Epigenetics—where the environment meets the genetics of disease: Interpreting high-throughput DNA Methylation data

Miles Benton

Institute of Environmental Science and Research, Wellington, New Zealand.
Genomics Research Centre, Griffith Health Institute, Griffith University, Australia.

WHBRS Young Investigators Meeting
Victoria University, Wellington

19th November 2012

Outline

1. Epigenetics Overview
2. Illumina 450K
3. Some Biology
4. Bringing it all together
5. Acknowledgements

Epigenetics: An Overview

- Epi-(GREEK:over,above) genetics
  - heritable changes in gene expression/phenotype
  - functionally relevant changes, doesn't modfiy underlying genomic sequence

Epigenetics: An Overview

- Epi-(GREEK:over,above) genetics
  - heritable changes in gene expression/phenotype
  - functionally relevant changes, doesn't modfiy underlying genomic sequence

- Several mechanisms:

Epigenetics: An Overview

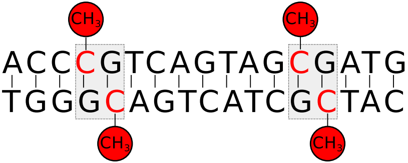
- Epi-(GREEK:over,above) genetics
  - heritable changes in gene expression/phenotype
  - functionally relevant changes, doesn't modfiy underlying genomic sequence
- Several mechanisms:
  1. Histone modifications (Acetylation and Methylation)

Epigenetics: An Overview

- Epi-(GREEK:over,above) genetics
  - heritable changes in gene expression/phenotype
  - functionally relevant changes, doesn't modfiy underlying genomic sequence
- Several mechanisms:
  1. Histone modifications (Acetylation and Methylation)
  2. **DNA Methylation**

A brief overview of DNA Methylation

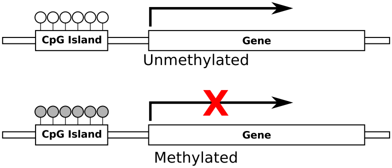
- 5'methylation of cytosine
- typically at CpG dinucleotide



A brief overview of DNA Methylation

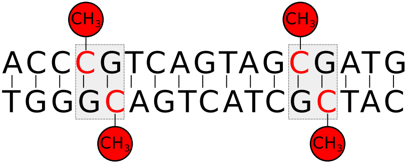
- 5’methylation of cytosine
- typically at CpG dinucleotide

Result is a reduction of gene expression (though there are exceptions)

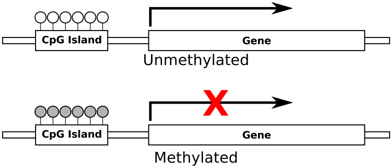


A brief overview of DNA Methylation

- 5'methylation of cytosine
- typically at CpG dinucleotide



Result is a reduction of gene expression (though there are exceptions)



- Methylation is heritable - imprinting
- **BUT** is also dynamic: response to environment

The Illumina 450K Human Methylation Chip

- Illumina 450K Human Methylation Chip
  - Genome Wide: has >450,000 probes (individual CpG sites)
  - only recently publications starting to emerge
  - analysis needs to catch up

The Illumina 450K Human Methylation Chip

- Illumina 450K Human Methylation Chip
  - Genome Wide: has >450,000 probes (individual CpG sites)
  - only recently publications starting to emerge
  - analysis needs to catch up

- DNA is Bisulfite converted, fragmented and hybridised.

The Illumina 450K Human Methylation Chip

- Illumina 450K Human Methylation Chip
  - Genome Wide: has >450,000 probes (individual CpG sites)
  - only recently publications starting to emerge
  - analysis needs to catch up

- DNA is Bisulfite converted, fragmented and hybridised.
- Difference in intensity between meth/unmeth probes gives Beta.

The Illumina 450K Human Methylation Chip

- Illumina 450K Human Methylation Chip
  - Genome Wide: has >450,000 probes (individual CpG sites)
  - only recently publications starting to emerge
  - analysis needs to catch up

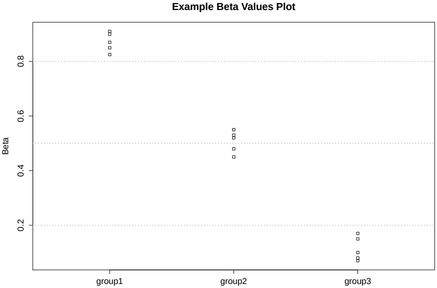
- DNA is Bisulfite converted, fragmented and hybridised.
- Difference in intensity between meth/unmeth probes gives Beta.

Beta Values

$$\beta = \frac{meth}{meth + unmeth + 100}$$

Values range between 0-1

- think of as percent methylation



The Illumina 450K Human Methylation Chip

- Illumina 450K Human Methylation Chip
  - Genome Wide: has >450,000 probes (individual CpG sites)
  - only recently publications starting to emerge
  - analysis needs to catch up

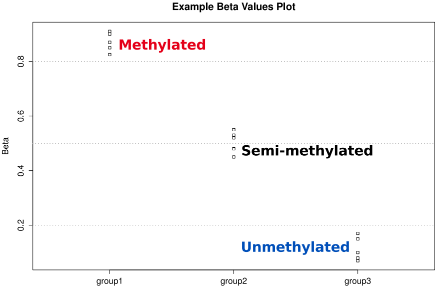
- DNA is Bisulfite converted, fragmented and hybridised.
- Difference in intensity between meth/unmeth probes gives Beta.

Beta Values

$$\beta = \frac{meth}{meth + unmeth + 100}$$

Values range between 0-1

- think of as percent methylation



Study Design is Crucial

Need to really think about study design:

- Same old story: Quality in equals Quality out
  - good bioinformatics will never make up for poor lab technique and/or poor study design

Study Design is Crucial

Need to really think about study design:

- Same old story: Quality in equals Quality out
  - good bioinformatics will never make up for poor lab technique and/or poor study design
- Gender - sex specific methylation patterns

Study Design is Crucial

Need to really think about study design:

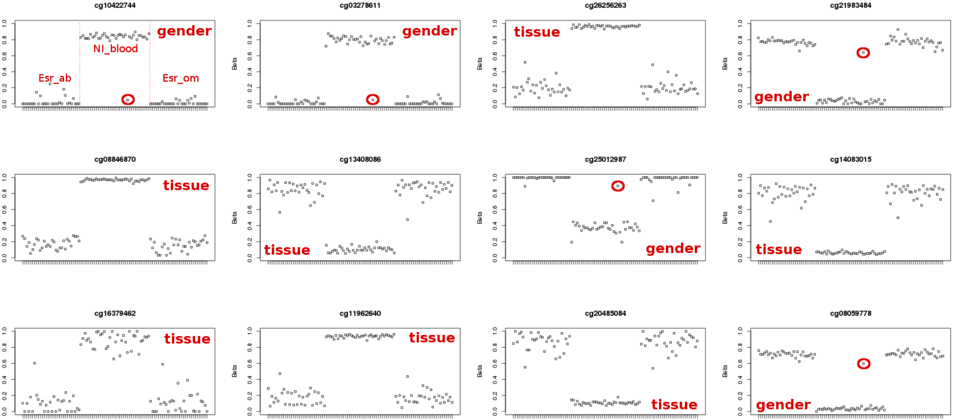
- Same old story: Quality in equals Quality out
  - good bioinformatics will never make up for poor lab technique and/or poor study design
- Gender - sex specific methylation patterns
- Origin of sample (i.e. blood, tissue)
  - mixed cell populations

Study Design is Crucial

Need to really think about study design:

- Same old story: Quality in equals Quality out
  - good bioinformatics will never make up for poor lab technique and/or poor study design
- Gender - sex specific methylation patterns
- Origin of sample (i.e. blood, tissue)
  - mixed cell populations
- Case/Control? Paired-samples? Numbers?

Example: Gender and Tissue specific markers



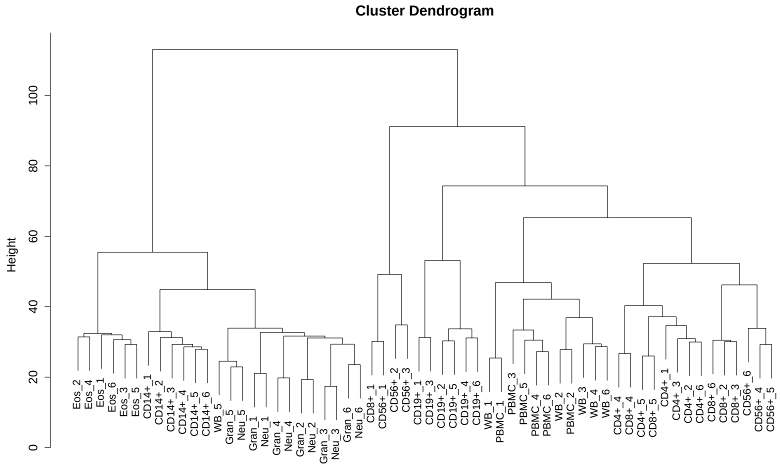
The Biology

Time for the interesting part...

The technology in action

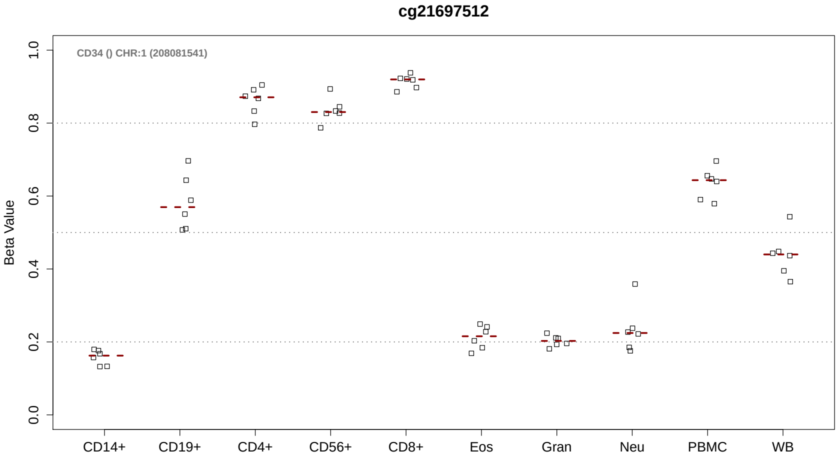
Investigating Methyation Profiles in Blood Cells

Publicly available data set (6 'healthy' males) - cell sorted



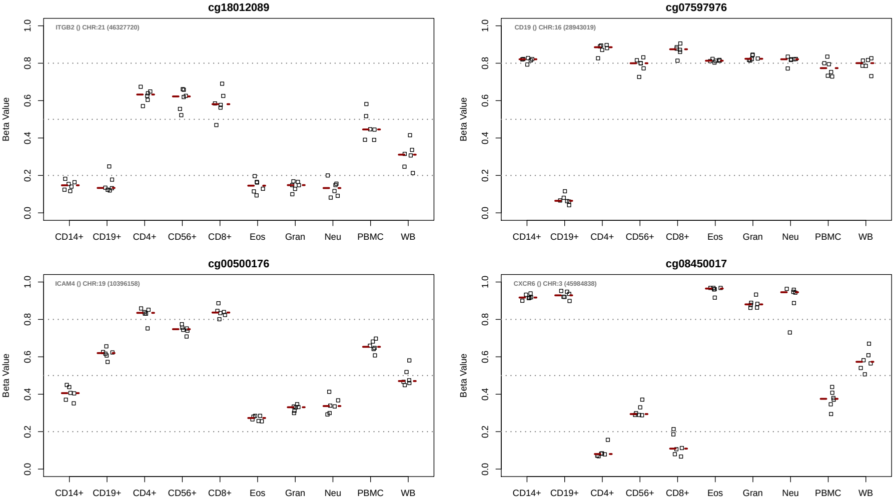
Investigating Methylation Profiles in Blood Cells

Publicly available data set (6 'healthy' males) - cell sorted



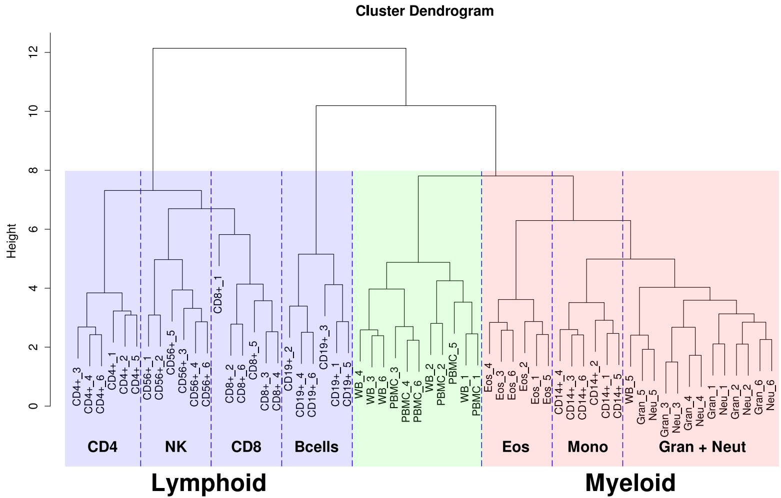
Investigating Methyation Profiles in Blood Cells

Publicly available data set (6 'healthy' males) - cell sorted

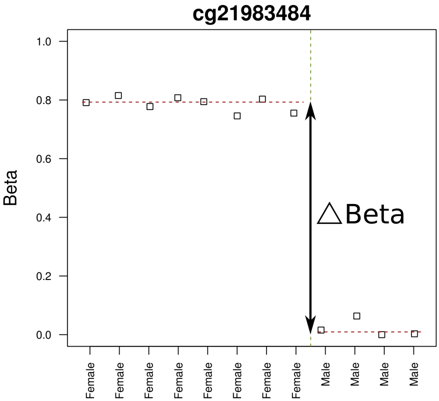


Investigating Methyation Profiles in Blood Cells

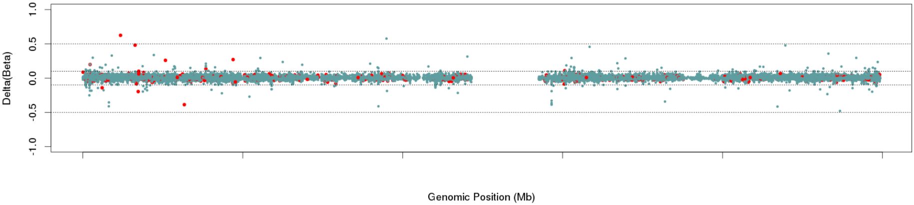
Publicly available data set (6 'healthy' males) - cell sorted



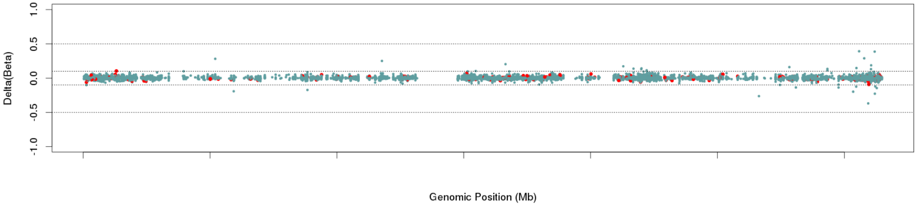
Genome Wide Methylation and Disease Grouping



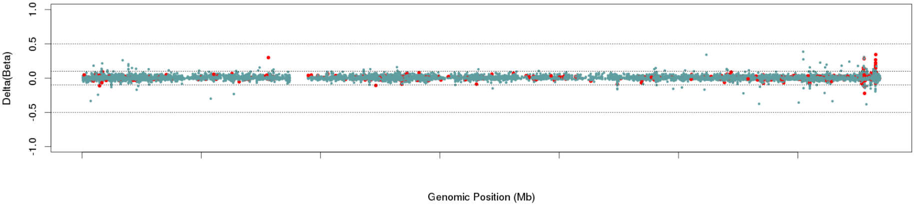
Genome Wide Methylation and Disease Grouping



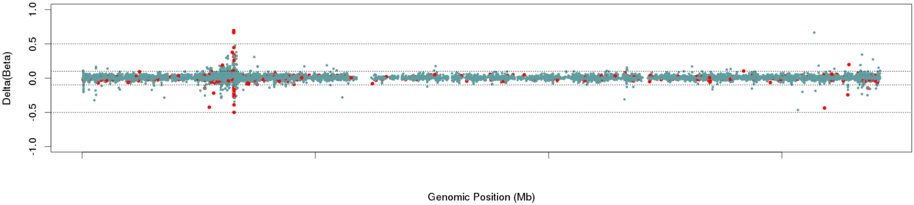
Genome Wide Methylation and Disease Grouping



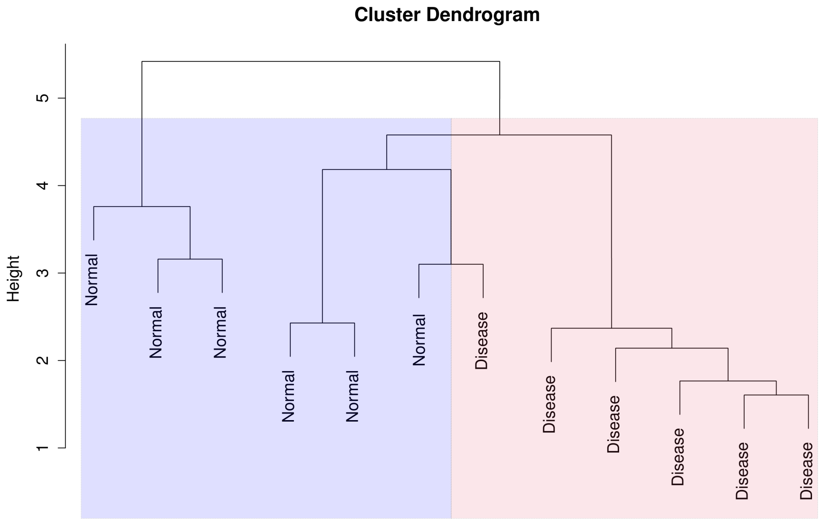
Genome Wide Methylation and Disease Grouping



Genome Wide Methylation and Disease Grouping



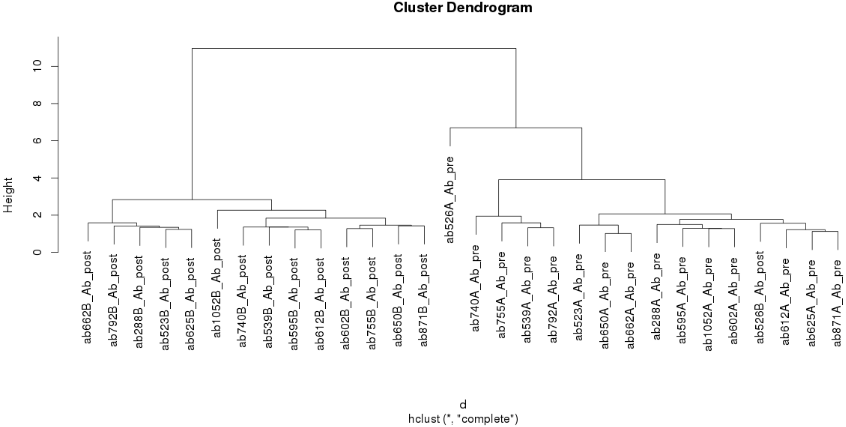
Genome Wide Methylation and Disease Grouping



ESR/Wakefield Obesity Project

ESR/Wakefield Obesity Project:

-15 patients pre/post-op gastric bypass (female; abdominal fat tissue)



Overlaying Genomic Data Sets

ESR/Wakefield Obesity Project:

Overlaying Genomic Data Sets

ESR/Wakefield Obesity Project:

- multiple layers of genomic data
  - mRNA data
  - miRNA data
  - methylation data

Overlaying Genomic Data Sets

ESR/Wakefield Obesity Project:

- multiple layers of genomic data
  - mRNA data
  - miRNA data
  - methylation data
- this allows multidimensional analysis

Overlaying Genomic Data Sets

ESR/Wakefield Obesity Project:

- multiple layers of genomic data
  - mRNA data
  - miRNA data
  - methylation data
- this allows multidimensional analysis

- Identified a specific miRNA that is significantly differentially methylated in abdominal pre vs post

Overlaying Genomic Data Sets

ESR/Wakefield Obesity Project:

- multiple layers of genomic data
  - mRNA data
  - miRNA data
  - methylation data
- this allows multidimensional analysis

- Identified a specific miRNA that is significantly differentially methylated in abdominal pre vs post

- the same miRNA shows significant differential expression

Overlaying Genomic Data Sets

ESR/Wakefield Obesity Project:

- multiple layers of genomic data
  - mRNA data
  - miRNA data
  - methylation data
- this allows multidimensional analysis

- Identified a specific miRNA that is significantly differentially methylated in abdominal pre vs post

- the same miRNA shows significant differential expression

- Pathways analysis indicates that 13 of the top differentially expressed mRNA are potentially regulated by the above miRNA

Overlaying Genomic Data Sets

ESR/Wakefield Obesity Project:

- multiple layers of genomic data
  - mRNA data
  - miRNA data
  - methylation data
- this allows multidimensional analysis
- Identified a specific miRNA that is significantly differentially methylated in abdominal pre vs post
- the same miRNA shows significant differential expression
- Pathways analysis indicates that 13 of the top differentially expressed mRNA are potentially regulated by the above miRNA
- We have identified significant correlations between all layers of data
  - suggests strong biological relevance

Conclusions

- We have designed and validated a novel QC and analysis pipeline

Conclusions

- We have designed and validated a novel QC and analysis pipeline
- This allows us to demonstrated the utility of high-throughput methylation technology in terms of disease association
  - able to group based on differential expressed methylation profiles

Conclusions

- We have designed and validated a novel QC and analysis pipeline

- This allows us to demonstrated the utility of high-throughput methylation technology in terms of disease association
  - able to group based on differential expressed methylation profiles

- Attention to study design will allow increased quality and application
  - i.e. multi-cell populations in tissue/whole blood

Conclusions

- We have designed and validated a novel QC and analysis pipeline
- This allows us to demonstrated the utility of high-throughput methylation technology in terms of disease association
  - able to group based on differential expressed methylation profiles
- Attention to study design will allow increased quality and application
  - i.e. multi-cell populations in tissue/whole blood
- Applying multiple layers of genomic data shows potential to increase our understanding of complex biological pathways/networks

Acknowledgements

ESR/Wakefield Obesity Project: Dr Donia Macartney-Coxson, Prof Richard Stubbs, Angela Jones, Daniel Kay

Norfolk Island Health Study: Prof Lyn Griffiths

PHD Supervisors: Dr Rod Lea, Prof Lyn Griffiths

Co Supervisors: Dr Donia Macartney-Coxson, Dr Geoff Chambers

Obesity Project Sample Running: George Washington University, Washington DC

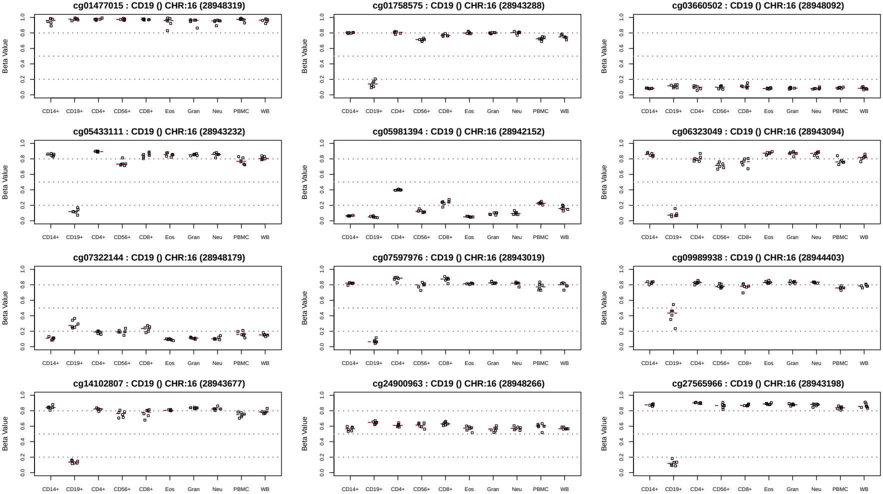
Genomics Research Centre: Dr Heidi Sutherland, Michelle Hanna, Dr Bridget Maher, Dr David Eccles

Institutes: Institute of Environmental and Scientific Research, Wakefield Hospital, Griffith University, Victoria University of Wellington

All the patients for their consent to being part of the respective studies

Investigating Methyation Profiles in Blood Cells

Publicly available data set (6 'healthy' males) - cell sorted


cg01477015 : CD19 () CHR:16 (28948319)
cg01758575 : CD19 () CHR:16 (28943288)
cg03660502 : CD19 () CHR:16 (28948092)
cg05433111 : CD19 () CHR:16 (28943232)
cg05981394 : CD19 () CHR:16 (28942152)
cg06323049 : CD19 () CHR:16 (28943094)
cg07322144 : CD19 () CHR:16 (28948179)
cg07597976 : CD19 () CHR:16 (28943019)
cg09989938 : CD19 () CHR:16 (28944403)
cg14102807 : CD19 () CHR:16 (28943677)
cg24900963 : CD19 () CHR:16 (28948266)
cg27565966 : CD19 () CHR:16 (28943198)