EPIGENETICS
targeting the mediator between environment and phenotype

Although it is still early days in terms of our understanding of epigenetics, the fast development of new tools and technologies to define genome-wide epigenetic variations in humans has the potential to enable effective new epigenetic therapies and diagnostic tests for a wide range of diseases beyond and including cancer.

By Dr John M. Rosenfeld, Dr Kan Saito and Dr Michael Sturges

It is well known that the bad habits of an individual such as overeating or cigarette smoking are likely to result in a shorter life span for that individual. Recent studies in epigenetics suggest that at least some of these bad behaviours may also impact the health of that individual's progeny to disease. The Överkalix study (an isolated community in northern Sweden) found that a single winter of overeating as a child can biologically predispose one's children and grandchildren to diabetes and heart disease, resulting in lifespans that are decades shorter than their peers[1]. The Avon Longitudinal Study of Parents and Children (ALSPAC) found that sons and grandsons of men who smoke during pre-puberty are at greater risk for obesity and other health problems[2]. How is it possible that a decision made by a child can have such catastrophic consequences for multiple generations of offspring so quickly? Conventional wisdom traditionally held that all heredity was specified by the genetic information encoded in an individual's DNA sequence and acquired changes in cell behaviour were due to changes in that sequence over time. However, observations such as those found in the Överkalix and ALPAC studies occurred far too quickly to be explained by natural selection; genetic mutations are clearly not the entire story.

Epigenetics is a second layer of hereditary information that causes changes in gene expression patterns. These changes are not a result of mutations to the DNA sequence itself, but instead result from environmental factors that alter expression through chemical modification of the DNA and chromatin associated proteins. These types of changes (particularly methylation of DNA) may also explain differences in phenotypes of identical twins[3]. Epigenetics is ubiquitous in nature, occurring in

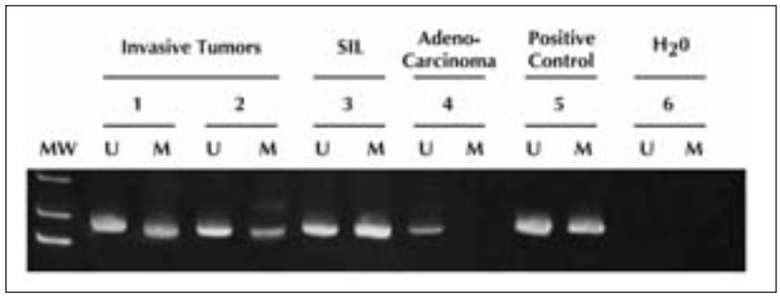

Figure 1
Methylation-specific PCR of the p16 gene in two invasive carcinomas, a squamous intraepithelial lesion (SIL) and an adenocarcinoma of the cervix. Results indicate that both invasive carcinomas and the SIL example contain methylated CpGs within a region of the p16 CpG Island while the adenocarcinoma sample is free of methylation at the p16 locus

bacteria, protists, fungi, plants and animals, and rapid advances being made in the field is contributing to our understanding of transcriptional regulation, nuclear organisation, embryonic development, ageing and disease.

Biologists have known for decades that epigenetics, particularly DNA methylation and chromatin remodelling, is critical for the development of multi-cellular eukaryotic organisms, providing an explanation for how distinct cell types in the body develop *in utero* to fulfill a specialised function despite each cell containing identical DNA. Each individual has only one genome but multiple epigenomes, which differ by cell and tissue type. Although the first epigenetic abnormality associated with cancer was discovered in 1983[4], only recently has there been significant effort to identify the role of epigenetics in the development of disease. This is because scientists now have advanced tools and technologies available for studying the epigenome in a relatively high-throughput fashion. They also realise that although it may take many generations for a genome to evolve, it takes only the addition of a methyl group to change an epigenome – and epigenomes change over the lifetime of an individual. As a result, today epigenetic studies are complementing genetic approaches to the molecular basis of disease and providing insight into the age-old question: Can I control my fate?

Bypassing natural selection

Epigenetics has been described as the link between nature and nurture because epigenetic changes occur in response to environmental signals, including hormones, nutrients, stress and cellular damage (eg, smoking and sun exposure). There are three classes of epigenetic information that effect transcriptional and post-transcriptional regulation of genes: DNA methylation, chromatin associated protein post translational modification and certain non-coding RNAs that are associated with chromatin. Although it has been confirmed that patterns of gene expression can be retained and passed on to descendents, evidence for how epigenetic memory occurs has been lacking. Understanding the mechanisms that promote cellular diversity using DNA sequences that are identical from cell to cell and how each contributes to epigenetic memory will impact diverse areas of research, from agriculture to human health and disease.

DNA methylation

DNA methylation, the covalent addition of a methyl group (CH_3) to the nucleotide cytosine, is involved in the regulation of many cellular processes, including chromosome stability, chromatin structure, X chromosome inactivation, embryonic development and transcription. When a methyl group is added via DNA methyltransferases to the C-5 carbon of the DNA base cytosine, it projects into the major groove of the double helix and blocks transcription factors from binding to the promoter region of a gene. The addition of the methyl group can also serve as a docking station for proteins that bind to the methylated sequence and interact with additional protein modifiers of chromatin structure. DNA methylation is maintained during cell division (and thus passed on to progeny) at dinucleotide C-G (CpG) via the enzyme DNA methyltransferase I. During DNA replication, DNA methyltransferase I seeks out hemi-methylated DNA and places a new methyl group on the daughter CpG. Methionine, an essential amino acid, is the source of methyl groups in this reaction and is converted to a biologically active donor state through a pathway that involves folic acid.

Approximately 1% of the genome consists of 500-2,000 base pair CpG-rich areas, also known as CpG islands, and roughly 60% of all gene promoters include CpG islands. Most of the CpG islands with these promoters are unmethylated and thus actively transcribed[5]. One example of the importance of DNA methylation in altering the phenotype of an organism was demonstrated in 2003 with publication of the agouti mouse (A^{vy})[6]. If expressed continuously, the agouti gene gives mice yellow coats, they become obese in adulthood and are predisposed towards diabetes and cancer. Two groups of identical pregnant agouti mice were fed two different diets: one group received a diet rich in B vitamins (folic acid and vitamin B12) and the other group did not. The B vitamins acted as methyl donors, which caused methyl groups to attach much more frequently to the agouti gene *in utero*, thereby preventing its expression (in this case, a protective effect). As a result, the mothers who received B vitamins throughout pregnancy

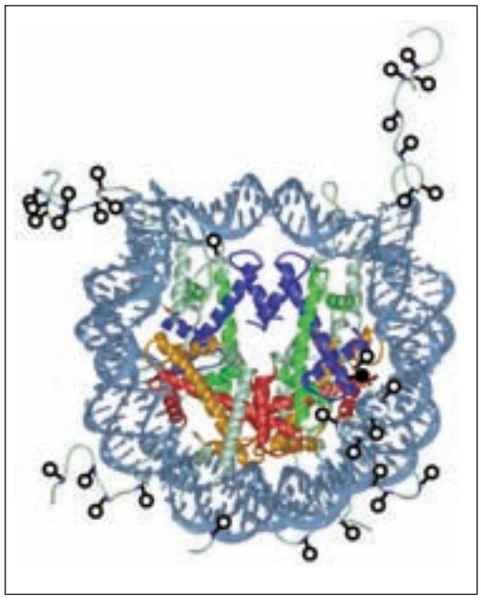

Figure 2
Crystal structure of the nucleosome, the basic subunit of chromatin. DNA appears in light blue and histones are depicted as coloured helices. Post-translational modifications of the histones are represented by black 'lollipops'. Figure courtesy of Dr Karolin Luger, Colorado State University

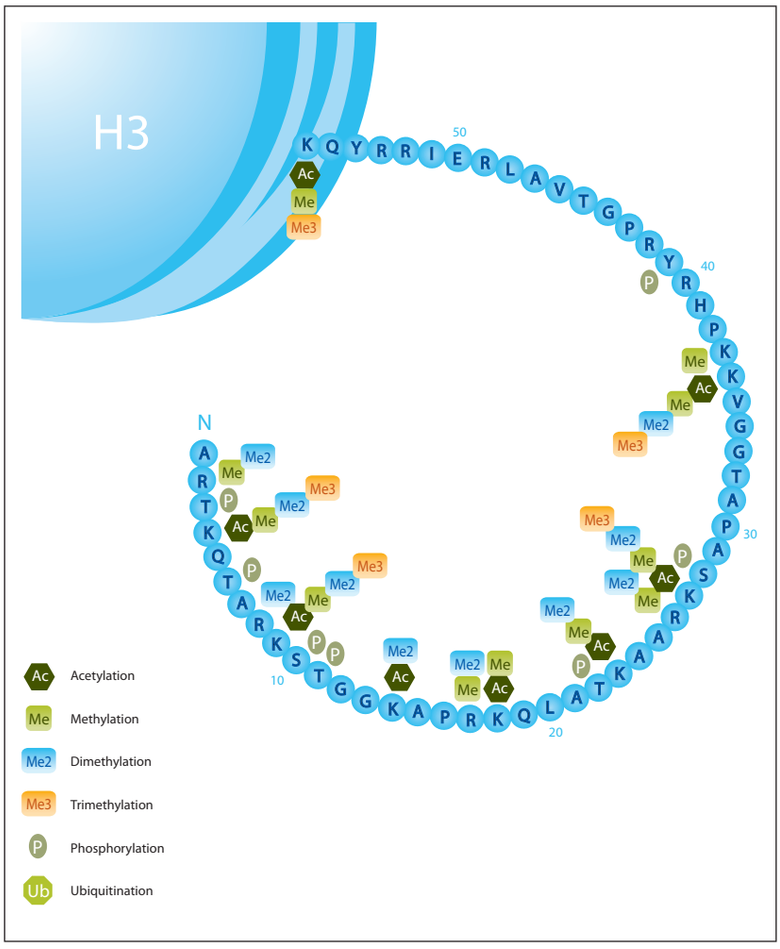

Figure 3: Histone H3 modifications

gave birth to healthy mice, in contrast to the mothers who did not receive this special diet.

In a follow up study, the same pregnant agouti mice (A^{vy}) were exposed to bisphenol A (BPA), a building block of polycarbonate plastics and epoxy resins used to make consumer goods ranging from water bottles to dental sealants[7]. BPA was found to significantly reduce DNA methylation in A^{vy} mice by 31%, resulting in the birth of more obese, yellow offspring who display a higher incidence of diabetes and cancer as adults. Although it remains to be demonstrated conclusively in humans, the use of the A^{vy} agouti mouse as an epistable biosensor has proven to be a valuable animal model in demonstrating the influence of environmental exposure to modulation of gene expression and this particular study highlights a possible connection between the increase in plastics in our environment and the rising incidence of obesity.

Today, aberrant DNA hypermethylation is known to play a role in a number of diseases, including many types of cancer, which can result when tumour suppressor genes are subject to inappropriate epigenetic inactivation[8]. Hypermethylation is found in every type of human cancer and is the most well characterised epigenetic change to occur in tumours[9].

There are two approaches to DNA methylation analysis. Typing technologies target a small number of loci across many samples, and involve the use of techniques such as methylation-specific PCR (**Figure 1**), methylation-sensitive restriction enzymes and mass spectrometry. Genome-wide methylome profiling technologies, on the other hand, include bisulfite sequencing (unmethylated cytosine residues are converted to uracil), methylated DNA immunoprecipitation sequencing (MeDIP-Seq), methylated DNA capture by affinity purification sequencing (MethylCap-Seq) and methylation-sensitive restriction enzyme sequencing (MRE-Seq). Advancements made in these mapping technologies has made possible the large number of recently published genome-wide methylome profiles in a wide variety of species and in the presence of various environmental signals[10]. Current efforts are focused on sequencing the location of every methylated cytosine in an organism's DNA, the 'methylome'[11]. In addition, various DNA methyltransferase assays are available, allowing researchers to study the activity level of these enzymes in various species

Chromatin modifications

Chromatin, a complex of DNA and associated proteins in the nucleus, is a higher ordered structure

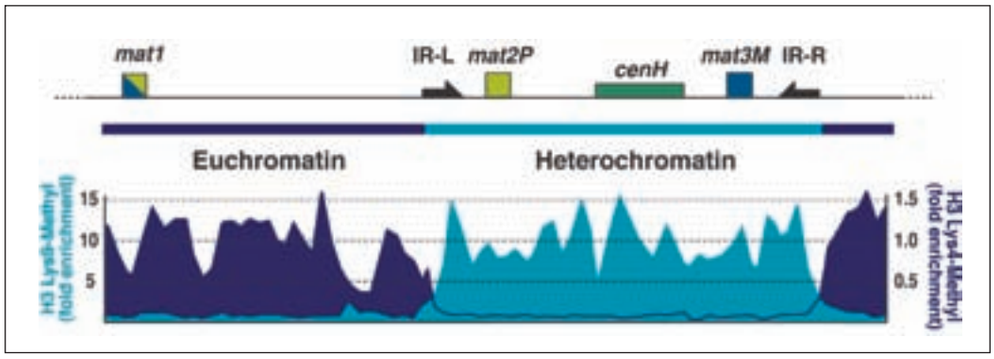

Figure 4
Regions of transcriptionally active euchromatin (blue bar) and inactive heterochromatin (red bar) are indicated below the schematic of the locus. Shown graphically at the bottom of the figure, histone H3 lysine 4 methylation (blue) and H3 lysine 9 methylation (red) are associated with discrete regions of this locus, as determined by chromatin immunoprecipitation. Lysine 4 methylation is limited to the active regions of euchromatin, whereas high levels of lysine 9 methylation are detected only in transcriptionally silent heterochromatin. Figure courtesy of Shiv Grewal, NCI

that allows cells to package their DNA, provides scaffolding for cell division and enables control of gene expression. The fundamental unit of chromatin is the nucleosome, in which DNA wraps around an octamer of four core histone proteins, two molecules each of histone H2A, H2B, H3 and H4 (**Figure 2**). Histone H1 associates with chromatin outside the nucleosome and regulates chromatin structure. Histones are subject to a variety of post-translational modifications, including acetylation of lysine, methylation of arginine and lysine, phosphorylation of serine, ubiquitinylation of lysine, and citrullination of arginine (**Figure 3**). Histone modifications, which occur primarily within the amino-terminal tails, are highly correlated with active DNA processes, including transcription, repair, and replication and it is becoming increasingly clear that these modifications represent regulatory events that govern the accessibility and function of the genome. In *S. pombe*, for example, lysine 4 methylation is limited to the transcriptionally active regions of euchromatin (lightly packed form of chromatin), whereas high levels of lysine 9 methylation are detected only in transcriptionally silent heterochromatin (**Figure 4**). The mechanism of maintaining chromatin modifications during cell division is not well understood.

There are two proposed mechanisms by which histone modifications are believed to mediate transcriptional activity[12]. The first mechanism proposes that these modifications alter the electrostatic charge of the histone, affecting the affinity between the histone and DNA. For example, histone acetylation relaxes the chromatin, allowing easier access for RNA polymerase and transcription factors. The second mechanism proposes that these modifications create protein binding sites, such as bromodomains and chromodomains, which recruit other proteins that recognise acetylated and methylated lysines, respectively, and these recruited proteins then promote transcription.

In addition to post-translational modifications, each histone protein (except H4) has a range of variants that differ in their amino acid sequence mainly in the N-terminal region. These variants are typically expressed at very low levels and are believed to provide novel structural and functional properties of the nucleosome. Although the exact role of these variants remains unclear, it is believed that their presence contributes to epigenetic memory.

Studies in models of Huntington Disease (HD) have shown that mutant Huntington affects histone acetyltransferase activity, suggesting that aberrant epigenetic activity may play a role in the development of this disease[13]. Furthermore, studies are showing a potential therapeutic role for histone deacetylase inhibitors in HD[14]. Further advances are being made in our understanding of the role of chromatin modification in disease through the use of various tools and technologies that provide a powerful means to explore these markers, including genome-wide profiling of DNA-protein interactions by chromatin immunoprecipitation (ChIP) coupled with high throughput sequencing (ChIP-Seq) (**Figure 5**) and histone modification-specific antibodies. ChIP is proving to be a powerful technique for studying protein-DNA complexes and analysing histone modifications. Antibodies specific to the modification under study enrich for regions of chromatin that contain the modification, which is then detected using quantitative PCR or microarray technology.

Non-coding RNA

Eukaryotic genomes transcribe up to 90% of their genomic DNA, however, only 1-2% of these transcripts encode for proteins. The vast majority is therefore transcribed as non-coding RNAs (ncRNAs) which display tissue-specific expression patterns and sub-cellular locations. On an evolutionary scale, the amount of non-coding RNA markedly increases along with the complexity of

the organism, a pattern that is not observed with protein coding DNA regions. For example, the percentage of the genome coding for proteins is ~90% in prokaryotes, ~68% in yeast, ~25% in nematodes, ~17% in insects, ~9% in pufferfish, ~2% in chickens and 1% in mammals[15]. These data strongly implicate ncRNA in the evolution of complex organisms. In addition, as genome-wide association studies continue to identify regions of the genome associated with complex diseases, it is becoming increasingly apparent that many variations that correlate with disease state occur in non-protein coding regions of the genome, some of which likely express regulatory RNAs[16].

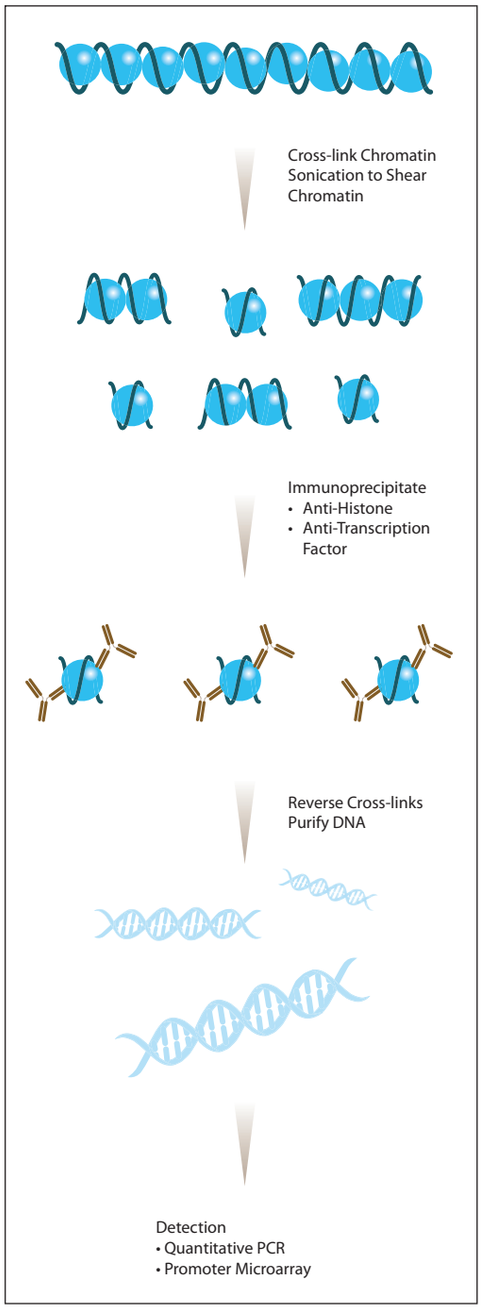

Figure 5
Genome-wide profiling of DNA-protein interactions by chromatin immunoprecipitation (ChIP)

Although the functional mechanisms remain poorly understood, there is growing evidence that many ncRNA transcripts are involved in epigenetic regulation, playing important roles in DNA methylation, chromatin modification, X-chromosome inactivation, genomic imprinting and paramutation. The fact that most ncRNAs, including small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs), small interfering RNAs (siRNAs), microRNAs (miRNAs), and piwi-interacting RNAs (piRNAs), are expressed at significantly lower levels than mRNAs further suggests that these transcripts fulfill regulatory functions. The deregulation of several ncRNAs has been linked to many types of solid cancers, as well as lymphomas and leukaemias.

RNA-binding protein immunoprecipitation (RIP), an RNA analogue of the ChIP application described above, can be used to identify the association of any type of RNA molecule with specific nuclear or cytoplasmic binding proteins (**Figure 6**). These experiments involve immunoprecipitation of endogenously formed RNA-protein complexes, identification of mRNAs (and potentially non-coding RNAs associated with them) and direct measurement using quantitative RT-PCR, microarray analysis (RIP-chip) and second-generation sequencing based platforms (RIP-Seq).

Stem cells and ageing

Epigenetic regulation is a critical mechanism for determining the fate of embryonic and adult stem cells. The pluripotent nature of embryonic stem cells is regulated by epigenetic mechanisms, relying on Polycomb group proteins to reversibly silence genes required for differentiation into specialised tissue[17]. Induced pluripotent stem cells (iPSCs), obtained by genetic reprogramming of somatic cells to an embryonic stem cell-like state, were generally assumed to be functionally equivalent to their ESC counterparts. A number of studies have demonstrated that this is not the case, however, showing that iPSCs retain 'epigenetic memory' of the donor tissue from which they were derived. Adult stem cells, which reside in most mammalian tissues, maintain tissue homeostasis and help repair and regenerate tissue in response to damage. Adult stem cells have their own unique epigenetic signature, which changes upon cellular differentiation[18]. The normal process of ageing results in a decline of stem cell function.

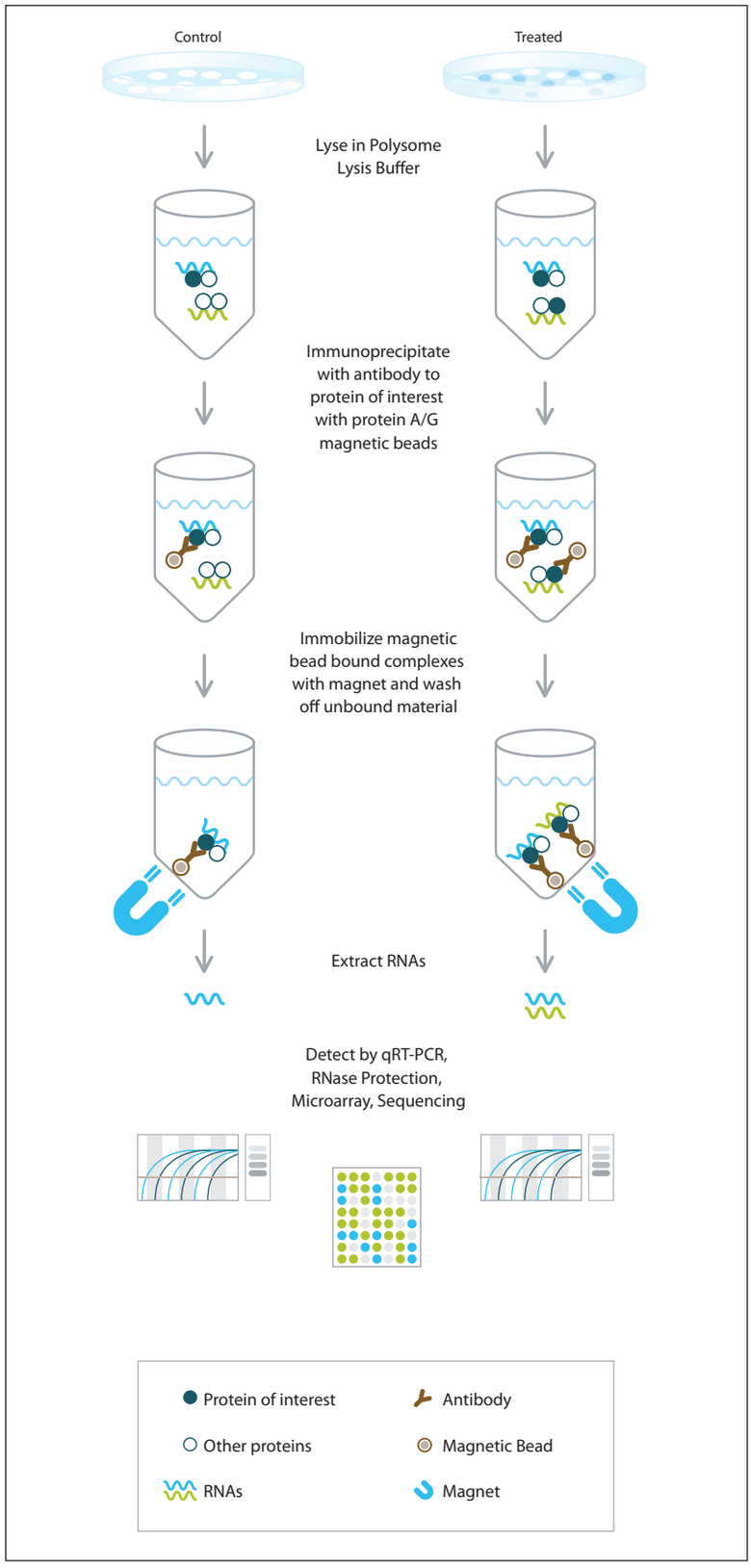

Figure 6: RNA-binding protein immunoprecipitation

Ageing can be thought of as how many times our stem cells have had to divide. In older individuals, there is an accumulation of these epigenetic events that is easily measurable in DNA. Every time a stem cell divides for either normal homeostasis or to repair an injury, it ages a little more with a decline in function. Stem cell ageing explains why adults who have been exposed to sun damage, for example, have skin that appears older than similar aged adults who have had minimal exposure to the sun. Every time the skin peels, the damage needs to be repaired by stem cells, which must divide. And every time a stem cell divides there is a finite chance of some sort of epigenetic alteration.

Unlike acquired DNA mutations, epigenomic changes are reversible. If they were permanent, ESCs would never be able to differentiate into specialised cells crucial for normal human functioning. Given the influence of extrinsic environmental factors on the epigenome and the reversibility of these modifications, it is possible that ageing can be modified at the epigenetic level. In fact, sirtuins, a class of proteins that possess histone deacetylase activity, have been under intense investigation since the discovery that they can extend the lifespan of yeast, worms and flies. The role of sirtuins in promoting human longevity and avoiding or delaying the onset of age-associated disorders is currently under investigation.

Targeting the epigenome

Dysregulated epigenetic modifications are now believed to play important roles in the onset and progression of human diseases, and it has been suggested that epigenetic changes may give rise to disease much more frequently than genetic variants[19]. In fact, a number of diseases and developmental disorders have been found to involve, and possibly be caused by, epigenetic alterations, including Beckwith-Wiedemann, Angelman, Prader-Willi, and Silver-Russell syndromes, Type I and Type II diabetes, autism, neurological disorders, obesity, almost all cancers and even cocaine addiction. From a drug discovery perspective, the fact that the epigenome is dynamic and the modifications reversible, targeting the epigenome to change the patterns of gene expression in an attempt to cure disease is a very real possibility.

A number of epigenetic drugs have already reached the market. Two DNA hypomethylation agents, Vidaza (azacitidine) and Dacogen (decitabine), are FDA-approved for the treatment of myelodysplastic syndromes (MDS), characterised by frequent epigenetic abnormalities and rapid evolution into acute myeloid leukaemia. Two histone

deacetylase (HDAC) inhibitors, Zolinza (vorinostat) and Istodax (romidepsin), are FDA-approved for the treatment of cutaneous T-cell lymphoma (CTCL). In addition, valproic acid, an HDAC inhibitor, is widely used to treat epilepsy and bipolar disorder, and is also being used for the treatment of migraine headaches and schizophrenia.

The principle of epigenetic therapy is to reverse pathological gene expression changes. There have been concerns that epigenetic therapy could adversely affect normal cells through reactivating, for example, imprinted genes, or even leading to the development of cancer. To date, these concerns have not been substantiated *in vivo*.

Novel epigenetic targets such as histone methylases and demethylases, histone acetylases, and readers (bromodomains and chromodomains) are currently the focus of drug discovery and development efforts. It is expected that more than six additional epigenetic drugs and a number of new diagnostic tests for cancer based on epigenetic biomarkers will receive FDA approval over the next decade.

Conclusion

Although our understanding of epigenetics remains in its infancy, the rapid development of new tools and technologies to define genome-wide epigenetic variations in humans will soon put this field on the same trajectory as genomics following improvements made in DNA sequencing technologies. The potential for effective new epigenetic therapies and diagnostic tests for a wide range of diseases beyond cancer is holding great promise. DDW

Dr John Rosenfeld *has held a scientific position in research and development at EMD-Millipore since 2003. Currently, he manages chromatin biology product development, innovating gene regulation research tools for life scientists, in particular kits and reagents for ChIP and RNA-binding protein immunoprecipitation (RIP). Dr Rosenfeld performed postdoctoral studies with Dr Ronald Evans of the Salk Institute of Biological Studies, and obtained his PhD in molecular biology and biochemistry from the University of California Irvine, in the laboratory of Dr Timothy Osborne. Dr Rosenfeld's scientific career and publication record has focused on transcription in mammalian systems, and he continues to develop research tools to advance understanding of epigenetic mechanisms of gene control.*

Dr Kan Saito *has been a Research and Development Scientist at EMD Millipore since 2007. Currently, he is in charge of the development of kits and reagents for RNA-binding protein immunoprecipitation (RIP) in Dr John Rosenfeld's group. Prior to joining EMD Millipore, Dr Saito was Seinor Postdoctoral Fellow in Dr Tomas Mustelin's laboratory at Burnham Institute for Medical Research. Dr Saito also performed a postdoctoral study in Dr Jean-Pierre Kinet's laboratory at Harvard Medical School as a research fellow of the Japan Society for the Promotional Science (JSPS) after he received a PhD in molecular biology from Tokyo Medical and Dental University. Dr Saito's research has been focused on signal transduction mechanisms and gene regulation mechanism of immune systems, and is now focusing on the development of next generation research tools for biological study, particularly in the emerging mechanisms of epigenetic and post transcriptional regulation.*

Dr Michael R. Sturges *has served as Senior Product Manager for Epigenetics at EMD Millipore since 2009. For the past 10 years he has worked in the bioscience industry in marketing and product development. This experience is complemented by more than 12 years of hands-on scientific research at both academic and commercial organisations. Dr Sturges earned his PhD at the University of California, Santa Cruz, and conducted his postdoctoral studies in the Department of Pediatrics in the laboratory of Dr Charles Roberts Jr at Oregon Health and Sciences University in Portland Oregon. Dr Sturges is currently applying his experience and knowledge of genomics, proteomics and molecular biology to drive the development of improved and novel kits, assays, antibodies and proteins that enable epigenetic research.*

References

1 Bygren, L, Kaati, G, Edvinsson, S. Acta Biotheoretica 2001; 49:53-59.
2 Pembrey, M et al. Eur J Hum. Genet. 2006; 14:159-166.
3 Jirtle and Skinner. Nat. Rev. Genet. 2007; 8:253-262.
4 Feinberg, A, Vogelstein, B. Nature 301:89-92,1983.
5 Clark, SJ. Hum. Mol. Genet. 2007; 16:R88-R95.
6 Waterland, R, Jirtle, R. Mol. Cell Biol. 2003; 23:5293-300.
7 Dolinoy et al. Proc. Natl. Acad. Sci. USA 2007; 104:13056-13061.
8 Ting, A, McGarvey, K, Baylin, S. Genes & Dev. 2006; 20:3215-3231.
9 Ting, A, McGarvey, K, Baylin, S. Genes & Dev. 2006; 20:3215-3231.
10 Zu, S, Han, L, Zhao, Z. Epigenetics 2011; 6:1-7.
11 Lister, R et al. Nature 2009; 462:315-322.
12 Lim, SJ et al. Bioinform 2010; 4:331-337.
13 Sadri-Vakili and Cha. Nat. Clin. Prac. Neurol. 2006; 2:330-338.
14 Sadri-Vakili and Cha. Nat. Clin. Prac. Neurol. 2006; 2:330-338.
15 Waterland and Jirtle. Mol. Cell. Biol. 2003; 23:5293-5300.
16 Manolio, T. N Engl J Med, 2010; 363:166-176.
17 Lunyak and Rosenfeld. Hum. Mol. Genet. 2008; 17:R28-R36.
18 Liu and Rando, J. Cell Biol. 2011; 193:257-266.
91 Simmons, D. Nature Education 2008; 1:1.