Lecture No: 4. DNA Methylation and miRNA

In this lecture we will discuss two of the most important regulatory processes of gene regulation; DNA methylation and microRNA (miRNA).

DNA Methylation

Although, DNA methylation does not relate to chromatin structure and function, it controls gene expression at the transcriptional level.

DNA methylation, not to be confused with histone methylation, is an archetypal epigenetic mark. It is borne by the genetic material but does not influence its sequence. It can regulate genomic activities, and can be maintained through mitosis and meiosis.

DNA methylation is essential in mammals: its loss leads to growth arrest or apoptosis in normal cells as well as in cancer lines. The presence of DNA methylation is absolutely required for embryonic development in mouse. The key role of DNA methylation is to control gene expression, and methylated sequences undergo transcriptional repression.

The DNA of mammals can be methylated on cytosines within the CpG dinucleotides (Fig. 1). The added methyl groups protrude in the major groove of DNA. When the DNA is symmetrically methylated, both methyls face the same direction and are close to one another. The addition of methyl groups changes the biophysical characteristics of the DNA and has two effects: it inhibits the recognition of DNA by some proteins and permits the binding of others. (Vaissie` re, *et al*., 2008).

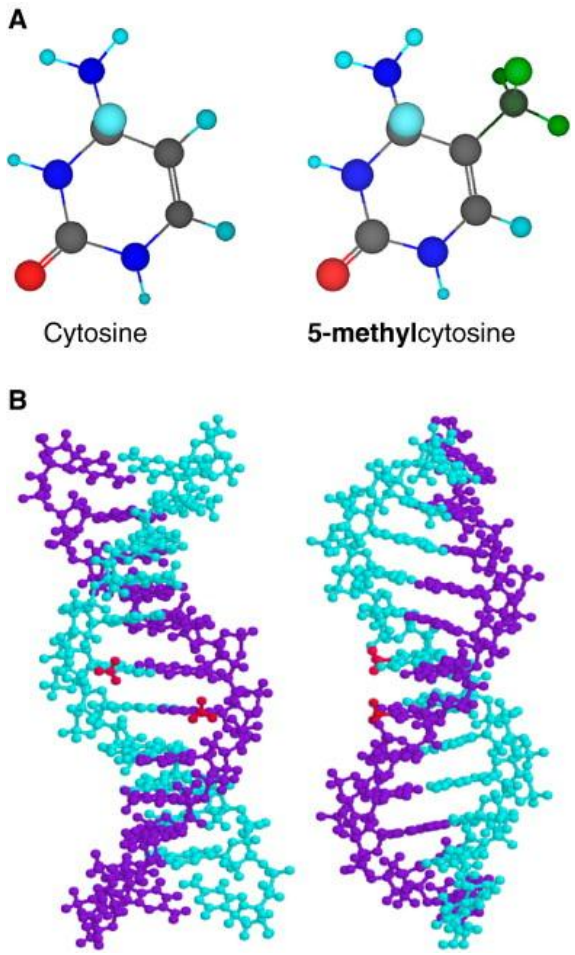

Fig 1: DNA methylation

DNA methyl-transferases (DNMTs)

The modification is brought about by enzymes called DNA methyltransferases (DNMTs). There are three such enzymes in mammals: DNMT1, DNMT3a, and DNMT3b. DNMT3L is structurally related, but is catalytically inactive and serves as a cofactor for DNMT3a and DNMT3b. The protein DNMT2 also has sequence similarity to these enzymes, but its function is quite different; it will not be discussed further here. Extensive enzymology studies have yielded important insight into the function of these enzymes. Notably, it was found that DNMT1 has preferential activity for hemi-methylated DNA over unmethylated DNA. It seems likely that, most of the time, DNMT3a and DNMT3b, aided by DNMT3L, set up the new imprints on previously naked DNA. For this they are called “de novo” methyltransferases. After DNA replication, methylated DNA becomes hemi-methylated, and DNMT1 would be

the main player in making it fully methylated again. It is thus called the "maintenance" enzyme.

Table 1. The proteins involved in setting up and interpreting the methylation mark.

De novo	**Maintenance**	**Cofactor**
DNMT3a	DNMT1	DNMT3L
DNMT3b		
DNA methylation binding proteins		
MBD	Zinc finger	SRA
MeCP2	Kaiso	UHRF1
MBD1	ZBTB4	UHRF2
MBD2	ZBTB38	
MBD4		

Top panel: the enzymes that methylate DNA in mammals. Bottom panel: the three families of proteins that bind methylated DNA in mammals.

The methyl mark is translated into transcriptional repression by the action of proteins that recognize methylated DNA and inhibit gene expression by creating a repressive chromatin structure. Three families of proteins specifically recognize methylated DNA (Table 1). The first family contains MBD1, MBD2, MBD4, and MeCP2; these proteins share a related DNA binding domain called Methyl-binding Domain (MBD). The second family contains the Zinc-finger proteins Kaiso, ZBTB4, and ZBTB38 . These proteins are bifunctional: they bind methylated DNA, but also some non-methylated consensus sequences. Finally, the third family comprises UHRF1 and UHRF2 (also known as ICBP90 and NIRF), which bind methylated DNA through their SET-and-RING-Finger-Associated (SRA) domain.

An important question, discussed at length in an excellent recent review is that of the redundancy between methyl-binding proteins. Their degree of sequence specificity is poorly characterized, and it is unclear whether they can all bind the same target loci, or whether they have distinct targets. Even if the proteins do share some targets, they could be functionally different for other reasons. For instance, they could have different DNA-binding affinities. Also, the different proteins could be expressed at different times or places. Finally they could have different protein or nucleotide interactors that could possibly recruit them to different compartments of the nucleus (Vaissie` re, *et al*., 2008).

Targets of DNA methylation differ in normal and cancer cells

In normal cells, three main types of targets are repressed by DNA methylation. First: parentally imprinted genes, i.e. genes that are expressed differentially from the maternal and the paternal chromosome. They are key regulators of embryonic development and adult life. In most cases the inactive allele is marked by DNA methylation, and monoallelic expression is lost in the absence of methylation . As an aside, recent data indicates that many genes may be expressed monoallelicaly in somatic cells, but it is yet unclear if this depends at all on DNA methylation. Second: the transposons and other repeated sequences that constitute a large fraction of the mammalian genome. Third: a number of genes are methylated in a tissue-specific manner. An interesting subset of those is the Cancer/Testis (C/T) antigens, which are unmethylated and expressed in the testis, and methylated and repressed in all other tissues.

DNA methylation is deregulated in cancer. Tumor cells often have an abnormal pattern of DNA methylation where some tumor suppressor genes are methylated and inactive. Conversely, some normally methylated sequences, such as repeated DNA, imprinted genes, and C/T antigens, can become demethylated. Abnormal DNA methylation is an early causal event during cellular transformation. Demethylating agents can re-establish the expression of silenced tumor suppressor genes and have been approved for clinical use against some leukemias.

MicroRNA (miRNA)

microRNAs (miRNAs) and short interfering RNAs (siRNAs), are key components of an evolutionarily conserved system of RNA-based gene regulation in eukaryotes. They are involved in many molecular interactions, including defence against viruses and regulation of gene expression during development. miRNAs interfere with expression of messenger RNAs encoding factors that control developmental timing, stem cell maintenance, and other developmental and physiological processes in plants and animals. miRNAs are negative regulators that function as specificity determinants, or guides, within complexes that inhibit protein synthesis (animals) or promote degradation (plants) of mRNA targets (Carrington *et al*., 2003).

The biogenesis of miRNAs starts in the nucleus and continues in the cytoplasm. In the nucleus, miRNA genes are transcribed by RNA polymerase II (pol II). The primary miRNA transcripts (pri-miRNAs) contain cap structures as well as poly(A) tails, which are the unique properties of class II gene transcripts. The pri-miRNAs produce hairpins of 60-80 nucleotides, bearing 2 nt 3` overhang (pre-miRNAs) by RNas III (Drosha). Drosha is a member of three members RNasIII enzyme family. Within the context of pri-miRNAs, RNA stem–loops with a large, unstructured terminal loop (X10 nt) are the preferred substrates for Drosha cleavage, and that Drosha then cleaves B22 nt away from the loop/stem junction (Zeng *et al*., 2005).

The pre-miRNAs are then exported to the cytoplasm by an exportin 5-mediated and RanGTPase-driven process (Bohnsack *et al*., 2004; and Yi *et al*., 2003). Exportin 5 is a member of the karyopherin family of nucleocytoplasmic transport factors that depend on a cofactor, termed Ran, for their function. Ran is a GTPase and only binds to karyopherins when in the GTP bound from. Exportin 5 forms a nuclear heterotrimer with Ran-GTP and the pre-miRNA. In addition to inducing pre-miRNA nuclear export, Exp5 binding also stabilizes nuclear pre-miRNAs. In any event, once the Exp5/Ran-GTP/pre-miRNA complex has passed through the nuclear pore and reached the cytoplasm, the Ran-GTP is hydrolyzed to Ran-GDP, resulting to the release of pre-miRNA (Fig 2) (Cullen. 2004).

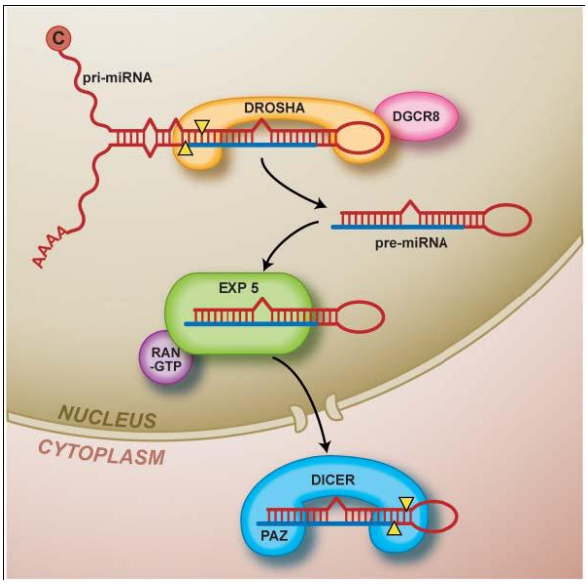

Fig 2. Key Steps in miRNA Biogenesis .The mature miRNA is indicated in blue whereas yellow triangles indicate processing sites (Cullen. 2004).

In the cytoplasm, the pre-miRNAs are processed by a cytoplasmic RNase III, Dicer, to yield double-stranded siRNA of 21 nt, bearing 2 nt 3` overhang. The PAZ domain of the Dicer binds to the 2nt 3` overhang present in the base of the pre-miRNA hairpin while the dsRNA domain binds to the stem and defines the distance of the cleavages from the base (Zhang, 2004). The processing of pre-miRNA by Dicer is ATP dependent process followed by the phosphorylation of the 5`- ends of the siRNA duplex, resulting in the formation of an inactive ~360 kDa RISC/siRNA complex, then ATP dependent unwinding of the siRNA duplex from the 5` end of the antisense strand and RISC is activated (McManus *et al*., 2002). Following RISC activation, the antisense strand of the unwind siRNA guides the siRNA–RISC* complex to the target mRNA. The guide antisense strand base pairs with the target mRNA, forming an A-form helix, and the RISC* protein complex recognizes the major groove of the A-form helix. The recognition and cleavage of the RNA target is ATP independent process (Nyka¨ nen *et al*., 2001). At the final step, the target mRNA is cleaved by RISC*. RISC* is then recycled to catalyze another cleavage event.

Investigation of protein function using RNAi

There are many ways to induce RNAi action in cells. Elbashir *et al*. have shown, for the frst time, siRNA-mediated gene silencing in mammalian cells. They found that the introduction of exogenous 21-nucleotide siRNAs duplexes into mammalian cells specifically suppress expression of endogenous and heterologous genes in different mammalian cell lines (Elbashir *et al*., 2001). The advantage of oligonucleotide-based siRNA is that in some lines of cells, the transfection efficiency is 90% and is higher than that obtained with plasmids (Myagish *et al*., 2003). On the other hand, the synthetic 29-mer shRNAs more potent inducers of RNAi than small interfering RNAs. The transfection of optimized Dicer substrates results in an improved RNAi response. Dicer cleavage of chemically synthesized short hairpin RNAs (shRNAs) with 29-base-pair stem and 2-nucleotide 3` overhangs produced predictable homogeneous small RNAs comprising the 22 bases at the 3` end of the stem. Maximal inhibition of target genes was achieved at lower concentrations and silencing at 24 h was often greater (Siolas *et al*., 2005). The efficiency of transfection of cells with siRNA depends on the type of cell, and RNAi seems to be sustained for only a limited period of time (Myagishi *et al*., 2003).

In 2002, Zeng and his co-researchers have shown that designed miRNA, by substituting the stem sequences of the mir-30 precursor with unrelated base paired sequences, could inhibit the expression of mRNA containing a complementary target site. These results indicate that the sequence of the precursor does not contribute to miRNA production and thus might be particularly suitable as “vectors” for novel miRNA production. These miRNAs can be designed to specifically inactivate the expression of selected target genes in human cells. The designed miRNAs can be produced from transfected DNA in human cells and these miRNAs can induce the specific degradation of a cognate mRNA target, similar to transfected siRNAs. This approach offers the transfection of miRNA expression plasmids which is simple and inexpensive and can result in a continous miRNA production, thus presumably leading to stable inhibition of target mRNA expression (Zeng *et al*., 2002).

The advantage of plasmid-based siRNA is that it is possible to eliminate those cells that have not been transfected with plasmids by selection for antibiotic-resistance

genes. Moreover, RNAi continues for much longer periods when plasmid-based siRNAs are used. Viral vectors allow delivery of siRNA expression cassettes into cells at high efficiencies of transfection, and in the case of lentivirus and retrovirus, it is easy to generate stable knockdown cells via integration of the viral vector into the genome (Myagishi *et al*., 2003).

Designing RNAi experiments

To induce high efficiency RNAi inside mammalian cells, well designed siRNA should be used. So a systematic analysis of siRNA was performed to determine the characteristics associated with siRNA functionality, therby induce effective silencing of the desired gene.

The analyses revealed that most highly functional siRNA have a G/C content that ranged between 38% to 52%, and to target regions of extended low to moderate GC content (Reynolds *et al*., 2004; Kirchner *et al*., 1998; and Amarzguioui *et al*., 2004). On the other hand G/C at the 5` end of the sense strand, and the absence of any GC stretches of more than 9 nt in length induce highly effective gene silencing in mammalian cells (Tei *et al*., 2004). Moreover, the low internal stability of siRNA at the 5` terminus of the antisense strand and at the 3` terminus of the sense strand is prerequisite for effective silencing (Cullen. 2004; Reynolds *et al*., 2004; Kirchner *et al*., 1998). This asymmetry probably important for duplex unwinding and efficient antisense entry into RISC (Amarzguioui *et al*., 2004 and T ei *et al*., 2004). RNAi was also induced with chemical modifications that stabilized interactions between A–U base pairs, demonstrating that these types of modifications may enhance mRNA targeting efficiency in allele-specific RNAi. Modifications altering the structure of the A-form major groove of antisense siRNA–mRNA duplexes abolished RNAi, suggesting that the major groove of these duplexes was required for recognition by activated RISC.

In addition, siRNA sequences that contain internal repeats or palindromes may form internal fold-back structures. These hairpin-like structures may exit in equilibrium with the duplex form, reducing the effective concentration and silencing potential of the siRNA (Reynolds *et al*., 2004; Kirchner *et al*., 1998).

Further analysis of the relation-ship between the sequence and structure, in particular, a tight structure, of the target RNA and the activities of siRNAs has been studied. The efficacy of siRNA is reduced when the target site is embedded within a tight RNA structure. Furthermore, the siRNA activity appears mostly to depend on the target sequence itself, with surrounding sequences having no major effects (Amarzguioui *et al*., 2004). Also, the results of other study suggest that it may not be critical to consider the target site's secondary structure, as the best algorithms only consider the sequence alone. (Sætrom *et al*., 2004).

The application of an algorithm incorporating all these criteria significantly improves potent siRNA selection. This highlights the utility of rational design for silencing potent siRNAs and facilitating functional gene knockdown studies (Reynolds *et al*., 2004). Many algorithms have been published recently, and they base their predictions on such different features as duplex stability, sequence characteristics, mRNA secondary structure, and target site uniqueness (Sætrom *et al*., 2004).

Delivery of RNAi into cells

To achieve efficient stable transformation of mammalian cells by DNA transfection, two factors must be considered: efficient delivery of DNA into the cell nuclei to promote its integration into the host chromosome and sufficient amounts of expression of the transduced gene to allow the cell to survive and grow in the course of selection (Chen and Okayama, 1987). Taking all together with that using a carrier, is better than treating cells with naked phosphorothioate oligonucleotides. Efforts to develop methods for functionally delivering polynucleotides into living cells have continued steadily over the past several years. Effective methods include the use of calium phosphate, liposomes, retroviral vectors, reconstituted viral envelopes and electroporation. In addition, there are several procedures which employ polycations such as poly-lysin, DEAE- dextran and polyrnithinc. (Felgner and Ringold, 1989). Many carriers are commercially available, including lipofectin, lipofectase, cytofectin (serum stable), Star- Burst dendrimers of many generations, cationic porphyrins, and others (Stein. 1999). Chen and Okayama, have developed a simple procedure

involving the use of calcium phosphate-mediated DNA transfection and new marker vectors, which achieve extremely efficient transformation of mammalian cells.
the extremely high transformation efficiency is contributed to The low pH of the 2 x BBS buffer. They suggested that the reason for the enhancement; perhaps the structure of the calcium phosphate-DNA complex promotes uptake by a larger number of cells, promotes more efficient uptake by cells, or promotes preservation of the DNA while it is en route to the nucleus. In addition to the low pH, the amount of DNA and the level of CO2 (and possibly a component of the medium) are important for the formation of the appropriate calcium phosphate-DNA complex. A transition from course to fine precipitates occurred at the optimum DNA concentration (Chen and Okayama, 1987).

Cationic lipid reagents have proven to be one of the most efficient methods for the transfection of nucleic acid molecules into cultured cells. A synthetic cationic Lipid, N-[1-(2, 3-dioleyloxy) propylj-N,N,N-trimethylammonium Chloride (DOTMA) has been created that forms unilamellar liposomes which complex with DNA and RNA for the transfection of mammalian cells. DOTMA is positively charged lipid forms liposomes in aqueous environment (Felgner and Ringold, 1989). Moreover, Malone *et al*., used DOTMA incorporated into a liposome (lipofectin) for RNA transfection in tissue culture cells. The RNA/lipofectin complex can be used to introduce RNA into a wide variety of cells. However, one limitation of the lipofectin procedure is the toxicity associated with the positively charged lipids. For this reason it is prudent to establish the optimal RNA-to-lipofectin ratio for the desired cell type (Malone *et al*., 1989). Currently, the cationic lipid reagent with the highest transfection efficiency in the widest variety of cells is LIPOFECTAMINE PLUS Reagent. LIPOFECTAMINE 2000 Reagent was compared to LIPOFECTAMINE PLUS Reagent in a variety of cells to assess its activity. For many cells, LIPOFECTAMINE 2000 Reagent resulted in the highest transfection activity. In most cells, transfections with LIPOFECTAMINE 2000 Reagent in the presence or absence of serum in the medium had similar activities (Ciccarone *et al*., 2000). Thus, a dramatic increase in transfection efficiency can be obtained by simply repeating transfection with the use of a common polycationic lipid. The effects of the carriers themselves on cells are not generally known. It should be kept in mind that

`antisense' may be caused by the summation of effects of the carrier plus the oligomer, as the carriers dissociate from the oligomer intracellularly (Stein. 1999).

Limitations of RNAi

Although the effects of dsRNA-mediated interference are potent and specific there are several limitations that should be taken into account when designing RNA-interference-based experiments. First, a sequence shared between several closely related genes may interfere with several members of the gene family. Second, it is likely that a low level of expression will resist RNA-mediated interference for some or all genes, and that a small number of cells will likewise escape these effects. Third, the introduction of siRNA into cells induces interferon response. Fourth, the presence of the hairpin secondary structure in the siRNA expression vectors interferes the standard sequencing reactions.

The specificity of gene silencing by siRNA in cultured human cells has been characterized using gene profiling. Transcript profiles revealed siRNA-specific rather than target specific signature, including direct silencing of nontargeted genes containing as few as eleven contigous, nucleotides of identity to the siRNA. These results demonstrate that siRNA may cross-react with targets of limited sequence similarity. On the other hand Jen *et al*., have showed that siRNA-induced gene silencing of transient or stably expressed mRNA is highly gene-specific and does not produce secondary effects detectable by genomewide expression profiling. Moreover, the specificity of siRNA has been investigated by applying gene expression profiling. Several siRNAs were designed against different regions of the same target gene for three different targets. Their effects on cells were compared by using DNA microarrays to generate gene expression signatures. When the siRNA design and transfection conditions were optimized, the signatures for different siRNAs against the same target were shown to correlate very closely, whereas the signatures for different genes revealed no correlation. These results indicate that siRNA is a highly specific tool for targeted gene knockdown, establishing siRNA-mediated gene silencing as a reliable approach for large-scale screening of gene function and drug target validation (Fire et., ; Jackson, ; Sledz *et al*., ; Chi, *et al*., ; and Semizarov *et al*.,).

References

Carrington *et al*., (2003) Role of mcroRNAs in Plant and Animal Development. Science 301, 336-338.

Zeng *et al*., (2005) Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. The EMBO Journal 24, 138–148.

Bohnsack *et al*., (2004). Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. RNA 10: 185-191.

Yi *et al*., (2003) Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes & Dev.* 17: 3011-3016

Cullen (2004) Transcription and Processing of Human microRNA Precursors. Molecular Cell, 16, 861–865

McManus et al., (2002) GENE SILENCING IN MAMMALS BY SMALL INTERFERING RNAS . *NATURE REVIEWS | GENETICS* 3 , 737- 747.

Nyka¨ nen *et al*., (2001) ATP Requirements and Small Interfering RNA Structure in the RNA Interference Pathway. *Cell*. 107, 309–321.

Elbashir *et al*., (2001) Duplexes of 21±nucleotide RNAs mediate RNA interference in cultured mammalian cells. *NATURE* . 411, 494-498.

Myagishi and Tairak (2003) Strategies for Generation of an siRNA Expression Library Directed Against the Human Genome. *Oligonucleotides* 13, 325–333.

Siolas *et al*., (2005) Synthetic shRNA as potent RNAi triggers. Nature Biotechnology 23, 227- 231.

Reynolds *et al*., (2004) Rational siRNA design for RNA interference. *Nature Biotechnology. 22, 326- 330.*

Amarzguioui and Prydz (2004) An algorithm for selection of functional siRNA sequences. *Biochemical and Biophysical Research Communications. 316, 1050–1058.*

Kirchner *et al*., (1998) Secondary structure dimorphism and interconversion between hairpin and duplex form of oligoribonucleotides. *Antisense Nucleic Acid Drug Dev. 8(6). 507-16.*

Tei *et al*., (2004) Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. *Nucleic Acids Research. 32 (3), 936-948.*

Sætrom and Snøve (2004) A comparison of siRNA efficacy predictors. *Biochemical and Biophysical Research Communications. 321, 247–253.*

Chen and Okayama (1987) High-Efficiency Transformation of Mammalian Cells by Plasmid DNA. *Molecular and Cellular Biology. 2745-2752.*

Felgner and Ringold (1989) Cationic liposome-mediated transfection. *Nature 337. 387-388.*

Stein (1999) Two problems in antisense biotechnology: in vitro delivery and the design of antisense experiments. *Biochimica et Biophysica Acta. 1489, 45-52.*

Malone *et al*., (1989) Cationic liposome-mediated RNA transfection. *Pro. Natl. Acad. Sci. USA. 86, 6077-6081*.

Ciccarone *et al*., (2000) *LIPOFECTAMINE™ 2000 Reagent For Rapid, Efficient Transfection of Eukaryotic Cells. Focus. 2 1(2),*

Fire, et., ( ) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans.*

Jackson (….) Expression profiling reveals off target gene regulation by RNAi.

Sledz (……) Activation of the interferon system by short-interfering RNAs.

Chi *et al*., (…..) Genomewide view of gene silencing by small

Semizarov *et al*., (……..) Specificity of short interfering RNA determined through gene expression signatures