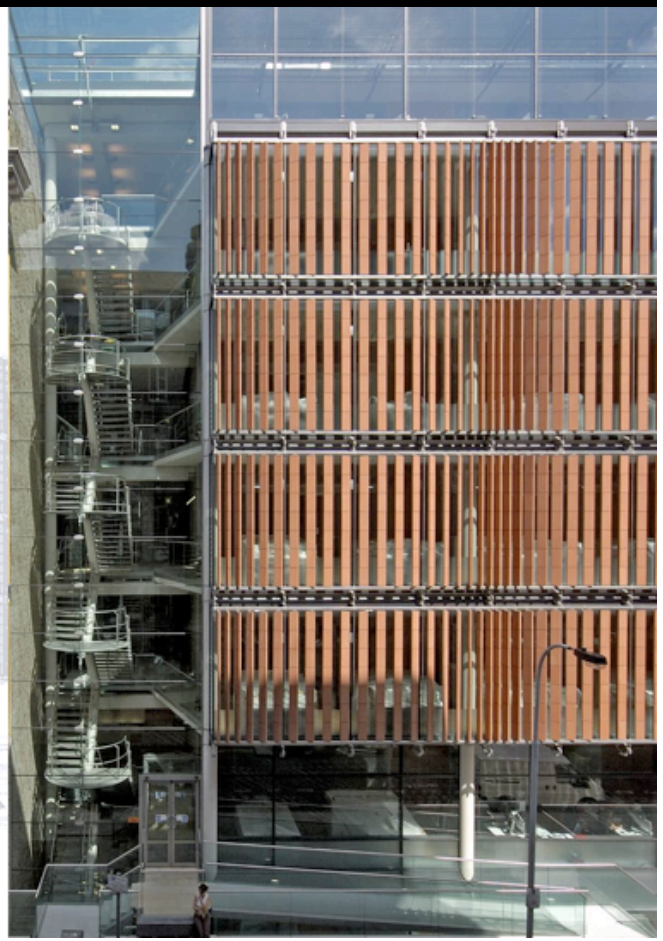
Reverse Phenotyping

Towards an integrated (epi)genomic approach to complex phenotypes and common disease

next-gen GWAS

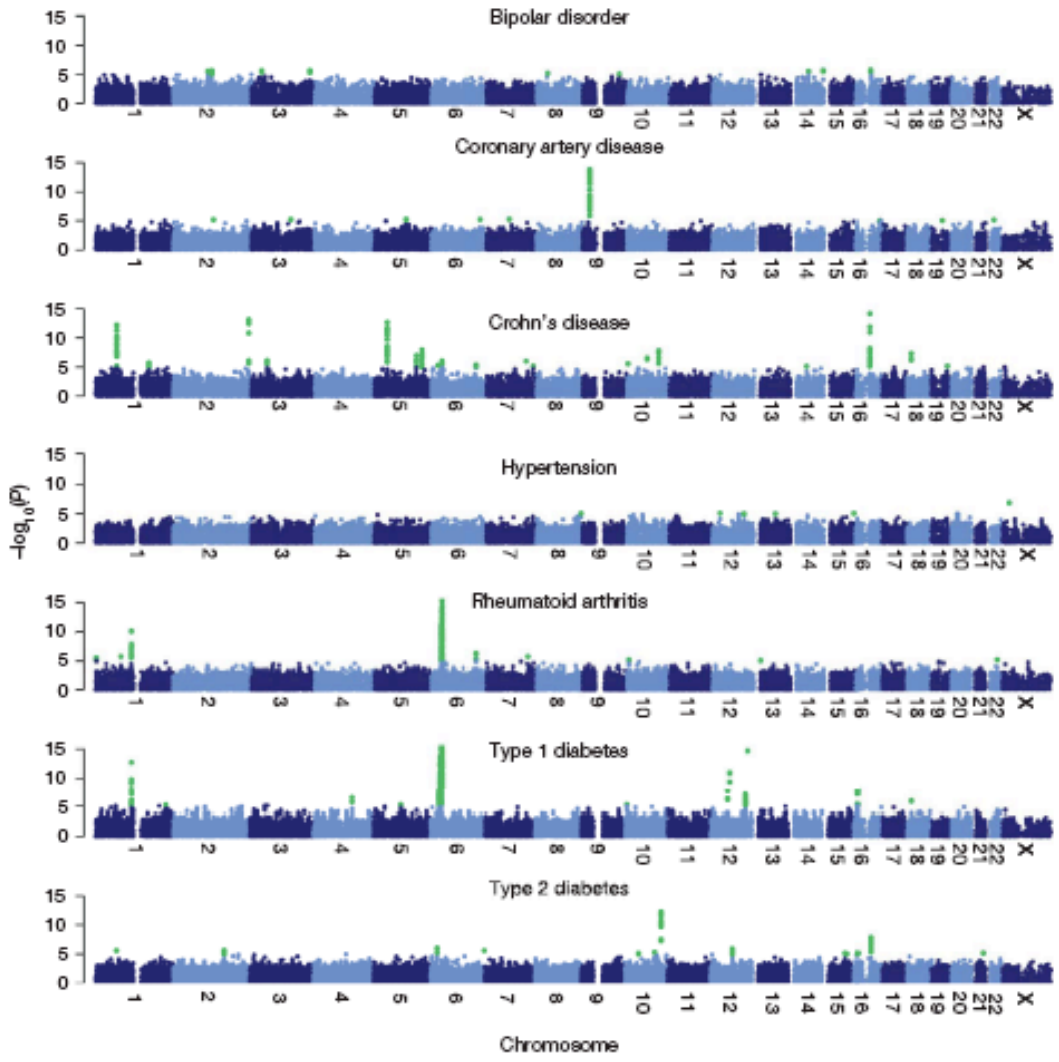
Stephan Beck
Medical Genomics
UCL Cancer Institute
University College London
s.beck@ucl.ac.uk

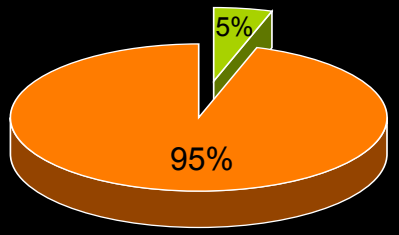


Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls

The Wellcome Trust Case Control Consortium*

NATURE|Vol 447|7 June 2007



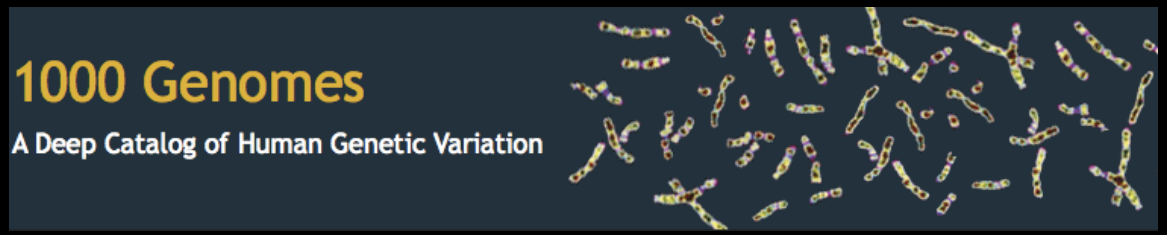


Effect Size

The case of the missing heritability

When scientists opened up the human genome, they expected to find the genetic components of common traits and diseases. But they were nowhere to be seen. **Brendan Maher** shines a light on six places where the missing loot could be stashed away.

rare variation



Genomics shifts focus to rare diseases

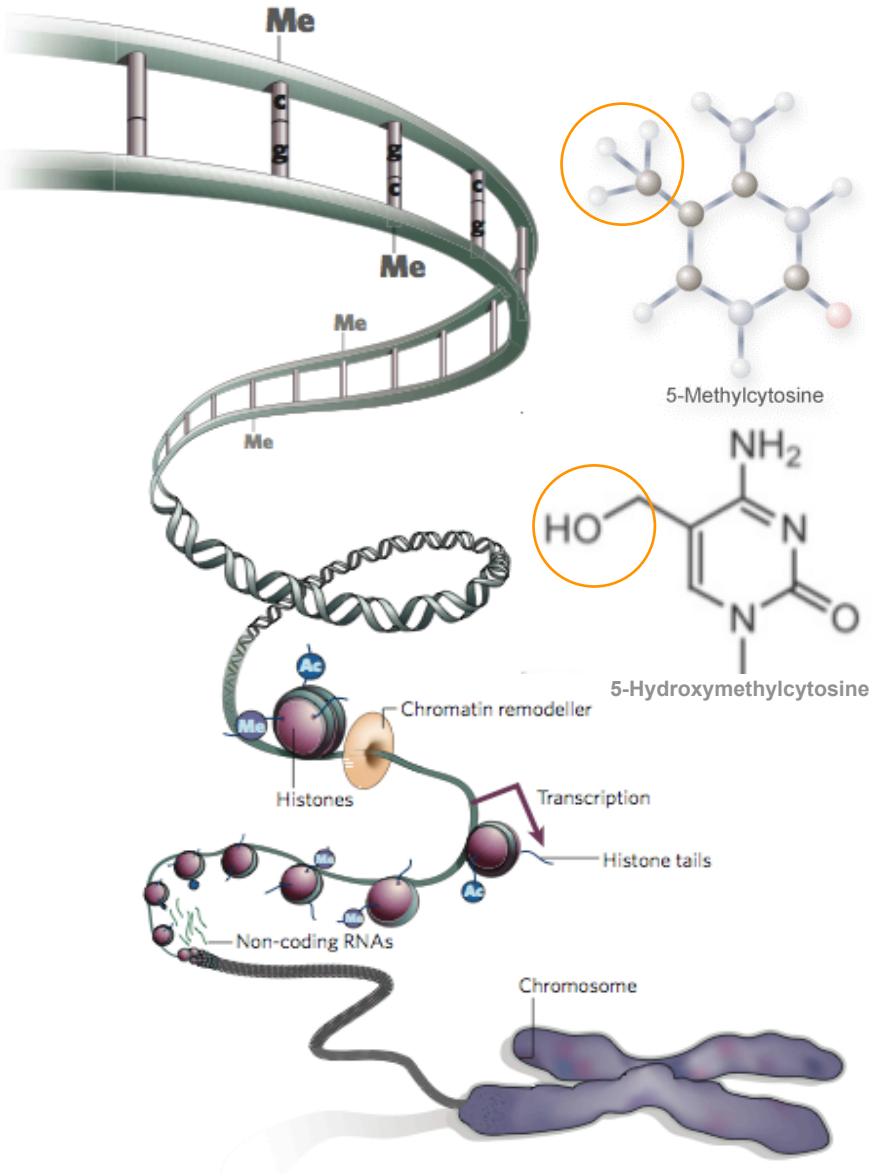
COLD SPRING HARBOR, NEW YORK

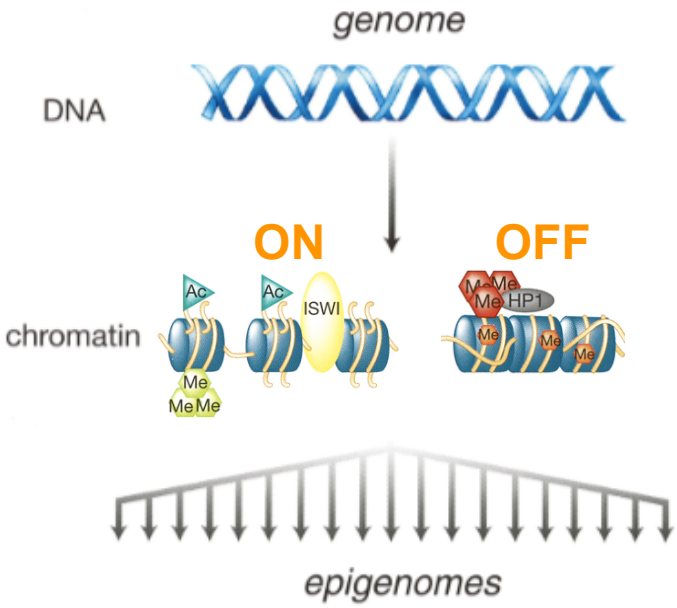
Genome sequencing may finally be living up to its promise of pinpointing genetic mutations that bear on treatment for individual patients. But the breakthroughs are not coming from the DNA analysis of common diseases with complex genetic origins, which has been the obsession of genomics for nearly the past decade. Instead, many genome scientists are turning back to study rare disorders that are traceable to defects in single genes, and whose causes have remained a mystery.

The change is partly a result of frustration with the disappointing results of genome-wide association studies (GWAS). Rather than sequencing whole genomes, GWAS studies examine a subset of DNA variants in thousands of unrelated people with common diseases. Now, however, sequencing costs are dropping, and whole genome sequences can quickly provide in-depth information about individuals, enabling scientists to locate genetic mutations that underlie rare diseases by sequencing a handful of people.

"Years ago, people were using families and mapping approaches to distil down to a region where they thought a causative gene was,"

epigenetic variation





the challenge

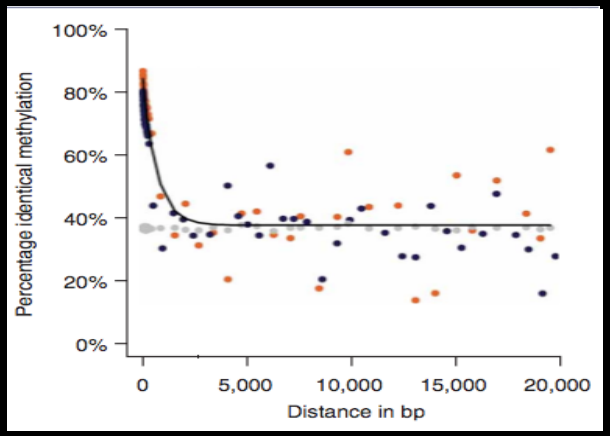
"*needle-in-a-haystack*" problem

Genotype

- 10-15 M SNPs
- 1 M tagSNPs

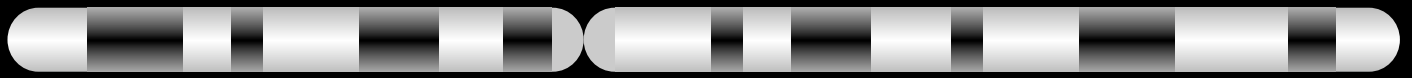
Epigenotype

- 30 M MVPs (28,112,194 NCBI36)
- 3 M tagMVPs (?)



methylation profiling approaches

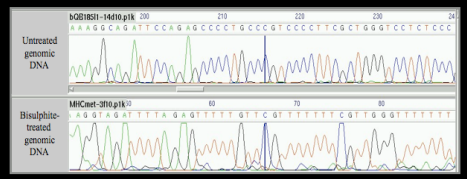
Genome



candidate approach

Samples

1000s



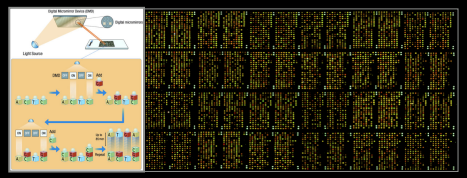
'Bis-seq'
bisulphite sequencing
e.g. ABI-3700 platform

Coverage (0.1-1%)

genome-wide approach

Samples

100s



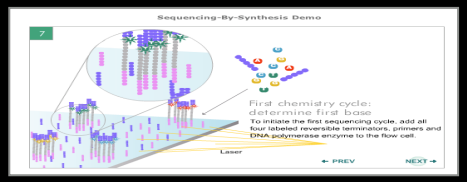
'MeDIP-chip'
immunoprecipitation & chip
e.g. Nimblegen platform

Coverage (1-10%)

whole-genome approach

Samples

10s



'MeDIP-seq'
immunoprecipitation & sequencing
e.g. Solexa platform

Coverage (10-100%)

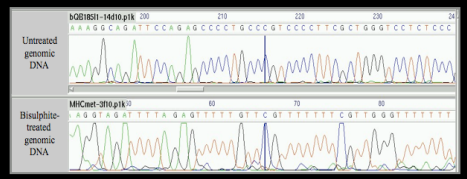
methylation profiling approaches

Genome

candidate approach

Samples

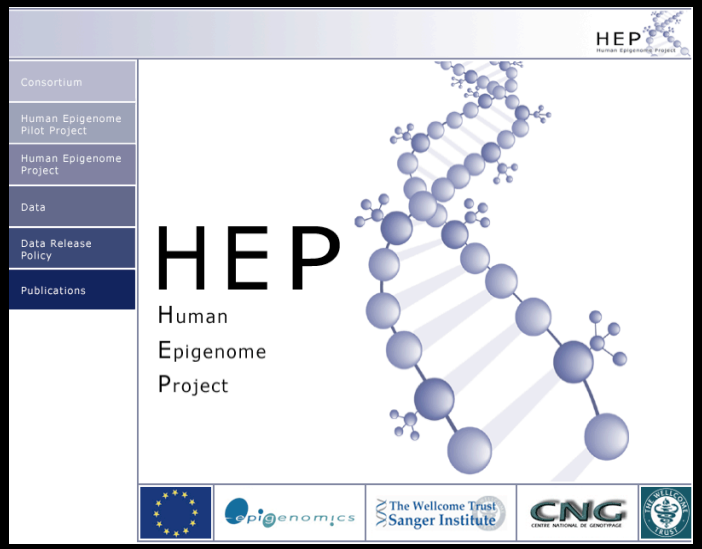
1000s

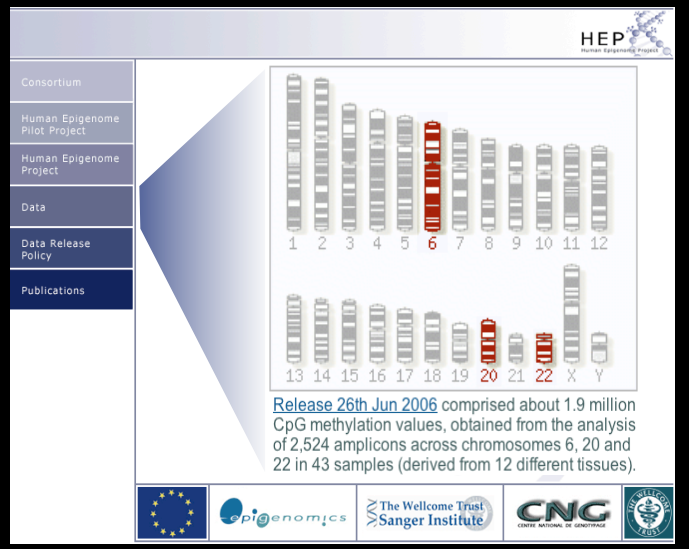


'Bis-seq'

bisulphite sequencing
e.g. ABI-3700 platform

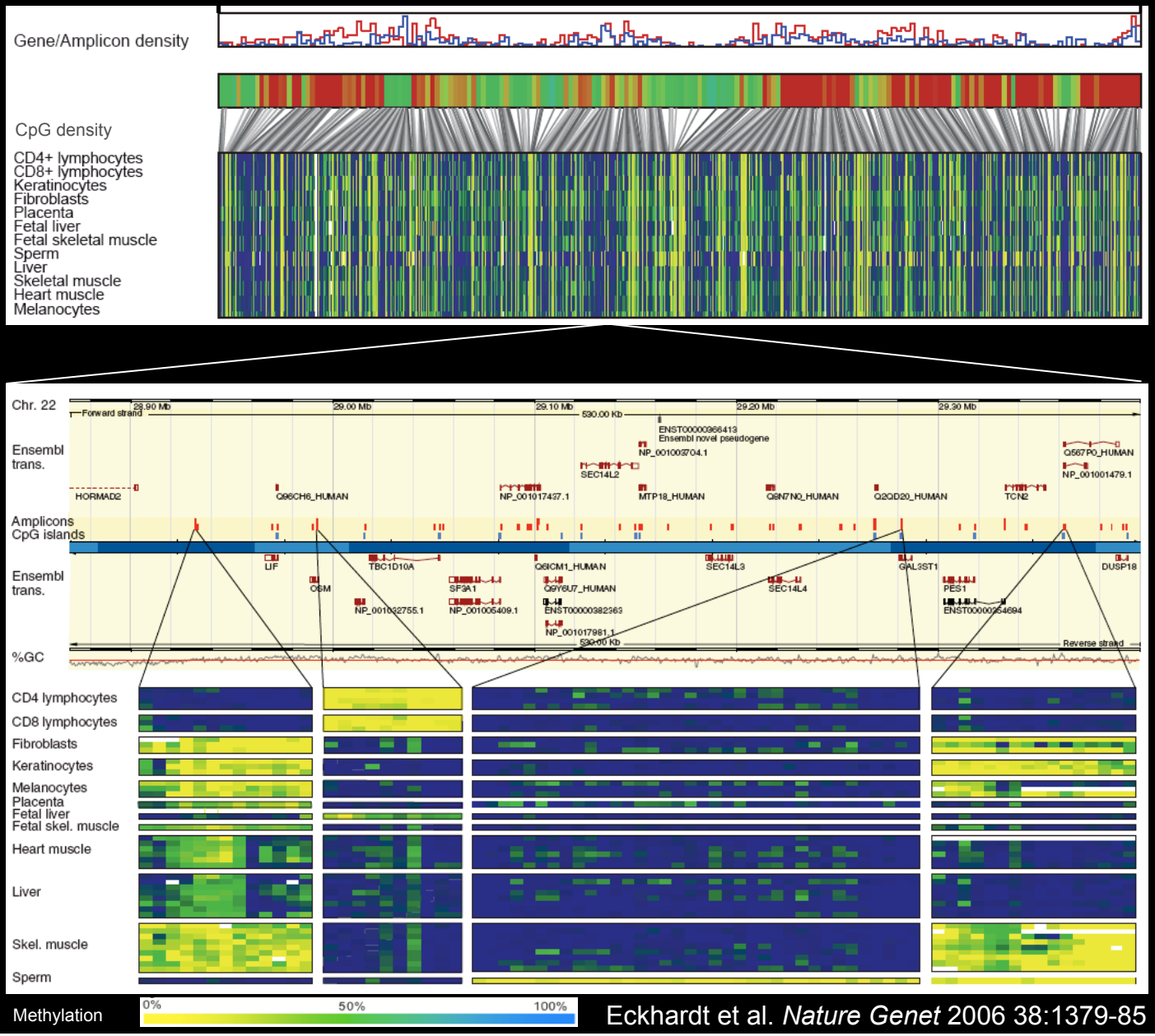
Coverage (0.1-1%)





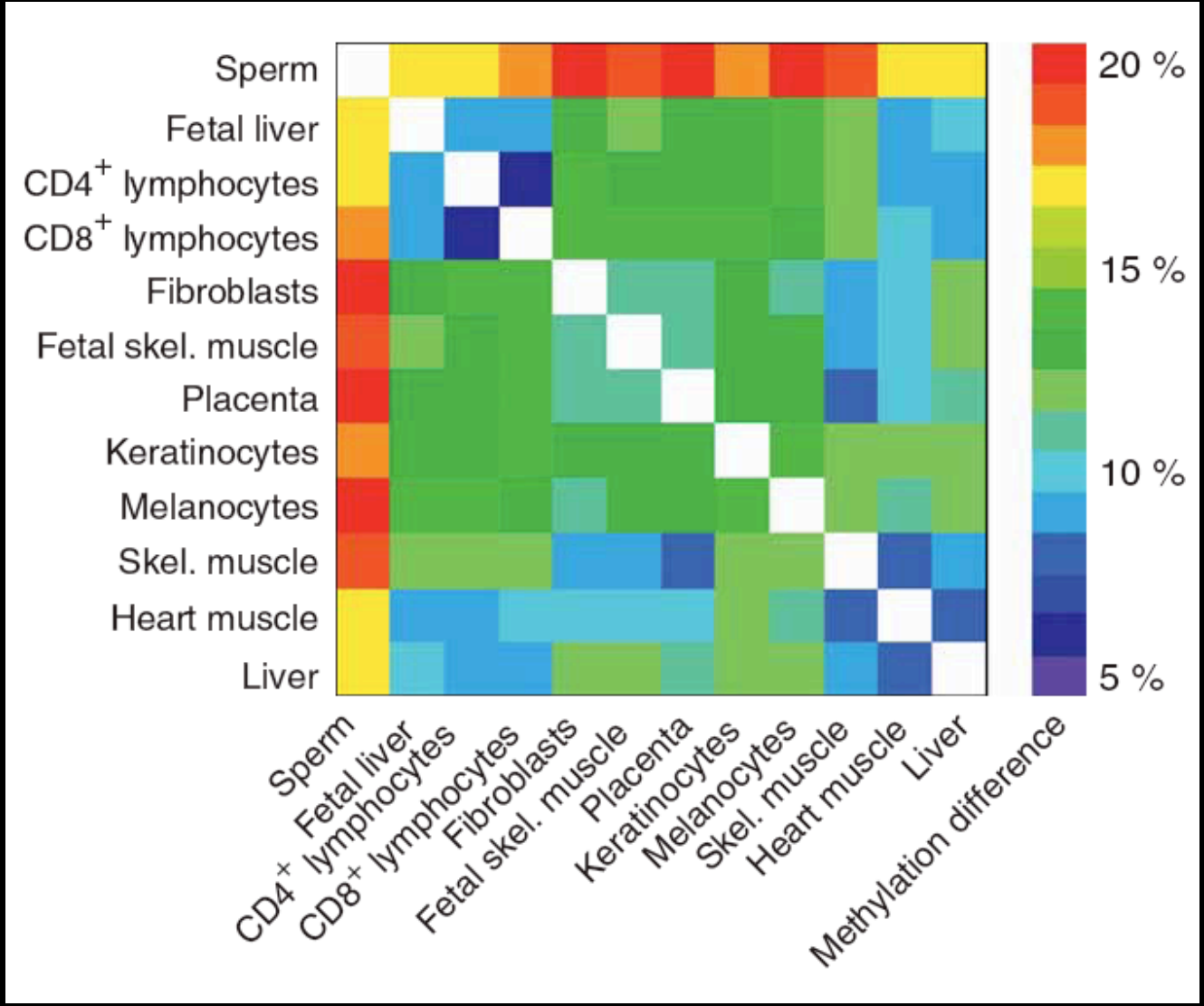
Eckhardt et al. 2006 *Nature Genet* 38:1379-85

methylation profile of chr 22



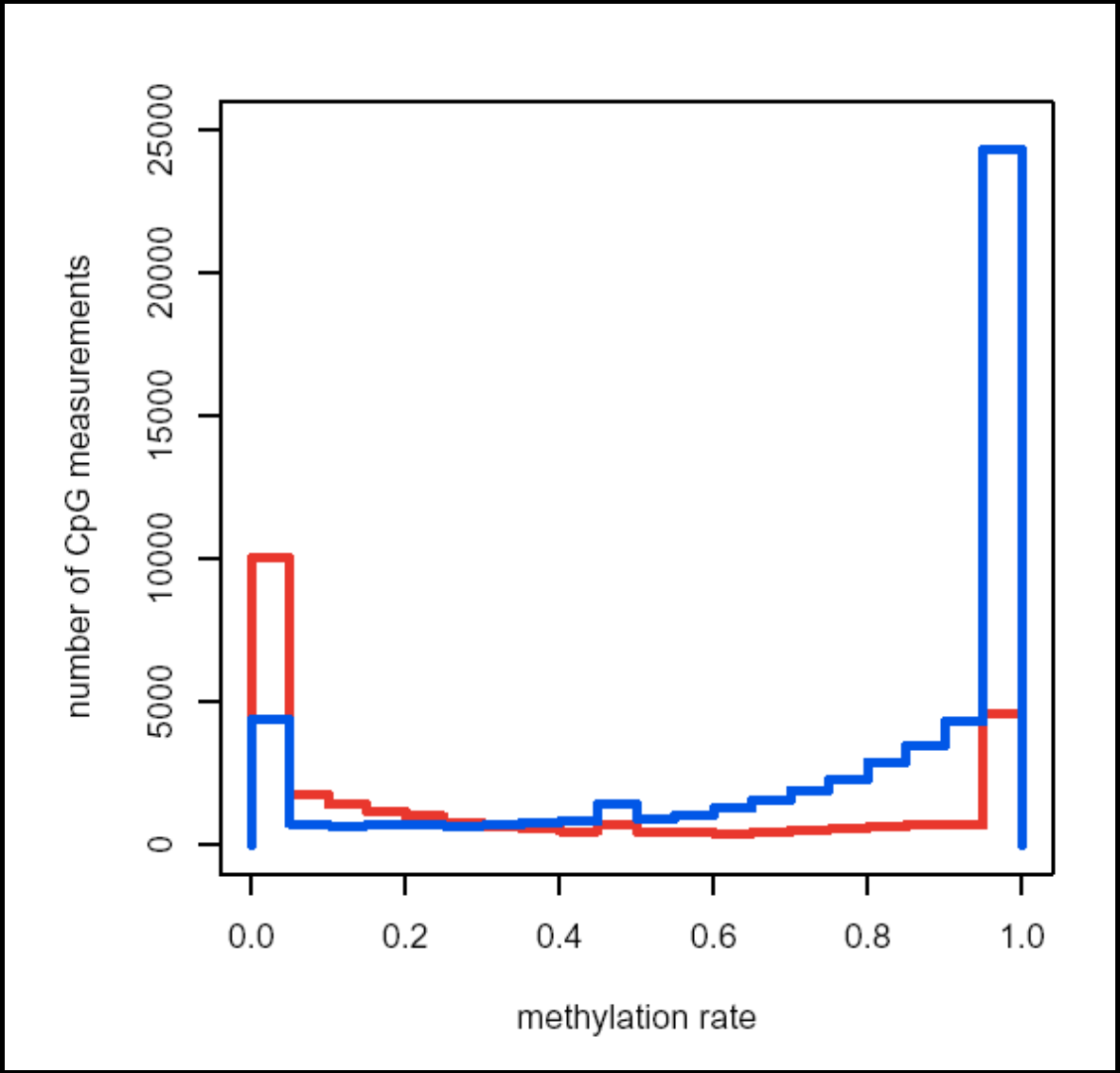
Eckhardt et al. *Nature Genet* 2006 38:1379-85

tissue-specific methylation



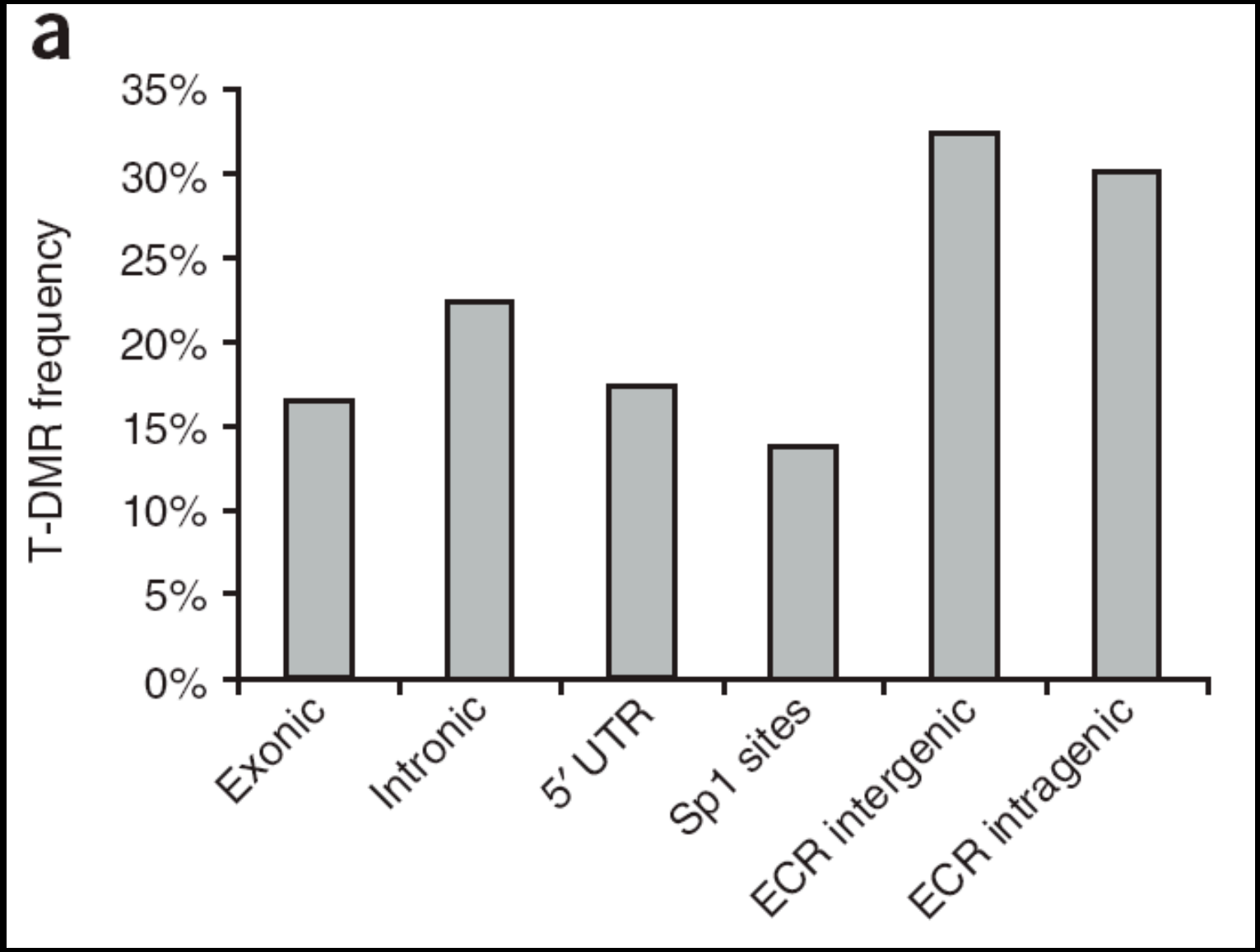
Eckhardt et al. 2006 *Nature Genet* 38:1379-85

distribution



Rakyan et al. 2004 *PLoS Biol* 2(12):e405
Eckhardt et al. 2006 *Nature Genet* 38:1379-85

target sites



Eckhardt et al. 2006 *Nature Genet* 38:1379-85

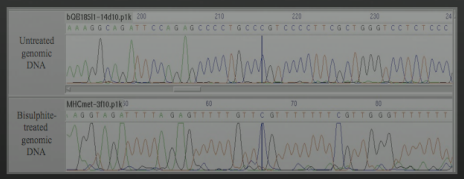
methylation profiling approaches

Genome

candidate approach

Samples

1000s



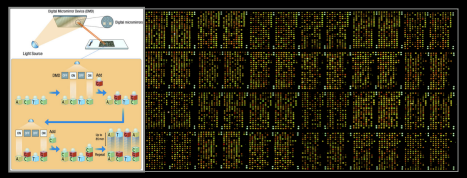
'Bis-seq'
bisulphite sequencing
e.g. ABI-3700 platform

Coverage (0.1-1%)

genome-wide approach

Samples

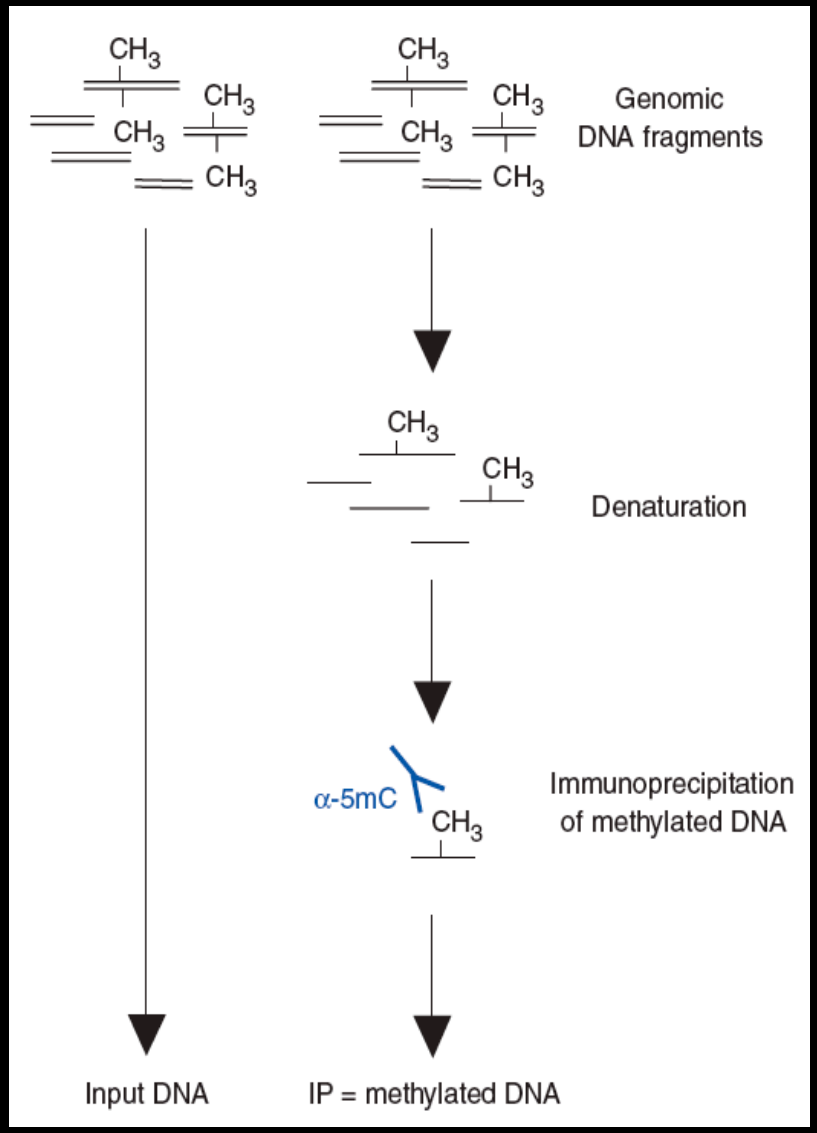
100s

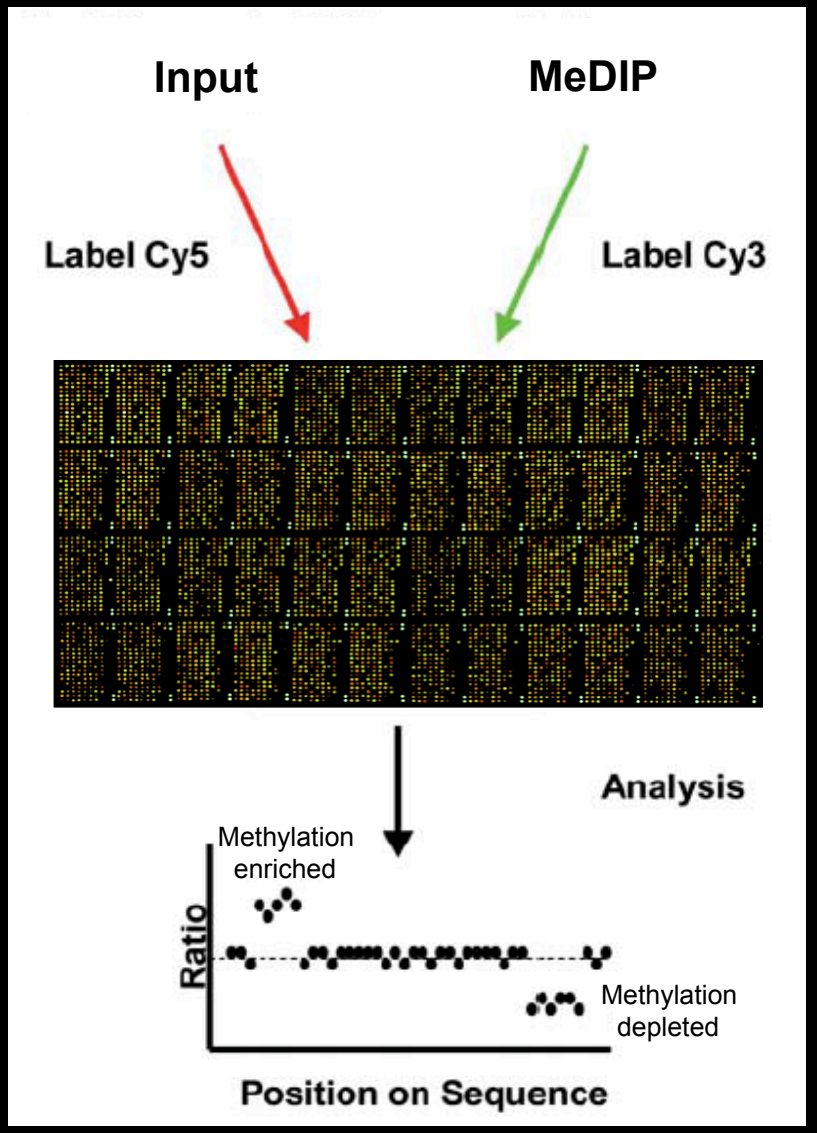


'MeDIP-chip'
immunoprecipitation & chip
e.g. Nimblegen platform

Coverage (1-10%)

genome-wide MeDIP-chip assay

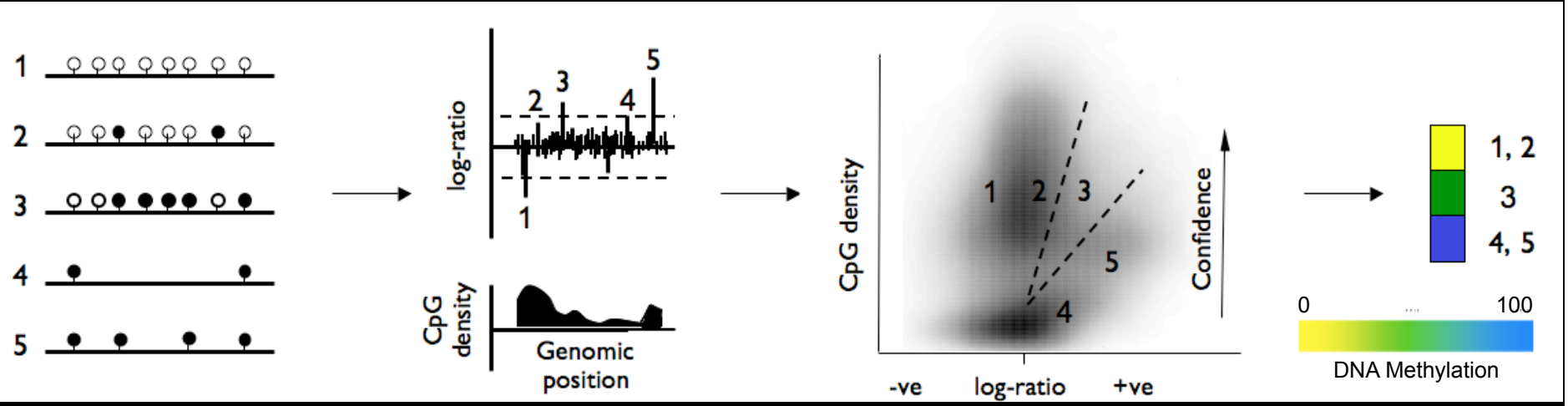




Weber et al. 2005 *Nat Genet.* 37:853-862

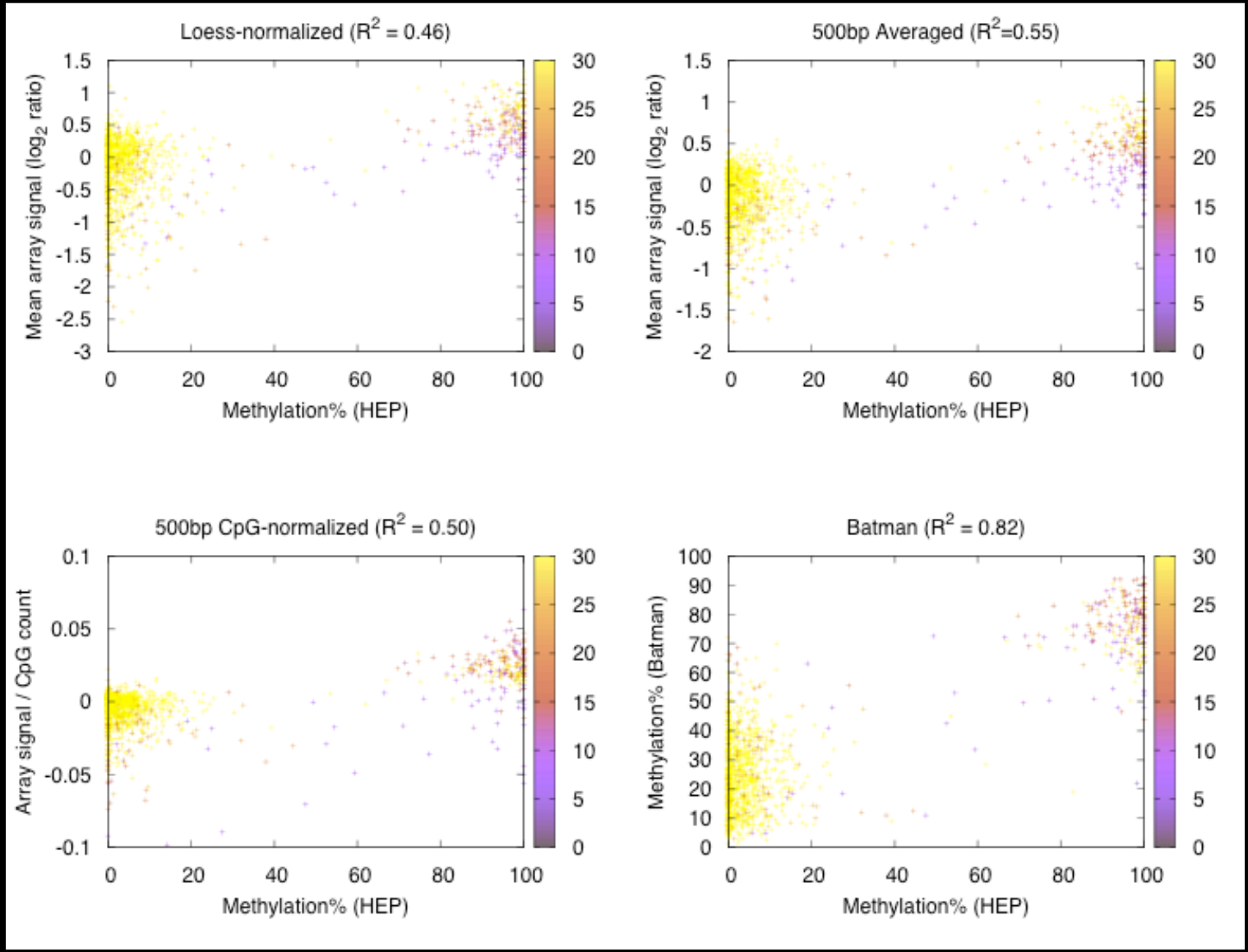
BATMAN

(Bayesian Tool for Methylation Analysis)



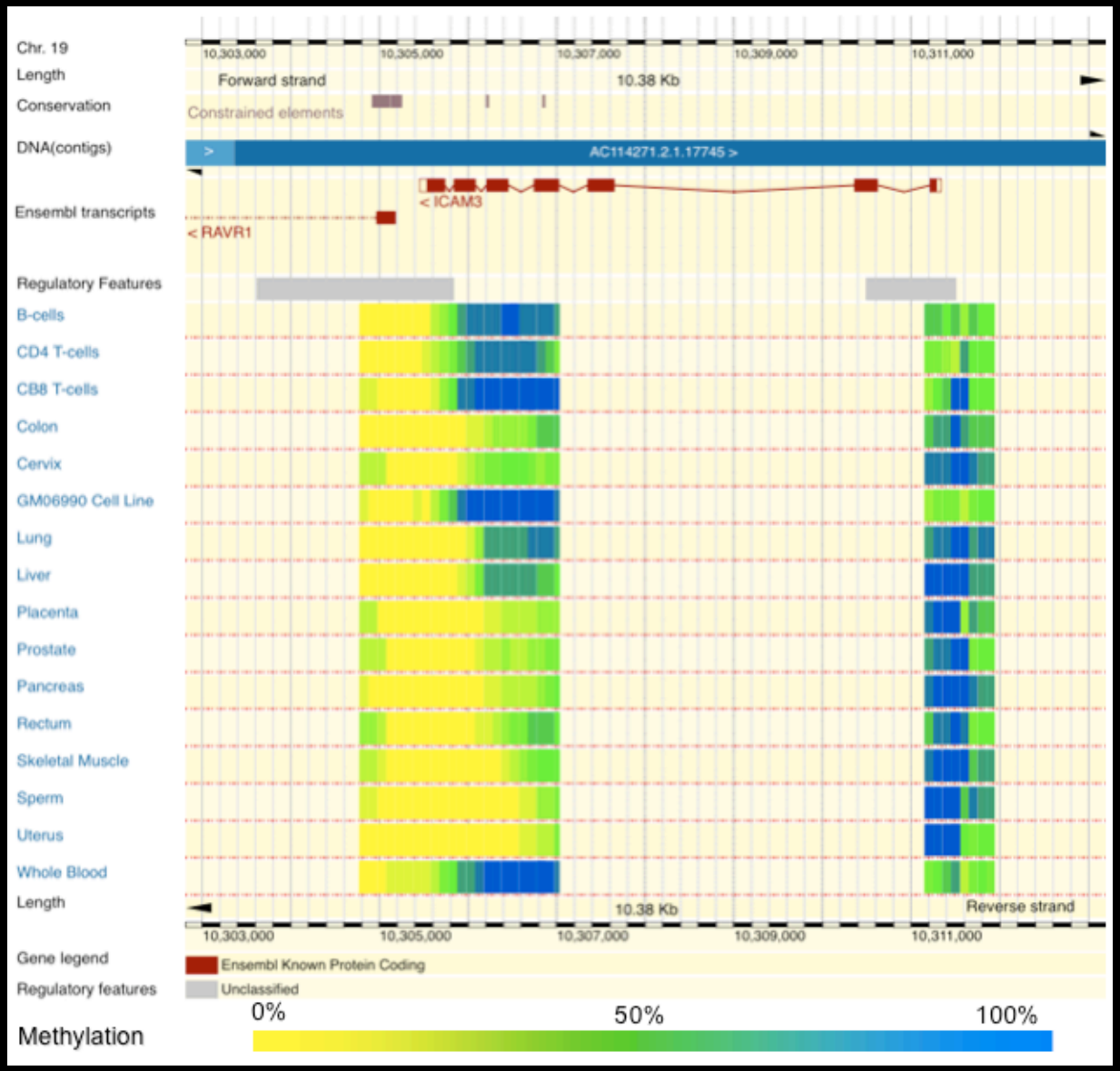
Down et al. *Nature Biotech* 2008 26:779-85

BATMAN performance



Down et al. *Nature Biotech*, July 8, 2008

genome-wide methylation profiles



Rakyan et al., *Genome Res,* 2008 18:1518-29

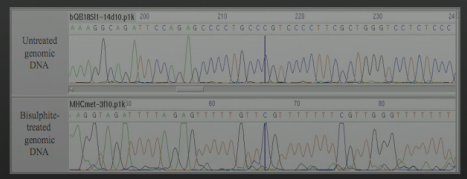
methylation profiling approaches

Genome

candidate approach

Samples

1000s



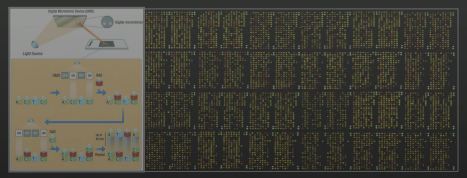
'Bis-seq'
bisulphite sequencing
e.g. ABI-3700 platform

Coverage (0.1-1%)

genome-wide approach

Samples

100s



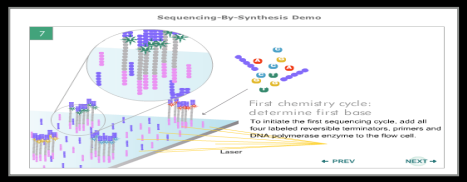
'MeDIP-chip'
immunoprecipitation & chip
e.g. Nimblegen platform

Coverage (1-10%)

whole-genome approach

Samples

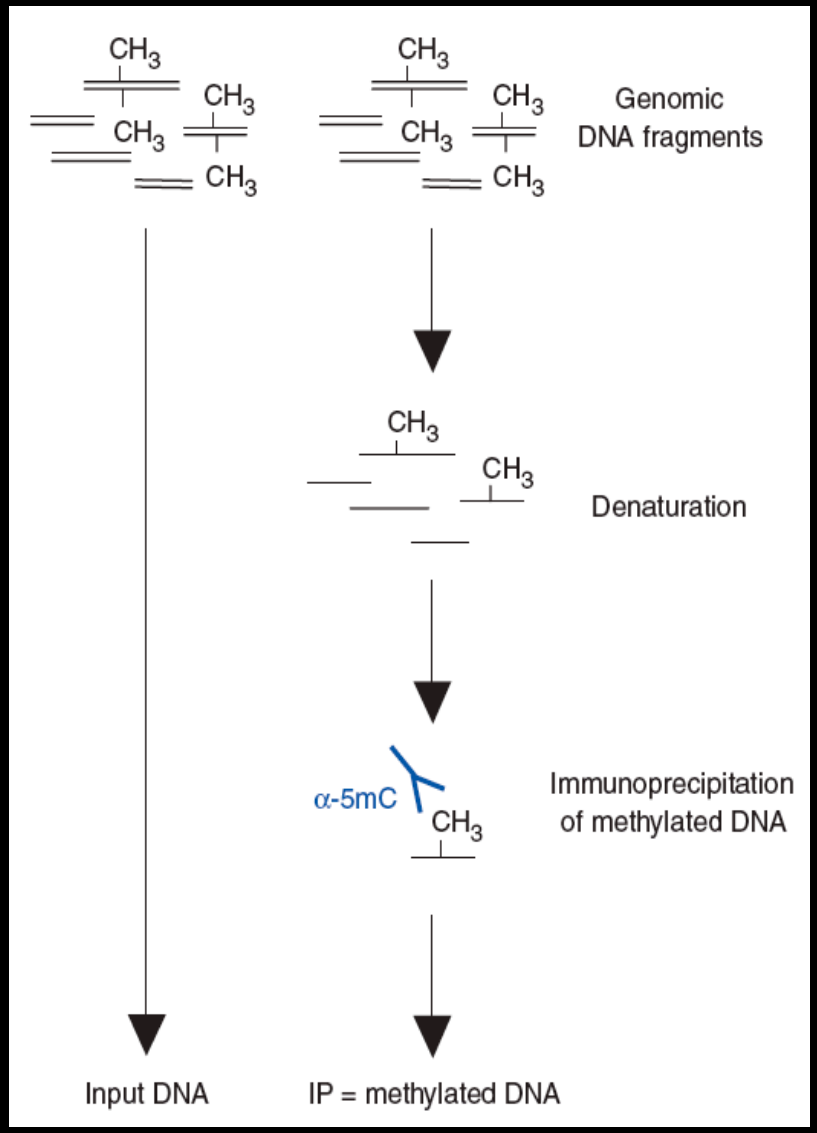
10s



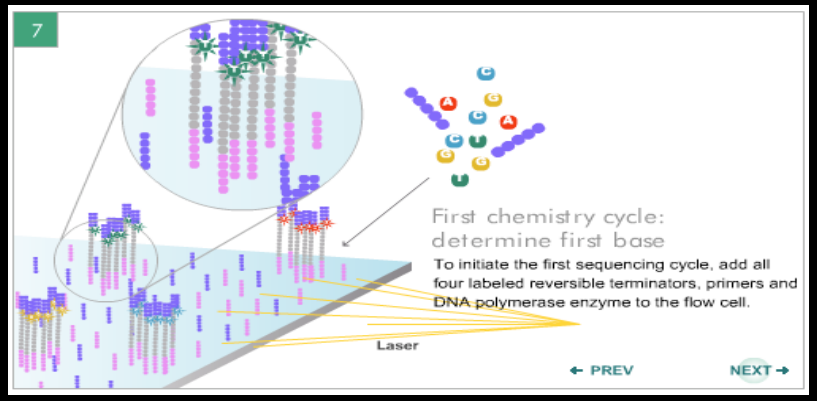
'MeDIP-seq'
immunoprecipitation & sequencing
e.g. Solexa platform

Coverage (10-100%)

whole-genome MeDIP-seq assay

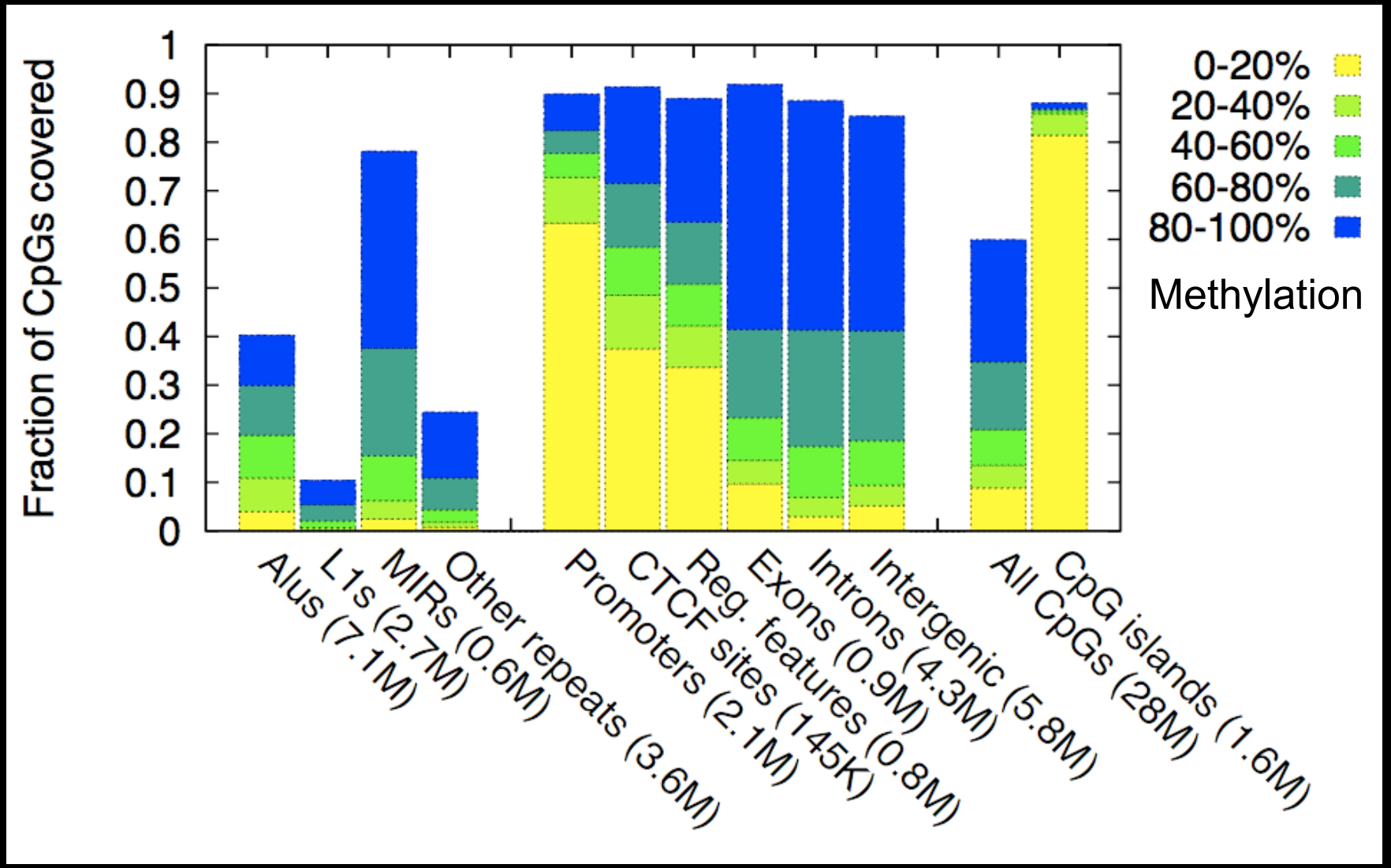


Weber et al. 2005 *Nat Genet.* 37:853-862

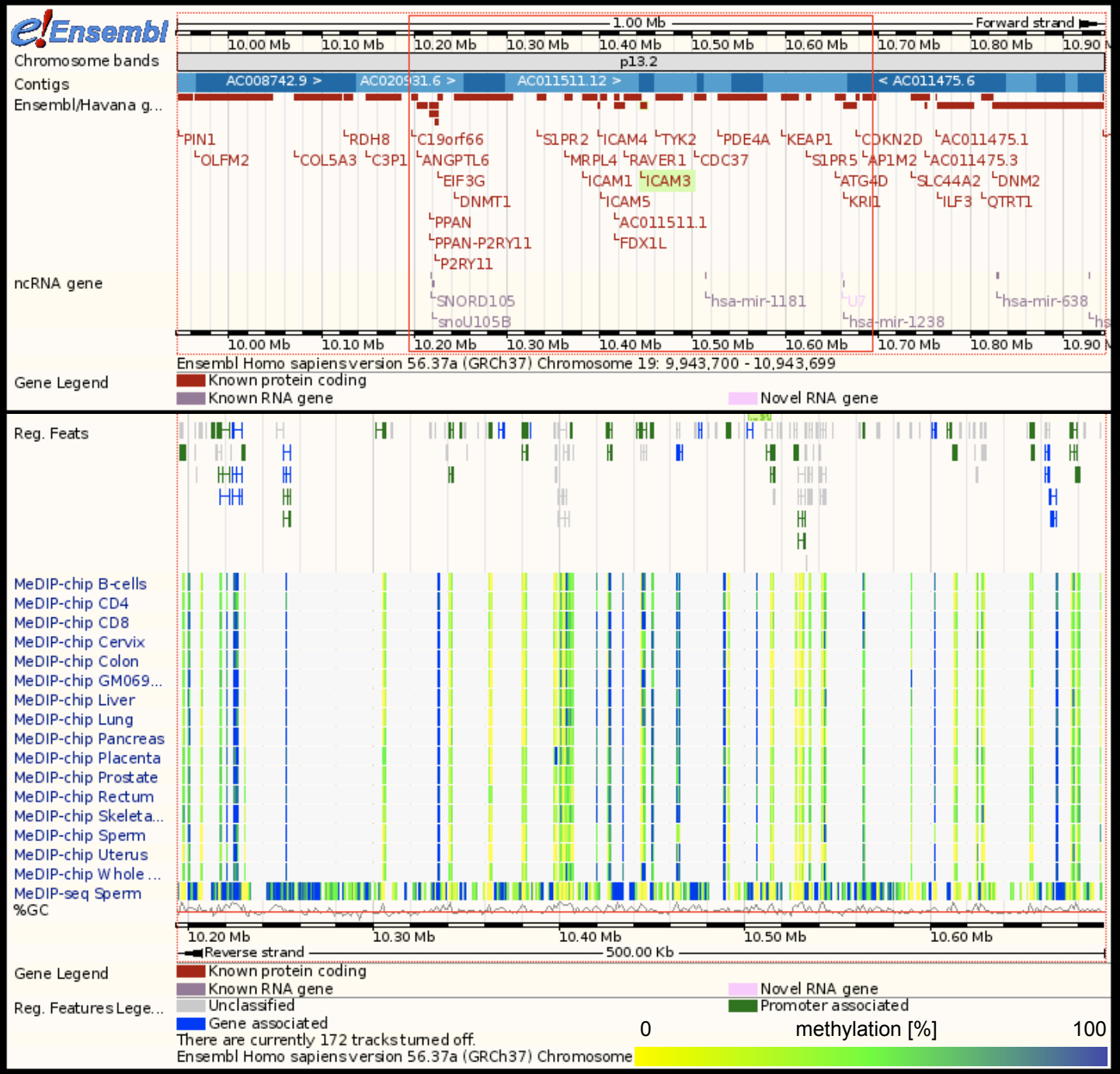


www.illumina.com

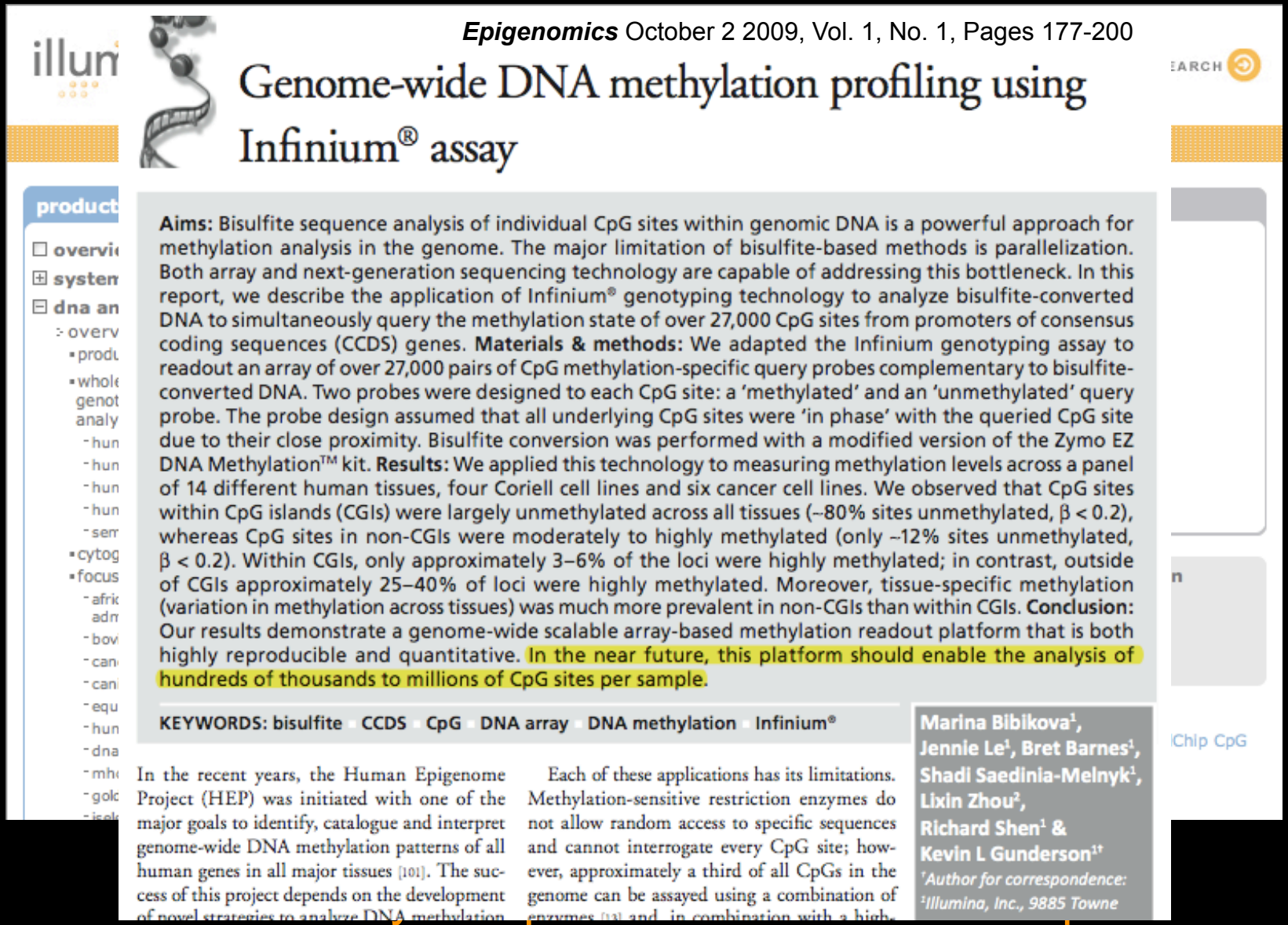
methylome of human male germline



Down et al. *Nature Biotech* 2008 26:779-85


Ensembl
1.00 Mb
Forward strand
10.00 Mb
10.10 Mb
10.20 Mb
10.30 Mb
10.40 Mb
10.50 Mb
10.60 Mb
10.70 Mb
10.80 Mb
10.90 Mb
Chromosome bands
p13.2
Contigs
AC008742.9 >
AC011511.12 >
< AC011475.6
Ensembl/Havana g...
PIN1
RDH8
C19orf66
S1PR2
ICAM4
TYK2
PDE4A
KEAP1
CDKN2D
AC011475.1
OLFM2
COL5A3
C3P1
ANGPTL6
MRPL4
RAVER1
CDC37
S1PR5
AP1M2
AC011475.3
EIF3G
ICAM1
ICAM3
ATG4D
SLC44A2
DNM2
DNMT1
ICAM5
KRI1
ILF3
QTRT1
PPAN
AC011511.1
PPAN-P2RY11
FDX1L
P2RY11
ncRNA gene
SNORD105
hsa-mir-1181
U7
hsa-mir-638
snoU105B
hsa-mir-1238
Ensembl Homo sapiens version 56.37a (GRCh37) Chromosome 19: 9,943,700 - 10,943,699
Gene Legend
Known protein coding
Known RNA gene
Novel RNA gene
Reg. Feats
MeDIP-chip B-cells
MeDIP-chip CD4
MeDIP-chip CD8
MeDIP-chip Cervix
MeDIP-chip Colon
MeDIP-chip GM069...
MeDIP-chip Liver
MeDIP-chip Lung
MeDIP-chip Pancreas
MeDIP-chip Placenta
MeDIP-chip Prostate
MeDIP-chip Rectum
MeDIP-chip Skeleta...
MeDIP-chip Sperm
MeDIP-chip Uterus
MeDIP-chip Whole ...
MeDIP-seq Sperm
%GC
Reverse strand
500.00 Kb
Unclassified
Promoter associated
Gene associated
0
methylation [%]
100
There are currently 172 tracks turned off.

methylome on a chip



illun

product

overvi

syster

dna an

overv

produ

whole genot analy

hun

hun

hun

hun

sem

cytog

focus

afric adn

bov

can

cani

equ

hun

dna

mh

gol

isel

Epigenomics October 2 2009, Vol. 1, No. 1, Pages 177-200

Genome-wide DNA methylation profiling using Infinium® assay

Aims: Bisulfite sequence analysis of individual CpG sites within genomic DNA is a powerful approach for methylation analysis in the genome. The major limitation of bisulfite-based methods is parallelization. Both array and next-generation sequencing technology are capable of addressing this bottleneck. In this report, we describe the application of Infinium® genotyping technology to analyze bisulfite-converted DNA to simultaneously query the methylation state of over 27,000 CpG sites from promoters of consensus coding sequences (CCDS) genes. **Materials & methods:** We adapted the Infinium genotyping assay to readout an array of over 27,000 pairs of CpG methylation-specific query probes complementary to bisulfite-converted DNA. Two probes were designed to each CpG site: a 'methylated' and an 'unmethylated' query probe. The probe design assumed that all underlying CpG sites were 'in phase' with the queried CpG site due to their close proximity. Bisulfite conversion was performed with a modified version of the Zymo EZ DNA Methylation™ kit. **Results:** We applied this technology to measuring methylation levels across a panel of 14 different human tissues, four Coriell cell lines and six cancer cell lines. We observed that CpG sites within CpG islands (CGIs) were largely unmethylated across all tissues (~80% sites unmethylated, $\beta < 0.2$), whereas CpG sites in non-CGIs were moderately to highly methylated (only ~12% sites unmethylated, $\beta < 0.2$). Within CGIs, only approximately 3–6% of the loci were highly methylated; in contrast, outside of CGIs approximately 25–40% of loci were highly methylated. Moreover, tissue-specific methylation (variation in methylation across tissues) was much more prevalent in non-CGIs than within CGIs. **Conclusion:** Our results demonstrate a genome-wide scalable array-based methylation readout platform that is both highly reproducible and quantitative. In the near future, this platform should enable the analysis of hundreds of thousands to millions of CpG sites per sample.

KEYWORDS: bisulfite · CCDS · CpG · DNA array · DNA methylation · Infinium®

Marina Bibikova[1], Jennie Le[1], Bret Barnes[1], Shadi Saedinia-Melnyk[1], Lixin Zhou[2], Richard Shen[1] & Kevin L Gunderson[1†]

[†]Author for correspondence:

[1]Illumina, Inc., 9885 Towne

In the recent years, the Human Epigenome Project (HEP) was initiated with one of the major goals to identify, catalogue and interpret genome-wide DNA methylation patterns of all human genes in all major tissues [101]. The success of this project depends on the development of novel strategies to analyze DNA methylation

Each of these applications has its limitations. Methylation-sensitive restriction enzymes do not allow random access to specific sequences and cannot interrogate every CpG site; however, approximately a third of all CpGs in the genome can be assayed using a combination of enzymes [13] and, in combination with a high-

EARCH

n

Chip CpG

nature nanotechnology

ARTICLES

PUBLISHED ONLINE: 22 FEBRUARY 2009 | DOI: 10.1038/NNANO.2009.12

Continuous base identification for single-molecule nanopore DNA sequencing

James Clarke[1], Hai-Chen Wu[2], Lakmal Jayasinghe[1,2], Alpesh Patel[1], Stuart Reid[1] and Hagan Bayley[2]*

A single-molecule method for sequencing DNA that does not require fluorescent labelling could reduce costs and increase sequencing speeds. An exonuclease enzyme might be used to cleave individual nucleotide molecules from the DNA, and when coupled to an appropriate detection system, these nucleotides could be identified in the correct order. Here, we show that a protein nanopore with a covalently attached adapter molecule can continuously identify unlabelled nucleoside 5′-monophosphate molecules with accuracies averaging 99.8%. Methylated cytosine can also be distinguished from the four standard DNA bases: guanine, adenine, thymine and cytosine. The operating conditions are compatible with the exonuclease, and the kinetic data show that the nucleotides have a high probability of translocation through the nanopore and, therefore, of not being registered twice. This highly accurate tool is suitable for integration into a system for sequencing nucleic acids and for analysing epigenetic modifications.

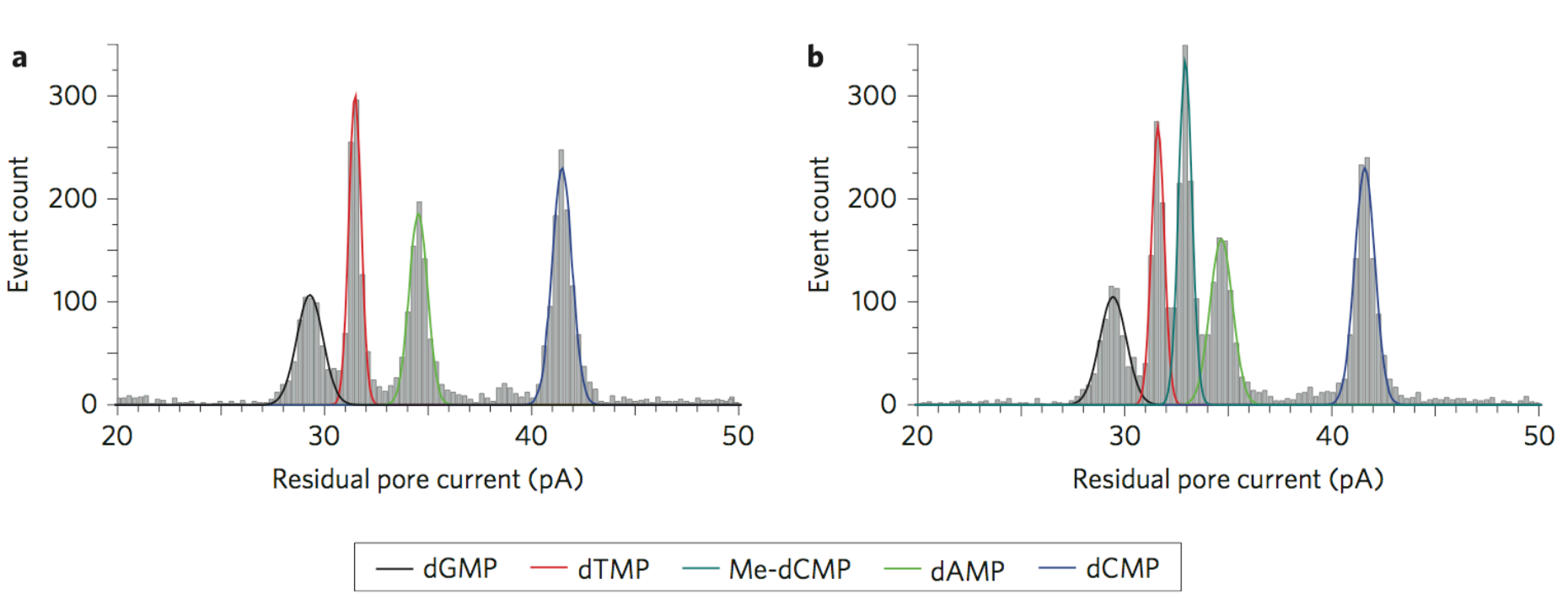


Table 1. Major projects, resources and initiatives dedicated to epigenomic research

'Epigenome' efforts	Start	Goals	URL
Human Epigenome Project (HEP)	2000	The HEP aims to identify, catalogue and interpret genome-wide DNA methylation patterns of all human genes in all major tissues.	http://www.epigenome.org
Encyclopedia of DNA Elements (ENCODE)	2003	ENCODE aims to carry out a project to identify all functional elements in the human genome sequence.	http://www.genome.gov/10005107
Epigenome Network of Excellence (NoE)	2004	The NoE aims to create a virtual core institute. Specific aims include (i) to advance scientific discoveries through a joint research programme, (ii) to integrate young colleagues through the NET-programme and (iii) to establish an open dialogue by building an interactive Website.	http://www.epigenome-noe.net
National Methylome 21 (NAME21)	2005	NAME21 aims to generate a first comprehensive DNA methylation map of all genes on human chromosome 21 using bisulphite sequencing technologies.	http://www.faculty.iu-bremen.de/ajeltsch/name
Epigenetic Treatment of Neoplastic Disease (EPITRON)	2005	EPITRON aims to define and validate epigenetic cancer treatment. Sepific aims include (i) to define epigenetic alterations in cancer, (ii) to identify therapeutic targets and (iii) to develop epi-drugs.	http://www.epitron.eu
Highthroughput Epigenetic Regulatory Organization In Chromatin (HEROIC)	2005	HEROIC aims to advance knowledge of chromatin function. Specific aims include (i) to decipher epigenetic profiles, transcription factor networks and nuclear organization; (ii) to focus on mouse ES cells and derivatives and (iii) to develop bioinformatics Tools.	http://www.heroic-ip.eu
Epigenetic Control of the Mammalian Genome (GEN-AU)	2006	GEN-AU aims to better understand the epigenetic control of mammalian genomes. Specific aims include (i) to profile histone modifications, (ii) to study imprinting and X chromosome inactivation and (iii) to identify polycomb-trithorax response elements.	http://www.gen-au.at
AACR Human Epigenome Taskforce and Alliance for the Human Epigenome and Disease (AHEAD)	2006	AACR Human Epigenome Taskforce developed the blueprint for an international human epigenome project and developed a timetable for the implementation of the AHEAD project.	http://www.aacr.org/home/scientists/working-groups-task-forces/task-forces/human-epigenome-task-force.aspx
NIH Roadmap: Epigenomics	2008	The Roadmap Epigenomics Program aims to generate comprehensive reference maps and new technology for epigenomic analysis. Specific aims include (i) to create an international committee; (ii) to develop standardized platforms, procedures and reagents for epigenomics research; (iii) to conduct demonstration projects to evaluate how epigenomes change; (iv) to develop new technologies for single cell epigenomic analysis and *in vivo* imaging of epigenetic activity and (v) to create a public data resource to accelerate the application of epigenomics approaches.	http://nihroadmap.nih.gov/epigenomics/

NIH Epigenome Roadmap: $190M for 5 years (2008-2013)

5 Awards for Epigenome Mapping/Coordination Centres

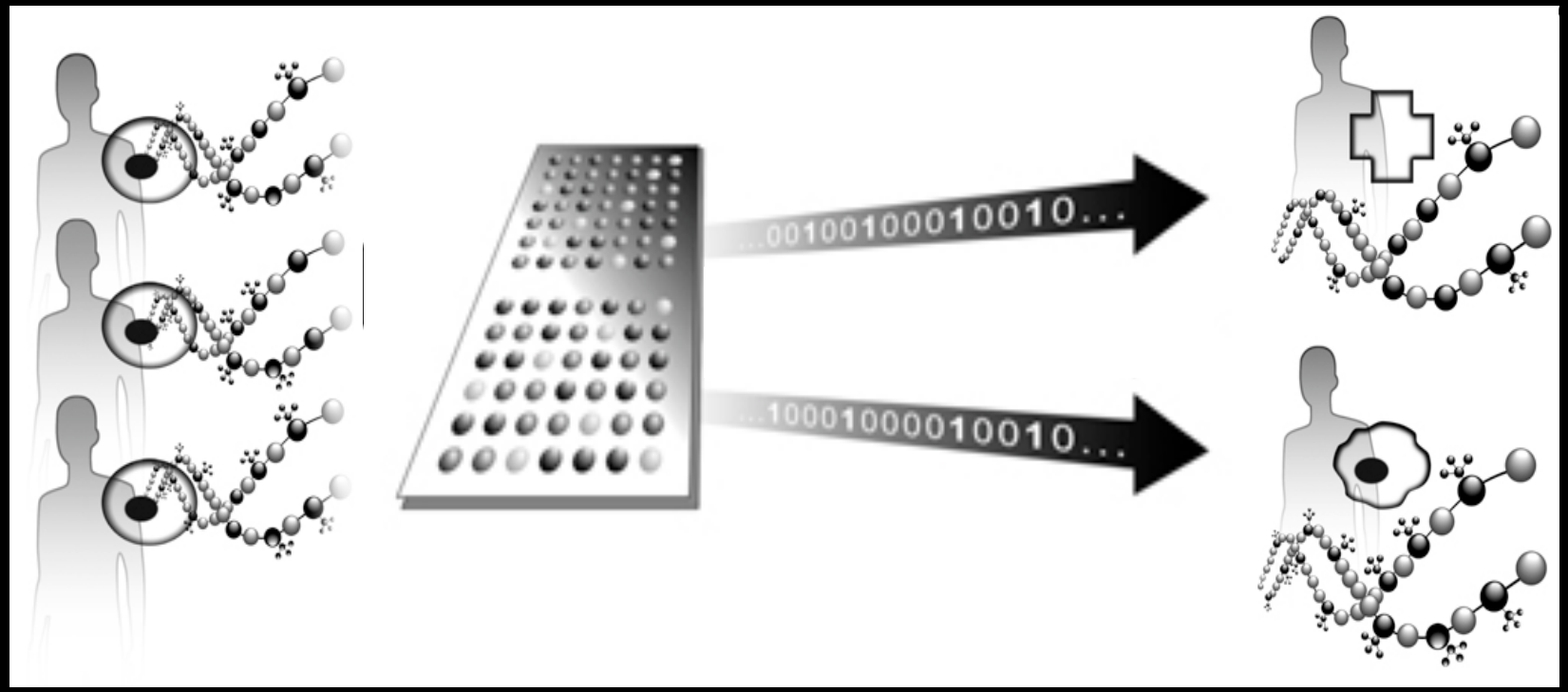
9 Awards for Technology Development in Epigenetics

7 Awards for Discovery of Novel Epigenetic Marks

22 Awards for Epigenomics of Human Health and Disease

Cancer, Alzheimer's, Atherosclerosis, Autism, Hypertension, Bipolar Disorder, Asthma, Lupus Erythematosus, Schizophrenia, Kidney Disease, Muscular Dystrophy and others

genome-wide association studies

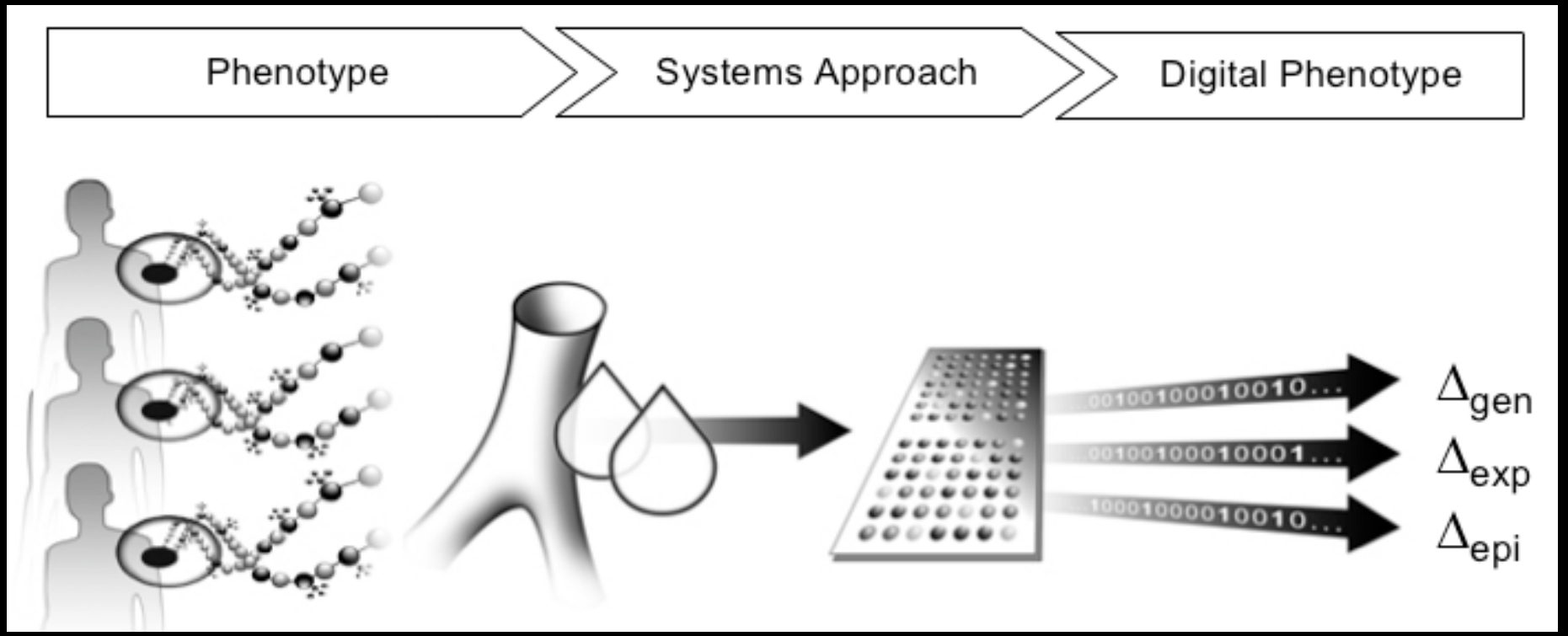


- Type 1 Diabetes
- Type 2 Diabetes
- Inflammatory Bowel Disease

} **WTCCC**

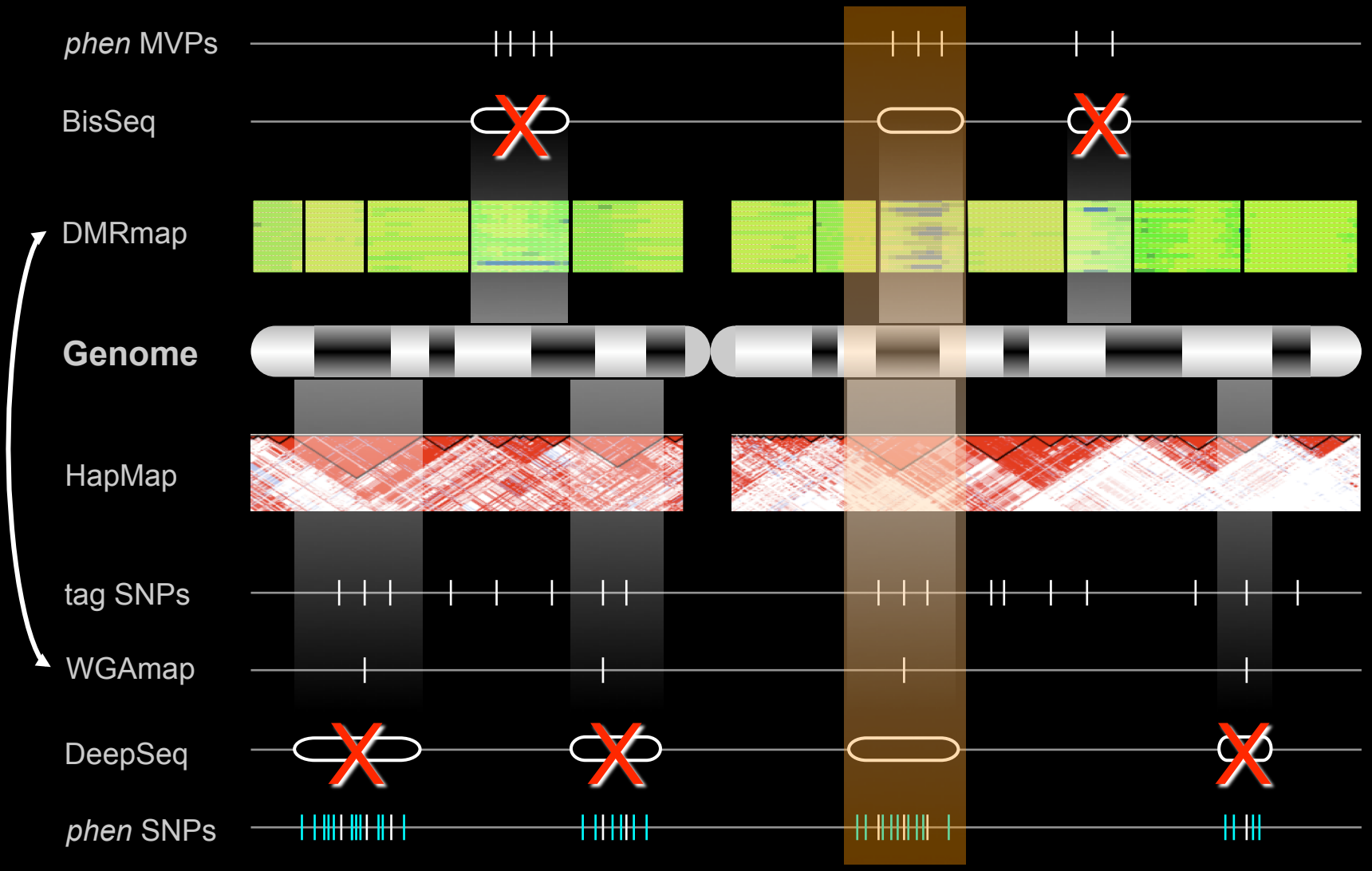
- Cancer (Sarcomas, NET, etc)

integrated (epi)genetic approach



reverse phenotyping

integrated (epi)genomic approach



examples . . .

Cell *129*, 879–890, June 1, 2007 ©2007 Elsevier Inc. 879 Cell

Downregulation of *Death-Associated Protein Kinase 1* (*DAPK1*) in Chronic Lymphocytic Leukemia

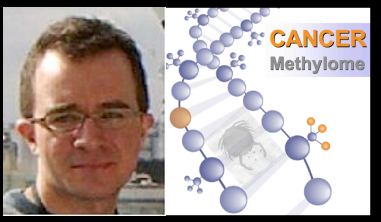
Aparna Raval,[1,10] Stephan M. Tanner,[1] John C. Byrd,[2] Elizabeth B. Angerman,[1] James D. Perko,[1] Shih-Shih Chen,[1] Björn Hackanson,[1,8] Michael R. Grever,[2] David M. Lucas,[2] Jennifer J. Matkovic,[2] Thomas S. Lin,[2] Thomas J. Kipps,[6] Fiona Murray,[7] Dennis Weisenburger,[4] Warren Sanger,[4] Jane Lynch,[4] Patrice Watson,[4] Mary Jansen,[4] Yuko Yoshinaga,[3] Richard Rosenquist,[7] Pieter J. de Jong,[3] Penny Coggill,[5] Stephan Beck,[5] Henry Lynch,[4] Albert de la Chapelle,[1,9,*] and Christoph Plass[1,9,*]

nature genetics published online 22 June 2008; doi:10.1038/ng.174

Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation

Kristi Kerkel[1], Alexandra Spadola[2], Eric Yuan[1], Jolanta Kosek[1], Le Jiang[1], Eldad Hod[3], Kerry Li[1], Vundavalli V Murty[1,3], Nicole Schupf[4], Eric Vilain[5,6], Mitzi Morris[7], Fatemeh Haghighi[7] & Benjamin Tycko[1,3]

cancer methylome project



Andrew Feber

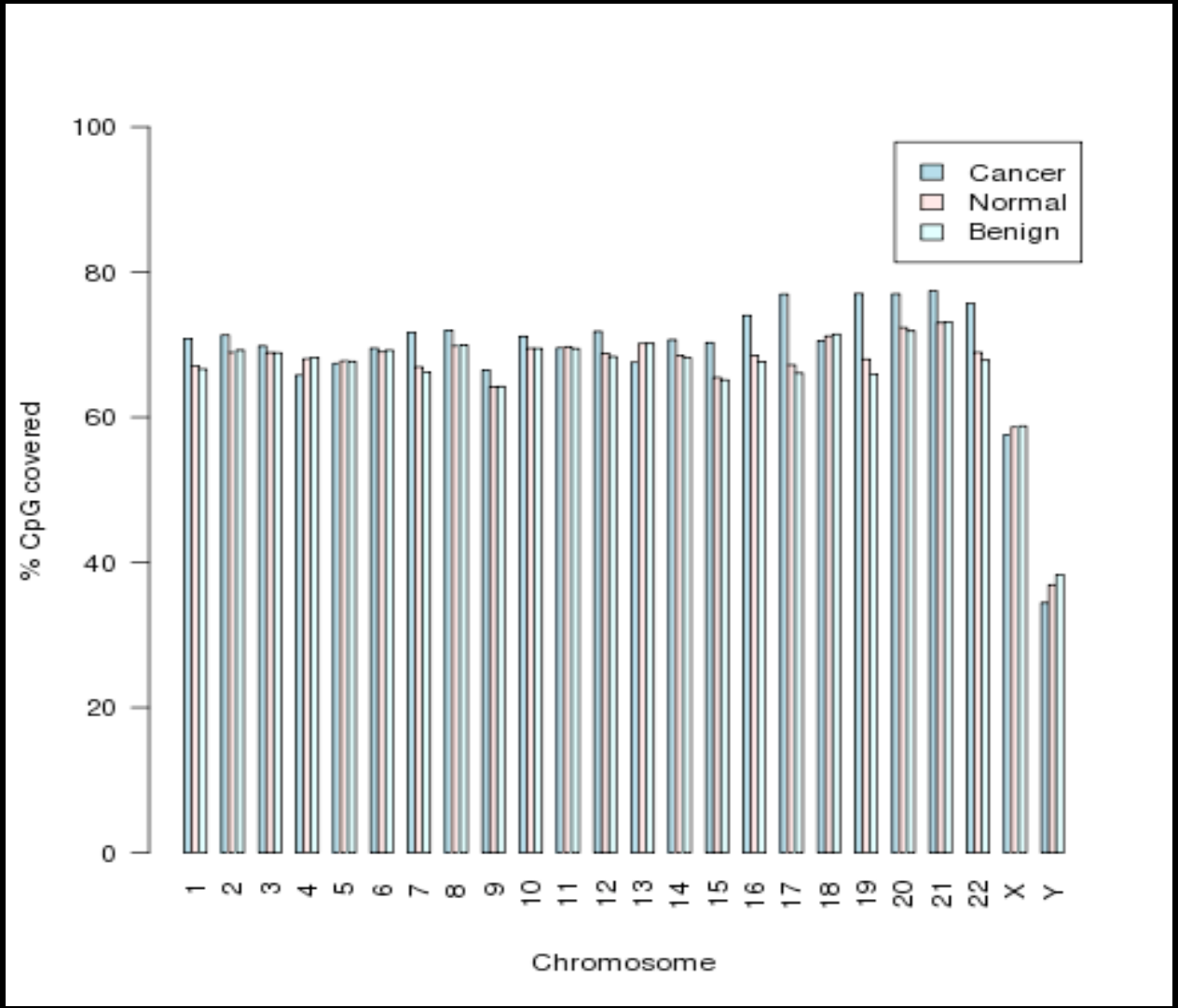
Neurofibroma

- common type of benign tumours affecting NF1 patients
- progression to malignant form is rare
- mechanism unknown, no molecular markers

Study Design

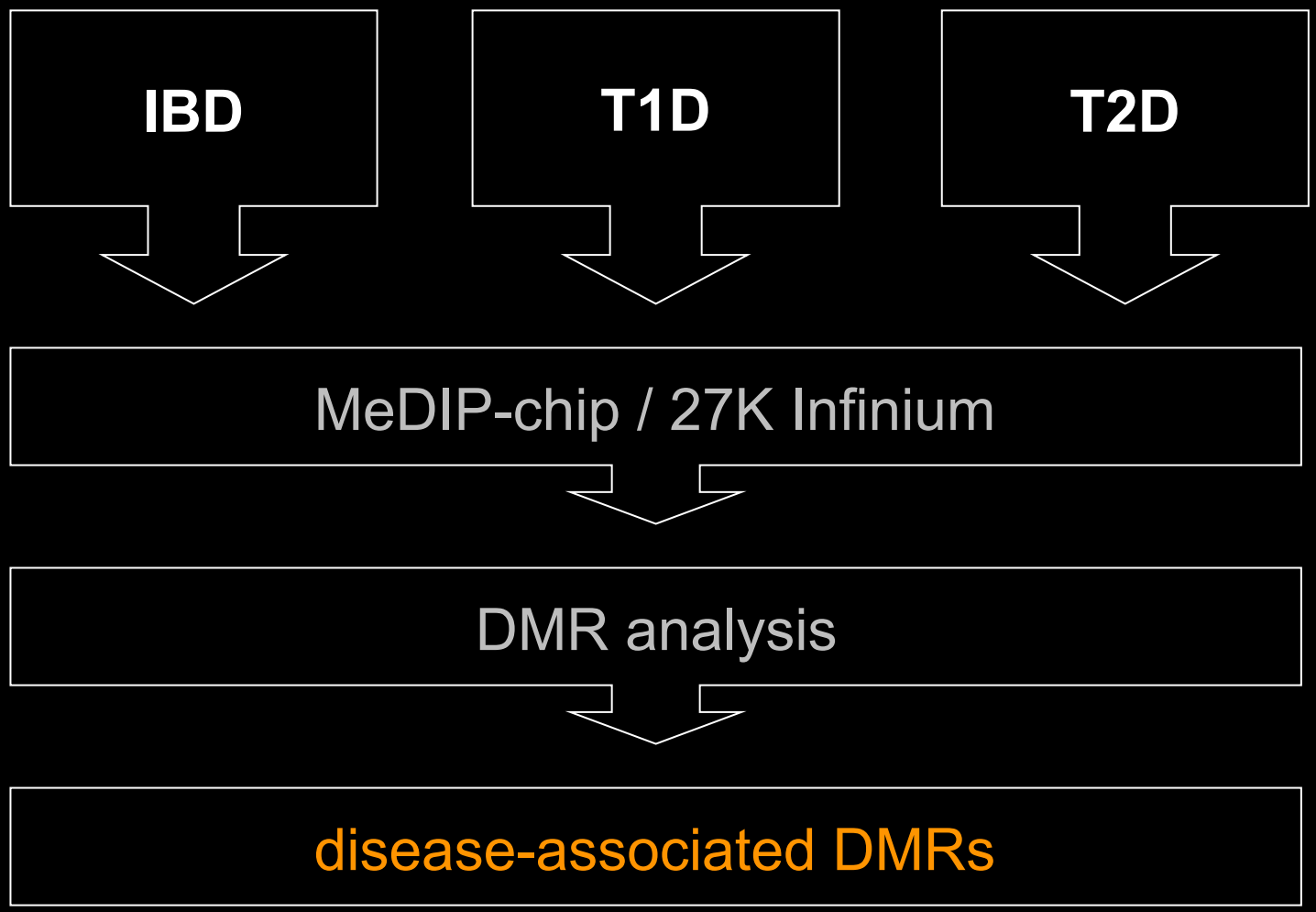
– pooled samples stratified for NF1 mutations
– control (n = 6, pooled)
– benign (n = 10, pooled)
– malignant (n = 10, pooled)
– Approach: MeDIP-seq

NF methylome – CpG coverage



Feber et al. unpublished

common diseases

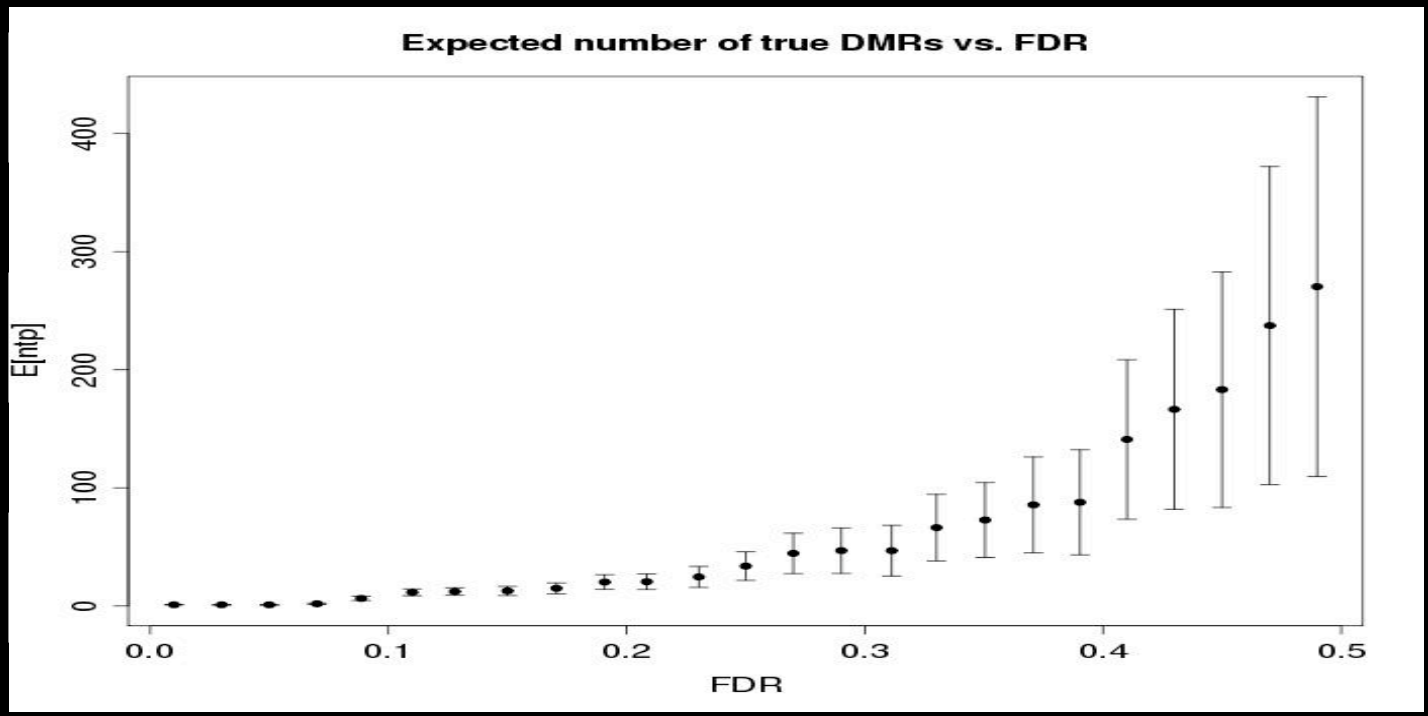


common diseases

IBD

T1D

T2D



Bäckdahl et al, unpublished

conclusions

- Technologies for DNA methylation analysis are available and working
- DNA methylation is stable, specific and ‘essentially’ binary
- Disease-associated DMRs exist in cancer and common disease and can be identified in tissue and blood
- Case for integrated (epi)genomic GWA studies

Acknowledgements

Technology

UCL-CI
Lee Butcher
Pawan Dhami
Andrew Feber
Andrew Teschendorff
Chrissie Thirlwell
Gareth Wilson

ICMS
Vardhman Rakyan

Gurdon
Thomas Down

EBI
Paul Flicek

Sanger
Dan Turner

CRI
Simon Tavaré

Illumina
Gary Schroth
Zhang Lu

IBD
Liselotte Bäckdahl
Philip Rosenstiel
Gareth Wilson
Andrew Teschendorff
Thomas Down
Robert Haesler
Stefan Schreiber

NF
Andrew Feber
Nadege Presneau
Bernadine Idowu
Gareth Wilson
Adrienne Flanagan

T1D
Vardhman Rakyan
Huriya Beyan
Mohammed Hawa
Thomas Down
Siarhei Maslau
David Leslie

NET
Chrissie Thirlwell
Martyn Caplin
Brian Davidson
Tim Meyer
Tom Kurzawinski
Steve Periera
Andrew Teschendorff
Daniel Hochhauser

T2D
Chris Bell
Sarah Finer
Cecilia Lindgren
Gareth Wilson
Andrew Teschendorff
Graham Hitman
Vardhman Rakyan
Panos Deloukas
Pelin Akan
Andrew Hattersley
Tim Frayling
Mark Walker
Mike Sampson
Ian Morison
Mark McCarthy