Services Performed

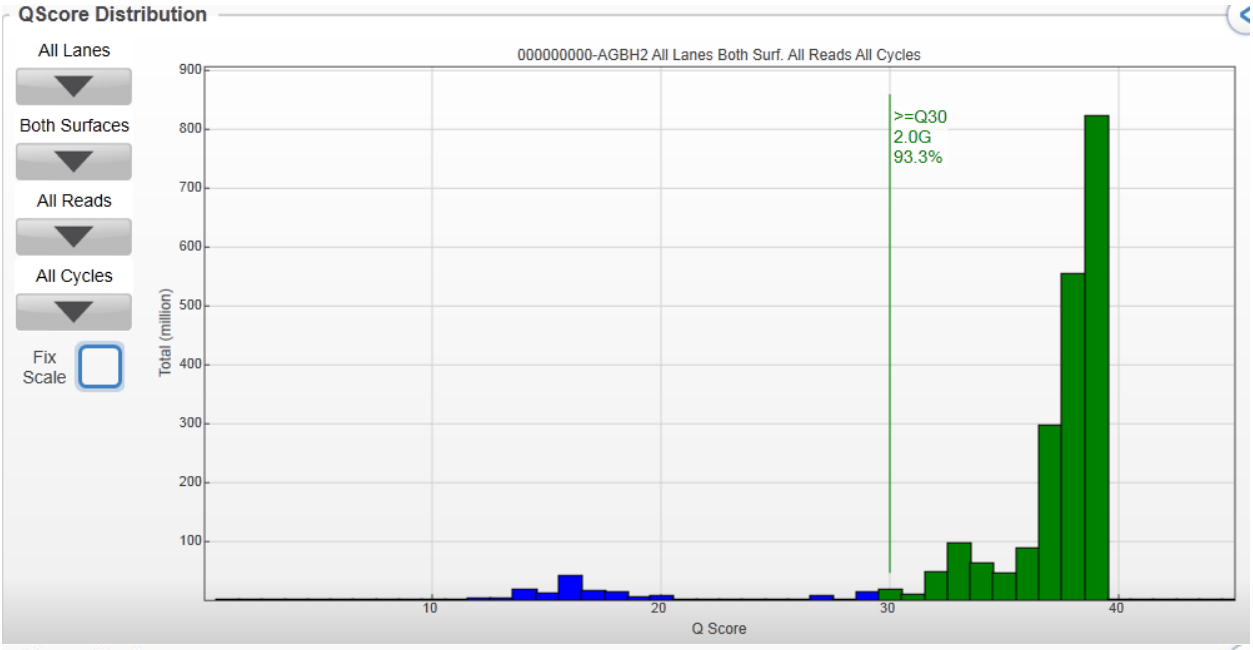
The following checklist confirms the steps of the Targeted Bisulfite Sequencing Service that were performed on your samples.

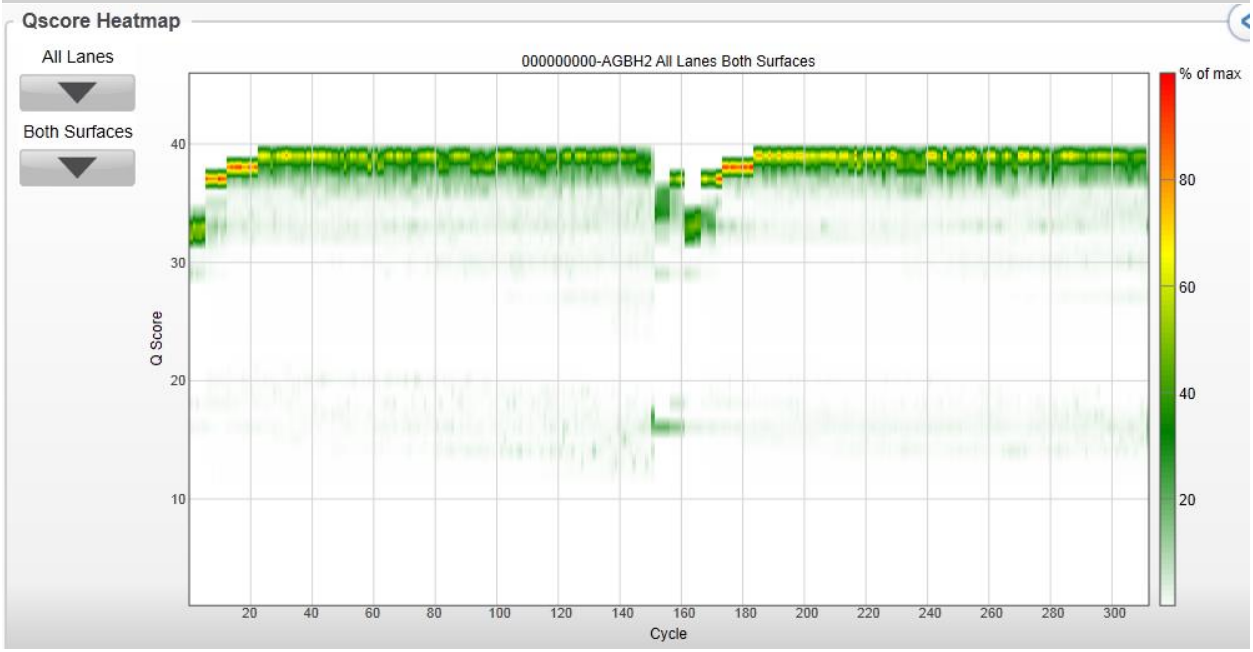
SERVICE	
Sample Received	✓
Sample Quality Evaluated	✓
Primers Designed and Validated	✓
Targeted Amplification	✓
Next-Gen Sequencing	✓
Sequence Quality Check	✓
Bioinformatics Processing	✓
Data/Results	✓

Data/Results Report

Sequencing Run Summary

Following sample handling, targeted amplification, and NGS library indexing, sequence data was generated on an Illumina MiSeq with the following run metrics:





Bioinformatics Analysis: Alignment and Methylation Calling
Sequence data was demultiplexed and assessed for sample read coverage, mapping efficiency, unique CpG coverage, and bisulfite conversion rate. For example,

	Total Read Number	Mapping Efficiency	Unique CpGs	Avg. CpG Coverage	Bisulfite Conversion Rate
Sample 1	75,753 read pairs	78%	96	3,920X	>99%
Sample 2	75,685 read pairs	76%	96	3,773X	>99%
Sample 3	73,894 read pairs	76%	111	3,225X	>99%
Sample 4	66,183 read pairs	78%	96	3,484X	>99%

Complete details for individual samples can be found by logging into your project's data page (http://epidata.zymoresearch.com) using your username and password:

Username: customer@emaildomain.com
Password: Password

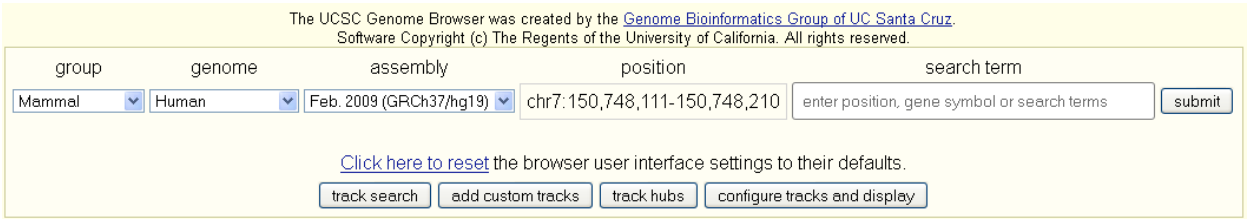
Excel Table
In the email, you will find an excel table entitled "all_count_10_onesheet". The file includes the read data and CpG methylation calling summarizing the CpG browser track files for your project. The methylation ratio (meth_ratio) is calculated by using methylated_CpG_count/total_CpG_count. Furthermore, in the excel table you will find the coordinates for each of the detected CpG sites across all of your project's samples. If one CpG site was detected in at least one sample with at least 10 reads, we listed that CpG site in the table. For samples where the CpG was not detected (<10 reads), we left the meth_ratio and total_CpG_count column blank. Please note that the CpG position counting in the amplicon starts with position 0 (e.g. 0, 1, 2, 3, 4, 5…).

The coordinates, methylated_CpG_count, and total_CpG_count for CpG sites in each sample are also listed in the browser track files, which can be downloaded from your project's data page. By loading the browser tracks into the UCSC genome browser, you will be able to visualize the data.

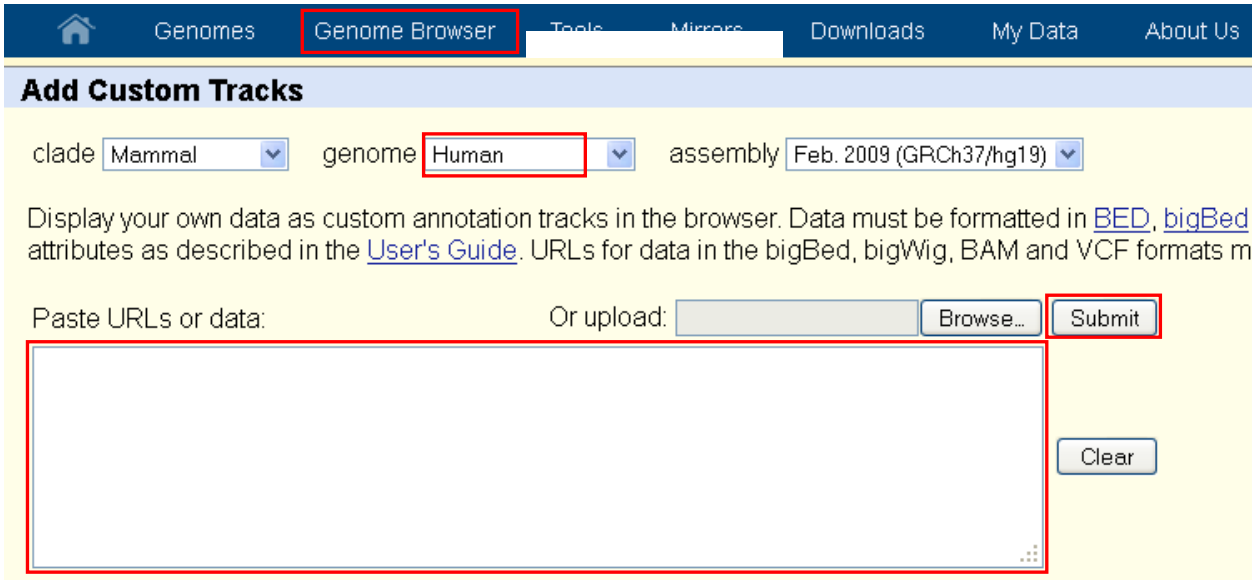
Instructions for Visualization of Genome Browser Tracks

- Go to UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway)
- Select the **genome** and **assembly** specific to your samples and then click **add custom tracks**

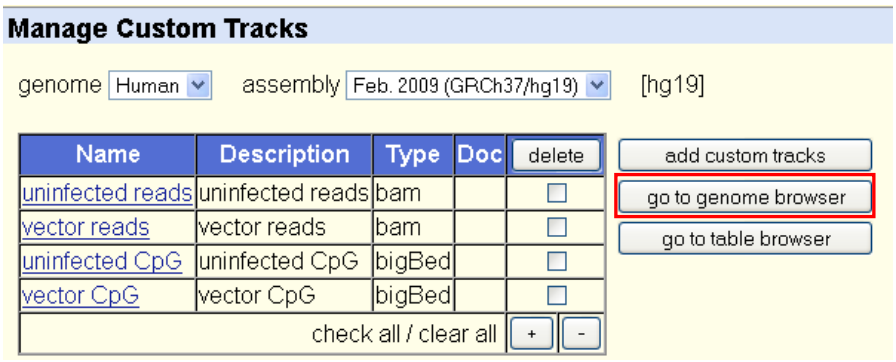
Note: Current **assembly** used for Human is "**Feb. 2009 (GRCh37/hg19)**"



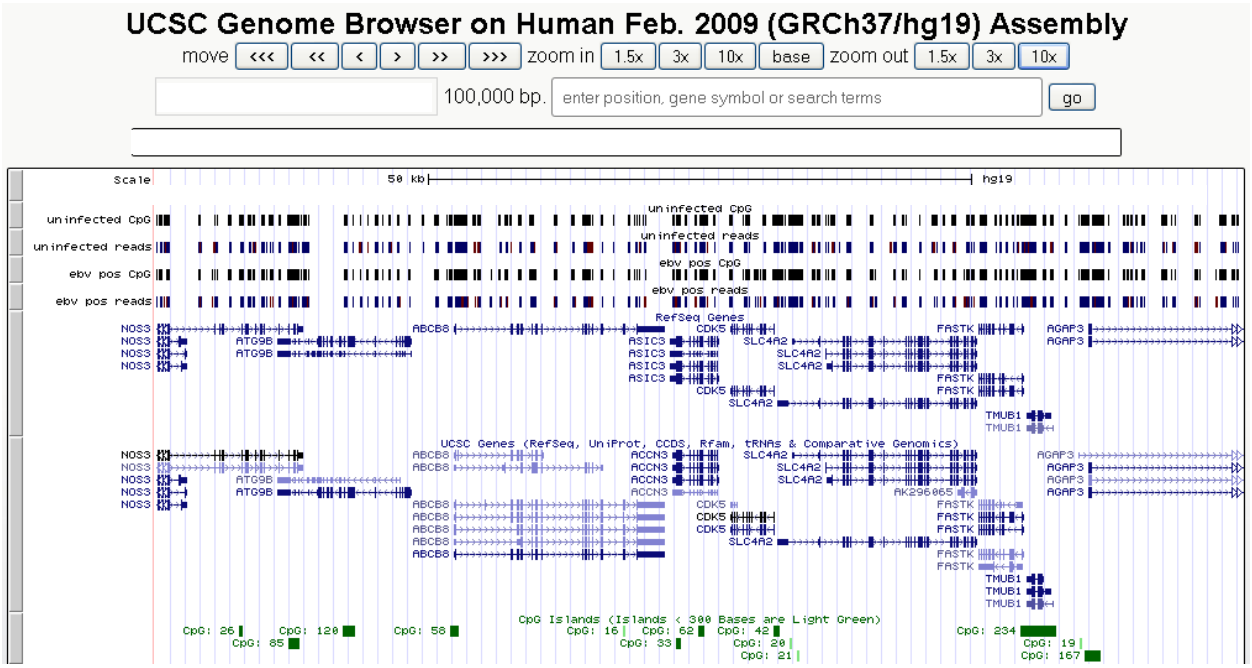
- Paste the links provided (one at a time) and then **Submit** to upload your browser track (this may take some time)



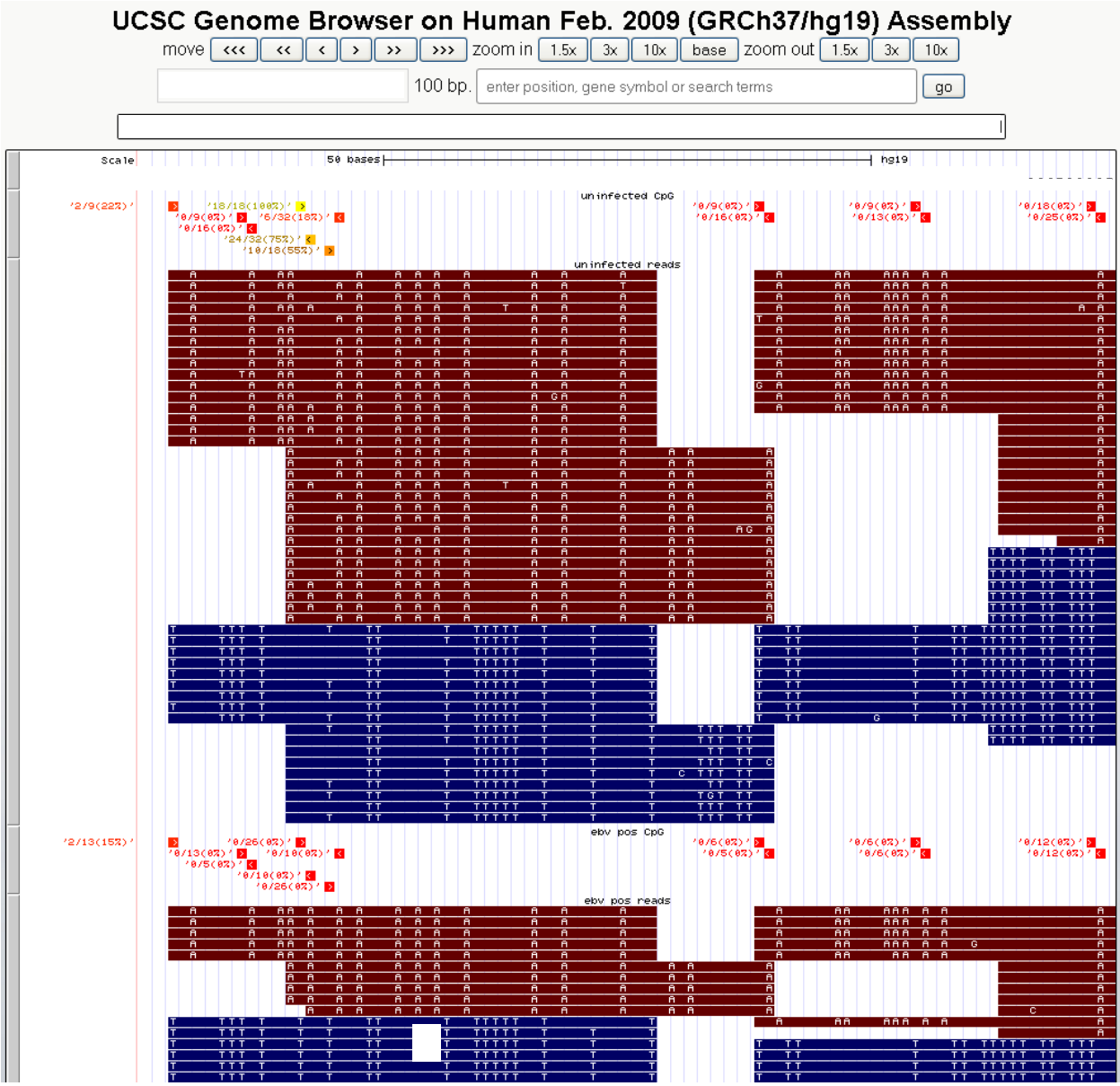
- Click **go to genome browser** once all desired tracks are loaded and navigate the genome to look for areas of interest



- One example for genomic region overview



- Zoomed in on a Differentially Methylated Region



The **CpG Methylation** track:

1. Shows the CpG DNA methylation ratio, percent methylation, and sequencing read coverage depth at each site
2. Yellow corresponds to higher methylation levels, while red corresponds to lower methylation levels. Intermediate methylation levels are represented with orange.
3. The label on the left of a CpG, say ‘4/31’, indicates 4 reads were methylated in 31–fold sequencing read depth of coverage

The **Read** track:

1. Shows the aligned reads with mismatches (conversion events included as mismatches).
2. Blue reads align to the '+' stand and red reads align to the '– 'strand.

Note: Once the browser tracks are loaded, can be visualized in the UCSC Genome Browser simply by copying and pasting the chromosome (**chrom**), starting position (**chromstart**), and ending position (**chromend**) directly from Microsoft Excel into the "enter position or search terms" box near the top and clicking **go** (or pressing Enter on your keyboard). The genome browser may have trouble loading or not be able to load at all if you have many custom tracks expanded when you are trying to navigate to different areas of the genome. If problems occur, try altering the display density of your tracks to "dense" or "squish" and then try again. Additionally, please note that the coordinates in the Excel file will be offset by one nucleotide relative to the coordinates in the UCSC Genome Browser.

Terminology:

Feature – annotated genomic features such as genes, gene promoters, exons, introns, and CpG Islands

Meth ratio – short for methylation ratio – the measured number of methylated cytosines divided by total number of cytosines covered at that site

Chrom – short for chromosome

Chromstart – chromosomal coordinates of the beginning of the site

Chromend – chromosomal coordinates of the end of the site

Strand – indicates read alignment to either the '+' or '-' strand

Total CpG count – read coverage for the sample at that site

Meth diff – short for methylation difference; the difference in methylation ratios between two samples at the specified site determined by subtracting the value for the second sample (to the right) from the value for the first sample (to the left)

pvalue – quantification of the statistical significance of the methylation difference by either the Student's t-test or the Fisher's Exact Test, depending on the analysis type

Methylation Classification Definitions (for interpretation of the CpG result tables):
HYPERMETHYLATED: 0-33% more methylated than reference (p-value < 0.05)

HYPOMETHYLATED: 0-33% less methylated than reference (p-value < 0.05)

STRONGLY HYPERMETHYLATED: 33-100% more methylated than reference (p-value < 0.05)

STRONGLY HYPOMETHYLATED: 33-100% less methylated than reference (p-value < 0.05)

INSIGNIFICANT: Statistically insignificant difference in methylation (p-value > 0.05)

Downloading Raw Data

We advise that you download all raw data within two months of project completion. You can do this through your project's data website by clicking on the,

- "Raw Data" tab to access the raw sequence reads in FASTQ format
- "Alignment" tab to access the reads aligned to the indicated reference genome in BAM or BAI format
- "Browser tracks" tab to access the methylation track in bigbed (.bb) format

The files are compressed and are too large to open on most computers. Therefore, to open these files, you will need to use a Linux machine and have some bioinformatics background. The FASTQ files will probably need to be uploaded to the GEO data prior to publication of the data – so if you need help, please contact us.

ZYMO RESEARCH CORP.

17062 Murphy Ave. ▪ Irvine, CA 92614

Phone: 1-888-882-9682 ▪ 1-949-679-1190 ▪ Fax: 1-949-266-9452

info@zymoresearch.com ▪ www.zymoresearch.com