Epigenetics and Human Health

More information about this series at
http://www.springer.com/series/8561

Alexander Meissner • Jörn Walter
Editors

Epigenetic Mechanisms in Cellular Reprogramming

Springer

Editors
Alexander Meissner
Dpt. of Stem Cell and Regenerative Biol
Harvard University Broad Institute
Bauer Laboratory
Cambridge
Massachusetts
USA

Jörn Walter
Universität des Saarlandes
FR84. Biosciences
Dept. of Genetics & Epigenetics
Saarbrücken
Germany

ISSN 2191-2262 ISSN 2191-2270 (electronic)
ISBN 978-3-642-31973-0 ISBN 978-3-642-31974-7 (eBook)
DOI 10.1007/978-3-642-31974-7
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014958446

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

During development the genome of the fertilised egg is utilised to create a whole organism with a rich diversity of cell types. While the underlying sequence remains unchanged, each cell type and developmental stage is reflected in a unique epigenome. This coordinated process of developmental epigenetic programming begins in the germ line (early primordial germs cells, PGCs) and cumulates in the creation of the specialised gametes. Post-fertilisation massive epigenetic reprogramming establishes the totipotent zygote and pluripotent cells of the early embryo (inner cell mass, ICM). The latter can be explanted into cell culture and give rise to pluripotent embryonic stem cells (ESCs) that can be maintained over long periods.

Over the last decade epigenetic reprogramming processes have been widely studied in the zygote, in PGCs and in ESCs. The research focused on various aspects of this topic, most of them being reflected in the selected articles of this volume including (1) understanding reprogramming events at the level of DNA and histone modifications, (2) the physiological parameters and enzymes that control the initiation, the entry and exit from pluripotency, and (3) the differences/similarities of epigenetic reprogramming mechanisms in various pluripotent and totipotent cells.

The detailed knowledge of the underlying reprogramming mechanisms is of great importance for many research areas in human health and disease ranging from stem cell biology to cancer. Examples are a controlled understanding of the cell intrinsic reprogramming mechanisms activated during the in vitro generation of induced pluripotent stem cells (iPSCs) from somatic cells and the erroneous reprogramming mechanisms in somatic (stem) cells leading to massive epigenetic changes in cancer.

This volume compiles a series of articles featuring the current knowledge of molecular events accompanying processes of epigenetic reprogramming. The articles focus on mechanisms operating during early embryonic development, the events that are defining the entry into and exit from pluripotency in ESCs and the implications of such mechanisms for aberrant reprogramming in the course of cancer. The reader will obtain a detailed view of the molecular changes occurring

at various epigenetic levels of histone and DNA modifications. All articles feature references to the important discoveries in the field over the last decade. A glossary at the end will help the reader to navigate through many of the specific terms used in epigenetic research.

Cambridge, MA — Alex Meissner
Saarbrücken, Germany — Jörn Walter

Glossary

Acetylation The introduction, via an enzymatic reaction, of an acetyl group to an organic compound, for instance to *histones* or other proteins.

Agouti gene The agouti gene (A) controls fur colour through the deposition of yellow pigment in developing hairs. Several variants of the gene exist, and for one of these (Agouti Variable Yellow, A^{vy}) the expression levels can be heritably modified by *DNA methylation*.

Alleles Different variants or copies of a gene. For most genes on the chromosomes, there are two copies: one copy inherited from the mother, the other from the father. The DNA sequence of each of these copies may be different because of genetic polymorphisms.

Assisted reproduction technologies (ART) The combination of approaches that are being applied in the fertility clinic, including *IVF* and *ICSI*.

5-Azacytidine A cytidine analogue in which the 5 carbon of the cytosine ring has been replaced with nitrogen. 5-azacytidine is a potent inhibitor of mammalian *DNA methyltransferases*.

Bivalent chromatin A chromatin region that is modified by a combination of histone modifications such that it represses gene transcription, but at the same time retains the potential of acquiring gene expression.

Bisulphite genomic sequencing A procedure in which bisulphite is used to deaminate cytosine to uracil in genomic DNA. Conditions are chosen so that 5-methylcytosine is not changed. PCR amplification and subsequent DNA sequencing reveal the exact position of cytosines which are methylated in genomic DNA.

Blastocyst The blastocyst is a structure formed in the early development of mammals. It is the last stage of preimplantation development in mammals and it is comprised of outer cell layer—trophoblast, which later develops into placenta, and of inner cell mass (see ICM), which gives rise to the embryonic tissues. ICM is

attached to inner side of the hollow basket-shaped structure, formed by trophectoderm (trophoblast cell layer).

Bromo domain Protein motif found in a variety of nuclear proteins including transcription factors and HATs involved in transcriptional activation. Bromo domains bind to histone tails carrying acetylated lysine residues.

Brno nomenclature Regulation of the nomenclature of specific histone modifications formulated at the Brno meeting of the NoE in 2004. Rules are: <Histone><amino acid position><modification type><type of modification>. Example: H3K4me3 = trimethylated lysine-4 on histone H3.

Cell fate The programmed path of differentiation of a cell. Although all cells have the same DNA, their cell fate can be different. For instance, some cells develop into brain, whereas others are the precursors of blood. Cell fate is determined in part by the organisation of *chromatin*—DNA and the histone proteins—in the nucleus.

Cellular Memory (epigenetic) Specific active and repressive organisations of chromatin can be maintained from one cell to its daughter cells. This is called *epigenetic inheritance* and ensures that specific states of gene expression are inherited over many cell generations.

ChIP see *chromatin immunoprecipitation*.

ChIP on chip After chromatin immunoprecipitation, DNA is purified from the immunoprecipitated chromatin fraction and used to hybridise arrays of short DNA fragments representing specific regions of the genome.

ChIP Seq Sequencing of the totality of DNA fragments obtained by ChIP to determine their frequency and position on the genome. Sequencing is usually preceded by PCR amplification of ChIP-derived DNA to increase its amount.

Chromatid In each somatic cell generation, the genomic DNA is replicated in order to make two copies of each individual chromosome. During M phase of the cell cycle, these copies—called chromatids—are microscopically visible one next to the other, before they get distributed to the daughter cells.

Chromatin The nucleo-protein complex constituting the chromosomes in eukaryotic cells. Structural organisation of chromatin is complex and involves different levels of compaction. The lowest level of compaction is represented by an extended array of *nucleosomes*.

Chromatin remodelling Locally, the organisation and compaction of chromatin can be altered by different enzymatic machineries. This is called chromatin remodelling. Several chromatin remodelling proteins move *nucleosomes* along the DNA and require ATP for their action.

Chromo domain (chromatin organisation modifier domain) Protein–protein interaction motif first identified in *Drosophila melanogaster HP1* and *polycomb group proteins*. Also found in other nuclear proteins involved in transcriptional silencing and heterochromatin formation. Chromo domains consist of approx. 50 amino acids and bind to histone tails that are methylated at certain lysine residues.

Chromosomal domain In higher eukaryotes, it is often observed that in a specific cell type, chromatin is organised (e.g. by *histone methylation*) the same way across hundreds to thousands of kilobases of DNA. These 'chromosomal domains' can comprise multiple genes that are similarly expressed. Some chromosomal domains are controlled by *genomic imprinting*.

Chromatin immunoprecipitation (ChIP) Incubation of chromatin fragments comprising one to several nucleosomes, with an antiserum directed against particular (histone) proteins or covalent modifications on proteins. After ChIP, the genomic DNA is purified from the chromatin fragments brought down by the antiserum and analysed.

CpG dinucleotide A cytosine followed by a guanine in the sequence of bases of the DNA. *Cytosine methylation* in mammals occurs at CpG dinucleotides.

CpG island A small stretch of DNA, of several hundred up to several kilobases in size, that is particularly rich in *CpG dinucleotides* and is also relatively enriched in cytosines and guanines. Most CpG islands comprise promoter sequences that drive the expression of genes.

Cytosine methylation In mammals, DNA methylation occurs at cytosines that are part of *CpG dinucleotides*. As a consequence of the palindromic nature of the CpG sequence, methylation is symmetrical, i.e. affects both strands of DNA at a methylated target site. When present at promoters, it is usually associated with transcriptional repression.

Deacetylation The removal of acetyl groups from proteins. Deacetylation of histones is often associated with gene repression and is mediated by histone deacetylases (HDACs).

DNA demethylation Removal of methyl groups from DNA. This can occur 'actively', i.e. by an enzymatically mediated process, or 'passively', when methylation is not maintained after DNA replication.

'de novo' DNA methylation The addition of methyl groups to a stretch of DNA which is not yet methylated (acquisition of 'new' DNA methylation).

DNA methylation A biochemical modification of DNA resulting from addition of a methyl group to either adenine or cytosine bases. In mammals, methylation is essentially confined to cytosines that are in *CpG dinucleotides*. Methyl groups can be removed from DNA by DNA demethylation.

DNA methyltransferase Enzyme which puts new (de novo) *methylation* onto the DNA, or which maintains existing patterns of DNA methylation.

Dosage compensation The X chromosome is present in two copies in the one sex, and in one copy in the other. Dosage compensation ensures that in spite of the copy number difference, X-linked genes are expressed at the same level in males and females. In mammals, dosage compensation occurs by inactivation of one of the X chromosomes in females.

Embryonic stem (ES) cells Cultured cells obtained from the inner cell mass of the blastocyst, and for human ES cells, possibly also from the epiblast. These cells are pluripotent; they can be differentiated into all different somatic cell lineages. ES-like cells can be obtained by dedifferentiation in vitro of somatic cells (see *iPS cells*).

Endocrine disruptor A chemical component which can have an antagonistic effect on the action of a hormone (such as on oestrogen) to which it resembles structurally. Some pesticides act as endocrine disruptors and have been found in animal studies to have adverse effects on development, and for some, to induce altered *DNA methylation* at specific loci. A well-characterised endocrine disruptor is *Bisphenol-A*, a chemical used for the productions of certain plastics.

Enhancer A small, specialised sequence of DNA which, when recognised by specific regulatory proteins, can enhance the activity of the promoter of a gene (s) located in close vicinity.

Epi-alleles Copies of a DNA sequence or a gene which differ in their epigenetic and/or expression states without the occurrence of a genetic mutation.

Epiblast The population of cells in the inner cell mass (see ICM) of a mammalian blastocyst. It is formed when ICM develops into the embryonic disc, consisting of two layers: the adjacent to the trophoblast epiblast and the adjacent the blastocoele (blastocyst cavity) hypoblast.

Epigenesis The development of an organism from fertilisation through a sequence of steps leading to a gradual increase in complexity through differentiation of cells and formation of organs.

Epigenetics The study of heritable changes in gene function that arise without an apparent change in the genomic DNA sequence. Epigenetic mechanisms are involved in the formation and maintenance of cell lineages during development, and, in mammals, in *X-inactivation* and *genomic imprinting*, and are frequently perturbed in diseases.

Epigenetic code Patterns of DNA methylation and histone modifications can modify the way genes on the chromosomes are expressed. This has led to the idea that combinations of epigenetic modifications can constitute a code on top of the genetic code which modulates gene expression.

Epigenetic inheritance The somatic inheritance, or inheritance through the germ line, of epigenetic information (changes that affect gene function, without the occurrence of an alteration in the DNA sequence).

Epigenetic marks Regional modifications of DNA and chromatin proteins, including *DNA methylation* and histone methylation, that can be maintained from one cell generation to the next and which may affect the way genes are expressed.

Epigenetic reprogramming The resetting of *epigenetic marks* on the genome so that these become like those of another cell type, or of another developmental stage. Epigenetic reprogramming occurs for instance in *primordial germ cells*, to bring them back in a 'ground state'. Epigenetic reprogramming and dedifferentiation also occur after *somatic cell nuclear transfer*.

Epigenome The epigenome is the overall epigenetic state of a particular cell. In the developing embryo, each cell type has a different epigenome. Epigenome maps represent the presence of DNA methylation, histone modification and other chromatin modifications along the chromosomes.

Epigenotype The totality of epigenetic marks that are found along the DNA sequence of the genome in a particular cell lineage or at a particular developmental stage.

Epimutation A change in the normal epigenetic marking of a gene or a regulatory DNA sequence (e.g. a change in DNA methylation) which affects gene expression.

Euchromatin A type of chromatin which is lightly staining when observed through the microscope at interphase. Euchromatic *chromosomal domains* are loosely compacted and relatively rich in genes. The opposite type of chromatin organisation is *heterochromatin*.

Genomic imprinting An epigenetic phenomenon which affects a small subset of genes in the genome and results in mono-allelic gene expression in a parent-of-origin dependent way (for a given pair of alleles uniformly either the maternally or paternally derived copy is active).

Germ line specific stem cells Cells derived from undifferentiated germ cells which can be maintained without alterations in their characteristics through many cell divisions.

Heterochromatin A type of chromatin which is darkly staining when observed through the microscope at interphase. Heterochromatic chromosomal domains, found in all cell types, are highly compacted, are rich in repeat sequences, and show little or no gene expression. Extended regions of heterochromatin are found close to centromeres and at telomeres.

Histone acetylation Post-translational modification of the ε-amino group of lysine residues in histones catalysed by a family of enzymes called *histone acetyltransferases (HATs)*. Acetylation contributes to the formation of

decondensed, transcriptionally permissive chromatin structures and facilitates interaction with proteins containing *bromo domains*.

Histone acetyltransferase (HAT) An enzyme that acetylates (specific) lysine amino acids on histone proteins.

Histone code Theory that distinct chromatin states of condensation and function are marked by specific histone modifications or specific combinatorial codes (see also epigenetic code).

Histone deacetylase (HDAC) An enzyme that removes acetyl groups from histone proteins. This increases the positive charge of histones and enhances their attraction to the negatively charged phosphate groups in DNA.

Histone demethylase (HDM) Proteins catalysing the active enzymatic removal of methyl groups from either lysine or arginine residues of histones. Prominent examples are LSD1 and Jumonji proteins.

Histone methylation Post-translational methylation of amino acid residues in histones catalysed by *histone methyltransferases (HMTs)*. Histone methylation is found at arginine as mono- or di-methylation and lysine as mono-, di- or tri-methylation. Modifications are described depending on the position and type of methylation (mono, di, tri-methylation) according to the *Brno nomenclature*. Different types of methylation can be found in either open trancriptionally active or silent (repressive) chromatin (*histone code*). Methylated lysine residues are recognised by proteins containing *chromo domains*.

Histone methyltransferase (HMT) Enzymes catalysing the transfer of methyl groups from S-adenosyl-methionine (SAM) to lysine or arginine residues in histones.

Intracytoplasmic sperm injection (ICSI) Capillary-mediated injection of a single sperm into the cytoplasm of an oocyte followed by activation to promote directed fertilisation.

Imprinted genes Genes that show a parent-of-origin specific gene expression pattern controlled by epigenetic marks that originate from the germ line.

Imprinting see *genomic imprinting*.

Imprinted X-inactivation Preferential inactivation of the paternal X chromosome in rodents (presumably also humans) during early embryogenesis and in the placenta of mammals.

Imprinting control region (ICR) Region that shows germ line derived parent-of-origin dependent epigenetic marking which controls the imprinted expression of neighbouring imprinted genes.

Inner cell mass (ICM) Cells of the inner part of the blastocyst forming the embryo proper. Inner cell mass cells are the source for ES cells.

Induced pluripotent stem cells (iPS) Cells with an ES cell-like pluripotent potential derived from differentiated somatic cells by in vitro reprogramming. Reprogramming is triggered by the activation of pluripotency factor genes and cultivation in ES cell medium. iPS cells are capable to generate all cell types of an embryo.

In vitro Fertilisation (IVF) Fertilisation of a surgically retrieved oocyte in the laboratory, followed by a short period of in vitro cultivation before the embryo is transferred back into the uterus to allow development to term.

Isoschizomers Restriction enzymes from different bacteria which recognise the same target sequence in DNA. Often these enzymes respond differently to methylation of bases within their target sequence, which may make them important tools in DNA methylation analysis. Thus, *Msp*I cuts both CCGG and C5mCGG, whereas *Hpa*II cuts only the unmethylated sequence.

Locus control region (LCR) Region marked by insulator functions and DNase hypersensitive sites. LCRs contain binding sites for insulator proteins and enhancer binding proteins. LCRs control the domain-specific developmentally regulated expression of genes by long-range interactions with gene promoters.

Maintenance methylation Process that reproduces DNA methylation patterns between cell generations. Depends in mammals critically (though not exclusively) on the activity of the 'maintenance DNA methyltransferase' Dnmt1. This enzyme preferentially methylates hemimethylated CpG sites, generated by replication of symmetrically methylated CpG sequences (see Cytosine methylation), while originally unmethylated sites remain unmethylated upon replication.

Maternal effects Long-term effects on the development of the embryo triggered by factors in the cytoplasm of the oocyte.

Metastable epiallele Loci, whose epigenetic state is particularly labile, i.e. prone to be epigenetically modified in a variable and reversible manner. As a consequence of this lability, various phenotypes may derive from genetically identical cells, resulting in phenotypic mosaicism between cells (variegation) and also between individuals (variable expressivity).

Methyl-binding domain (MBD) Protein domain in Methyl-CpG-binding proteins (MBPs) responsible for recognising and binding to methylated cytosine residues in DNA. Proteins containing MBDs form a specific family of proteins with various molecular functions.

Methyl-CpG-binding proteins (MBPs) Proteins containing domains (such as MBD) binding to 5-methyl-cytosine in the context of CpG dinucleotides. MBPs mostly act as mediators for molecular functions such as transcriptional control or DNA repair.

Non-coding RNA (ncRNA) RNA transcripts that does not code for a protein. ncRNA generation frequently involves RNA processing.

Non-Mendelian inheritance Inheritance of genetic traits that do not follow Mendelian rules and/or cannot be explained in simple mathematically modelled traits.

Nuclear periphery Region around the nuclear membrane characterised by contacts of the chromosomes with the nuclear lamina.

Nuclear (chromosomal) territory Cell type-specific areas within the nucleus occupied by specific chromosomes during interphase (G1).

Nucleolus Specific compartments within the nucleus formed by rDNA repeat domains. Nucleoli are marked by specific heterochromatic structures and active gene expression.

Nucleosome Fundamental organisational unit of chromatin consisting of 147 base pairs of DNA wound around a histone octamer.

Oogenesis The process by which primary oocyte develops into mature ovum. In mammals primary oocytes are formed shortly before or shortly after the birth during the process called oocytogenesis.

Parthenogenesis A form of asexual reproduction in which growth and development of embryos occur without fertilisation, with only oocyte genome (in some very rare cases—only sperm genomes) contributing to the embryonic genotype. This form of reproduction occurs naturally in different plant, as well as animal (both invertebrates and vertebrates) species, but not in mammals. The mammalian egg can be artificially induced to undergo parthenogenetic development, but the resulting embryos are not capable of developing to term due to the restrictions imposed by genomic imprinting (see also: Genomic Imprinting).

Pluripotency Capacity of stem cells to form all cell types of an embryo including germ cells but not extraembryonic lineages (see Totipotency).

Polycomb group proteins Epigenetic regulator proteins forming multiprotein complexes (PRCs = polycomb repressive complexes). Polycomb group proteins possess enzymatic properties to control the maintenance of a suppressed state of developmentally regulated genes, mainly through histone methylation and ubiquitination.

Position effect variegation (PEV) A type of clonally heritable variability of gene expression which relies on epigenetic lability (see also metastable epialleles) associated with the particular position of a gene within the genome. PEV has first been observed in the context of gene translocations from euchromatic to heterochromatic environments and is a consequence of variable formation of heterochromatin across the respective locus. PEV may give rise to patches of cells with altered expression profiles. A classical example is represented by certain mutations in *Drosophila* leading to variegated eye pigmentation ('mottled eyes').

Primordial germ cell Mammalian cells set aside during early embryogenesis which migrate through the hind gut of the developing mammalian embryo into the 'Gonadenanlagen' to form founder cells of the latter germ line.

Pronucleus The haploid nucleus, which is formed from sperm or oocyte genomes upon the fertilisation and formation of a zygote (see Zygote). The sperm genome is transformed into paternal pronucleus; the maternal pronucleus originates from the oocyte genome. Both paternal and maternal pronuclei exist within the same ooplasm and parental chromosome remains separated until first metaphase stage.

Protamines Small, arginine-rich proteins that replace histones late in the haploid phase of spermatogenesis (during *spermiogenesis*). They are thought to be essential for sperm head condensation and DNA stabilisation. After fertilisation protamines are removed from paternal chromosomes in the mammalian zygote.

RNA interference (RNAi) Post-transcriptional regulatory effects on mRNAs (control of translation or stability) triggered by processed ds and ss small RNA (si-, mi-, piRNAs) molecules. Effects are propagated by enzymatic complexes such as RISC containing the small RNAs bound by Argonaute proteins.

SAHA Suberoylanilide hydroxamic acid, an inhibitor of certain histone deactylases, leading to enhanced levels of histone acetylation. See also *TSA*.

S-adenosylhomocysteine (SAH) Hydrolysed product formed after the methylation reaction catalysed by DNA and *histone methyltransferases* using SAM as methyl group donor. SAH is a competitive inhibitor of SAM for most methyltransferases.

S-adenosyl methionine (SAM) A cofactor for all DNA (DNMTs) and histone methyltransferases (HMTs) providing the methyl group added to either cytosines (DNA) or histones (arginine or lysine).

SET domain A domain found in virtually all lysine-specific *histone methyltransferases (HMTs)*. A protein–protein interaction domain required for HMT activity and modulation of chromatin structure, frequently associated with cysteine-rich Pre-SET and Post-SET domains.

Silencer Element in the DNA to which proteins bind that inhibit transcription of a nearby promoter. Silencer elements are recognised and bound by silencer proteins.

siRNAs small interfering RNAs, RNAs in the size range of 21–24 nucleotides derived from double-stranded long RNAs cleaved by Dicer. siRNAs are incorporated into the RISC complex to be targeted to complementary RNAs to promote cleavage of these mRNAs.

Somatic cell nuclear transfer (SCNT) Transfer of the nucleus of a somatic cell into an enucleated oocyte using a glass capillary to form an SCNT zygote. After activation of the zygote the genome of the nucleus derived from the somatic cells becomes reprogrammed to start development.

Spermatogenesis The process by which spermatogonia develop into mature spermatozoa. Spermatozoa (sperm) are the mature male gametes. Thus, spermatogenesis is the male version of gametogenesis.

Spermiogenesis The final stage of spermatogenesis which sees the maturation of spermatids into mature, motile spermatozoa (sperm). During this stage, cells no longer divide and undergo a major morphological transformation. In addition, at most of the genome, histone proteins are replaced by the more basic *protamines*.

Stem Cell Non-committed cell which has the capacity to self-renew and divide many times giving rise to daughter cells which maintain the stem cell function. Stem cells have the property to differentiate into specialised cells.

Totipotency Capacity of stem cells to produce all cell types required to form a mammalian embryo, i.e. embryonic and extraembryonic cells (*see Pluripotency*). Totipotent cells are formed during the first cleavages of the embryo.

TSA Trichostatin-A, an inhibitor of certain types of histone deacetylases.

Trithorax group proteins Proteins containing a trithorax like bromo domain: they are usually involved in recognising histone modifications marking transcriptionally active regions and contribute to maintenance of activity.

Trophoblast Cells of the blastoderm forming the placental tissues in mammals.

Uniparental Disomy The occurrence in the cell of two copies of a chromosome, or part of a chromosome, that are identical and of the same parental origin.

X-chromosome inactivation Epigenetically controlled form of *dosage compensation* in female mammals resulting in transcriptional silencing of genes on surplus X chromosomes. X-chromosome inactivation is triggered by the non-coding RNA Xist and manifested by various epigenetic modifications including histone methylation, histone deacetylation and DNA methylation.

Zygote The earliest developmental stage of an embryo. Results from the fusion of maternal (oocyte) and paternal (sperm) haploid gametes. This stage is often called ‘one-cell embryo’ stage.

Contents