

From Understanding to Mining DNA Methylation Data: a TCGA resource

Bodour Salhia, PhD
Assistant Professor

Integrated Cancer Genomics Division
Translational Genomics Research Institute

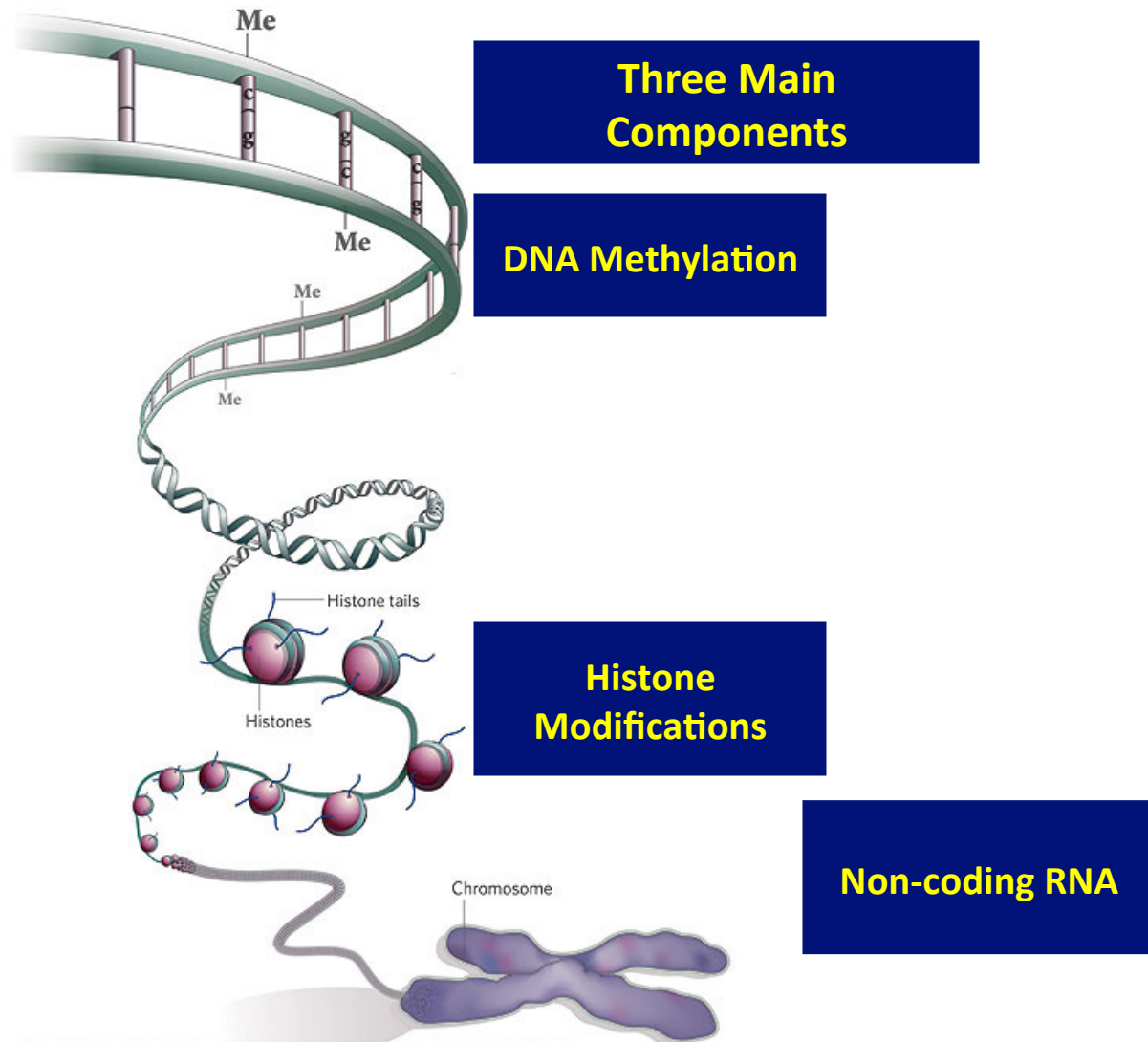
Outline

- Epigenetics
- DNA methylation
- Methylation Assays
- Illumina Bead Arrays
- TCGA methylation data – query data, downloading data
- Analysis

What is Epigenetics?

- The modern definition of epigenetics is modifications of the DNA or associated proteins, other than DNA sequence variation, that carry information content during cell division.
- A heritable modification in gene function or activity
- Epigenetic processes are genetic modifications that affect gene regulation by changing chromatin conformation

Epigenetic Code



Reprinted by permission from Macmillan Publishers Ltd.: *Epigenetics Unfinished Symphony* c2006

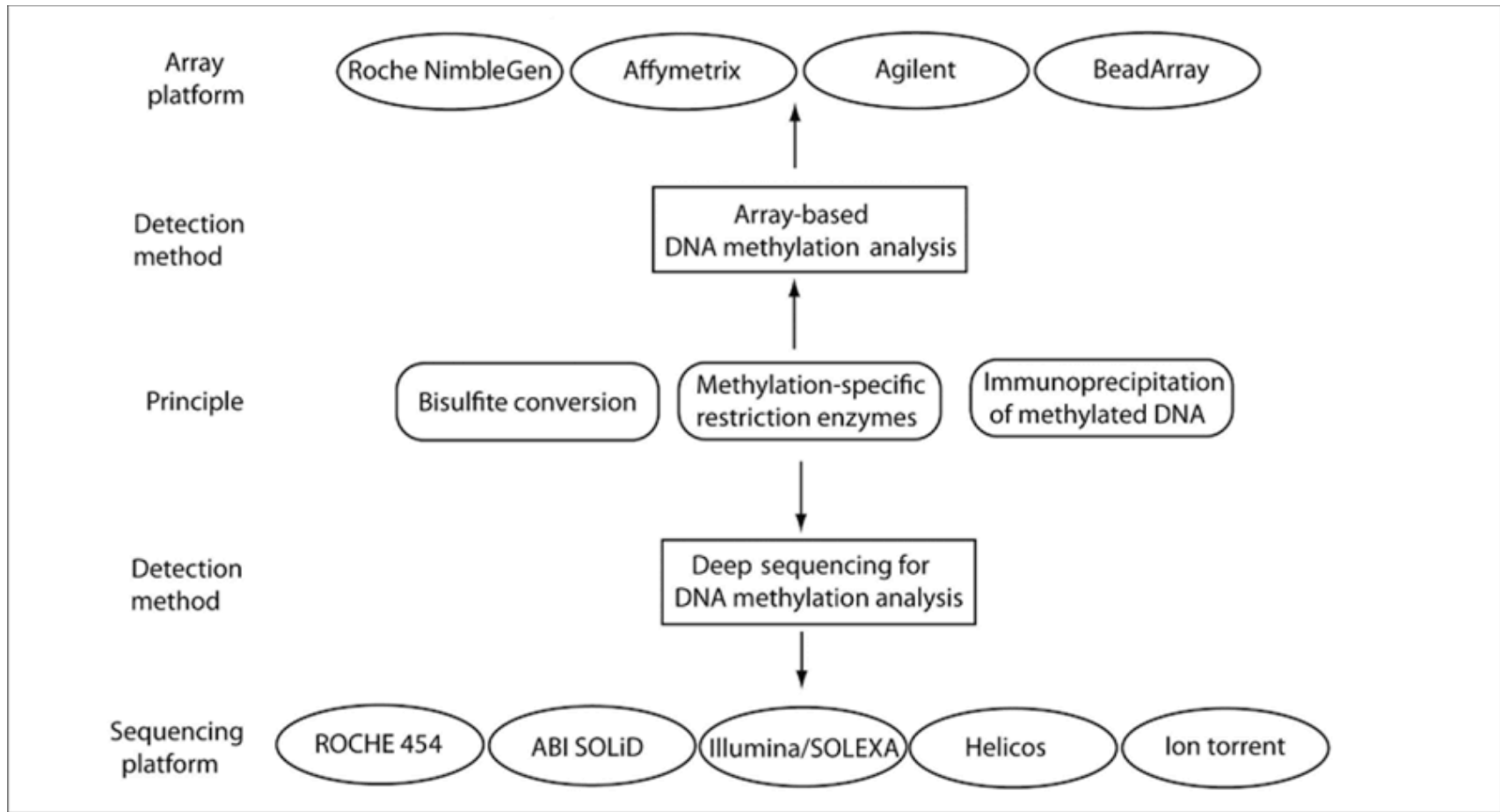
DNA Methylation

- Covalent addition of a methyl (CH_3) group to 5' position of cytosines in CpG dinucleotides
- Reaction is catalyzed by DNA methyltransferases (DNMT)
- In mammals, 70-80% of all CpG dinucleotides are methylated
 - most of this occurs in repetitive elements or regions of low CpG density
- CpG rich regions (CpG Islands):
 - often found in gene promoters 'generally' unmethylated (usual formal definition of a CpG island is a region with at least 200 bp and with a GC percentage that is greater than 50% and with an observed/expected CpG ratio that is greater than 60%)

DNA Methylation

- ~45 000 CpG islands - 50-60% of these are within promoter regions
- CpG islands are primarily unmethylated in normal tissue
- Methylation of CpG islands is associated with delayed replication, condensed chromatin and inhibition of transcription
- Critical for normal development, gene expression and genomic instability
- Aberrant methylation is a common feature of cancer (e.g. hypermethylation of tumor suppressor genes)

High-throughput DNA Methylation Platforms



Requirements for Assay Selection

1) Sample quantity

Low amounts of input DNA

2) Sample quality

3) Sample throughput

96-well formats

4) Gene throughput

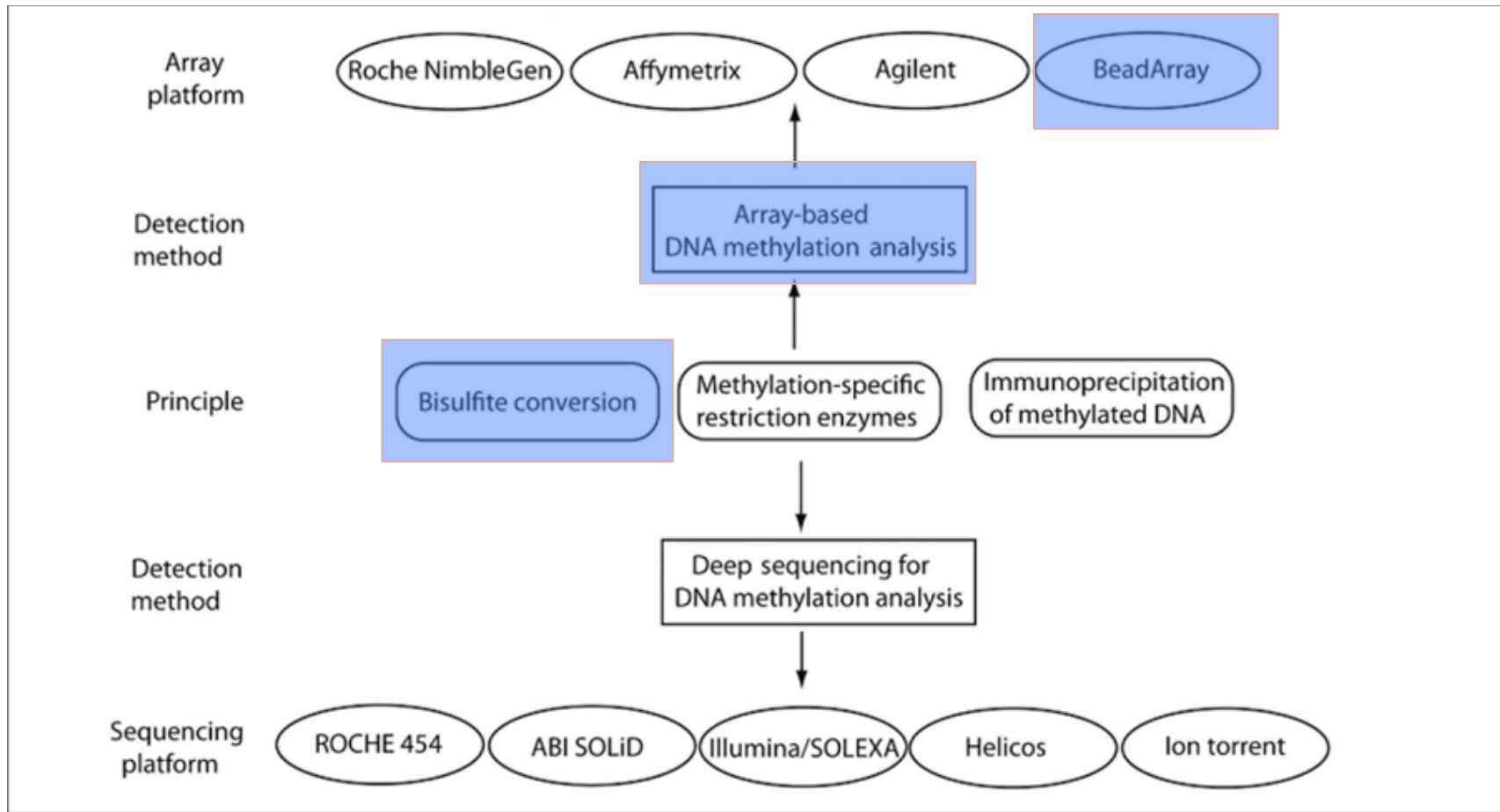
genome wide analysis

5) Avoid regional bias

6) Work flow

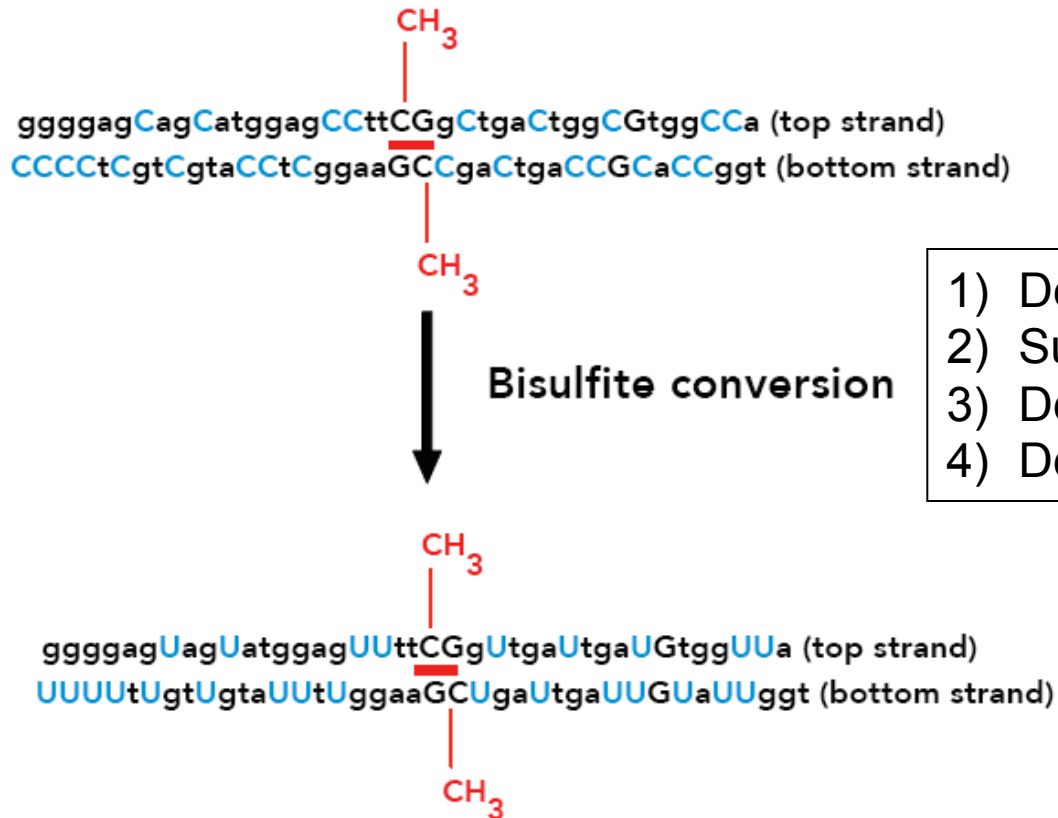
7) Cost

High-throughput DNA Methylation Platforms



Bisulfite Conversion

FIGURE 2: CHANGES IN DNA AFTER BISULFITE TREATMENT



- 1) Denature
- 2) Sulphonate
- 3) Deaminate
- 4) Desulphonate

Epigenetic code → Genetic code

Illumina Bead Array Assay for Methylation

A high-throughput single nucleotide polymorphism (SNP) genotyping system, was adapted for DNA methylation detection to identify bisulfite-converted genomic DNA

ILLUMINA METHYLATION PLATFORMS FOR DNA METHYLATION

- GoldenGate – ~1536 CpG loci
- HumanMethylation27 BeadChip - ~27 500 CpG loci
- 450K Infinium Methylation BeadChip - ~450K CpG loci

HumanMethylation27 BeadChip Kit

- 27, 578 CpG sites
- 14,475 consensus coding sequencing (CCDS) in the NCBI Database (Genome Build 36).
- 110 miRNA promoters
- 144 Methylation hotspots and 982 cancer-related genes
- 2-8 CpGs per target
- single CpG-site resolution, higher sample throughput (up to 12 samples per array),
- requires as little as 500 ng of bisulfite-converted DNA per sample.

HumanMethylation450K BeadChip Kit

- >485 000 methylation sites
- covering 99% of RefSeq genes
- >90% common with HM27K
- 96% of CpG islands and with an average of 17 CpG sites per gene
- Sites are distributed across promoter, 5'UTR, first exon, gene body and 3'UTR.

HumanMethylation450K BeadChip Kit

- CpG islands and shores
- CpG sites outside of CpG islands
- Non-CpG methylated sites identified in human stem cells
- Differentially methylated sites identified in tumor versus normal (multiple forms of cancer) and across several tissue types
- CpG islands outside of coding regions
- miRNA promoter regions
- Disease-associated regions identified through GWAS

Sample Locus:

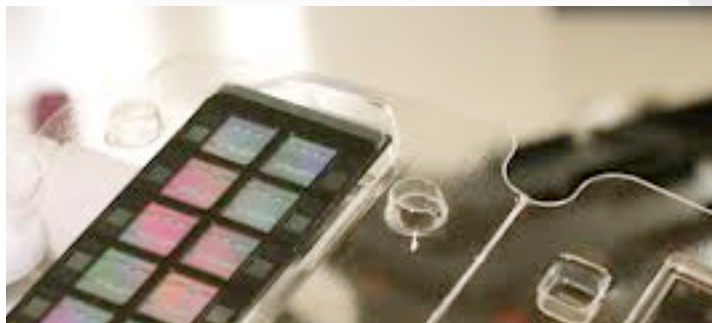
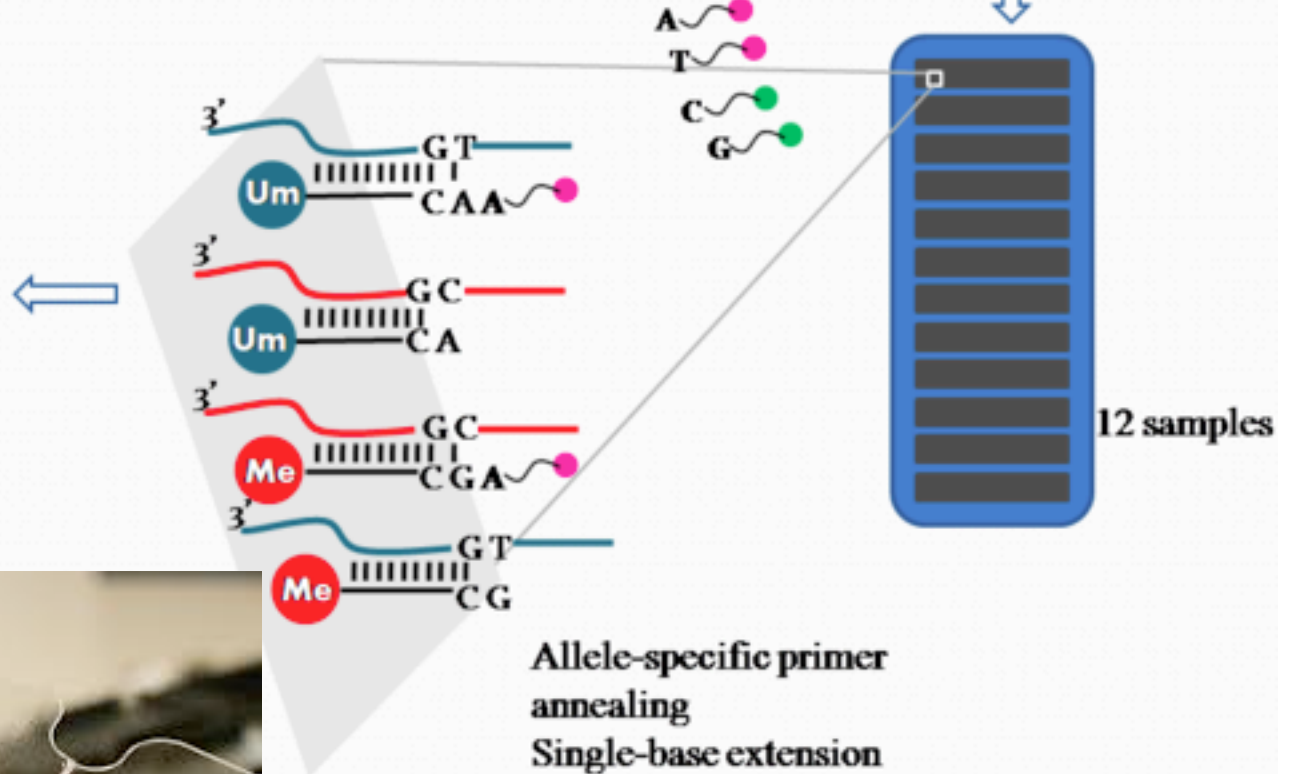
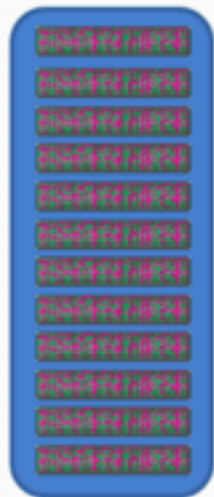
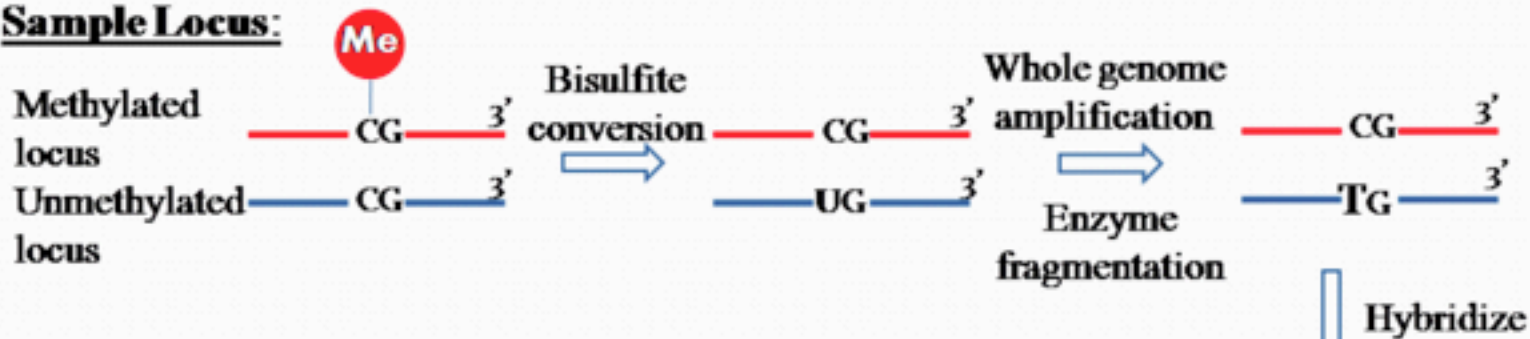
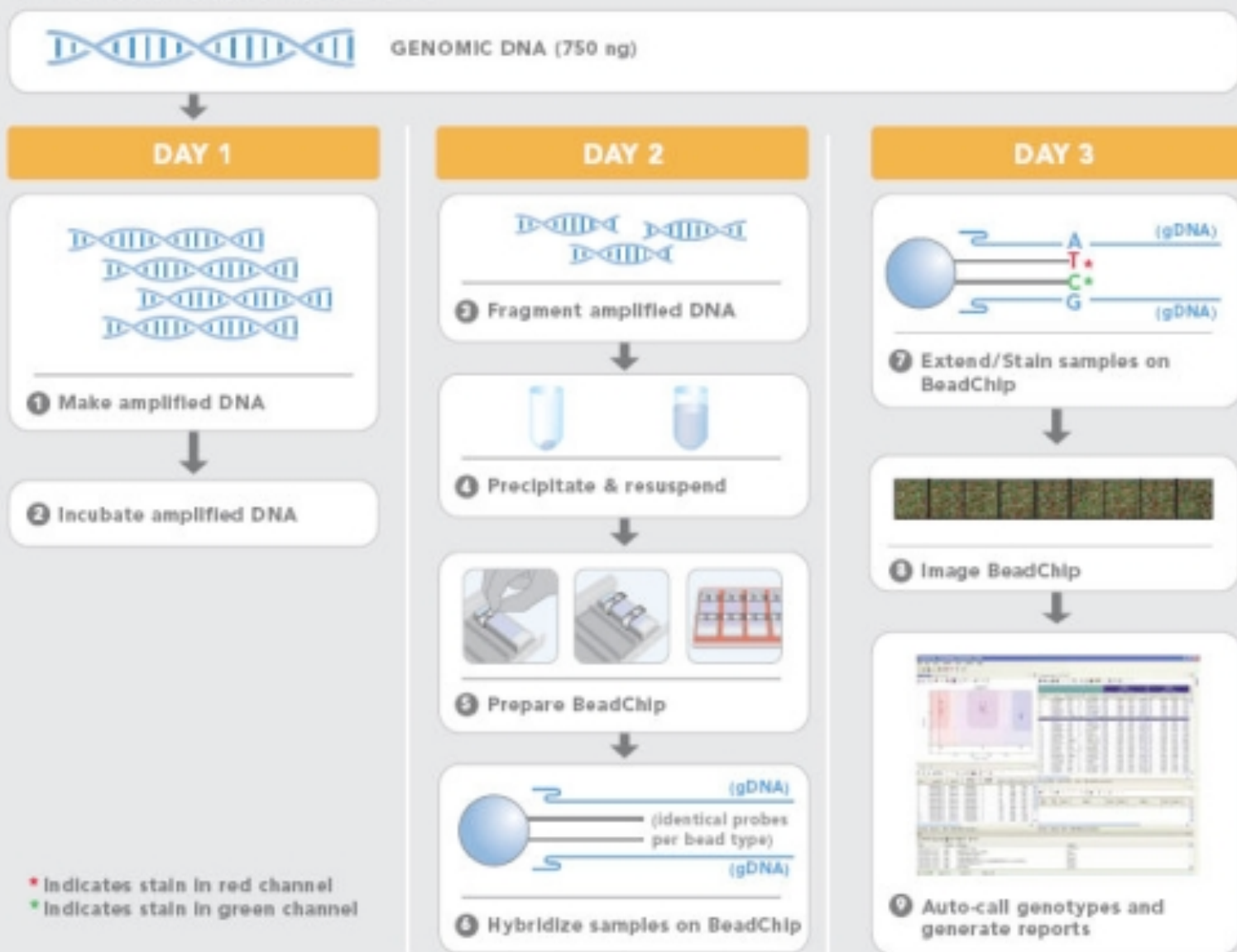
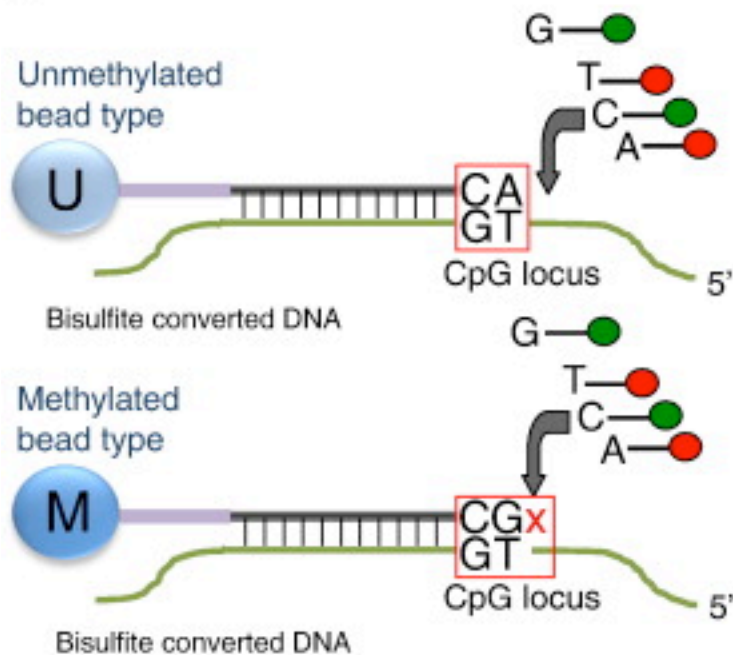
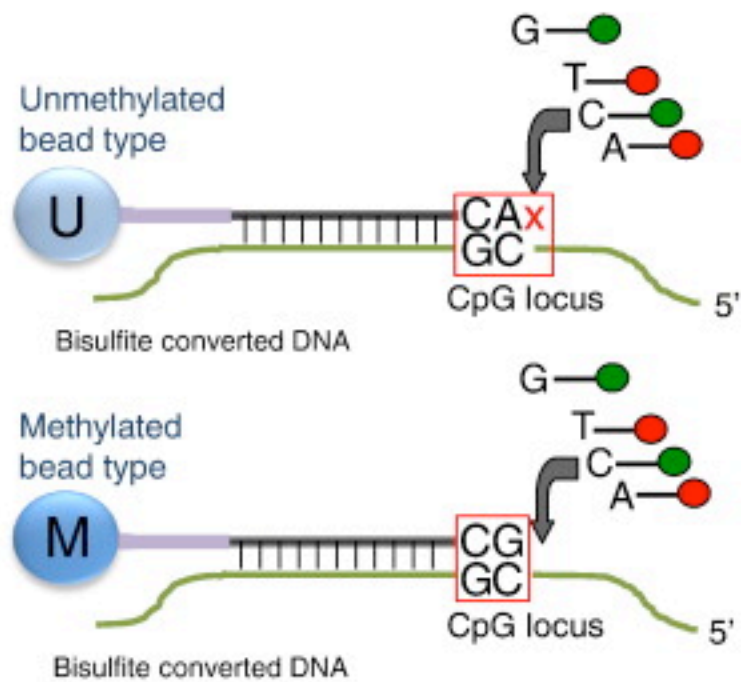


FIGURE 1: INFINIUM II ASSAY PROTOCOL



AHM27K
&
HM450K

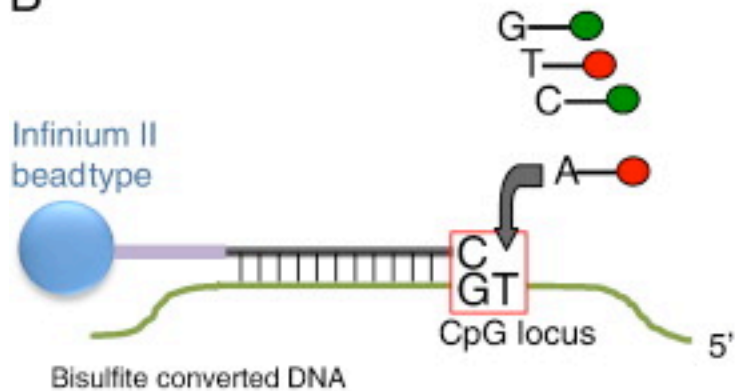
Unmethylated locus



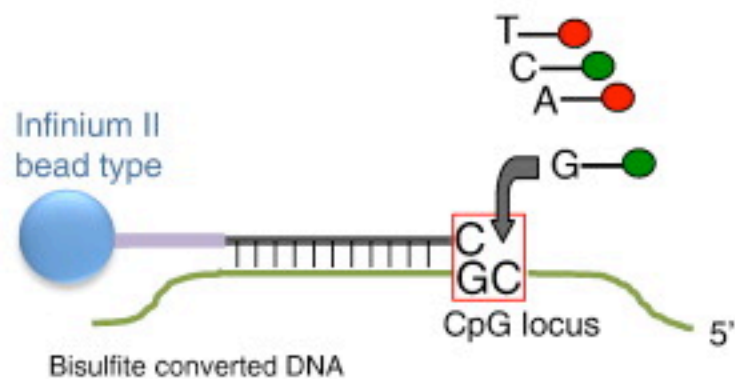
Methylated locus

B

HM450K



Unmethylated locus

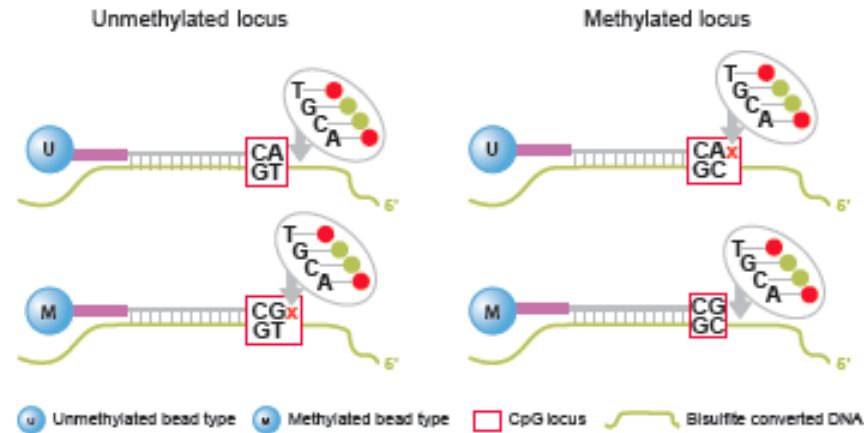


Methylated locus

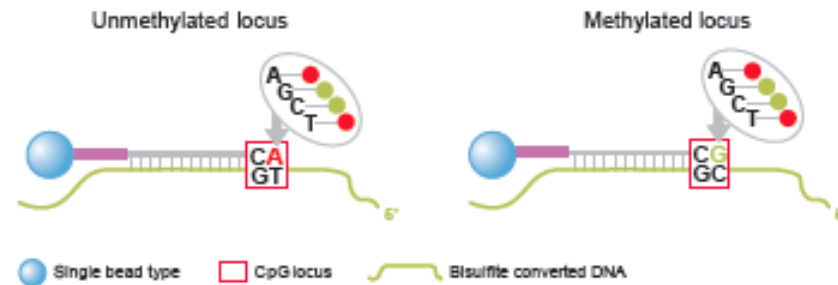
Figure 1: Broader Coverage Using Infinium I and II Assay Designs

HM27K
&
HM450K

A. Infinium I



B. Infinium II



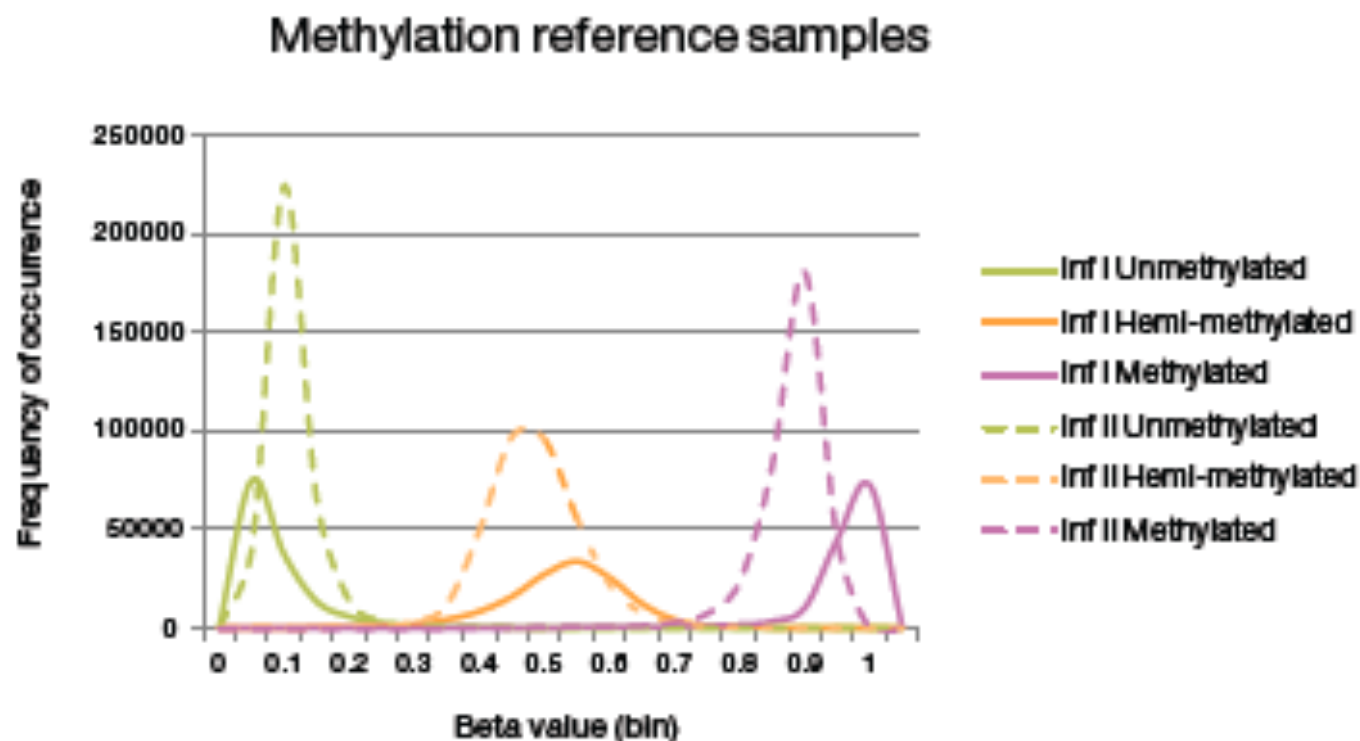
HM450K

The HumanMethylation450 BeadChip employs both Infinium I and Infinium II assays, enhancing its breadth of coverage.

A) Infinium I assay design employs two bead types per CpG locus, one each for the methylated and unmethylated states.

B) The Infinium II design uses one bead type, with the methylated state determined at the single base extension step after hybridization.

Figure 3: Infinium I and Infinium II Chemistries Cover the Full Spectrum of Methylation



The different chemistries of Infinium I and Infinium II assays result in distinct beta value distributions³.

- BeadArrays are scanned using the Illumina iScan technology, and the raw data are extracted using the R-based *methyumi* package²⁵ to calculate the beta value DNA methylation score for each probe and sample.

(can also use GenomeStudio Software, Illumina)

- The level of DNA methylation at each CpG locus is scored as beta (β) value calculated as $(M/(M+U))$, ranging from 0 to 1, with values close to 0 indicating low levels of DNA methylation and beta values close to 1 indicating high levels of DNA methylation.

Website: <http://cancergenome.nih.gov/>



National Cancer Institute

National Human Genome Research Institute



The Cancer Genome Atlas



Understanding genomics
to improve cancer care

[Launch Data Portal](#) | [Contact Us](#) | [For the Media](#)



Home

About Cancer Genomics

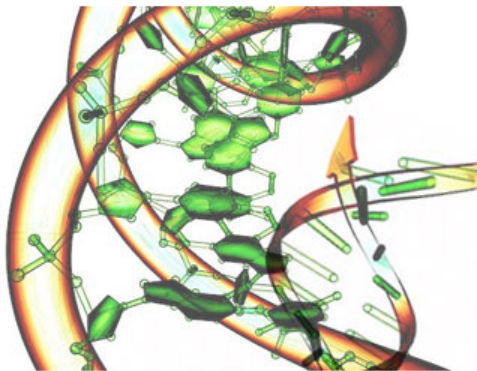
Cancers Selected for Study

Research Highlights

Publications

News and Events

About TCGA



Cancers Selected for Study

The Cancer Genome Atlas researchers are mapping the genetic changes in 20 cancers. Find out which cancers have been selected for study, the criteria for selection and the scientific questions being asked about each cancer.

[Learn More](#)



Researcher Profile



Symposium Videos



Cancers Selected for Study



About Cancer Genomics

Launch Data Portal

The Cancer Genome Atlas (TCGA) Data Portal provides a platform for researchers to search, download, and analyze data sets generated by TCGA.

Questions About Cancer

Visit www.cancer.gov

Call 1-800-4-CANCER

Use LiveHelp Online Chat

Multimedia Library

Images

Videos and Animations

Podcasts

Interactive

News Releases and Announcements

September 21, 2012

Study reveals genomic similarities between breast

Research Briefs

December 2012

Researchers Use TCGA Glioblastoma Data in the Discovery of a Novel Fusion Gene Implicated in a Subset of Brain Tumors



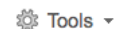
Search

- TCGA Home
- About TCGA
- TCGA Encyclopedia
- Applications
- Data Reports and Dashboards
- Web Services
- TCGA User Documentation
 - Publication Guidelines
 - TCGA Specifications
 - TCGA User's Guides**
- Frequently Asked Questions (FAQ)
- Support and Contacts
- DCC 2.0 Preview Release

TCGA User's Guides

Added by [Brenda Ayala](#), last edited by [Anna Chu](#) on Jul 21, 2011 ([view change](#)) [show comment](#)


None



15 Child Pages

- [Shipped Items Pending BCR Data Submission Report Web Service User Guide](#)
- [TCGA Annotations Manager User's Guide](#)
- [TCGA Annotations Web Service User's Guide](#)
- [TCGA Barcode to UUID Web Service User's Guide](#)
- [TCGA Biospecimen Metadata Browser User Guide](#)
- [TCGA Biospecimen Metadata Web Service User's Guide](#)
- [TCGA Bulk Download User's Guide](#)
- [TCGA Data Browser User's Guide](#)
- [TCGA Data Matrix User's Guide](#)
- [TCGA Data Matrix Web Service User's Guide](#)

Platform Design

 Please note: All links in the table will open in a new window.

Center	TCGA Platform Code [▲]	Platform Name	Sequence Download	TCGA ADF Download
broad.mit.edu	ABI	Applied Biosystems Sequence data	Primers	N/A
genome.wustl.edu	ABI	Applied Biosystems Sequence data	Primers	N/A
hgsc.bcm.edu	ABI	Applied Biosystems Sequence data	Primers	N/A
unc.edu	AgilentG4502A_07_1	Agilent 244K Custom Gene Expression G4502A-07-1	FASTA	TCGA ADF
unc.edu	AgilentG4502A_07_2	Agilent 244K Custom Gene Expression G4502A-07-2	FASTA	TCGA ADF
unc.edu	AgilentG4502A_07_3	Agilent 244K Custom Gene Expression G4502A-07-3	FASTA	TCGA ADF
mskcc.org	CGH-1x1M_G4447A	Agilent SurePrint G3 Human CGH Microarray Kit 1x1M	FASTA	TCGA ADF
broad.mit.edu	Genome_Wide_SNP_6	Affymetrix Genome-Wide Human SNP Array 6.0	FASTA	TCGA ADF
genome.wustl.edu	Genome_Wide_SNP_6	Affymetrix Genome-Wide Human SNP Array 6.0	FASTA	TCGA ADF
unc.edu	H-miRNA_8x15K	Agilent 8 x 15K Human miRNA-specific microarray	FASTA	TCGA ADF
unc.edu	H-miRNA_8x15Kv2	Agilent Human miRNA Microarray Rel12.0	FASTA	TCGA ADF
mskcc.org	HG-CGH-244A	Agilent Human Genome CGH Microarray 244A	FASTA	TCGA ADF
hms.harvard.edu	HG-CGH-244A	Agilent Human Genome CGH Microarray 244A	FASTA	TCGA ADF
hms.harvard.edu	HG-CGH-415K_G4124A	Agilent Human Genome CGH Custom Microarray 2x415K	FASTA	TCGA ADF

Products / **Infinium HumanMethylation450 BeadChip Kit**

Subscribe | | Follow us:

QUESTIONS

The Infinium HumanMethylation450 BeadChip's unique combination of comprehensive, expert-selected coverage, high sample throughput, and affordable price make it an ideal solution for epigenome-wide association studies (EWAS).

Powered by Illumina's revolutionary [Infinium Methylation Assay](#), this BeadChip allows researchers to interrogate > 485,000 methylation sites per sample at single-nucleotide resolution. Content was selected with the guidance of a consortium of methylation experts comprising 22 members that represent 19 institutions worldwide. It covers 99% of RefSeq genes, with an average of 17 CpG sites per gene region distributed across the promoter, 5'UTR, first exon, gene body, and 3'UTR. It covers 96% of CpG islands, with additional coverage in island shores and the regions flanking them. Further content categories requested by the Consortium include:

- CpG sites outside of CpG islands
- Non-CpG methylated sites identified in human stem cells
- Differentially methylated sites identified in tumor versus normal (multiple forms of cancer) and across several tissue types
- CpG islands outside of coding regions
- miRNA promoter regions

This BeadChip includes > 90% of the content contained on the [HumanMethylation27 BeadChip](#). Its 12-sample per array format is compatible with automation, enabling up to 96 samples to be run in parallel (with full LIMS support available) and requiring only 500 ng of input (1 µg with automation). A [protocol enabling compatibility with FFPE samples](#) is now also available. This unique combination of features makes the HumanMethylation450 BeadChip a powerful tool to drive exciting new study designs and further fuel the rapid evolution of epigenetics research.

- [Genotyping Studies with Illumina Omni BeadChips](#)

Catalog IDs: WG-314-1003, WG-314-1001, WG-314-1002



HM-450 BeadChip (24 samples) ▾

Catalog #: **WG-314-1003**

Quantity:

ADD TO CART

Interested in receiving newsletters, case studies, and information on new applications? Enter your email address below.

First

Last

Email

https://tcga-data.nci.nih.gov/tcga/



National Cancer Institute

National Human Genome Research Institute



The Cancer Genome Atlas
Data Portal



Understanding genomics
to improve cancer care

[TCGA Home](#) | [Contact Us](#) | [For the Media](#)

Home

Query the Data

Download Data

Tools

About the Data

Publication Guidelines

Home

TCGA Data Portal Overview

We provide 3 ways to download data: The Cancer Genome Atlas (TCGA) Data Portal provides a platform for researchers to search, download, and analyze data sets generated by TCGA. It contains clinical information, genomic characterization data, and high-throughput sequencing analysis of the tumor genomes.

The TCGA Data Portal does not host lower levels of sequence data. NCI's [Cancer Genomics Hub \(CGHub\)](#) is the new secure repository for storing, cataloging, and accessing sequence related data. New users must still apply for authorized access through NCBI's [Database of Genotypes and Phenotypes \(dbGaP\)](#).

Query the Data >

Search summarized data for
genes, patients and pathways

Download Data >

Choose from three ways to
download data

Available Cancer Types	# Patients with Samples	# Downloadable Tumor Samples	Date Last Updated (mm/dd/yy)
Acute Myeloid Leukemia [LAML]	202	200	02/08/13
Bladder Urothelial Carcinoma [BLCA]	153	153	02/12/13
Brain Lower Grade Glioma [LGG]	222	222	02/08/13

Announcements

01/03/2013 - DCC Software Released

The software release scheduled for today has been completed and the TCGA Portal should be accessible.

A complete list of the issues addressed in this release can be found on the TCGA Wiki [Release Notes](#) and for those with JIRA access the tickets covered in this release can be found on the wiki [here](#).

If you have any questions or concerns about this release, contact tcga-dcc-binf-l@list.nih.gov.

11/13/2012 - DCC Software Released

The software release scheduled for this morning has been completed and the Data Portal has been returned to normal operation. A complete list of the issues addressed in this release can be found on [the TCGA Wiki](#).

If you have any questions or concerns about this release, contact tcga-dcc-binf-l@list.nih.gov.

Data Browser

The output of this application reflects DCC data as of October 2010. Efforts are currently underway to update this site.

IMPORTANT: Data downloaders are urged to use the data annotation search interface (<https://tcga-data.nci.nih.gov/annotations/>) to query the case, sample, and aliquot identifiers in their download to obtain the latest information associated with their data.

Genes

Participants

Pathways

Search Criteria

 "Use same search criteria on all tabs"

Disease Type

GBM - Glioblastoma multiforme

Copy Number - Genes

- Select to Add -

DNA Methylation

- Select to Add -

Genes

 All Genes Chromosome Region [Add](#) Gene List

MGMT

Clear

Copy Number - miRNAs

- Select to Add -

HumanMethylation27

>= 0

Frequency >= 40 %

Gene Expression

- Select to Add -

Validated Somatic Mutations

- Select to Add -

Participants

 All Participants Participants List

miRNA Expression

- Select to Add -

Correlations

- Select to Add -

Data Browser

The output of this application reflects DCC data as of October 2010. Efforts are currently underway to update this site.

IMPORTANT: Data downloaders are urged to use the data annotation search interface (<https://tcga-data.nci.nih.gov/annotations/>) to query the case, sample, and aliquot identifiers in their download to obtain the latest information associated with their data.



Genes | **Participants** | **Pathways**

Search Criteria

Disease

GBM

Platform Type

DNA Methylation (methylation probe)

Center / Platform

jhu-usc.edu / HumanMethylation27

Criteria

>= 0, Frequency >= 40%

Gene List

MGMT

[Modify Search Criteria](#)

Search Results

Genes

Rows per Page:

100

[Export Table](#)

Results as Percentages Results as Ratios

Results 1 - 22 of 22

<input type="checkbox"/> Gene	methylation probe	HumanMethylation27	Gene Location
<input type="checkbox"/> MGMT	cg18488970	100%	Chr 10: 131155443 - 131455773
	cg03271907	100%	Chr 10: 131155443 - 131455773
	cg16698623	100%	Chr 10: 131155443 - 131455773
	cg17686260	100%	Chr 10: 131155443 - 131455773
	cg02381948	100%	Chr 10: 131155443 - 131455773
	cg10333959	100%	Chr 10: 131155443 - 131455773
	cg07453748	100%	Chr 10: 131155443 - 131455773
	cg20778669	100%	Chr 10: 131155443 - 131455773
	cg02941816	100%	Chr 10: 131155443 - 131455773
	cg02941816	100%	Chr 10: 131155443 - 131455773

Data Browser

The output of this application reflects DCC data as of October 2010. Efforts are currently underway to update this site.

IMPORTANT: Data downloaders are urged to use the data annotation search interface (<https://tcga-data.nci.nih.gov/annotations/>) to query the case, sample, and aliquot identifiers in their download to obtain the latest information associated with their data.



Genes | **Participants** | **Pathways**

Search Criteria

Disease

GBM

Platform Type

DNA Methylation (methylation probe)

Center / Platform

jhu-usc.edu / HumanMethylation27

Criteria

>= 0, Frequency >= 40%

Gene List

MGMT

[Modify Search Criteria](#)

Search Results

Genes | **Participants for cg18488970**

Rows per Page: 25

[Export Table](#)

Results 1 - 25 of 251

<input type="checkbox"/> Participant	HumanMethylation27
<input type="checkbox"/> TCGA-02-0001	0.674
<input type="checkbox"/> TCGA-02-0003	0.706
<input type="checkbox"/> TCGA-02-0006	0.844
<input type="checkbox"/> TCGA-02-0007	0.758
<input type="checkbox"/> TCGA-02-0009	0.660
<input type="checkbox"/> TCGA-02-0010	0.786
<input type="checkbox"/> TCGA-02-0011	0.264
<input type="checkbox"/> TCGA-02-0014	0.224

Data Browser

The output of this application reflects DCC data as of C

IMPORTANT: Data downloaders are urged to use the c identifiers in their download to obtain the latest informa

Genes **Participants** **Pathways**

Search Criteria

Disease

GBM

Platform Type

Expression-Genes (gene)

DNA Methylation (methylation probe)

Gene List

MGMT

Search Results

Genes **Participants for cg18488970** ✕

Rows per Page: 25

<input type="checkbox"/> Participant ↑	HumanMethylation27
<input type="checkbox"/> TCGA-02-0001	0.674
<input type="checkbox"/> TCGA-02-0003	0.706
<input type="checkbox"/> TCGA-02-0006	0.844
<input type="checkbox"/> TCGA-02-0007	0.758
<input type="checkbox"/> TCGA-02-0009	0.660
<input type="checkbox"/> TCGA-02-0010	0.786
<input type="checkbox"/> TCGA-02-0011	0.264
<input type="checkbox"/> TCGA-02-0014	0.224
<input type="checkbox"/> TCGA-02-0021	0.679

identifiers in their download to obtain the latest info

Genes **Participants** **Pathways**

Search Criteria

Disease

GBM

Platform Type

Expression-Genes (gene)

DNA Methylation (methylation probe)

Gene List

MGMT

annotations/) to query the case, sample, and aliquot

[Help](#)

Criteria

<= -0.5 or >= 0.5, Frequency >= 40%

>= 0, Frequency >= 40%

Search Results

Genes **Participants for cg18488970** ✕

Rows per Page: 25

<input type="checkbox"/> Participant ↑	AgilentG4502A_07 log2 tumor/normal ratio
<input type="checkbox"/> TCGA-02-0009	-2.103
<input type="checkbox"/> TCGA-02-0010	-0.957
<input type="checkbox"/> TCGA-02-0011	-0.551
<input type="checkbox"/> TCGA-02-0014	-1.101
<input type="checkbox"/> TCGA-02-0015	-1.148
<input type="checkbox"/> TCGA-02-0016	-1.398
<input type="checkbox"/> TCGA-02-0026	-0.785
<input type="checkbox"/> TCGA-02-0028	-1.994
<input type="checkbox"/> TCGA-02-0033	-0.913
<input type="checkbox"/> TCGA-02-0037	-0.593
<input type="checkbox"/> TCGA-02-0038	0.635

Download Data

The TCGA Data Portal does not host lower levels of sequence data. NCI's [Cancer Genomics Hub \(CGHub\)](#) is the new secure repository for storing, cataloging, and accessing sequence related data. New users must still apply for authorized access through NCBI's [Database of Genotypes and Phenotypes \(dbGaP\)](#)

We provide 3 ways to download data:

Method	What it offers	When to use it
Data Matrix	Select and download subsets of data by center, platform and data types. Includes: Level 1, 2 and 3 data Access the FAQ	Use when: <ul style="list-style-type: none">You want to download data as tab-delimited textYou only want a subset of the data
Bulk Download	A form that helps you locate files in the data archives. Includes: Level 1, 2, 3 and limited level 4 data	Use when: <ul style="list-style-type: none">You want to download bulk datasets as provided by the research centers
Access HTTP Directories <ul style="list-style-type: none">Open-access HTTP DirectoryControlled-Access HTTP Directory	Direct access to the HTTP directories where the data archives are stored. Includes: Level 1, 2, 3 and limited level 4 data. Login is required for the Controlled-access HTTP Directory. See controlled-access requirements .	Use when: <ul style="list-style-type: none">You know how to use HTTP directories and you prefer to find files yourself rather than use the Bulk Download form

▼ In This Section

[Download Data](#)[Data Matrix](#)[Bulk Download](#)[Open-Access HTTP Directory](#)[Controlled-Access HTTP Directory](#)

Controlled-Access Requirements

The controlled access data tier contains clinical data and individually unique information. This tier requires user certification for data access.

[Controlled access requirements](#)

User Guides and Help

[Data Matrix User's Guide](#)  [↗](#)[TCGA Data Guide](#) [↗](#)

Relationship of Data Levels to Data Types

Data Type	Data Subtypes	Technology Platform(s)	Level 1	Level 2	Level 3	Important Metadata
DNA Methylation		HumanMethylation27 HumanMethylation450 IlluminaDNAMethylation_OMA002_CPI IlluminaDNAMethylation_OMA003_CPI	Raw signal intensities of probes <u>File type:</u> tab-delimited (.txt) and binary [(.idat) (applies to HumanMethylation450 platform)]	Calculated beta values <u>File type:</u> tab-delimited (.txt)	Calculated beta values mapped to genome, per sample <u>File type:</u> tab-delimited (.txt)	Experimental protocol, including calculation methods, is included in the MAGE-TAB archive Probe information is contained in the Array design files for each platform

Data Matrix

Select initial matrix filter settings. To view all data, click [here](#) or click "Apply" without choosing any settings. (Note: unfiltered matrix is large and can take some time to load.)

Filter Settings

Select a disease:

BRCA – Breast invasive carcinoma

Data Type:

All
Clinical
DNA Methylation
Expression–Genes

Batch Number:

All
Batch 0
Batch 47
Batch 56

Data Level:

- Level 1
 Level 2
 Level 3

Availability:

- Available
 Pending
 Not Available

Center/Platform:

TCGA (HumanMethylation27) (All)
BI (Genome_Wide_SNP_6)
JHU_USC (HumanMethylation27)
JHU_USC (HumanMethylation450)
MDA (MDA_RPPA_Core)

Sample:

ID Matches:

TCGA- -- -- [Remove](#)

[Add Row](#)

Paste Sample List:

Upload Sample List:

 [Browse...](#)


Access Tier:

- All
 Protected
 Public


Tumor/Normal:

- Tumor - matched
 Tumor - unmatched
 Normal - matched
 Normal - unmatched

Submitted Since
(Date):

mm/dd/yyyy 

Submitted Up To
(Date):

mm/dd/yyyy 

- Only show samples with data available for all columns

Get web service URL for this filter



Apply

Clear

▶ In This Section

Data Matrix Datasets

BRCA Data Matrix
 Options:

[Help](#)

Color Cells By:

Scroll Size:

Legend:

- TN Tumor, Matched Normal
- T Tumor, no Matched Normal
- NT Normal, Matched Tumor
- N Normal, no Matched Tumor
- No Available Dataset

Red = protected data

Batch/Sample	Level	JHU-USC HumanMethylation27	JHU-USC HumanMethylation450	Methyl
Batch 47	TCGA-A7-A0DC-01	<input type="checkbox"/>	<input checked="" type="checkbox"/>	TN
Batch 72	TCGA-BH-A0BC-01	<input type="checkbox"/>	<input checked="" type="checkbox"/>	TN
	TCGA-BH-A0DH-01	<input type="checkbox"/>	<input checked="" type="checkbox"/>	TN
	TCGA-BH-A0DQ-01	<input type="checkbox"/>	<input checked="" type="checkbox"/>	TN
Batch 80	TCGA-BH-A0B2-01	<input type="checkbox"/>	<input checked="" type="checkbox"/>	TN
Batch 85	TCGA-BH-A0B7-01	<input checked="" type="checkbox"/>	<input type="checkbox"/>	TN
	TCGA-BH-A0BL-01	<input checked="" type="checkbox"/>	<input type="checkbox"/>	TN
	TCGA-BH-A0BQ-01	<input checked="" type="checkbox"/>	<input type="checkbox"/>	TN
	TCGA-BH-A0BW-01	<input checked="" type="checkbox"/>	<input type="checkbox"/>	TN

Data Download

Enter E-mail Address:

Re-Enter E-mail Address:

Estimated Uncompressed Size: 33.702 MB

70 Gb maximum allowable size

Archive Options:

 Flatten directory structure

Please enter and confirm your e-mail address. Upon selecting "Download", your files will be tar'd and gzip'd. When completed, an e-mail will be sent to you with a link to your file. This file will remain on the server for 24 hours. A link to the file will also appear in the browser window.

Select files to include in your archive:

- METADATA
 - JHU_USC (HumanMethylation27)
 - selected_samples::jhu-usc.edu_BRCA.HumanMethylation27.1.6.0.idf.txt (1.495 KiB)
 - selected_samples::jhu-usc.edu_BRCA.HumanMethylation27.1.6.0.srdf.txt (706.87 KiB)
- DNA Methylation
 - JHU_USC (HumanMethylation27)
 - Level 3
 - TCGA-BH-A0B7-01::jhu-usc.edu__HumanMethylation27__TCGA-BH-A0B7-01A-12D-A112-05__methylation_analysis.txt (1.223 MiB)
 - TCGA-BH-A0BL-01::jhu-usc.edu__HumanMethylation27__TCGA-BH-A0BL-01A-11D-A112-05__methylation_analysis.txt (1.223 MiB)
 - TCGA-BH-A0BO-01::jhu-usc.edu__HumanMethylation27__TCGA-BH-A0BO-01A-23D-A12E-05__methylation_analysis.txt (1.223 MiB)
 - TCGA-BH-A0BQ-01::jhu-usc.edu__HumanMethylation27__TCGA-BH-A0BQ-01A-21D-A112-05__methylation_analysis.txt (1.223 MiB)
 - TCGA-BH-A0BW-01::jhu-usc.edu__HumanMethylation27__TCGA-BH-A0BW-01A-11D-A112-05__methylation_analysis.txt (1.224 MiB)
 - TCGA-BH-A0DE-01::jhu-usc.edu__HumanMethylation27__TCGA-BH-A0DE-01A-11D-A112-05__methylation_analysis.txt (1.221 MiB)
 - TCGA-BH-A0DL-01::jhu-usc.edu__HumanMethylation27__TCGA-BH-A0DL-01A-11D-A112-05__methylation_analysis.txt (1.224 MiB)
 - TCGA-BH-A0DO-01::jhu-usc.edu__HumanMethylation27__TCGA-BH-A0DO-01B-11D-A12E-05__methylation_analysis.txt (1.222 MiB)
 - TCGA-BH-A0DT-01::jhu-usc.edu__HumanMethylation27__TCGA-BH-A0DT-01A-21D-A12E-05__methylation_analysis.txt (1.223 MiB)
 - TCGA-BH-A0H5-01::jhu-usc.edu__HumanMethylation27__TCGA-BH-A0H5-01A-21D-A112-05__methylation_analysis.txt (1.221 MiB)
 - TCGA-BH-A18F-01::jhu-usc.edu__HumanMethylation27__TCGA-BH-A18F-01A-11D-A12E-05__methylation_analysis.txt (1.221 MiB)
 - TCGA-BH-A18J-01::jhu-usc.edu__HumanMethylation27__TCGA-BH-A18J-01A-11D-A12E-05__methylation_analysis.txt (1.222 MiB)

The archive you created is available at https://tcga-data.nci.nih.gov/tcgafiles/ftp_auth/distro_ftpusers/anonymous/userCreatedArchives/b53f7424-105b-45d3-b031-099128b9de02.tar.gz

It will be available for download for 24 hours, after which it will be deleted from our servers.

The following filter settings were used for the search criteria:

Disease: BRCA

Platform Type: DNA Methylation

Center (Platform): JHU_USC (HumanMethylation27), JHU_USC (HumanMethylation450)

Level: 3

Availability: Available

Archive Processing Details:

Total file processing time: 00:50

Total archive generation time: 00:07

Total processing time: 00:57

When added to queue: 1:18:44 PM

Time waiting in queue: 00:00

Metadata file

Investigation Title	TCGA Analysis of DNA Methylation for BRCA using Illumina Infinium HumanMethylation27 platform						
Experimental Design	disease_state_design						
Experimental Design Term Source REF	MGED Ontology						
Experimental Factor Type	disease						
Experimental Factor Type Term Source REF	MGED Ontology						
Person Last Name	Laird						
Person First Name	Peter						
Person Mid Initials	W						
Person Email	plaird@usc.edu						
Person Phone	323.442.7890						
Person Address	USC Epigenome Center, University of Southern California, CA 90033, USA						
Person Affiliation	University of Southern California						
Person Roles	submitter						
Quality Control Types	real_time_PCR_quality_control						
Quality Control Types Term Source REF	MGED Ontology						
Replicate Type	bioassay_replicate_reduction						
Replicate Type Term Source REF	MGED Ontology						
Public Release Date	Fri Oct 26 18:04:07 2012						
Protocol Name	jhu-usc.edu:labeling:HumanMethylation27:01	jhu-usc.edu:hybridization	jhu-usc.edu:image_acquisition:HumanMethylation27:01				
Protocol Type	labeling	hybridization	scan				
Protocol Term Source REF	MGED Ontology	MGED Ontology	MGED Ontology				
Protocol Description	Illumina Infinium HumanMethylation27 Labeled Extract	Illumina Infinium HumanMethylation27	Illumina Infinium HumanMethylation27 Scan Protocol				
Protocol Parameters							
SDRF Files	jhu-usc.edu_BRCA.HumanMethylation27.1.6.0.sdrf.txt						
Term Source Name	MGED Ontology	caArray					
Term Source File	http://mged.sourceforge.net/ontologies/MGEDontology.php	http://caarraydb.nci.nih.gov/					
Term Source Version	1.3.1.1	2007-01					

▼ b53f7424-105b-45d3-b031-099128b9de02

▼ DNA_Methylation

▼ JHU_USC__HumanMethylation27

▼ Level_3

- 📄 jhu-usc.edu__HumanMethylation27_TCGA-E2-A15M-01A-11D-A12E-05__methylation_analysis.txt
- 📄 jhu-usc.edu__HumanMethylation27_TCGA-E2-A15L-01A-11D-A12E-05__methylation_analysis.txt
- 📄 jhu-usc.edu__HumanMethylation27_TCGA-E2-A158-01A-11D-A12E-05__methylation_analysis.txt
- 📄 jhu-usc.edu__HumanMethylation27_TCGA-E2-A153-01A-12D-A12E-05__methylation_analysis.txt
- 📄 jhu-usc.edu__HumanMethylation27_TCGA-E2-A14Y-01A-21D-A12E-05__methylation_analysis.txt
- 📄 jhu-usc.edu__HumanMethylation27_TCGA-BH-A18V-01A-11D-A12E-05__methylation_analysis.txt
- 📄 jhu-usc.edu__HumanMethylation27_TCGA-BH-A18U-01A-21D-A12E-05__methylation_analysis.txt
- 📄 jhu-usc.edu__HumanMethylation27_TCGA-BH-A18S-01A-11D-A12E-05__methylation_analysis.txt
- 📄 jhu-usc.edu__HumanMethylation27_TCGA-BH-A18R-01A-11D-A12E-05__methylation_analysis.txt
- 📄 jhu-usc.edu__HumanMethylation27_TCGA-BH-A18Q-01A-12D-A12E-05__methylation_analysis.txt
- 📄 jhu-usc.edu__HumanMethylation27_TCGA-BH-A18P-01A-11D-A12E-05__methylation_analysis.txt
- 📄 jhu-usc.edu__HumanMethylation27_TCGA-BH-A18N-01A-11D-A12E-05__methylation_analysis.txt
- 📄 jhu-usc.edu__HumanMethylation27_TCGA-BH-A18M-01A-11D-A12E-05__methylation_analysis.txt
- 📄 jhu-usc.edu__HumanMethylation27_TCGA-BH-A18L-01A-32D-A12E-05__methylation_analysis.txt
- 📄 jhu-usc.edu__HumanMethylation27_TCGA-BH-A18K-01A-11D-A12E-05__methylation_analysis.txt
- 📄 jhu-usc.edu__HumanMethylation27_TCGA-BH-A18J-01A-11D-A12E-05__methylation_analysis.txt
- 📄 jhu-usc.edu__HumanMethylation27_TCGA-BH-A18F-01A-11D-A12E-05__methylation_analysis.txt
- 📄 jhu-usc.edu__HumanMethylation27_TCGA-BH-A0H5-01A-21D-A112-05__methylation_analysis.txt
- 📄 jhu-usc.edu__HumanMethylation27_TCGA-BH-A0DT-01A-21D-A12E-05__methylation_analysis.txt
- 📄 jhu-usc.edu__HumanMethylation27_TCGA-BH-A0DC-01B-11D-A12E-05__methylation_analysis.txt
- 📄 jhu-usc.edu__HumanMethylation27_TCGA-BH-A0DL-01A-11D-A112-05__methylation_analysis.txt
- 📄 jhu-usc.edu__HumanMethylation27_TCGA-BH-A0DE-01A-11D-A112-05__methylation_analysis.txt
- 📄 jhu-usc.edu__HumanMethylation27_TCGA-BH-A0BW-01A-11D-A112-05__methylation_analysis.txt
- 📄 jhu-usc.edu__HumanMethylation27_TCGA-BH-A0BQ-01A-21D-A112-05__methylation_analysis.txt
- 📄 jhu-usc.edu__HumanMethylation27_TCGA-BH-A0BO-01A-23D-A12E-05__methylation_analysis.txt
- 📄 jhu-usc.edu__HumanMethylation27_TCGA-BH-A0BL-01A-11D-A112-05__methylation_analysis.txt
- 📄 jhu-usc.edu__HumanMethylation27_TCGA-BH-A0B7-01A-12D-A112-05__methylation_analysis.txt

▼ METADATA

▼ JHU_USC__HumanMethylation27

- 📄 jhu-usc.edu_BRCA.HumanMethylation27.1.6.0.idf.txt
- 📄 jhu-usc.edu_BRCA.HumanMethylation27.1.6.0.sdrf.txt

📄 file_manifest.txt

Label	Identifier for	Value	Value description	Possible values
Project	Project name	TCGA	TCGA project	TCGA
TSS	Tissue source site	02	GBM (brain tumor) sample from MD Anderson	See Code Tables Report
Participant	Study participant	0001	The first participant from MD Anderson for GBM study	Any alpha-numeric value
Sample	Sample type	01	A solid tumor	Tumor types range from 01 - 09, normal types from 10 - 19 and control samples from 20 - 29. See Code Tables Report for a complete list of sample codes
Vial	Order of sample in a sequence of samples	C	The third vial	A to Z
Portion	Order of portion in a sequence of 100 - 120 mg sample portions	01	The first portion of the sample	01-99
Analyte	Molecular type of analyte for analysis	D	The analyte is a DNA sample	See Code Tables Report
Plate	Order of plate in a sequence of 96-well plates	0182	The 182nd plate	4-digit alphanumeric value
Center	Sequencing or characterization center that will receive the aliquot for analysis	01	The Broad Institute GCC	See Code Tables Report

Level 3 Methylation Data

A	B	C	D	E	F	G
barcode	probe name	beta value	gene symbol	chromosome	position	
TCGA-E2-A158-11A-22D-A12E-05	cg03817621	0.89120883	A1CF	10	52645399	
TCGA-E2-A158-11A-22D-A12E-05	cg03586879	0.03511464	A2BP1	16	6068831	
TCGA-E2-A158-11A-22D-A12E-05	cg19378133	0.03900802	A2BP1	16	6068835	
TCGA-E2-A158-11A-22D-A12E-05	cg12058490	0.38538459	A2M	12	9268379	
TCGA-E2-A158-11A-22D-A12E-05	cg03490200	0.87331792	A2ML1	12	8975196	
TCGA-E2-A158-11A-22D-A12E-05	cg27653134	0.68920504	A2ML1	12	8975309	
TCGA-E2-A158-11A-22D-A12E-05	cg09744051	0.14460686	A4GALT	22	43115937	
TCGA-E2-A158-11A-22D-A12E-05	cg07393322	0.03561774	A4GALT	22	43117318	
TCGA-E2-A158-11A-22D-A12E-05	cg17687282	0.83103132	A4GNT	3	137851111	
TCGA-E2-A158-11A-22D-A12E-05	cg18931888	NA	A4GNT	3	137851874	
TCGA-E2-A158-11A-22D-A12E-05	cg23032316	0.06880887	AAAS	12	53715148	
TCGA-E2-A158-11A-22D-A12E-05	cg00559473	0.13789196	AAAS	12	53715592	
TCGA-E2-A158-11A-22D-A12E-05	cg05303448	NA	AACS	12	125549535	
TCGA-E2-A158-11A-22D-A12E-05	cg03013422	0.07758769	AACS	12	125550350	
TCGA-E2-A158-11A-22D-A12E-05	cg05564657	0.89838289	AADAC	3	151531375	
TCGA-E2-A158-11A-22D-A12E-05	cg25159668	0.70372119	AADACL2	3	151450877	
TCGA-E2-A158-11A-22D-A12E-05	cg12506971	0.01723199	AADAT	4	171011264	
TCGA-E2-A158-11A-22D-A12E-05	cg01654582	0.09069938	AADAT	4	171011577	
TCGA-E2-A158-11A-22D-A12E-05	cg22940988	0.90857662	AAGAB	15	67538446	
TCGA-E2-A158-11A-22D-A12E-05	cg01272601	0.02228249	AAK1	2	69871685	
TCGA-E2-A158-11A-22D-A12E-05	cg21613754	0.89502439	AAMP;PNKD	2	219133847	
TCGA-E2-A158-11A-22D-A12E-05	cg14864603	NA	AANAT	17	74462846	
TCGA-E2-A158-11A-22D-A12E-05	cg09382492	0.59121998	AANAT	17	74463681	
TCGA-E2-A158-11A-22D-A12E-05	cg25432323	0.03732347	AARS	16	70322827	
TCGA-E2-A158-11A-22D-A12E-05	cg09102714	0.0369942	AARS	16	70323758	
TCGA-E2-A158-11A-22D-A12E-05	cg16356516	0.94363319	AARS;EXOSC	16	70286390	
TCGA-E2-A158-11A-22D-A12E-05	cg04429140	NA	AARSD1;RUN	17	41132427	
TCGA-E2-A158-11A-22D-A12E-05	cg02426410	0.01823468	AASDH	4	57253603	
TCGA-E2-A158-11A-22D-A12E-05	cg03289369	0.02222164	AASDH	4	57253815	
TCGA-E2-A158-11A-22D-A12E-05	cg26289824	0.01362622	AASDH;PPT;K	11	105948485	
TCGA-E2-A158-11A-22D-A12E-05	cg06667406	0.93599915	AASS	7	121773889	
TCGA-E2-A158-11A-22D-A12E-05	cg11857445	0.01748868	AATF	17	35305905	

Meaning of “NA”

- *Level 2:* P values >0.05 were not considered to be significantly different from background
- *Level 3:*
 - 1) contain known single nucleotide polymorphisms (SNPs) after comparison to the dbSNP database (Build 130),
 - 2) contain repetitive sequence elements that cover the targeted CpG locus in each 50 bp probe sequence
 - 3) are not uniquely aligned to the human genome (NCBI build 36.1) at 20 nucleotides at the 3' terminus of the probe sequence
 - 4) span known regions of small insertions and deletions (indels) in the human genome (dbSNP build 130).
 - 5) 450K probes that are annotated as having a SNP within 10 base pairs of the interrogated locus (HM27 carryover or recently discovered, dbSNP build 131)

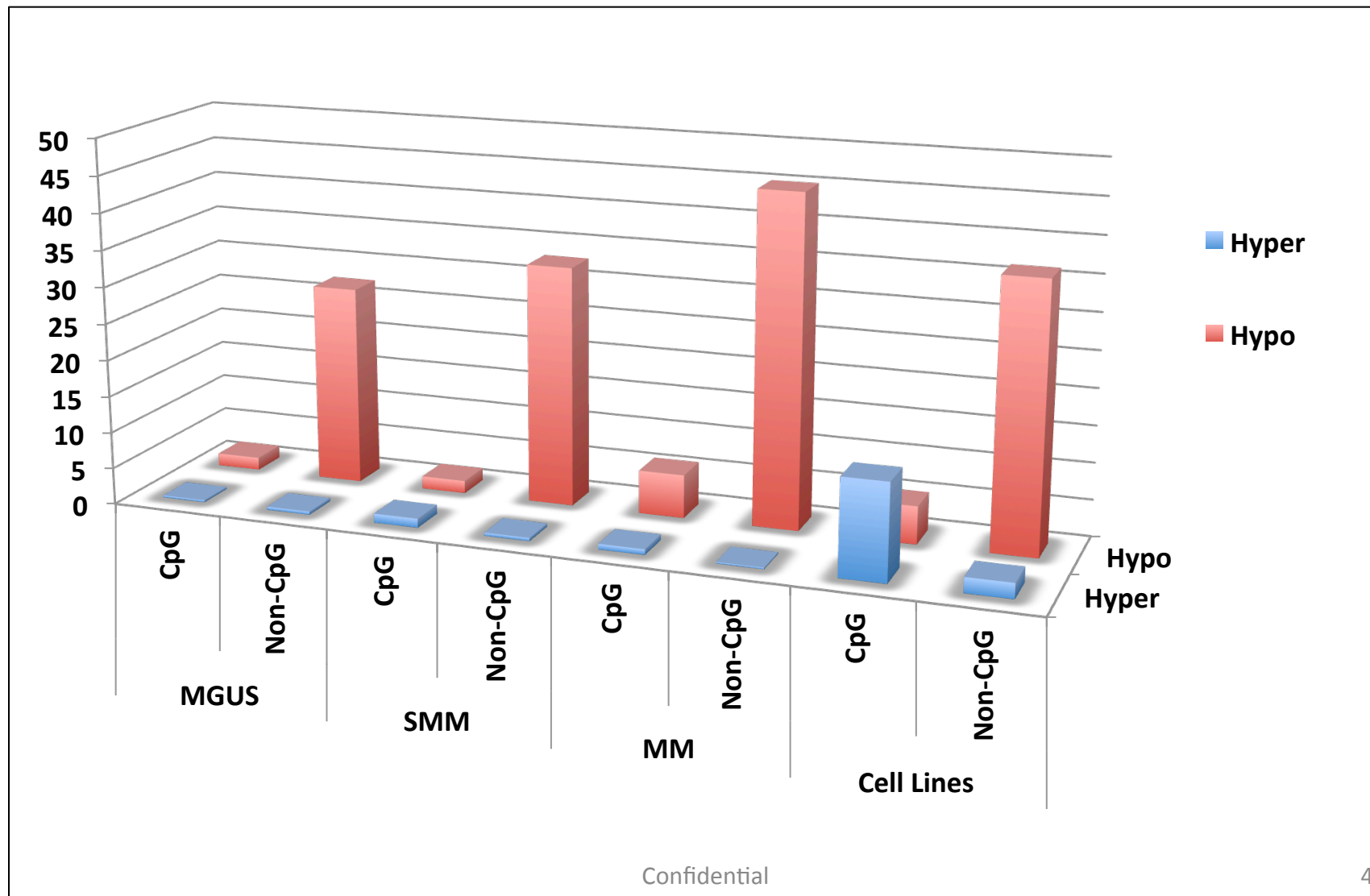
Methylation Analysis

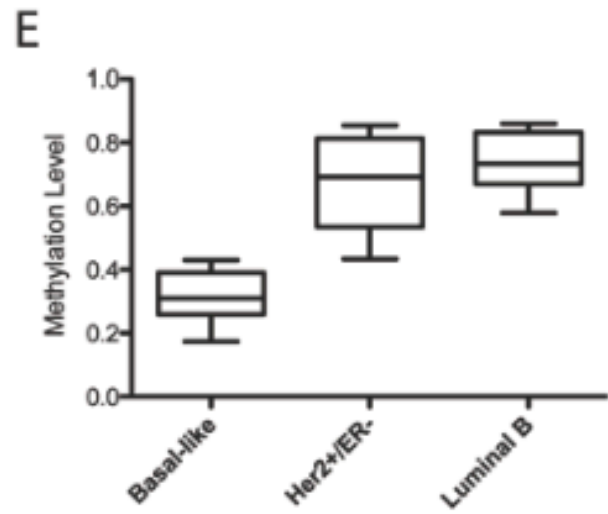
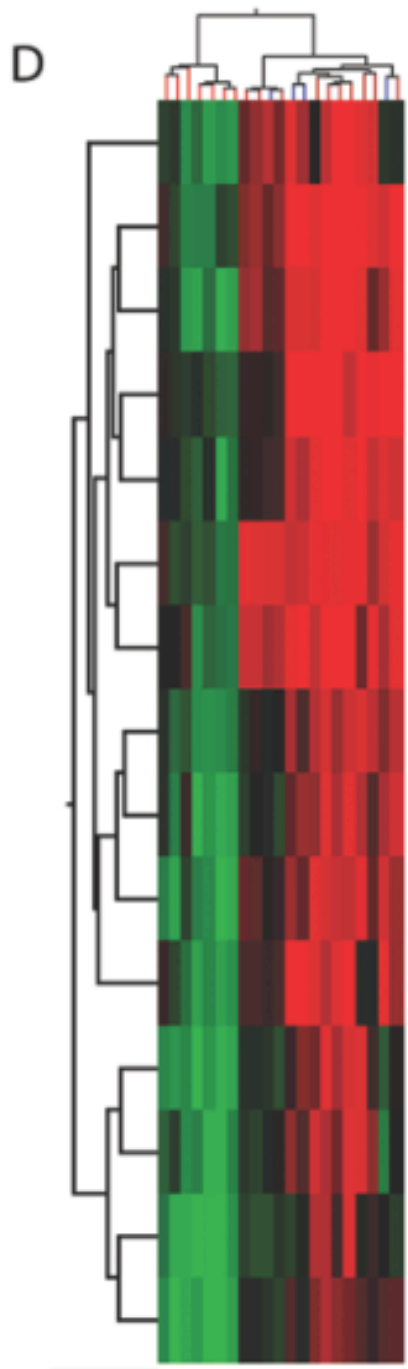
- Differentially methylated locus (DML) = a locus that differed by more than 20% when compared to normal samples and survived statistical significance ($p < 0.05$) - Delta beta value
- Distribution
- Clustering
- Regional

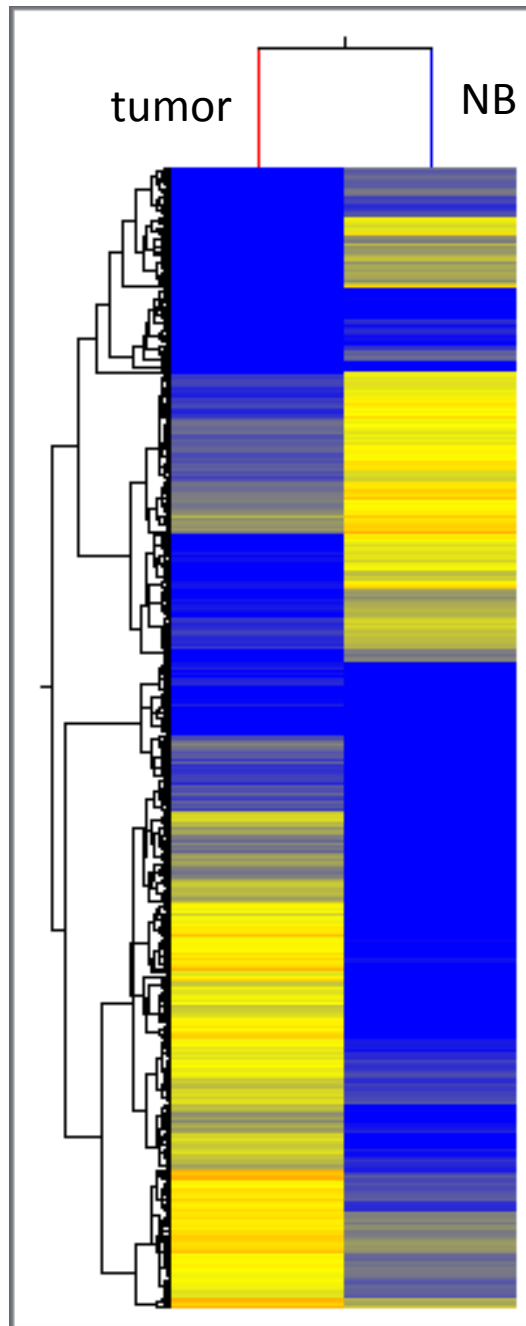
Calculating Delta Beta

A	B	C	D	E	F	G	H	I	J
barcode	probe name	TCGA-E2-A15L-01A-11D-A12E-05	TCGA-E2-A15L-11A-31D-A12E-05		Delta Beta	gene symbol	chromosome	position	
TCGA-E2-A15L-11A-31D-A12E-05	cg03817621	0.575350973	0.856666238		-0.281315264	A1CF	10	52645399	
TCGA-E2-A15L-11A-31D-A12E-05	cg03586879	0.019496835	0.02869236		-0.009195524	A2BP1	16	6068831	
TCGA-E2-A15L-11A-31D-A12E-05	cg19378133	0.02087879	0.028950632		-0.008071842	A2BP1	16	6068835	
TCGA-E2-A15L-11A-31D-A12E-05	cg12058490	0.672173305	0.583852259		0.088321046	A2M	12	9268379	
TCGA-E2-A15L-11A-31D-A12E-05	cg03490200	0.8836703	0.925979392		-0.042309092	A2ML1	12	8975196	
TCGA-E2-A15L-11A-31D-A12E-05	cg27653134	0.540717343	0.737272427		-0.196555084	A2ML1	12	8975309	
TCGA-E2-A15L-11A-31D-A12E-05	cg09744051	0.639105947	0.187013633		0.452092314	A4GALT	22	43115937	
TCGA-E2-A15L-11A-31D-A12E-05	cg07393322	0.051436812	0.098567333		-0.047130521	A4GALT	22	43117318	
TCGA-E2-A15L-11A-31D-A12E-05	cg17687282	0.90041134	0.886003219		0.014408121	A4GNT	3	137851111	
TCGA-E2-A15L-11A-31D-A12E-05	cg18931888	NA	NA		#VALUE!	A4GNT	3	137851874	
TCGA-E2-A15L-11A-31D-A12E-05	cg23032316	0.071153716	0.065851264		0.005302452	AAAS	12	53715148	
TCGA-E2-A15L-11A-31D-A12E-05	cg00559473	0.104738793	0.09135475		0.013384043	AAAS	12	53715592	
TCGA-E2-A15L-11A-31D-A12E-05	cg05303448	NA	NA		#VALUE!	AACS	12	125549535	
TCGA-E2-A15L-11A-31D-A12E-05	cg03013422	0.055966058	0.046417893		0.009548165	AACS	12	125550350	
TCGA-E2-A15L-11A-31D-A12E-05	cg05564657	0.933856334	0.920946575		0.012909759	AADAC	3	151531375	
TCGA-E2-A15L-11A-31D-A12E-05	cg25159668	0.750388159	0.736759151		0.013629007	AADACL2	3	151450877	
TCGA-E2-A15L-11A-31D-A12E-05	cg12506971	0.212389395	0.031680508		0.180708887	AADAT	4	171011264	
TCGA-E2-A15L-11A-31D-A12E-05	cg01654582	0.30447307	0.0696277		0.23484537	AADAT	4	171011577	
TCGA-E2-A15L-11A-31D-A12E-05	cg22940988	0.878932473	0.846253186		0.032679287	AAGAB	15	67538446	
TCGA-E2-A15L-11A-31D-A12E-05	cg01272601	0.023453979	0.02317117		0.00028281	AAK1	2	69871685	
TCGA-E2-A15L-11A-31D-A12E-05	cg21613754	0.742997272	0.885247952		-0.14225068	AAMP;PNKD	2	219133847	
TCGA-E2-A15L-11A-31D-A12E-05	cg14864603	NA	NA		#VALUE!	AAANAT	17	74462846	
TCGA-E2-A15L-11A-31D-A12E-05	cg09382492	0.603188247	0.735198229		-0.132009982	AAANAT	17	74463681	
TCGA-E2-A15L-11A-31D-A12E-05	cg25432323	0.062172579	0.052612175		0.009560404	AARS	16	70322827	
TCGA-E2-A15L-11A-31D-A12E-05	cg09102714	0.049683192	0.040494976		0.009188216	AARS	16	70323758	
TCGA-E2-A15L-11A-31D-A12E-05	cg16356516	0.903233731	0.927964813		-0.024731082	AARS;EXOSC	16	70286390	
TCGA-E2-A15L-11A-31D-A12E-05	cg04429140	NA	NA		#VALUE!	AARSD1;RUN	17	41132427	
TCGA-E2-A15L-11A-31D-A12E-05	cg02426410	0.01515914	0.017289161		-0.002130021	AASDH	4	57253603	
TCGA-E2-A15L-11A-31D-A12E-05	cg03289369	0.022498336	0.021194693		0.001303644	AASDH	4	57253815	
TCGA-E2-A15L-11A-31D-A12E-05	cg26289824	0.015773546	0.015030109		0.000743436	AASDHPTT;K	11	105948485	
TCGA-E2-A15L-11A-31D-A12E-05	cg06667406	0.918425985	0.922686373		-0.004260388	AASS	7	121773889	
TCGA-E2-A15L-11A-31D-A12E-05	cg11857445	0.023918472	0.020008869		0.003909603	AATF	17	35305905	
TCGA-E2-A15L-11A-31D-A12E-05	cg11857445	0.023918472	0.020008869		0.003909603	AATF	17	35305905	

Genic Hypomethylation occurs during progression to MM







Practical

- Navigate TCGA home page and access portal
- Query Data
- Download Data

bsalhia@tgen.org