Automating whole-genome DNA methylation analyses

A step toward identifying subtle epigenetic variation in cancer and other complex disease

Lee M Butcher[†]*, Ignacio Mazon[‡], Jean-Jacques Goval[‡], Juana Magdalena[‡] & Stephan Beck[†]

[†]UCL Cancer Institute, University College London, London, WC1E 6BT, UK. [‡] Diagenode sa, Tour GIGA, 3ème étage, 1 Avenue de l'Hopital, B-4000, Liège, BELGIUM.

*email: l.butcher@ucl.ac.uk, tel: 020 7679 6004, web: www.ucl.ac.uk/cancer/research-groups/medical-genomics/index.htm

Background

Although a cell's actions are determined by the genetic code (the DNA sequence), the extent to which they are realised is **determined by epigenetics**.

DNA methylation is one epigenetic mark critical to development and it refers to the addition of $-CH_3$ group to cytosine (see fig1). Of particular clinical relevance, DNA methylation can be reversed using drugs, which makes disease-causing methylation profiles a priority for establishing drug targets.

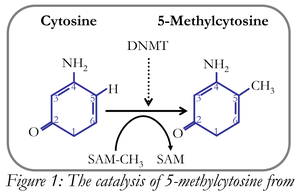


Figure 1: The catalysis of 5-methylcytosine from cytosine by DNA methyltransferase (DNMT)

In mammals it occurs at CG dinucleotides (CpGs), which cluster in 'islands' that most typically occur at gene promoters. However**, recent evidence**[1] suggests that disease-causing DNA methylation occurs not at gene promoters or CpG islands but in nearby **shores** – regions up to several kilobases up- or downstream of the promoter.

Current state of the art

Currently, most DNA methylation assays are either **not global enough**, or are **biased toward CpG islands**. Moreover, the effects of DNA methylation polygenic and pleiotropic, which loosely translates to **subtle!** Therefore forthcoming studies require surveys that encompass islands, shores (and more) using **an increasing number of individuals**. This requires a scalable assay on a truly genome-wide level.

The technique most suitable to tackling this problem is ***MeDIP-seq***[2]. Although it captures most of the genome, it is laborious and its sensitivity not is fully known.

Aims

The **aims** of this project are twofold:

1. Reduce labour through automation

2. Assess the limits of *MeDIP* for use in large-scale analyses

Automation of Immunoprecipitation techniques

Immunoprecipitation is technique that uses an antibody to capture your target of choice. **Methylated DNA immunoprecipitation (*MeDIP*)**[3] is a whole genome DNA methylation assay that enriches for the methylated fraction. It can be combined with microarrays (*MeDIP*-chip) or second-generation sequencing (*MeDIP*-seq) to generate methylome profiles (methylomes).

To automate the hands-on aspect of *MeDIP* ('*Auto MeDIP*' & '*Auto ChIP*'), **Diagenode** recently introduced a robotic liquid handler (see fig2). We are currently testing this machine using a series of custom assays with qPCR, microarrays and, ultimately, Illumina Solexa 2nd-generation sequencing.

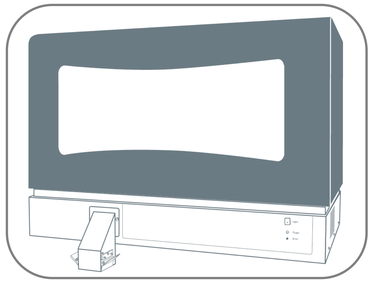
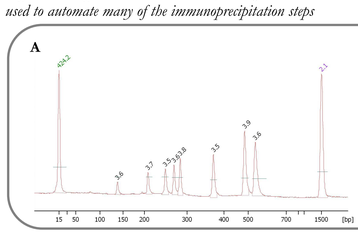


Figure 2: Diagenode SX-8G IP Star, the liquid handler used to automate many of the immunoprecipitation steps

Methods

***qPCR-**based assessments*

- We generated a **spike-cocktail** of 8 *in vitro* methylated dsDNA fragments
- Spikes were λ-DNA, against background of human gDNA
- Spiked-in across a molarity gradient ($1x10^6 \rightarrow 7x10^3$ molecules)



B

Size (bp)	No. CpGs	CpG density	Conc. (ng/µl)	Molarity (nM/ltr)
138	2	1.4%	0.33	3.6
209	1	0.5%	0.51	3.7
249	3	1.2%	0.58	3.5
270	5	1.9%	0.64	3.6
284	20	7.0%	0.7	3.8
371	10	2.7%	0.87	3.5
485	15	3.1%	1.25	3.9
536	20	3.7%	1.29	3.6

Figure 3: Electropheragram showing the 8 in vitro *methylated dsDNA fragments (A), and their properties (B). Following enrichment,* MeDIP *and INPUT fractions are qPCR'd to estimate recovery due to* MeDIP. *As a qualitative indicator of* MeDIP *success (quality control; QC), future substantive work employs a quality control assay involving PCR for an unmethylated version of the 284bp fragment and a methylated 536bp fragment.*

***Microarray**-based assessments*

- Methylated and unmethylated human gDNA mixed in 0%, 25%, 50%, 75% and 100% ratios
- Quality controlled using λ-DNA spike-ins (see fig3 legend)
- Hybridized to 2Kb-res microarray, tiling entire 4Mb MHC region on 6p

Results

The *Auto MeDIP* workflow **qualitatively distinguished** between methylated and unmethylated fragments (QC; see fig4).

Samples passing QC were then subject to qPCR to ascertain how many CpG sites are required per fragment (and in what quantity) for *Auto MeDIP* to detect them (see fig5).

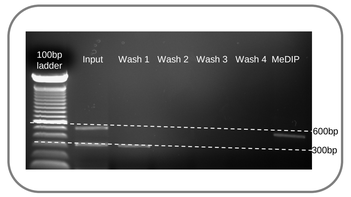


Figure 4: Example of QC. Input shows both methylated and unmethylated fragments. Most of unmethylated fragment from MeDIP *comes out in the first wash. MeDIP selectively enriches for the methylated fragment.*

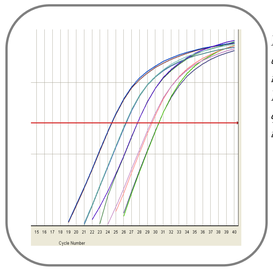


Figure 5: Anticipated qPCR results showing clear separation of recovery for decreasing (right to left) amounts of a single methylated region. Input (not shown) is also run in a typical qPCR, which allows recovery of what was input to be quantified.

We chose to extend our analyses to microarrays as a step toward understanding the biases of *MeDIP* on a **quasi-genome-wide scale**.

MeDIP and Input samples were differentially labelled and co-hybridized to MHC microarrays. Following scanning, log2 enrichment ratios were transformed into absolute methylation estimates using the Batman algorithm[4]. We expect to see a linear response across all loci on the microarray (see fig6).

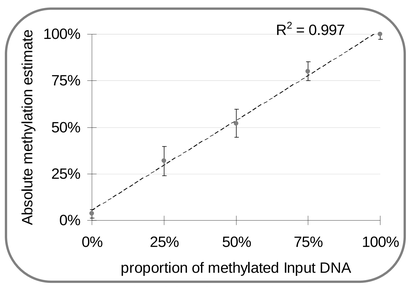


Figure 6:Hypothetical data illustrating observed versus mean methylation scores for loci on a 4Mb custom array

Future work

We aim test the *Auto MeDIP* workflow with Illumina Solexa sequencing using both control DNA and, as a proof-of-principle, experimental samples. These experiments will help fine-tune the methylation-scoring algorithms for use in future studies.

By optimising *Auto MeDIP* – and understanding its limits – we will have the frontend of a high-throughput DNA methylation analysis pipeline in place. This will then facilitate fast, accurate and reliable generation of methylomes, which will hopefully reveal novel methylation events that can lead to diagnostic and therapeutic advances.

If you are interested in future collaboration, do not hesitate to contact myself or Stephan Beck.

Collaborators: diagenode

Funding: wellcome trust

Footnotes and references:

[1] R.A. Irizarry *et al. Nat. Genet.* **41**, 178-186 (2009)

[2] A step-by-step animation of the *MeDIP* workflow is on my webpage

[3] M. Weber *et al. Nat. Genet.* **39**, 457-466 (2007)

[4] T. Down *et al. Nat. Biotechnol.* **26**, 779-785 (2008)