

EDITORIAL

The allure of the epigenome

Naomi Attar

The Oxford English Dictionary dates the word 'genome' to 1926 but has not yet seen fit to include the term 'epigenome' in its illustrious pages (where it would reside handsomely between 'epigenist' and 'epigenous'). This differential treatment reflects a state of affairs in which the epigenome has yet to infiltrate the popular imagination, but belies the recent explosion of epigenomics in the scientific literature, where as many as 70% of all abstracts featuring 'epigenome' have been published within the last three years.

The present-day proliferation in epigenomics, in the exploration of the dynamic regulatory layers that insulate the genome's static DNA sequence, has been enabled by novel high-throughput techniques for interrogating the positioning of DNA (hydroxy)methylation, histone marks and open chromatin. The ready availability of genomic data, without which we could not map the location of these features, has provided the essential context needed to make biological sense of the high-throughput data, and so propel epigenomics to the forefront of mainstream biology.

In this special issue, *Genome Biology* presents a collection of articles that describe a diverse range of novel insights into epigenomes, from human disease to ciliate reproduction to the containment of endogenous retroviruses. We also include a number of methods that will improve and simplify the study of epigenomics, in particular the computational steps that follow data generation. Finally, a selection of review and comment articles give an overview of current and future directions in epigenomics research.

A new software toolbox

The availability of new high-throughput methods creates a demand for software tools to process and analyze the overwhelming flow of unintelligible raw data that will inevitably be produced. *Genome Biology* has a proud history of publishing the most popular examples of such tools, with high profile examples including Bowtie [1] (next-generation sequencing data), MACS [2] (ChIP-seq

data) and DEseq [3] (RNA-seq data). The challenge of designing bioinformatics tools for the ever expanding number of DNA methylation genome-wide profiling methods has been taken up by many bioinformatics labs [4]. In the past few months, for example, *Genome Biology* has published SWAN [5], a method for reducing technical variation in data from the cutting-edge Illumina HumanMethylation450 BeadChip platform, and Bis-SNP [6], a method for calling SNPs in bisulfite sequencing data, which also has the advantage of improving the accuracy of methylation calls.

The special issue now adds a number of powerful new tools to assist those working with DNA methylation data. BatMeth [7] is a mapper for Illumina and SOLiD bisulfite sequencing reads, and is faster and more accurate than existing methods. BSmooth [8] takes your bisulfite sequencing data and tells you where the differentially methylated regions are, even if the coverage of your reads is low. methylKit [9] is a suite of tools for the annotation and statistical analysis of DNA methylation or hydroxymethylation data. EpiExplorer [10] is a webserver for super-fast epigenome browsing, interactive exploration and hypothesis generation, with output transferrable to other tools, such as Galaxy and the Genome HyperBrowser, for further analysis. The special issue even includes a new method for cheaper, simpler and more high-throughput RRBS (reduced representation bisulfite sequencing) [11], so we anticipate plenty more data being generated as inputs for these tools.

DNA methylation where it ought not be

The widespread presence of DNA methylation across the tree of life does have a few notable exceptions: for example, the model nematode species *Caenorhabditis elegans* is a *bona fide* methylation-free zone. By extension, conventional thinking has held that DNA methylation is missing throughout the nematode phylum. But researchers studying the nematode *Trichinella spiralis* now shatter this assumption as they describe, for the first time, the presence of DNA methylation in the Nematoda phylum [12]. Unlike the free-living *C. elegans*, *T. spiralis* is a parasitic organism, and is notable as a widespread food-borne pathogen that poses a problem in both human health and agriculture. *T. spiralis* belongs to a basal clade of Nematoda, having diverged from *C. elegans*

Correspondence: naomi.attar@genomebiology.com
Genome Biology, BioMed Central, 236 Gray's Inn Road, London WC1X 8HB, UK

several hundred million years ago in the late Precambrian [13], and so it seems possible that it has retained a DNA methylation machinery that may have been present in the ancestral nematode.

DNA methylation in *T. spiralis* can readily be explained by the presence of a (previously unnoticed) putative DNA methyltransferase in its genome, specifically a *DNMT3* homolog – a gene that intriguingly has no known counterpart in any other nematode [12]. But trickier to explain is the new report, also in this issue, of DNA methylation in the ciliate *Oxytricha trifallax*, an organism that has no recognizable DNA methyltransferase [14].

The study in *O. trifallax* finds that methylation acts as a marker for sites of DNA elimination during the genome rearrangement that is a signature feature of ciliated protist biology. Conversion to hydroxymethylation is a key step in this process and, unlike the mysteriously missing methyltransferase, putative homologs of the hydroxymethylation-generating Tet enzymes can in fact be found in the *O. trifallax* genome [14].

If these stories of weird methylation have left you scratching your head, we include a Research Highlight to explain the new findings in an easily digestible form, as well as to place them in the context of existing work [15].

DNA methylation in human health: cause or correlation?

Changes in promoter methylation have frequently been reported in cancers, although how these alterations relate to carcinogenesis is not clear. This special issue includes a comprehensive analysis of CpG island methylation in the promoters of several human cancers, from which it is concluded that hypermethylation occurs at sites already methylated in pre-cancerous tissue, and so is not likely to be contributing to gene silencing [16].

In addition to hypermethylation, hypomethylation at some promoters has also been associated with cancer. Another study in this issue shows that the hypomethylation previously observed in a rodent model of non-genotoxic carcinogenesis is the product of an active demethylation mechanism that begins by converting methylation sites to hydroxymethylation intermediates [17]. An accompanying Research Highlight argues for a new model of cancer epigenetics, in light of recent revelations about hydroxymethylation [18].

Whether causative or correlative, promoter DNA methylation changes that are associated with cancer can undoubtedly serve as useful biomarkers, as is demonstrated in an article that successfully identifies prognostic methylation biomarkers from neuroblastoma samples [19]. In the study, data from two genome-wide assaying methods were integrated with literature mining to pinpoint the biomarkers, which were then validated using a methylation-specific PCR assay [19].

Effective use of cancer DNA methylation biomarkers depends upon the availability of tumor cell samples. For other diseases, it may be possible to test a more convenient cell type, such as blood, where sampling of the affected tissue would be more invasive. However, such a strategy relies on the assumption that the DNA methylation biomarker in question is present across multiple tissues. A recent article in *Genome Biology* used matched brain and blood samples from the same individuals to show that some instances of inter-individual variation in DNA methylation are replicated in both tissues [20]. An article in this issue now adds support to the hypothesis that blood tests may give insight into DNA methylation changes in the brain [21]. In this case, the DNA methylation changes studied were those that are acquired during the course of human aging, and included sites in the genome associated with early-onset Alzheimer's disease [21].

Cancer and aging are far from the only areas of human health in which methylation has been implicated, and this issue also explores DNA methylation in human heart failure. In a perhaps surprising study, significant hypomethylation specific to satellite repeat elements is reported, with hypomethylation events being accompanied by a strong upregulation of the corresponding transcript [22]. Why one class of repetitive element in particular is implicated, with no significant methylation changes observed in any other class, is an intriguing question.

Epigenome death match: endogenous retroviruses versus genes

The havoc that repeat elements are liable to wreak upon our genomes, as it seems may sometimes be the case in human heart failure, is for the most part suppressed by DNA hypermethylation. The authors of one article in this issue wondered whether any leakage of DNA methylation from repeat elements into gene promoters might occur when a gene is located in close proximity to a repeat sequence. In an examination of endogenous retroviruses in mice, it was observed that, in most cases, the spreading of DNA methylation was contained before it reached gene promoters [23]. How does the genome prevent these promoters from being hypermethylated? Evidence is provided in the study for a defensive barrier formed by H3K4me3 and CTCF enrichment, with H3K4me3 – typically representative of euchromatin – even encroaching into the endogenous retrovirus element itself [23].

Every histone tells a story

Deciphering the genetic code was a task that occupied some of the world's sharpest minds (and neck-tie designers [24]) for more than a decade, but understanding the instruction underlying each codon has proved to be a

far simpler task than unraveling the meaning of each histone modification. A study in this issue uses ChIP-seq and mass spectrometry to characterize the histone mark H2A.Z in mouse and human embryonic stem cells, as well as in neuronal progenitor cells. The study clarifies the biology of a mark where previous work had reported a confusingly diverse array of functions, and identifies a novel dually modified form of H2A.Z, with both N-terminal acetylation and C-terminal ubiquitination [25].

Another research article in this issue focuses on H3K27me3, finding that this marker of Polycomb-mediated gene silencing is associated with paralog divergence following whole genome duplication in *Arabidopsis* [26]. The study is well complemented by a Review article on epialleles and plant evolution [27] – epigenomics plays a particularly intriguing role in plants, given that their lack of mobility relative to animals favors a potent epigenome capable of rapid responses to environmental changes.

Future work on histones will be aided by CHANCE, a user-friendly software package for assessing several different quality metrics in ChIP-seq data [28]. Obvious measures, such as the strength of the pull-down, form the core of CHANCE, but handy extras include a facility to compare your data to that produced by the ENCODE project, to see how well they match. CHANCE highlights the need for high quality ChIP-seq data; happily, recent methodological innovations have dramatically improved the resolution of this technique, as is highlighted in a Review article that also appears in this issue [29].

Concluding remarks

Our special issue hopes to give a flavor, to both the experienced and the novice epigenomicist, of the many exciting directions taken by contemporary research seeking to understand and to exploit the epigenome. *Genome Biology's* editorial team would like to thank the many scientists who collectively responded to our call for papers with a deluge of submissions, the many peer reviewers who – as ever – have been generous with their time and intellectual creativity, and our Guest Editor Dr Alexander Meissner, whose advice throughout his tenure has been invaluable and who has kindly written his own Editorial as an alluring introduction to the epigenomics field [30].

Published: 23 October 2012

References

- Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** *Genome Biol* 2009, **10**:R25.
- Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS: **Model-based Analysis of ChIP-Seq (MACS).** *Genome Biol* 2008, **9**:R137.
- Anders S, Huber W: **Differential expression analysis for sequence count data.** *Genome Biol* 2010, **11**:106.
- Bock C: **Analysing and interpreting DNA methylation data.** *Nat Rev Genet* 2012, **13**:705-719.
- Maksimovic J, Gordon L, Oshlack A: **SWAN: Subset-quantile Within Array Normalization for Illumina Infinium HumanMethylation450 BeadChips.** *Genome Biol* 2012, **13**:R44.
- Liu Y, Siegmund KD, Laird PW, Berman BP: **Bis-SNP: Combined DNA methylation and SNP calling for Bisulfite-seq data.** *Genome Biol* 2012, **13**:R61.
- Lim JQ, Tennakoon C, Li G, Wong E, Ruan Y, Wei CL, Sung WK: **BatMeth: improved mapper for bisulfite sequencing reads on DNA methylation.** *Genome Biol* 2012, **13**:R82.
- Hansen KD, Langmead B, Irizarry RA: **BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions.** *Genome Biol* 2012, **13**:R83.
- Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE: **methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles.** *Genome Biol* 2012, **13**:R87.
- Halachev K, Bast H, Albrecht F, Lengauer T, Bock C: **EpiExplorer: live exploration and global analysis of large epigenomic datasets.** *Genome Biol* 2012, **13**:R96.
- Boyle P, Clement K, Gu H, Smith ZD, Ziller M, Fostel JL, Holmes L, Meldrum J, Kelley F, Gnirke A, Meissner A: **Gel-free multiplexed reduced representation bisulfite sequencing for large-scale DNA methylation profiling.** *Genome Biol* 2012, **13**:R92.
- Gao F, Liu X, Wu X, Wang X, Gong D, Lu H, Song Y, Wang J, Du J, Liu S, Han X, Tang Y, Yang H, Jin Q, Zhang X, Liu M: **Differential DNA methylation in discrete developmental stages of the parasitic nematode *Trichinella spiralis*.** *Genome Biol* 2012, **13**:R100.
- Mitreva M, Jasmer DP, Zarlenga DS, Wang Z, Abubucker S, Martin J, Taylor CM, Yin Y, Fulton L, Minx P, Yang S-P, Warren WC, Fulton RS, Bhonagiri V, Zhang X, Hallsworth-Pepin K, Clifton SW, McCarter JP, Appleton J, Mardis ER, Wilson RK: **The draft genome of the parasitic nematode *Trichinella spiralis*.** *Nat Genet* 2011, **43**:228-235.
- Bracht JR, Perlman DH, Landweber LF: **Cytosine methylation and hydroxymethylation mark DNA for elimination in *Oxytricha trifallax*.** *Genome Biol* 2012, **13**:R99.
- Yi S: **Birds do it, bees do it, worms and ciliates do it too: DNA methylation from unexpected corners of the tree of life.** *Genome Biol* 2012, **13**:174.
- Sproul D, Kitchen RR, Nestor CE, Dixon JM, Sims AH, Harrison DJ, Ramsahoye BH, Meehan RR: **Tissue of origin determines cancer-associated CpG island promoter hypermethylation patterns.** *Genome Biol* 2012, **13**:R84.
- Thomson JP, Lempiäinen H, Hackett JA, Nestor CE, Müller A, Bolognani F, Oakeley EJ, Schübeler D, Terranova R, Reinhardt D, Moggs JG, Meehan RR: **Non-genotoxic carcinogen exposure induces defined changes in the 5-hydroxymethylome.** *Genome Biol* 2012, **13**:R93.
- Song CX, He C: **Balance of DNA methylation and demethylation in cancer development.** *Genome Biol* 2012, **13**:173.
- Decock A, Ongenaert M, Hoebeek J, De Preter K, Van Peer G, Criekinge WV, Ladenstein R, Schulte JH, Noguera R, Stallings RL, Van Damme A, Laureys G, Vermeulen J, Van Maerken T, Speleman F, Vandesompele J: **Genome-wide promoter methylation analysis in neuroblastoma identifies prognostic methylation biomarkers.** *Genome Biol* 2012, **13**:R95.
- Davies MN, Volta M, Pidsley R, Lunnon K, Dixit A, Lovestone S, Coarfa C, Harris RA, Milosavljevic A, Troakes C, Al-Sarraj S, Dobson R, Schalkwyk LC, Mill J: **Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood.** *Genome Biol* 2012, **13**:R43.
- Horvath S, Zhang Y, Langfelder P, Kahn RS, Boks MPM, van Eijk K, van den Berg LH, Ophoff RA: **Aging effects on DNA methylation modules in human brain and blood tissue.** *Genome Biol* 2012, **13**:R97.
- Haider S, Cordeddu L, Robinson E, Movassagh M, Siggens L, Vujic A, Choy MK, Goddard M, Lio P, Foo R: **The landscape of DNA repeat elements in human heart failure.** *Genome Biol* 2012, **13**:R90.
- Rebollo R, Miceli-Royer K, Zhang Y, Farivar S, Gagnier L, Mager DL: **Epigenetic interplay between mouse endogenous retroviruses and host genes.** *Genome Biol* 2012, **13**:R89.
- Nobelprize.org: How the code was cracked.** [<http://www.nobelprize.org/educational/medicine/gene-code/history.html>].
- Ku M, Jaffe JD, Koche RP, Rheinbay E, Endoh M, Koseki H, Carr SA, Bernstein BE: **H2A.Z landscapes and dual modifications in pluripotent and**

- multipotent stem cells underlie complex genome regulatory functions. *Genome Biol* 2012, **13**:R85.
26. Berke L, Sanchez-Perez GF, Snel B: Contribution of the epigenetic mark H3K27me3 to functional divergence after whole genome duplication in *Arabidopsis*. *Genome Biol* 2012, **13**:R94.
 27. Colot V, Weigel D: Epialleles in plant evolution. *Genome Biol* 2012, **13**:249.
 28. Diaz A, Nellore A, Song JS: CHANCE: comprehensive software for quality control and validation of CHIP-seq data. *Genome Biol* 2012, **13**:R98.
 29. Zentner GE, Henikoff S: Surveying the epigenomic landscape, one base at a time. *Genome Biol* 2012, **13**:250.
 30. Meissner A: What can epigenomics do for you? *Genome Biol* 2012, **13**:420.

doi:10.1186/gb-2012-13-10-419

Cite this article as: Attar N: The allure of the epigenome. *Genome Biology* 2012, **13**:419.