

---

# Methyl-MiniSeq Service Report: Genome-Wide Bisulfite Sequencing

## Workflow Checklist

---

Sample Received	✓
Sample Quality Check	✓
Sample Prepared for Sequencing	✓
Next-Gen Sequencing	✓
Sequence Quality Check	✓
Bioinformatics Analysis	✓
Data/Results	✓

## Materials & Methods

---

The samples were processed and analyzed using the Classic RRBS Service: Genome-wide bisulfite sequencing (Zymo Research, Irvine, CA).

**DNA Extraction:** If DNA extraction was performed, one of two different DNA extraction kits was used depending on the sample amount. The manufacturer's instructions were followed, unless otherwise stated. The kit used in this project is marked below.

- Quick-DNA Plus Kit Miniprep Kit (Zymo Research, Irvine, CA)
- Quick-DNA Plus Kit Microprep Kit (Zymo Research, Irvine, CA)
- N/A (DNA Extraction Not Performed)

Additional Notes: NA

**Methyl-MiniSeq® Library preparation:** 200 ng of starting input genomic DNA was sequentially digested with 60 units of TaqI followed by 30 units of MspI (NEB) and then purified with Zymo Research [DNA Clean & Concentrator™-5](#) (Cat#: D4003). Fragments were ligated to pre-annealed adapters containing 5'-methyl-cytosine instead of cytosine according to Illumina's specified guidelines. Adaptor-ligated fragments of 150-250 bp and 250-350 bp in size were recovered from a 2.5% NuSieve 1:1 agarose gel using [Zymoclean™ Gel DNA Recovery Kit](#) (Cat#: D4001). The fragments were then bisulfite-treated using the [EZ DNA Methylation-Lightning™ Kit](#) (Cat#: D5030). Preparative-scale PCR was performed and the resulting products were purified with [DNA Clean & Concentrator™-5](#) (Cat#: D4003) for sequencing on an Illumina HiSeq.

**Sequence alignments and data analysis:** Sequence reads from bisulfite-treated Classic RRBS libraries were identified using standard Illumina base calling software and then raw FASTQ files were adapter, filled-in nucleotides, and quality trimmed using [TrimGalore 0.6.4](#). [FastQC 0.11.8](#) was used to assess the effect of trimming and overall quality distributions of the data. Alignment to the [hg19](#) reference genome was performed using [Bismark 0.19.0](#). Methylated and unmethylated read totals for each CpG site were called using MethylDackel 0.5.0. The methylation level of each sampled cytosine was estimated as the number of reads reporting a C, divided by the total number of reads reporting a C or T. Fisher's exact test was performed for groups without replicates (Tumor\_2 vs NAT\_2, Tumor\_3a vs NAT\_3a, Tumor\_4 vs NAT\_4 ) and T-test was performed for groups having replicates (Tumor vs NAT). These tests were performed on each CpG site that has at least five reads coverage. In order to identify significantly differential methylated cytosines, methylation value between groups has to be minimum 10% with a p-value (unadjusted) of < 0.05. Promoter, gene body, and CpG island annotations were added for each CpG included in the comparison where available.

## Sample Information

---

Sample ID	Original Sample Label	Total Read Number (M)	Mapping Efficiency (%)
in013_1	Tumor_4	17.763210	46.98%
in013_2	NAT_4	22.375362	41.51%
in013_3	Tumor_2	19.062438	44.74%
in013_4	NAT_2	18.124028	39.87%
in013_5	Tumor_3a	21.587385	53.46%
in013_6	NAT_3a	22.246479	47.33%

## Tables and Graphical Data Information

### Terminology

<b>Chr/pos/Site</b>	Chromosomal coordinates for the site
<b>Sample Name</b>	DNA methylation ratio that's the measured number of methylated cytosine divided by total number of cytosines covered at that site
<b>methdiff</b>	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)
<b>pvalue</b>	Quantification of the statistical significance of the methylation difference by either the Student's t-test, the Fisher's Exact test, or ANOVA depending on the analysis type
<b>Padj_fdr</b>	FDR-adjusted p-value of the test
<b>Padj_slim</b>	Slim-adjusted p-value of the test
<b>Promoter/exon/intron</b>	Gene annotation of a particular region
<b>Cgi</b>	A 'Y' indicates the region falls within a CpG island
<b>Meth_cov</b>	The total number of methylated cytosine covered at that site
<b>Total_cov</b>	The total number of cytosines covered at that site

### Graphical Data Definition

<b>Heatmaps with Hierarchical Clustering</b>	Samples are grouped based on similarity for the top 3000 or all differentially methylated CpG sites covered in the assay. Red represents high levels of DNA methylation, and blue represents low levels of DNA methylation.
<b>Coverage Plots</b>	The charts show the sequencing coverage of CpG Sites in the samples
<b>Violin Plots (Labelled "MethByAnno_...")</b>	Display the overall methylation level of individual sample for each annotated genomic feature (CpG islands aka "cgi", gene bodies, promoters). A small circle inside the violin represents the median methylation ratio, the black bars extending from the circle represent the interquartile range (25 <sup>th</sup> and 75 <sup>th</sup> percentiles), and the whiskers extend out to 1.5x of the interquartile range. The width of the "violin" represents a smoothed probability density of CpG sites with a specific methylation ratio.

## Project Data

---

### Comparison Results

#### *Top 3000 Differential CpG Sites (XLSX format)*

##### ***Tumor vs NAT***

[https://epiquest.s3.amazonaws.com/epiquest\\_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/diff\\_xlsx/Tumor\\_vs\\_NAT\\_top3000.xlsx](https://epiquest.s3.amazonaws.com/epiquest_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/diff_xlsx/Tumor_vs_NAT_top3000.xlsx)

##### ***Tumor\_2 vs NAT\_2***

[https://epiquest.s3.amazonaws.com/epiquest\\_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/diff\\_xlsx/Tumor\\_2\\_vs\\_NAT\\_2\\_top3000.xlsx](https://epiquest.s3.amazonaws.com/epiquest_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/diff_xlsx/Tumor_2_vs_NAT_2_top3000.xlsx)

##### ***Tumor\_3a vs NAT\_3a***

[https://epiquest.s3.amazonaws.com/epiquest\\_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/diff\\_xlsx/Tumor\\_3a\\_vs\\_NAT\\_3a\\_top3000.xlsx](https://epiquest.s3.amazonaws.com/epiquest_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/diff_xlsx/Tumor_3a_vs_NAT_3a_top3000.xlsx)

##### ***Tumor\_4 vs NAT\_4***

[https://epiquest.s3.amazonaws.com/epiquest\\_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/diff\\_xlsx/Tumor\\_4\\_vs\\_NAT\\_4\\_top3000.xlsx](https://epiquest.s3.amazonaws.com/epiquest_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/diff_xlsx/Tumor_4_vs_NAT_4_top3000.xlsx)

#### *All Differential CpG Sites (CSV format)*

##### ***Tumor vs NAT***

[https://epiquest.s3.amazonaws.com/epiquest\\_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/diff\\_csv/Tumor\\_vs\\_NAT.csv](https://epiquest.s3.amazonaws.com/epiquest_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/diff_csv/Tumor_vs_NAT.csv)

##### ***Tumor\_2 vs NAT\_2***

[https://epiquest.s3.amazonaws.com/epiquest\\_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/diff\\_csv/Tumor\\_2\\_vs\\_NAT\\_2.csv](https://epiquest.s3.amazonaws.com/epiquest_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/diff_csv/Tumor_2_vs_NAT_2.csv)

##### ***Tumor\_3a vs NAT\_3a***

[https://epiquest.s3.amazonaws.com/epiquest\\_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/diff\\_csv/Tumor\\_3a\\_vs\\_NAT\\_3a.csv](https://epiquest.s3.amazonaws.com/epiquest_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/diff_csv/Tumor_3a_vs_NAT_3a.csv)

##### ***Tumor\_4 vs NAT\_4***

[https://epiquest.s3.amazonaws.com/epiquest\\_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/diff\\_csv/Tumor\\_4\\_vs\\_NAT\\_4.csv](https://epiquest.s3.amazonaws.com/epiquest_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/diff_csv/Tumor_4_vs_NAT_4.csv)

## Project Data (Cont.)

---

### *Heatmaps (show top 3000 differential sites)*

#### ***Tumor vs NAT***

[https://epiquest.s3.amazonaws.com/epiquest\\_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/Tumor\\_vs\\_NAT.png](https://epiquest.s3.amazonaws.com/epiquest_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/Tumor_vs_NAT.png)

#### ***Tumor\_2 vs NAT\_2***

[https://epiquest.s3.amazonaws.com/epiquest\\_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/Tumor\\_2\\_vs\\_NAT\\_2.png](https://epiquest.s3.amazonaws.com/epiquest_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/Tumor_2_vs_NAT_2.png)

#### ***Tumor\_3a vs NAT\_3a***

[https://epiquest.s3.amazonaws.com/epiquest\\_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/Tumor\\_3a\\_vs\\_NAT\\_3a.png](https://epiquest.s3.amazonaws.com/epiquest_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/Tumor_3a_vs_NAT_3a.png)

#### ***Tumor\_4 vs NAT\_4***

[https://epiquest.s3.amazonaws.com/epiquest\\_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/Tumor\\_4\\_vs\\_NAT\\_4.png](https://epiquest.s3.amazonaws.com/epiquest_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/Tumor_4_vs_NAT_4.png)

## Project Data (Cont.)

---

### Global Plots

#### *Read Coverage Distributions*

[https://epiquest.s3.amazonaws.com/epiquest\\_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/Coverage.pdf](https://epiquest.s3.amazonaws.com/epiquest_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/Coverage.pdf)

#### *Global Methylation Distribution in Annotation Regions (Violin Plots)*

##### **CpG Island**

[https://epiquest.s3.amazonaws.com/epiquest\\_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/MethByAnno\\_cgi.pdf](https://epiquest.s3.amazonaws.com/epiquest_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/MethByAnno_cgi.pdf)

##### **Gene Body**

[https://epiquest.s3.amazonaws.com/epiquest\\_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/MethByAnno\\_genes.pdf](https://epiquest.s3.amazonaws.com/epiquest_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/MethByAnno_genes.pdf)

##### **Promoter**

[https://epiquest.s3.amazonaws.com/epiquest\\_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/MethByAnno\\_promoter.pdf](https://epiquest.s3.amazonaws.com/epiquest_in013/NM7wkUqufpLLDK45GKHJJK/CG/output/MethByAnno_promoter.pdf)

### Raw Data

#### *Trimmed FASTQ Files*

List of Downloadable URLs:

[https://epiquest.s3.amazonaws.com/epiquest\\_in013/NM7wkUqufpLLDK45GKHJJK/download\\_fastq.txt](https://epiquest.s3.amazonaws.com/epiquest_in013/NM7wkUqufpLLDK45GKHJJK/download_fastq.txt)

#### *BAM Alignments*

[https://epiquest.s3.amazonaws.com/epiquest\\_in013/NM7wkUqufpLLDK45GKHJJK/download\\_bam.txt](https://epiquest.s3.amazonaws.com/epiquest_in013/NM7wkUqufpLLDK45GKHJJK/download_bam.txt)

#### *Methylation Calling Tables*

[https://epiquest.s3.amazonaws.com/epiquest\\_in013/NM7wkUqufpLLDK45GKHJJK/CG/download\\_cytosine\\_report.txt](https://epiquest.s3.amazonaws.com/epiquest_in013/NM7wkUqufpLLDK45GKHJJK/CG/download_cytosine_report.txt)

## Genome Browser Tracks

---

Please see on the next pages detailed instructions for loading custom tracks in UCSC Genome Browser.

### ***BigWig Group Averages***

Track files were produced in BigWig format by taking the average of samples within each group which had a coverage  $\geq 5$  at each individual cytosine. A custom tracks file contains the URLs for each track file, which can be cut and pasted into the UCSC Genome Browser for viewing.

Download Custom Tracks File:

[https://epiquest.s3.amazonaws.com/epiquest\\_in013/NM7wkUqufpLLDK45GKHJJK/CG/tracks/tracks\\_avg.txt](https://epiquest.s3.amazonaws.com/epiquest_in013/NM7wkUqufpLLDK45GKHJJK/CG/tracks/tracks_avg.txt)

### **Genome Tracks for Individual Samples with Methylation Ratio and Coverage**

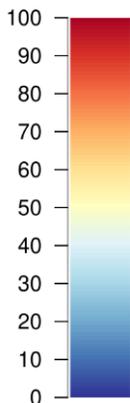
Track files were generated in BigBed format. The tracks show each covered CpG site with coverage of  $\geq 5$  reads in the genome for each sample. Each CpG site is labeled with a fraction, which represents the number of methylated cytosines over total number of cytosines covered by sequencing reads at that site. A greater number of total reads at a cytosine increases confidence in the resulting methylation ratio. CpG sites are colored using a color scale (see below) matching the methylation ratio (the division of methylated reads over total reads for each CpG). The tracks can be viewed on UCSC Genome Browser using the same procedure as for the group average tracks above.

Download Custom Tracks File:

[https://epiquest.s3.amazonaws.com/epiquest\\_in013/NM7wkUqufpLLDK45GKHJJK/CG/tracks/tracks\\_meth.txt](https://epiquest.s3.amazonaws.com/epiquest_in013/NM7wkUqufpLLDK45GKHJJK/CG/tracks/tracks_meth.txt)

Color Scale representing CpG methylation level in genome tracks for individual samples:

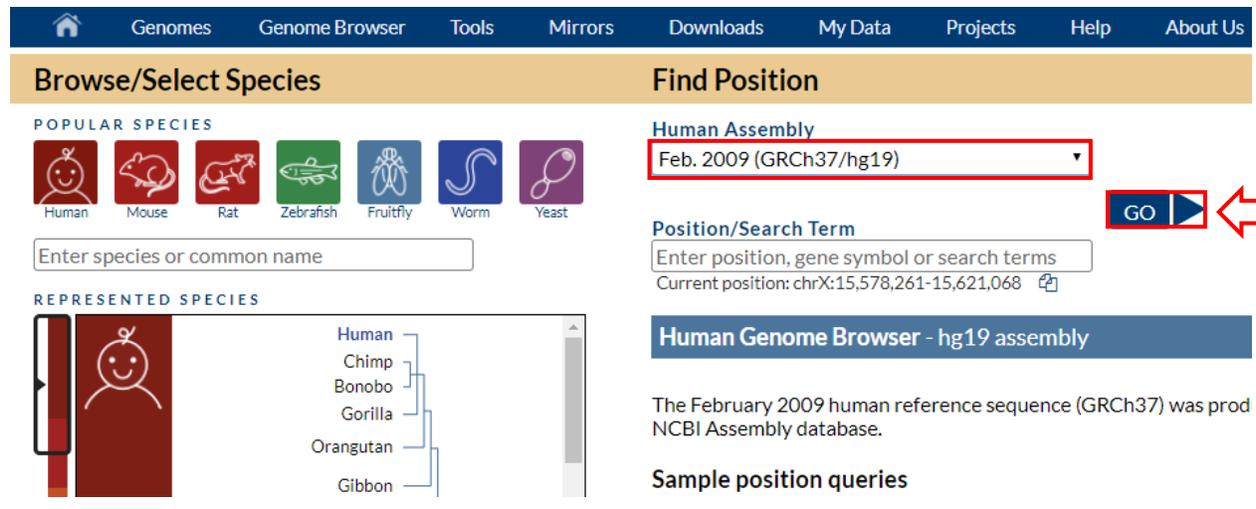
**% Methylation**



# Genome Browser Tracks Visualization

**To Visualize genome browser tracks on UCSC, please follow steps below:**

1. Go to [UCSC Genome Browser](#), 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)"

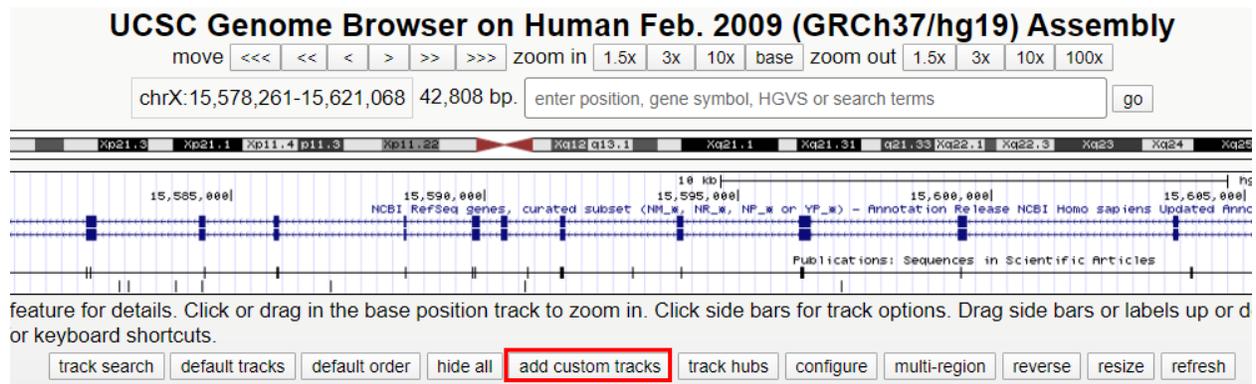


The screenshot shows the UCSC Genome Browser interface. At the top is a navigation bar with links: Home, Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, Projects, Help, and About Us. Below this is a header with two main sections: "Browse/Select Species" and "Find Position".

**Browse/Select Species:** This section includes "POPULAR SPECIES" with icons for Human, Mouse, Rat, Zebrafish, Fruitfly, Worm, and Yeast. There is a text input field "Enter species or common name". Below that is "REPRESENTED SPECIES" with a tree diagram showing Human, Chimp, Bonobo, Gorilla, Orangutan, and Gibbon.

**Find Position:** This section has a "Human Assembly" dropdown menu currently set to "Feb. 2009 (GRCh37/hg19)". Below it is a "Position/Search Term" input field with the text "Enter position, gene symbol or search terms" and "Current position: chrX:15,578,261-15,621,068". A "GO" button is highlighted with a red box and a red arrow points to it from the right.

Below the search fields is a blue bar labeled "Human Genome Browser - hg19 assembly". Underneath, it says "The February 2009 human reference sequence (GRCh37) was produced by the NCBI Assembly database." and "Sample position queries".



The screenshot shows the UCSC Genome Browser tracks visualization for Human Feb. 2009 (GRCh37/hg19) Assembly. The title is "UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly".

Navigation controls include "move" (with left and right arrows), "zoom in" (1.5x, 3x, 10x, base), and "zoom out" (1.5x, 3x, 10x, 100x). The current position is "chrX:15,578,261-15,621,068" with "42,808 bp." and a search input field "enter position, gene symbol, HGVS or search terms" with a "go" button.

The tracks include: "Xp21.3, Xp21.1, Xp11.4, p11.3, Xp11.22, Xq12, q13.1, Xq21.1, Xq21.31, q21.33, Xq22.1, Xq22.3, Xq23, Xq24, Xq25".

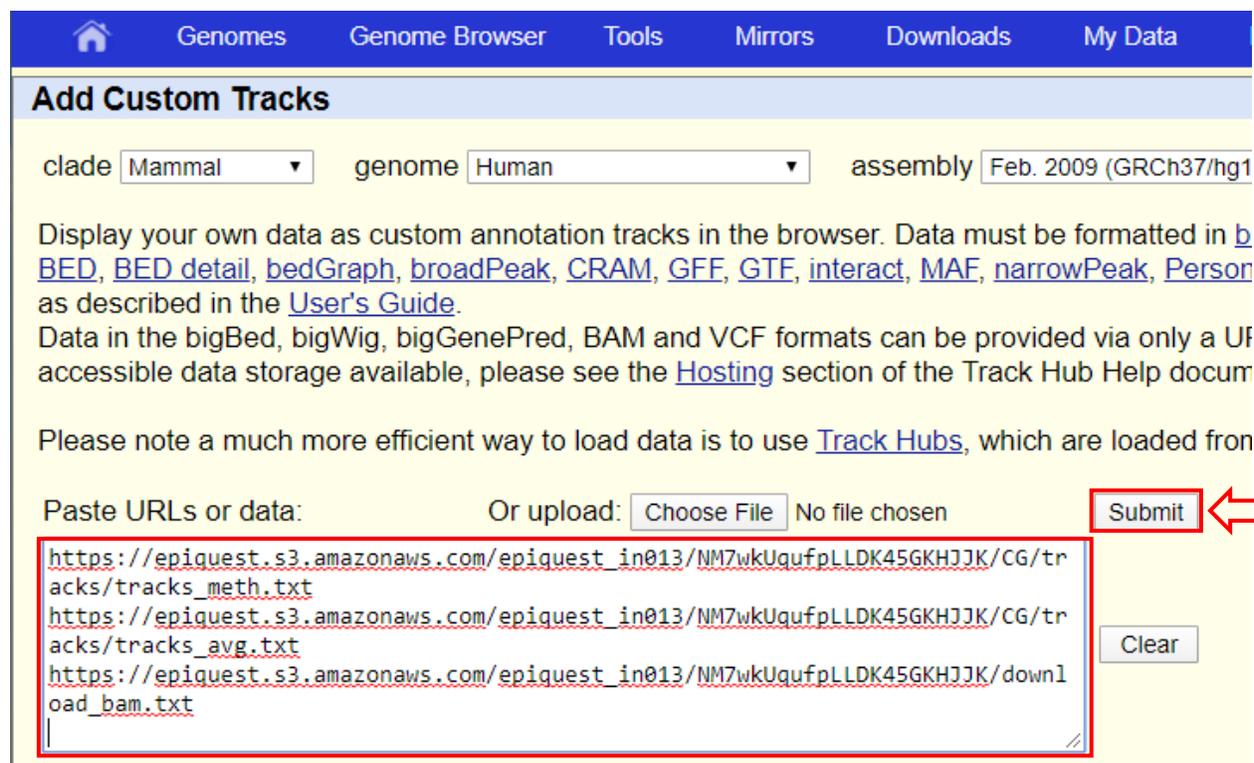
Gene annotations include "15,585,000", "15,590,000", "15,595,000", "15,600,000", and "15,605,000". The track is labeled "NCBI RefSeq genes, curated subset (NH\_w, NR\_w, NP\_w or YP\_w) - Annotation Release NCBI Homo Sapiens Updated Ann".

Other tracks include "Publications: Sequences in Scientific Articles".

At the bottom, there are several buttons: "track search", "default tracks", "default order", "hide all", "add custom tracks" (highlighted with a red box and a red arrow pointing up), "track hubs", "configure", "multi-region", "reverse", "resize", and "refresh".

## Genome Browser Tracks Visualization (Cont.)

2. Paste the links provided and then **Submit** to upload your browser track (this may take some time)



clade  genome  assembly

Display your own data as custom annotation tracks in the browser. Data must be formatted in [BED](#), [BED detail](#), [bedGraph](#), [broadPeak](#), [CRAM](#), [GFF](#), [GTF](#), [interact](#), [MAF](#), [narrowPeak](#), [Person](#) as described in the [User's Guide](#).

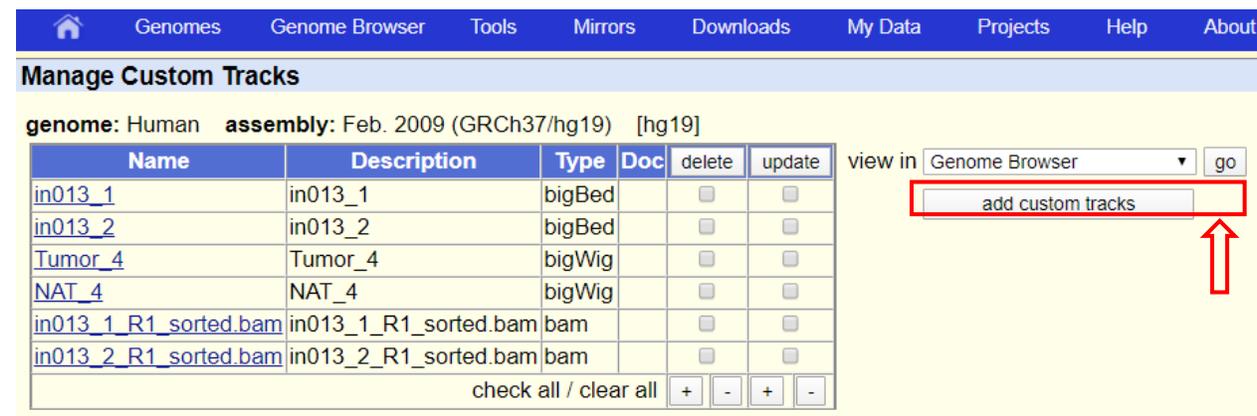
Data in the bigBed, bigWig, bigGenePred, BAM and VCF formats can be provided via only a UF accessible data storage available, please see the [Hosting](#) section of the Track Hub Help document.

Please note a much more efficient way to load data is to use [Track Hubs](#), which are loaded from

Paste URLs or data: Or upload:  No file chosen

```
https://epiquest.s3.amazonaws.com/epiquest_in013/NM7wkUqufpLLDK45GKHJJK/CG/tracks/tracks_meth.txt
https://epiquest.s3.amazonaws.com/epiquest_in013/NM7wkUqufpLLDK45GKHJJK/CG/tracks/tracks_avg.txt
https://epiquest.s3.amazonaws.com/epiquest_in013/NM7wkUqufpLLDK45GKHJJK/download_bam.txt
```

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



genome: Human assembly: Feb. 2009 (GRCh37/hg19) [hg19]

Name	Description	Type	Doc	delete	update
<a href="#">in013_1</a>	in013_1	bigBed		<input type="checkbox"/>	<input type="checkbox"/>
<a href="#">in013_2</a>	in013_2	bigBed		<input type="checkbox"/>	<input type="checkbox"/>
<a href="#">Tumor_4</a>	Tumor_4	bigWig		<input type="checkbox"/>	<input type="checkbox"/>
<a href="#">NAT_4</a>	NAT_4	bigWig		<input type="checkbox"/>	<input type="checkbox"/>
<a href="#">in013_1_R1_sorted.bam</a>	in013_1_R1_sorted.bam	bam		<input type="checkbox"/>	<input type="checkbox"/>
<a href="#">in013_2_R1_sorted.bam</a>	in013_2_R1_sorted.bam	bam		<input type="checkbox"/>	<input type="checkbox"/>

view in

check all / clear all

## Genome Browser Tracks Visualization (Cont.)

### Example of methylation tracks and sequence read tracks in the genome browser:

The **CpG Methylation** track:

1. The fraction denotes methylation ratio which is the number of methylated cytosines over total number of cytosines covered by sequencing reads at that site.
2. Red corresponds to higher methylation levels, while blue corresponds to lower methylation levels.
3. Methylation ratio is displayed next to each CpG site when display density is set to “pack” (as in the example above) or all the way to the left when display density is set to “full”.

The **Read** track:

1. Shows the aligned reads with mismatches (unmethylated cytosines converted to thymines are shown as mismatches).
2. Blue reads align to the “+” strand and red reads align to the “-” strand.

The genome browser may have trouble loading or not be able to load when there are many custom tracks expanded and you are trying to navigate to different areas of the genome. If problems occur, try altering the display density of the tracks to “dense” or “squish” and then try again.

