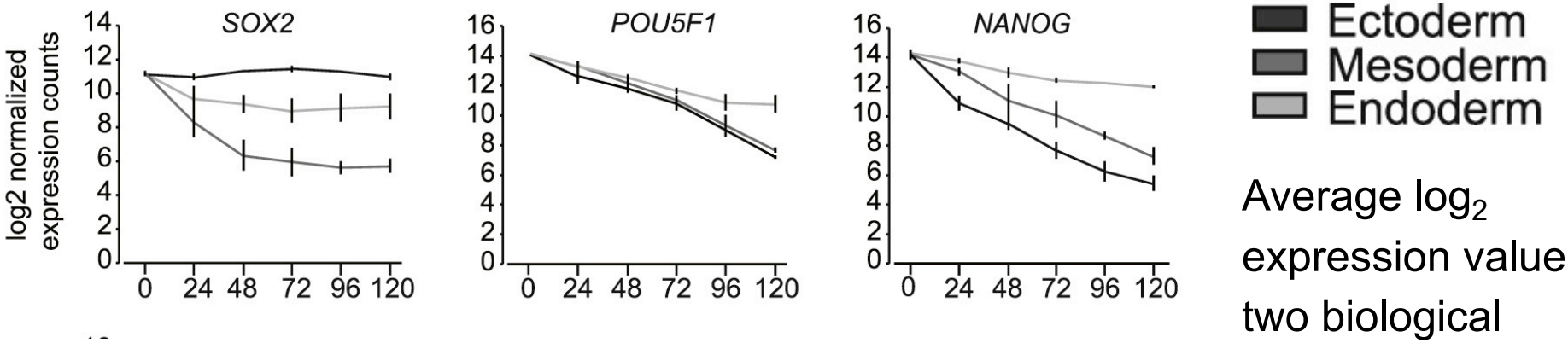
These 4 TFs can exert a dominant role in reconstructing
the transcriptional regulatory network of ES cells.

A further 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.

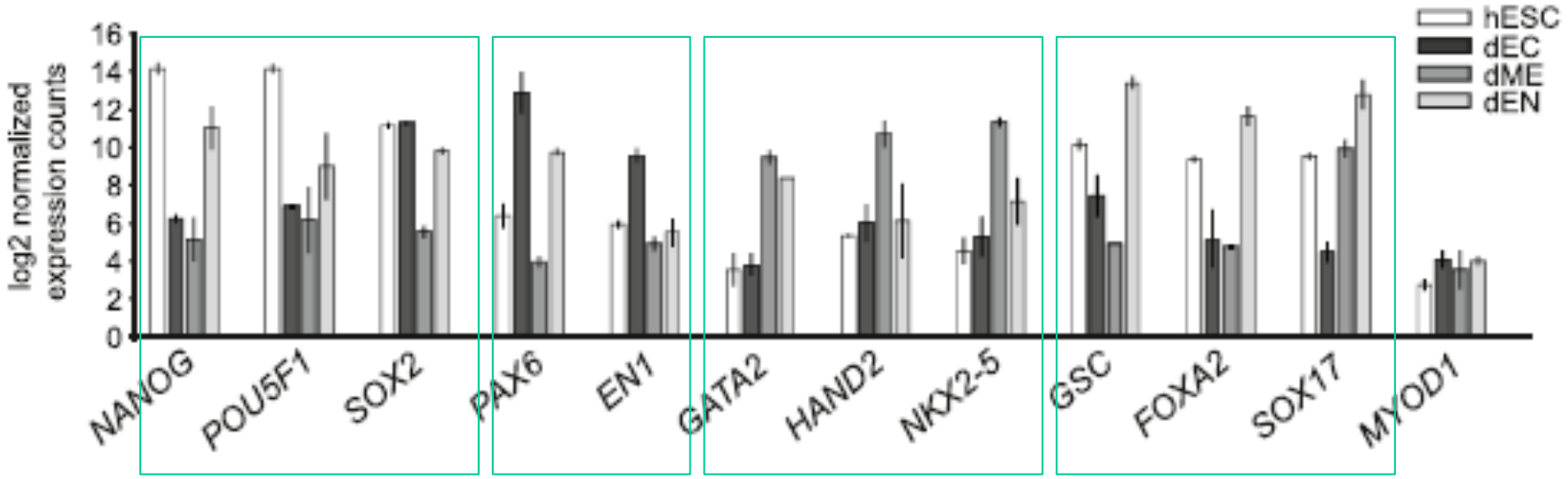
Yamanaka factors (for cell reprogramming): Oct4 (Pou5f1), Sox2, cMyc, and Klf4

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



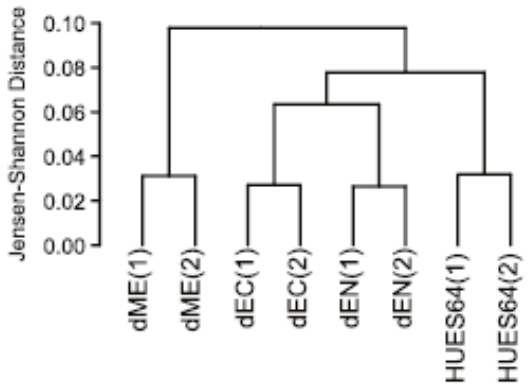
Pluripotency genes | EC markers | ME markers | EN markers

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)

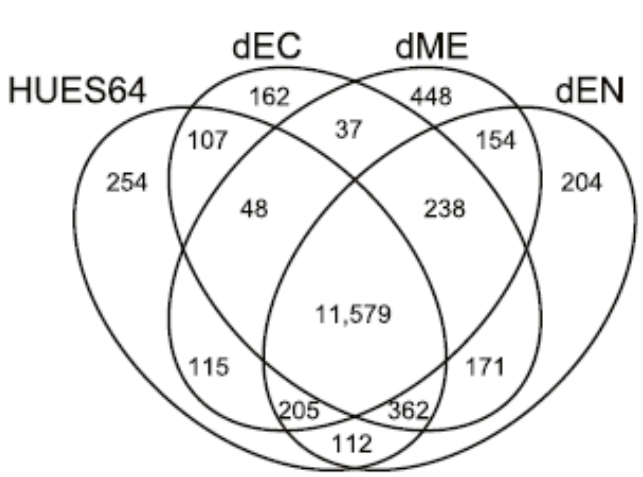
Transcriptional relationship between lineages



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!

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 the following **informative chromatin states**

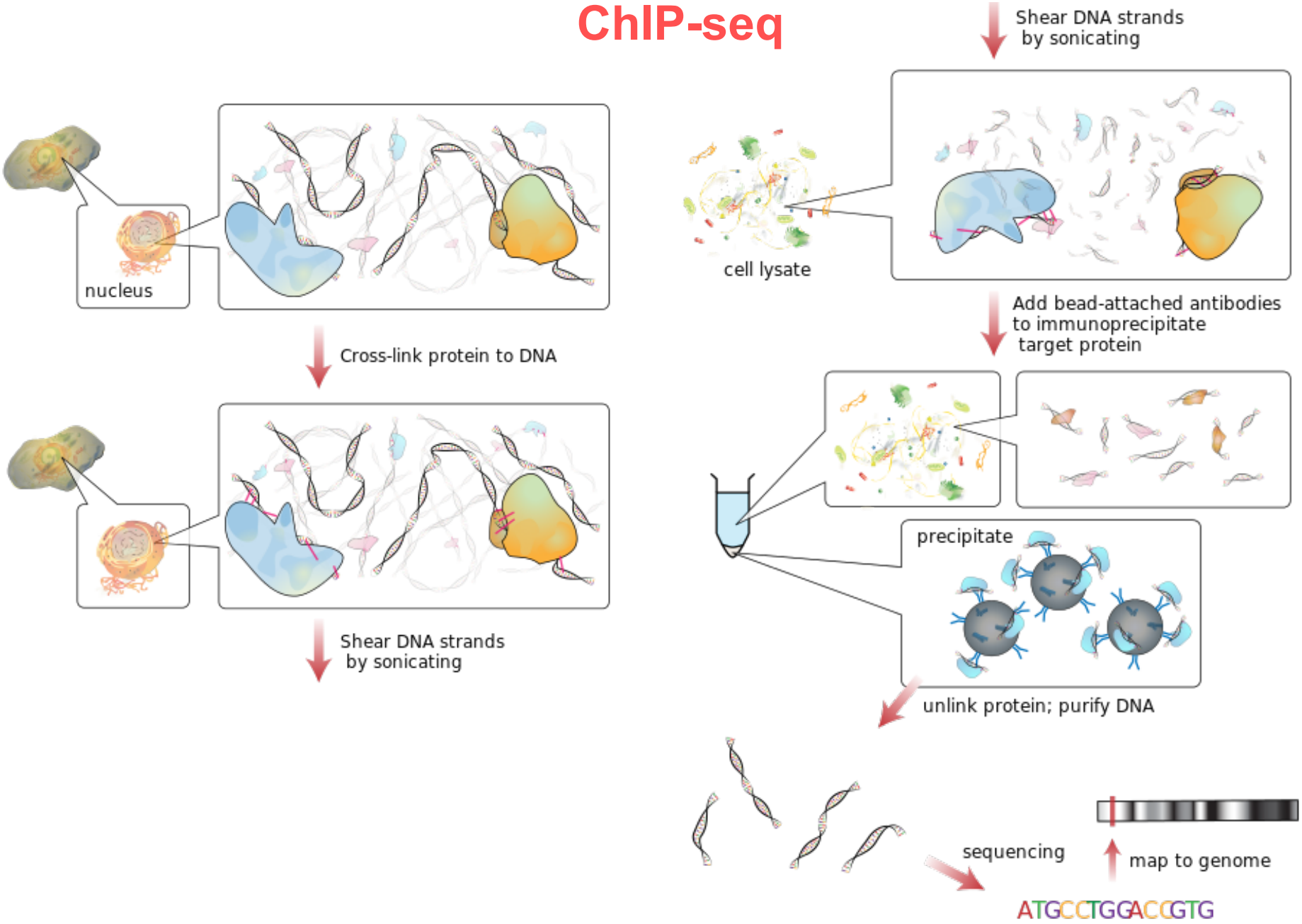
- H3K4me3+H3K27me3 (bivalent/poised promoter);
  „Poised“ genes: RNA-Polymerase II is located at their promoters in the absence of ongoing transcription, the genes are loaded to be transcribed soon
- H3K4me3+H3K27ac (active promoter); gene is actively transcribed
- H3K4me3 (initiating promoter);
- H3K27me3+H3K4me1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

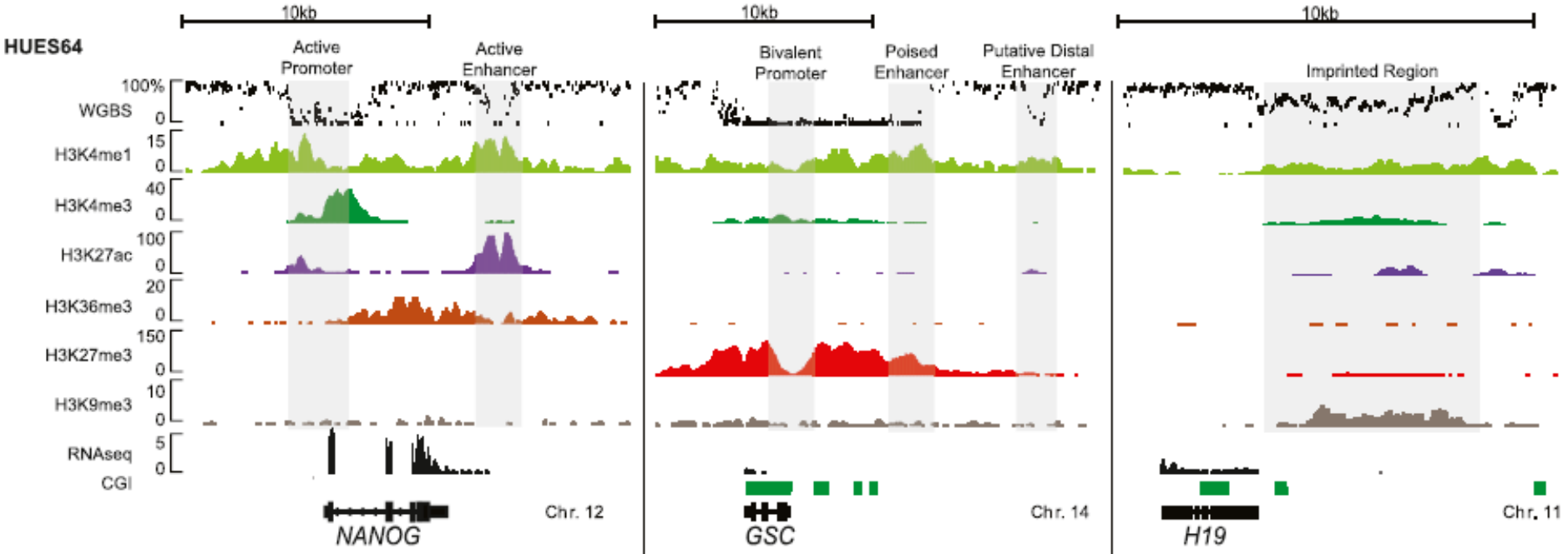
ChIP-seq



www.wikipedia.org

Epigenetic Data for hESC

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



Shown: data for the undifferentiated hESC line HUES64 at 3 loci: NANOG, GSC, and H19 (imp). 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**.

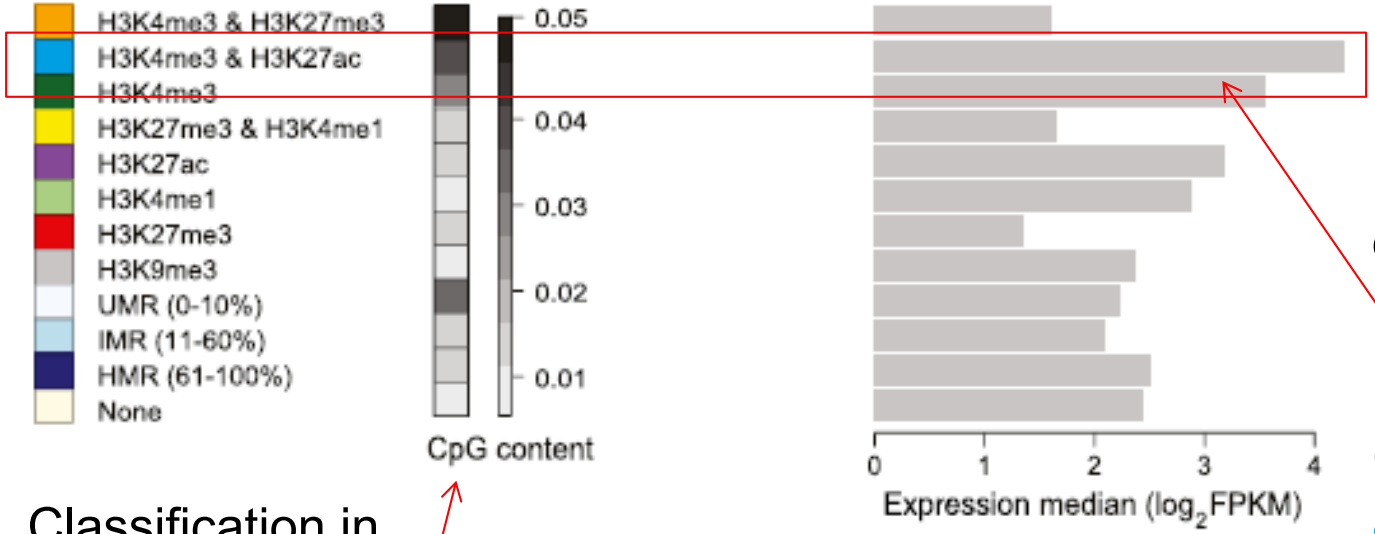
Same data was also collected for dEC, dME, and dEN cells (ca. 12 million cells each)

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 epige-netic states:

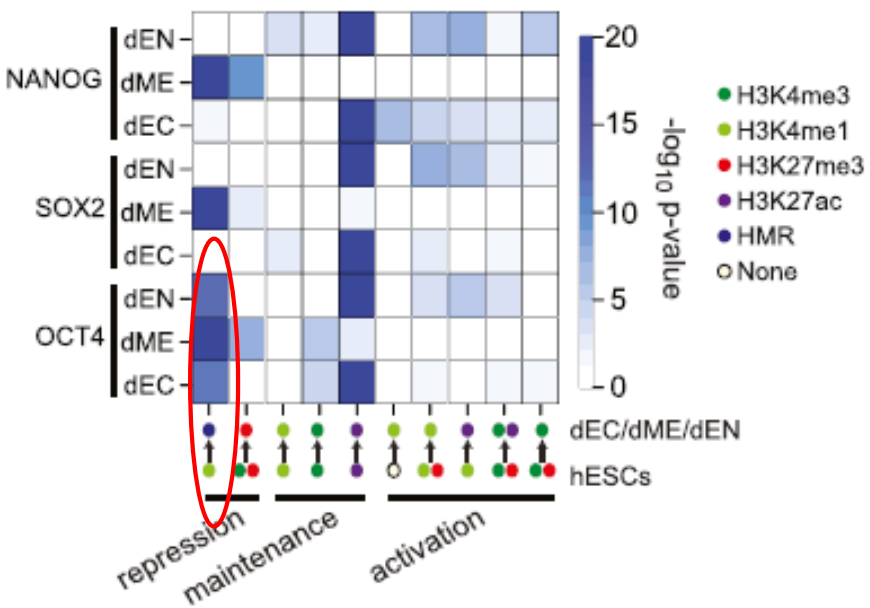
The combination of **H3K4me3** and **H3K27me3** exhibits the **highest CpG content**.

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.
Regions enriched for H3K27ac in hESCs that maintain this state in dEN or dEC are likely to be bound by SOX2 and NANOG.

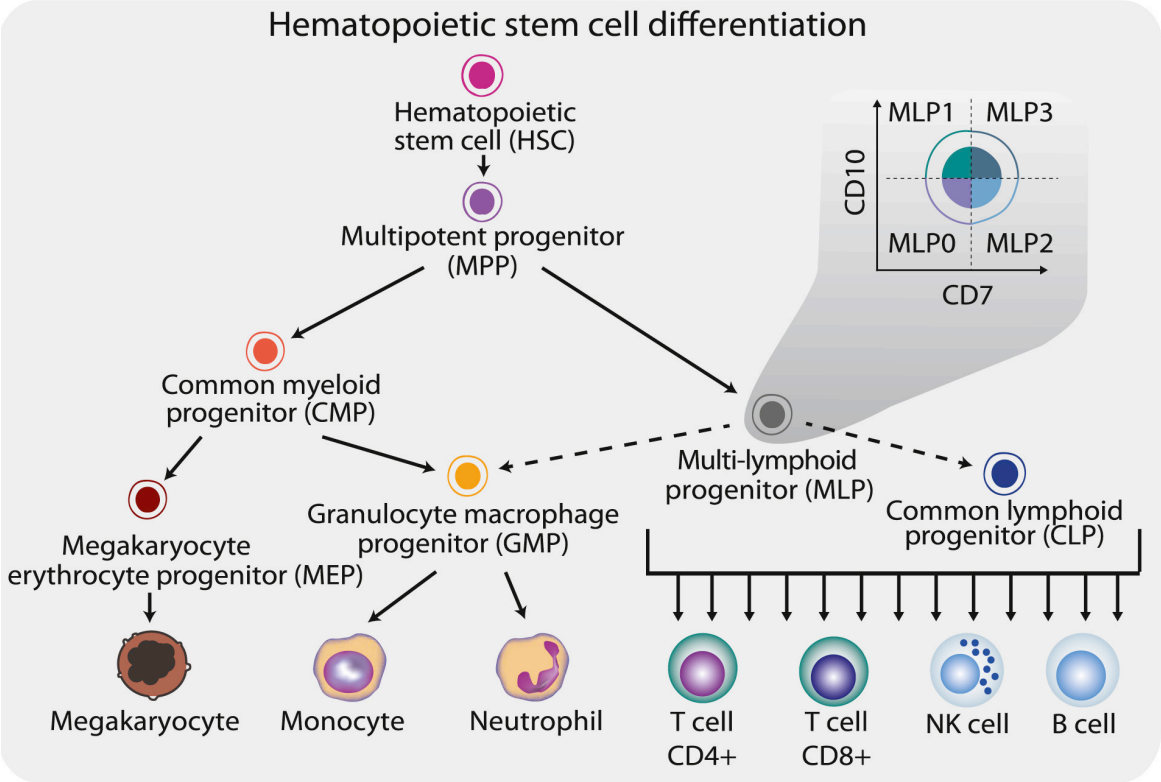
Enrichment of OCT4, SOX2, and NANOG within various classes of dynamic genomic regions that change upon differentiation of hESC.

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

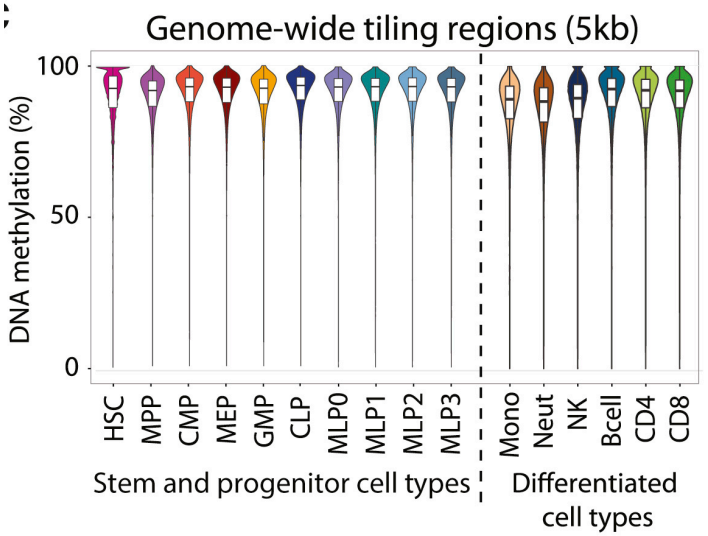
Epigenetic dynamics are categorized into 3 major classes:
- repression (loss of H3K4me3 or H3K4me1 and acquisition of H3K27me3 or DNAme),
- maintenance of open chromatin marks (H3K4me3, H3K4me1, and H3K27ac), and
- activation of previously repressed states.

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

DNA methylation levels during hematopoiesis



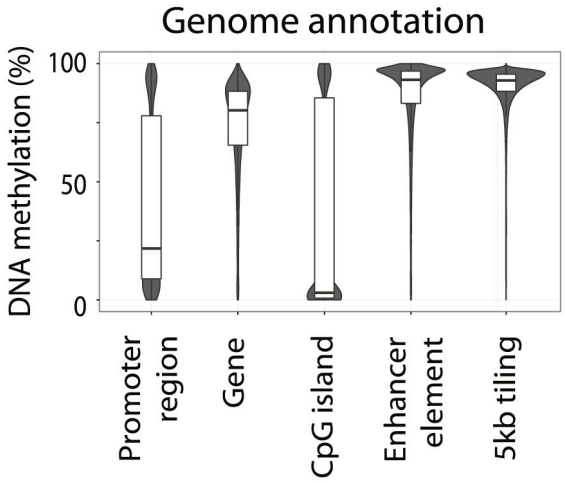
(Left) single-cell whole genome bisulfite sequencing for 17 hematopoietic cell types (multiple types of HSCs).



(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.

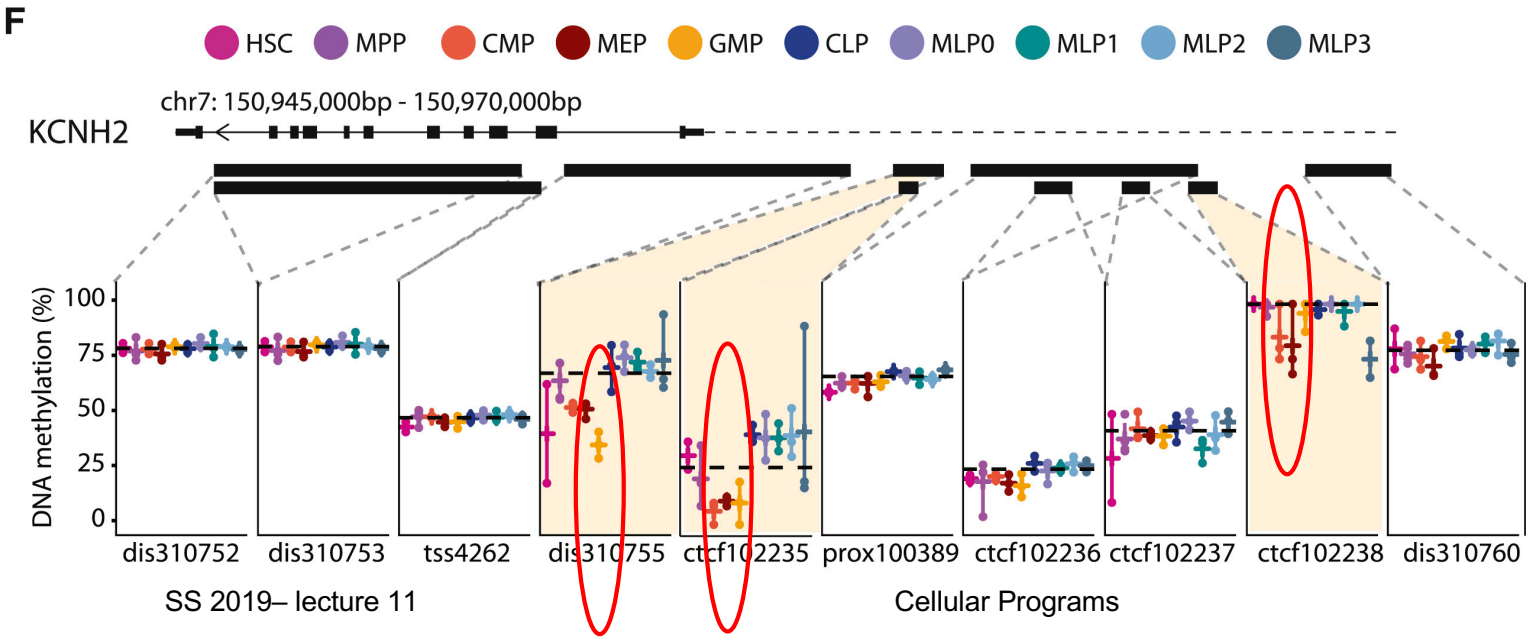
Farlik M et al. Cell Stem Cell (2016) 19:808-822

Local variation of DNA methylation levels



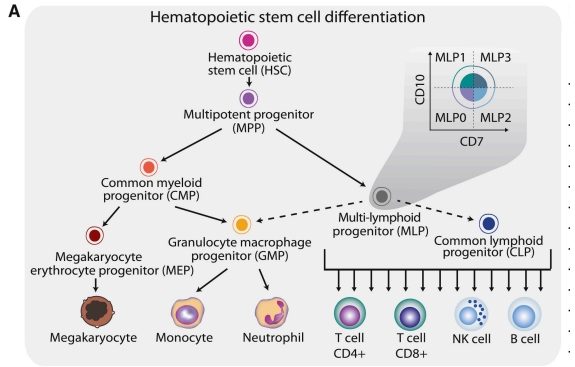
Typical behavior observed: high levels of DNA methylation in most parts of the genome; locally reduced levels at gene promoters and CpG islands

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.

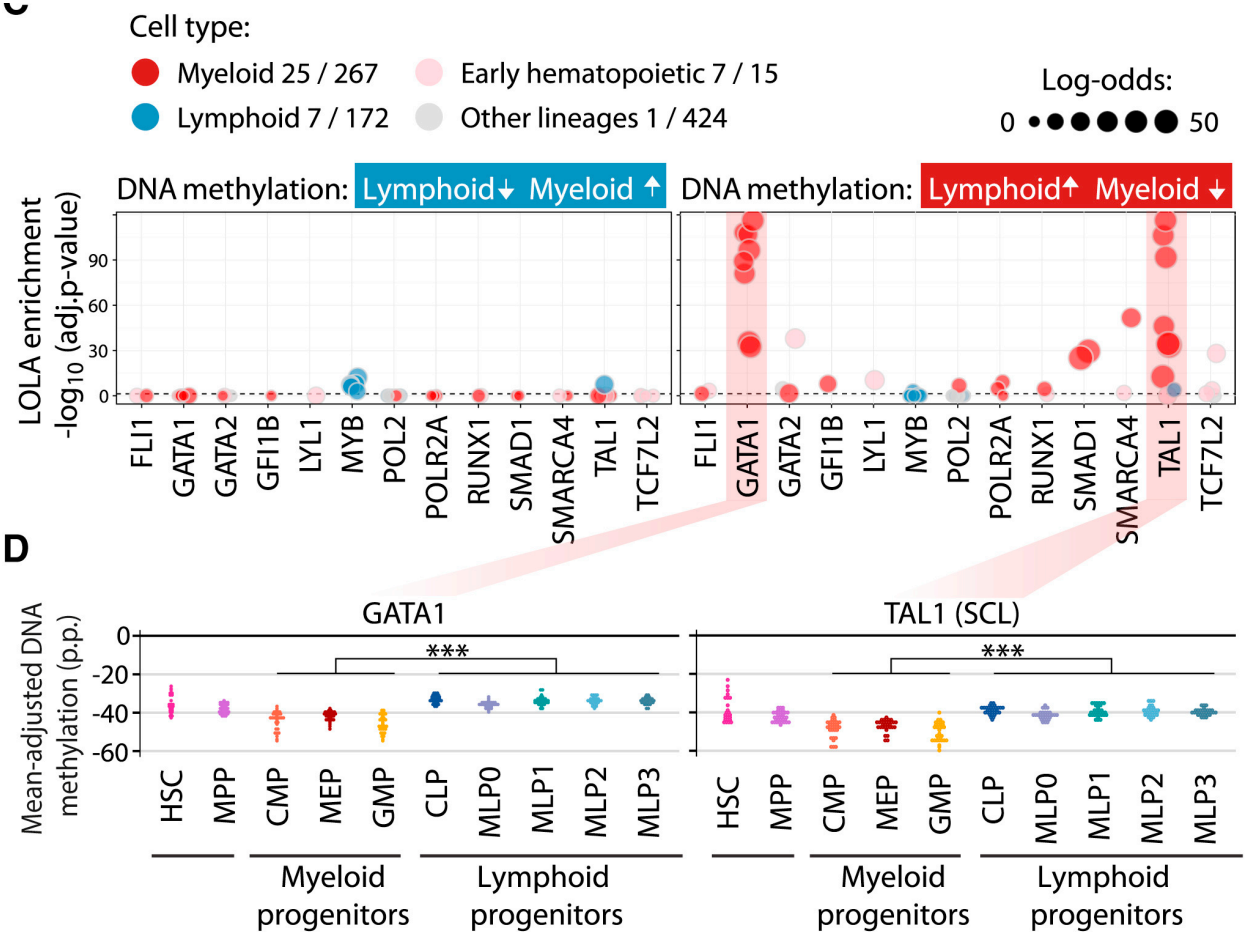


Farlik M et al. Cell Stem Cell (2016) 19:808-822

Myeloid-Lymphoid Lineage Choice



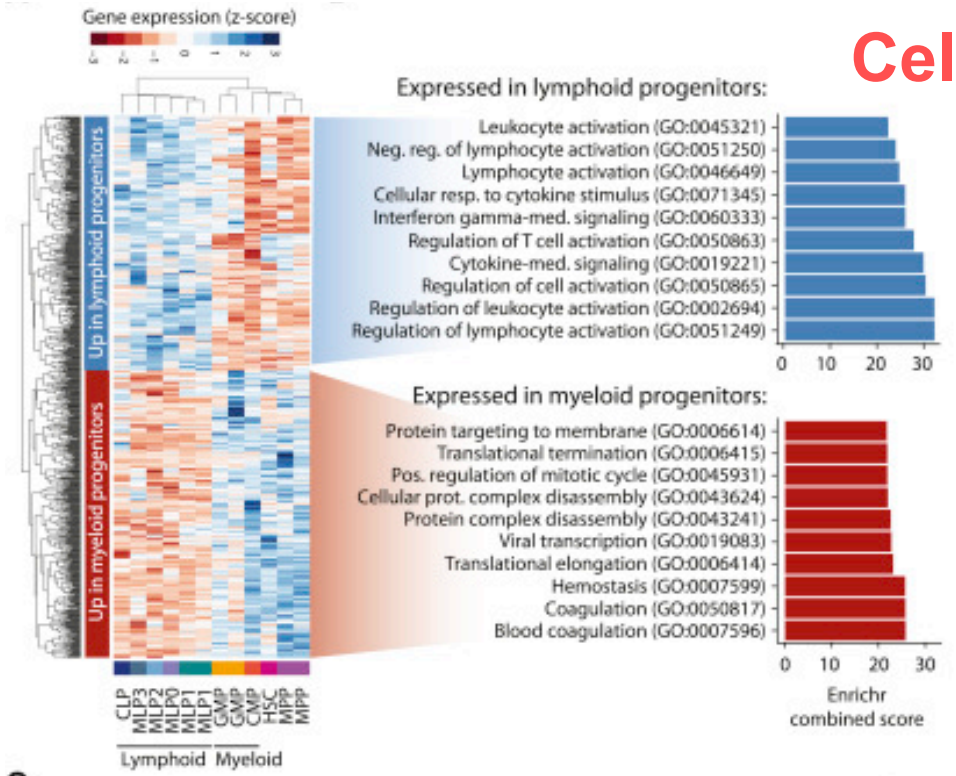
Differentially methylated regions between myeloid and lymphoid progenitors were enriched for binding sites of 11 transcription factors and for RNA polymerase II binding in hematopoietic cells

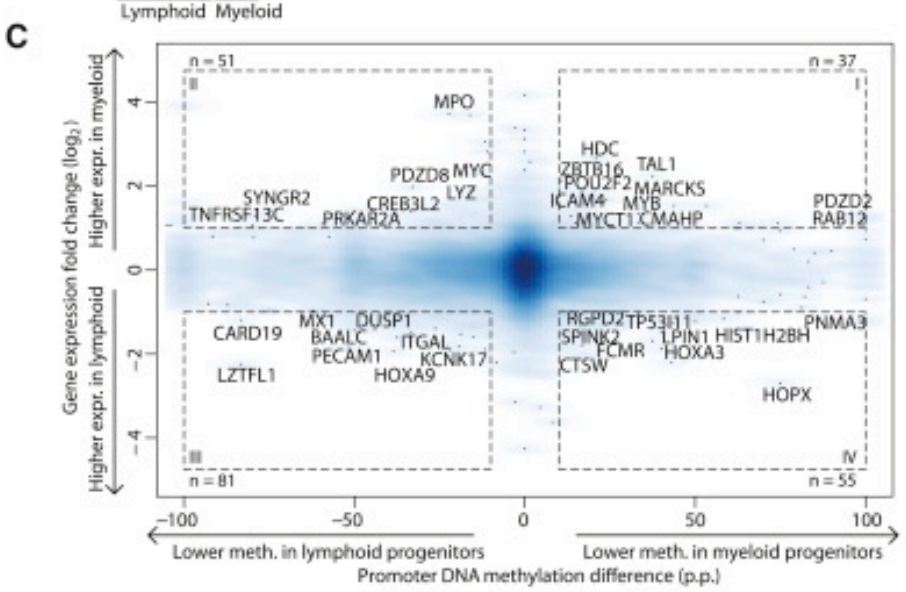


Strongest effects for GATA1 and TAL1.

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

→ The relationship between DNA methylation and gene expression is quite complex and still not clearly understood.

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

miRNAs

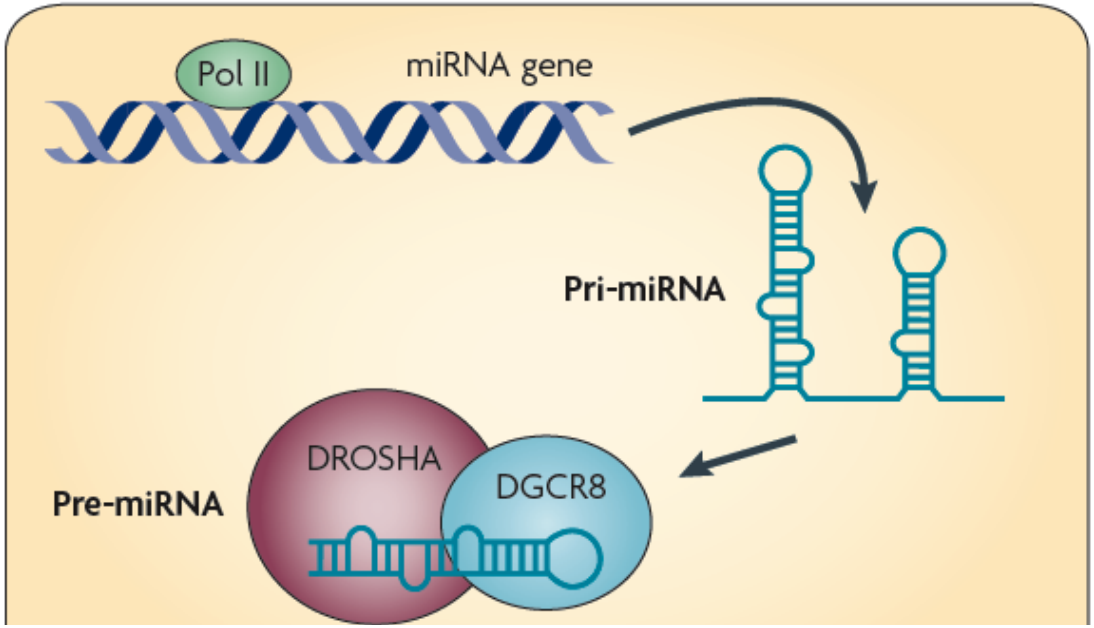
microRNAs (miRNA) are **single-stranded RNA** molecules of 21-23 nucleotides in length.

miRNAs have a crucial role in regulating gene expression.

Remember: miRNAs are encoded by DNA but not translated into protein (non-coding RNA).

www.wikipedia.org

Overview of the miRNA network



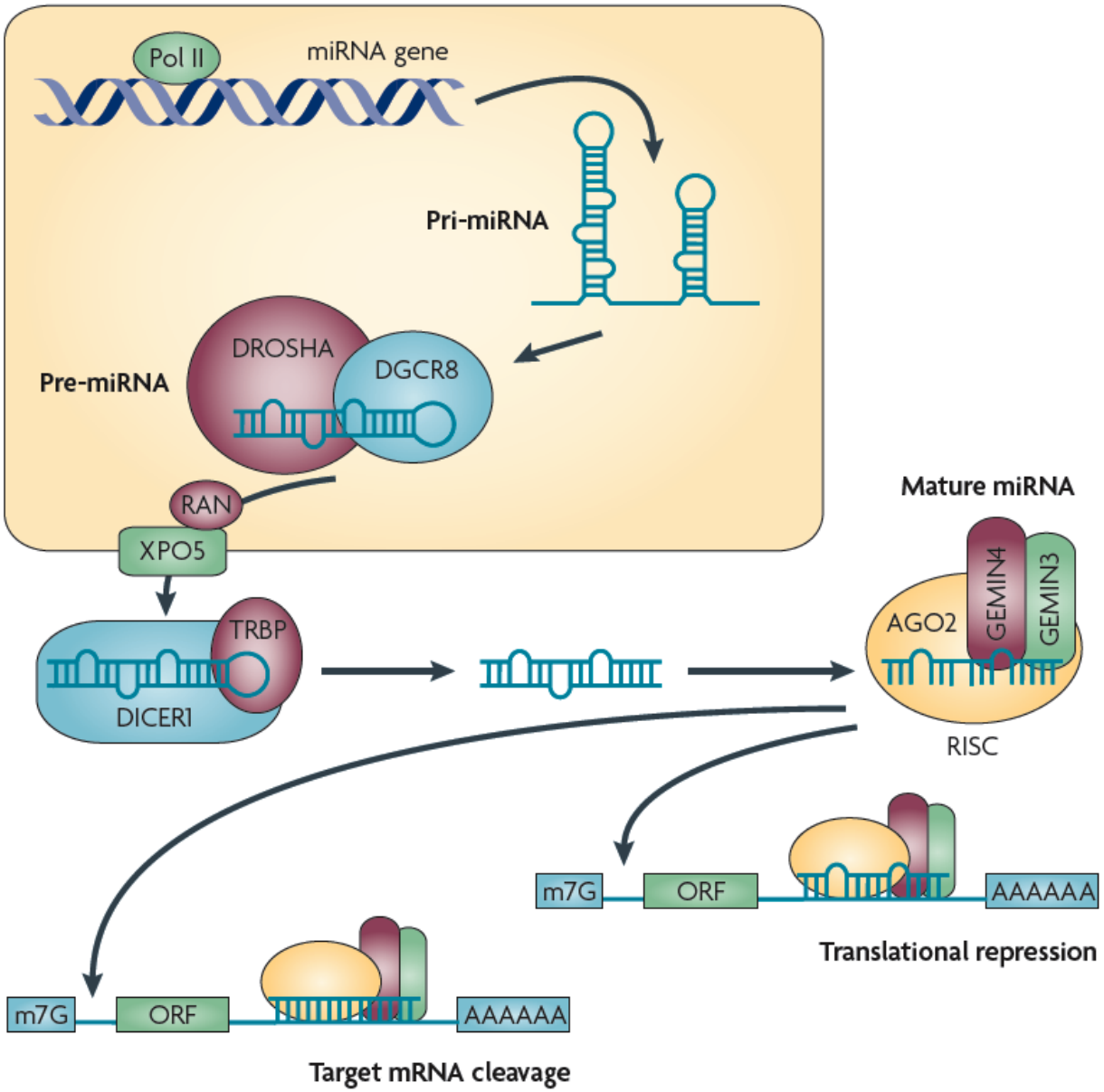
RNA polymerase II (Pol II) produces a 500–3,000 nucleotide transcript, called the primary microRNA **(pri-miRNA)**.

pri-miRNA is then cropped to form a **pre-miRNA** hairpin of ~60–100 nucleotides in length by a multi-protein complex that includes the protein **DROSHA**.

AA, poly A tail;
m7G, 7-methylguanosine cap;
ORF, open reading frame.

Ryan et al. Nature Rev. Cancer (2010) 10, 389

Overview of the miRNA network

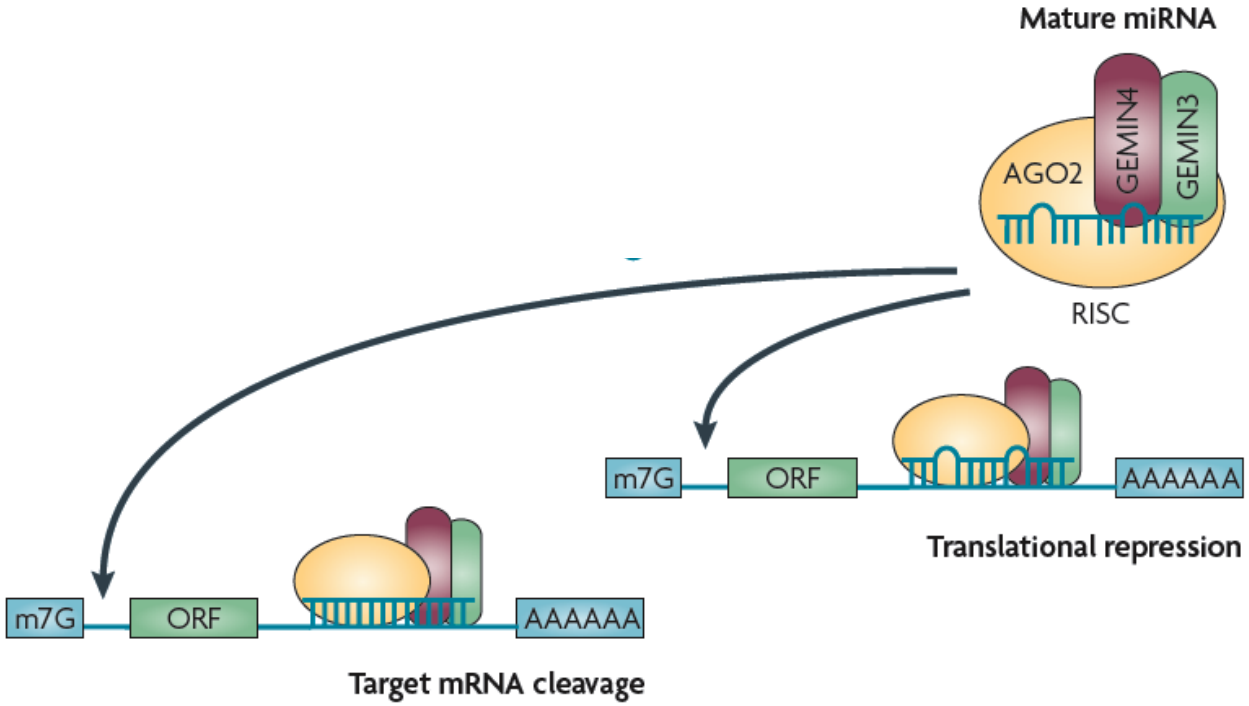


This double-stranded **pre-miRNA** hairpin structure is **exported** from the nucleus by RAN GTPase and exportin 5 (XPO5).

Finally, the pre-miRNA is cleaved by the protein **DICER1** to produce two miRNA strands:
- a mature miRNA sequence, approximately 20 nt in length,
- and its short-lived complementary sequence, which is denoted miR.

Ryan et al. Nature Rev. Cancer (2010) 10, 389

Overview of the miRNA network



The thermodynamic stability of the miRNA duplex termini and the identity of the nucleotides in the 3′ overhang determines which of the single strand miRNA is incorporated into the RNA-inducing silencing complex (**RISC**).

The RISC complex is then targeted by the miRNA to the target 3′ untranslated region of a mRNA sequence to facilitate **repression** and **cleavage**.

The main function of miRNAs is to down-regulate gene expression of their target mRNAs.

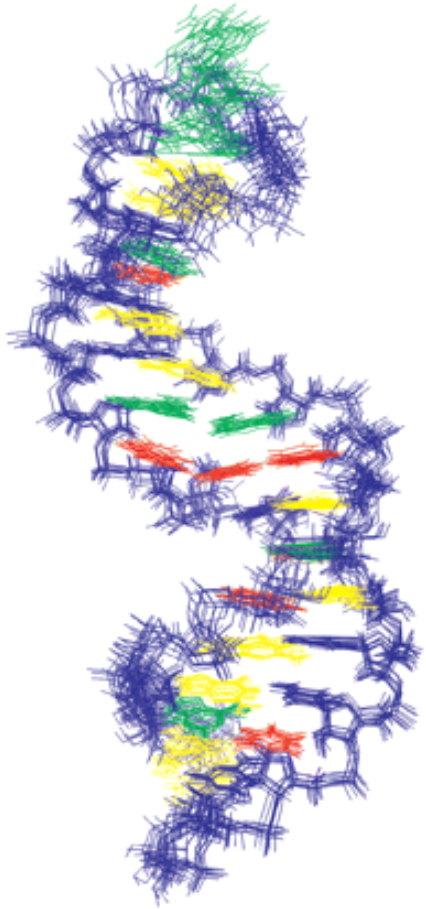
Ryan et al. Nature Rev. Cancer (2010) 10, 389

miRNAs

Mature miRNA molecules are partially complementary to one or more mRNA molecules.

Fig. shows the solution NMR-structure of *let-7* miRNA:*lin-41* mRNA complex from *C. elegans*

Cevec et al. *Nucl. Acids Res. (2008) 36: 2330.*



miRNAs typically have **incomplete base pairing** to a target and inhibit the translation of **many different mRNAs** with similar sequences.

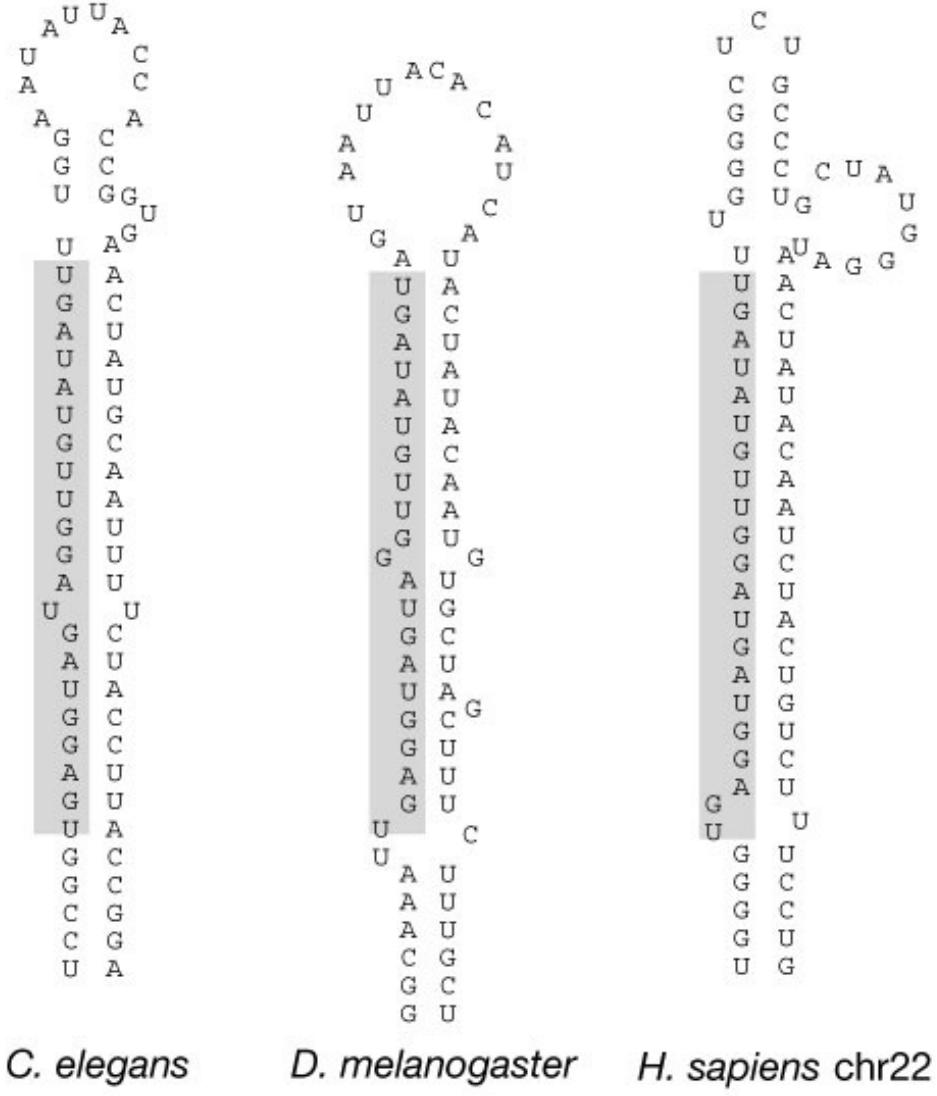
www.wikipedia.org

discovery of let7

The first two known microRNAs, lin-4 and let-7, were originally discovered in the nematode *C. elegans.*
There, they control the timing of stem-cell division and differentiation.

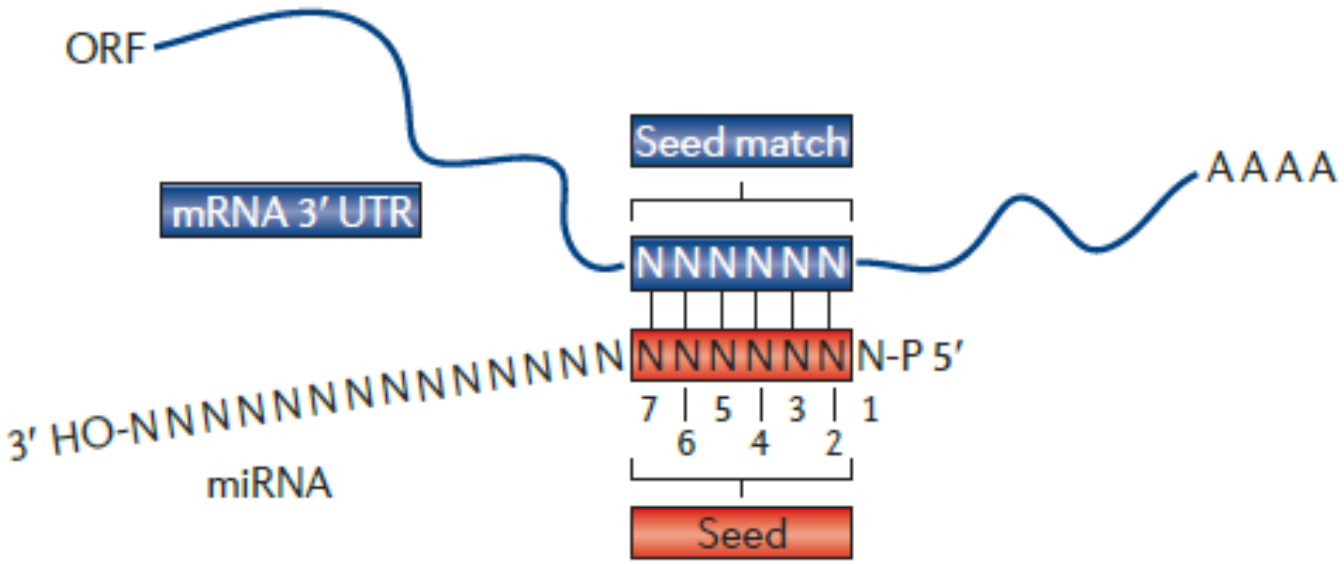
let-7 was subsequently found as the first known human miRNA.

let-7 and its family members are **highly conserved** across species in sequence and function.
Misregulation of let-7 leads to a less differentiated cellular state and the development of cell-based diseases such as cancer.



Pasquinelli et al. Nature (2000) 408, 86
www.wikipedia.org

miRNAs recognize targets by Watson-Crick base pairing



Animal miRNAs recognize **partially complementary** binding sites which are generally located in 3' UTRs of mRNA.

Complementarity to the 5' end of the miRNA – the "**seed**" sequence containing nucleotides 2-7 – is a major determinant in target recognition and is sufficient to trigger silencing.

Huntzinger, Izaurralde, Nat. Rev. Genet. 12, 99 (2011)

Tissue signature enrichment levels

Paper #9 contains something about microRNAs, pluripotency and cancerogenesis:

MicroRNA miR-302 inhibits the tumorigenecity of human pluripotent stem cells by coordinate suppression of the CDK2 and CDK4/6 cell cycle pathways
SL Lin, DC Chang, SY Ying, D Leu, DTS Wu
Cancer Res. 70, 9473-9482 (2010)
http://cancerres.aacrjournals.org/content/70/22/9473.long