

# Epigenomic Data Lab – Hands on Session for Epigenome-Wide DNA Methylation Data Analysis

Andres Cardenas, Harvard Medical School

Elena Colicino, Columbia University, Mailman School of Public Health

Allan Just, Icahn School of Medicine at Mount Sinai

# What to expect:

- We assume some basic working knowledge of R
  - We will help you individually as much as possible
  - Let us know if you are having issues so you don't fall behind
  - We also assume some basic biology knowledge
  - Don't be afraid to ask questions!
  
- Issues you might encounter
  - You have the wrong version of R and required packages
  - You only have 4 GB of RAM and your computer is getting stuck
  - The graphing is being slow or giving you errors (small plotting window)

# **Parts to the workshop**

- Part 1- Short intro to technical aspects of the 450k
- Part 2- Preprocessing of the data
- Part 3- Analysis of the data (methodology)
- Part 4- Interpretation of the results

# Infinium HumanMethylation450 BeadChip

Microarray includes  
485,512 methylation sites  
across the human genome



Covers 99% of RefSeq genes  
but still <2% of all CpGs

Bisulfite treated DNA  
hybridizes to targeted  
probes → fluorescence

$$\beta\text{-value} = \frac{M}{M+U+\varepsilon}$$

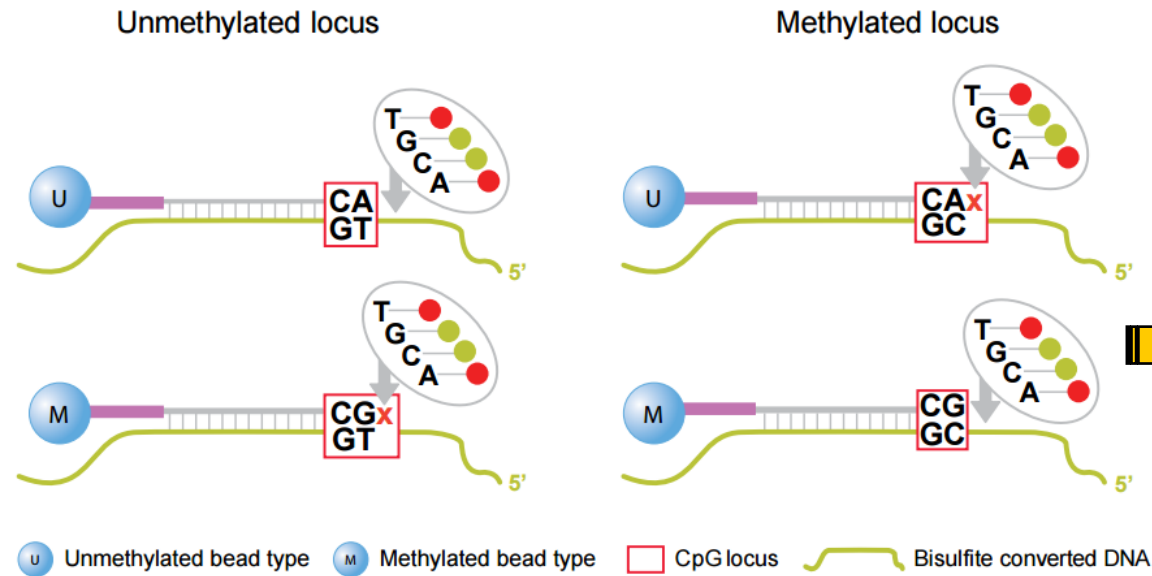
$M$ : signal from methylated CpG target

$U$ : signal from unmethylated CpG target

Average  $\beta$  is an interval-scaled quantity between 0-1 that is interpreted as the fraction of DNA molecules whose target CpG is methylated

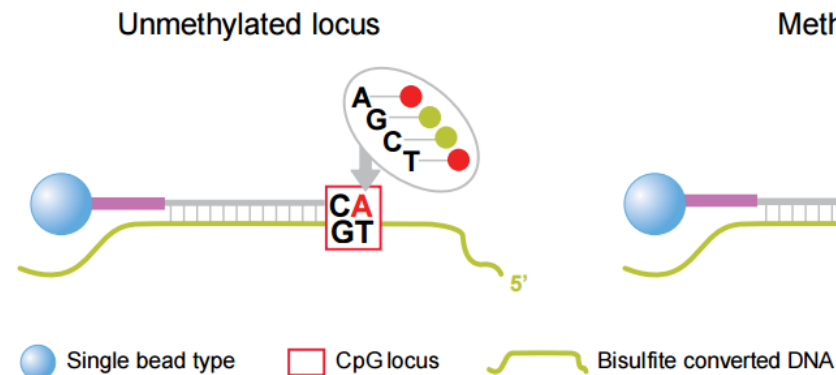
# Infinium assay: two types of chemistry

## Infinium I



- 2 beads
  - Same color channel
  - **Red vs Red**
  - **Green vs Green**
- Makes up ~1/4 of 450k**

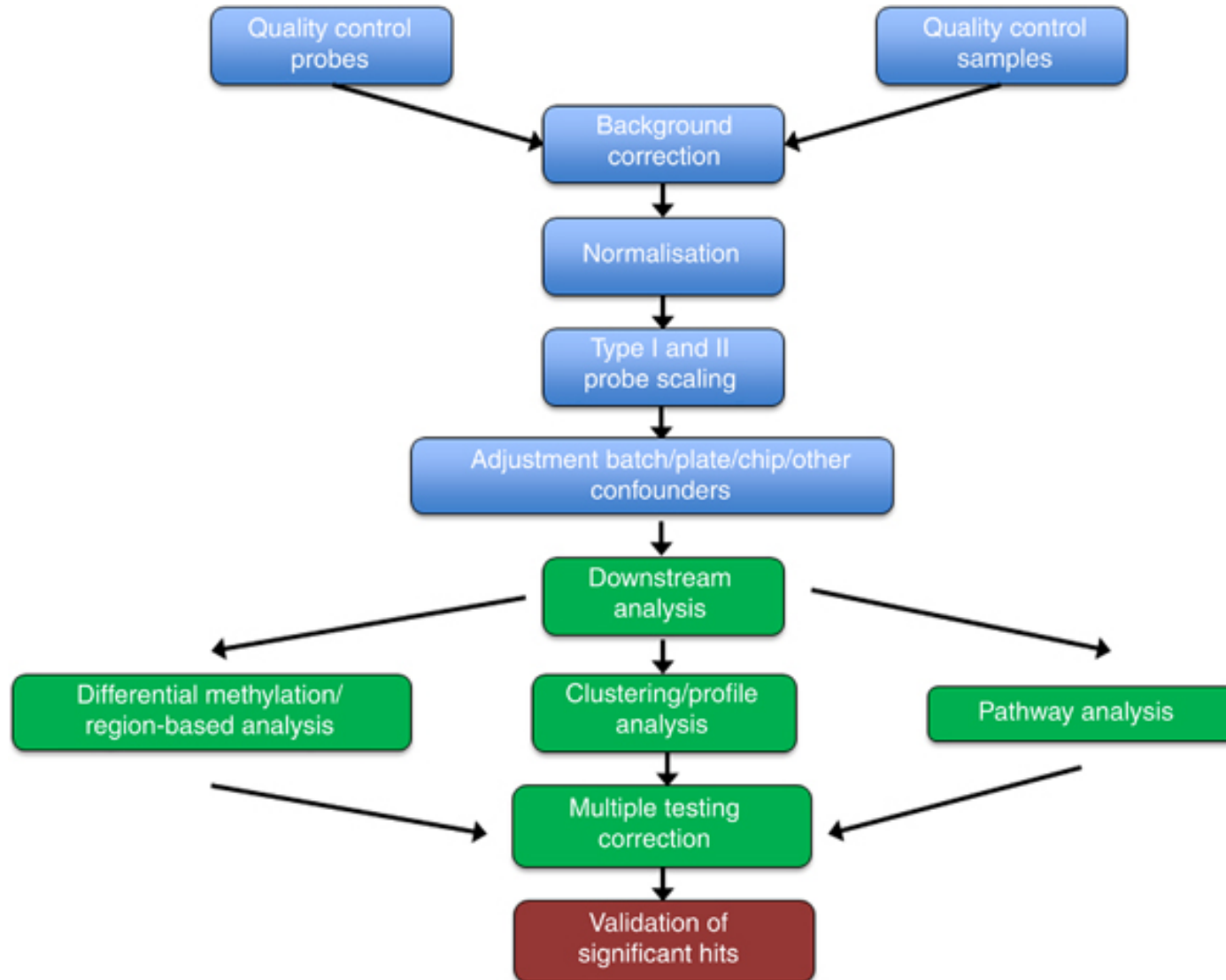
## Infinium II



- 1 bead type
  - Two different colors
  - **Red** and **Green**
- Makes up ~3/4 of 450k**

Used in slightly different contexts  
Lesser dynamic range

# 450K Processing & Analysis





## **Steps in the workflow today:**

- Estimating blood cell proportions
- Background correction
- Probe type adjustment
- Detection p-value filtering
- Batch effect adjustment
- Checking sex
- Filtering polymorphic, cross-reactive, and allosomal probes
- Fitting an EWAS
- Genomic inflation and interpretation of EWAS results
- Differentially methylated region analysis

## **Steps we aren't showing today:**

- Plating & study design
- Import of idat files
- Identity checks / sample swaps
- QC visualization – see shinyMethyl
- Enrichment analysis or Pathway analysis



## Methylation microarray analyses in the future: 450k replaced with EPIC array (850k) in late 2015

- Same chemistries (mostly added Infinium II probes)
- Chip design has 8 rows & 1 column instead of 6R x 2C
- Adds content in enhancer regions, outside CpG islands, and more...
- Most of 450k included on EPIC – but drops 31,389 cg probes
- Minfi updating to address normalization and handling
- More work needed on integrating 450k and 850k (e.g. can't impute methylation as is done with GWAS chips of deferring density)