Infinium Methylation 450K: Array overview and clinical utilization

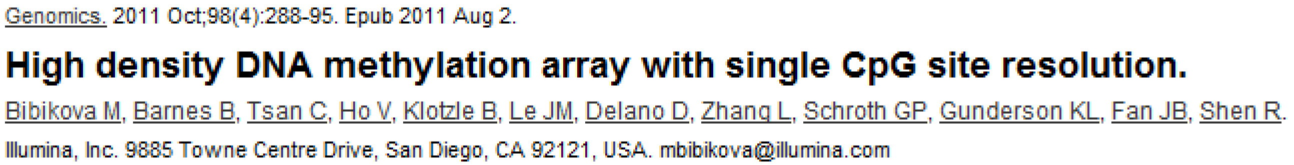
The 3rd Epigenome Informatics Workshop

Bekim Sadikovic, PhD
Assistant Professor
Molecular and Human Genetics
Baylor College of Medicine

Overview

- Describe the features of the 450K array
- Present some of the first literature published on validation of the array
- Discuss the performance differences between the 2 chemistries on the array
- Overview our current work on clinical validation and database development at the Baylor Medical Genetics Laboratories
- Present some early data

Notable early publications

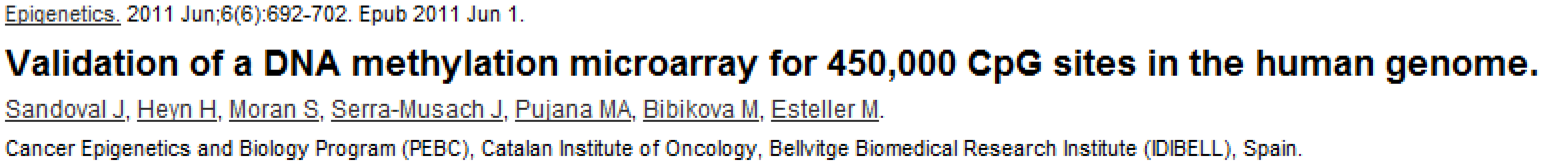


Genomics. 2011 Oct;98(4):288-95. Epub 2011 Aug 2.

High density DNA methylation array with single CpG site resolution.

Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL, Fan JB, Shen R.

Illumina, Inc. 9885 Towne Centre Drive, San Diego, CA 92121, USA. mbibikova@illumina.com

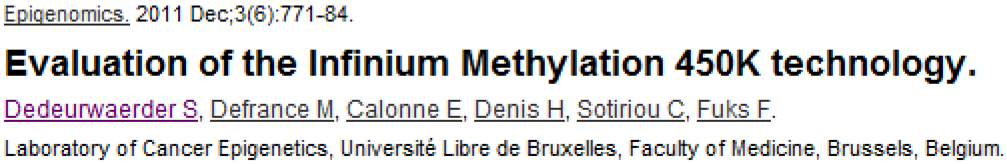


Epigenetics. 2011 Jun;6(6):692-702. Epub 2011 Jun 1.

Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome.

Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, Esteller M.

Cancer Epigenetics and Biology Program (PEBC), Catalan Institute of Oncology, Bellvitge Biomedical Research Institute (IDIBELL), Spain.

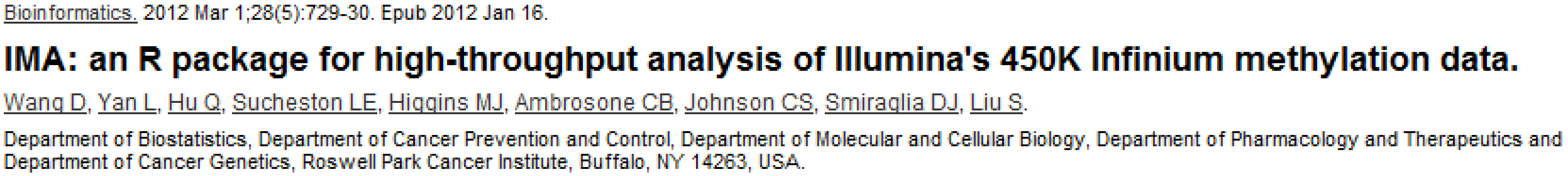


Epigenomics. 2011 Dec;3(6):771-84.

Evaluation of the Infinium Methylation 450K technology.

Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F.

Laboratory of Cancer Epigenetics, Université Libre de Bruxelles, Faculty of Medicine, Brussels, Belgium.



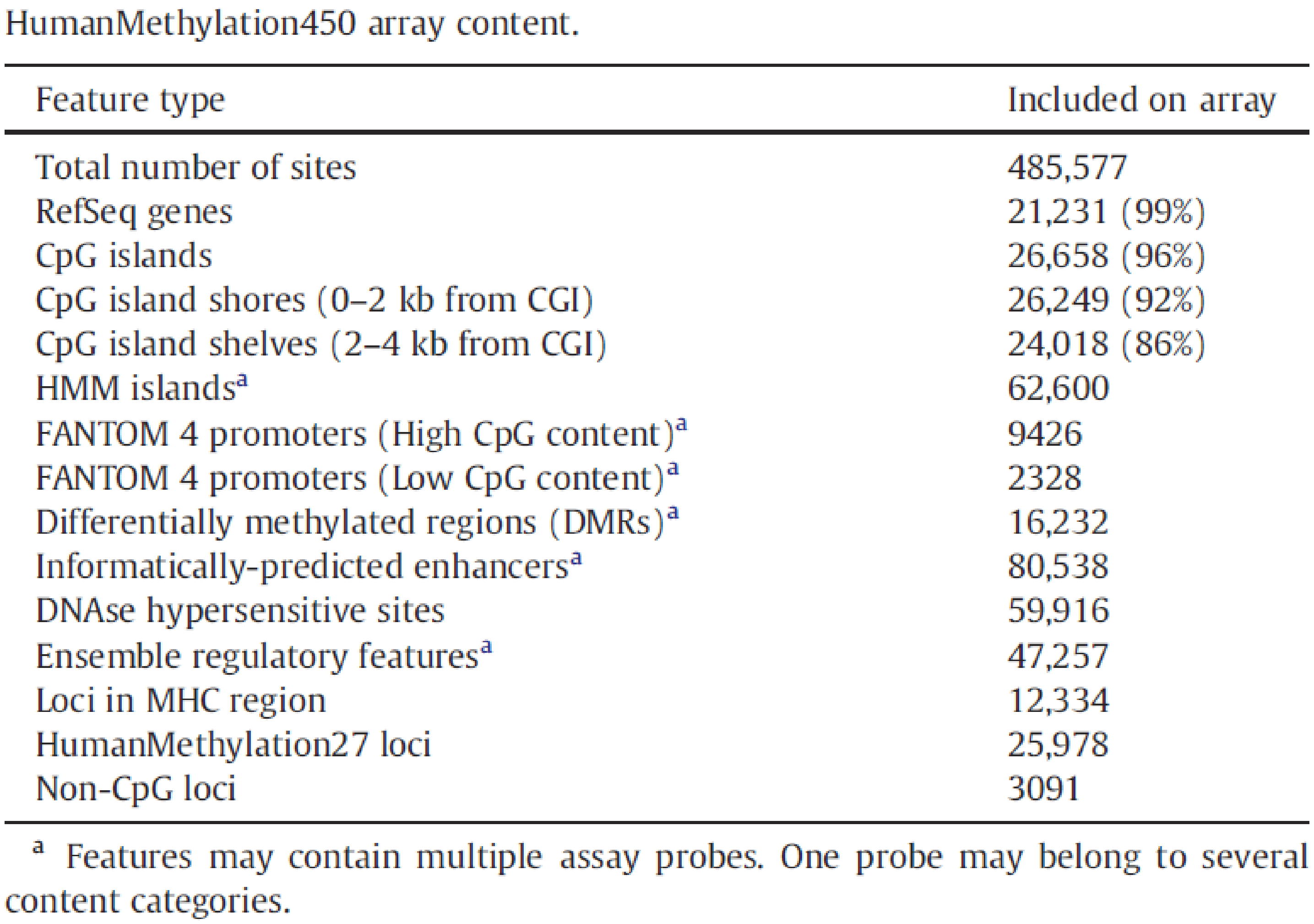
Bioinformatics. 2012 Mar 1;28(5):729-30. Epub 2012 Jan 16.

IMA: an R package for high-throughput analysis of Illumina's 450K Infinium methylation data.

Wang D, Yan L, Hu Q, Sucheston LE, Higgins MJ, Ambrosone CB, Johnson CS, Smiraglia DJ, Liu S.

Department of Biostatistics, Department of Cancer Prevention and Control, Department of Molecular and Cellular Biology, Department of Pharmacology and Therapeutics and Department of Cancer Genetics, Roswell Park Cancer Institute, Buffalo, NY 14263, USA.

Genomic probe distribution



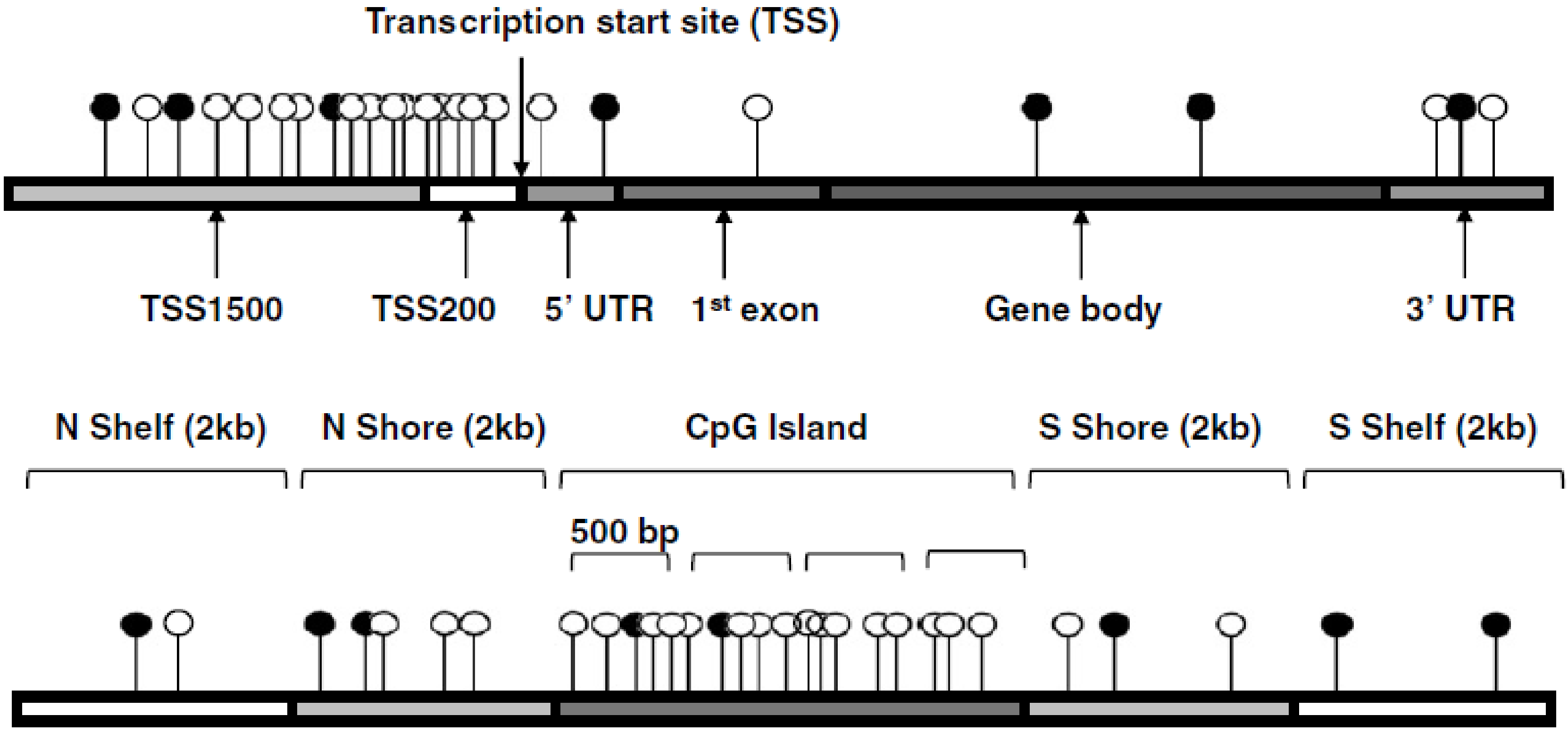
HumanMethylation450 array content.

Feature type	Included on array
Total number of sites	485,577
RefSeq genes	21,231 (99%)
CpG islands	26,658 (96%)
CpG island shores (0–2 kb from CGI)	26,249 (92%)
CpG island shelves (2–4 kb from CGI)	24,018 (86%)
HMM islands[a]	62,600
FANTOM 4 promoters (High CpG content)[a]	9426
FANTOM 4 promoters (Low CpG content)[a]	2328
Differentially methylated regions (DMRs)[a]	16,232
Informatically-predicted enhancers[a]	80,538
DNAse hypersensitive sites	59,916
Ensemble regulatory features[a]	47,257
Loci in MHC region	12,334
HumanMethylation27 loci	25,978
Non-CpG loci	3091

[a] Features may contain multiple assay probes. One probe may belong to several content categories.

Bibikova et al 2011

Gene/CpG island probe distribution



Bibikova et al 2011

Gene/CpG island probe distribution

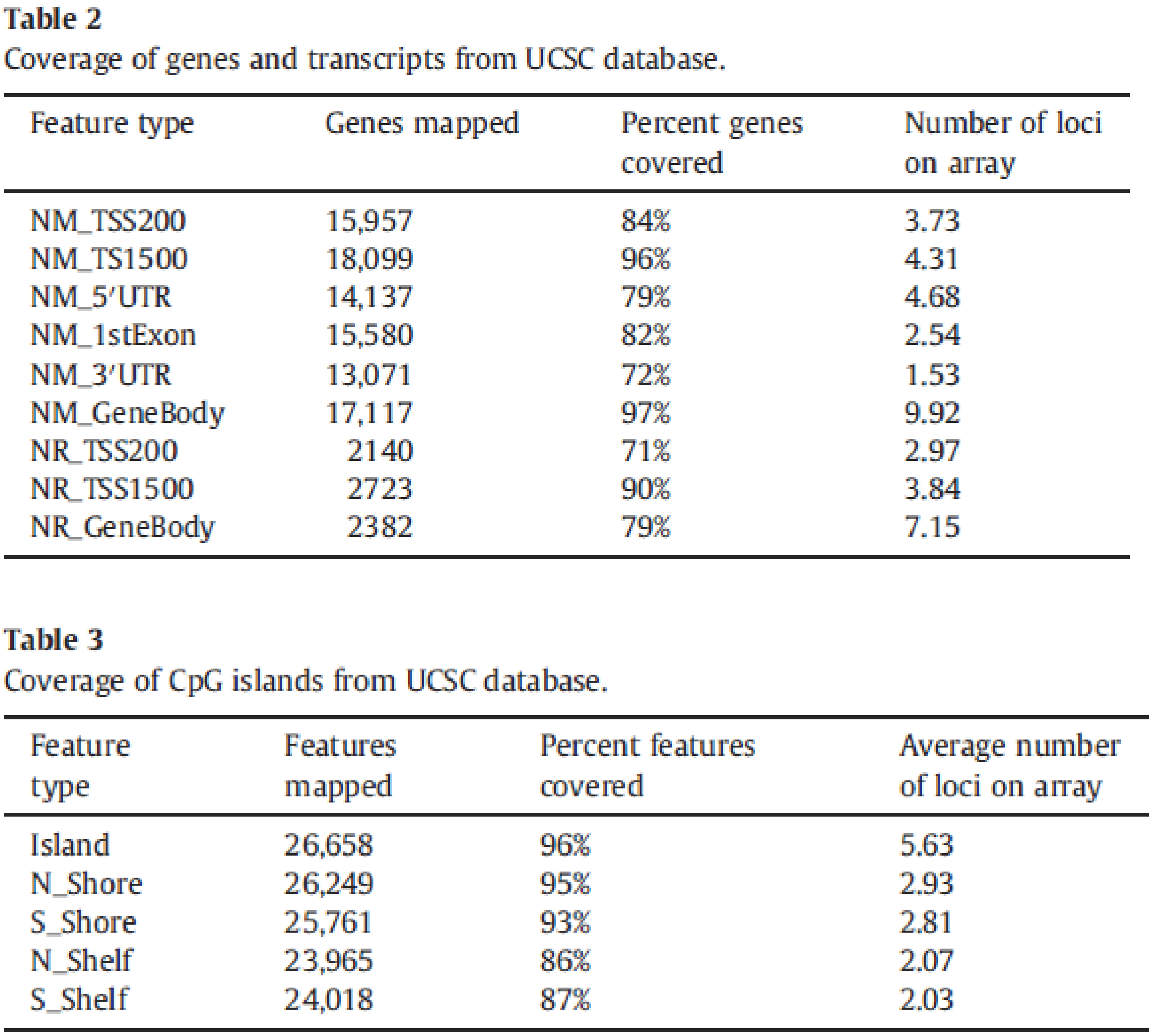


Table 2
Coverage of genes and transcripts from UCSC database.

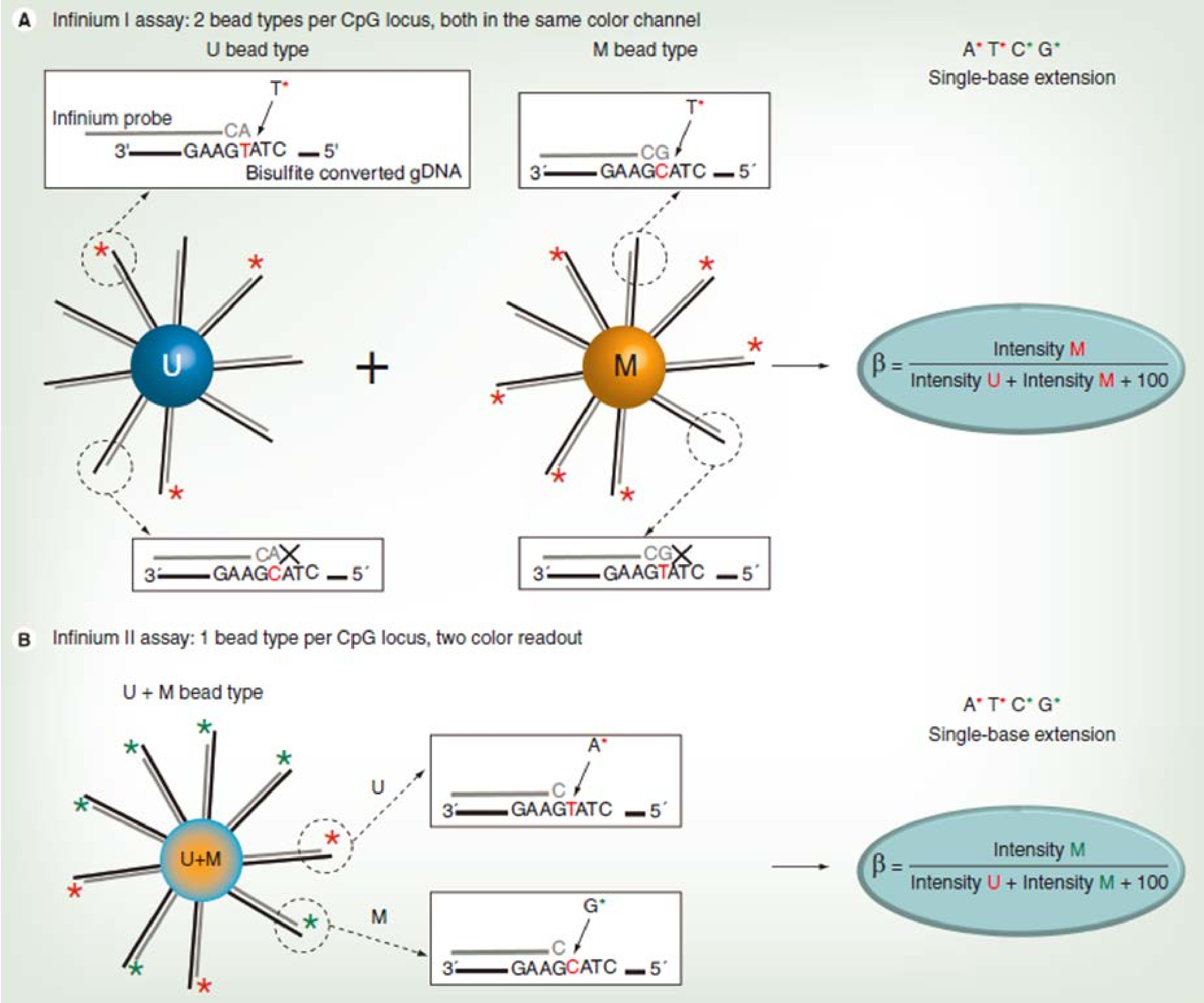
Feature type	Genes mapped	Percent genes covered	Number of loci on array
NM_TSS200	15,957	84%	3.73
NM_TS1500	18,099	96%	4.31
NM_5′UTR	14,137	79%	4.68
NM_1stExon	15,580	82%	2.54
NM_3′UTR	13,071	72%	1.53
NM_GeneBody	17,117	97%	9.92
NR_TSS200	2140	71%	2.97
NR_TSS1500	2723	90%	3.84
NR_GeneBody	2382	79%	7.15

Table 3
Coverage of CpG islands from UCSC database.

Feature type	Features mapped	Percent features covered	Average number of loci on array
Island	26,658	96%	5.63
N_Shore	26,249	95%	2.93
S_Shore	25,761	93%	2.81
N_Shelf	23,965	86%	2.07
S_Shelf	24,018	87%	2.03

Bibikova et al 2011

Infinium chemistry – 2 types on 450K



Infinium I

-1 probe/bead U/M

-2 beads

-2 channels, same color

Infinium II

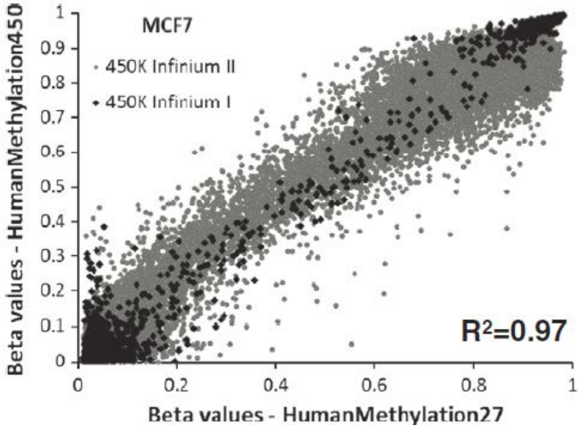
-2 probes/bead (2?)

-1 bead

-1 channel, 2 colors

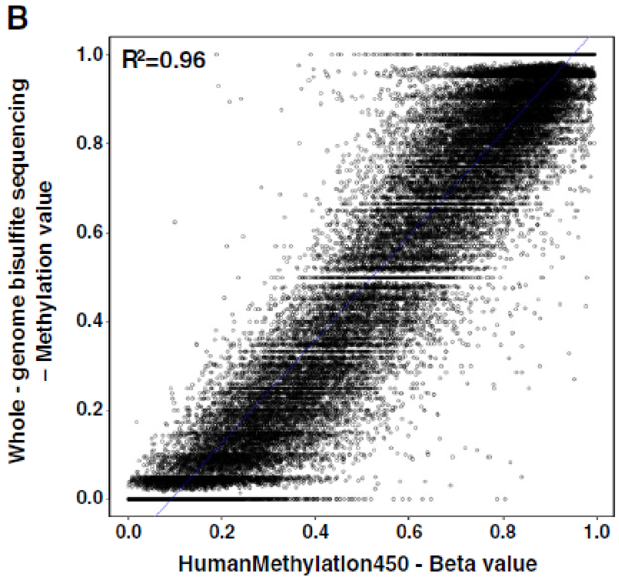
Dedeurwaerder et al 2011

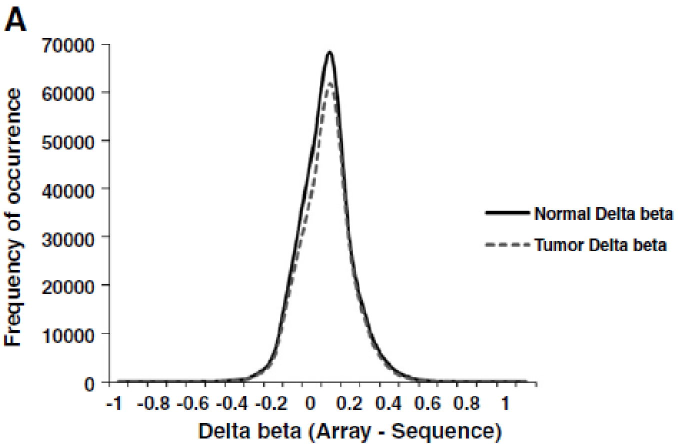
Comparison to 27K and BisSeq



Very good correlation to both

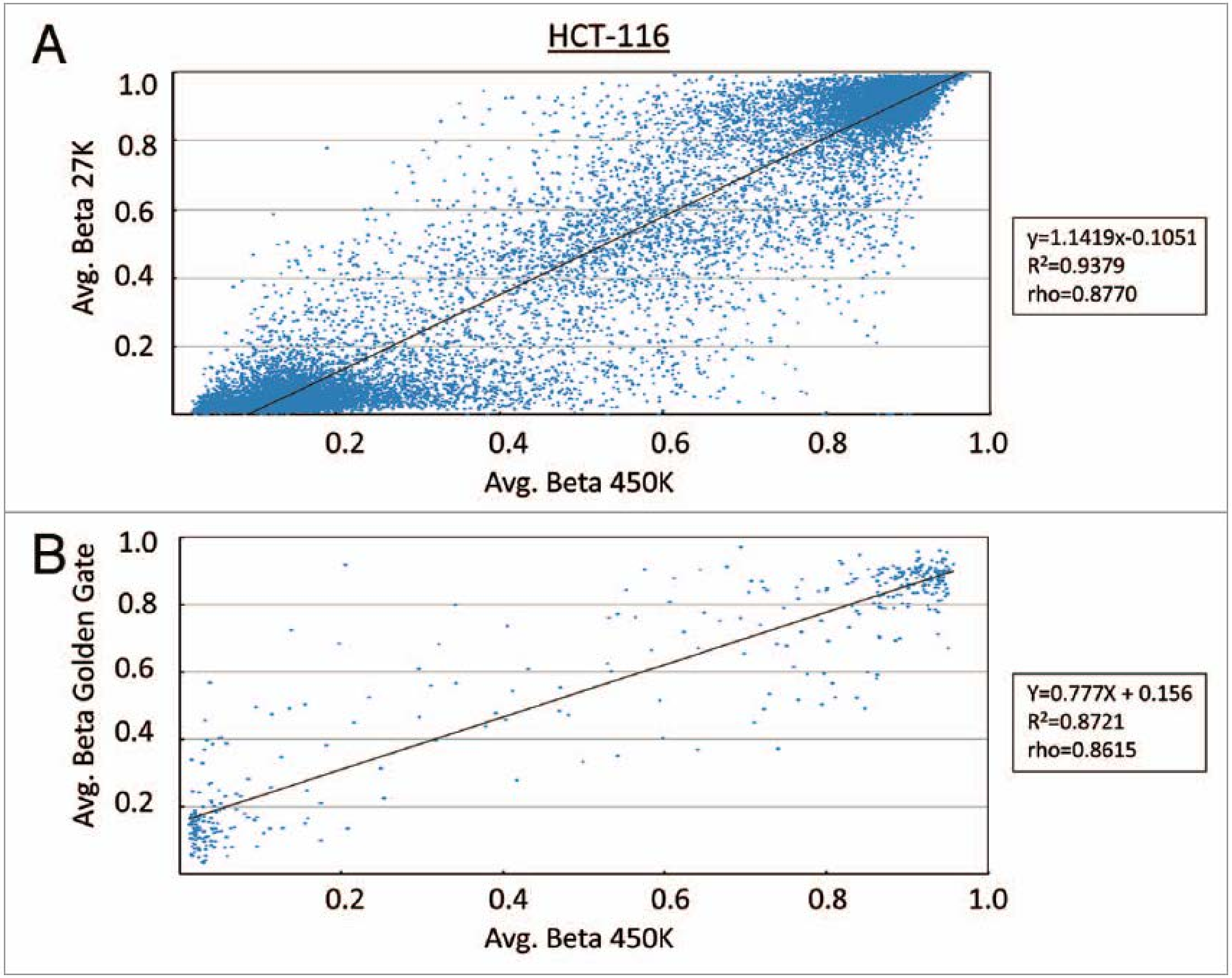
Also, some variation attributed to BisSeq





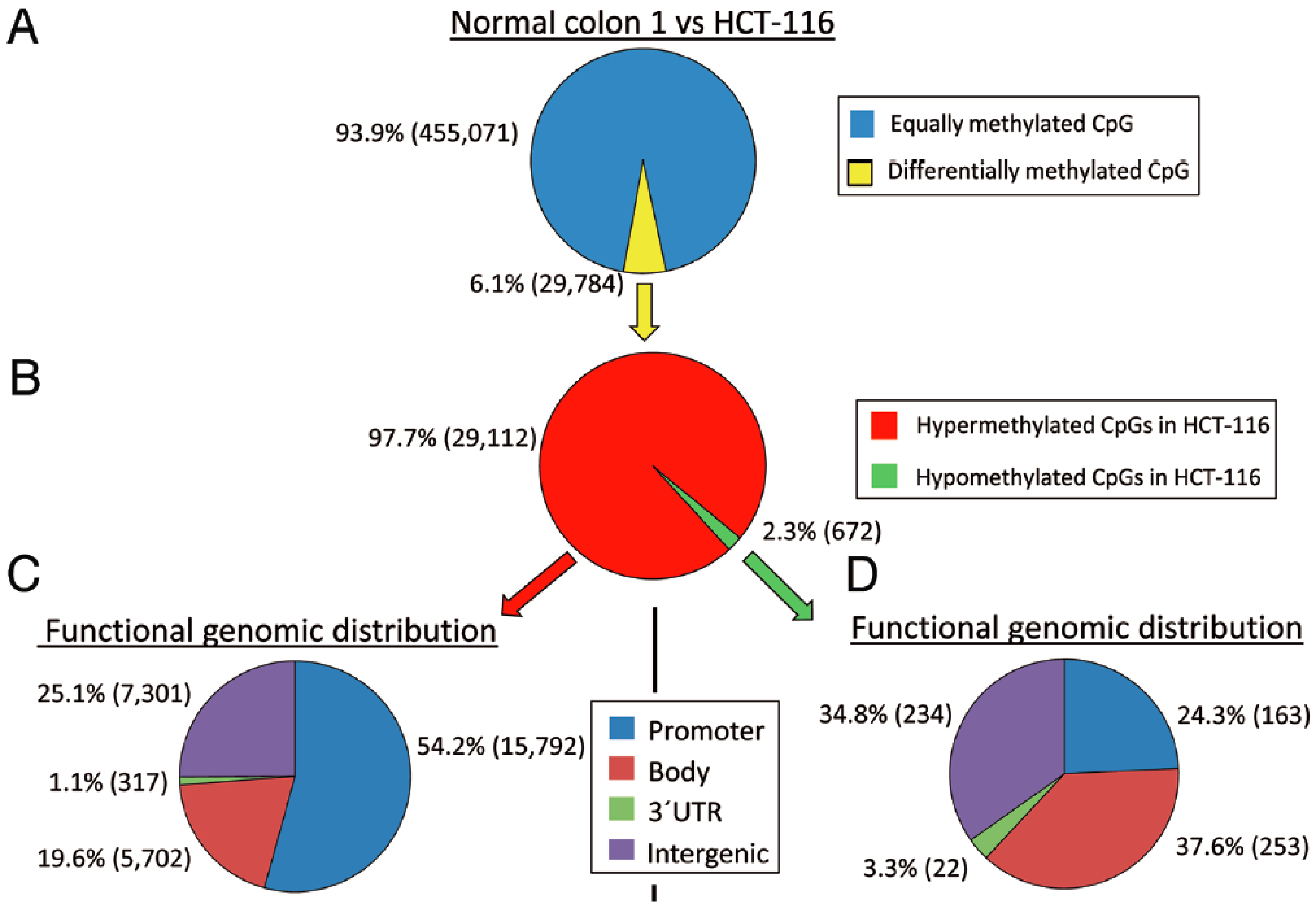
Bibikova et al 2011

Comparison to 27K and Golden gate



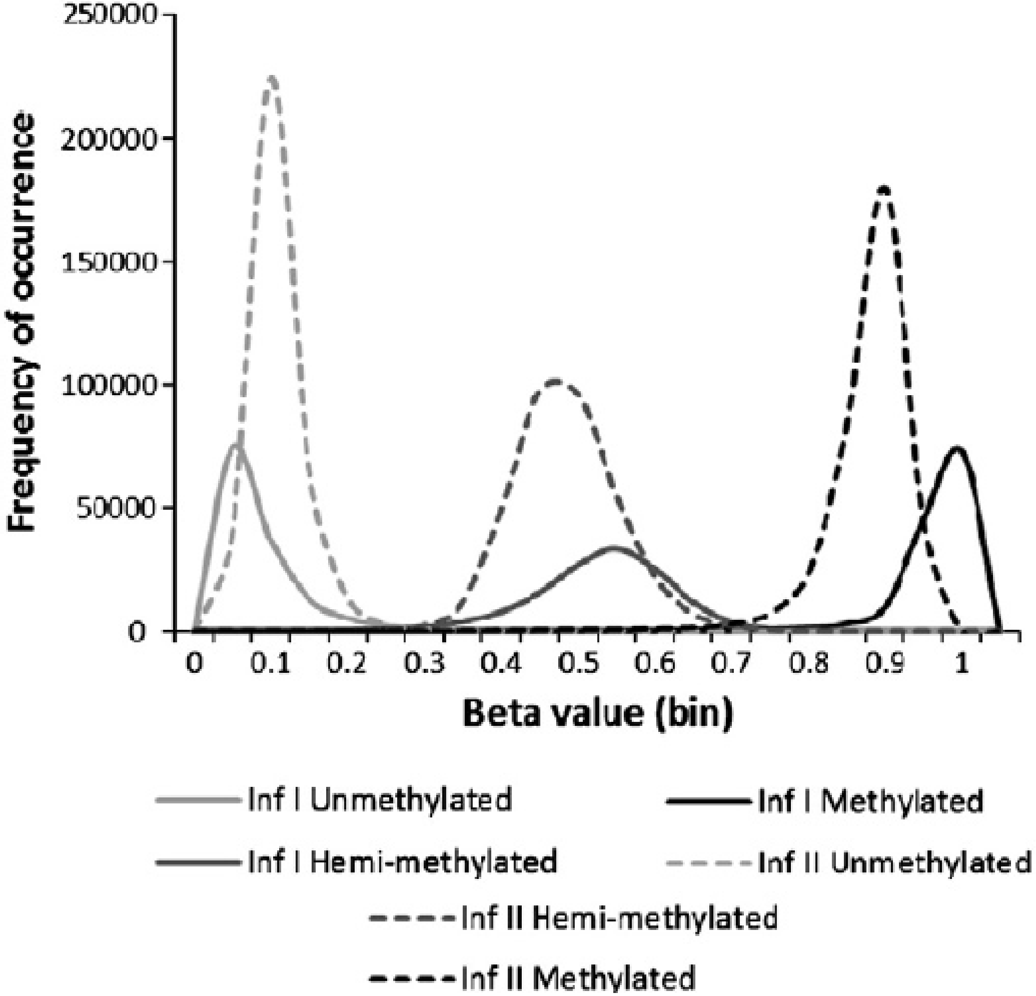
Sandoval et al 2011

450K: cancer vs normal tissue



Sandoval et al 2011

But.. difference in chemistry performance



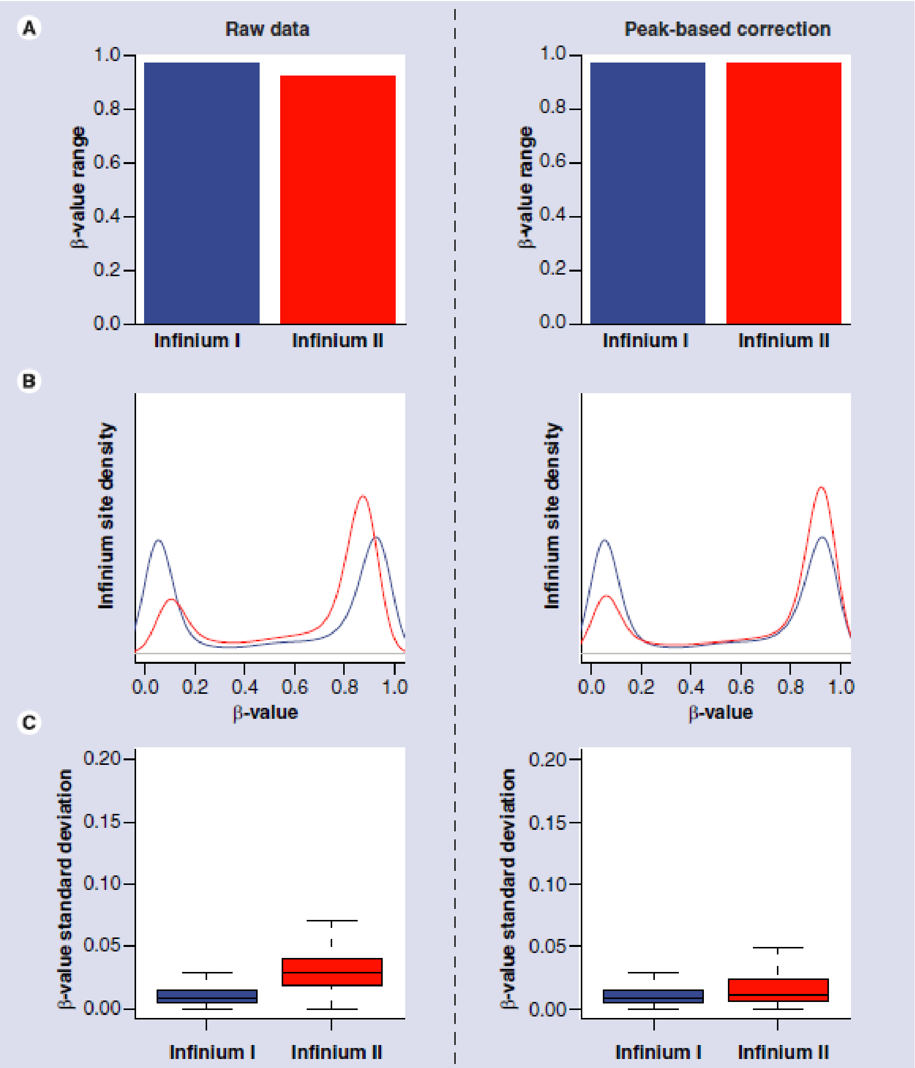
Infinium II has lower dynamic range than Infinium I

i.e. Hypomethylated probes not quite 0, and hypermethylated CpG loci not quite 100

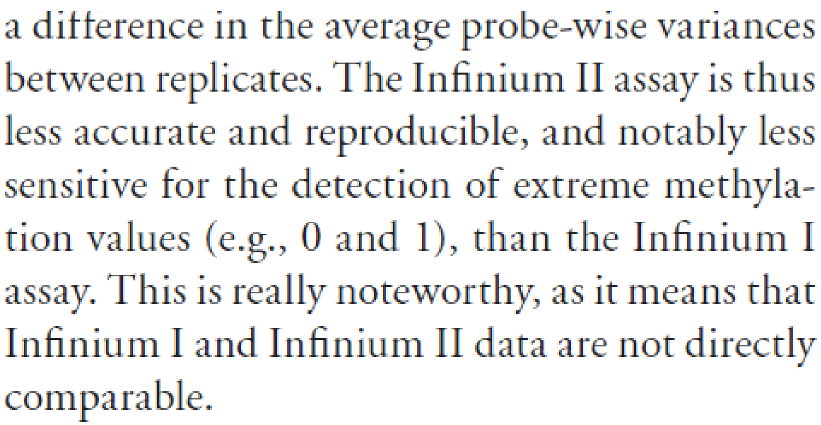
Which is right??

Bibikova et al 2011

Infinium II peak matching algorithm



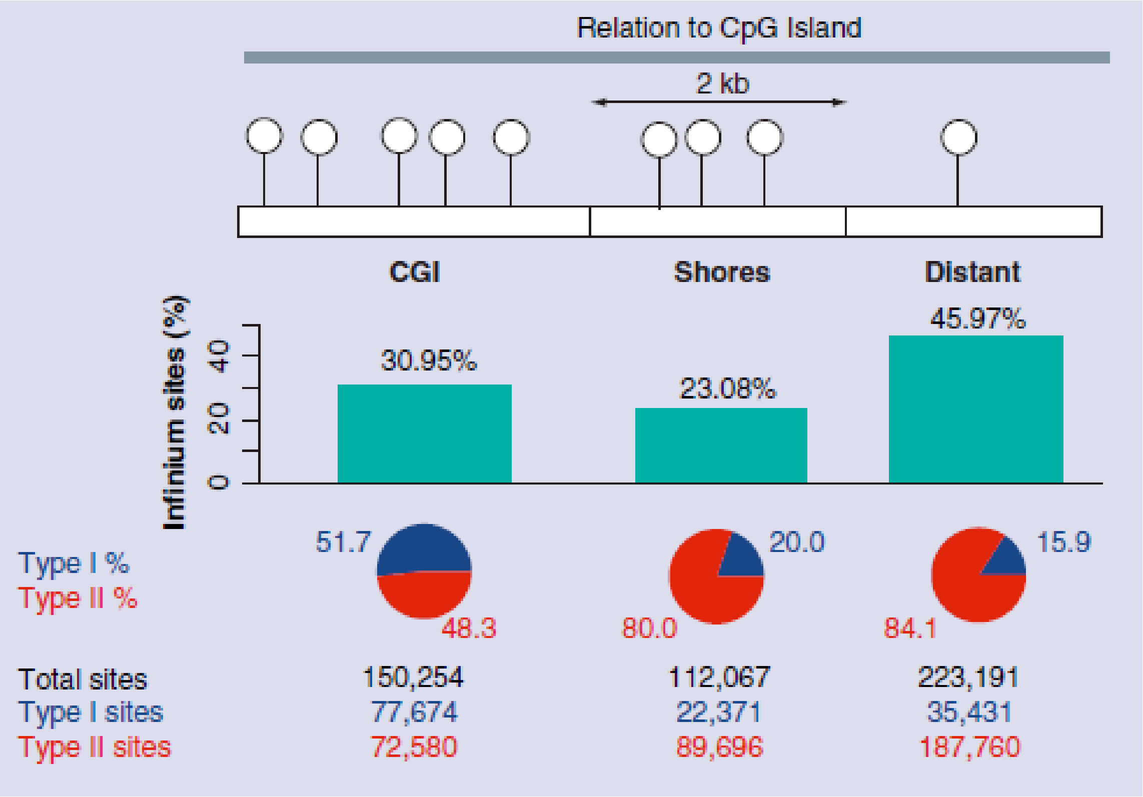
Lower dynamic range, slightly higher standard deviation in signal, and slightly lower accuracy in validation of few loci with BisPyroSeq resulted in conclusion that Infinium II is inferior, hence use algorythm to “normalize” to Infinium I peaks.



a difference in the average probe-wise variances between replicates. The Infinium II assay is thus less accurate and reproducible, and notably less sensitive for the detection of extreme methylation values (e.g., 0 and 1), than the Infinium I assay. This is really noteworthy, as it means that Infinium I and Infinium II data are not directly comparable.

Dedeurwaerder et al 2011

However..

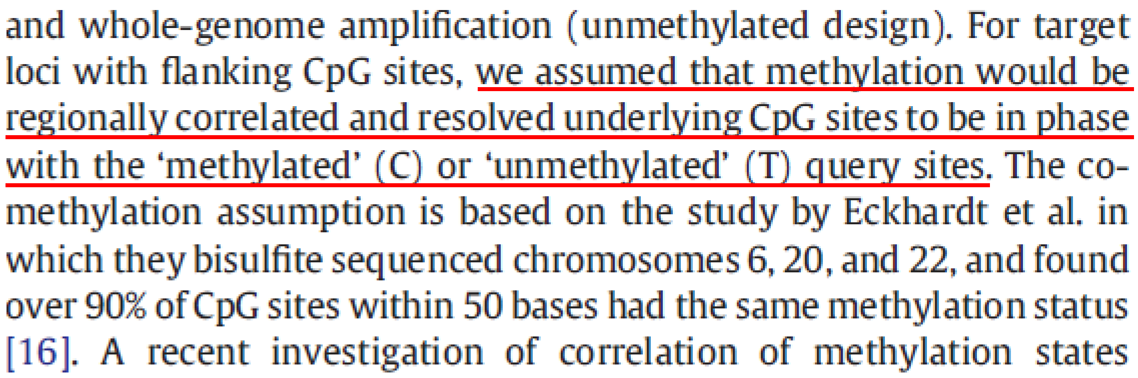


There are probe distribution differences.

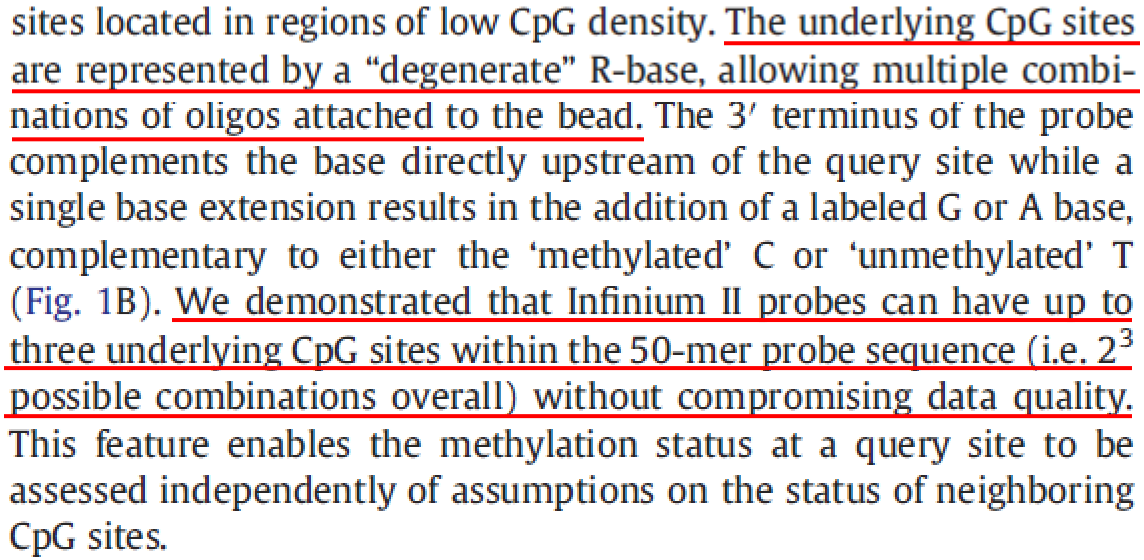
Why?

Dedeurwaerder et al 2011

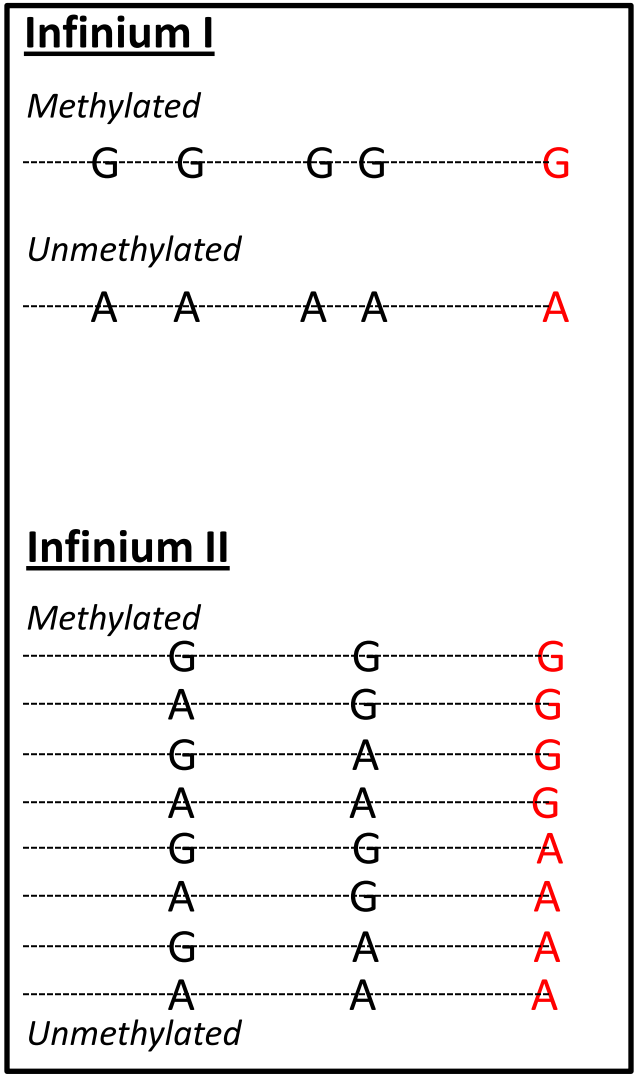
Infinium I designed with *a priori* bias



and whole-genome amplification (unmethylated design). For target loci with flanking CpG sites, we assumed that methylation would be regionally correlated and resolved underlying CpG sites to be in phase with the 'methylated' (C) or 'unmethylated' (T) query sites. The co-methylation assumption is based on the study by Eckhardt et al. in which they bisulfite sequenced chromosomes 6, 20, and 22, and found over 90% of CpG sites within 50 bases had the same methylation status [16]. A recent investigation of correlation of methylation states



sites located in regions of low CpG density. The underlying CpG sites are represented by a "degenerate" R-base, allowing multiple combinations of oligos attached to the bead. The 3′ terminus of the probe complements the base directly upstream of the query site while a single base extension results in the addition of a labeled G or A base, complementary to either the 'methylated' C or 'unmethylated' T (Fig. 1B). We demonstrated that Infinium II probes can have up to three underlying CpG sites within the 50-mer probe sequence (i.e. 2^3 possible combinations overall) without compromising data quality. This feature enables the methylation status at a query site to be assessed independently of assumptions on the status of neighboring CpG sites.



Infinium I

Methylated

G G G G G

Unmethylated

A A A A A

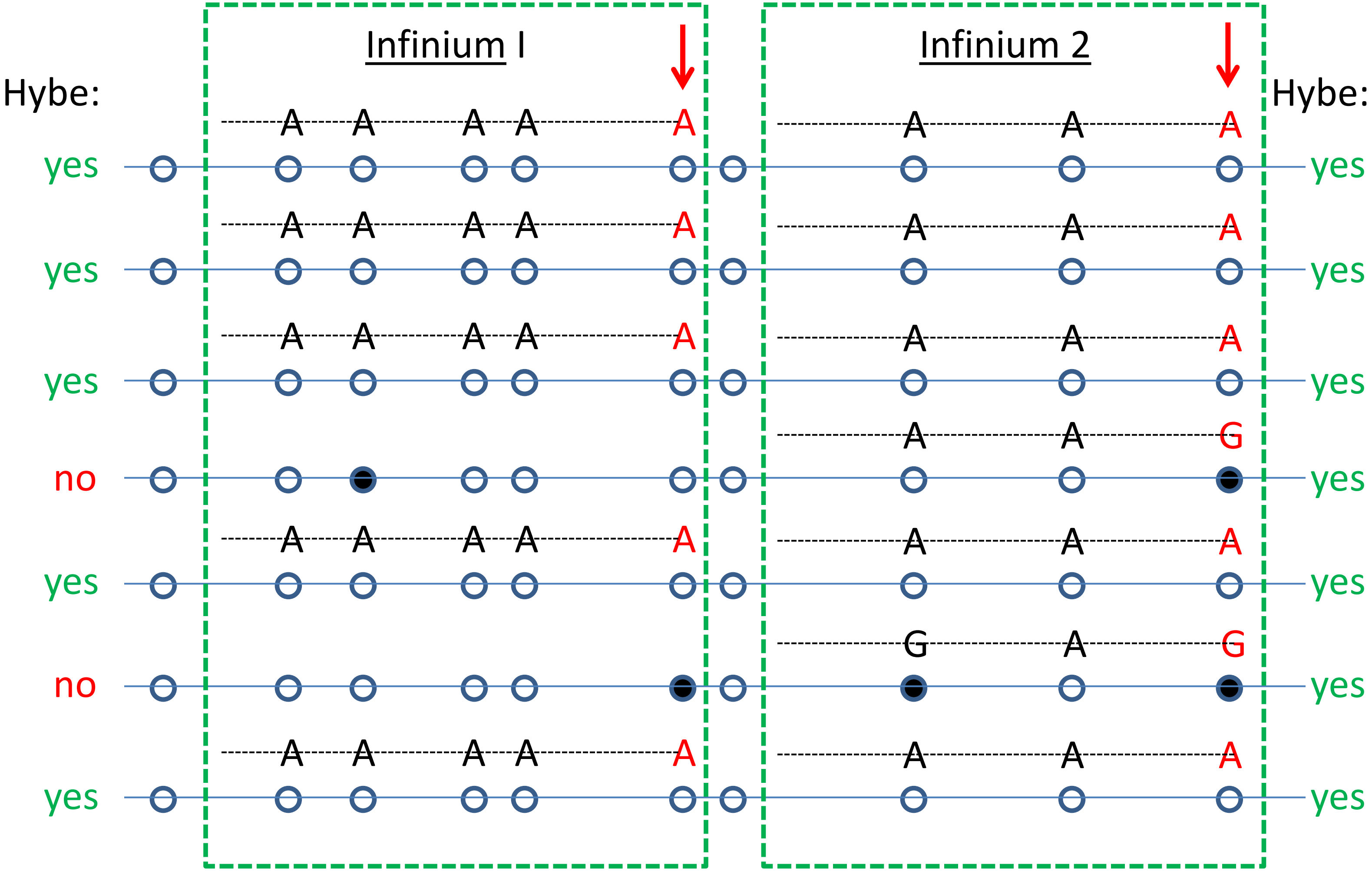
Infinium II

Methylated

G G G
A G G
G A G
A A G
G G A
A G A
G A A
A A A

Unmethylated

An “unmethylated” CpG island



Actual Meth: 85%
Measured: 100%

Actual Meth: 71%
Measured: 71%

To norm or not to norm?

- Methylation differences a few percent (av. 3-5%)

- St. Dev. between measurements ~3% (I), ~7% (II)
  - Variation possibly due to bead design in II: up to 8 different probes/1 bead, bead production variability?, also multiple target DNAs-hybe bias?

- Lower dynamic range in II:
  - May actually represent closer to true biological state

- "*at the end it doesn't really matter*" – Linkin Park
  - Biased by my clinical work, but unless near 50% meth difference, don't talk to me about an effect on gene expression
  - Maybe slight biases in Infinium I vs II, in opposite direction, together result in closer to actual biological state?

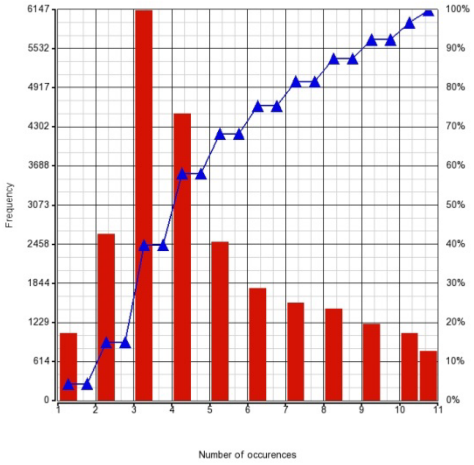
450K – a clinical diagnostic array

- Baylor Medical Genetics Labs to launch 450K as a clinical diagnostic array
- Prelim clinical validation studies completed:
  - 150 pediatric peripheral blood samples processed:
    - Various imprinting/UPD disorders, normal ctrls, ped. cancer + MCA, autism
  - 150 brain tissue:
    - Ped. autism, adult schiz. and bipolar, normal ctrls.
  - 20 Embrionic stem cells
- Ongoing studies:
  - 400 ped. peripheral blood samples
    - Ctrls, imprinting/UPD, CMA negative, others (suspected epigenetic etiology)
  - Adult cancer: 1-200 each: Colon, MDS, Breast, normal tissue ctrls
- Goal: to offer as a Tier I test for suspected epi. etiology, and as augment/follow up to negative CMA, genome seq.

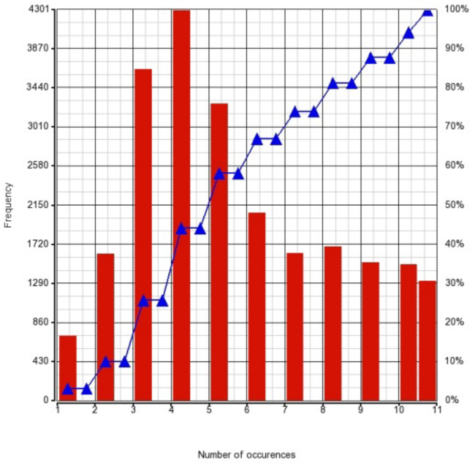
Clinically important technical considerations

Probe coverage

CpG island

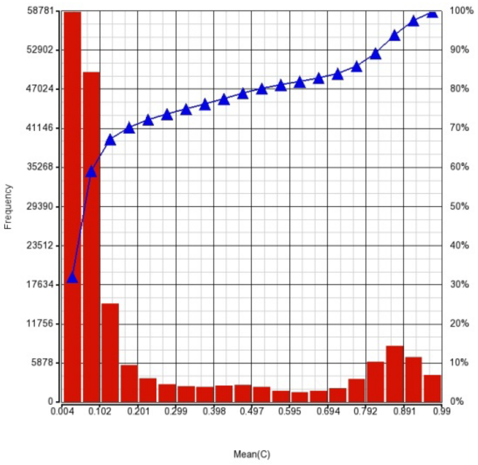


CpG island +200bp

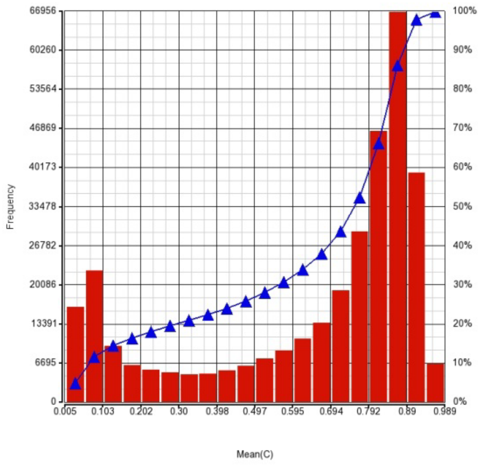


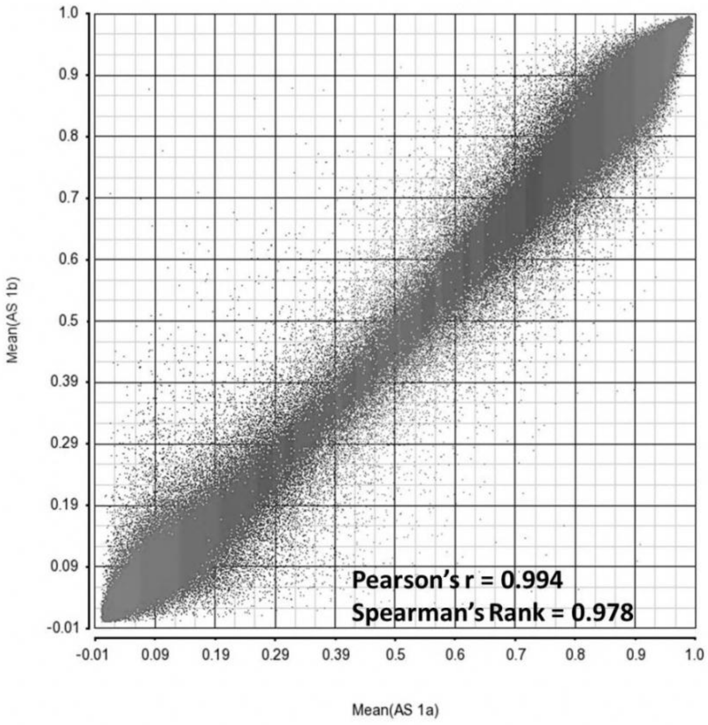
Methylation levels

CpG island



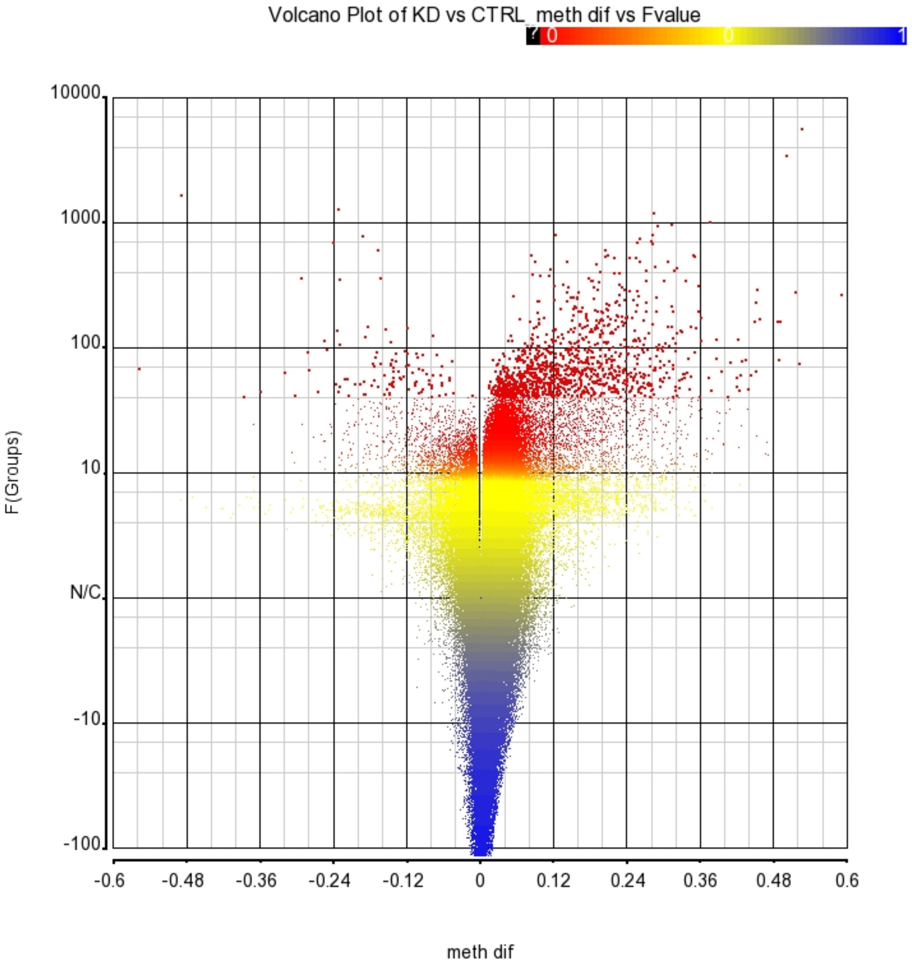
Non-CpG island





Highly reproducible FULL technical reps in patient samples

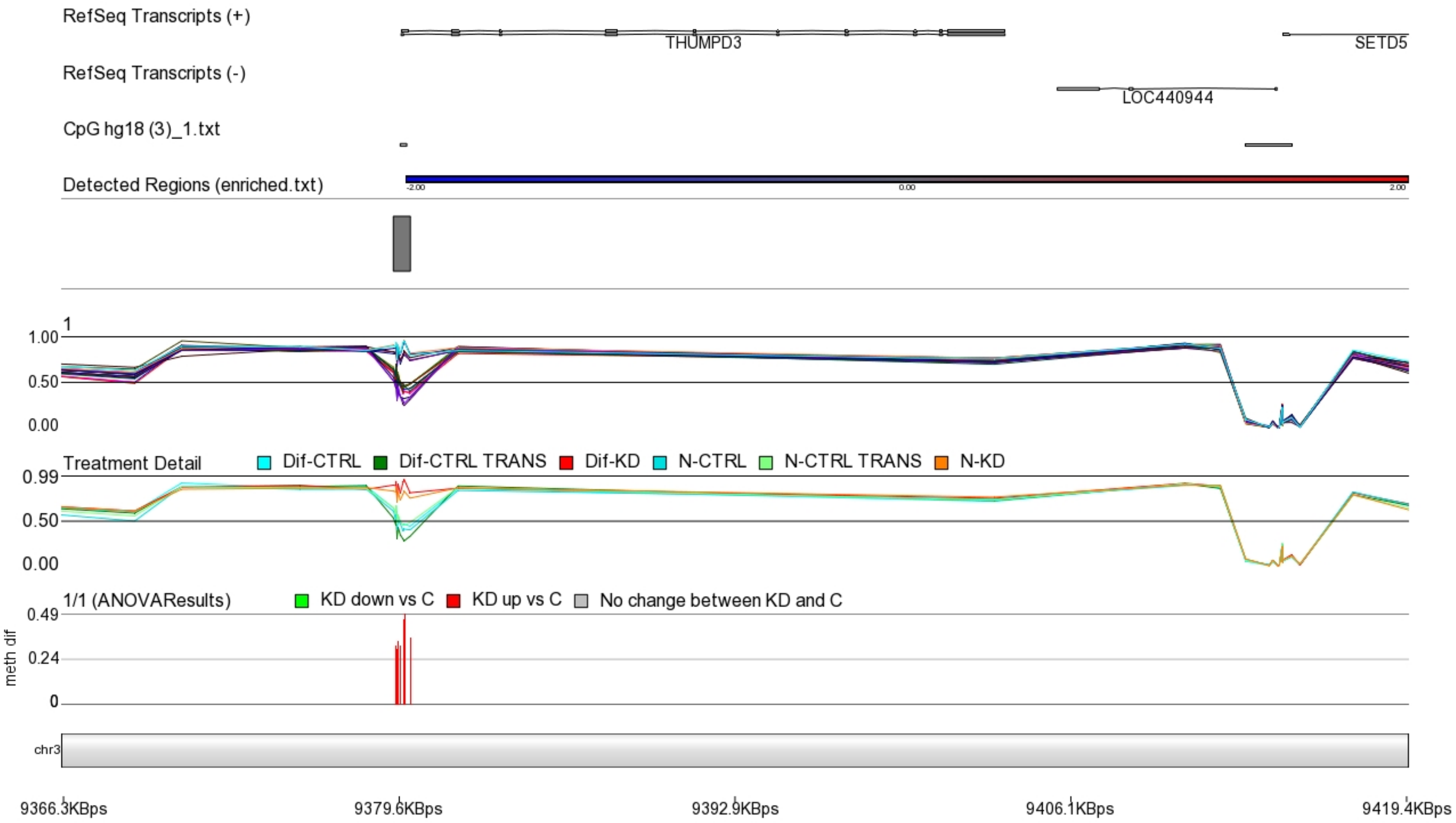
A good algorithm requires



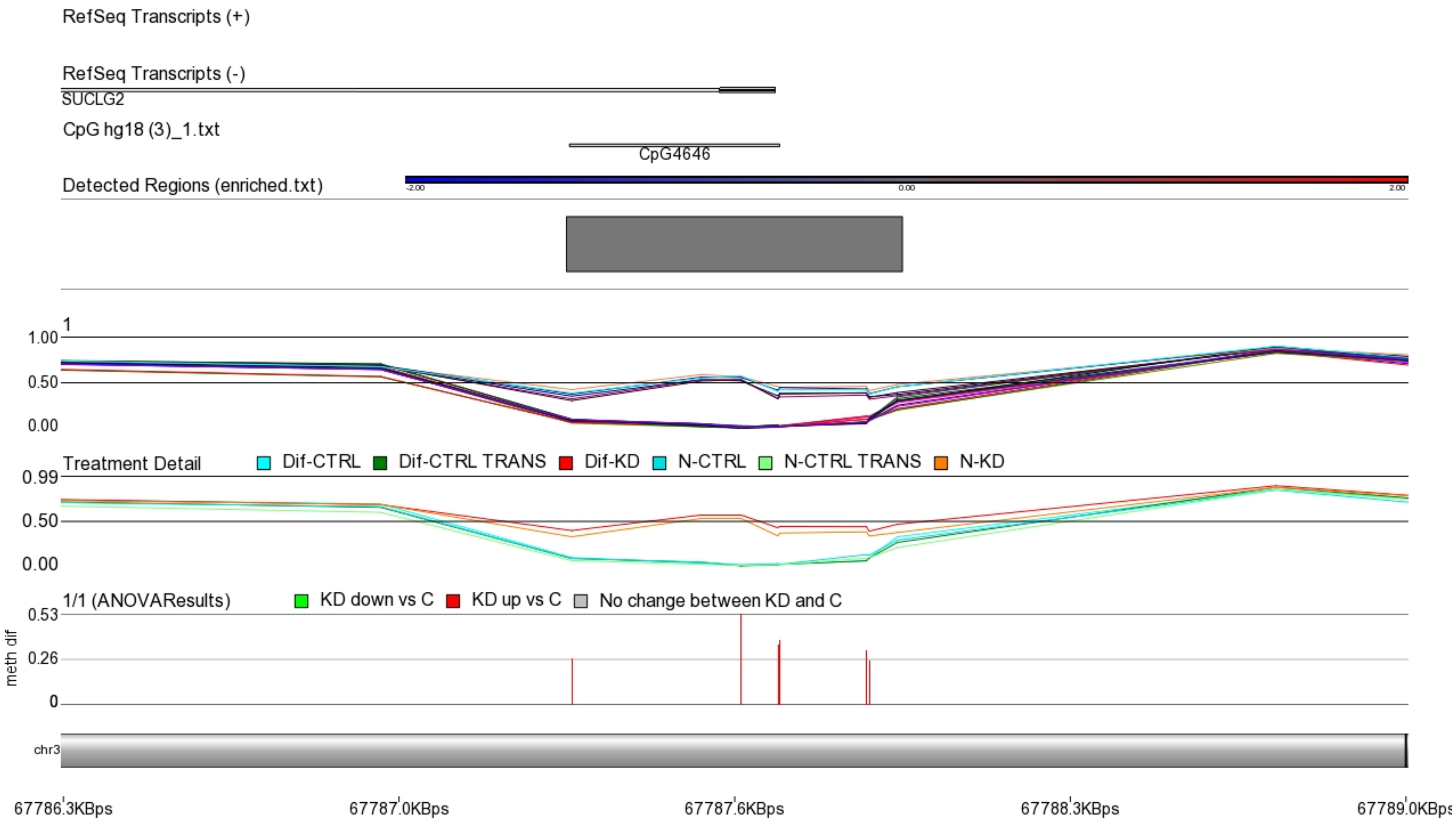
1. Statistics/Informatics
   - “hard”:
     - P-value
     - Meth difference
     - Signal to noise
   - “soft”
     - CpG island overlap
     - Gene overlap

2. Large control database
   - Tissue, age

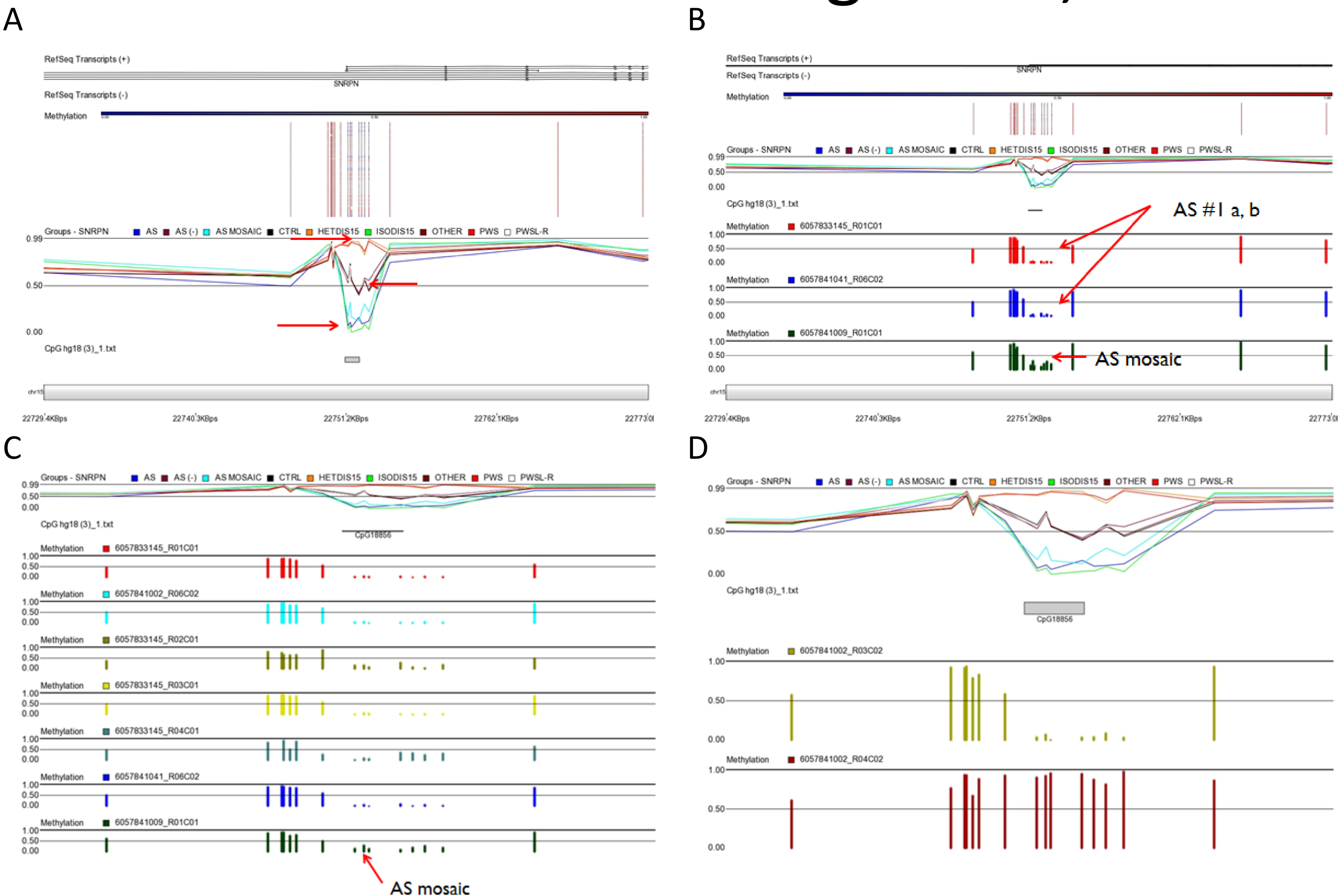
Ex – ES experiment



Ex – ES experiment



Clinical validation – Angelman, UPD

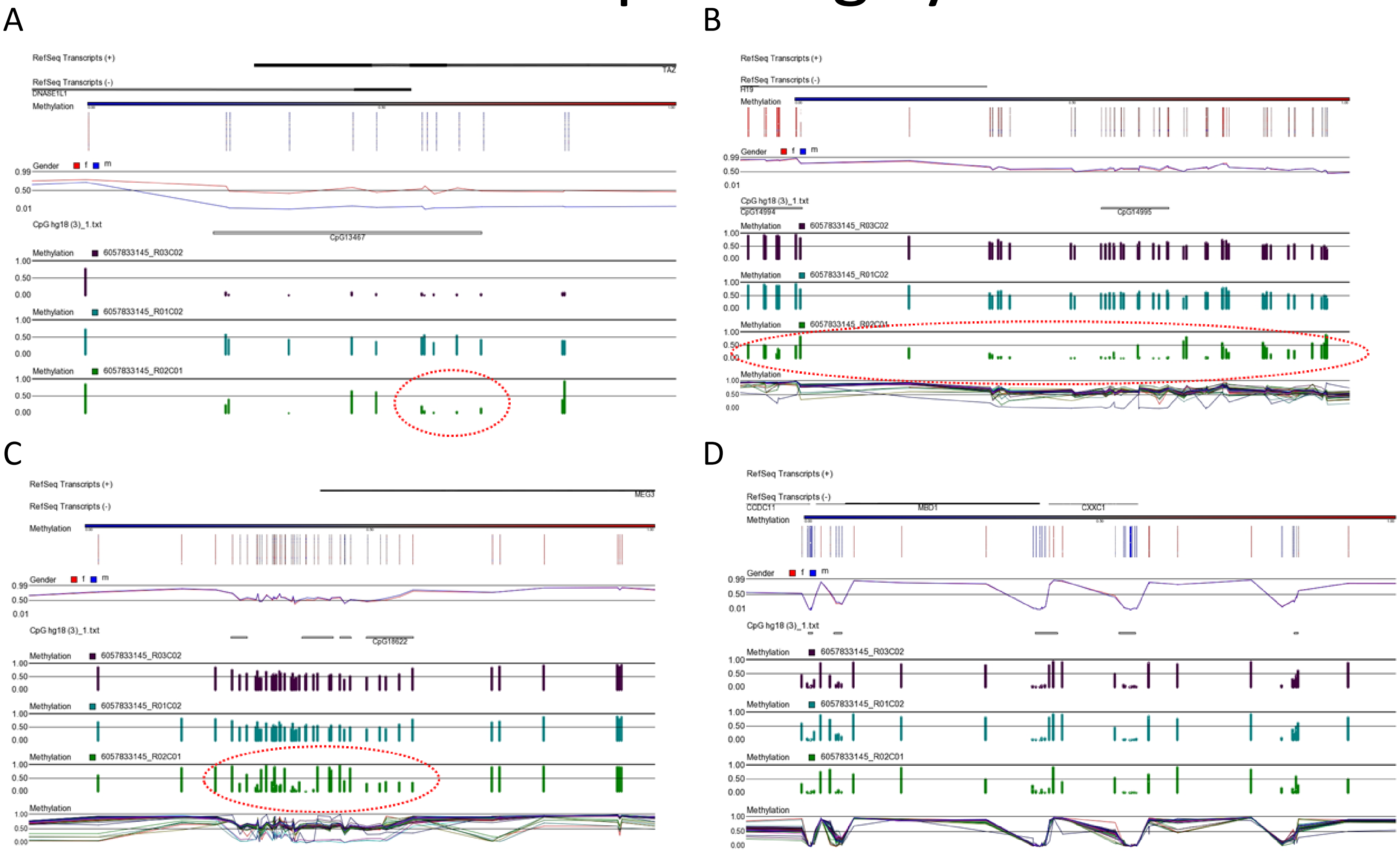


AS #I a, b

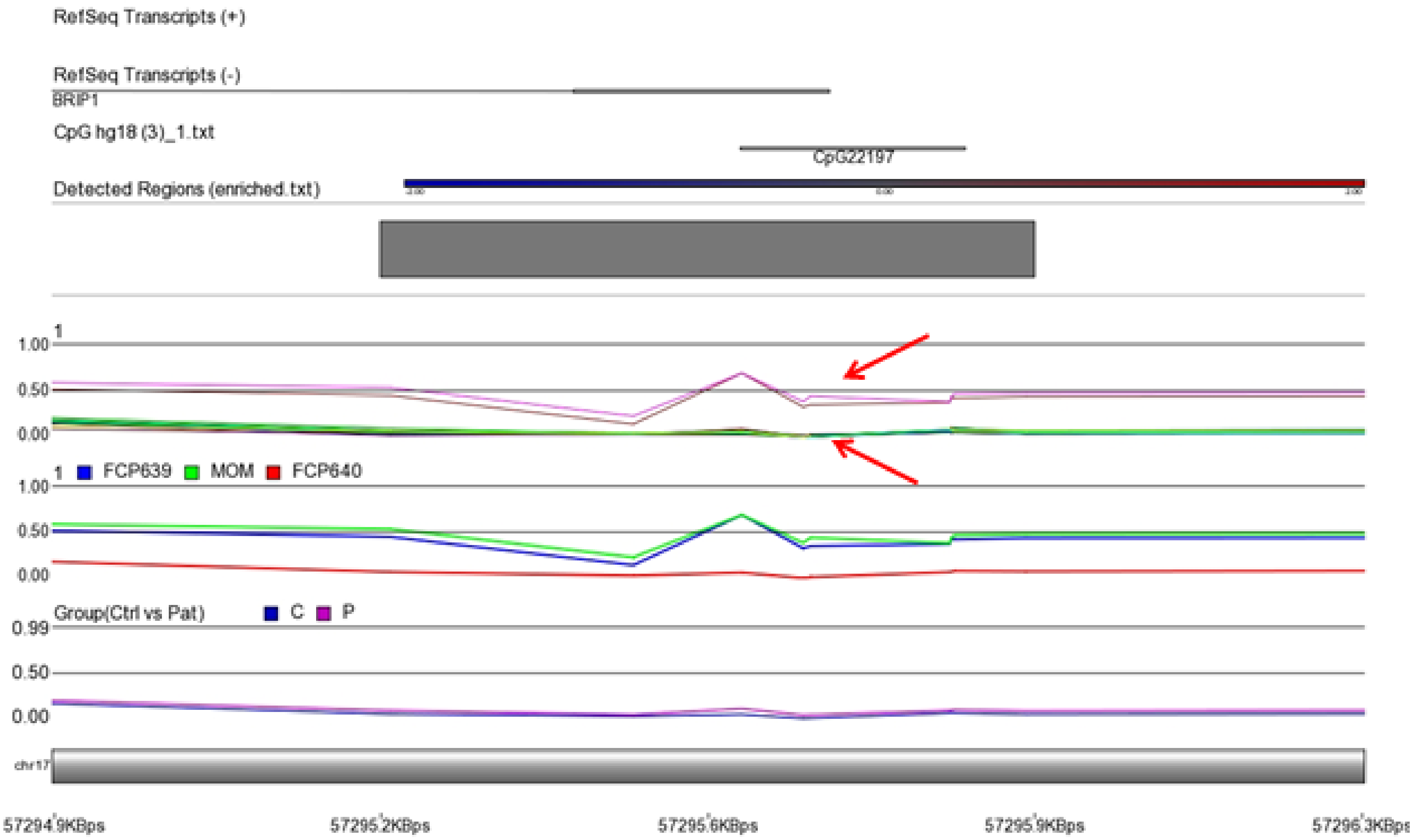
AS mosaic

AS mosaic

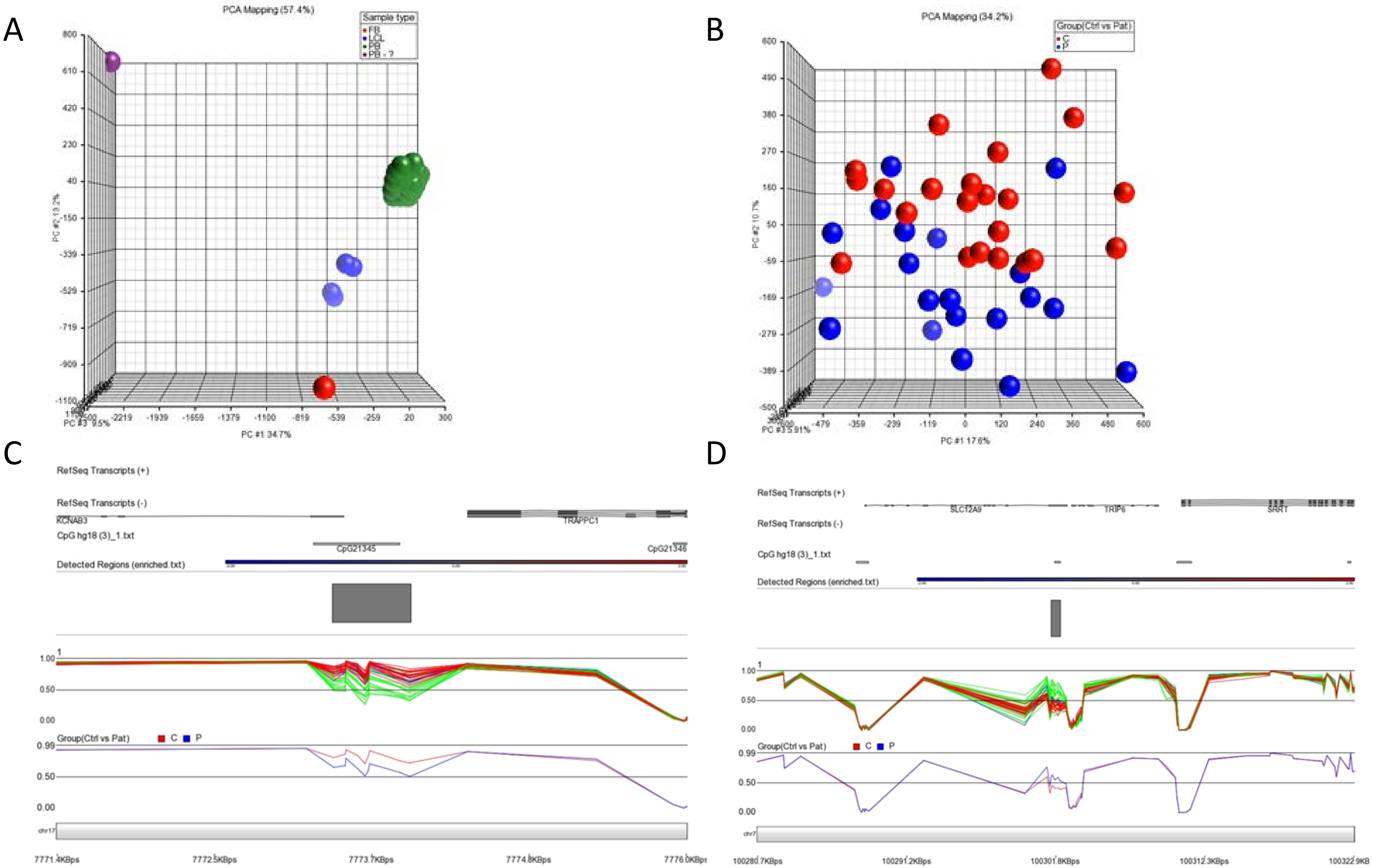
Multiple epigenetic defects in patients with known imprinting syndromes



Epigenetic lesions in known genetic syndrome loci in pediatric patients



Dealing with tissue, age, and inter-individual variation



Acknowledgements

- Art Beaudet Lab
- Sharon Plon Lab
- Igna Van den Veyver Lab
- Joanna Viszniewska
- MGL Lab