Methylation Analysis Tools

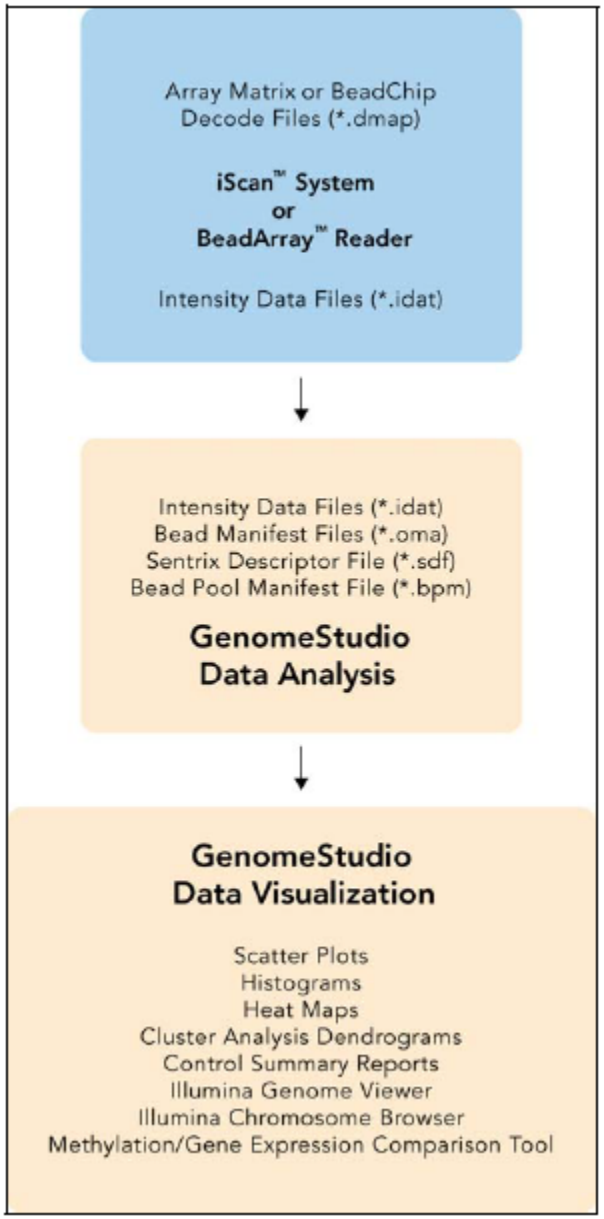
David Crossman, Ph.D.
UAB Heflin Center for Genomic Science

Immersion Course
Tuesday, April 9, 2013

Methylation Analysis Tools

- Command line
  - R (Bioconductor)
    - Lumi
    - SWAN
    - IMA
    - SQN
    - BMIQ
    - And many more…
- Graphical User Interface (GUI)
  - GenomeStudio

GenomeStudio pipeline



RESEARCH PAPER

Epigenetics 8:3, 333–346; March 2013; © 2013 Landes Bioscience

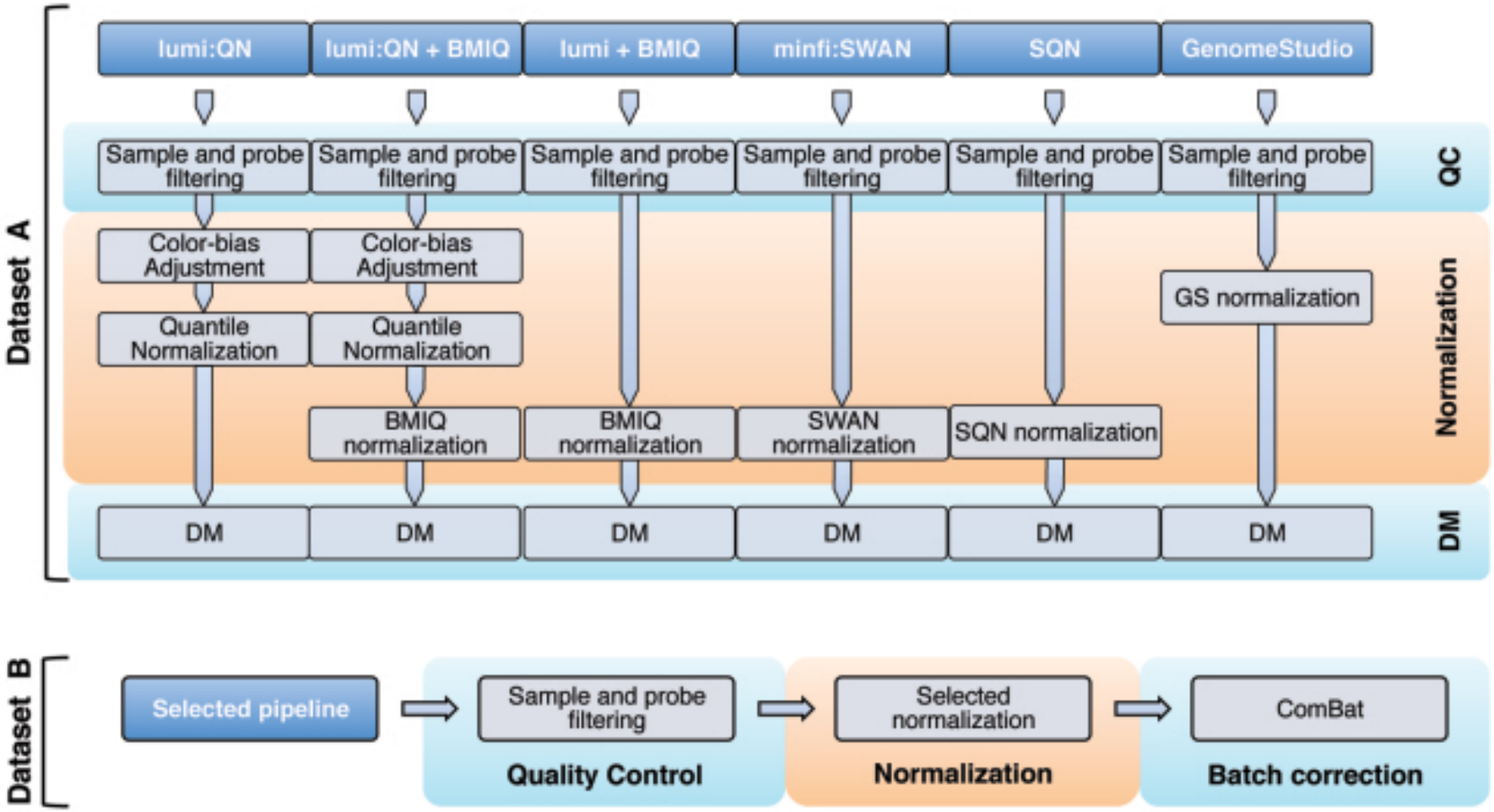
An evaluation of analysis pipelines for DNA methylation profiling using the Illumina HumanMethylation450 BeadChip platform

Francesco Marabita,[1] Malin Almgren,[2] Maléne E. Lindholm,[3] Sabrina Ruhrmann,[2] Fredrik Fagerström-Billai,[4] Maja Jagodic,[2] Carl J. Sundberg,[3] Tomas J. Ekström,[2] Andrew E. Teschendorff,[5] Jesper Tegnér[1] and David Gomez-Cabrero[1,6,*]

[1]Unit of Computational Medicine; Center for Molecular Medicine; Department of Medicine; Karolinska Institutet; Stockholm, Sweden; [2]Center for Molecular Medicine; Department of Clinical Neuroscience; Karolinska Institutet; Stockholm, Sweden; [3]Department of Physiology and Pharmacology; Karolinska Institutet; Stockholm, Sweden; [4]Department of Biosciences and Nutrition; Karolinska Institutet; Stockholm, Sweden; [5]Statistical Genomics Group; Paul O'Gorman Building; UCL Cancer Institute; University College London; London, UK; [6]Bioinformatics Infrastructure for Life Sciences; Stockholm, Sweden

Keywords: technical variability, DNA methylation, microarray, Illumina 450K, normalization

The proper identification of differentially methylated CpGs is central in most epigenetic studies. The Illumina HumanMethylation450 BeadChip is widely used to quantify DNA methylation; nevertheless, the design of an appropriate analysis pipeline faces severe challenges due to the convolution of biological and technical variability and the presence of a signal bias between Infinium I and II probe design types. Despite recent attempts to investigate how to analyze DNA methylation data with such an array design, it has not been possible to perform a comprehensive comparison between different bioinformatics pipelines due to the lack of appropriate data sets having both large sample size and sufficient number of technical replicates. Here we perform such a comparative analysis, targeting the problems of reducing the technical variability, eliminating the probe design bias and reducing the batch effect by exploiting two unpublished data sets, which included technical replicates and were profiled for DNA methylation either on peripheral blood, monocytes or muscle biopsies. We evaluated the performance of different analysis pipelines and demonstrated that: (1) it is critical to correct for the probe design type, since the amplitude of the measured methylation change depends on the underlying chemistry, (2) the effect of different normalization schemes is mixed, and the most effective method in our hands were quantile normalization and Beta Mixture Quantile dilation (BMIQ) and (3) it is beneficial to correct for batch effects. In conclusion, our comparative analysis using a comprehensive data set suggests an efficient pipeline for proper identification of differentially methylated CpGs using the Illumina 450K arrays.



Epigenetics 8:3, 333-346; March 2013