HM450 arrays

Kim Siegmund
kims@usc.edu

Illumina HumanMethylation450

Figure from http://www.illumina.com

485,577 Sample Probes

Target Site

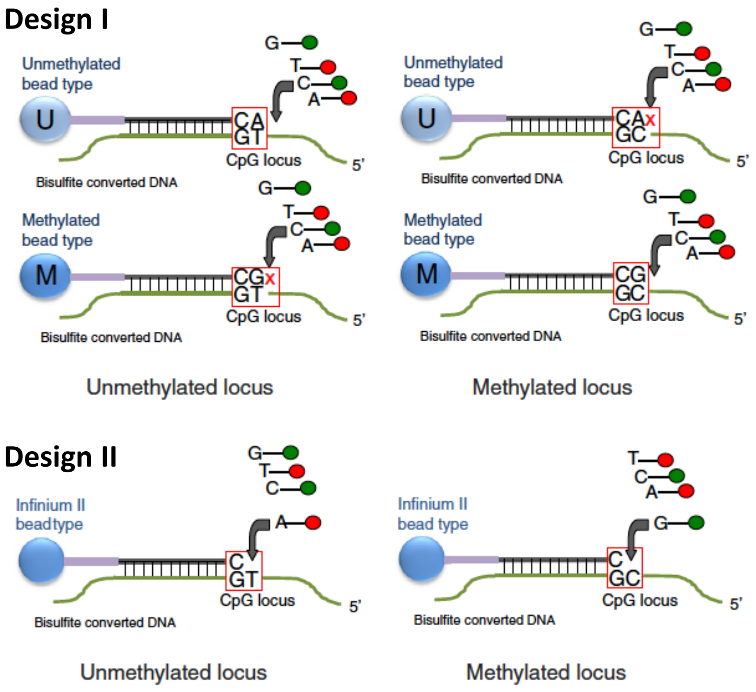
- 482,421 CpG
- 3,091 CpH
- 65 rs

Color of bead types:

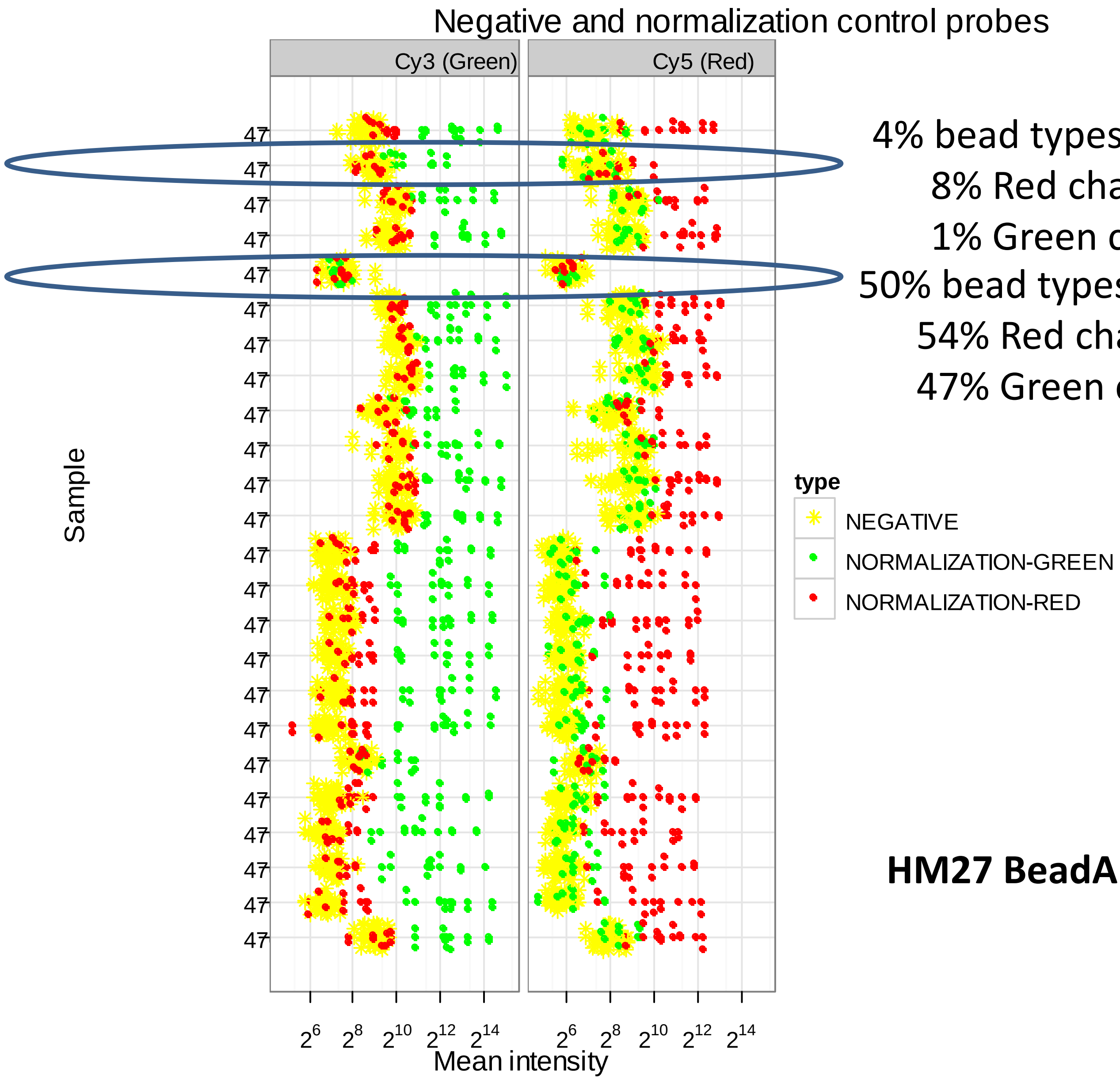
350,076 (70%) Both (**M**,**U**) Design II

46,298 (10%) Green (**M**,**U**) Design I

89,203 (20%) Red (**M**,**U**) Design I



QC: Plot Control Probs



4% bead types failed

8% Red channel

1% Green channel

50% bead types failed

54% Red channel

47% Green channel

HM27 BeadArrays

Preprocessing HM Data

0. Image analysis Summarize Avg M, Avg U, nbeadsM,…
1. Data import BioC pkgs: methylumi, lumi, minfi
2. Quality assessment
3. Background correction
4. Dye-bias equalization - new for HM450
5. Summarization Beta-value or M-value ?
6. Normalization Quantile Normalization (Sun et al. 2011)
   adjust batch, DNA input, BSC efficiency (Teschendorff et al. 2009)
   Subset-Within Array Normalization (Makismovic et al., 2012)
7. Batch-effect correction ComBat (Sun et al. 2011)

DNA methylation measures

$$\text{DNA methylation frequency} = \frac{M}{M + U + c}$$

where M = avg fluorescent intensity from methylated allele
U = avg fluorescent intensity from unmethylated allele
c is a constant used by Illumina (c=100)

M-value = $\log_2(M/U)$

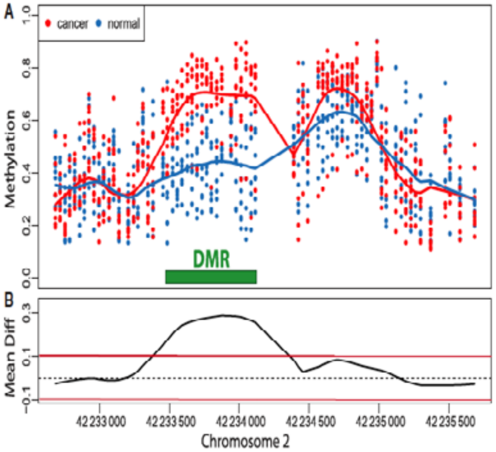
Question: What is the best scale for analysis?

Differential DNAm Analysis
Spatially Correlated Features

- DNA methylation shows correlation at < 1000 bps

  Eckhardt et al. (2006) Nature Genetics

- CHARM:



Jaffe et al. (2012)
Int J Epidemiol

- HM450 Array: Median distance between adjacent probes ~300 bps

Differential DNAm Analysis
Spatially Correlated Features

- Marginal tests
  - Non-Homogeneous Markov Model

    Kuan & Chiang, Biometrics 2012
  - Smooth coefficients to identify regions (Bump hunting)

    Jaffe et al. Int J Epidemiol 2012

- High-dimensional Regression Analysis
  - Penalized Regression

    Sun & Wang, Bioinf 2012

    Teschendorff & Widschwendter 2012

Enrichment Testing

DNA methylation motivates new contexts for enrichment testing

Tests for enrichment by:

- GO terms, pathways ('traditional')
- genomic context/coordinates
  - CGI TSS
  - Polycomb-group protein (Widschwendter et al. 2007)
  - Nuclear lamina-associated domains (Berman et al. 2011)