DNA Methylation Profiling using MSNP

Venkatraman E. Seshan

Department of Biostatistics / HICCC
Columbia University

June 16, 2009

Statistical Genomics
Institute of Mathematical Sciences
National University of Singapore

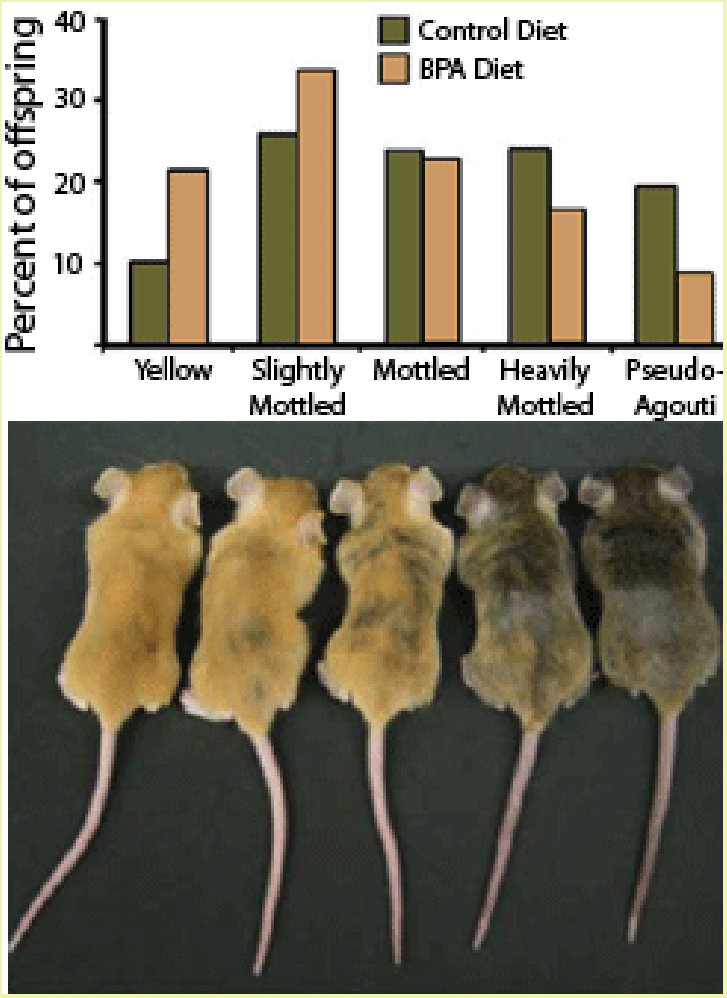
Acknowledgement: Ben Tycko, MD, PhD

DNA: A (adenine), C (cytosine), G (guanine), T (thymine)

- **Epigenetics addresses the question: if the DNA sequence in all the cells are the same how do they act differently?**
- **DNA Methylation is one of the epigenetic mechanisms.**
- **Involves attaching a methyl group to DNA mostly cytosine.**
- **Converts cytosine to 5-methylcytosine which can't be read.**
- **Methylation is faithfully reproduced during cell division.**
- **Cancer - imbalance - global hypo & CpG island hyper**

  hypo-methylating drug decitabine.

Bisphenol-A pre-natal exposure, Dolinoy et al PNAS 2007



- **In mammalian genomes methylation occurs**

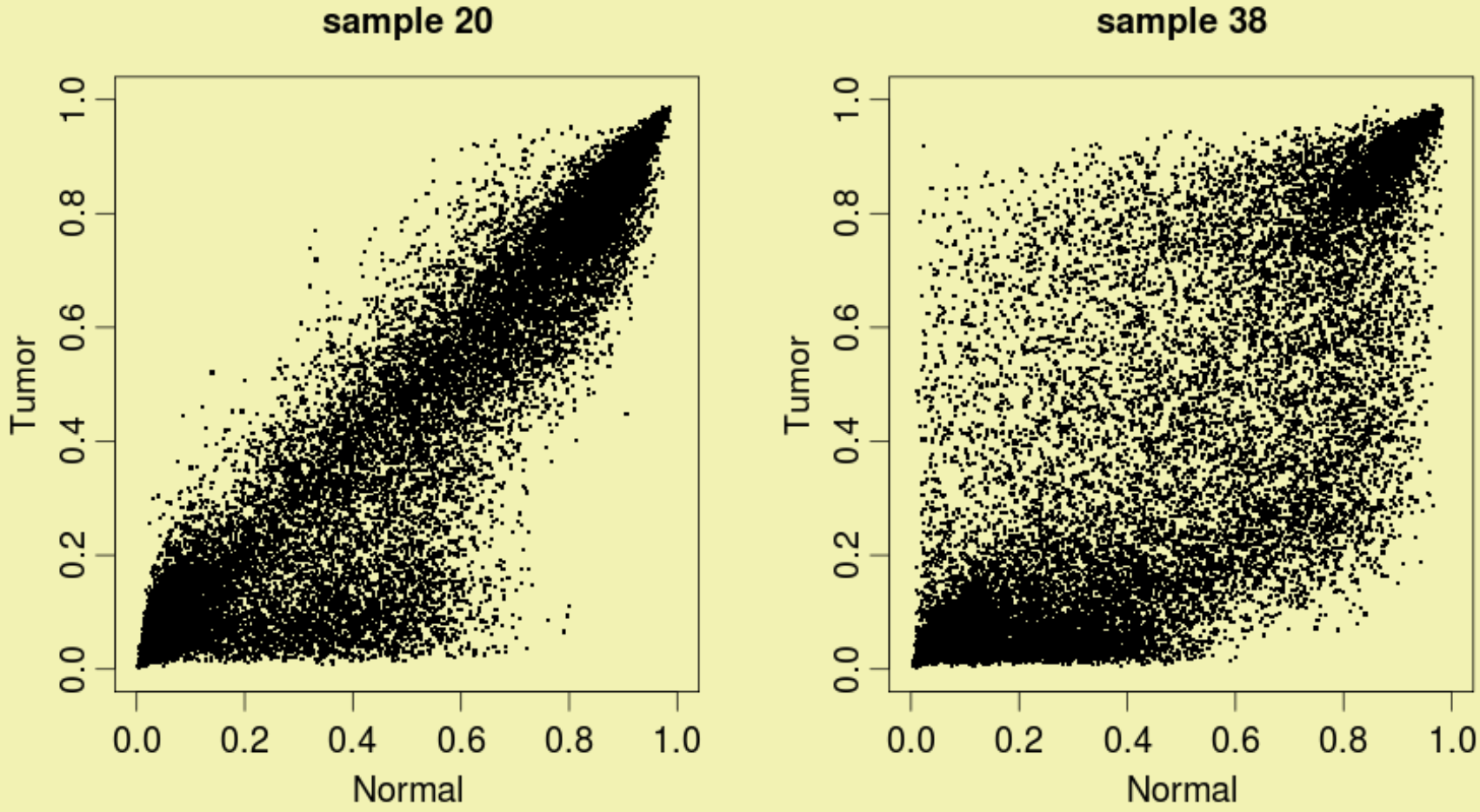
  almost exclusively in CpG (p = phosphodiester bond).

- **CpG less frequent than product probabilities ($\sim 4\%$).**
- **Methylation: transition mutations C $\leftrightarrow$ T (and A $\leftrightarrow$ G).**
- **CpG islands are DNA stretches (>500 bp) with high GC content and greater than expected frequency of CpG.**
- **coding regions of genome methylated although CpG poor.**
- **CpG common in promoter regions but unmethylated.**

- **Bisulfite conversion:** sodium bisulfite treatment of DNA converts unmethylated cytosine to uracil. PCR amplification replaces uracil with thymine. (e.g. Illumina)
- **Methylation sensitive restriction enzymes** are used to enrich for either methylated or unmethylated DNA which are then used on micro arrays. (Affy U133, Agilent/Nimblegen aCGH, Affy SNP arrays)
- **Affinity purification** purify methylated DNA using methyl binding domain (MBD) that bind to methylated CG.

Ref: Zilberman and Henikoff (*Development* 2007)

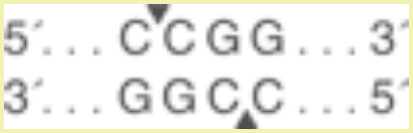
- Sample is first bisulfite converted, then fragmented and finally whole genome amplified.
- Bead bound probes used to detect presence of C (methylated) and T (converted from unmethylated C).
- 27578 pairs of probes corresponding to methylation loci.
- The ratio of C to the sum is the methylation proportion.
- Normal and tumor sample done separately to measure differential methylation.



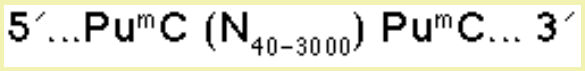
Direct sequencing methods (454, Solexa, Solid) use this approach

- Cuts DNA at specific recognition nucleotide sequence.
- It can be methylation sensitive i.e. uncut if C is methylated in recognition sequence.

Hpa-II/Msp-I



McrBC



Hpa-II sensitive to CpG methylation; Msp-I not

Enzymes used in in affymetrix arrays

Nsp-I

Sty-I

Xba-I

Single letter code: R = A or G; Y = C or T; W = A or T; N = A or C or G or T

- DNA isolated and fragmented with Nsp-I or Sty-I.
- Fragmented DNA is ligated with adaptor.
- Undergoes whole genome amplification.
- Hybridized to SNP chip:
  - ⋆ 500k arrays are one 250k chip each with Nsp-I and Sty-I.
  - ⋆ GWS5 and GWS6 are single chips with mixture of equal quantities of Nsp-I and Sty-I processed DNA.
- Scanned, analyzed ...

How does a SNP array work?

- DNA fragments are defined by restriction enzymes sites.
- Adaptor is recognition sequence specific.
- Amplification of DNA fragments defined above.
- Those containing target SNP or CNV hybridize to probes.
- Intensities dependent on:
  - ⋆ relative quantity of DNA - SNP, CNV.
  - ⋆ probe affinity.
  - ⋆ non-specific hyb.

SNP array for profiling methylation

- SNP array(s) with Nsp-I/Sty-I treated sample.
- SNP array(s) with Nsp-I/Sty-I +Hpa-II treated sample
- SNP array with Nsp-I/Sty-I +Msp-I treated sample.
- Hpa-II/Msp-I sites between two Nsp-I/Sty-I sites.
  - ⋆ Msp-I cuts all such DNA fragments.
  - ⋆ Hpa-II cuts only unmethylated DNA fragments.
  - ⋆ Cut fragments not amplified dropping DNA quantity.

Ref: Yuan et al., *Cancer Res.*, 2006; Kerkel et al., *Nature Genet.*, in press

SNP/CNV probes with Hpa-II/Msp-I sites

within Nsp-I/Sty-I DNA fragments:

- **Nsp-I/Sty-I treated sample gives total DNA $S_d + S_{ns}$.**
- **+Hpa-II treated sample gives methylated DNA $S_m + S_{ns}$.**
- **+Msp-I treated sample gives non-specific DNA S_{ns}.**

Other SNP/CNV probes all three have the "same" intensity

$$\textbf{Methylated fraction} = S_m/S_d$$

Caveat: method profiles methylation of CpG in CCGG.

SNP/CNV probes affected by Hpa-II/Msp-I treatment

- Test whether DNA fragment contains Hpa-II site.
- SNP/CNV genomic positions in Affymetrix annotation.
- Flanking Nsp-I/Sty-I sites at SNP/CNV position.
- Match recognition sequence against human genome (3Gb).

fasta file / R code using scan & grep.

Class I = no Hpa-II site	Class II = has Hpa-II site
Class III = Hpa-II site altered by SNP	Class IV = Hpa-II site depends on Nsp-I/Sty-I

Steps using `aroma.affymetrix`

- **Allelic crosstalk calibration.**
- **Quantile normalization.**
- **Copy number probe level modeling**
  - **allele specific data by not merging them.**
- **Fragment length normalization.**
- **Extract chip effect.**

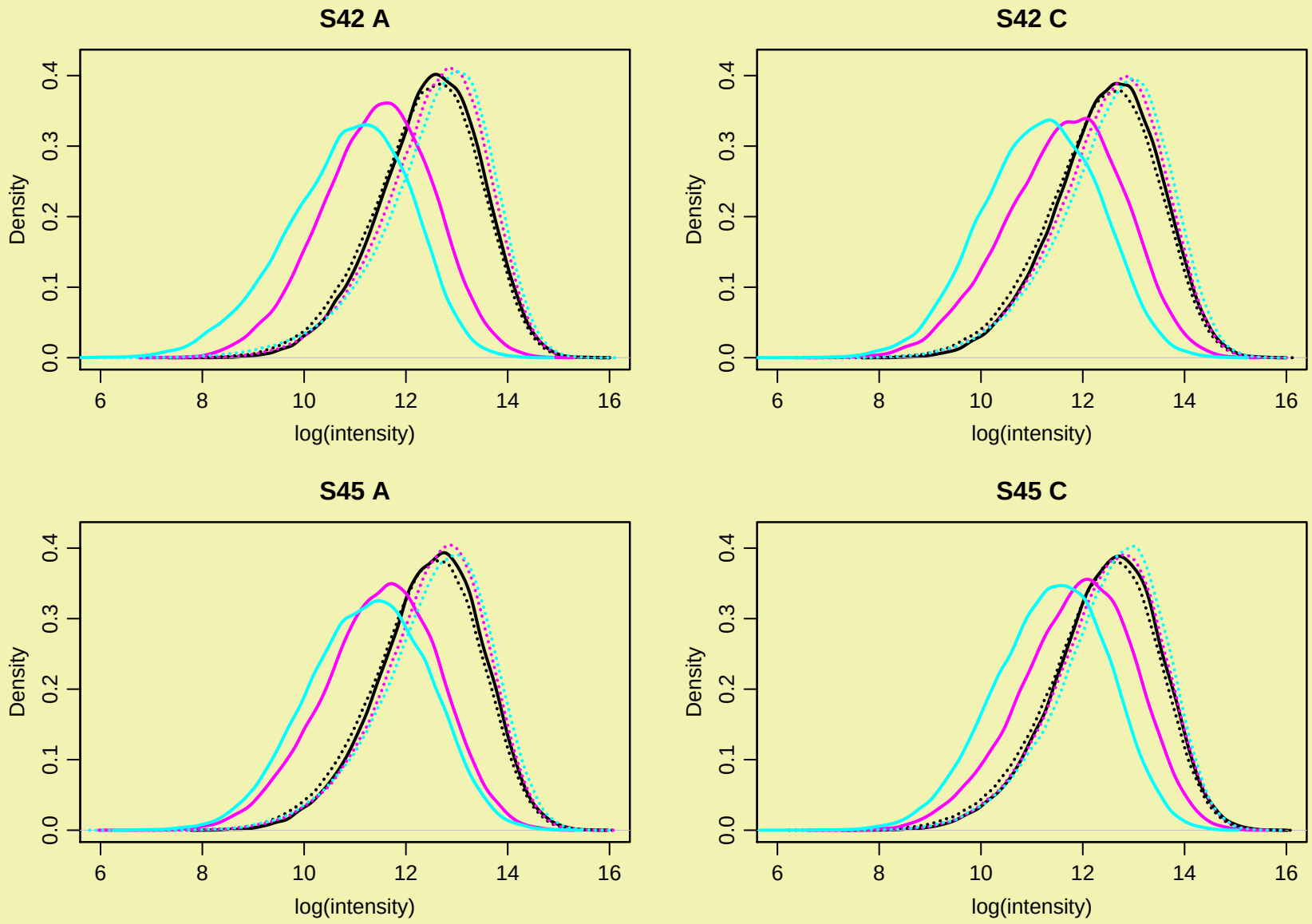
Ref: H. Bengtsson et al, *Tech Report 745, Bioinformatics 2008*

MSNP example

- **Profile global loss of methylation in 2 cases.**
- **Myofibroblasts from polypoid intestinal-type carcinoma and normal gastric tissue 10cm away.**
- **Affy 250k chips using Sty-I .**
- **Each tissue – 5 arrays: 2 Sty-I, 2 Hpa-II and 1 Msp-I.**
- **All 20 chips analyzed together**

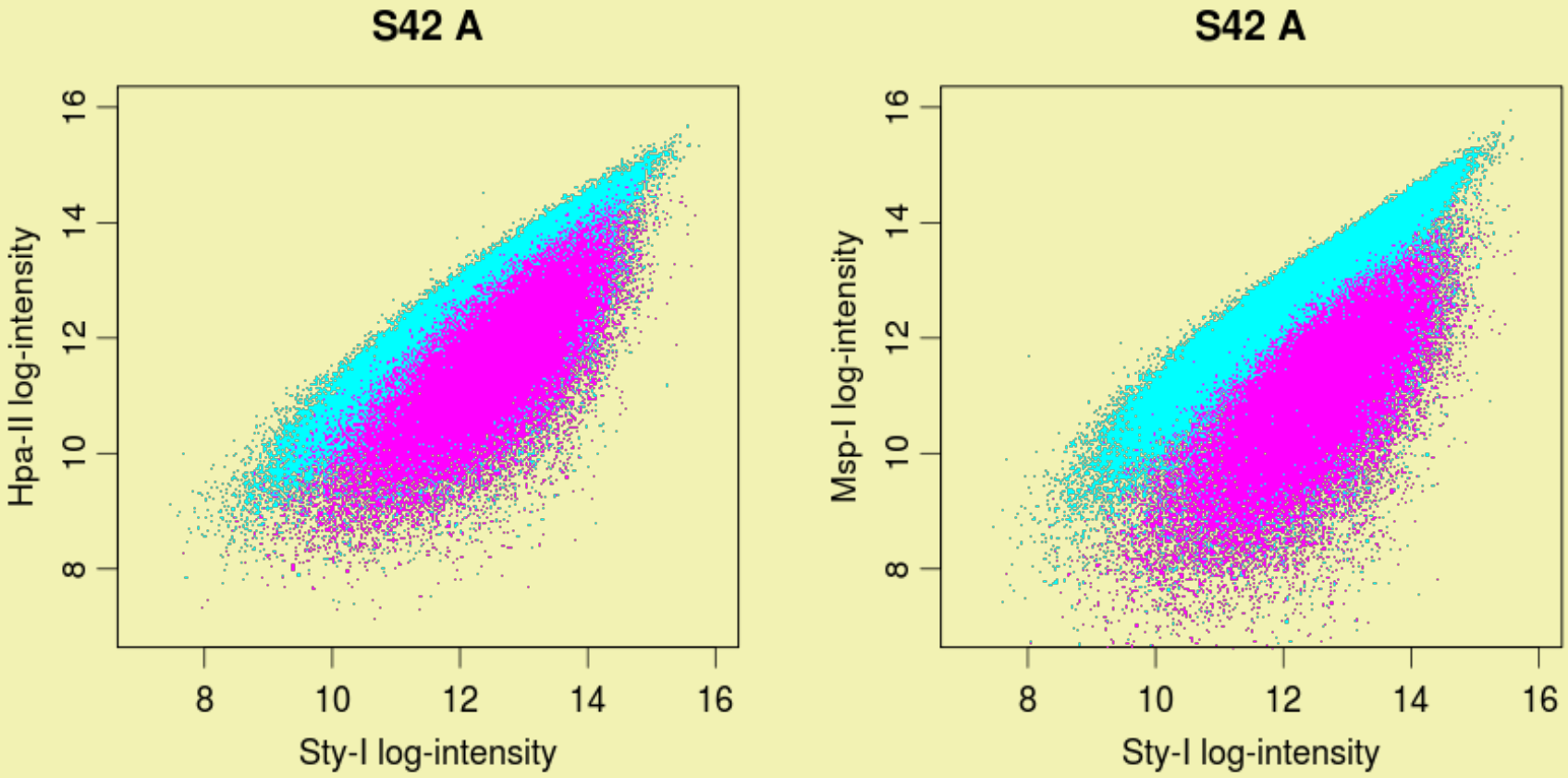
Ref: Jiang et al, *Cancer Research*, **in press**

MSNP example



Black = Sty−I, Magenta = Hpa−II, Cyan = Msp−I

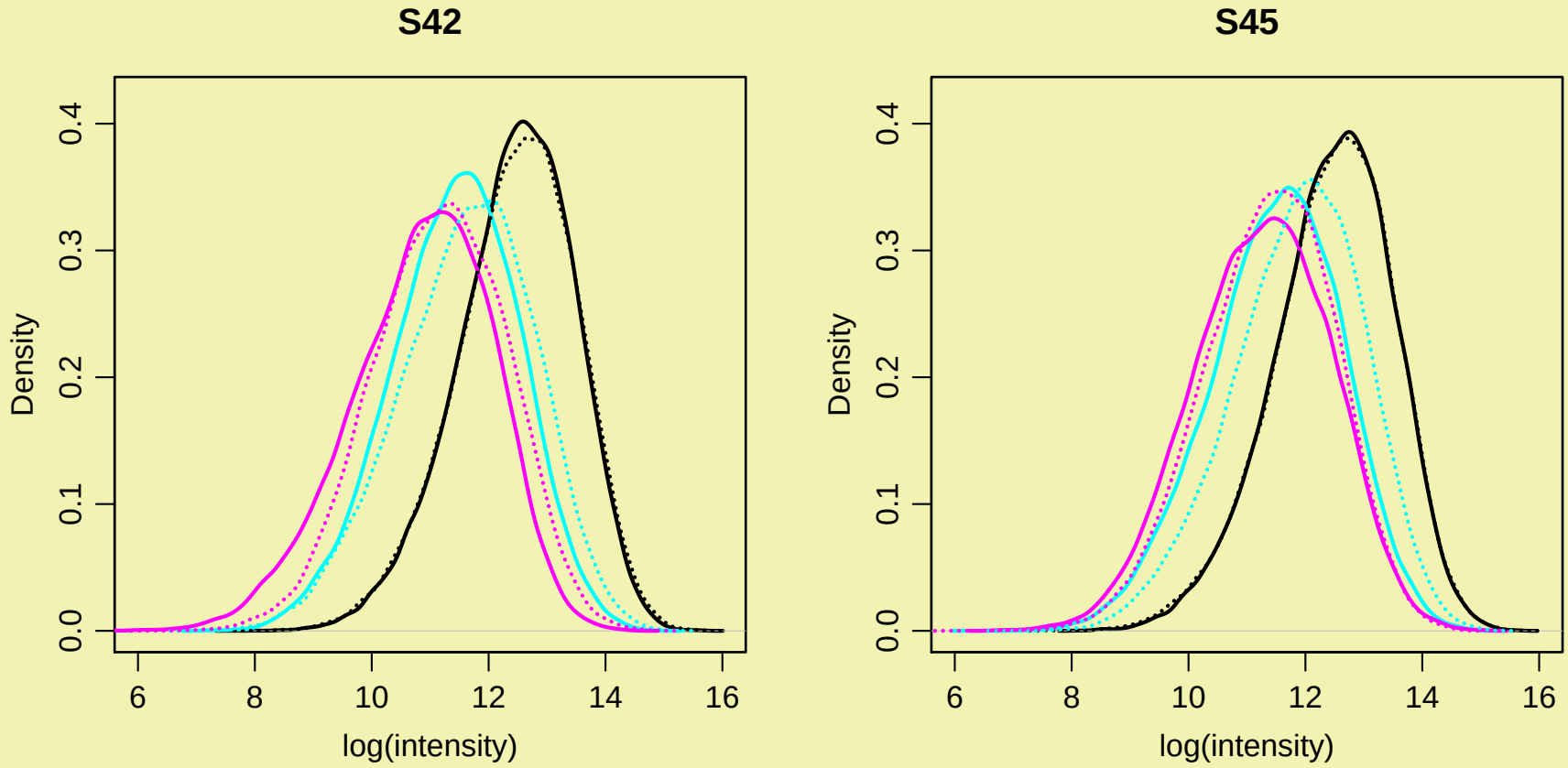
MSNP example



Cyan = Class I, Magenta = Class II

Note the shift in the density of magenta blob.

MSNP example



Black = Sty-I, Magenta = Hpa-II, Cyan = Msp-I; Solid = A, dashed = C

Note the shift in the density of Hpa-II curves.

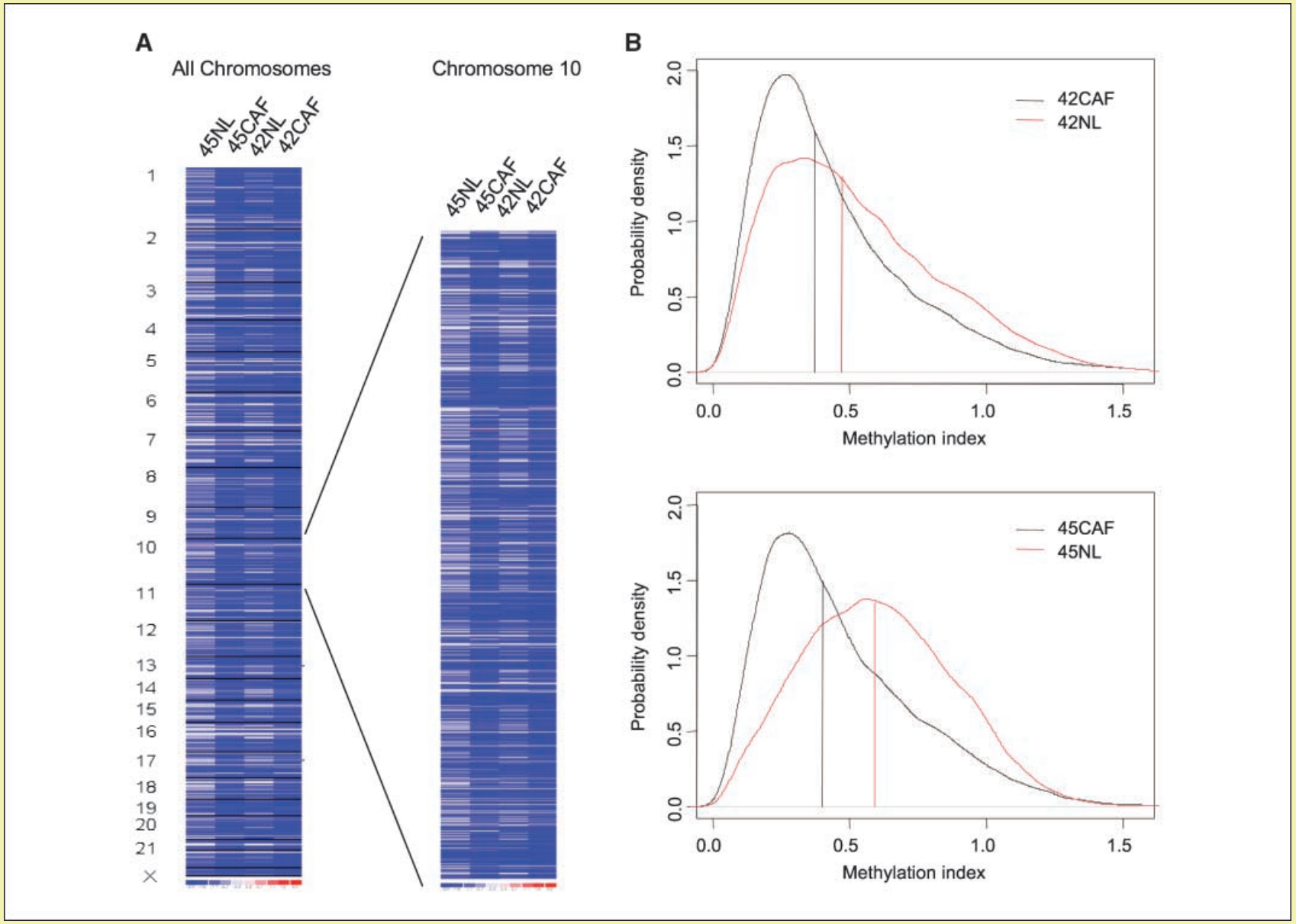
- **Intensity of Sty-I chips give $S_d + S_{ns}$**
- **Intensity of Sty-I + Hpa-II chips give $S_m + S_{ns}$**
- **Intensity of Sty-I + Msp-I chips give S_{ns}**
- **Methylation index =**

  [Sty-I - (Sty-I + Hpa-II)]/[Sty-I - (Sty-I + Msp-II)]

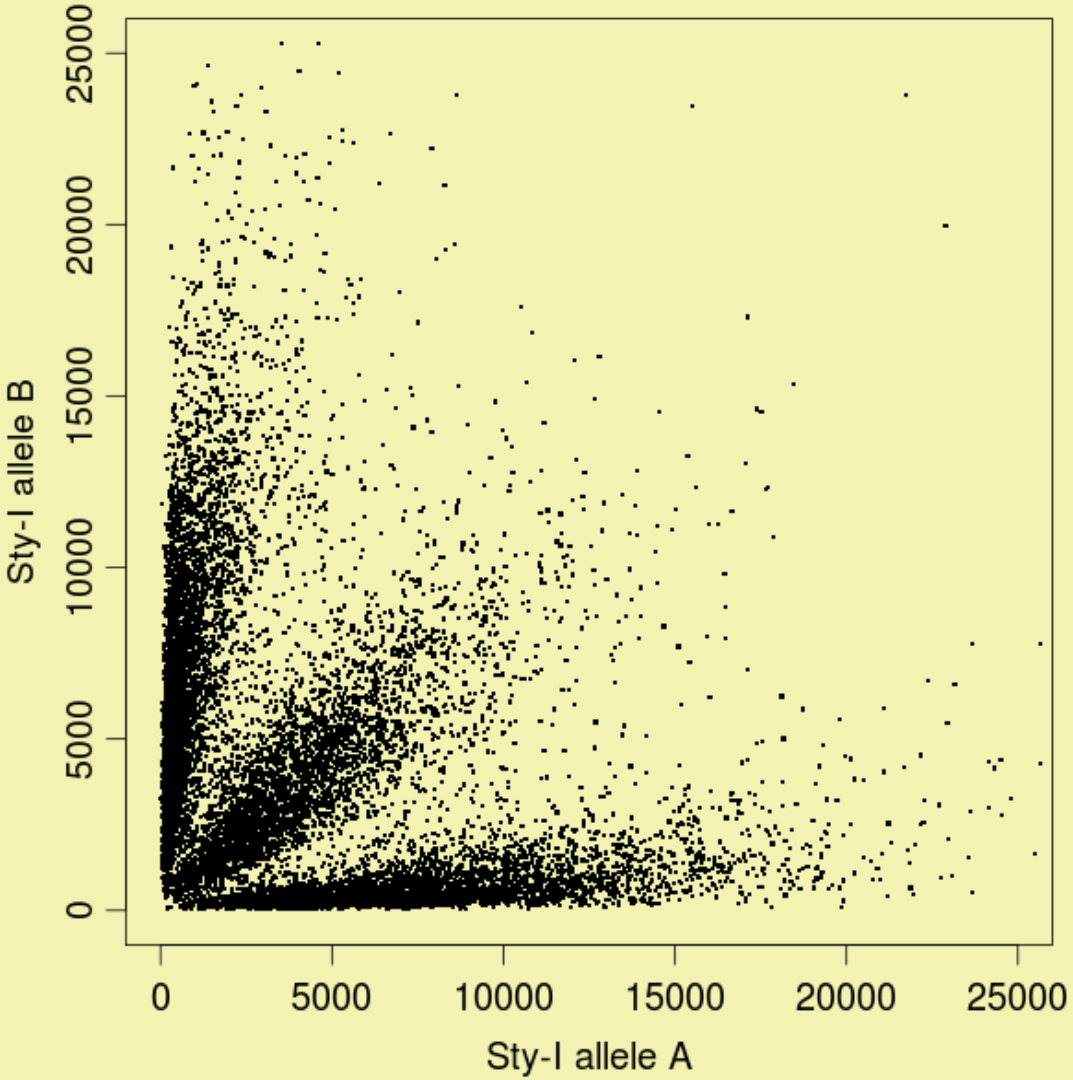
Issue: often Sty-I + Msp-I signal is very close to Sty-I alone

Empirical Class 2 Msp-I should cut the signal by at least half.

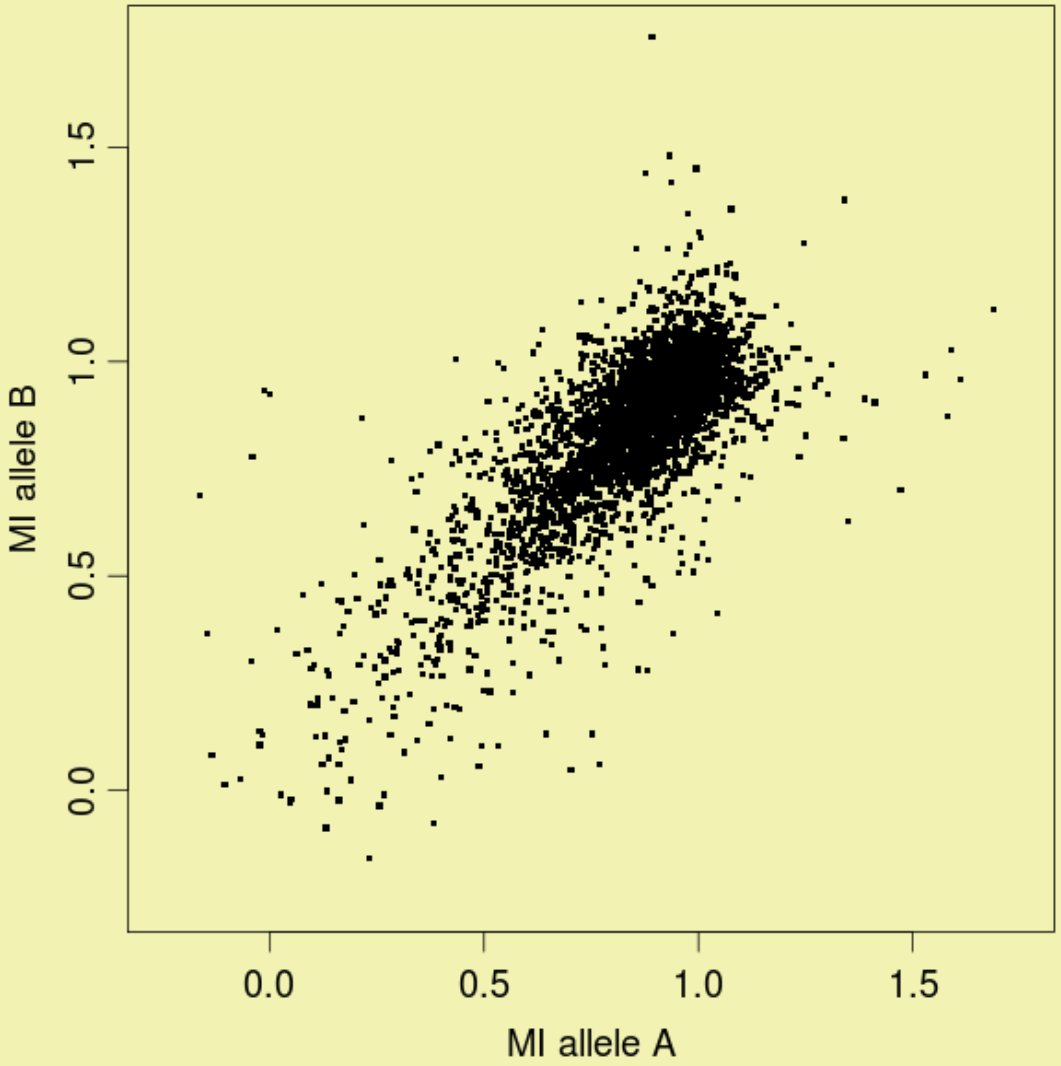
Methylation Index



What about ASM? - Pick out the heterozygous SNPs



Calculate allele specific methylation indices and plot



- Everything we have done is a per sample analysis
- We can treat MI and ASM as raw data and analyze

  a la gene expression
- Identify loci that are differentially methylated or prognostic

  show allelic imbalance in methylation

MSNP a promising method for methylation profiling.

Fine tune the analysis flow.