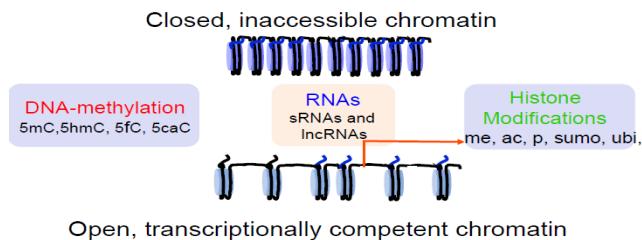


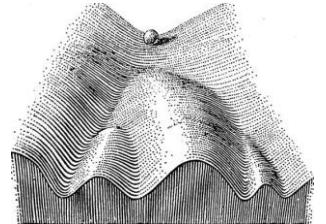
1 Epigenetics

Definition Reversible modifications of DNA or chromatin that affect the functions of the genome or the chromosomes in a heritable way.

Epigenetic layers



Epigenesis Differentiation of a multicellular organism from a totipotent stem cell that occurs in waves of divisions and differentiation. The > 200 cell types of our body contain the identical genome, but use it differently due to epigenetic modifications.



Canalisation The genotype and environment form a high dimensional epigenetic landscape and the differentiation of a cell from a totipotent cell is symbolized by a ball rolling down the hill. Evolution favours mechanisms that form cell type specific canals.

Non-mendelian Effects

- sex-specific effects of inheritance like genomic imprinting and X-chromosome dosage compensation
- variable penetrance or variable expressivity of genetic traits, which explains stable environmental effects
- inheritance of non-genetic effects over generations

Molecular Basis of Epigenetics The DNA of chromosomes is packed by histones in nucleosome units. The different histone types are linked to different functional regions. Histones and the DNA can both contain secondary epigenetic modifications that are reversible. Each cell type develops a characteristic pattern of epigenetic modifications along the chromosomes. A change in the environment alters the epigenetic landscape (DNA-methylome) of stem cells.

Epigenetics and Evolution Epigenetic mechanisms add new layers of genetic control

Epigenesis: increases the regulatory complexity at a constant number of genes

Genetic immunity: protect genomes against genetic alterations (by other genes)

Speciation: establish genetic barriers

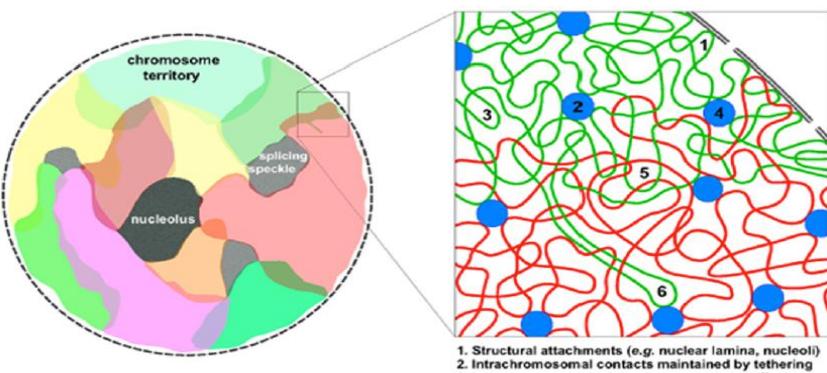
Adaption: allow rapid adaptation to environmental changes without genetic alterations

Epigenetics and Disease

- immunology (autoimmune reactions, immunological memory)
- blood related diseases
- age related diseases (Alzheimer, Parkinson,...)
- solid tumours
- complex genetic syndromes
- prenatal development and reproduction (environmental influences, reprogramming, aging)

2 Chromatin and Mechanisms of Chromatin Control

Chromosome Location Chromosomes are located in distinct territories in the nucleus that are dynamic and change between cell types and developmental stages. In the interphase they are transcriptionally active and non-replicative and form distinct TADs with different functions and chromatin condensation states. They are attached to the nuclear envelope (lamina) through proteins and specific regions.



Topology Associated Domains (TADs)

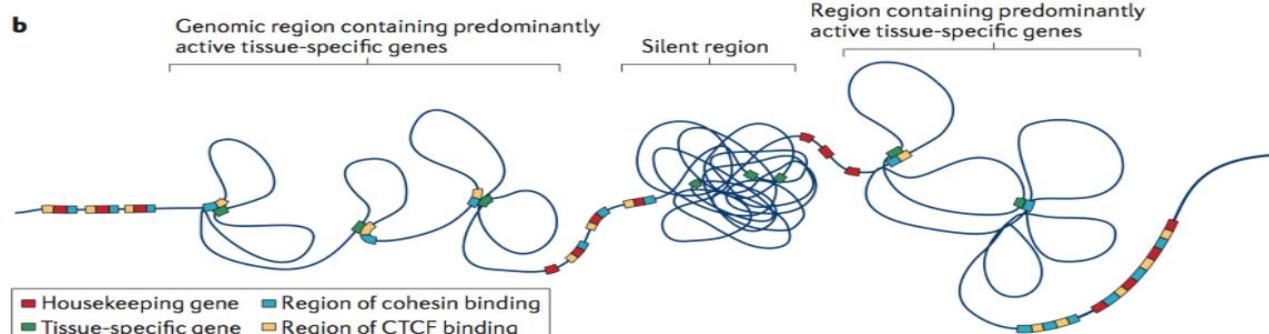
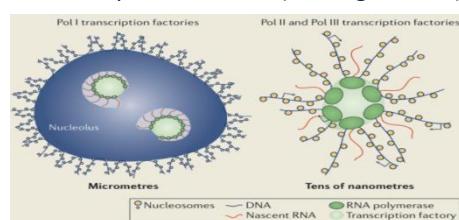
Regions that show high levels of interaction within the region and little or no interaction with neighbouring regions.

Central regions: high levels of chromatin interactions, tissue-specific genes and their enhancers. The interactions of the enhancers and the promoters is facilitated by CTCF (cohesin and CCCTC-binding factor)

Border regions: housekeeping genes, generally with enhancers and with high levels of CTCF that seems to prevent interactions between TADs

Transcription Factories

Gene promoters in domains are organised in transcription factories ("co-regulation").

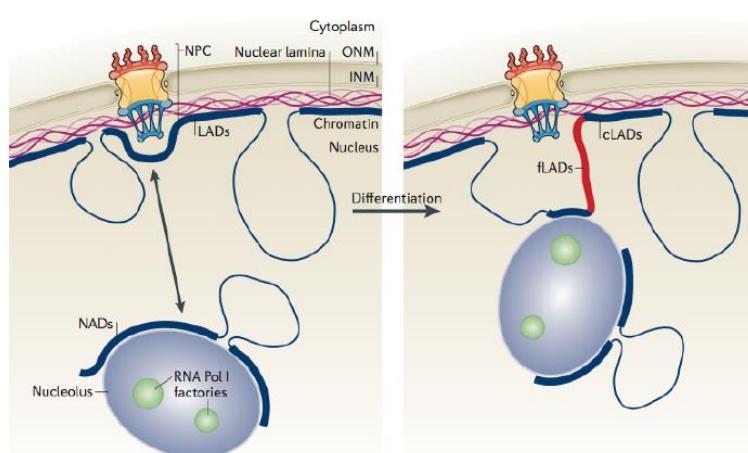


States of Chromatin

Euchromatin: less densely packed domains in the interphase nucleus, transcriptionally active

Heterochromatin: more densely packed domains in the interphase nucleus, not transcriptionally active

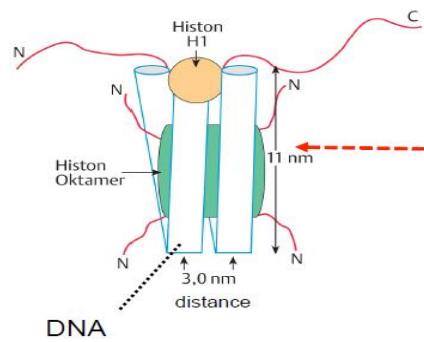
- *Facultative heterochromatic* regions are only condensed under certain conditions (differentiation) or phases of the cell cycle
- *Constitutive heterochromatic* regions are condensed in all phases and in all cell types. They are usually gene-poor and contain more repetitive elements. Blocks of them are found in the vicinity of centromeres or telomeres.



Lamina-associated domains (LADs)

Regions of heterochromatin that are bound by the nuclear lamina that contain silent genes. Regions of euchromatin loop out into the interior of the nucleus. As cells differentiate, constitutive LADs (cLADs) remain associated with the lamina, whereas facultative LADs (fLADs) become detached as their genes become active.

Nucleosome Basic unit of chromatin. It consists of 145-160bp of DNA wrapped around a histone octamer (2 identical tetramers each composed of 4 different histone proteins H2A, H2B, H3 and H4). A linker histone (H1) can be bound at the entry/exit of the DNA into the nucleosome. Histone “tails” are shown as red lines protruding from the core octamer.



Histones Most abundant proteins in the nucleus that are small (102-135AA), very basic, evolutionary well conserved, and have enriched number of arginines (R) and lysines (K). The protruding N- and C-terminal regions (tails) have unordered flexible structures and many (+)-charged amino acids. The Tails are the major docking stations for histone-DNA and histone-protein interactions that are modulated by post-translational modifications.

Histone Modifications

Most modifications are established or removed when histones are in nucleosomes.

<i>Residues</i>	<i>Functions regulated</i>
<u>Acetylation:</u> arginine (R), lysine (K)	transcription, repair, replication, condensation – open and active chromatin is acetylated
<u>Phosphorylation:</u> arginine (R), lysine (K)	transcription, repair, condensation
<u>Ubiquitination:</u> lysine (K)	transcription, repair – hub for protein docking
<u>Methylation:</u> serine (S), threonine (T)	transcription, repair

Histone type	modified amino acid	chromatin-effect: predominant localisation
H2A	K119Ub	Closing of the Chromatin: silencing of gene-domains
H3	K4me1	Active regulatory elements: Enhancer/Silencer
	K4me3	Local opening and activation of transcription: promoter
	K9Ac	Local opening and activation of transcription: promoter
	K9me2	Compaction of chromatin and transcriptional silencing
	K9me3	Gene-associated (facultative) heterochromatin
	K36me3	Gene-poor (constitutive) heterochromatin
H4	K27me3	Elongation/Progression of transcription: gene body
	K27ac	Local gene silencing: promoter
	R3me1/2	Local activation of genes: promoter
	K20me1	Activation of Transcription: promoter
	K20me3	Association with Transcription and repair: gene body
		Association with Transcription and repair: gene body

Prominent/important modification types of histone H2A, H3 and H4 in mammals/human. The positions and the modification type are presented according to the „Brno Nomenclature“: histone – aminoacid (K, R) – position (in the primary sequence) – type and number of modifications (me1 =mono, me2 =di, me3=tri-methylated). An example: H3K9me2 = Histone H3 modified at lysine 9 with two methyl groups. The right part of the table describes the effect of the modification on chromatin and the primary localisation in chromosomes.

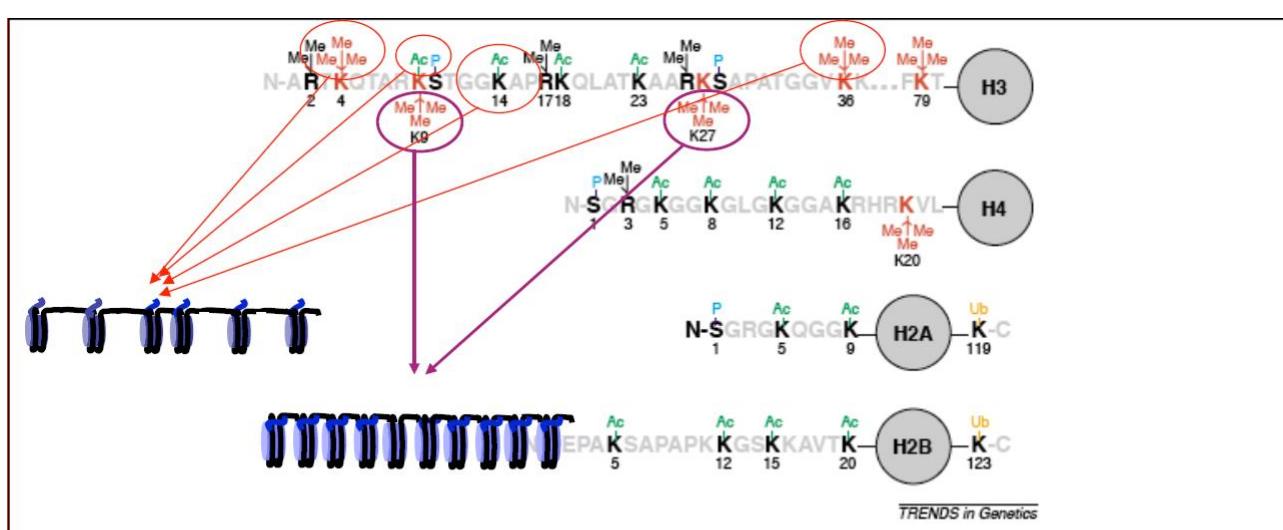
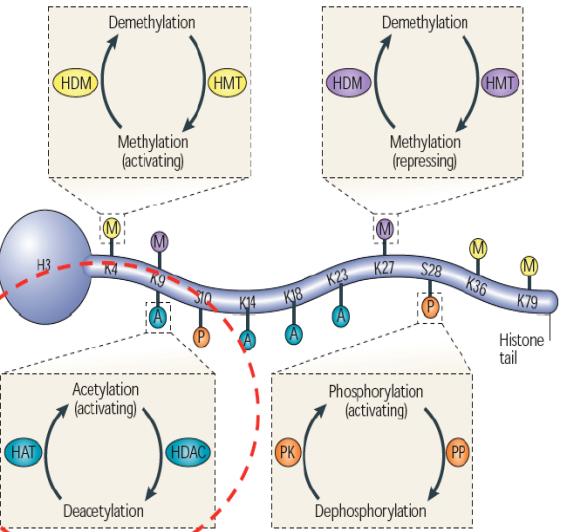
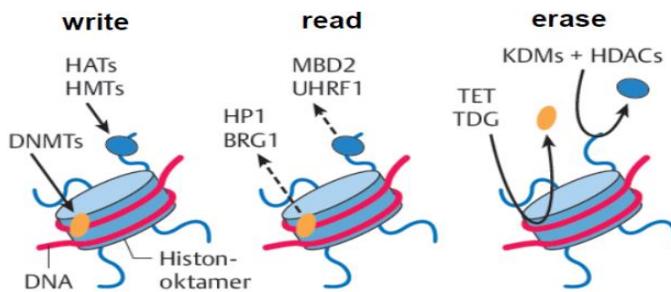


Figure 1. The known post-translational covalent modifications of histones (H2A, H2B H3 and H4). Lysine (K) methylation (Me) is represented in red. Acetylation (Ac), phosphorylation (P) and ubiquitination (Ub) are indicated in green, blue and orange, respectively. Arginine (R) methylation is represented in black. Modifications shown above each amino acid correlate with activation, whereas the lysine methylation shown below correlates with repression. The highest degree of methylation possible for each residue is displayed. Note that in H4, K20 methylation might also participate in activation.

Histone modifications are reversibly written, erased or read by sets of highly specific enzymes. The recognition depends on the amino-acid composition and the general protein structure context. Removal of a modification can either occur directly or by “exchanging” a modified histone by a non-modified histone. Arginine, Lysine and the polar Threonine and Serine residues are the major targets of (known) modifications.

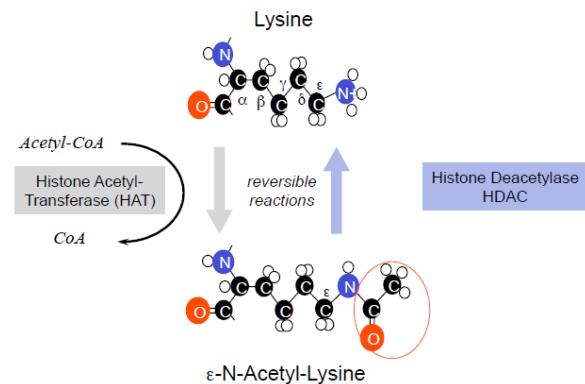


Histone Acetylation

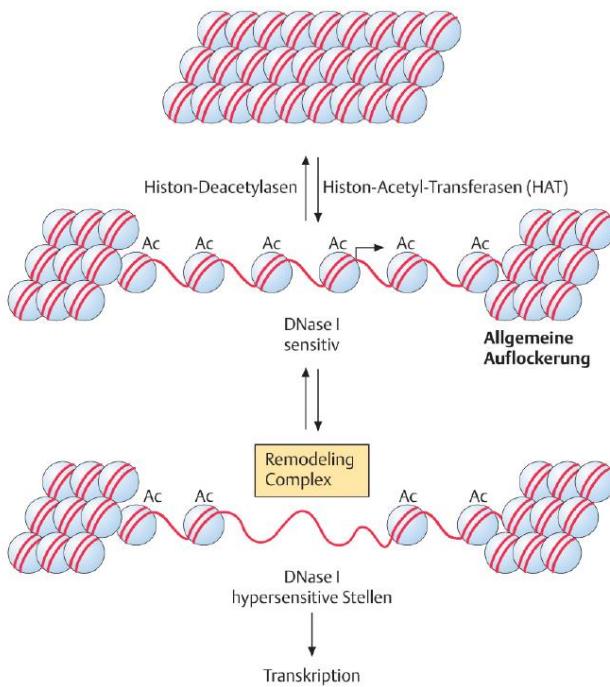
HATs and HDACs are often within complexes of transcription regulators and may also (de-)acetylate non-histone proteins and may modulate the activity of transcription factors in chromatin.

Histone-Acetyl-Transferase (HAT): They transfers acetyl-groups from Acetyl-Co-A to Lysine and are frequently associated with transcription-activating complexes like CBP/P300. Lysine-acetylation in histones H3 and H4 facilitates the opening of chromatin as the acetylation weakens the interaction of nucleosomes with DNA so that the nucleosomes can be moved.

Histone Deacetylases (HDAC): They remove the N-acetyl-groups from lysine residues in nucleosomes and are often associated with repressor complexes or heterochromatising factors. It can be inhibited by small molecules like Trichostatin A (TSA) or Vorinostat (SAHA) that are used for cancer treatment.



Opening or closing of chromatin for transcription:

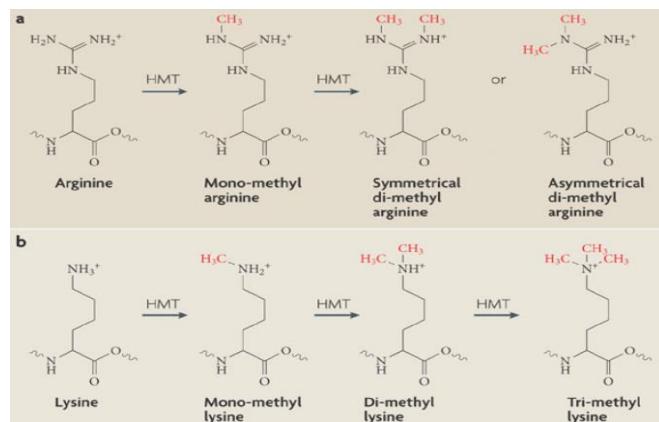


Histone Methylation

and lysines contributes to stable chromosome structures.

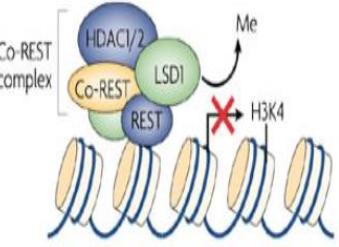
Histone-Methyl-Transferase (HMT): They contain a SET domain, often also chromo- or bromo-domains and they and PRMTs use the co-factor S-adenosyl-methionine (SAM) for the methylation reaction. They catalyse histone modifications.

PRMT: Protein-Arginine-Methyl-Transferase



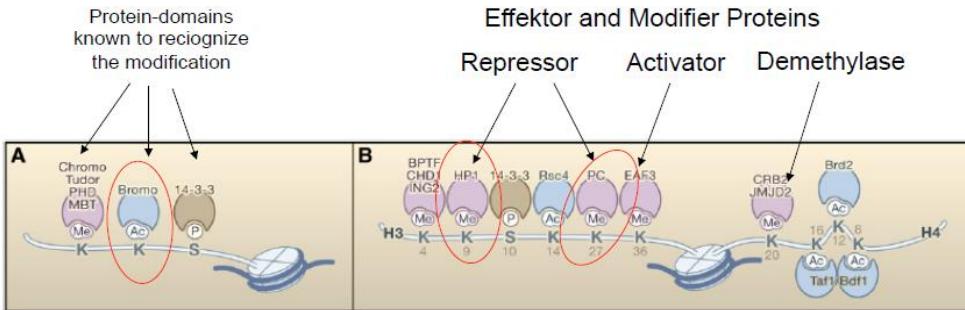
Histone-Demethylation Histone-Demethylase (HDM) removes the methylation.

LSD1: Lysine demethylases that are recruited to chromatin by repressor complexes to locally demethylate H3K4me3. It uses *Flavin Adenine Dinucleotide* (FAD) as a redox cofactor to covert and cleave formaldehyde from the histone. REST is a neuronal specific transcription factor. It forms together with Co-REST, HDAC1/2 and LSD1 a repressor complex binding to promoters of genes which have to be silenced upon development, i.e. neuron specific genes in non-neuronal cell types. Histone H3 becomes demethylated at K4 and deacetylated = transcriptional silent (repressed).



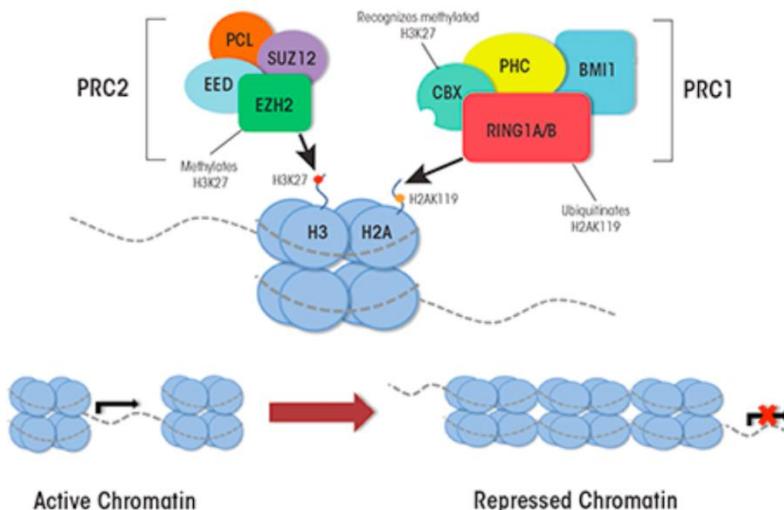
Jumonji-type (Umj1): Lysine demethylases use α -ketoglutarate as a cofactor. The cofactor is converted to succinate to finally oxidise CH_3 and to release formaldehyde from the histone.

Reading of Histone Modifications



- A) Domains used for the recognition of methylated lysines, acetylated lysines or phosphorylated serines.
- B) Proteins that associate preferentially with modified versions of histone H3 and H4.

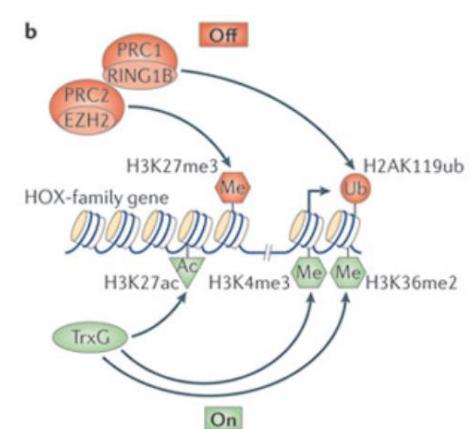
Heterochromatin Protein 1 (HP1): It stabilises heterochromatin structures and interacts with other heterochromatin forming proteins.



Polycombgroup-Protein (PcG): They are key components in the large PRC complexes.

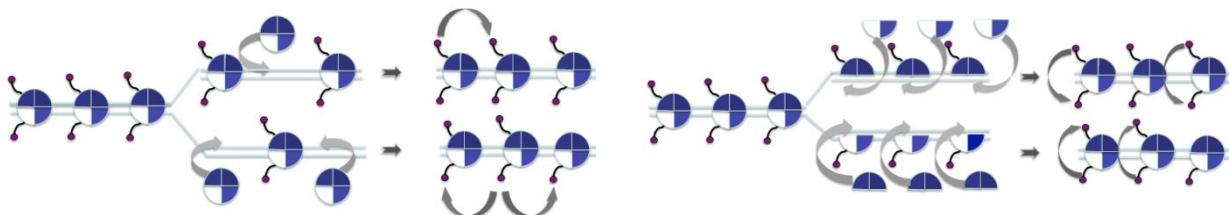
PRC complexes: They bind to polycomb response elements (PREs). PRC1 complexes contain Ring1b, an enzyme that ubiquitinylates histone H2B. PRC2 complexes contain histone modifying enzymes such as EzH. Together they form and stabilise local "facultative" heterochromatin at promoters of development genes (e.g. Hox genes). They are essential to silence such genes in cells where they should not be expressed.

Trithorax-Group-Proteins (TRX): They are transcriptional activators and contain characteristic BROMO domains through which they attract transcriptional regulators/activators. TRX are often in a complex with HMTs and HATs to locally establish open chromatin at promoters.



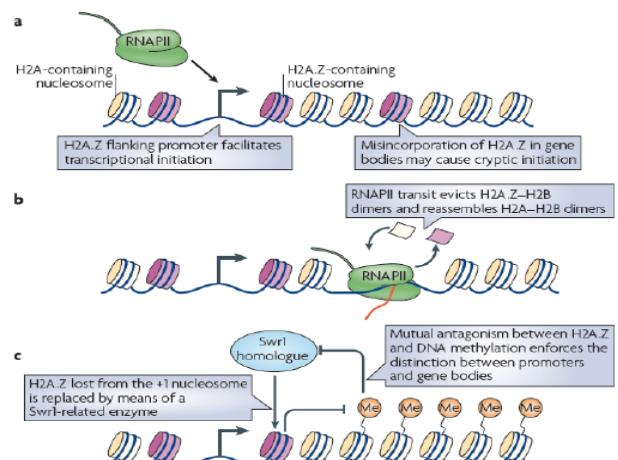
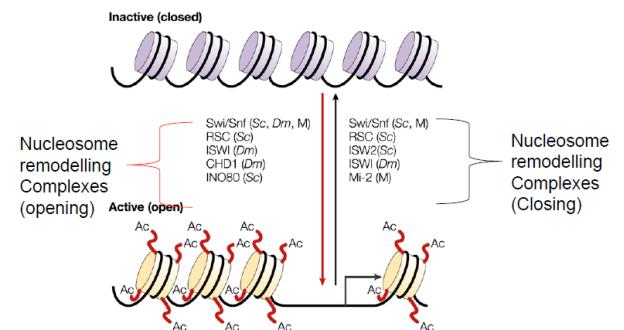
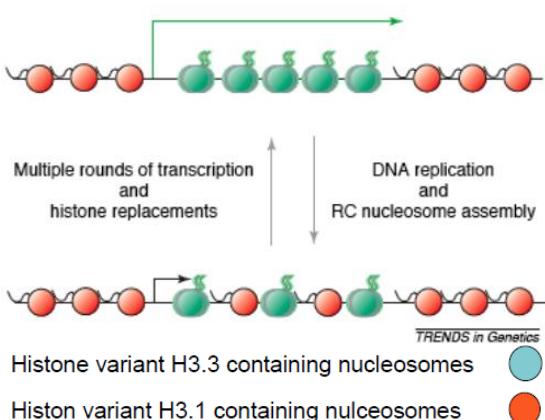
Nucleosome Assembly Unmodified histones are assembled by general histone "chaperones", like **Nucleosome Assembly Factor 1** (NAP1), into newly formed nucleosomes. Histone synthesis is activated before and during S-phase. During replication the number of nucleosomes is halved on the semi-conservatively replicated DNA. Nucleosome assembly factors such as NAP1 or CAF1 deposit the new nucleosomes onto the DNA. Single modified histone of dimers/tetramers can be exchanged by specific chaperons with unmodified histones or histone variants.

Model I and II: Inheritance of histone modifications after replication



Nucleosome Remodelling: The density and presence, like distribution and spacing, of nucleosomes in chromatin (on the DNA) is altered by enzymes in **Chromatin Remodelling Complexes** (CRC). These complexes are important to locally open the chromatin, by disrupting DNA-histone contacts, to induce transcription. Nucleosomes can also be moved along the DNA (or the DNA along the nucleosomes) to release or hide certain DNA sequences for recognition. The central “moving” enzymatic unit are ATP dependent catalytic subunits like ISWI or SNF.

Histone Variants There exist variant forms, with usually only a few changed amino acids, for all histones. Histone H3 has 4 main variants: H3.1, H3.2, H3.3 and CanpA, while histone H2A has 3 main variants: H2A, H2AZ, γH2A. These variants are placed in specific regions in the chromatin and are linked to specific functions. For instance, transcription is linked to the accumulation of the histone variant H3.3 and active promoters are marked by the presence of the histone variant H2AZ.



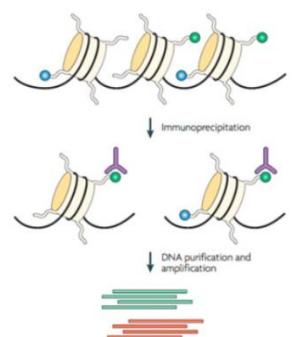
Methods for the Analysis of the Total Content of Modified Histones of a Cell

Immunofluorescence: Powerful technique that utilises fluorescent-labelled antibodies to detect specific target antigens, such as histones and histone modifications.

Western-blot: It is a widely used analytical technique used to detect specific proteins in a sample of tissue homogenate or extract. It uses gel electrophoresis to separate native proteins by 3-D structure or denatured proteins by the length of the polypeptide. The histone proteins are then transferred to a membrane, where they are stained with antibodies specific to the target protein and histone modifications.

Mass-spectroscopy: It is an analytical technique that sorts ions based on their mass (or “weight”) and results in a mass spectrum that will give a picture of the exact chemical composition of a sample.

Chromatin-Immunoprecipitation (ChIP): DNA-binding proteins are crosslinked to DNA and then the chromatin is isolated and the DNA with the bound proteins is sheared into small fragments. Antibodies bind specifically to the DNA-binding proteins to isolate the complex by precipitation. The cross-linking is reversed to release the DNA and digest the proteins. Then use PCR to amplify specific DNA sequences and analyse the locus and use NGS for a genome wide analysis.



3 Molecular Control of DNA Methylation

DNA-Methylation Specificity

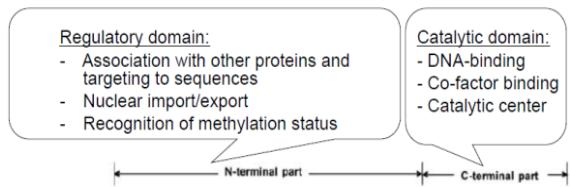
The DNA of most organisms contains modified cytosine bases (C5-Methyl-Cytosin = 5mC). Bacteria have enzymes modifying N4-Methyl-Cytosin (N4-mC) or N6-Methyl-Adenine (N6-mA). DNA-methylation occurs in a sequence context and is usually symmetrical on both strands. In bacteria many different enzymes are found which all have different sequence specificities with methylated bases.

CpG Islands in Mammals

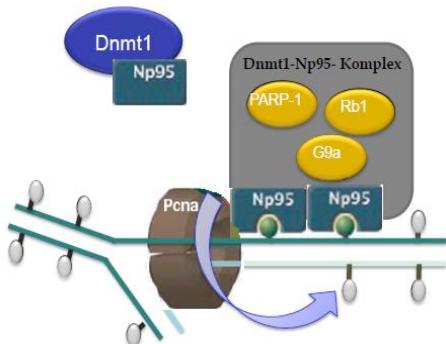
The CpG dinucleotide is with 2% relatively rare and not randomly distributed. Most CpGs are found in so called CpG islands which are mostly not methylated and occur in the 5' end of promoters in ~50% of all genes. Exceptions are CpG islands on the Xi and some imprinted genes which are methylated in one of the alleles. In stem cells cytosine can be methylated at non-CpG positions such as CNG or any CA(N). The methylation does not affect the base pairing.

Mammalian DNA-Methyl-Transferase (DNMT)

All DNA-methylations are post-replicative, almost exclusively confined to the ^{met}CpG sequence context and enzymatically catalysed by the DNMTs that interact with other proteins which direct them to the position they methylate.



DNMT1 It is a maintenance methyltransferase responsible for the inheritance (copying) of methylation patterns upon cell division. As such it has a major function during the S-Phase (DNA-replication) and prefers hemimethylated DNA as a substrate. It is important to regulate coordinated gene expression during development, silence promoters of "junk DNA" and stabilise genomic imprinting.



DNMT1 has a long amino-terminal domain for interactions with regulatory/modifying protein (-complexes) such as URHF1/PCNA for replication, HDACs for changes to histone-modification and co-repressors for targeting of the methylation reaction. Knockout of DNMT1 cause a genome wide demethylation and results in embryonic lethality.

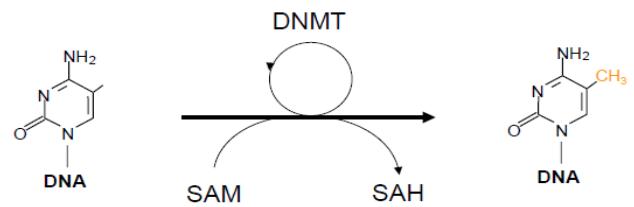
Regulation of DNMT1 Maintenance Activity: NP95 (URHF1) coordinates the setting and spreading of heterochromatic information through interaction with chromatin modifiers such as G9A, which is a H3K9me2/3 specific histone methyltransferase.

DNMT3A and DNMT3B They are de novo methyltransferases, but can also methylate hemimethylated DNA substrates. Both are very important in recognising and methylating repetitive elements in the genomes and are essential for development, despite their functions partially being redundant. Single knockouts are partially viable but double KOs cause an early embryonic lethality with defects in organ development.

DNMT3L It has no intrinsic methyltransferase activity, but guides DNMT3A and DNMT3B to targets. It is important for establishing maternal imprints during oogenesis and paternal imprints during spermatogenesis.

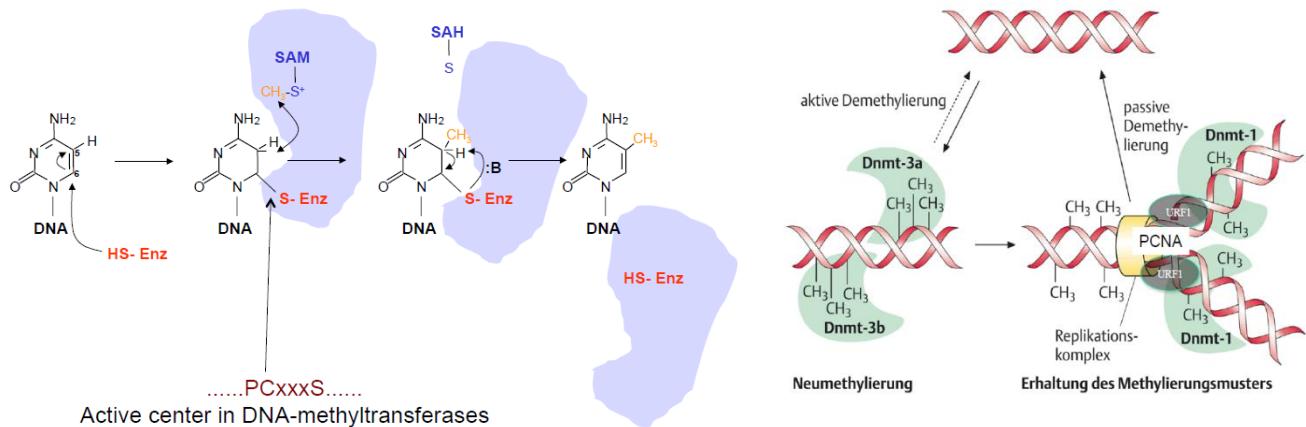
DNMT2 It has a typical DNA-methyltransferase structure, but it specifically methylates a cytosine near the anticodon loop in tRNA that protects the rRNA from degradation.

Enzymatic reaction 5'-CpG-3' dinucleotide in the DNA is recognised by the DNMT and a methyl-group is transferred from the methyl-group donor **S-Adenosyl-Methionine (SAM or "AdoMet")** to the carbon 5 (C5) of the cytosine ring. SAM is converted to **S-Adenosyl-Homocysteine (SAH)**.



DNA-Methylation Transitions

Passive demethylation leads to an unmethylated DNA. Active de novo methylation by DNMT3A and DNMT3B introduces new methylations that are then maintained by DNMT1.



Reading and Interpreting DNA-Methylation

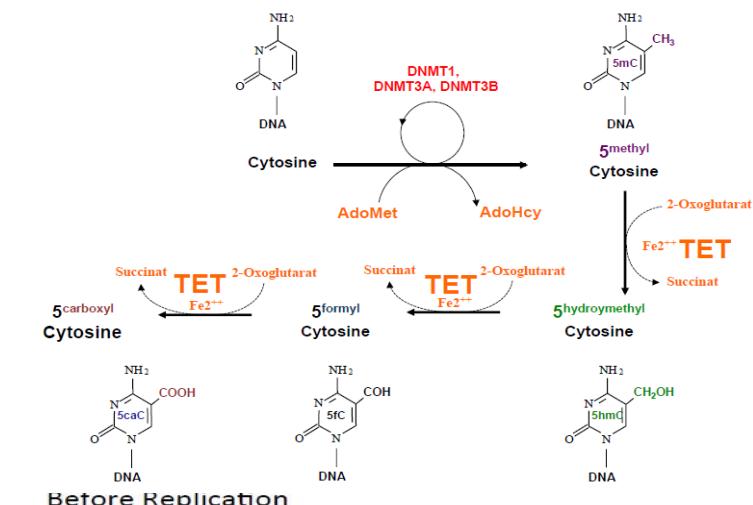
Methylated DNA Binding Domain (MBD): They are specific binding proteins that recognise DNA-methylation and are often present in chromatin modifying (repressor) complexes such as NuRD. Interpretation and reading of DNA methylation is mediated by methyl-CpG binding proteins containing MBD domains for 5meCytosine recognition. They mediate a “crosstalk” between heterochromatic histone- and DNA-modifications. Examples are MBD1, MBD2, MBD3, MBD4, MeCP2 and some methylation specific transcription factors like Kaiso and ZBTB33.

Detection Methods for DNA-Methylation

Direct: via selective chemical conversion of cytosines or modified cytosines and sequencing/array based assays.

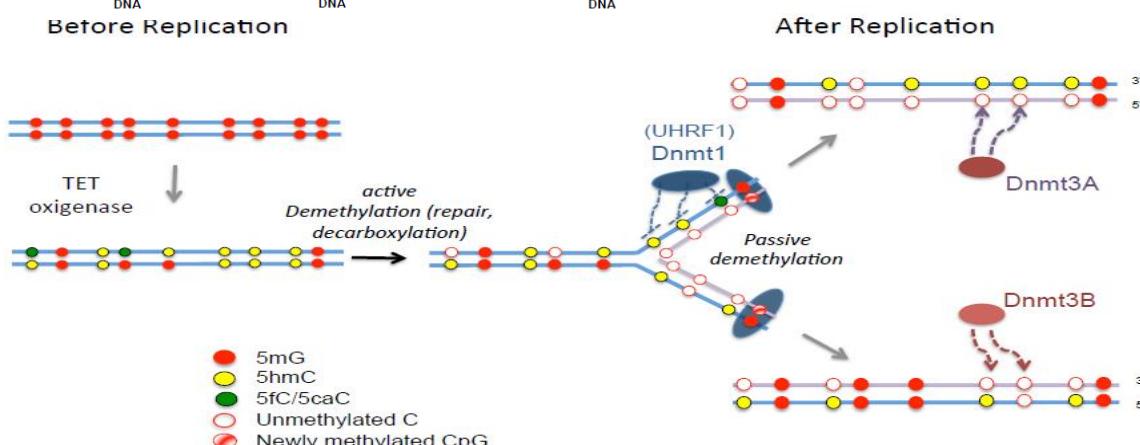
Indirect: By binding of 5mC specific antibodies and “pull down” of bound DNA followed by sequencing or arrays.

Oxidation of DNA-Methylation of Cytosines in Mammals

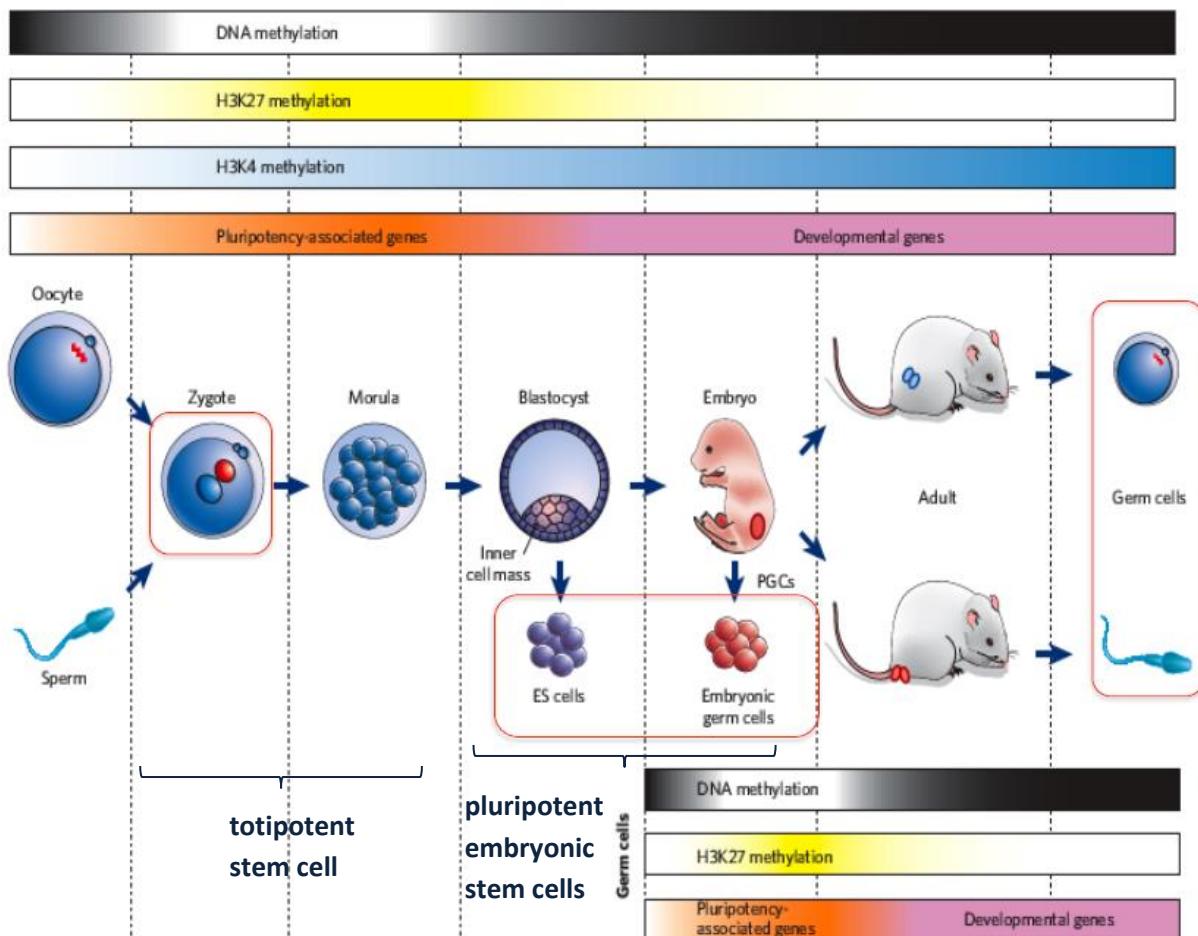


Oxidised forms are exclusively formed on 5mC. 5hmC, 5fC and 5caC are less abundant and their presence varies greatly from cell type to cell type. The highest amounts of 5hmC and 5fC are found in embryonic stem cells and in adult brains (neurons).

5hmC, 5fC and 5caC play a role in gene expression control, changes in the chromatin structure, control of epigenetic inheritance (= maintenance) and the induction of direct, active demethylation.



4 Epigenetic Programs During Development



Stem Cells Multipotent, somatic stem cells have a limited cultivation potential with the exception of induced pluripotent stem cells. Pluripotent embryonic stem cells can be cultivated indefinitely as they are immortal and can generate a complete new animal. Stem cells maintain a functional plasticity, have the capacity to undergo unlimited self-renewal and can differentiate into various cell types.

Development During development histone modifications and DNA-methylation are erased in the zygote to reach the epigenetic ground state of totipotency. Immortal, totipotent stem cells like the zygote then differentiate into specialised cells. The differentiation is controlled by changes in the gene expression in a temporal and cell type specific manner. Multipotent, somatic stem cells remain as precursors for cell renewal and differentiation, e.g. in haematopoiesis.

Genome-wide DNA-methylation changes during development and maternal and paternal chromosomes show asymmetric 5mC and 5hmC methylation, as 5mC is converted to 5hmC in the paternal pro-nucleus but not or less in the maternal one. The knockdown of TET3 affects the 5hmC conversion in the paternal pro-nucleus.

DNA-methylation changes and histone modifications occur side by side.

Epigenetic Reprogramming in the Zygote

1. The incoming sperm chromosomes are packaged in protamines, histone like proteins. The sperm membrane breaks down and the protamines are evicted. Then the paternal chromosomes are repackaged by histones coming from the egg cytoplasm. The histones are initially only partially acetylated on H4 but not methylated at lysines of the histone H3 tail. As a result, a new paternal pro-nucleus is formed with “normal” chromatin that is the naïve, unmodified ground state of nucleosomes before replication.
2. In G1, the paternal chromatin acquires gradual H3K4 methylation but not H3K9. The absence of H3K9me2 allows TET to recognise DNA methylation and convert 5mC to 5hmC.

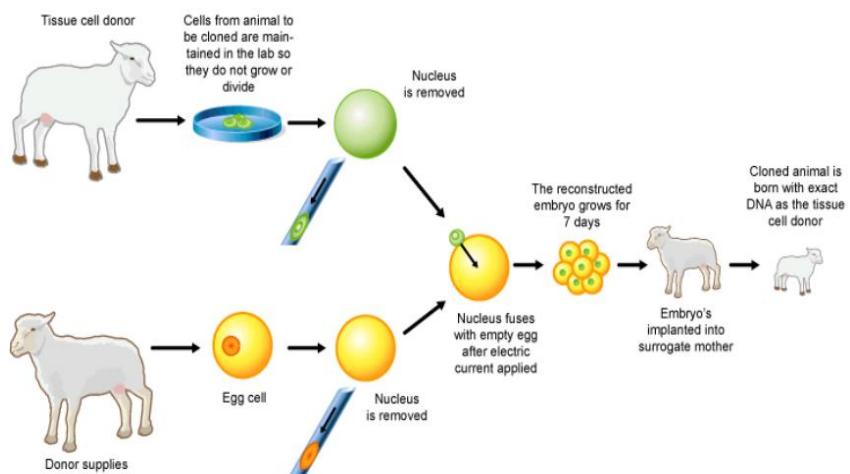
3. The maternal chromosomes have a higher level of histone modifications and 5mC is not oxidised to 5hmC, because the presence of H3K9me3 prevents the oxidation by TET3 by recruiting the Stella protein.

Hairpin-Bisulfite-Sequencing (HBS) shows that DNA-methylation is reduced after DNA replication and that hemimethylated CpGs accumulate, where only one DNA strand is methylated. The conclusion is that maintenance methylation is blocked.

Induced Pluripotent Stem Cells (iPS) Grow human primary cells and transfect them with gene cassettes expressing the four transcription factors Oct4, Sox2, Klf4 and c-Myc. Their overexpression induces the activation of endogenous stem cell programs and results in induced pluripotent stem cells that can then be used for patient specific cell therapy, drug screening and disease models. This reprogramming into stem cell states requires extensive epigenetic reprogramming.

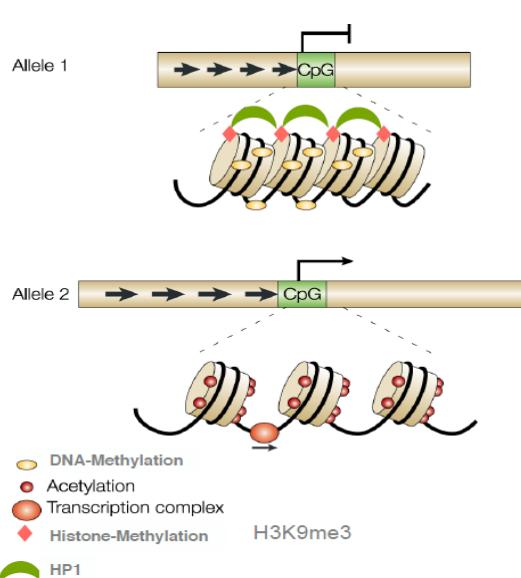
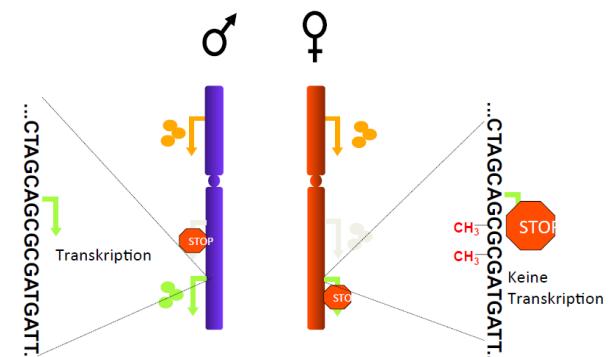
Nuclear Cloning and Epigenetic Reprogramming (Zygote)

The maintenance of embryonic stem cell (ES) pluripotency is dependent on the inhibition of the GSK3/FGFR signalling pathway. Inhibition through the 2i medium causes a genome-wide reduction of DNA-methylation. 2i cultivated ES-cells are reminiscent of pluripotent ICM cells of the natural blastocyst and DNA-demethylation is probably caused by 5hmC maintenance blocking and downregulation of de novo methylation (DNMT3A, DNMT3B and DNMT3L).



Epigenetic Reprogramming in the Germ Line

Genomic Imprinting Marking of genes by DNA-methylation in germ cells that are specific to the parental origin. Almost all 35000 genes are biallelic, meaning that they are expressed from both the mother and the father inherited chromosomes. Only 100 genes are expressed exclusively from one chromosome while the other is silent and differ epigenetically and transcriptionally. They regulate placental development and nutrient transfer from the mother to the foetus, the fetal growth and the behaviour after birth.



Uncontrolled imprinting leads to mild/severe developmental disturbances and the generation of complex syndromic diseases. The reason for that is an unbalanced expression of the genes in question.

Molecular Control

Allele 1 has DNA- as well as histone-methylation (H3K9me3) at the CpG island at the transcription start of the gene. HP1 binds to the histone-modifications and the result is tightly packed and transcriptionally inactive heterochromatin.

Allele 2 has no histone- or DNA-methylation at the CpG island at the transcription start of the gene, resulting in loosely packed euchromatin that can be transcribed by the transcription complex.

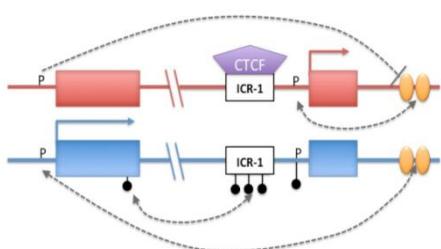
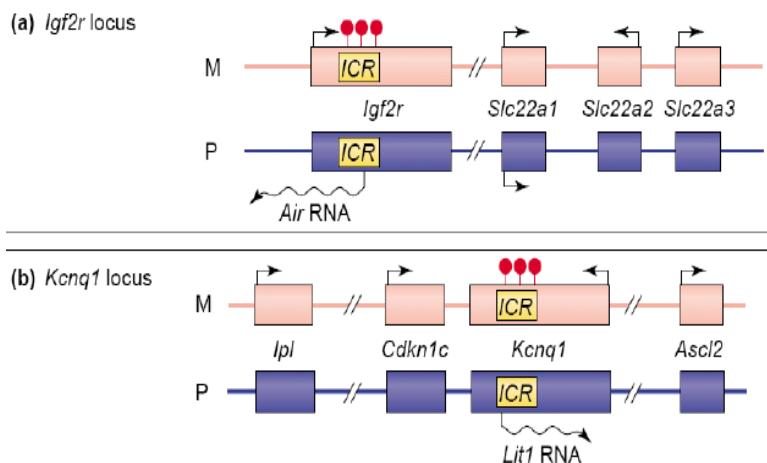
Imprinting Control Regions (ICRs)

They are Differentially Methylated Regions (DMRs) whose differential methylation plays a causal role in imprinting and function as regulators. They are also marked by differentially methylated chromatin on the active and inactive chromosome. They are essential for controlling the epigenetic status of many genes in the clusters of imprinted genes by mediating the silencing of one of the chromosomes. Most ICRs are **maternally** and **few paternally derived** and they inherit a germline specific methylation imprint.

They are usually **CpG islands** and function either as **promoters** or **domain boundary elements** separating enhancers from promoters. The mechanisms of transcription control at each imprinted gene or region are slightly different for different imprinted genes or clusters.

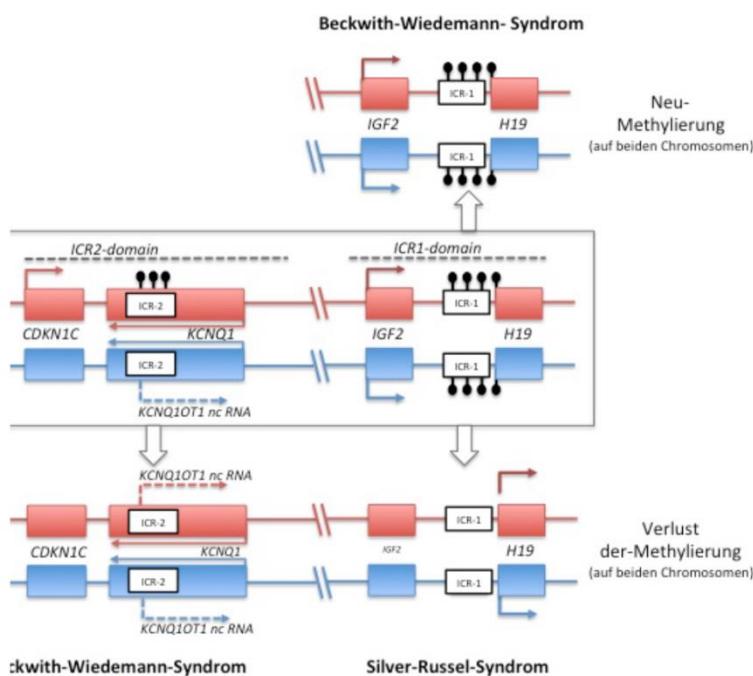
Regulation of Imprinting in the Promoter of an (Antisense) Transcript to the Imprinted Gene:

The **antisense transcripts Air and Lit1** originate in **introns**. Both antisense transcripts are made on the **paternal chromosome**, while the **DNA-methylation imprint** repressing the transcription of these antisense transcripts is on the **maternal chromosome**. Both ICRs are CpG rich promoters and the CpG is methylated on the maternal and unmethylated on the paternal chromosome. Both ICRs are established in the maternal germ line as regulatory **imprints**. The antisense AIR RNA induces silencing in cis by its transcription through the sense promoter of the paternal IGF2R by DNA-methylation on the paternal Igf2R promoter (not shown).



Function as an Insulator: The ICR is located between two genes and controls the domain structure allowing or preventing the interaction of promoter and enhancer. The ICR function depends on the binding of the CTCF insulator to the ICR. If CTCF binds the enhancer can interact with the H19 promoter and causes strong H19 expression. If CTCF is not bound (methylated maternal chromosome) the enhancer engages with the Igf2 promoter causing strong, exclusive Igf2 expression.

Consequences of Abnormal Imprinting at ICRs:



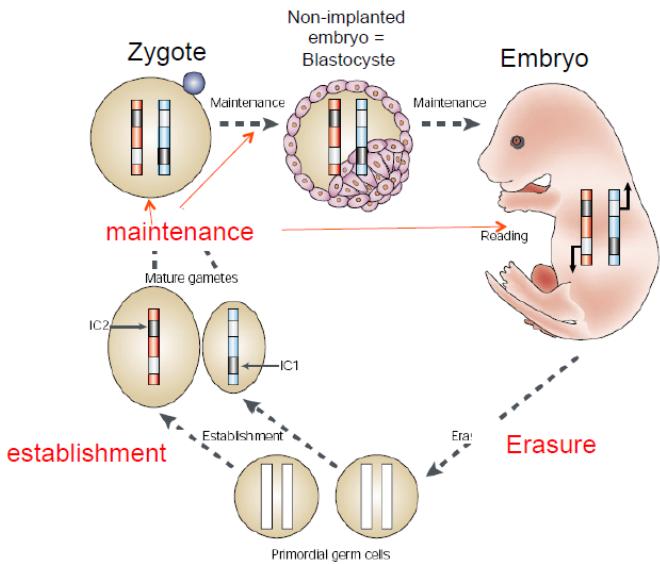
Both new methylation of the ICR1 domain on both chromosomes and loss of methylation of the ICR2 domain on both chromosomes lead to **Beckwith-Wiederman-Syndrome**, which results in fetal overgrowth and higher cancer risk. Loss of methylation of the ICR1 domain on both chromosomes leads to **Silver-Russel-Syndrome** which leads to underweight at birth.

Life Cycle of Methylation Imprints

Imprinting markers undergo a cycle of erasing, setting and maintaining per generation.

Erasure: Early germ cells have the “normal”, methylated epigenome of a somatic cell. Upon development into primordial germ cells when the migrating cells enter the gonads, they dramatically loose DNA-methylation and completely demethylate imprinted genes, resulting in the lowest DNA-methylation status. This is completed by embryonic day 12-13.

Establishment: Sex-specific imprints are established by **de novo methylation** by **DNMT3A** and **DNMT3L** during the development of germ cells into mature oocytes or sperm cells. The oocytes are in **meiotic arrest** and methylation occurs during their growth, while in spermatocytes the methylation occurs **before meiosis**. Most primary imprints (= ICRs) are established in the female germ line.



Maintenance: Parent specific methylation imprints are maintained after fertilisation.

Translation: The primary ICR imprint is translated into neighbouring genes/promoters during development.

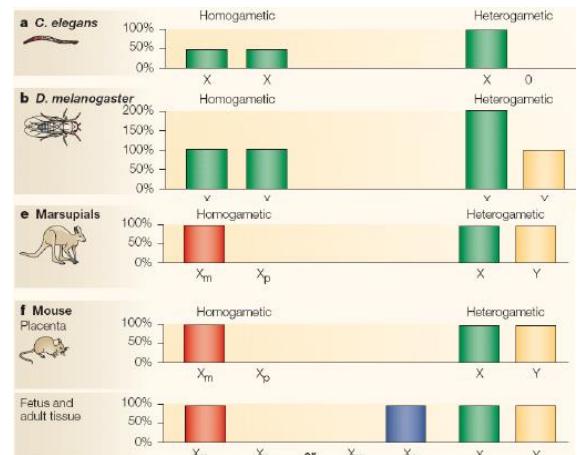
Development of Uniparental Embryos Embryos derived from zygotes with **two paternal genomes** develop very poorly and show an **enhanced growth of extraembryonic** (placental/yolk sack) tissues. Embryos derived from zygotes with **two maternal genomes** develop **rather normally until mid-stage embryogenesis**. However, they show a **reduced** development of the extraembryonic tissue (trophoblast/placenta and yolk sack). Both “uniparental” embryonic do not **survive to birth**; the development and normal phenotype of embryos requires a diploid set of two chromosomes, one from the father and one from the mother.

Uniparental Disomy It's the inheritance of a pair of chromosomes or a chromosomal region from a single parent. It can be created experimentally through balanced chromosomal translocations, or occur naturally by “correction” of aneuploidies (monosomies, trisomies) at very early stages of embryonic development. The “correction” refers to the accidental loss of a chromosome or accidental chromosome duplication during mitotic divisions.

Uniparental Disomy of chromosome 15: **Maternal disomy or paternal deletion** leads to **Prader-Willi-Syndrome** while **paternal disomy or maternal deletion** leads to **Angelman-Syndrome**.

Dosage Adjustment/Compensation on Sex Chromosomes In most organisms with heterogametic sex determination (X;Y) a genetic regulatory mechanism operates to equalize the activity of sex chromosomes such that the amount of **gene products from the X chromosome is equal in both sexes**. In mammals the gene dose of genes on the X-chromosomes is regulated such that most genes on one of the X-chromosomes are completely silenced.

In all species with dosage compensation, equalization of gene dosage is achieved at the level of transcription. This can be achieved by different mechanisms: In *C. elegans* hermaphrodites (XX), both X chromosomes are partially repressed. Drosophila males (XY) double the expression of genes along the X chromosome. In human females (XX), one chromosome is inactivated (XCI), resulting in a heterochromatic and largely genetically inactive Barr body. Any of these mechanisms results in balancing the relative gene expression between males and females (or, in the case of *C. elegans*, hermaphrodites and males).

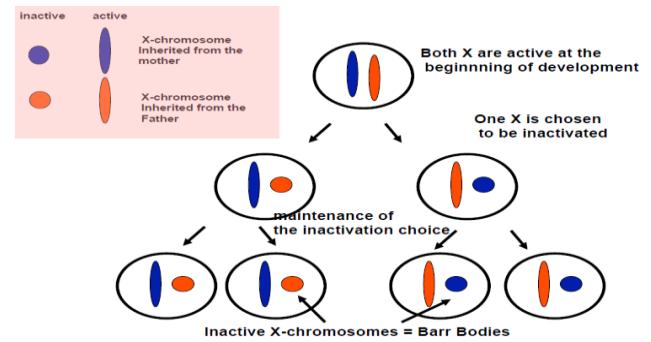


X-Chromosome Inactivation (XCI) in Mammals and Humans

Barr Body: Female (but not male) cells contain a heterochromatic structure called the Barr body. The Barr body is generally located on the periphery of the nucleus, replicates late and is an inactivated X-chromosome.

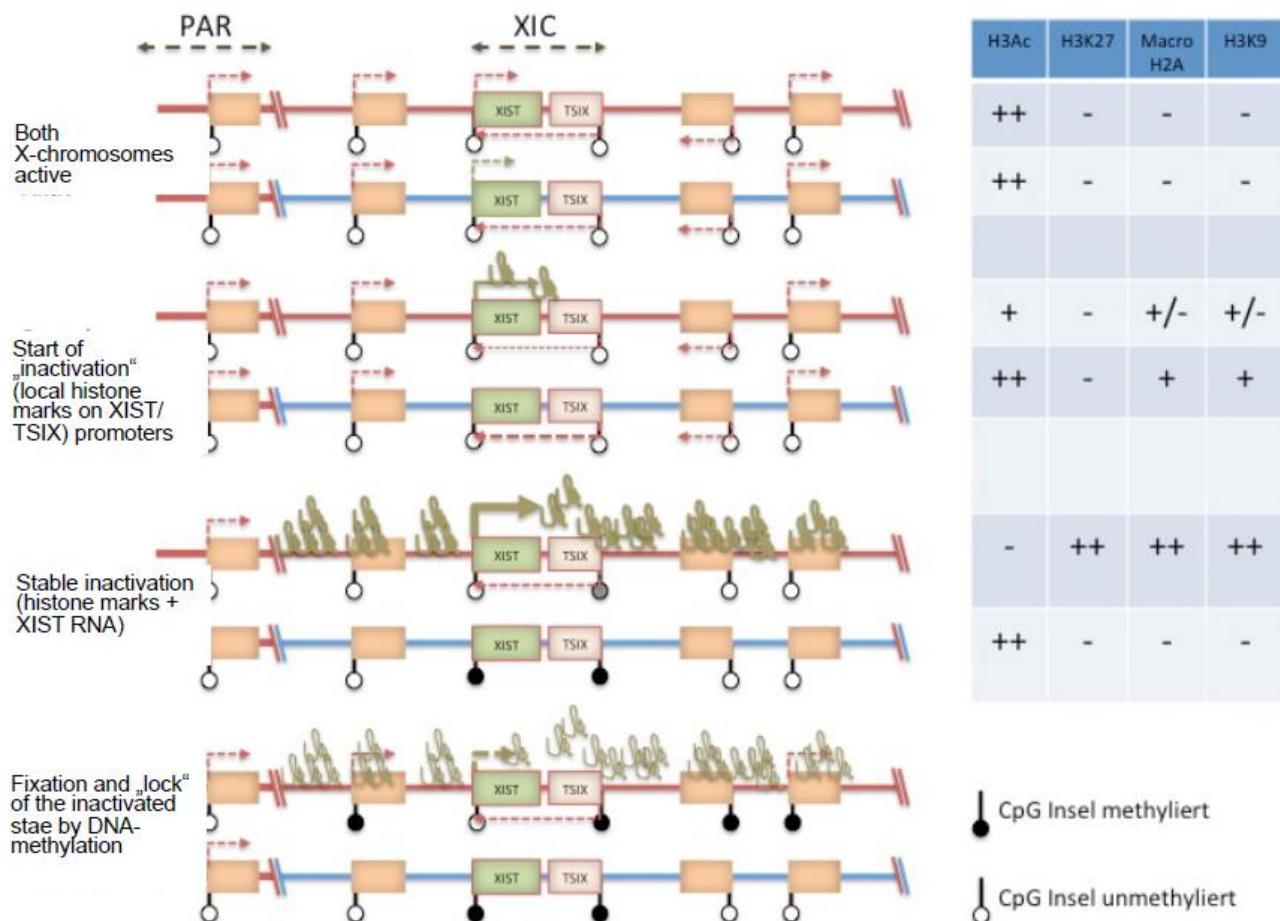
Lyon Hypothesis

In all female cells one of the two X-chromosomes is early in embryonic development randomly inactivated in individual cells, which results in condensation and the Barr body formation. The originally inactivated X-chromosome remains inactive in all daughter cells in subsequent cell divisions (clonal inactivation). As a result mammalian females are X-chromosome mosaics and even identical twin females can have quite different patterns of X-linked gene expression.



Molecular Mechanisms of X-Chromosome Inactivation

- Counting: Count the X-chromosomes; all but one become repressed per diploid autosome set.
- Choice: Which X-chromosome should become inactivated.
- Initiation: Initiate the inactivation by silencing the XIST allele.
- Spreading: The inactive state progresses by spreading heterochromatin and XIST RNA across the entire chromosome.
- Maintenance: Maintain the inactive state by introducing DNA-methylation in the CpG island promoters of X-linked genes on the inactive X-chromosome.



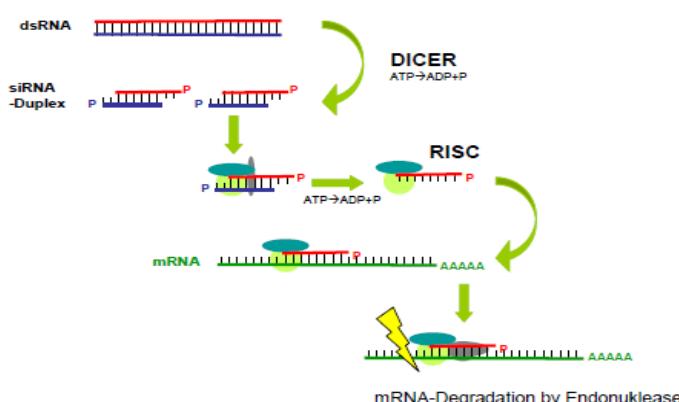
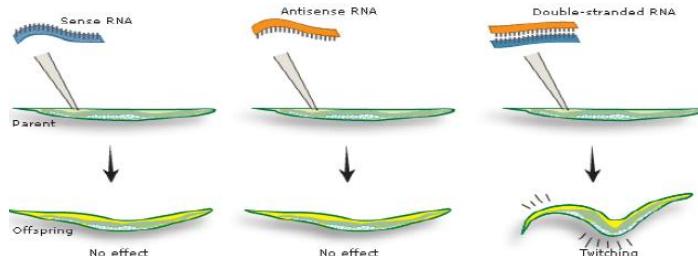
In embryonic stem cells both X-chromosomes are active, as Xist is repressed by the expression of the stem cell factors Oct4, Nanog, Sox2 and Rex1. During early differentiation they get downregulated and release Xist from repression and allows the initiation of the X-chromosome inactivation.

5 RNA Interference and RNA-induced Epigenetic Modifications

RNA Interference (RNAi) It's a *eukaryotic* mechanism modulating activity of gene expression using small RNA molecules that lead enzymes homology-dependent to the DNA or RNA. There are different RNAi pathways and with that many possibilities to knock down gene expression. In **Transcriptional Gene Silencing (TGS)** siRNA leads to chromatin and DNA modifications, while in **Post-Transcriptional Gene Silencing (PTGS)** siRNAs lead to mRNA degradation or repression of translation.

Fire and Mello Experiment

They injected RNA corresponding to a gene important for muscle function in the worm *C. elegans*. Single-stranded RNA (sense or antisense) had no effect. But double-stranded RNA caused the worm to twitch in a similar way to worms that lack a functional gene for the muscle protein.



Post-Transcriptional Gene Silencing (PTGS)

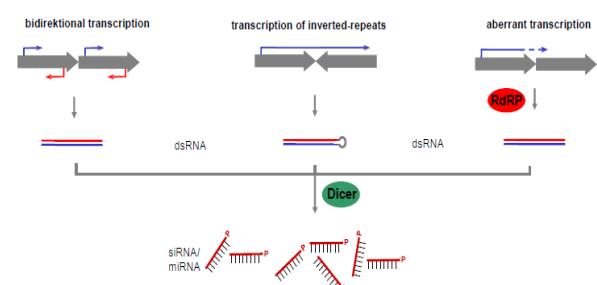
Double-stranded RNA (dsRNA) initiates RNAi by activating the DICER that cleaves the dsRNA into smaller siRNA-duplexes using ATP. The RNA-induced Silencing Complex (RISC), whose active components are Argonauts, binds to the siRNA and splits it into single strands. RISC and one siRNA-strand search the whole transcriptome until they find a 100% homologous position on an mRNA. The mRNA is then degraded by endonucleases at that position.

DICER: The DICER is an **endoribonuclease** that cleaves dsRNA into small, also double stranded RNA duplexes of 20-25nt with a 2nt overhang at the **3'-end** and a monophosphate.



Argonaut/Piwi: The Argonauts are components of RISC and bind specifically to individual classes of small RNAs. They interact with enzymes of subsequent processes and therefore transmit the biological function of small RNAs. Some Argonauts have catalytic functions like **RNAse III** and **SLICER-activity**.

siRNA and miRNA Synthesis The siRNA/miRNA synthesis happens by bidirectional transcription, transcription of inverted repeats or aberrant transcription that results in dsRNA. siRNAs are cleaved from the dsRNA by the DICER. They often **cis-regulate** processes and are exogenously applied dsRNA to trigger RNAi phenotypes. miRNAs are **cleaved from hairpins** and their synthesis differs between animals and plants as there are **more steps in the nucleus of plants**.



mRNA Sequencing

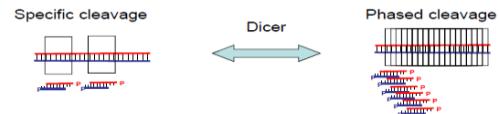
1. total RNA extraction
 2. 1st strand cDNA
 3. 2nd strand cDNA
 4. cDNA fragmentation
 5. adapter ligation
 6. PCR amplification
 7. sequencing
- Diagram illustrating the mRNA sequencing process: total RNA extraction leads to 1st strand cDNA, which is then converted to 2nd strand cDNA. This is followed by cDNA fragmentation, adapter ligation, PCR amplification, and finally sequencing.

Small RNA Sequencing

1. small RNAs gel extraction
 2. ligation of 5'- and 3'-adapters
 3. 1st strand cDNA synthesis
 4. PCR amplification
 5. sequencing
- Diagram illustrating the small RNA sequencing process: small RNAs are extracted from a gel, then ligated with 5'- and 3'-adapters. This is followed by 1st strand cDNA synthesis, PCR amplification, and finally sequencing.

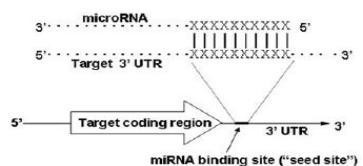
Small RNA Sequence Preferences

Sequence preferences can be due to specific DICER cleavage or due to specific Argonaute loading and subsequent stabilization.

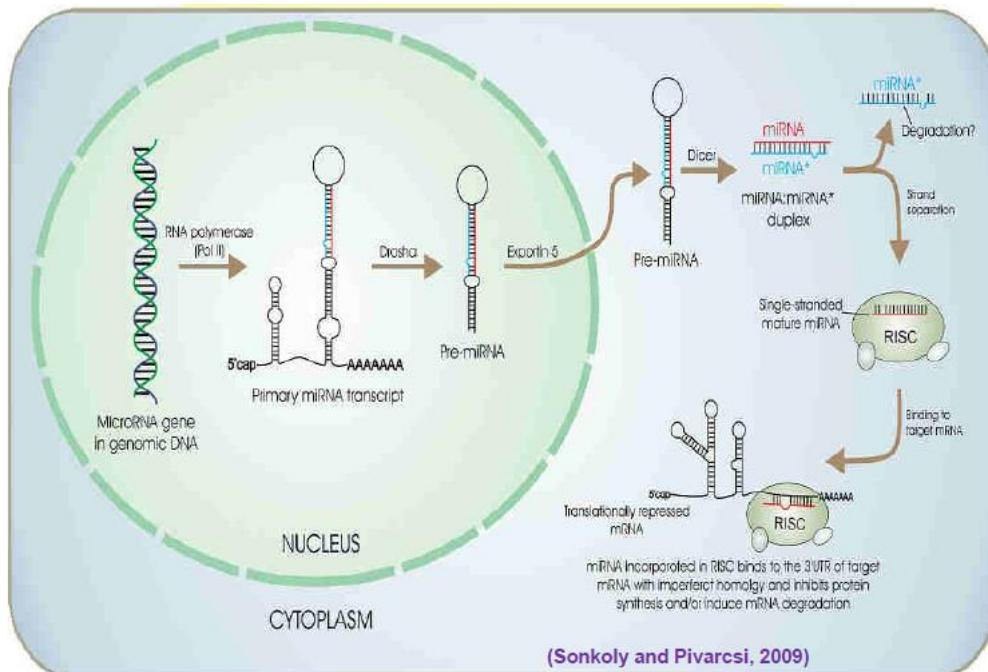


miRNAs miRNA genes encode for miRNA precursors (hairpins) and control gene expression in trans. The miRNA precursor hairpin RNAs are imperfect and Drosha and DICER specifically recognise mismatches that are not tolerated as they are required for precise recognition and cleavage; this cleavage is precise without phasing. **miRNAs don't bind 100% to target mRNAs and these mismatches are allowed and required in animals while it's not allowed in plants.**

miRNA "Seed" Sequence: 6-7nt signature region at the 5'-end that need to fit completely to the target and direct **miRISC** to target **3'-UTR of mRNA**. Secondary structures in mRNAs may inhibit targeting of miRNAs and seed sequences.

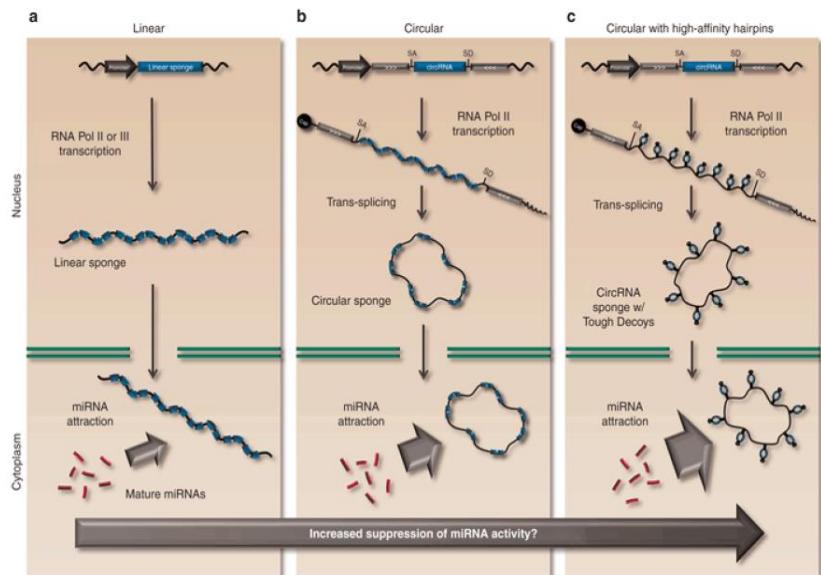


Classical miRNA Mechanism:



(Sonkoly and Pivarcsi, 2009)

Exogenous miRNA Sponging: It's an approach in gene therapy to reduce tumour specific miRNAs. This entails transformation of tissue and cells with constructs that produce an **RNA** which can **bind and catch miRNAs** quantitatively so that they cannot attack **mRNAs**. **Circular RNAs**, which are a product of **alternative trans-splicing events**, are indicated to be an **endogenous miRNA sponge** and therefore a negative regulator for miRNAs and a positive regulator of gene expression.



Artificial Induction of RNAi

It's possible to artificially induce RNAi through transformation of transgenes or by transformation of dsRNA or siRNA.

RNAi as a Tool Gene knockout mutants lead to a total knockout with **no proteins remaining**. The verification of this knockout happens on the **DNA-level** and it's not possible to verify further mutations. **Lethal knockouts** make it impossible to observe a phenotype. Knockout-RNAi-lines on the other hand allow a variable efficiency of knock down and it's possible to verify them on the **RNA-level** as well as verifying specific silencing or transitivity. Even with lethal knockdowns it's still possible to observe a phenotype in the less efficient RNAi-lines to characterise protein function.

siRNAs and siRNA Amplification Mechanisms

RNA-dependent RNA Polymerase (RdRP): It either **synthesises long dsRNAs** from **aberrant ssRNA templates** that the get cleaved into siRNAs, or it **directly synthesises siRNAs** from single strand mRNA templates.

Secondary siRNA Synthesis in Plants: At the end RdRP **synthesises dsRNA** in full length that then get cleaved into **siRNAs by DICER**, resulting in an amplification of the silencing signal.

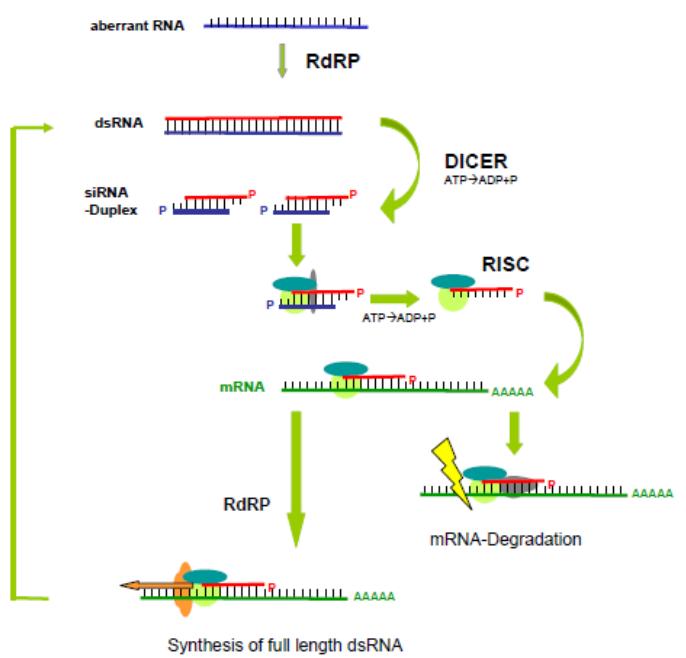
Secondary siRNA Synthesis in C. elegans:

At the end RdRP **directly synthesises secondary siRNAs** that then lead to mRNA degradation and the amplification of the silencing signal.

Secondary siRNAs and their amplification of the silencing signal is import for **spreading virus resistance**.

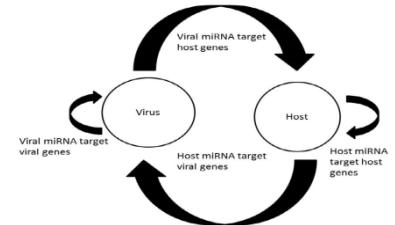
RNAi as an Antivirus Mechanism

RNA-based Immunity: **Viral RNA** can be the source of **siRNAs and miRNAs** as invading viral DNA is processed by DICER into siRNAs that then target the viral genome.



miRNA in Host-Viral Interaction: There are two classes of miRNA involved: the cellular, host miRNA which interferes with virus replication by targeting viral genes, and viral miRNAs which interfere in host gene expression by targeting host genes. The virus encoded miRNA genes use the host RNAi machinery to produce miRNAs and subsequent silencing.

Functions of Viral Encoded miRNAs: There are again two classes: The analogues of host miRNAs that copy events of host miRNA genes into the viral genome, and newly developed miRNAs that are specific to the virus. They prevent apoptosis, alter the cell cycle, serve viral immune evasion, transformation and/or the autoregulation of virus gene expression.



Exogenous dsRNA: Exogenous dsRNA, for example through oral uptake, can be used as an antiviral method if the dsRNA silences the viral mRNA.

Spreading of Silencing

Transitivity (RdRP related): It means the spread of the silencing signal beyond the original target, as siRNAs can have different, secondary siRNAs that in turn fit on other genes that then undergo a knockdown. This is also called "cross-silencing" and requires consideration.

Systemic RNAi: It means the spread of the silencing signal over the whole organism or to other cells and tissues. In plants the silencing signal seems to follow the **phloem flow** and the plant architecture influences the movement of the signal. It's an efficient RNA based immune system against virus infections. Systemic RNAi can also happen by oral uptake of dsRNA.

Transcriptional Gene Silencing and Heterochromatin Formation (TSG)

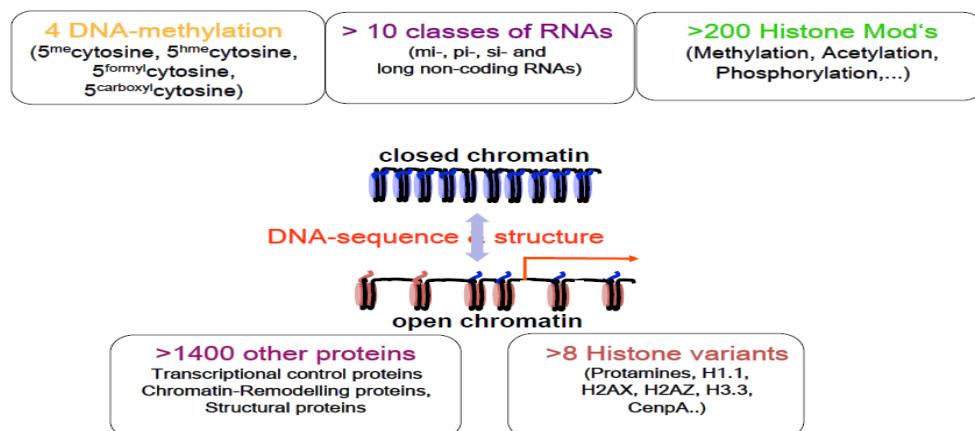
It's a form of transcription control where small RNAs recruit enzymes like HAT, Kinase, PRMT, HKMT that modify the chromatin and histones and lead to silent heterochromatin. The paradox is that transcriptional silent heterochromatin has to be transcribed in order to retain the heterochromatin state.

6 Functional Genomics and Systems Epigenetics

Genomics It's the analysis of genome composition and genome variation and is associated with diseases and phenotypes. Functional genomics is about individual genome variation, the epigenomic profiles and the transcriptome of cells, and the regulatory elements.

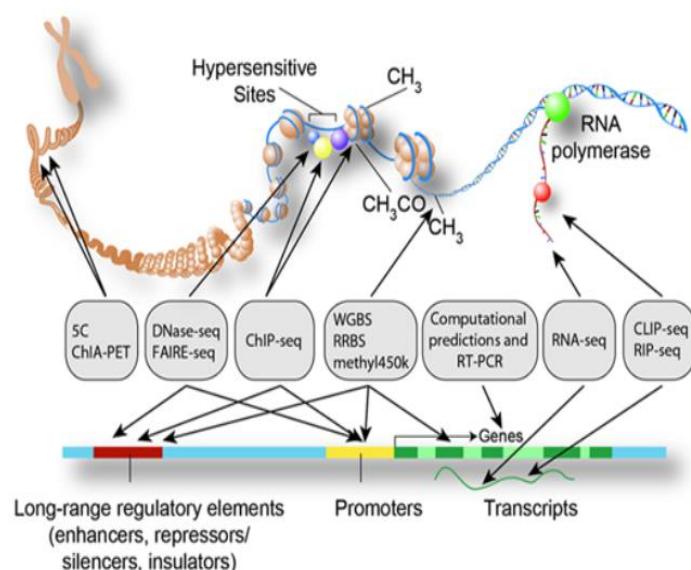
Epigenomics It's the analysis of epigenetic signatures, like chromatin- and DNA-methylation, in isolated cells and in tissues and is correlated with gene expression and disease states.

Epigenomic Features:



In most catalogue projects only a smaller, selected number of features is looked at. For example, only 6-7 histone mods instead of >200.

Encyclopedia of DNA Elements (ENCODE) It aims to identify all functional elements in the human genome.



ChIA-PET: Chromatin Interaction Analysis with Paired-End Tag Sequencing

FAIRE-seq: Formaldehyde-Assisted Isolation of Regulatory Elements

ChIP-seq: Chromatin Immunoprecipitation

WGBS: Whole-Genome Bisulfite Sequencing

RRBS: Reduced Representation Bisulfite Sequencing

Methyl450k: Methylation Bead Array

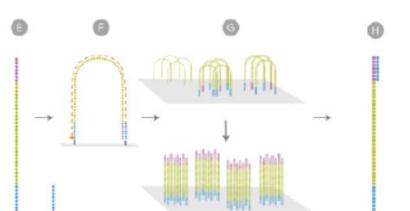
RT-PCR: Real Time Polymerase Chain Reaction

CLIP-seq: Cross-Linking Immunoprecipitation

RIP-seq: RNA Immunoprecipitation

Illumina Solexa Sequencing: HiSeq2500

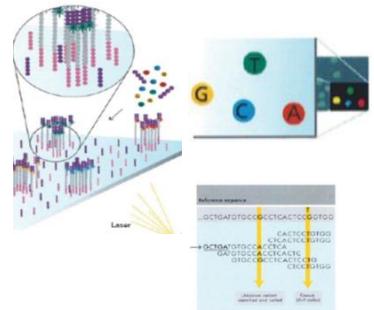
1. Library Preparation for 6h
 - a. fragment DNA
 - b. repair ends/acid A overhang
 - c. ligate adapters
 - d. selected ligated DNA



2. Cluster Generation for 5h
 - a. attack DNA to flow cell
 - b. perform bridge amplification
 - c. generate clusters
 - d. anneal sequencing primer

3. Sequencing

- Determine the first base by adding all four labelled reversible terminators, primers and DNA polymerase enzyme to the flow cell.
- Image the first base by capturing the image of emitted fluorescence after laser excitation from each cluster on the flow cell and record the identity of the first base for each cluster.
- Repeat for the remaining bases one base at a time.
- Align the data, compare it to a reference and identify sequence differences.



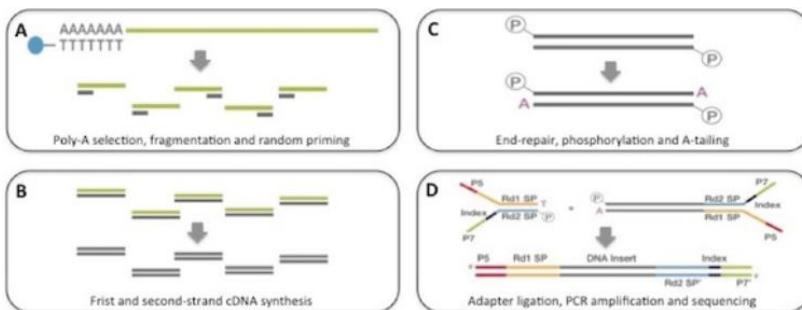
4. Demultiplex Conversion to FASTQ

5. FirstQC and FastQC: Initial quality control by checking for read duplications, sequence bias and overall quality

6. Trimming: remove low quality and adapter sequences

7. FastQ files

RNA-Seq



Library prep begins from 100ng-1ug of Total RNA which is poly-A selected (A) with magnetic beads. Double-stranded cDNA (B) is phosphorylated and A-tailed (C) ready for adapter ligation. The library is PCR amplified (D) ready for clustering and sequencing.

The data is generated from 3 libraries: mRNA, total (ribo-)RNA and small RNAs. It determines expression levels of transcriptions, circular and small RNAs, long non-coding transcripts and annotates alternative transcripts including promoter usage and alternative splicing.

The pipeline consists of mapping, QC and quantification.

Whole-Genome Bisulfite Sequencing (WGBS)

It's a genome wide DNA-methylation analysis that generates a base resolution map of modified cytosine, which normally occurs in CpG contexts, along the entire genome. Each cytosine position in a cell can be 0%, 50% or 100% methylated.

Isolate and fragment genomic DNA, treat it with sodium bisulfite that converts an unmethylated C to a T, generate a library and sequence at an average depth of 25-30x per base. Finally, align the reads to an "in silico bisulfite converted" reference genome and call the ratio of methylation at CpGs and non-CpG methylation.

Bisulfite Data Processing:

- input*: trimmed FASTQC files
- mapping
- realignment
- mark duplicate reads
- clip overlapping ends
- recalibrate quality values
- call methylation levels by separating SNPs from methylated cytosines by observing both strands; a methylation is T - G while a mutation is T - A
- general quality control with regard to mapping efficiency, duplication levels, # called sites, coverage and conversion rate

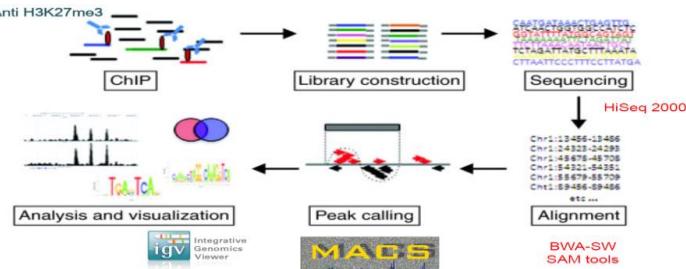
Chromatin Immunoprecipitation (ChIP-Seq)

It aims to find regions enriched by a certain histone mark.

- Isolate cells and crosslink ("fix") the chromatin using formaldehyde.
- Open up the cells and fragment the fixed chromatin by shearing to sizes of mono- or dinucleosomes.
- Use very specific antibodies against histone modifications and let them bind to the chromatin fragments.
- Wash away not bound antibodies and precipitate the bound fraction.
- Isolate the DNA from the bound fraction and prepare a sequencing library.
- Sequence 30-60 million reads per sample.
- Align the reads to a reference genome and call regions of sequence enrichment ("peaks").

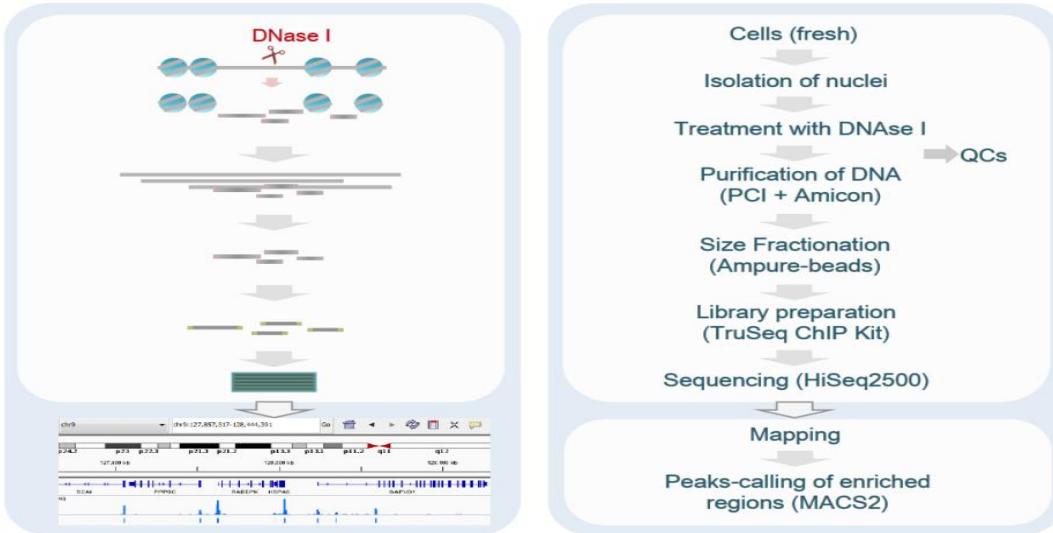
ChIP-Seq Processing:

1. *input*: FASTQ-files
2. Mapping
3. (mark/remove duplicate reads)
4. Peak calling
5. Quality control



DNase-Seq

It maps open chromatin regions.



Data Analysis

Exploratory / Descriptive: Analyse the **number of peaks**, **average methylation** and **distribution**. Use **FASTQC**, **genome browsers** or **R** for **visualization** and make sure that the data behaves as **expected** and that there are **no batch effects**, like failed sequencing runs, biological material and experimental factors.

Differential: Compare two data sets and analyse differential expression, methylation and accessibility. It is particularly useful when having a hypothesis and there are different approaches for different assays. For instance, for RNA-seq use differential expression, while for methylation compare the methylation levels using software like RnBEads or Metilence. For enrichment assays like ChIP-seq, DNase1 and ARAT you can use software like diffBind, diffReps or MANorm.

Integrative: Merge different types of data and analyse overlap and ChromHMM (**ChIP-seq segmentation**).

Downstream Analysis and Integration of Epigenomic Data

