Automating high-throughput methylomes

Lee M Butcher

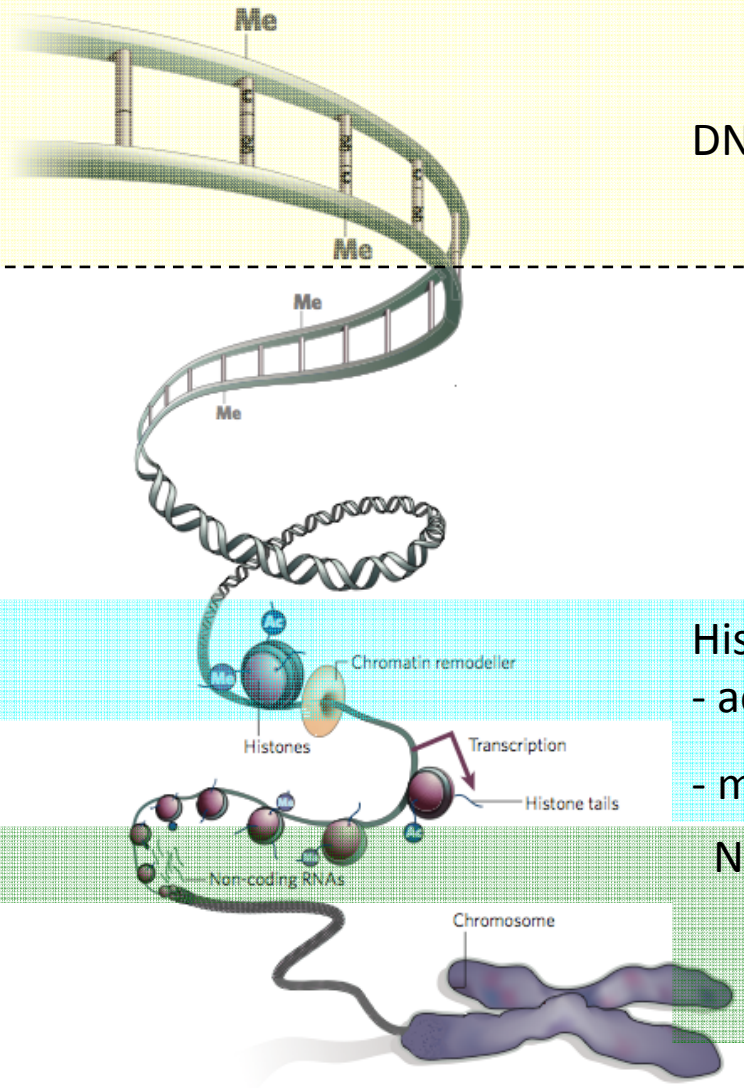
Postdoctoral Associate

UCL Cancer Institute

Introduction

Name:	Lee M Butcher
Position:	Postdoc
Team:	Medical Genomics
Supervisor:	Stephan Beck
Lab interests:	High-throughput technology
	Epigenetics/epigenomics
	Human complex disease/traits
	Mouse development

Epigenetics



DNA methylation

Histone modifications, e.g.,
- acetylation
- methylation

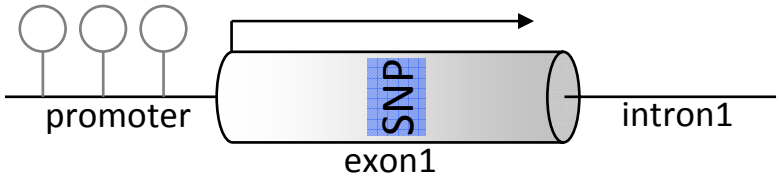
Non-coding RNA (ncRNA)

&

Micro RNA (miRNA)

Genomics and epigenomics

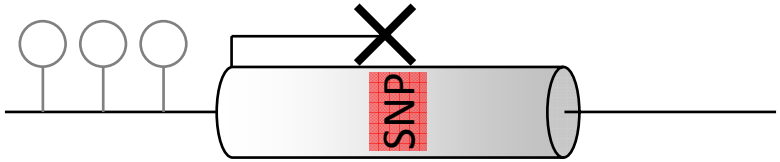
Unmethylated promoter



Transcription /

Healthy phenotype

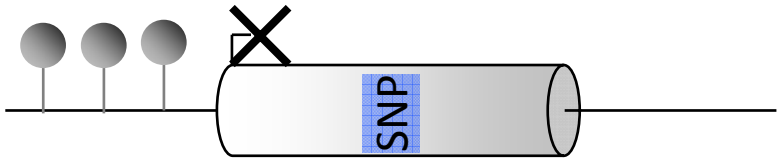
Unmethylated promoter



altered transcription /

Healthy-ish phenotype

Methylated promoter



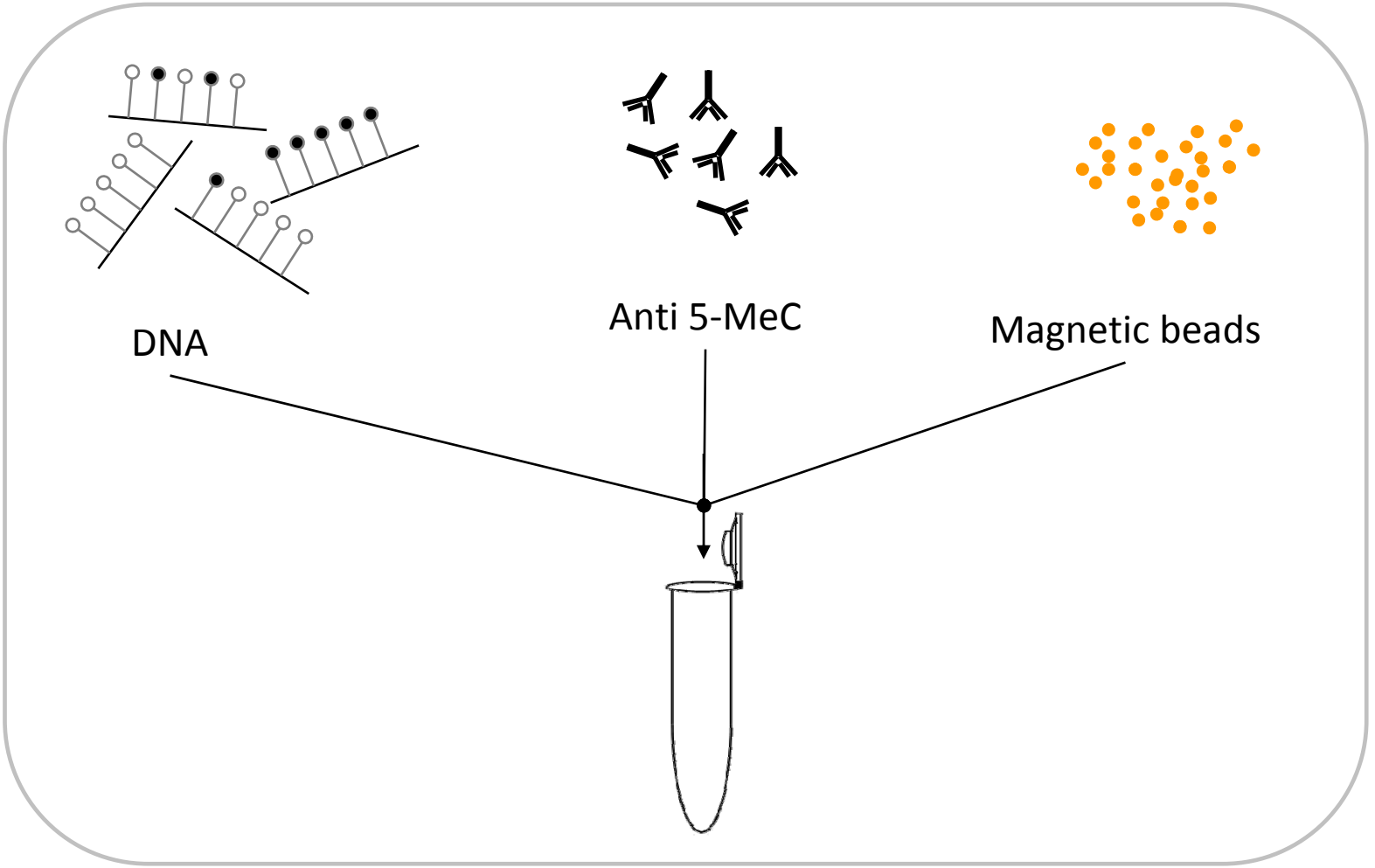
NO transcription /

Disease phenotype

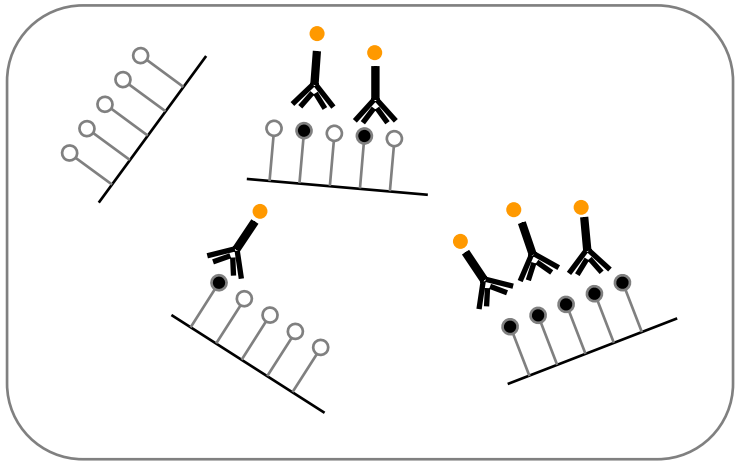
Challenges for epigenomic research

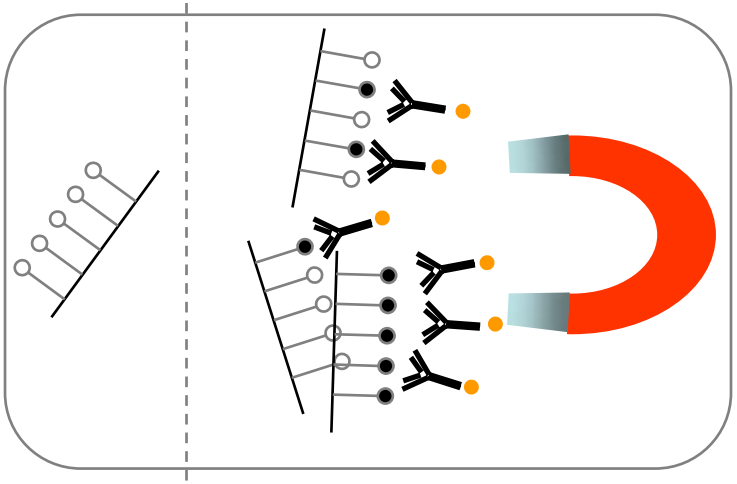
- Complex disease is influenced by multiple genes
  - Disease loci scattered at [largely] unknown genomic positions
  - Need an epigenetic assay that is:
    - Accurate
    - Affordable
    - High-throughput
    - Whole genome
- Assay of choice:
  - Methylated DNA immunoprecipitation (MeDIP; Weber et al. [2005] NatGen)
  - Decent balance between:
    - Throughput
    - Genomic coverage
    - Cost

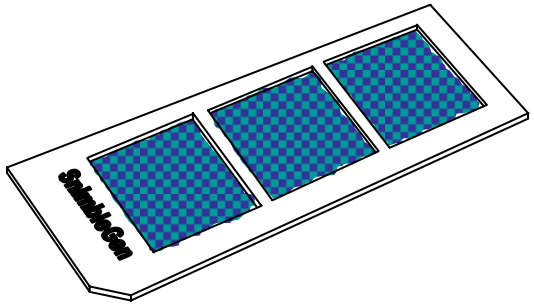
MeDIP assay – how it works



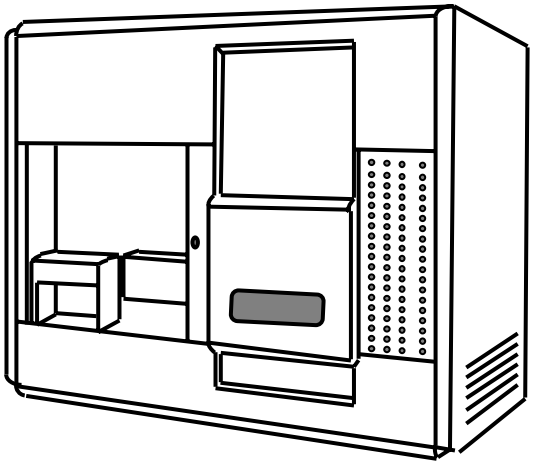
MeDIP assay – how it works







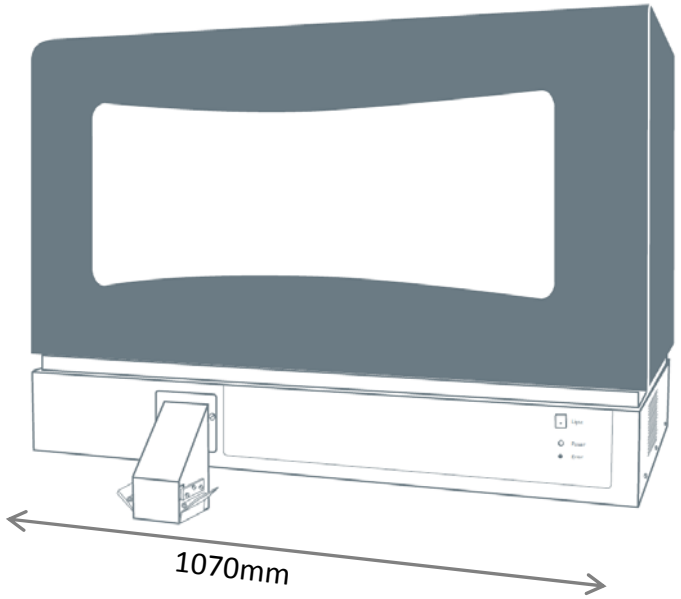
{MeDIP-Chip}



{MeDIP-Seq}

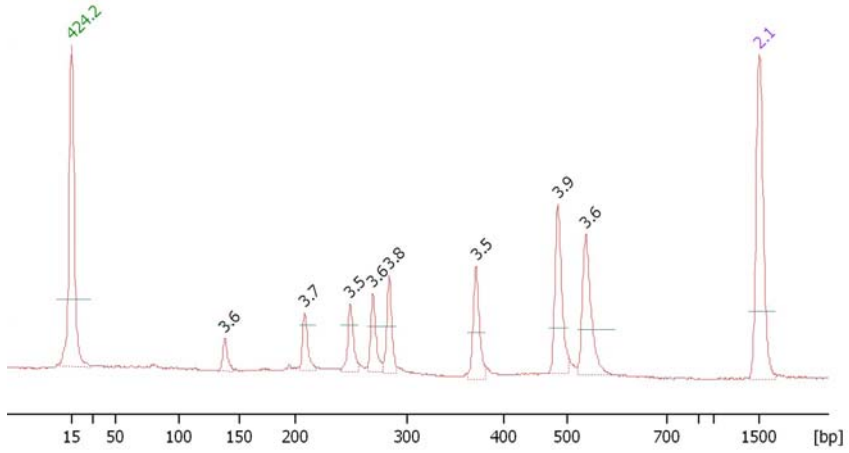
AutoMeDIP

- Diagenode
  - Automates up to 16 samples per run
    - Up to 24 samples every 24 hours (5 hour IP)
  - Hands on time saving = 75%
- Performance measured using:
  - *In vitro* methylated spike
  - qPCR



Spike-in made from λ-DNA

A

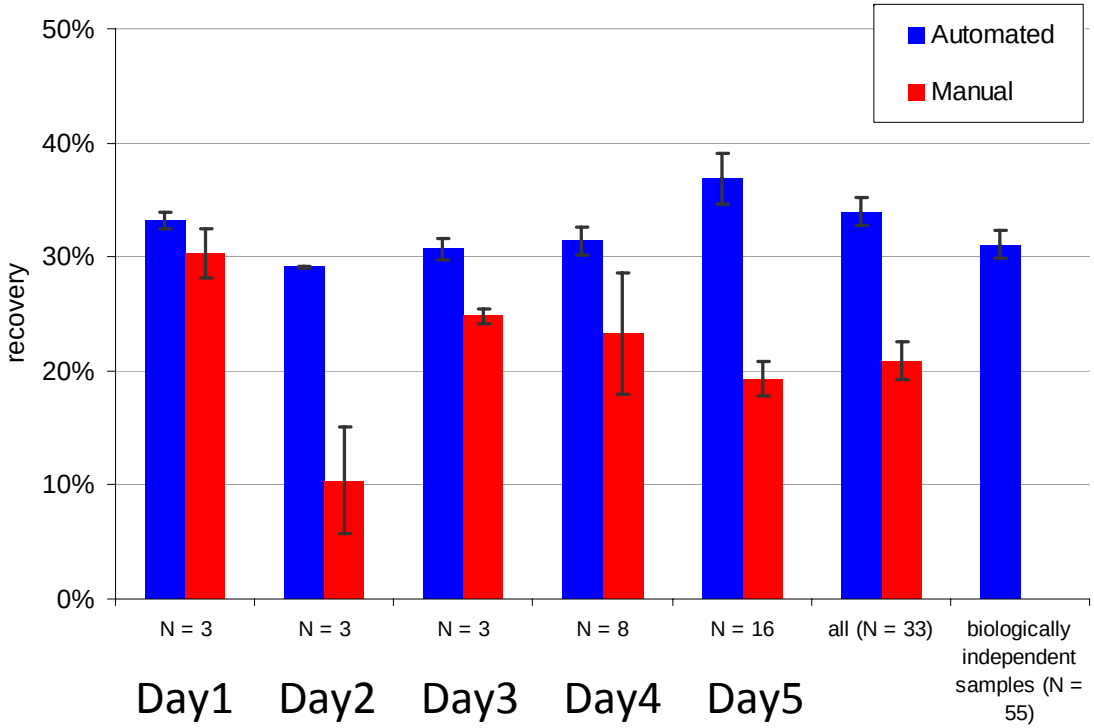


B

Size (bp)	No. CpGs	CpG density	Conc. (ng/µl)	Molarity (nM/ltr)
138	2	1.4%	0.33	3.6
197	1	0.5%	0.51	3.7
226	3	1.3%	0.58	3.5
270	5	1.9%	0.64	3.6
284	20	7.0%	0.7	3.8
360	10	2.8%	0.87	3.5
462	15	3.2%	1.25	3.9
400	20	5.0%	1.29	3.6

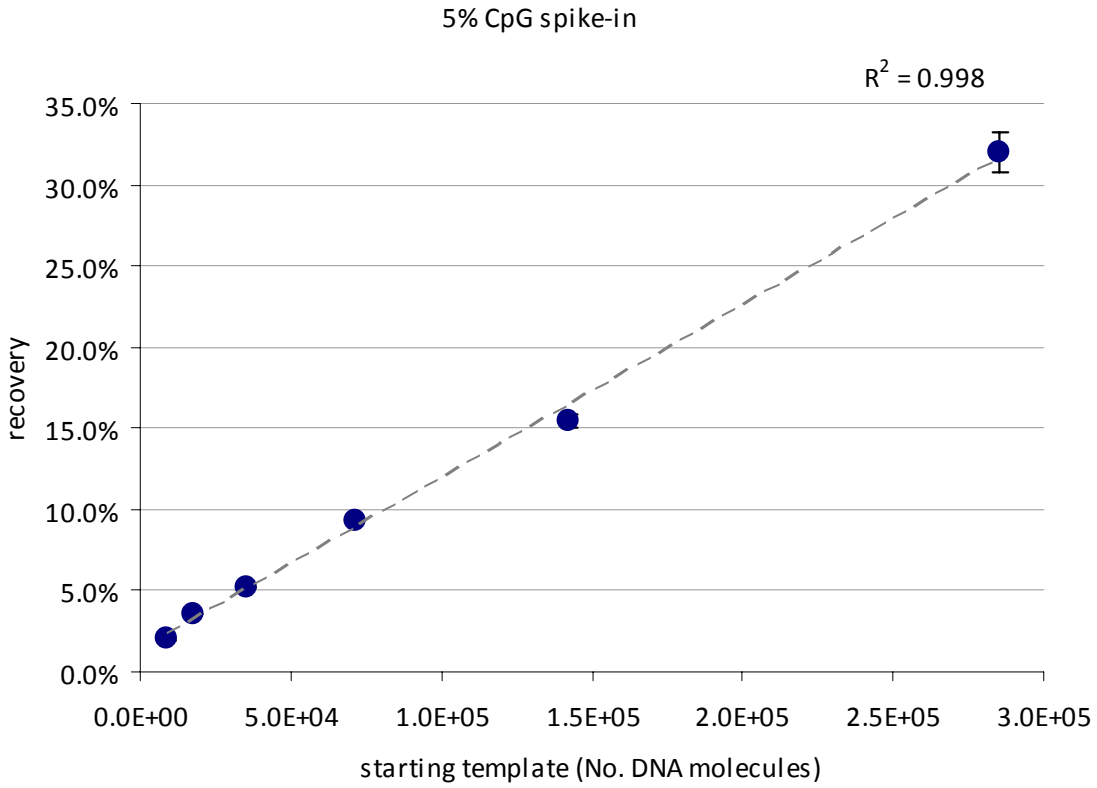
Results

Results – reproducibility



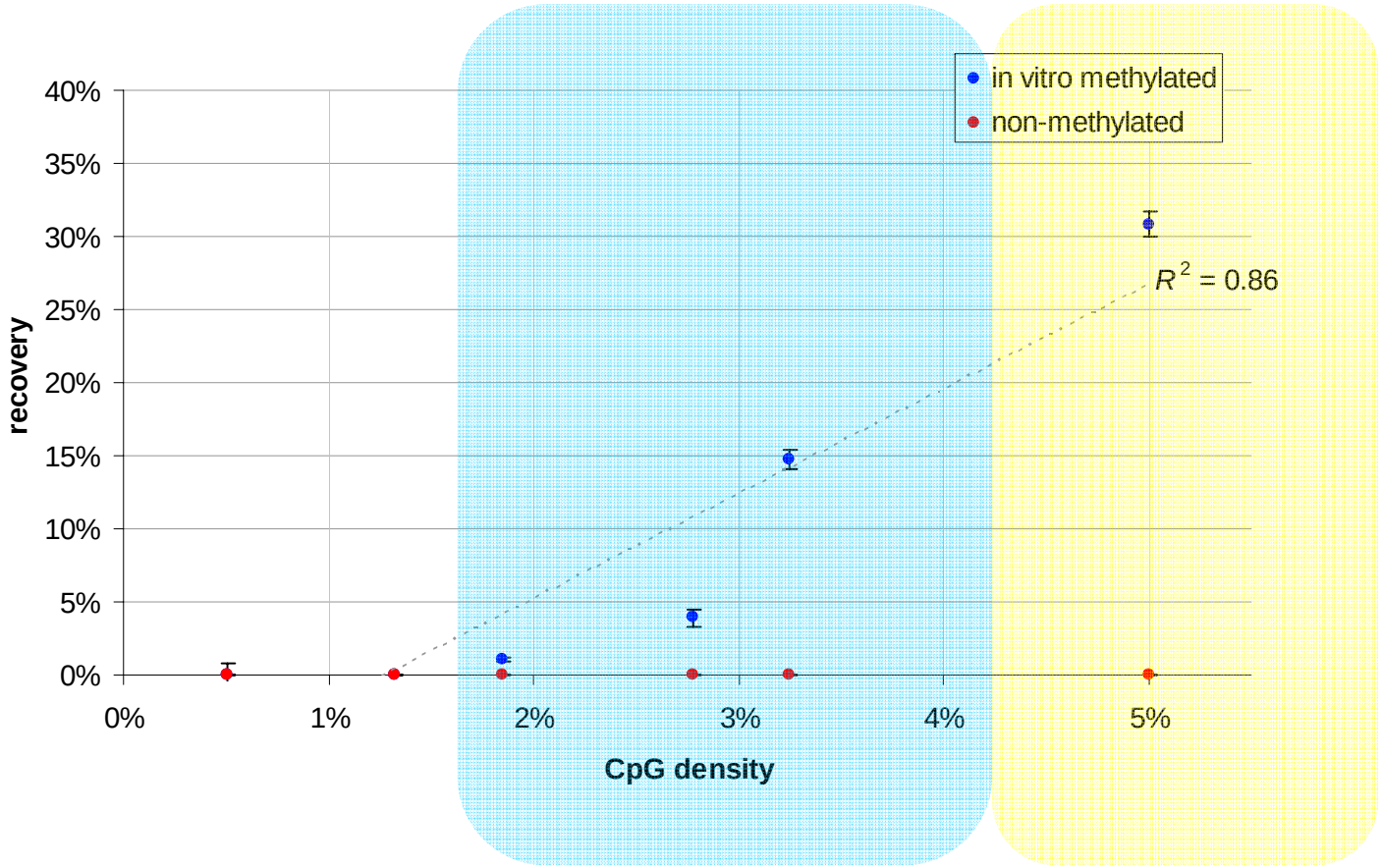
Butcher & Beck (2010) *Methods*, Apr 10

Results – sensitivity



Butcher & Beck (2010) *Methods*, Apr 10

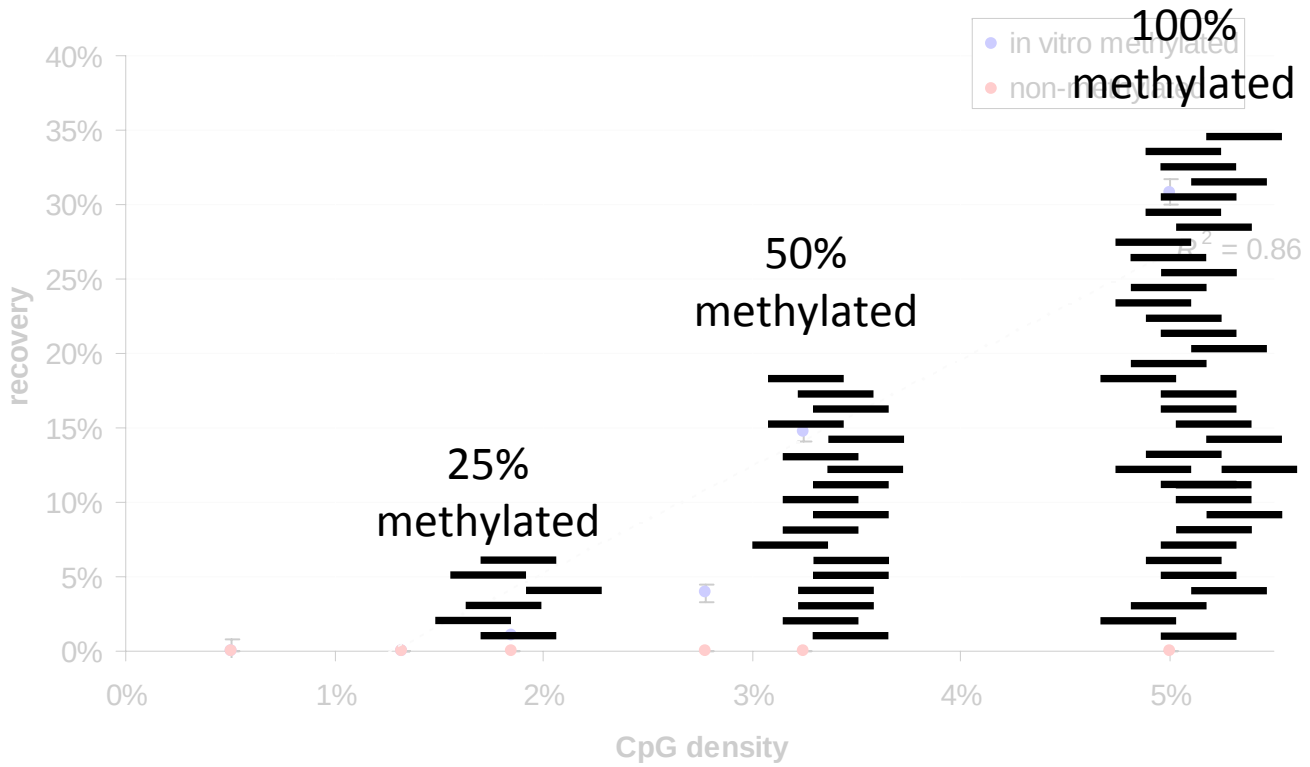
Results – specificity



CpG island shores CpG islands

Butcher & Beck (2010) *Methods*, Apr 10

Results – implications on alignment



Bioinformatic challenges…

- Enrichment bias means absolute methylation levels are difficult to quantitate
- Fortunately, help is at hand:

NATURE BIOTECHNOLOGY VOLUME 26 NUMBER 7 JULY 2008 779

A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis

Thomas A Down[1,8], Vardhman K Rakyan[2,8], Daniel J Turner[3], Paul Flicek[4], Heng Li[3], Eugene Kulesha[4], Stefan Gräf[4], Nathan Johnson[4], Javier Herrero[4], Eleni M Tomazou[3], Natalie P Thorne[5], Liselotte Bäckdahl[6], Marlis Herberth[7], Kevin L Howe[5], David K Jackson[3], Marcos M Miretti[3], John C Marioni[5], Ewan Birney[4], Tim J P Hubbard[3], Richard Durbin[3], Simon Tavaré[5] & Stephan Beck[6]

- Unfortunately…

$$f(A \mid m) = \prod_p G\left(A_p \mid A_{base} + r \sum_c C_{cp} m_c, v^{-1} \right)$$

MeDIP and non-CG methylation

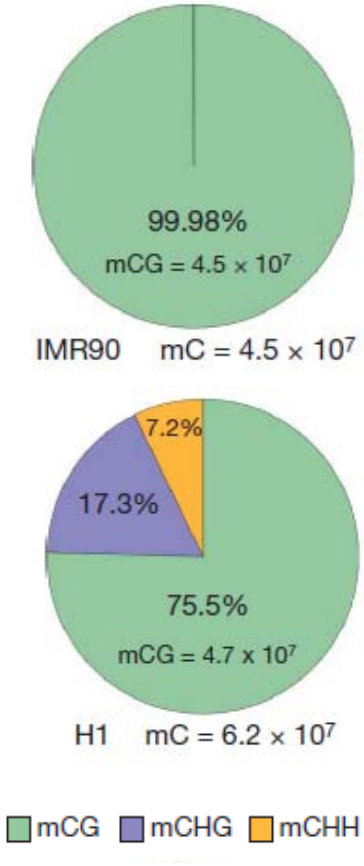
Vol 462 | 19 November 2009 | doi:10.1038/nature08514 nature

ARTICLES

Human DNA methylomes at base resolution show widespread epigenomic differences

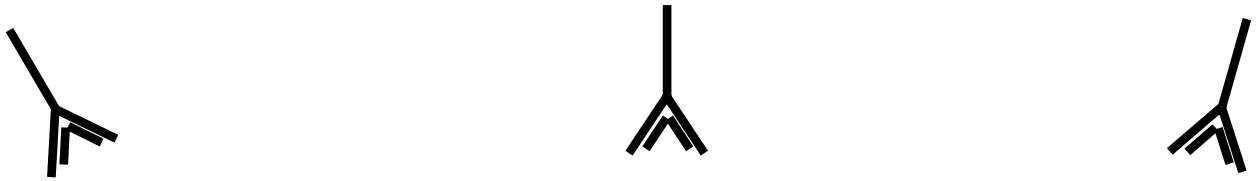
Ryan Lister[1]*, Mattia Pelizzola[1]*, Robert H. Dowen[1], R. David Hawkins[2], Gary Hon[2], Julian Tonti-Filippini[4], Joseph R. Nery[1], Leonard Lee[2], Zhen Ye[2], Que-Minh Ngo[2], Lee Edsall[2], Jessica Antosiewicz-Bourget[5,6], Ron Stewart[5,6], Victor Ruotti[5,6], A. Harvey Millar[4], James A. Thomson[5,6,7,8], Bing Ren[2,3] & Joseph R. Ecker[1]

DNA cytosine methylation is a central epigenetic modification that has essential roles in cellular processes including genome regulation, development and disease. Here we present the first genome-wide, single-base-resolution maps of methylated cytosines in a mammalian genome, from both human embryonic stem cells and fetal fibroblasts, along with comparative analysis of messenger RNA and small RNA components of the transcriptome, several histone modifications, and sites of DNA–protein interaction for several key regulatory factors. Widespread differences were identified in the composition and patterning of cytosine methylation between the two genomes. Nearly one-quarter of all methylation identified in embryonic stem cells was in a non-CG context, suggesting that embryonic stem cells may use different methylation mechanisms to affect gene regulation. Methylation in non-CG contexts showed enrichment in gene bodies and depletion in protein binding sites and enhancers. Non-CG methylation disappeared upon induced differentiation of the embryonic stem cells, and was restored in induced pluripotent stem cells. We identified hundreds of differentially methylated regions proximal to genes involved in pluripotency and differentiation, and widespread reduced methylation levels in fibroblasts associated with lower transcriptional activity. These reference epigenomes provide a foundation for future studies exploring this key epigenetic modification in human disease and development.



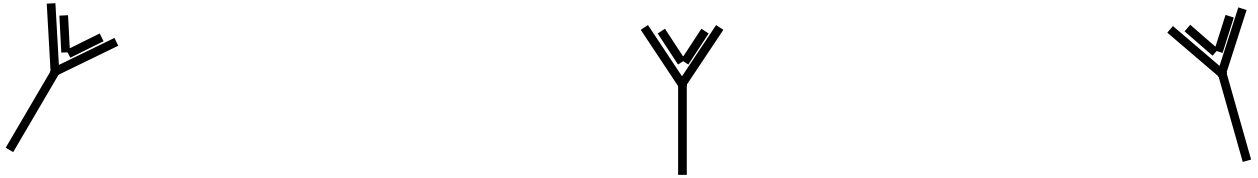
MeDIP and non-CG methylation

The CG context



5′ GCTTCTCTGGAGTGCGACAGGTTTGATGACAAAAAATTAGCGCAAGAAGACAAAAATCACCTTGCGCTAATGCT 3′

3′ CGAAGAGACCTCACGCTGTCCAAACTACTGTTTTTTAATCGCGTTCTTCTGTTTTTAGTGGAACGCGATTCCGA 5′



MeDIP and non-CG methylation

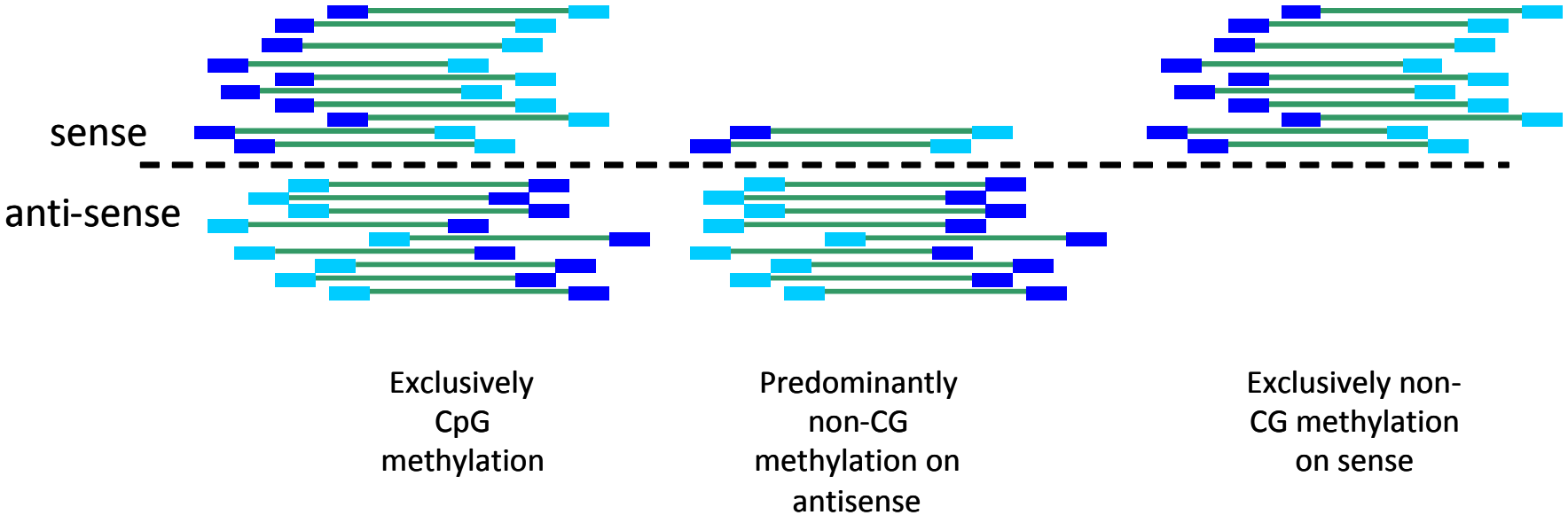
The non-CG context

5′ GCTTCTCTGGAGTGCGACAGGTTTGATGACAAAAAATTAGCGCAAGAAGACAAAAATCACCTTGCGCTAATGCT 3′

3′ CGAAGAGACCTCACGCTGTCCAAACTACTGTTTTTTAATCGCGTTCTTCTGTTTTTAGTGGAACGCGATTCCGA 5′

MeDIP and non-CG methylation

- Non-CG methylation will result in strand bias during sequencing



Applications of MeDIP

- We've collaborated on:
  - Two "large-scale" projects (*N*=37, *N*=60)
  - Four smaller-scale projects (*N*=3, *N*=6, *N*=7, *N*=12)
- Allocate 1 lane of flow cell (Illumina GAIIx)
  - Generate 40M reads
  - Average genome-wide CpG depth ≈ 1.2-1.5x
  - Average CpG depth covered by reads ≈ 3.3-4.4x
  - No. CpGs covered by reads ≈ 7.1M-10.9M (31.8% - 48.8%)

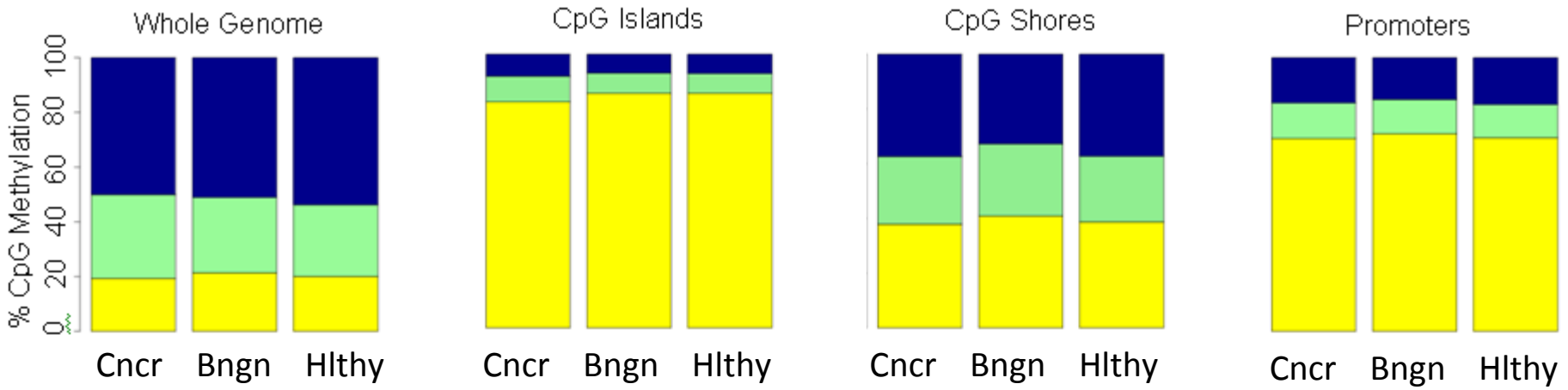
A cancer methylome

- Aim:
  - Define the methylome associated with a malignant phenotype
  - MeDIP-Seq to identify tumour specific differential methylation
  - Correlation with clinical end points
- Samples:
  - Pools of ten cases per sample cohort
    - Malignant Peripheral Nerve Sheath Tumors (MPNST) ***– Cancer***
    - Benign plexiform neurofibromas ***– Benign***
    - Normal Schwann Cell ***– Healthy***

Feber et al (2010) *Submitted*

Global changes in methylation

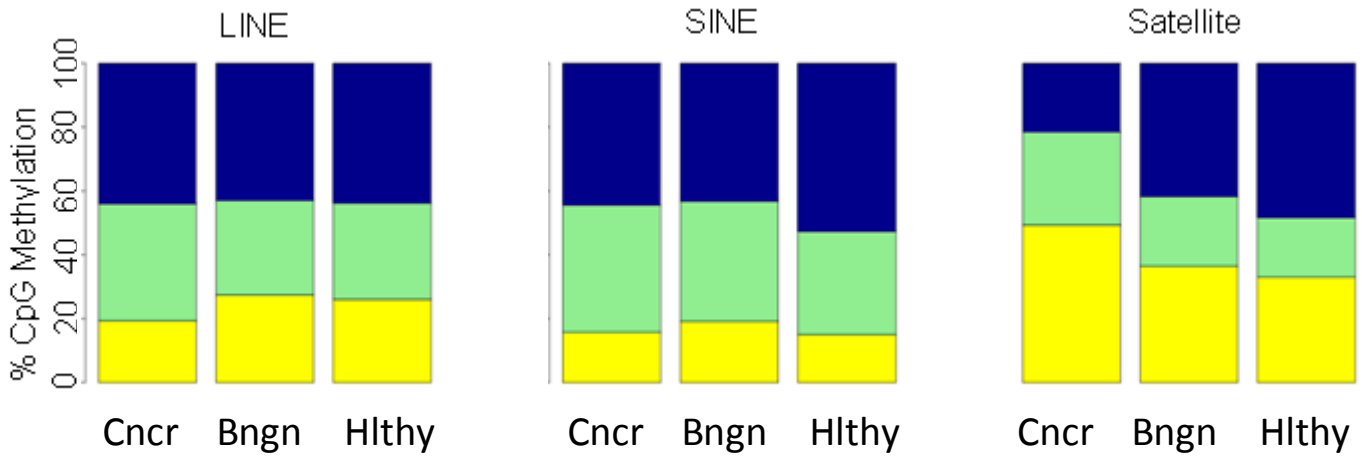
- Methylation state of each CpG site binned
  - Low (<40%), intermediate (40-60%) & high (>60%)
- Studies of other tumour types
  - hallmark of a cancer genome = global hypomethylation



Feber et al (2010) *Submitted*

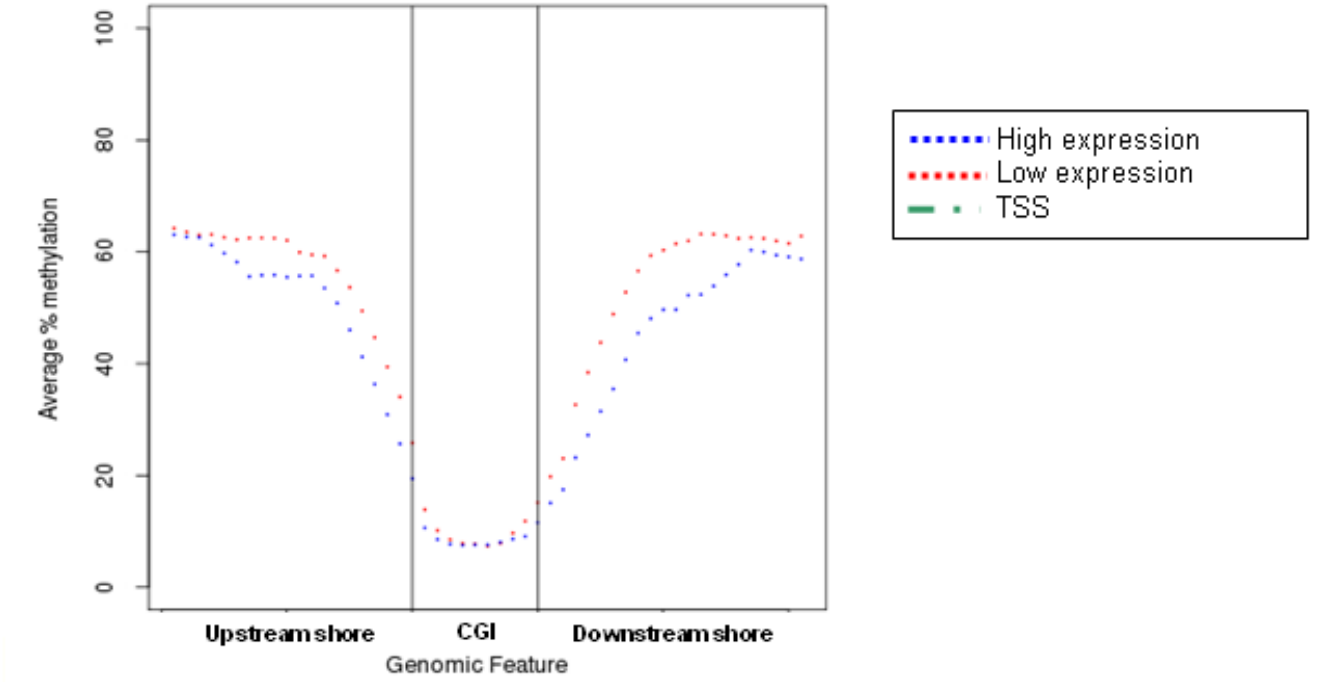
Global repeat methylation

- Methylation over LINE and SINE repeats change slightly
- Interestingly, LINE repeats appear to lose hypomethylated CpGs
- Satellite repeats show the largest change in global methylation



Feber et al (2010) *Submitted*

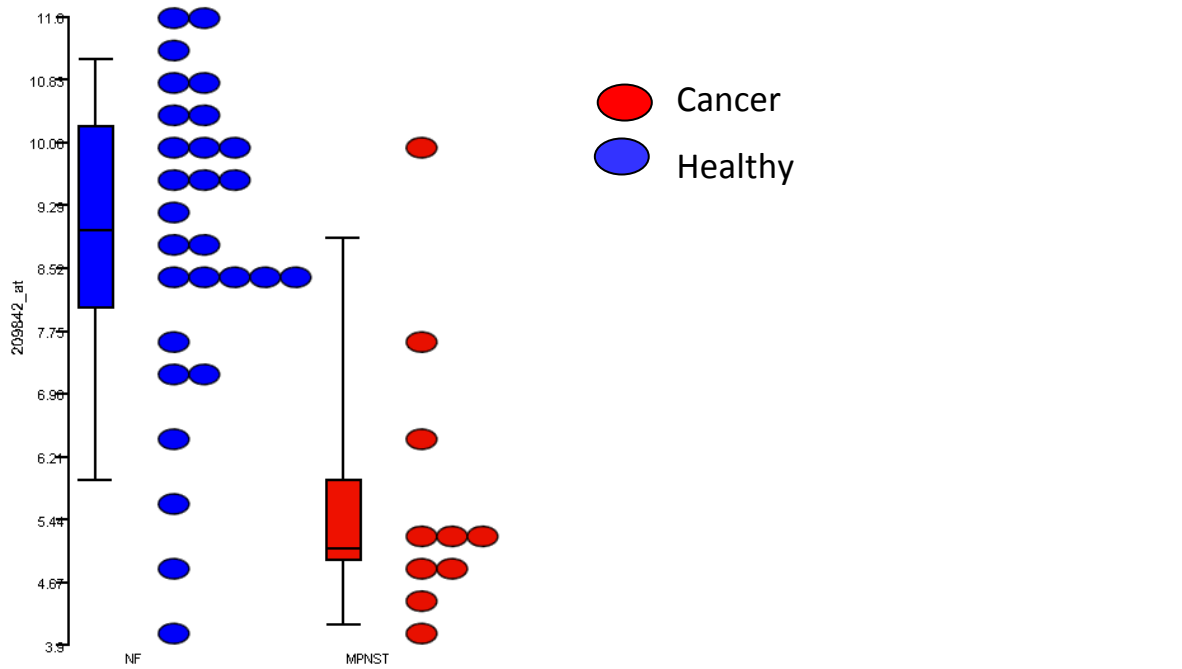
Effect of methylation at CGIs and CGI shores on gene expression

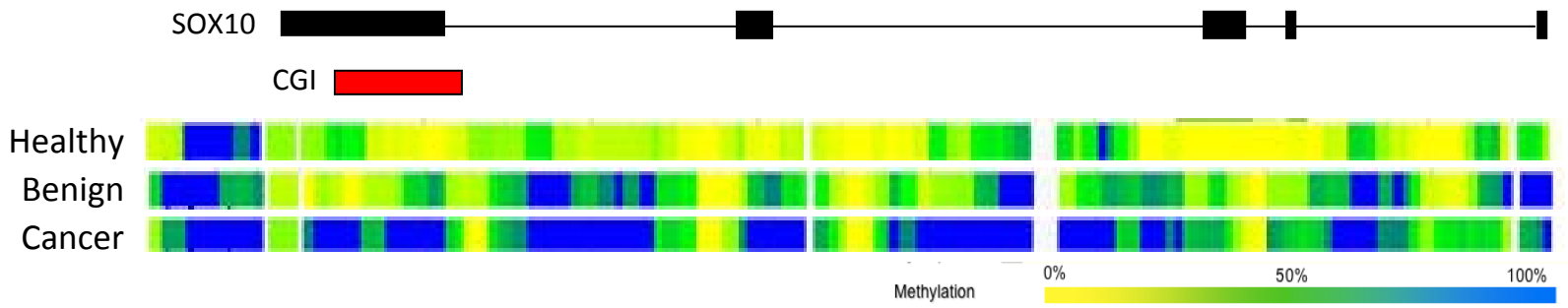


Feber et al (2010) *Submitted*

SOX10

Feber et al (2010) *Submitted*





Acknowledgements

Lab:

Funding: wellcome trust

Diagenode, BE:

Jean-Jacques Govel
Ignacio Mazon

Beijing Genomics Institute, CN:

Minghzi Ye
Li Ning
Qiuqing Zhang
Jun Wang

Babraham Institute, UK:

Steffi Seisenberger

Questions?

l.butcher@ucl.ac.uk

http://www.ucl.ac.uk/cancer/research-groups/medical-genomics/index.htm