Human Molecular Genetics, 2005, Vol. 14, Review Issue 1 **R3–R10**
doi:10.1093/hmg/ddi110

From genome to epigenome

Adele Murrell*, Vardhman K. Rakyan and Stephan Beck

The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Received January 14, 2005; Revised February 10, 2005; Accepted February 17, 2005

The success of the human genome sequencing project has created wide-spread interest in exploring the human epigenome in order to elucidate how the genome executes the information it holds. Although all (nucleated) human cells effectively contain the same genome, they contain very different epigenomes depending upon cell type, developmental stage, sex, age and various other parameters. This complexity makes it intrinsically difficult to precisely define 'an' epigenome, let alone 'the' epigenome. What is clear, however, is that in order to unravel any epigenome, existing and novel high-throughput approaches on the DNA, RNA and protein levels need to be harnessed and integrated. Here, we review the current thinking and progress on how to get from the genome to the epigenome(s) and discuss some potential applications.

INTRODUCTION

The term 'epigenetics' was first introduced by Conrad Waddington in the 1940s to describe 'the interactions of genes with their environment, which bring the phenotype into being' (1). The ensuing research on gene regulation during cell differentiation and development has raised most of the questions that are still at the core of epigenetics. Early epigenetic studies described heterochromatin as regions of the genome that have low gene density, contain satellite repeat elements and are late replicating. Subsequently, it was shown that heterochromatin and euchromatin are associated with distinct DNA methylation and histone modification patterns that correlate with particular states of gene activity, leading to the idea of an 'epigenetic code' that determines the chromatin state and, consequently, gene expression (2,3). More recently, an epigenetic memory system mediated by polycomb (pcG) and trithorax group (trxG) proteins to maintain silent chromatin states has been uncovered (4,5), and short hairpin RNAs have been shown to play a role in heterochromatin formation via an RNAi pathway (6,7), placing heterochromatin at the core of epigenetic silencing. Heritable variations in gene expression, such as gene silencing due to paramutation in plants, X-inactivation in mammals and genomic imprinting, have highlighted the complexity of gene regulation, so that nowadays epigenetics is simply defined as heritable changes in gene expression not attributable to nucleotide sequence variation.

Following the trend from local to global analyses, the term epigenomics was introduced for the study of epigenetic changes on a genome-wide basis (8). Fundamentally, epigenomics is the study of the effects of chromatin structure including the higher order of chromatin folding and attachment to the nuclear matrix, packaging of DNA around nucleosomes, covalent modifications of histone tails (acetylation, methylation, phosphorylation, ubiquitination) and DNA methylation. These epigenetic components are all amenable to genome-wide study, and integrated studies that correlate gene expression with DNA methylation and chromatin profiles need to be designed. However, it will be of limited value to study the epigenome in a generic cell. Traditional genomic resources such as cell lines undergo expression and methylation changes in culture, whereas primary tissues are often made up of numerous cell types. Moreover, epigenetic changes can also occur as a result of external factors such as age and diet. Thus, serious consideration needs to be given to origins of cell type as well as the developmental stage. Various tissues of diseased and healthy origin will need to be studied, and as a baseline, epigenetic profiles need to be established in 'normal' tissues.

The enormous interest in epigenetics has encouraged several groups to exploit whole genome approaches to embark upon characterizing the epigenome. Dedicated academic centres to study epigenetics are being formed in various universities, the largest being the Center for the Epigenetics of Common Human Disease at Johns Hopkins (http://www.hopkinsmedicine.org/epigenetics/). In Europe, two major international consortia of similar name but different and complementary aims have been formed. The Human Epigenome Project (HEP) (www.epigenome.org) is a joint effort by an international collaboration, which was established in 1999 with the aim to identify, catalogue and interpret genome-wide DNA methylation patterns and profiles of all

*To whom correspondence should be addressed. Tel: +44 1223494753; Fax: +44 1223494919; Email: amm@sanger.ac.uk

human genes in all major tissues. The Epigenome Network of Excellence (www.epigenome-noe.net) was established in 2004 with the aim to provide a portal to a vast array of epigenetics resources and information for both the scientists and the interested public and to advance epigenetic research (for human and model organisms) into chromatin modification, nucleosome dynamics, non-coding RNA and gene silencing, X-chromosome inactivation and imprinting, transcriptional memory, assembly and nuclear organization, cell fate and disease and epigenomic maps.

TECHNOLOGY PLATFORMS

Central to successfully resolving any epigenome is the availability of robust technologies and assays to generate quantifiable and reliable data that can be integrated with existing genome annotation. Figure 1 suggests how the key technology platforms of DNA methylation and chromatin profiling could be integrated with expression profiling platforms and existing genome data and bioinformatics to produce a multidimensional epigenome database. Direct comparison of the various technology platforms to find the most sensitive, accurate and highest resolution platform is not really meaningful because the endpoints are different (e.g. global methylation profiles across the whole genome or detailed methylation maps across preselected regions of the genome). Table 1 summarizes the major technology platforms with their individual methodologies. These methodologies enable large amounts of DNA methylation, chromatin and expression profiling data to be generated that can be processed through bioinformatics and be superimposed upon the genome to ultimately create an epigenome database.

DNA methylation profiling platforms

Gene regulation is influenced by interactions between histone modifications and DNA methylation. DNA methylation is proposed to recruit methyl-DNA binding proteins that associate with histone deacetylases (9,10). However, DNA methyltransferases can also target histone deacetylases, leading to histone modifications that are independent of methyl-DNA binding proteins (11). Regardless of where DNA methylation occurs in the cascade of histone modification, it remains the most accessible epigenomic feature because of its stability. If we are going to start looking for 'epigenetic signatures' associated with various diseases on the basis of genome-wide chromatin profiling or merely concentrate on specific loci or groups of genes, knowledge of where CpG methylation occurs within the genome (or epigenome) will be invaluable. The baseline normal methylation profile will also be helpful when we try to elucidate complex processes such as genomic imprinting, X-chromosome inactivation, gene regulation, chromatin structure, genome stability and complex multifactor diseases such as immune disorders and cancer.

For whole genome methylation analyses, the key parameters are whether we can analyse multiple CpGs in several genes at once and whether methylation levels can be quantified. Further consideration includes whether we intend covering the whole genome without preselecting specific regions and at what resolution we intend analyzing the results.

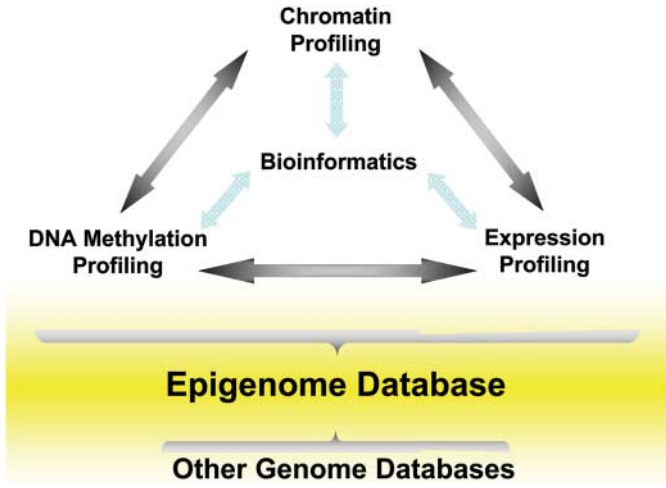

Figure 1. Potential scheme for integrated technology platforms to produce an epigenome database.

Technologies for DNA methylation analyses remain either PCR based (after bisulphite conversion of unmethylated cytosines to uracil) or methylation sensitive restriction enzyme based (reviewed in 12–14). Microarray technology and comparative genomic hybridization have further opened the field for high-throughput methylation analyses, and the various advantages and disadvantages have been extensively reviewed in the last 18 months (12,13,15,16). Modifications of the contemporary methodology include using matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) for quantitative detection of methylation after primer extension for discrimination between methylated and unmethylated CpGs on bisulphite treated DNA (17). MALDI-MS offers a high degree of automation and integration and allows for discrimination of methylation levels that differ by $\geq$5%. This sensitivity is almost matched by pyrosequencing protocols for bisulphite analyses which currently enable only a few CpGs to be analysed at a time. Pyrosequencing assays that enable high-throughput analyses of up to 10 CpGs in an amplicon size of 300 bp are presently being developed (18) and promise to be quantitative, but expensive. Real time PCR approaches may quantitatively detect methylation specific amplicons (19). At present, bisulphite conversion of DNA followed by PCR and sequencing remains the gold standard of methylation analysis. Bioinformatics programs to analyse raw sequence data for quantitative differences at individual CpGs and which can align multiple sequences to identify variable methylation are already in place for managing the large amount of data generated when bisulphite sequencing is done on a large scale (20).

As an alternative to high-throughput bisulphite analyses, microarray based methods have the advantage of being faster and do not necessarily rely on pretreatment of DNA with bisulphite. However, the limited availability of well-defined microarrays and the low resolution of methylation profiles obtained from microarray based methods presently hampers the widespread use of DNA microarray technology for global methylation profiling. The resolution obtained with genomic microarrays depends upon the size and content of the clones spotted, with BAC and cosmid clones giving 100 and 40 kb

Table 1. Comparison of current methodologies for use in high-throughput epigenome profiling

	Pretreatment of material	Detection	Prior sequence selection	Global (G)/ individual (I) CpG mapping	Whole genome coverage	Quantifiable
DNA methylation maps	Bisulphite treatment of DNA	PCR and sequencing	Y	I	N	N
		Pyrosequencing/MALDI-MS	Y	I	N	Y
		Q-PCR	Y	I	N	Y
		Hybridization to CpG and TpG oligo microarrays	Y	I	N	Y
	Restriction enzyme techniques					
	RLGS	Cloning into libraries and sequencing	N	G	Y	N
	McrBC/Meth sensitive RE MS-RDA	Hybridization to genomic microarrays	N	G	M	Y
	ChIP with antibodies to MBDs	Cloning into libraries and sequencing	N	G	Y	N
		Hybridization to genomic microarrays	N	G	M	Y
Histone modification maps	ChIP	Hybridization to genomic/ amplicon/oligo microarrays (ChIP-chip)	N		M	
		Cloning into libraries and sequencing	N		Y	
		ChIP-BS	Y		N	
Physical chromatin structure	Chromatin fractionation	Hybridization to genomic/ amplicon/oligo microarrays	N		M	

Y, yes; N, no; M, maybe depending on the coverage of the microarray.

resolution and also containing high amounts of repetitive DNA sequences. The most viable options for methylation analyses are between custom arrays of oligos or amplicons covering regions-of-interest. These arrays increase resolution but entail sequence preselection.

Methylation specific oligonucleotide arrays containing oligonucleotides that can distinguish between bisulphite converted TpG dinucleotides and methylated CpG dinucleotides have been described (16,21), and custom microarray panels using clones of CpG islands (CGI library clones) have been generated (22). The initial CGI library was isolated through use of affinity purified methyl CpG binding domain of MeCP2 (0.2–2 kb sized clones), and the initial experiments were performed prior to sequencing of the clones. Alternative methods of generating methylated DNA libraries rely on methylation sensitive restriction digestion followed by amplification protocols that enrich for methylated CGI (23,24). Most of the currently available DNA microarrays generally represent only a small fraction of the genome, and the size of arrayed fragments affects resolution. Thus, an array of 200 kb BAC clones will give relatively low resolution but may cover a larger region of the genome compared with PCR amplicon probes or oligo tilepaths across specific regions-of-interest.

Enzyme based global methylation methods are still being used as an alternative to bisulphite treatments. A number of restriction enzymes are methylation sensitive and do not recognize restriction sites with methylated cytosines. The use of McrBC, a GTP-requiring, modification-dependent endonuclease of *Escherichia coli* K-12, which specifically recognizes DNA sites of the form 5′ R(m)C 3′ has been employed to deplete methylation rich sequences while constructing plant libraries (25,26) and has been recently used to comprehensively analyse CGIs in chromosome 21q (27). Enzyme based isolation of methylated or unmethylated DNA has the advantage that the output material can more easily be cloned into libraries, enabling global analyses and avoiding biases introduced through preselection of sequences to be analysed. Methylation sensitive representational difference analysis (Me-RDA) has been one of the first technique used to specifically screen the whole genome for imprinted genes on the basis of differential methylation (28–30). Restriction landmark genomic scanning (RLGS) is a type of two-dimensional electrophoresis relying on restriction digestion with rare cutting methylation sensitive *Not*1, radioactive labelling and separating fragments in one direction and then followed by in-gel restriction with an enzyme of choice to obtain profiles for analyses (31–33). Downstream analyses include spot cloning and identification by sequencing, comparison to arrayed libraries of *Not*1–*Eco* RV fragments or computer based approaches (discussed subsequently). This technique and its applications for methylation detection have been extensively reviewed elsewhere (32).

Chromatin profiling platforms

Chromatin structure is an integral part of the epigenome and has started to be unravelled at genomic levels. Genome-wide maps of DNaseI hypersensitive sites have been described after sequencing libraries either enriched for or depleted of

hypersensitive DNA (34,35). These maps are useful in the genome-wide identification of regions containing regulatory elements. The first chromatin structure map of the whole human genome, mapping the distribution of compact and open chromatin fibre to the genome and correlating compaction status with gene density and expression status in lymphoblastoid cells, has recently been described (36). In this study, hybridization of density fractionated chromatin to genomic DNA microarrays enabled high resolution maps showing that compact chromatin fibres are not only composed of heterochromatin but also contain some active genes, whereas open chromatin fibers correlate with regions of high gene density rather than gene expression. Further functional read outs of chromatin structure include high resolution maps of replication timing on chromosomes 22 and 6 (37–39). It would be interesting to incorporate the methylation data obtained in the HEP with the replication timing data.

Microarray technology combined with chromatin immunoprecipitation (ChIP) procedures has been applied to study chromatin structure (ChIP-chip) (40). DNA methylation analyses can be followed by ChIP for histone modifications, methylbinding proteins, transcription factors, chromatin modifiers and secondary chromatin structure (41). One limitation of this technique is that DNA extracted after immunoprecipitation needs to be linearly amplified prior to hybridization. Protocols which do not rely on linear amplification are more likely to yield quantifiable results. The best methodology for ChIP-chip has not been established, and the majority of primary papers describe using this technology successfully in yeast because of the compact and non-repetitive nature of yeast genomes (reviewed in 42). In *Drosophila*, global patterns of histone acetylation and methylation have been mapped and correlated to gene expression using ChIP-chip technology (43). Bisulphite sequencing can also be done after chromatin immunoprecipitation (ChIP-BS) (44), which could be useful for examining DNA methylation status in combination with histone modification on a relatively large scale. Preparative ChIP to create libraries enriched for specific transcription factors or chromatin features (e.g., to look for genes regulated by the boundary element CTCF) (45) has been undertaken, but so far, these have not yet been fully sequenced and incorporated into a chromatin map.

The epigenome is not a linear system of neatly aligned nucleosomes subjected to histone modifications and DNA methylation changes. We also need to think about gene expression and genome usage in a multidimensional way taking into consideration long range interactions of regulatory regions and secondary chromatin structure. It is now proved that long range interactions between gene regulatory regions occur through chromatin loops and that the DNA methylation status may either depend upon the loop or influence the loop structure (46,47). Additionally, we should not only be trying to integrate different aspects of epigenetic regulation but also be interpreting this information within the context of the whole nucleus. In recent years, it has been shown that the genome is organized within dynamic chromosome territories, which impacts upon gene expression (48–52). Indeed, it is foreseeable that the next echelon of epigenome maps will be three-dimensional spatial maps of human chromosomes in the nucleus.

Expression profiling platforms

To correlate the epigenome with gene expression, quantitative measurements of expression are required. Currently, profiling of whole genome gene expression patterns is being widely performed in both basic and applied research, using techniques such as high-throughput microarrays and real-time PCR methods. Other techniques such as parallel signature sequencing on microbeads (53) and serial analysis of gene expression (reviewed in 54) also provide powerful quantitative approaches for determining expression levels. As RNA and protein levels are subjected to post-transcriptional and translational regulation, accurate correlations between epigenomic and expression profiles may be difficult to establish. Additionally, heritable variation in gene expression exists which may be due to sequence variation. An optimal approach would combine allelic gene expression data with a catalogue of candidate regulatory polymorphisms. ChIP technology using antibodies to RNA polymerase II (RNAII pol-ChIP) (55) can be used to establish nascent transcription profiles and chromatin profiles. If the epigenetic code holds true, then chromatin and DNA methylation profiles could eventually predict gene expression patterns. This will become evident as more DNA methylation and histone modification patterns correlated with particular states of gene activity emerge.

Bioinformatics

Bioinformatic approaches to genome-wide prediction of CpG methylation have been limited to *in silico* simulation analyses such as comparing virtual image restriction landmark genomic scanning with real RLGS in *Arabidopsis* and mice (31), but no complete CpG maps have yet been completed. Further computational search algorithms for epigenetic features have been applied to search for imprinting signatures (56,57) and pcG elements (58). As part of the Human Genome Project, bioinformatics was invaluable for integrating curated and computationally predicted genomic data into flexible, public databases such as ENSEMBL and the UCSC Genome Browser. Except for the DNA methylation databases (www.epigenome.org and www.methdb.de), as yet, no general epigenome database that integrates all epigenetic data derived from the DNA, RNA, chromatin and protein levels exists. One of the main problems hampering the development of such an epigenome database is the lack of primary databases (such as EMBL/GenBank/DDBJ) to which epigenetic data can be submitted.

EPIGENOME PROJECTS

Systematic whole genome DNA methylation studies are just beginning. The comprehensive survey of allelic CGI methylation across human chromosome 21q in peripheral blood indicated that up to 20% of the CGIs were methylated and also detected two novel imprinted loci as well a non-imprinted gene with monoalleleic methylation of its CGI (27). Large scale structure of genomic methylation profiles in the brain is currently being mapped using similar enzyme based methods for the fractionation of methylated and unmethylated domains of the genome (T. Bestor, personal communication).

Apart from the already mentioned chromatin fibre map (36) and the DNA replication maps, histone modifation maps are beginning to emerge, including a high resolution genomic ChiP-chip analyses of H3 Lys4 methylation and H3 Lys9/14 acetylation for human chromosomes 21 and 22 in a human hepatoma cell line (59). The same work included comparative human and mouse primary fibroblast maps for Lys4 methylation at selected loci (59). Smaller scale epigenome projects such as the intergrated profiling of gene expression and chromatin modifications (histone modifications and DNA methylation) in *Drosophila* (43) and *Arabidopsis* (60) have been undertaken using the previously described technology platforms. The latter authors show that heterochromatin in *Arabidopsis* is determined by transposable elements and related tandem repeats, under the control of the chromatin remodelling protein DDM1 (60,61).

Human epigenome project (HEP)

The HEP aims to systematically analyse DNA methylation in the regulatory regions of all known genes in most major cell types and their diseased variants along with high-density snapshots of non-genic regions spread evenly across the human genome. Methylation variable positions (MVPs) are thought to reflect gene activity, tissue type and disease state and are useful epigenetic markers revealing the dynamic state of the genome. MVPs are defined as CpG sites with statistical power to discriminate between different biological samples and/or states. Akin to single nucleotide polymorphisms (SNPs), MVPs will greatly advance our ability to elucidate and diagnose the molecular basis of human diseases.

As a pilot study, DNA methylation profiling was carried out on the human major histocompatibility complex (MHC), one of the most gene-dense regions in the human genome, containing genes with a high diversity of function located on chromosome 6 (61). For the pilot study, an integrated pipeline for high-throughput methylation analysis using bisulphite DNA sequencing, MVP discovery, epigenotyping by MALDI-MS and a public database were developed. DNA methylation levels within regulatory, exonic and intronic regions associated with 90 genes (i.e. >70% of all expressed genes within the MHC) were analysed in seven human tissues—adipose, brain, breast, liver, lung, muscle and prostate—with multiple samples from different individuals. For the DNA methylation profiling of the human MHC, regions with potential regulatory functions and CpGs dense regions of a gene were selected for sequencing in addition to CGIs. This selection was dependent on annotated sequence data and, whereas CGIs and CpG dense regions within genes were easily identifiable, the precise locations of all the promoters within the human MHC were unknown at the time this study was initiated. Therefore, sequences 2 kb upstream of all annotated start codons were also selected for study to ensure that promoters and upstream regulatory regions were included. Although the MHC is one of the most thoroughly annotated regions of the genome, the annotation is still ongoing, and we foresee that more sequences will be added to the HEP map as antisense genes and regulatory RNAs are identified.

The overall methylation profile of the MHC was shown to be bimodal with >90% of the regions tested being either hyper- or hypomethylated; however, heterogeneity at individual CpG sites was frequently observed. These results are similar to bimodal genomic methylation profiles observed previously by several authors (reviewed in 62) and confirm the results of others who have shown heterogeneous methylation profiles of individual genes *in vivo* (63,64). 80% of the CpGs analysed in the HEP pilot displayed methylation levels that varied by >20%, either between individuals and/or tissues. Upstream regions (5′-UTR, and promoter regions) of genes analysed were more likely to be hypomethylated compared with intragenic regions, and introns were less likely to be methylated than exons. Comparisons of DNA methylation with expression levels for MHC genes in several tissues indicated that hypermethylation of upstream regions was associated with gene silencing. A web-based ENSEMBL-like genome browser has been created for displaying HEP data which are publicly available (Fig. 2). Following further scale-up, methylation profiling of all known genes (around 3000) on chromosomes 6, 13, 20 and 22 are now underway.

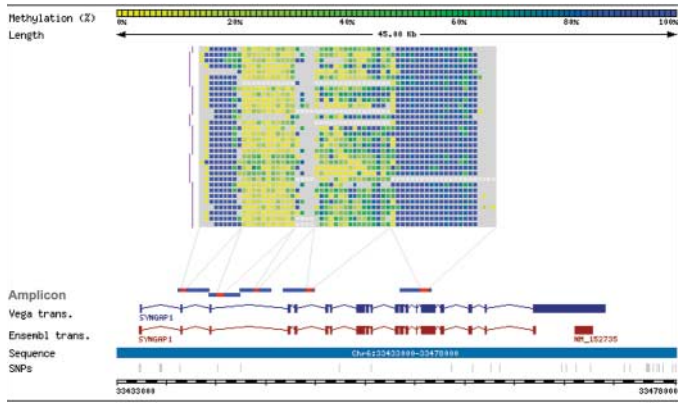

Figure 2. The HEP database, a web-based, ENSEMBL-like genome browser for displaying HEP data which is publicly available at http://www.epigenome.org/data/. The methylation levels are displayed in the form of a matrix. Each matrix contains the data obtained from all the samples of one amplicon. Each colour-coded square (yellow represents 0% methylation, blue represents 100% methylation and green represents intermediate levels) within the matrix represents one CpG site. Clicking on a square reveals the tissue source of the sample and the level of methylation observed at that particular CpG site. Grey squares indicate CpG sites for which methylation levels could not be determined. Each row of squares represents all the CpG sites for one sample of a particular amplicon and the samples are grouped into tissue type. The red bar indicates the genomic region analysed. Also shown are chromosome coordinates, CGI SNPs, ENSEMBL and high-quality manually curated VEGA transcript information. The HEP-database links to the ENSEMBL genome browser, providing additional information about the region-of-interest. The example shows amplicons within the *SynGAP 1* gene that correspond to regions that were determined to be hypomethylated (second amplicon from the left), hypermethylated (first and fifth amplicons) or heterogeneously methylated (fourth amplicon).

Applications of the epigenome

The future potential of the epigenome is wide-ranging. In addition to advancing basic research, the epigenome has immediate applications for diagnostics, and as epigenetic alterations are potentially reversible, it has potential appli-

cations for therapeutics as well. As a resource, the HEP will provide the normal baseline level of DNA methylation as a reference for subsequent profiling in the context of cancer and complex disease. Methylation profiling technologies promise to enable the characterization of distinct methylation signatures for complex diseases and various cancers with diagnostic implications. DNA methylation is now considered a potential biomarker in cancer (65,66). Cancers can be classified according to their degree of methylation, and those cancers with high degrees of methylation (the CpG island methylator phenotype) represent a clinically and aetiologically distinct group that is characterized by 'epigenetic instability' (reviewed in 66). Epigenetic therapies propose using global DNA methylation inhibitors to reverse gene silencing caused by altered methylation (67).

The HEP is essentially embarking on practical epigenotyping by identifying and classifying epigenetic marks that are transmitted vertically, for example, inter-individual variants and tissue-specific variants. In the first instance, we can ask whether epigenetic variation is less between monozygotic twins than siblings. Variation in gene expression between alleles is not restricted to regulation by genomic imprinting or X-inactivation. Indeed, allele expression variation is relatively common in humans and differentially expressed genes are distributed throughout the genome (68,69). Moreover, some alleles, known as epialleles, have variable expressivity in the absence of genetic heterogeneity because of their epigenetic states (reviewed in 70). The mechanisms responsible for variable gene expression can now be unravelled by relating epigenotype variation to genotype variation and haplotypes in normal individuals. In complex diseases, the frequency and disease onset time may be influenced by epigenetic variants and age-dependent epigenetic changes (71). It is conceivable that variation in methylation status of a gene could be affected by genotype either directly, where genetic variation could introduce or remove CpG sites which are susceptible to methylation or indirectly, by introducing sequences (e.g. repeat elements) that affect methylation in *cis*. Loss of imprinting of the *IGF2* gene is present in 10% individuals who showed no sign of the imprinted growth disorder, that is, Beckwith–Wiedeman syndrome (BWS). In BWS patients, specific haplotypes within the *IGF2* gene have been associated with loss of methylation at the locus (72), suggesting that epigenetic and genetic variation may act synergistically to influence a phenotype. For this, we suggest the introduction of 'hepitype' which combines haplotype and epitype information and allows dissecting out subtle epigenetic contributions to a given phenotype. The basic concept of hepitypes is illustrated in Figure 3.

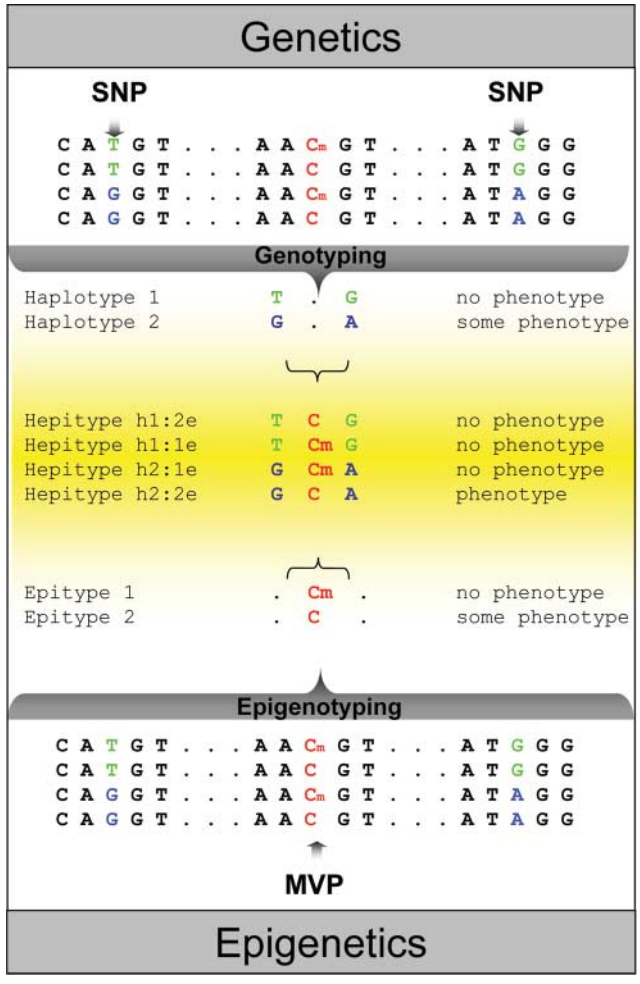

Figure 3. Basic concept of Hepitypes: genotyping for SNPs to define haplotypes (top panel) and epigenotyping for MVPs to define epitypes (bottom panel) can be brought together to define hepitypes which could be associated with a disease phenotype and which could otherwise go undetected by genotyping or epigenotyping in isolation.

OUTLOOK

While the discussion continues on what constitutes 'the' epigenome and how the underlying data can be integrated, international frameworks are now in place to make a start on those epigenetic data sets that can be defined and for which sufficiently mature technologies exist for genome-wide scale-up. DNA methylation profiling clearly has reached that stage and is likely to provide us the first glimpse into an epigenome, similar to EST sequences that provided the first view into the human genome. Considering the current excitement surrounding the epigenome, completion of the HEP is expected to be self-catalyzing and could happen before the end of this decade.

ACKNOWLEDGEMENTS

A.M. and S.B. were supported by the Wellcome Trust and V.R. was supported by a C.J. Martin Fellowship from the National Health and Medical Research Council of Australia.

REFERENCES

1. Waddington, C. (1942) The Epigenoptype. *Endeavour*, **1**, 18–20.
2. Jenuwein, T. and Allis, C.D. (2001) Translating the histone code. *Science*, **293**, 1074–1080.
3. Wang, Y., Fischle, W., Cheung, W., Jacobs, S., Khorasanizadeh, S. and Allis, C.D. (2004) Beyond the double helix: writing and reading the histone code. *Novartis Found. Symp.*, **259**, 3–17.

4. Orlando, V. (2003) Polycomb, epigenomes, and control of cell identity. *Cell*, **112**, 599–606.
5. Levine, S.S., King, I.F.G. and Kingston, R.E. (2004) Division of labor in polycomb group repression. *Trends Biochem. Sci.*, **29**, 478–485.
6. Jenuwein, T. (2002) Molecular biology. An RNA-guided pathway for the epigenome. *Science*, **297**, 2215–2218.
7. Lippman, Z. and Martienssen, R. (2004) The role of RNA interference in heterochromatic silencing. *Nature*, **431**, 364–370.
8. Beck, S., Olek, A. and Walter, J. (1999) From genomics to epigenomics: a loftier view of life. *Nat. Biotechnol.*, **17**, 1144.
9. Fuks, F., Hurd, P., Wolf, D., Nan, X., Bird, A. and Kouzarides, T. (2003) The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *J. Biol. Chem.*, **278**, 4035–4040.
10. Hashimshony, T., Zhang, J., Keshet, I., Bustin, M. and Cedar, H. (2003) The role of DNA methylation in setting up chromatin structure during development. *Nat. Genet.*, **34**, 187–192.
11. Burgers, W.A., Fuks, F. and Kouzarides, T. (2002) DNA methyltransferases get connected to chromatin. *Trends Genet.*, **18**, 275–277.
12. Huang, T.H.-M., Plass, C., Liang, G. and Laird, P.W. (2003) Epi meets genomics: technologies for finding and reading the 5th base. In Beck, S. and Olek, A. (eds), *The Epigenome. Molecular Hide and Seek.* Wiley-VCH GmbH & Co.KGaA, Germany, pp. 41–64.
13. Fazzari, M.J. and Greally, J.M. (2004) Epigenomics: beyond CpG islands. *Nat. Rev. Genet.*, **5**, 446–455.
14. Liu, Z.J. and Maekawa, M. (2003) Polymerase chain reaction-based methods of DNA methylation analysis. *Anal. Biochem.*, **317**, 259–265.
15. Hou, P., Ji, M., Li, S., He, N. and Lu, Z. (2004) High-throughput method for detecting DNA methylation. *J. Biochem. Biophys. Methods*, **60**, 139–150.
16. Gitan, R.S., Shi, H., Chen, C.M., Yan, P.S. and Huang, T.H. (2002) Methylation-specific oligonucleotide microarray: a new potential for high-throughput methylation analysis. *Genome Res.*, **12**, 158–164.
17. Tost, J., Schatz, P., Schuster, M., Berlin, K. and Gut, I.G. (2003) Analysis and accurate quantification of CpG methylation by MALDI mass spectrometry. *Nucleic Acids Res.*, **31**, e50.
18. Dupont, J.M., Tost, J., Jammes, H. and Gut, I.G. (2004) *De novo* quantitative bisulphite sequencing using the pyrosequencing technology. *Anal. Biochem.*, **333**, 119–127.
19. Thomassin, H., Kress, C. and Grange, T. (2004) MethylQuant: a sensitive method for quantifying methylation of specific cytosines within the genome. *Nucleic Acids Res.*, **32**, e168.
20. Lewin, J., Schmitt, A.O., Adorjan, P., Hildmann, T. and Piepenbrock, C. (2004) Quantitative DNA methylation analysis based on four-dye trace data from direct sequencing of PCR amplificates. *Bioinformatics*, **20**, 3005–3012.
21. Adorjan, P., Distler, J., Lipscher, E., Model, F., Muller, J., Pelet, C., Braun, A., Florl, A.R., Gutig, D., Grabs, G. *et al.* (2002) Tumour class prediction and discovery by microarray-based DNA methylation analysis. *Nucleic Acids Res.*, **30**, e21.
22. Cross, S.H., Clark, V.H. and Bird, A.P. (1999) Isolation of CpG islands from large genomic clones. *Nucleic Acids Res.*, **27**, 2099–2107.
23. Hatada, I., Kato, A., Morita, S., Obata, Y., Nagaoka, K., Sakurada, A., Sato, M., Horii, A., Tsujimoto, A. and Matsubara, K. (2002) A microarray-based method for detecting methylated loci. *J. Hum. Genet.*, **47**, 448–451.
24. Huang, T.H., Perry, M.R. and Laux, D.E. (1999) Methylation profiling of CpG islands in human breast cancer cells. *Hum. Mol. Genet.*, **8**, 459–470.
25. Bedell, J.A., Budiman, M.A., Nunberg, A., Citek, R.W., Robbins, D., Jones, J., Flick, E., Rholfing, T., Fries, J., Bradford, K. *et al.* (2005) Sorghum genome sequencing by methylation filtration. *PLoS Biol.*, **3**, e13.
26. Rabinowicz, P.D. (2003) Constructing gene-enriched plant genomic libraries using methylation filtration technology. *Methods Mol. Biol.*, **236**, 21–36.
27. Yamada, Y., Watanabe, H., Miura, F., Soejima, H., Uchiyama, M., Iwasaka, T., Mukai, T., Sakaki, Y. and Ito, T. (2004) A comprehensive analysis of allelic methylation status of CpG islands on human chromosome 21q. *Genome Res.*, **14**, 247–266.
28. Monk, D., Smith, R., Arnaud, P., Preece, M.A., Stanier, P., Beechey, C.V., Peters, J., Kelsey, G. and Moore, G.E. (2003) Imprinted methylation profiles for proximal mouse chromosomes 11 and 7 as revealed by methylation-sensitive representational difference analysis. *Mamm. Genome*, **14**, 805–816.
29. Smith, R.J. and Kelsey, G. (2001) Identification of imprinted loci by methylation: use of methylation-sensitive representational difference analysis (Me-RDA). *Methods Mol. Biol.*, **181**, 113–132.
30. Smith, R.J., Dean, W., Konfortova, G. and Kelsey, G. (2003) Identification of novel imprinted genes in a genome-wide screen for maternal methylation. *Genome Res.*, **13**, 558–569.
31. Matsuyama, T., Kimura, M.T., Koike, K., Abe, T., Nakano, T., Asami, T., Ebisuzaki, T., Held, W.A., Yoshida, S. and Nagase, H. (2003) Global methylation screening in the *Arabidopsis thaliana* and *Mus musculus* genome: applications of virtual image restriction landmark genomic scanning (Vi-RLGS). *Nucleic Acids Res.*, **31**, 4490–4496.
32. Rush, L.J. and Plass, C. (2002) Restriction landmark genomic scanning for DNA methylation in cancer: past, present, and future applications. *Anal. Biochem.*, **307**, 191–201.
33. Costello, J.F., Fruhwald, M.C., Smiraglia, D.J., Rush, L.J., Robertson, G.P., Gao, X., Wright, F.A., Feramisco, J.D., Peltomaki, P., Lang, J.C. *et al.* (2000) Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat. Genet.*, **24**, 132–138.
34. Crawford, G.E., Holt, I.E., Mullikin, J.C., Tai, D., Blakesley, R., Bouffard, G., Young, A., Masiello, C., Green, E.D., Wolfsberg, T.G. *et al.* (2004) Identifying gene regulatory elements by genome-wide recovery of DNase hypersensitive sites. *Proc. Natl Acad. Sci. USA*, **101**, 992–997.
35. Sabo, P., Humbert, R., Hawrylycz, M., Wallace, J., Dorschner, M., McArthur, M. and Stamatoyannopoulos, J. (2004) Genome-wide identification of DNaseI hypersensitive sites using active chromatin sequence libraries. *Proc. Natl Acad. Sci. USA*, **101**, 4537–4542.
36. Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N.P. and Bickmore, W.A. (2004) Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell*, **118**, 555–566.
37. White, E.J., Emanuelsson, O., Scalzo, D., Royce, T., Kosak, S., Oakeley, E.J., Weissman, S., Gerstein, M., Groudine, M., Snyder, M. *et al.* (2004) DNA replication-timing analysis of human chromosome 22 at high resolution and different developmental states. *Proc. Natl Acad. Sci. USA*, **101**, 17771–17776.
38. Woodfine, K., Fiegler, H., Beare, D.M., Collins, J.E., McCann, O.T., Young, B.D., Debernardi, S., Mott, R., Dunham, I. and Carter, N.P. (2004) Replication timing of the human genome. *Hum. Mol. Genet.*, **13**, 191–202.
39. Woodfine, K., Beare, D.M., Ichimura, K., Debernardi, S., Mungall, A.J., Fiegler, H., Collins, V.P. and Carter, N.P. (2005) Replication timing of human chromosome 6. *Cell Cycle*, **4**, e144.
40. Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E. *et al.* (2000) Genome-wide location and function of DNA binding proteins. *Science*, **290**, 2306–2309.
41. Widlak, P. (2004) DNA microarrays, a novel approach in studies of chromatin structure. *Acta Biochim. Pol.*, **51**, 1–8.
42. Hanlon, S.E. and Lieb, J.D. (2004) Progress and challenges in profiling the dynamics of chromatin and transcription factor binding with DNA microarrays. *Curr. Opin. Genet. Dev.*, **14**, 697–705.
43. Schubeler, D., MacAlpine, D.M., Scalzo, D., Wirbelauer, C., Kooperberg, C., van Leeuwen, F., Gottschling, D.E., O'Neill, L.P., Turner, B.M., Delrow, J. *et al.* (2004) The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes Dev.*, **18**, 1263–1271.
44. Matarazzo, M.R., Lembo, F., Angrisano, T., Ballestar, E., Ferraro, M., Pero, R., De Bonis, M.L., Bruni, C.B., Esteller, M., D'Esposito, M. *et al.* (2004) *In vivo* analysis of DNA methylation patterns recognized by specific proteins: coupling CHIP and bisulphite analysis. *Biotechniques*, **37**, 666–668.
45. Mukhopadhyay, R., Yu, W., Whitehead, J., Xu, J., Lezcano, M., Pack, S., Kanduri, C., Kanduri, M., Ginjala, V., Vostrov, A. *et al.* (2004) The binding sites for the chromatin insulator protein CTCF map to DNA methylation-free domains genome-wide. *Genome Res.*, **14**, 1594–1602.
46. Murrell, A., Heeson, S. and Reik, W. (2004) Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. *Nat. Genet.*, **36**, 889–893.
47. Horike, S., Cai, S., Miyano, M., Cheng, J.F. and Kohwi-Shigematsu, T. (2005) Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nat. Genet.*, **37**, 31–40.
48. Cremer, T. and Cremer, C. (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.*, **2**, 292–301.

49. Cremer, M., von Hase, J., Volm, T., Brero, A., Kreth, G., Walter, J., Fischer, C., Solovei, I., Cremer, C. and Cremer, T. (2001) Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells. *Chromosome Res.*, **9**, 541–567.
50. Osborne, C.S., Chakalova, L., Brown, K.E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J.A., Lopes, S., Reik, W. *et al.* (2004) Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat. Genet.*, **36**, 1065–1071.
51. Chambeyron, S. and Bickmore, W.A. (2004) Does looping and clustering in the nucleus regulate gene expression? *Curr. Opin. Cell Biol.*, **16**, 256–262.
52. Boyle, S., Gilchrist, S., Bridger, J.M., Mahy, N.L., Ellis, J.A. and Bickmore, W.A. (2001) The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum. Mol. Genet.*, **10**, 211–219.
53. Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M. *et al.* (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.*, **18**, 630–634.
54. Tuteja, R. and Tuteja, N. (2004) Serial analysis of gene expression (SAGE): unraveling the bioinformatics tools. *Bioessays*, **26**, 916–922.
55. Sandoval, J., Rodriguez, J.L., Tur, G., Serviddio, G., Pereda, J., Boukaba, A., Sastre, J., Torres, L., Franco, L. and Lopez-Rodas, G. (2004) RNAPol–ChIP: a novel application of chromatin immunoprecipitation to the analysis of real-time gene transcription. *Nucleic Acids Res.*, **32**, e88.
56. Wang, Z., Fan, H., Yang, H.H., Hu, Y., Buetow, K.H. and Lee, M.P. (2004) Comparative sequence analysis of imprinted genes between human and mouse to reveal imprinting signatures. *Genomics*, **83**, 395–401.
57. Weidman, J.R., Murphy, S.K., Nolan, C.M., Dietrich, F.S. and Jirtle, R.L. (2004) Phylogenetic footprint analysis of IGF2 in extant mammals. *Genome Res.*, **14**, 1726–1732.
58. Ringrose, L., Rehmsmeier, M., Dura, J. and Paro, R. (2003) Genome-wide prediction of polycomb/trithorax response elements in *Drosophila melanogaster*. *Dev. Cell*, **5**, 759–771.
59. Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., III, Gingeras, T.R. *et al.* (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*, **120**, 169–181.
60. Lippman, Z., Gendrel, A., Black, M., Vaughn, M., Dedhia, N., McCombie, W., Lavine, K., Mittal, V., May, B., Kasschau, K. *et al.* (2004) Role of transposable elements in heterochromatin and epigenetic control. *Nature*, **430**, 471–476.
61. Rakyan, V.K., Hildmann, T., Novik, K.L., Lewin, J., Tost, J., Cox, A.V., Andrews, T.D., Howe, K.L., Otto, T., Olek, A. *et al.* (2004) DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. *PLoS Biol.*, **2**, e405.
62. Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.*, **16**, 6–21.
63. Costello, J.F., Fruhwald, M.C., Smiraglia, D.J., Rush, L.J., Robertson, G.P., Gao, X., Wright, F.A., Feramisco, J.D., Peltomaki, P., Lang, J.C. *et al.* (2000) Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat. Genet.*, **24**, 132–138.
64. Kondo, T., Bobek, M.P., Kuick, R., Lamb, B., Zhu, X., Narayan, A., Bourc'his, D., Viegas-Pequignot, E., Ehrlich, M. and Hanash, S.M. (2000) Whole-genome methylation scan in ICF syndrome: hypomethylation of non-satellite DNA repeats D4Z4 and NBL2. *Hum. Mol. Genet.*, **9**, 597–604.
65. Esteller, M. and Herman, J.G. (2002) Cancer as an epigenetic disease: DNA methylation and chromatin alterations in human tumours. *J. Pathol.*, **196**, 1–7.
66. Issa, J.P. (2004) Opinion: CpG island methylator phenotype in cancer. *Nat. Rev. Cancer*, **4**, 988–993.
67. Egger, G., Liang, G., Aparicio, A. and Jones, P.A. (2004) Epigenetics in human disease and prospects for epigenetic therapy. *Nature*, **429**, 457–463.
68. Lo, H.S., Wang, Z., Hu, Y., Yang, H.H., Gere, S., Buetow, K.H. and Lee, M.P. (2003) Allelic variation in gene expression is common in the human genome. *Genome Res.*, **13**, 1855–1862.
69. Pastinen, T. and Hudson, T.J. (2004) Cis-acting regulatory variation in the human genome. *Science*, **306**, 647–650.
70. Rakyan, V.K., Blewitt, M.E., Druker, R., Preis, J.I. and Whitelaw, E. (2002) Metastable epialleles in mammals. *Trends Genet.*, **18**, 348–351.
71. Bjornsson, H.T., Fallin, M.D. and Feinberg, A.P. (2004) An integrated epigenetic and genetic approach to common human disease. *Trends Genet.*, **20**, 350–358.
72. Murrell, A., Heeson, S., Cooper, W.N., Douglas, E., Apostolidou, S., Moore, G.E., Maher, E.R. and Reik, W. (2004) An association between variants in the IGF2 gene and Beckwith-Wiedemann syndrome: interaction between genotype and epigenotype. *Hum. Mol. Genet.*, **13**, 247–255.