Joshua Lin
11/26/24

QBIO 490: Directed Research - Multi-Omic Analysis - Fall 2024 Review Project

Part 1: Review Questions

General Concepts

1. TCGA stands for The Cancer Genome Atlas, a program, which is federally funded by the National Institutes of Health in the National Cancer Institute and the National Human Genome Research Institute, that maintains a large public multi-omic database of cancer patient data. TCGA is important because it is a large database (larger sample sizes) and publicly accessible for anyone to perform research and even multi-omic data analysis on cancer patient data.
2. Some strengths of TCGA are that it has multi-omic data (clinical, mutation (genomics), RNA counts (transcriptomics), and methylation (epigenomics)) and potentially better accuracy in analysis through larger numbers of patient samples. Some weaknesses of TCGA are that it is a federally-funded public database, meaning that there could be less money and maintenance dedicated for TCGA and potentially more data pre-processing needed before analysis.

Coding Skills

1. To save a file to your GitHub repository, run the following:
   a. cd [repopath] # change to the directory of the GitHub repository at repopath
   b. git add [filepath] # add desired file located at filepath
   c. git commit -m [message] # committing the added file to repo
   d. git push # push committed file to repo
   e. git status # to ensure the process of pushing the file to the repository is going as intended
2. To run a package in R, run the following:
   a. if(!require(“[packagename]“)) install.packages(“[packagename]“) # installs package *packagename* if the packagename is not installed yet
   b. library([packagename]) # load in the R package
3. To use a *Bioconductor* package in R, run the following:
   a. if(!require(“BiocManager”)) install.packages(“BiocManager”)
   b. if(!require(“[packagename]”)) BiocManager::install(“[packagename]”)
   c. library([packagename])
4. Boolean indexing is the process of using a boolean mask/vector of the same length of the R dataframe columns/rows to pre-process (T = kept value, F = ignored value) an R dataframe. Some applications of boolean indexing are to subsetting data in a dataframe

by a selected variable, process data to remove unwanted values such as NAs, and obtaining specific data from the dataframe, etc.

5. A mockup sample dataframe named “clinical_data” could be:

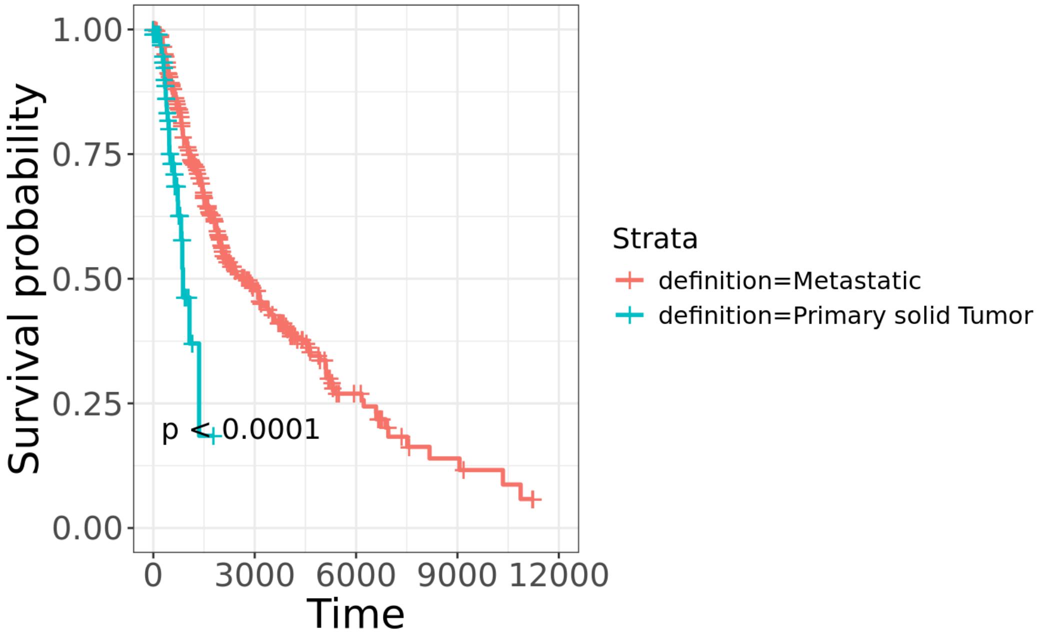
patient_id	age_at_diagnosis	sex	(*not part of the original dataframe)* Elderly
1	67	M	Not Elderly
2	84	F	Elderly

   a. ifelse(clinical_data$age_at_diagnosis < 70, ‘Not Elderly’, ‘Elderly’). This ifelse statement checks to see if the column ‘age_at_diagnosis’ in the dataframe ‘clinical_data’ is less than 70: if it is, then it will return ‘Not Elderly’, otherwise it will return ‘Elderly’.
   b. Elderly_mask <- ifelse(clinical_data$age_at_diagnosis < 70, ‘Not Elderly’, ‘Elderly’)
   clinical_data$Elderly <- clinical_data[Elderly_mask,]
   This boolean indexing creates a boolean/vector mask ‘Elderly_mask’ equal to the ifelse statement in part a, then a new column ‘Elderly’ in the clinical_data dataframe will be created and assigned values according to the boolean mask for each respective row/value.

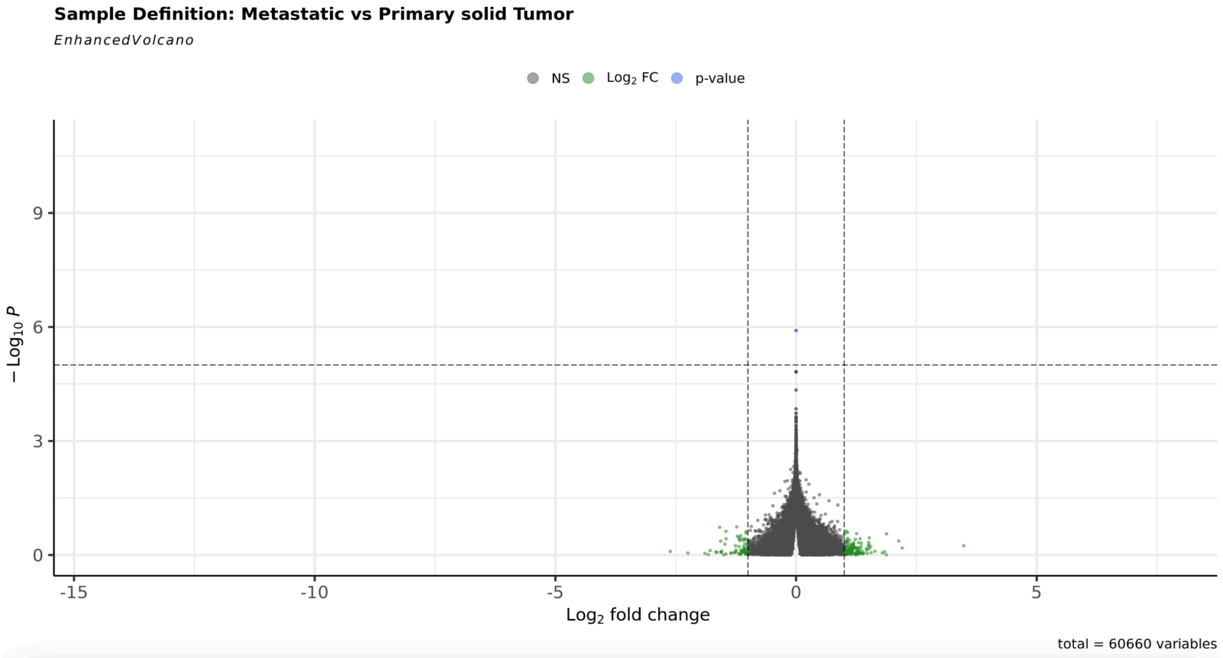
Part 2: SKCM Analysis

Exploration of Methylation Patterns and Effect on Transcription

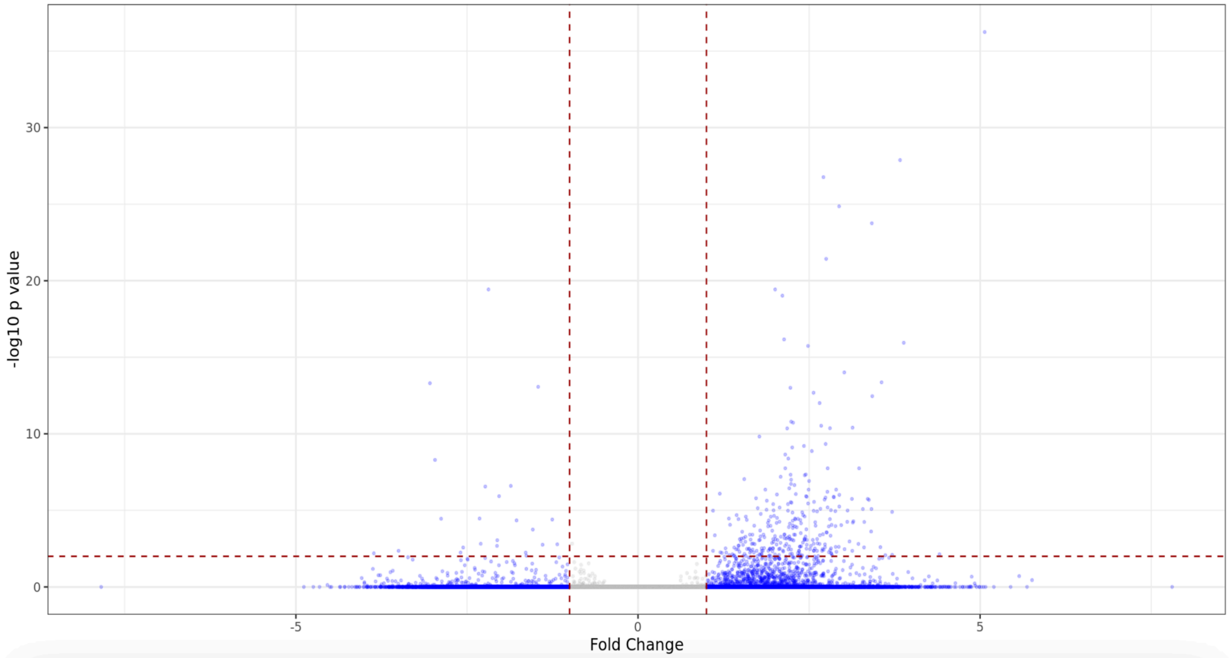
1. The following KM plot shows potential differences in survival between metastatic and non-metastatic patients:



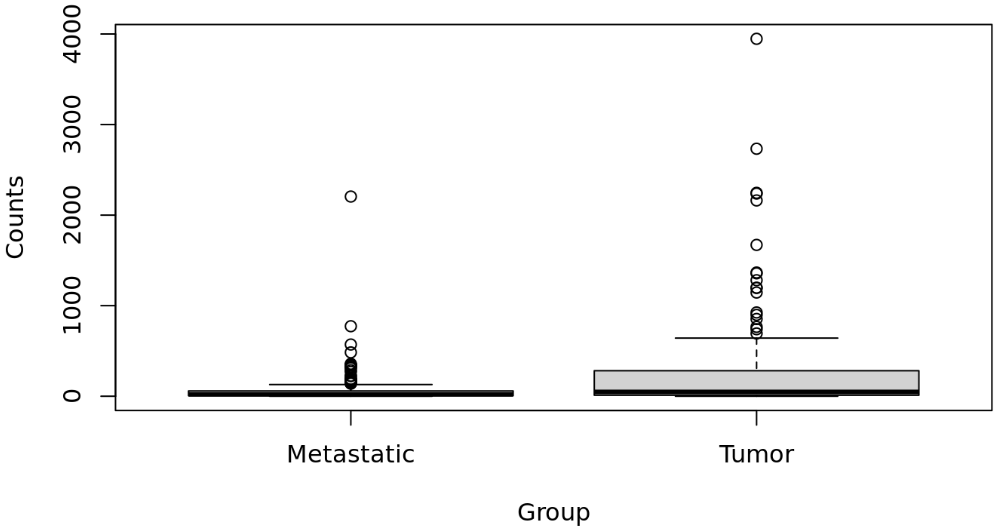
2. The following EnhancedVolcano plot shows differential expression between non-metastatic and metastatic patients, controlling for treatment effects, race, gender, and vital status (as covariates) using DESeq2:

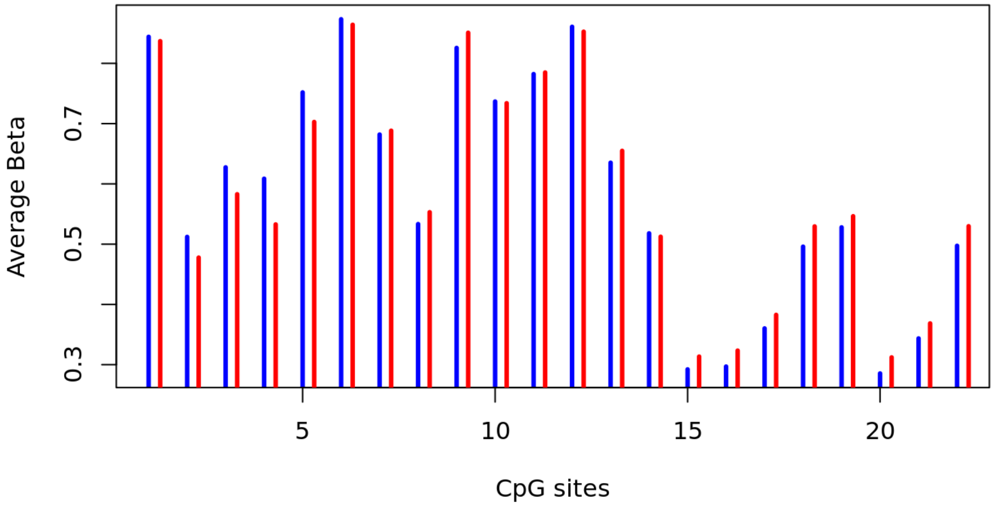


3. The following Volcano plot shows naive differential methylation between non-metastatic and metastatic patients:

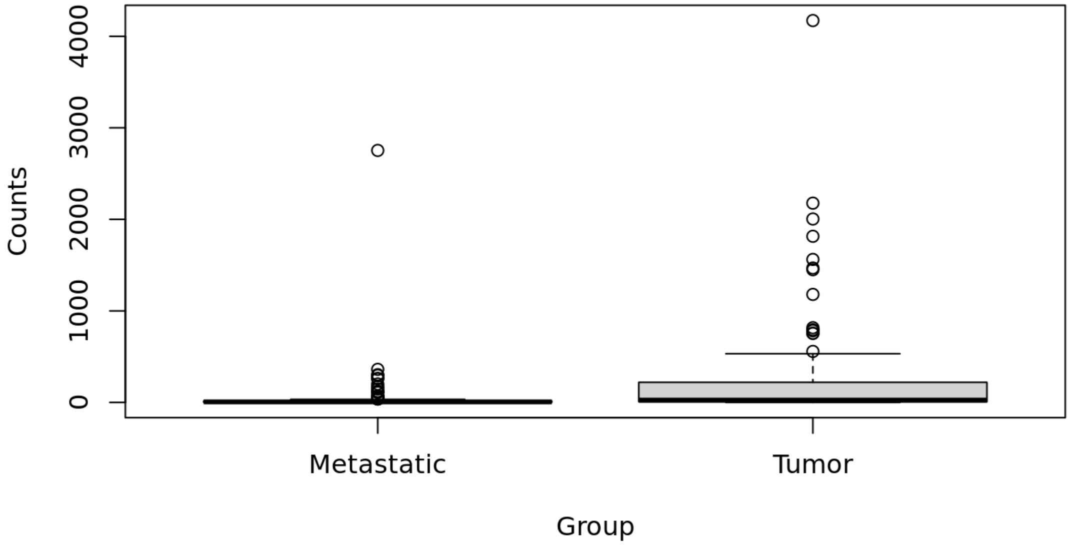


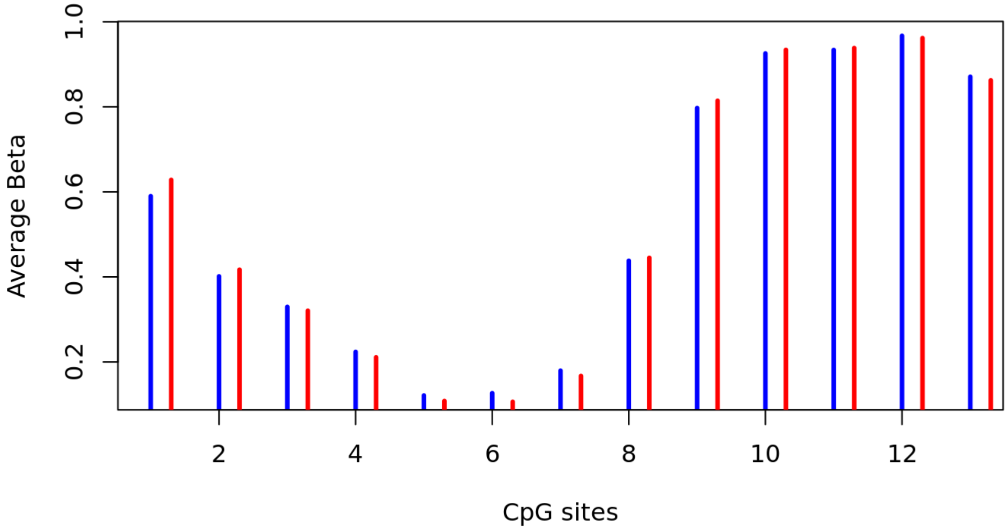
4. The following plots show a direct comparison of methylation status to transcriptional activity across non-metastatic and metastatic patients for 10 genes:
    a. TTC22




Average Beta
0.3
0.5
0.7
5
10
15
20
CpG sites

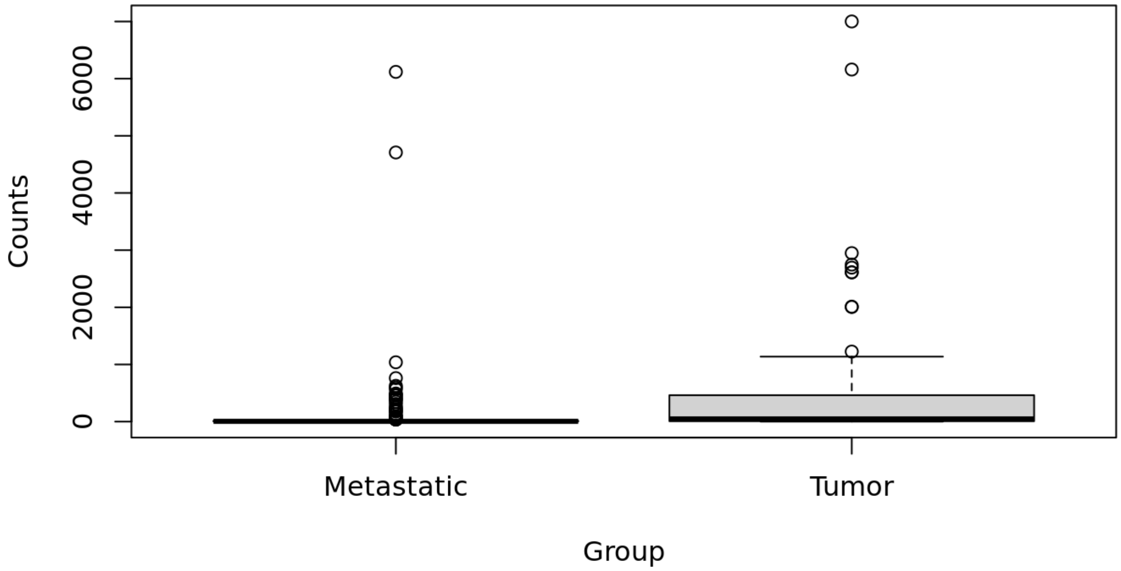
b. EPN3


Counts
0
1000
2000
3000
4000
Metastatic
Tumor
Group

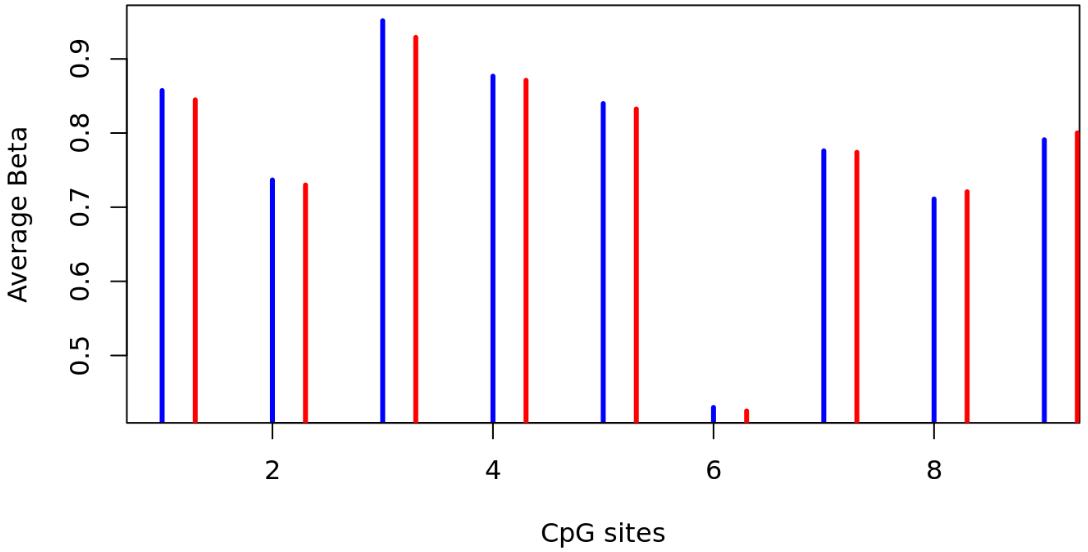


Average Beta
0.2
0.4
0.6
0.8
1.0
2
4
6
8
10
12
CpG sites

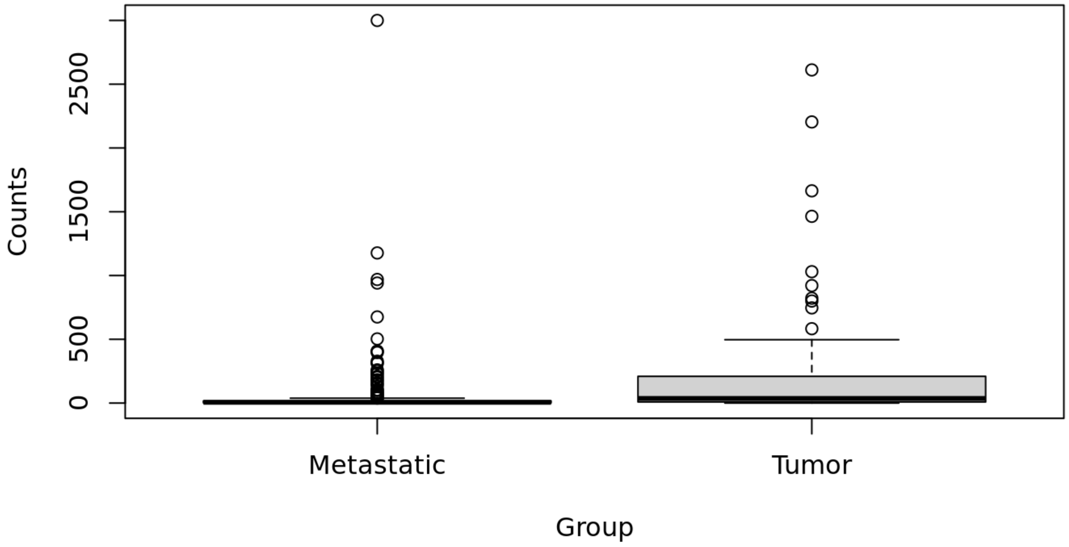
c. PRSS8

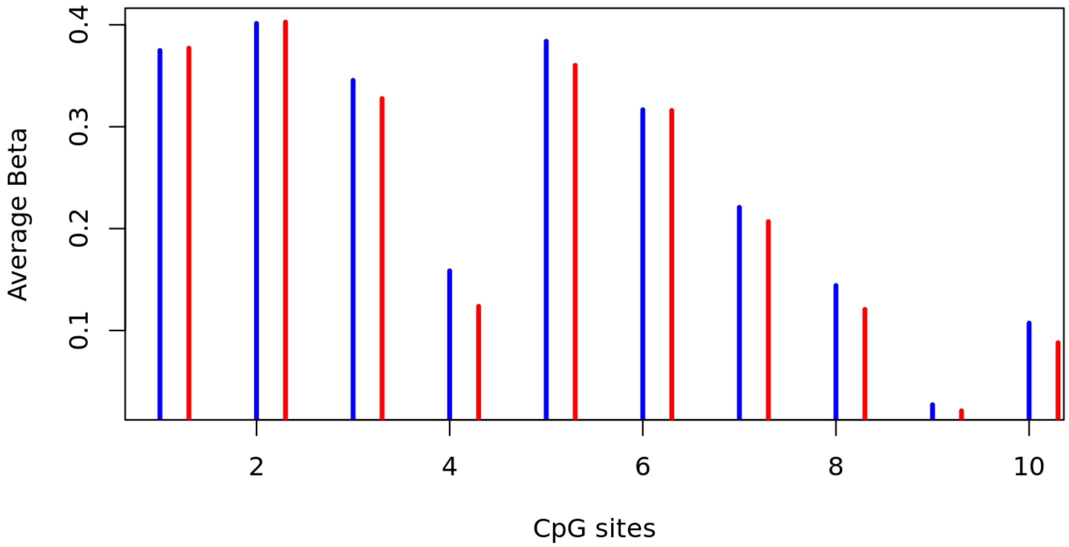


Counts
0
2000
4000
6000
Metastatic
Tumor
Group

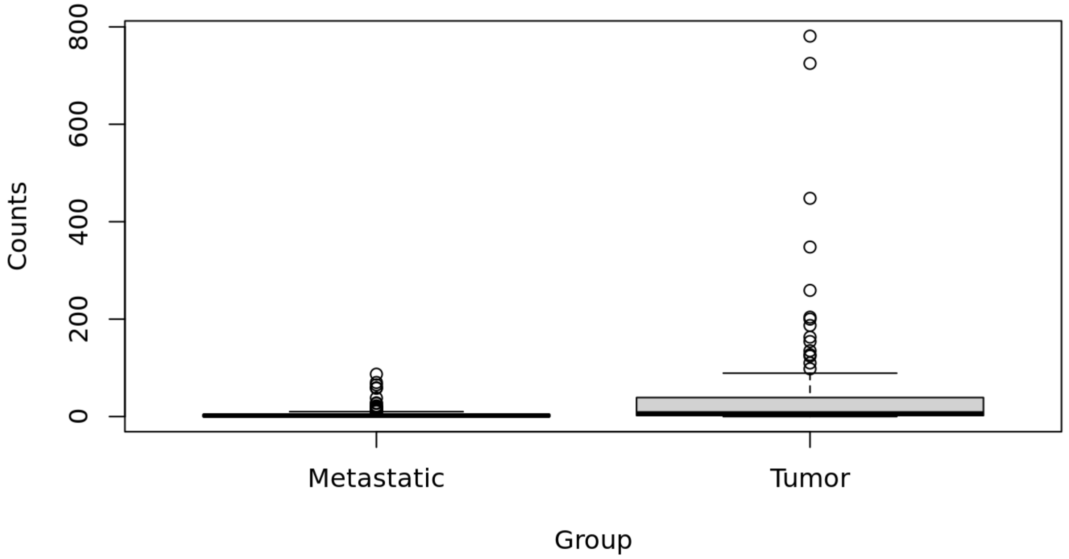

Average Beta
0.5
0.6
0.7
0.8
0.9
2
4
6
8
CpG sites

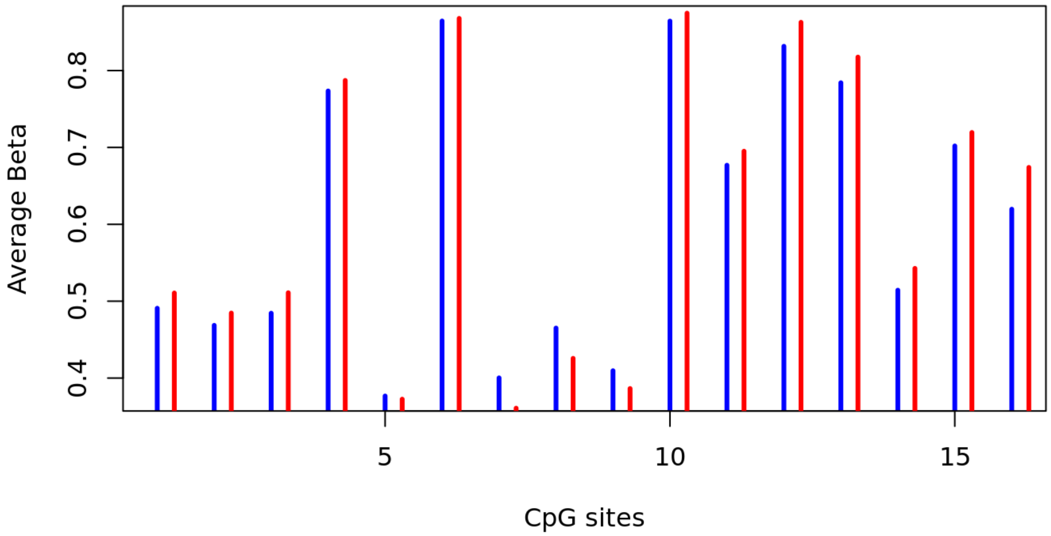
d. DLX3


Counts
0
500
1500
2500
Metastatic
Tumor
Group

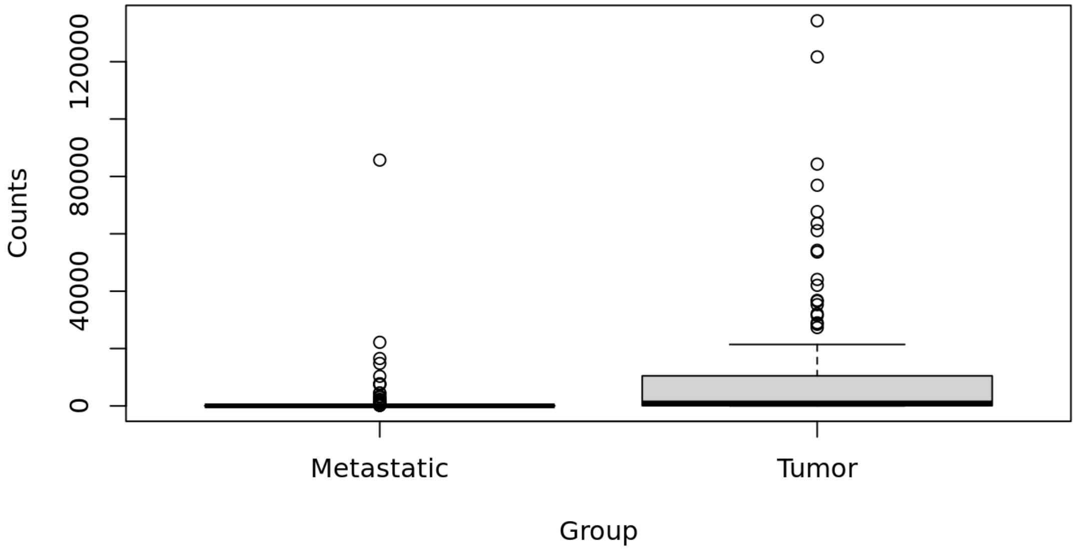

Average Beta
0.1
0.2
0.3
0.4
2
4
6
8
10
CpG sites

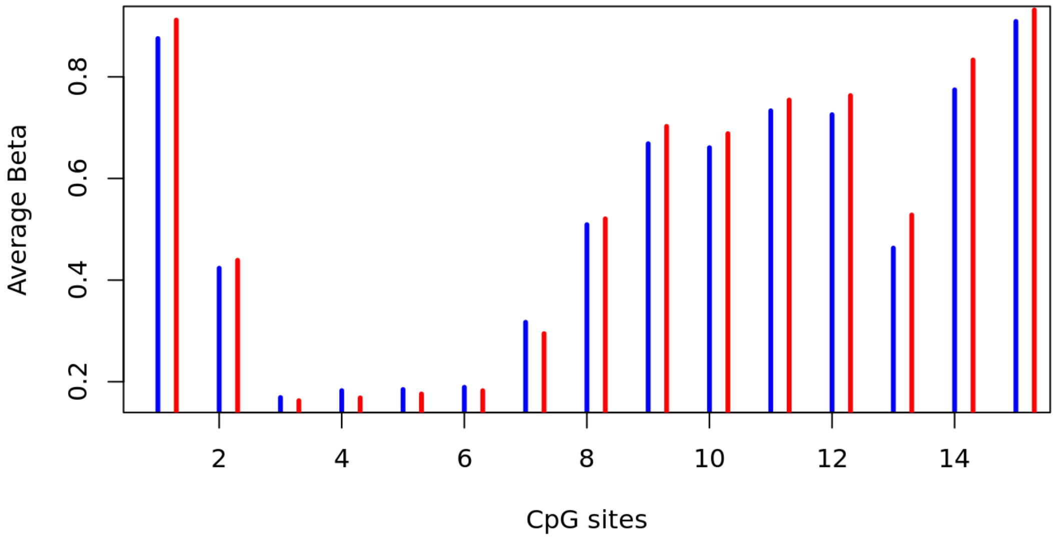
e. CNGB1


Counts
0
200
400
600
800
Metastatic
Tumor
Group

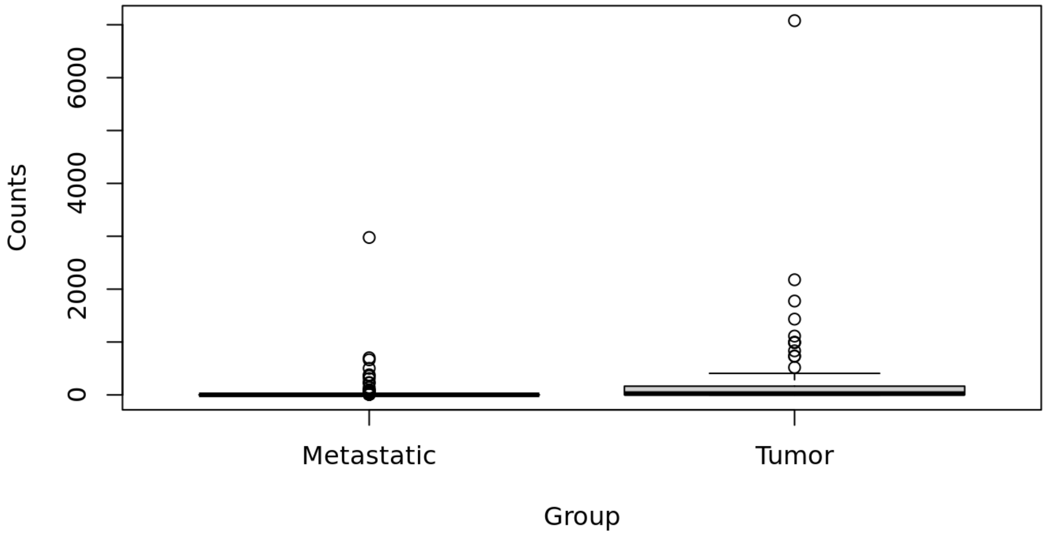

Average Beta
0.4
0.5
0.6
0.7
0.8
5
10
15
CpG sites

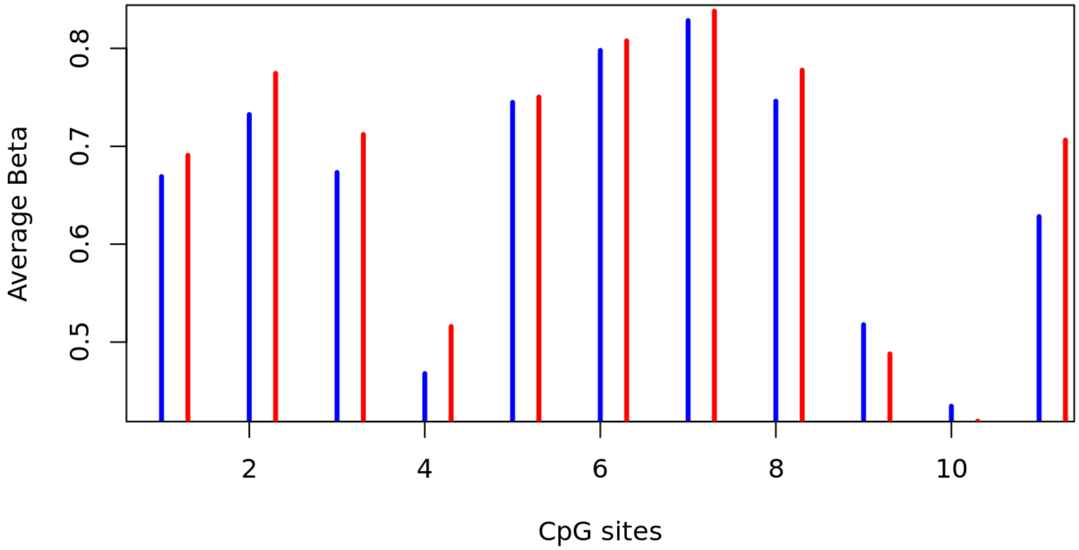
f. PKP1


Counts
0
40000
80000
120000
Metastatic
Tumor
Group

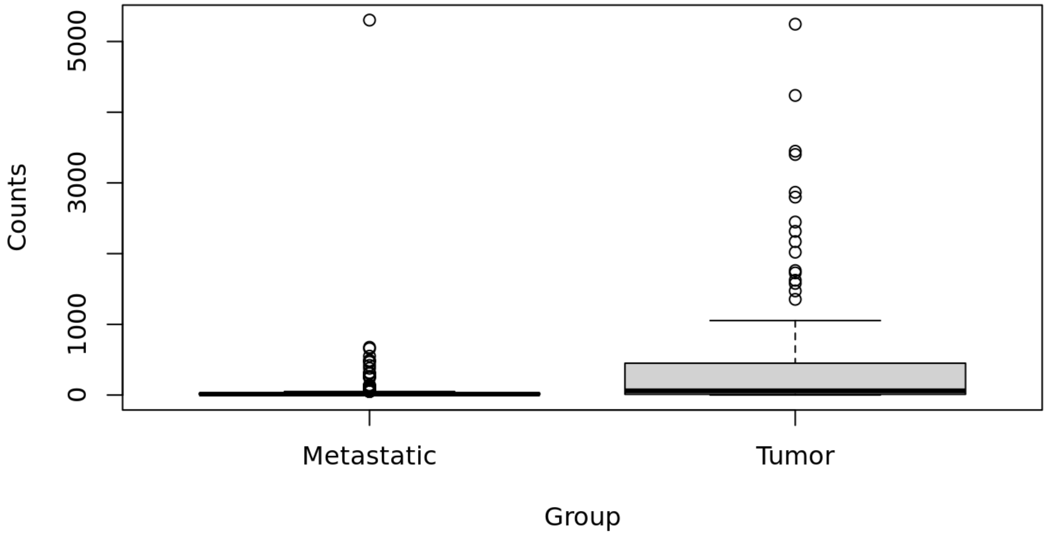

Average Beta
0.2
0.4
0.6
0.8
2
4
6
8
10
12
14
CpG sites

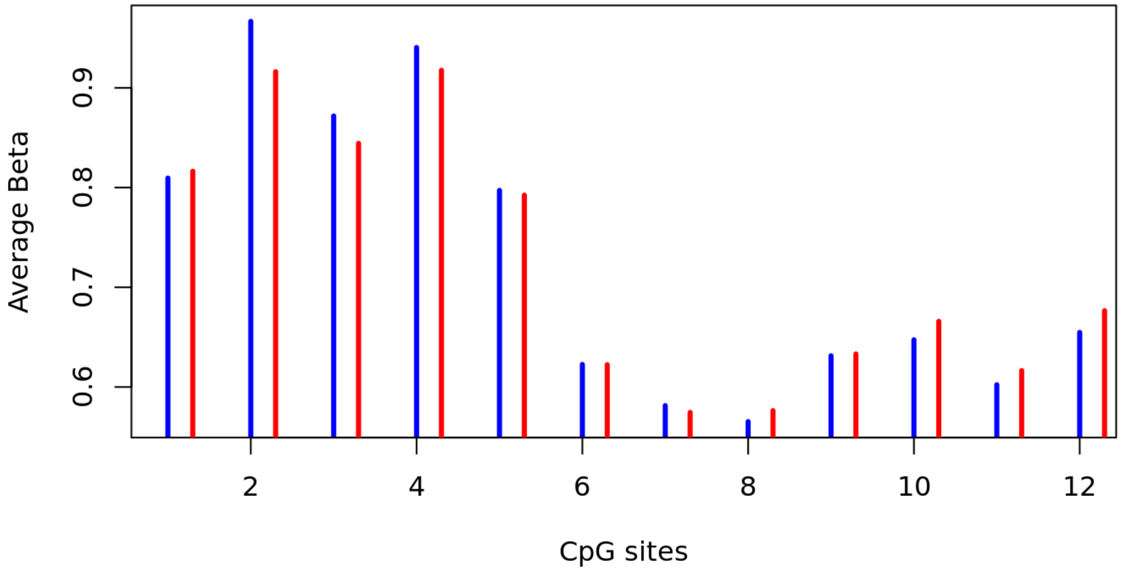
g. GRHL2


Counts
0
2000
4000
6000
Metastatic
Tumor
Group


Average Beta
0.5
0.6
0.7
0.8
2
4
6
8
10
CpG sites

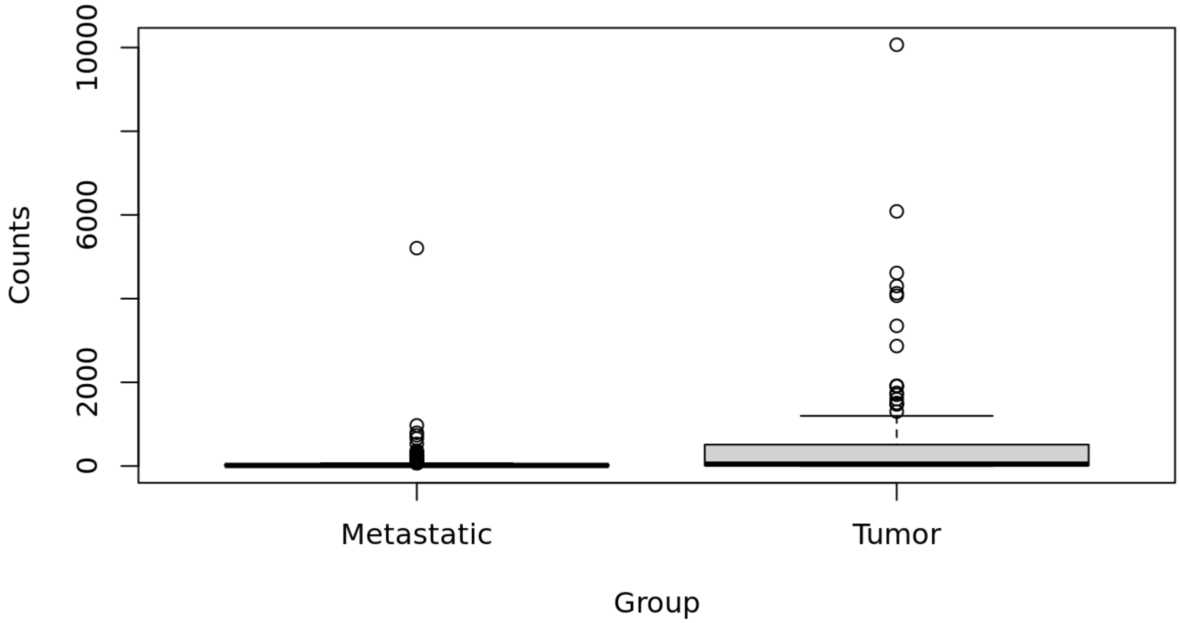
h. ESRP2

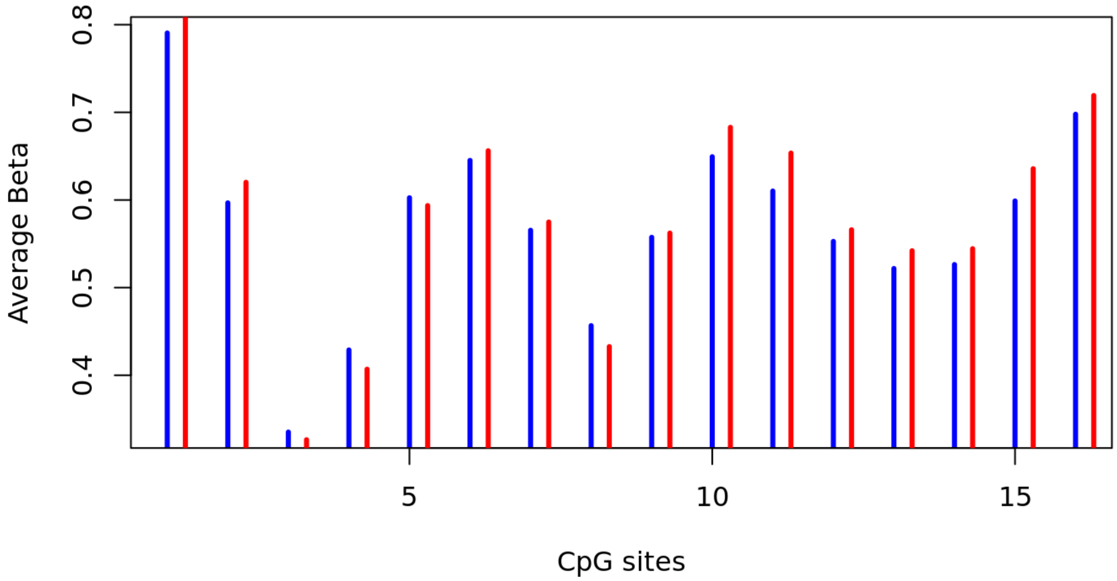

Counts
0
1000
3000
5000
Metastatic
Tumor
Group


Average Beta
0.6
0.7
0.8
0.9
2
4
6
8
10
12
CpG sites

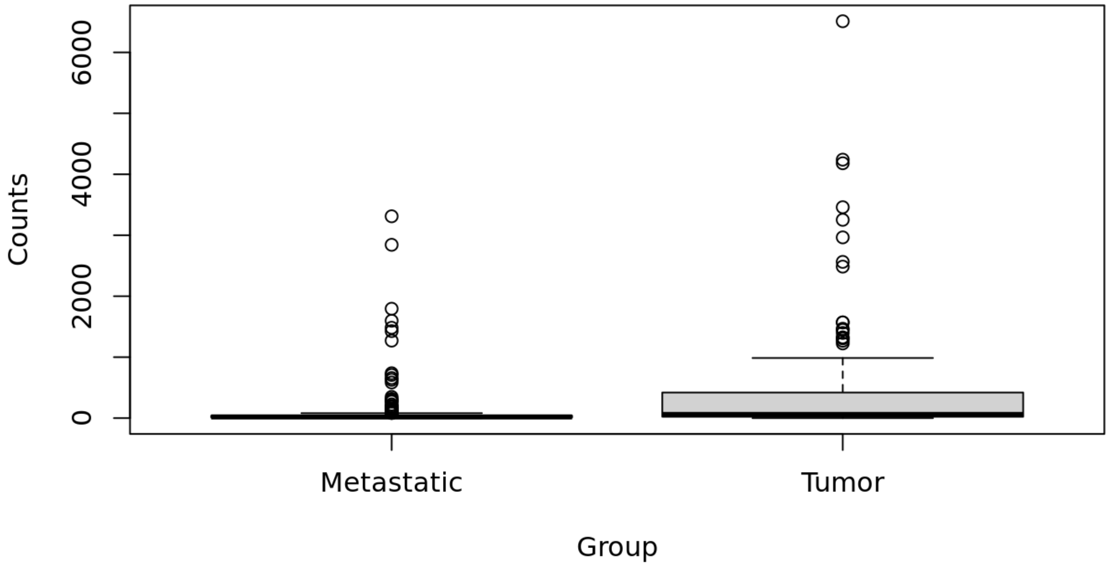
i. EPHX3

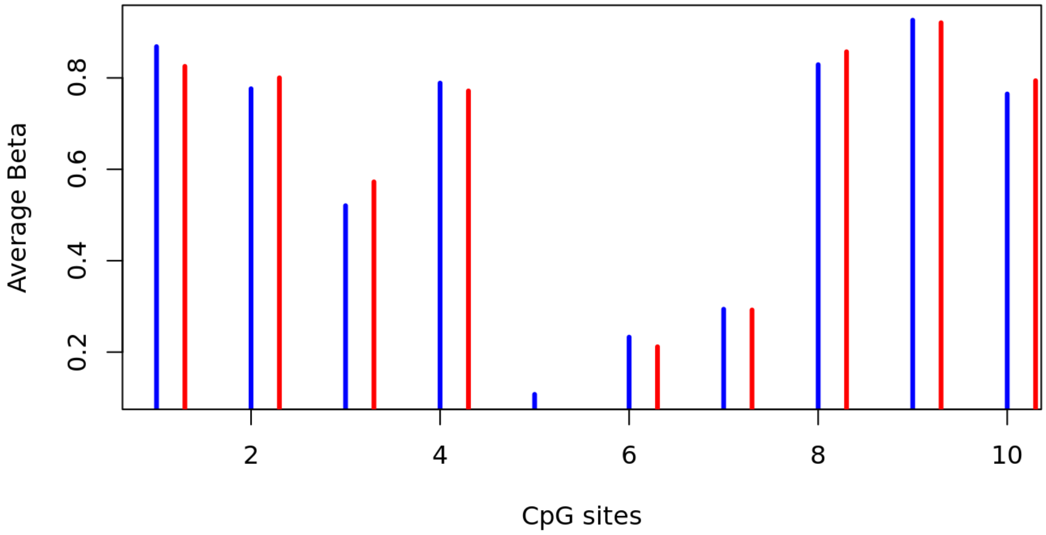
h


Counts
0
2000
6000
10000
Metastatic
Tumor
Group

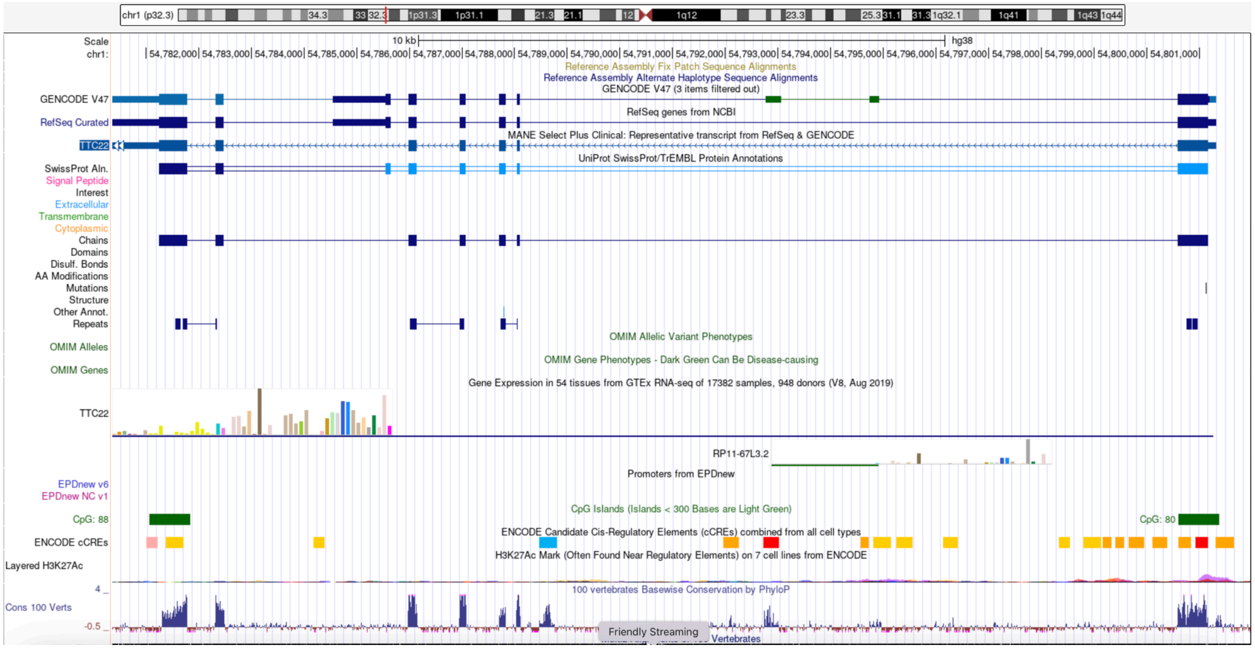

Average Beta
0.4
0.5
0.6
0.7
0.8
5
10
15
CpG sites

j. SCNN1A

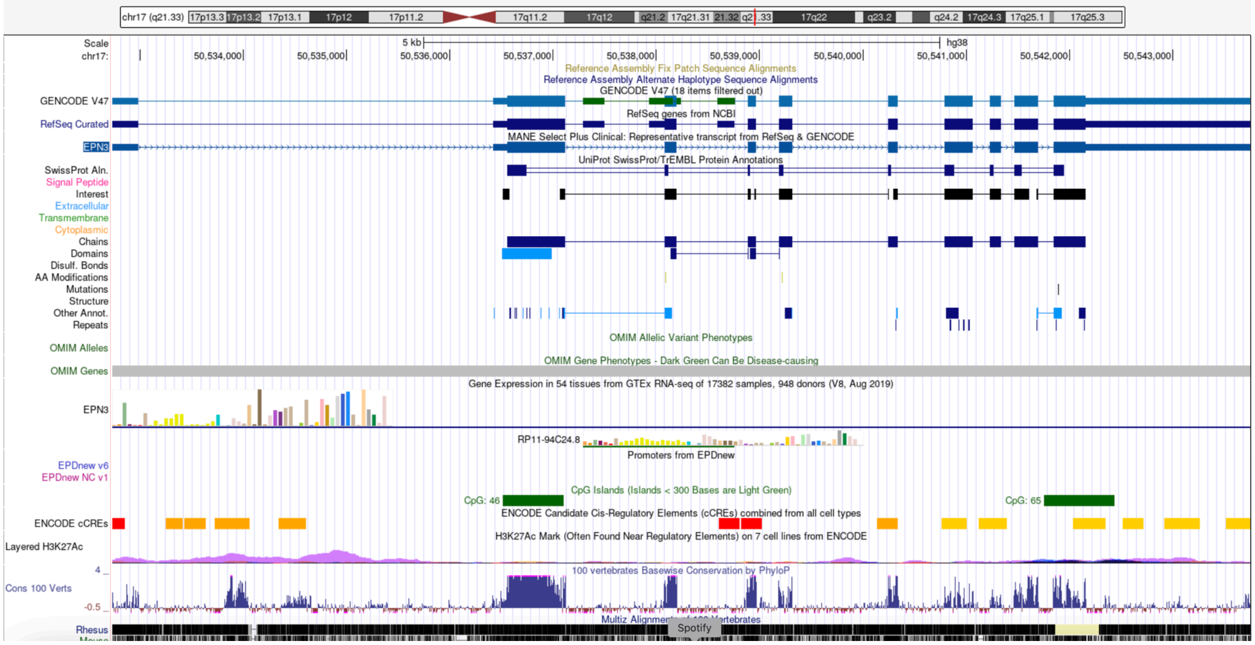

Counts
0
2000
4000
6000
Metastatic
Tumor
Group



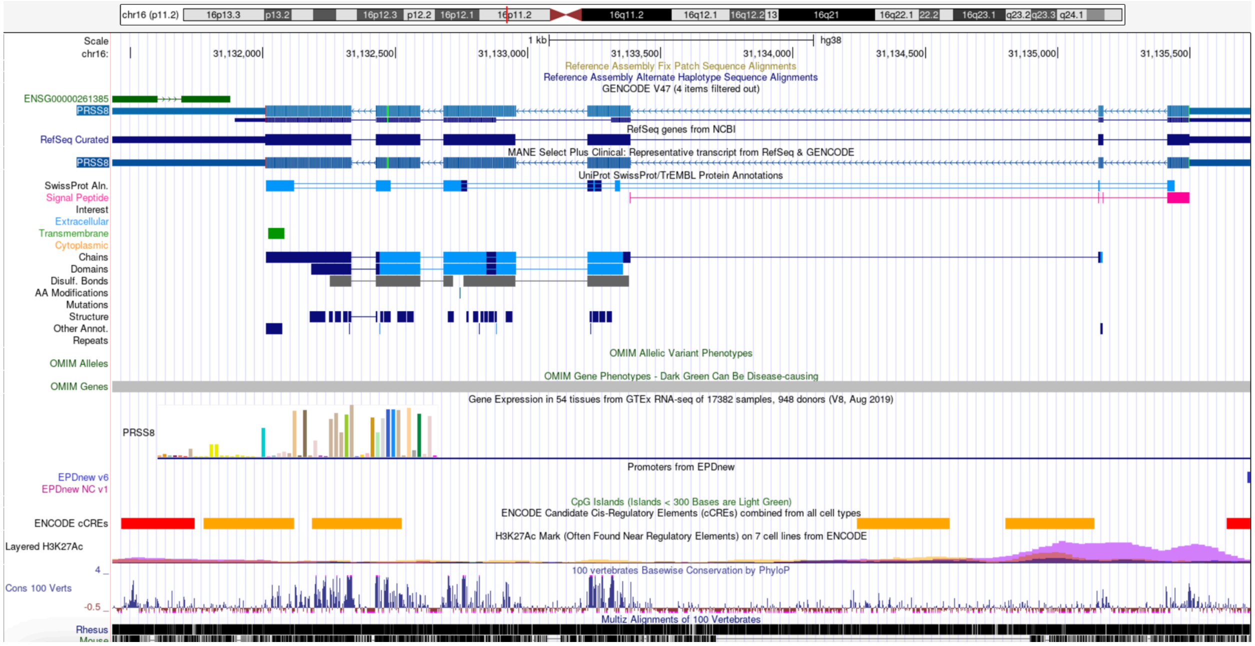
5. Using the UCSC genome browser for a few genes, the following plots visualize CpG sites and protein domains for 3 genes:
    a. TTC22


chr1 (p32.3)
Scale
chr1:
10 kb
hg38
54,782,000 | 54,783,000 | 54,784,000 | 54,785,000 | 54,786,000 | 54,787,000 | 54,788,000 | 54,789,000 | 54,790,000 | 54,791,000 | 54,792,000 | 54,793,000 | 54,794,000 | 54,795,000 | 54,796,000 | 54,797,000 | 54,798,000 | 54,799,000 | 54,800,000 | 54,801,000
Reference Assembly Fix Patch Sequence Alignments
Reference Assembly Alternate Haplotype Sequence Alignments
GENCODE V47 (3 items filtered out)
GENCODE V47
RefSeq genes from NCBI
RefSeq Curated
MANE Select Plus Clinical: Representative transcript from RefSeq & GENCODE
TTC22
UniProt SwissProt/TrEMBL Protein Annotations
SwissProt Aln.
Signal Peptide
Interest
Extracellular
Transmembrane
Cytoplasmic
Chains
Domains
Disulf. Bonds
AA Modifications
Mutations
Structure
Other Annot.
Repeats
OMIM Allelic Variant Phenotypes
OMIM Alleles
OMIM Gene Phenotypes - Dark Green Can Be Disease-causing
OMIM Genes
Gene Expression in 54 tissues from GTEx RNA-seq of 17382 samples, 948 donors (V8, Aug 2019)
TTC22
RP11-67L3.2
Promoters from EPDnew
EPDnew v6
EPDnew NC v1
CpG Islands (Islands < 300 Bases are Light Green)
CpG: 88
CpG: 80
ENCODE Candidate Cis-Regulatory Elements (cCREs) combined from all cell types
ENCODE cCREs
H3K27Ac Mark (Often Found Near Regulatory Elements) on 7 cell lines from ENCODE
Layered H3K27Ac
4 _
100 vertebrates Basewise Conservation by PhyloP
Cons 100 Verts
-0.5 _
Friendly Streaming
Vertebrates

    b. EPN3

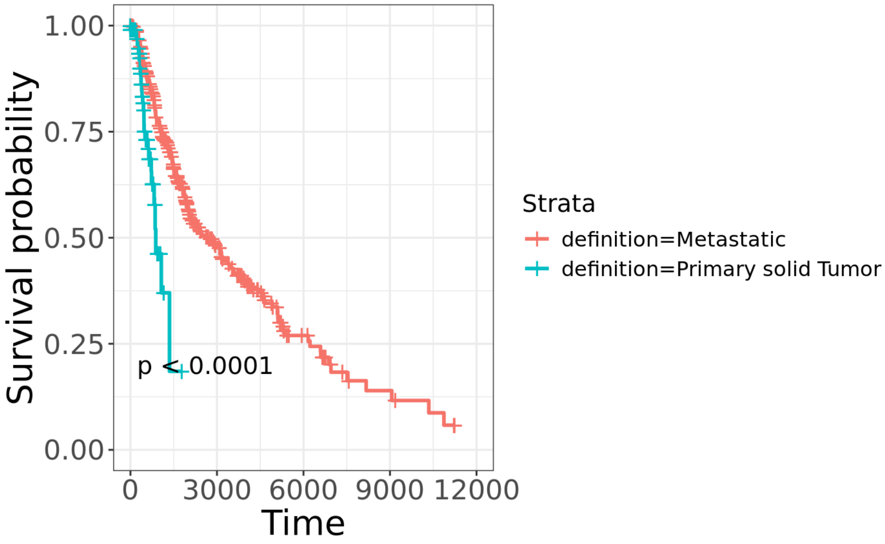

chr17 (q21.33)
17p13.3 17p13.2 17p13.1 17p12 17p11.2 17q11.2 17q12 q21.2 17q21.31 21.32 q21.33 17q22 q23.2 q24.2 17q24.3 17q25.1 17q25.3
Scale
chr17:
5 kb
hg38
50,534,000
50,535,000
50,536,000
50,537,000
50,538,000
50,539,000
50,540,000
50,541,000
50,542,000
50,543,000
Reference Assembly Fix Patch Sequence Alignments
Reference Assembly Alternate Haplotype Sequence Alignments
GENCODE V47 (18 items filtered out)
GENCODE V47
RefSeq genes from NCBI
RefSeq Curated
MANE Select Plus Clinical: Representative transcript from RefSeq & GENCODE
EPN3
UniProt SwissProt/TrEMBL Protein Annotations
SwissProt Aln.
Signal Peptide
Interest
Extracellular
Transmembrane
Cytoplasmic
Chains
Domains
Disulf. Bonds
AA Modifications
Mutations
Structure
Other Annot.
Repeats
OMIM Allelic Variant Phenotypes
OMIM Alleles
OMIM Gene Phenotypes - Dark Green Can Be Disease-causing
OMIM Genes
Gene Expression in 54 tissues from GTEx RNA-seq of 17382 samples, 948 donors (V8, Aug 2019)
EPN3
RP11-94C24.8
Promoters from EPDnew
EPDnew v6
EPDnew NC v1
CpG Islands (Islands < 300 Bases are Light Green)
CpG: 46
CpG: 65
ENCODE Candidate Cis-Regulatory Elements (cCREs) combined from all cell types
ENCODE cCREs
H3K27Ac Mark (Often Found Near Regulatory Elements) on 7 cell lines from ENCODE
Layered H3K27Ac
4
100 vertebrates Basewise Conservation by PhyloP
Cons 100 Verts
-0.5
Multiz Alignments of 100 Vertebrates
Spotify
Rhesus
Mouse

c. PRSS8


chr16 (p11.2)
16p13.3 p13.2 16p12.3 p12.2 16p12.1 16p11.2 16q11.2 16q12.1 16q12.2 13 16q21 16q22.1 22.2 16q23.1 q23.2 q23.3 q24.1
Scale
chr16:
1 kb
hg38
31,132,000
31,132,500
31,133,000
31,133,500
31,134,000
31,134,500
31,135,000
31,135,500
Reference Assembly Fix Patch Sequence Alignments
Reference Assembly Alternate Haplotype Sequence Alignments
GENCODE V47 (4 items filtered out)
ENSG00000261385
PRSS8
RefSeq genes from NCBI
RefSeq Curated
MANE Select Plus Clinical: Representative transcript from RefSeq & GENCODE
PRSS8
UniProt SwissProt/TrEMBL Protein Annotations
SwissProt Aln.
Signal Peptide
Interest
Extracellular
Transmembrane
Cytoplasmic
Chains
Domains
Disulf. Bonds
AA Modifications
Mutations
Structure
Other Annot.
Repeats
OMIM Allelic Variant Phenotypes
OMIM Alleles
OMIM Gene Phenotypes - Dark Green Can Be Disease-causing
OMIM Genes
Gene Expression in 54 tissues from GTEx RNA-seq of 17382 samples, 948 donors (V8, Aug 2019)
PRSS8
Promoters from EPDnew
EPDnew v6
EPDnew NC v1
CpG Islands (Islands < 300 Bases are Light Green)
ENCODE Candidate Cis-Regulatory Elements (cCREs) combined from all cell types
ENCODE cCREs
H3K27Ac Mark (Often Found Near Regulatory Elements) on 7 cell lines from ENCODE
Layered H3K27Ac
4
100 vertebrates Basewise Conservation by PhyloP
Cons 100 Verts
-0.5
Multiz Alignments of 100 Vertebrates
Rhesus
Mouse

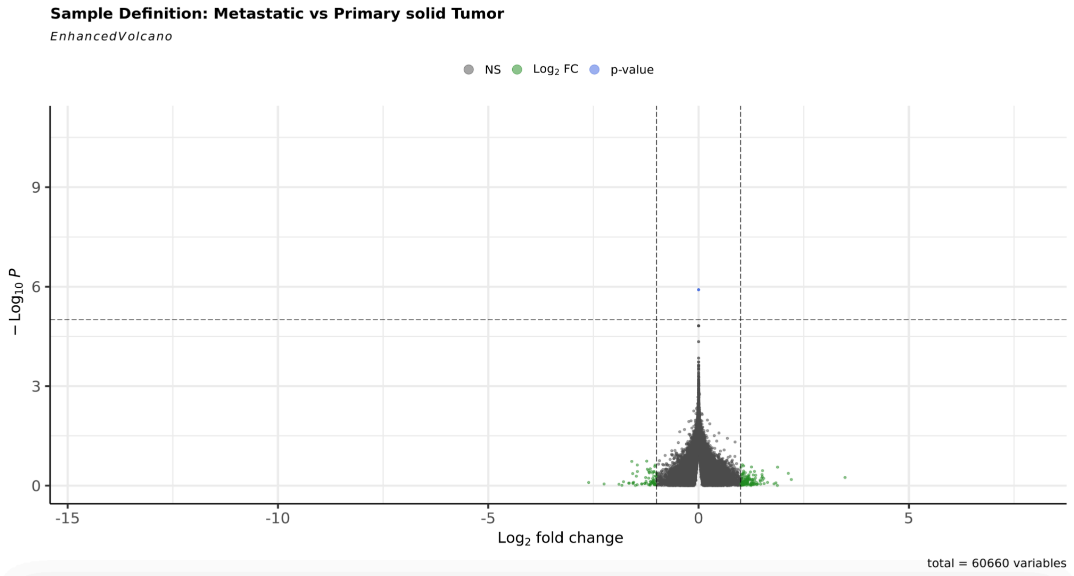
Part 3: Results and Interpretations:

1. Differences in survival between metastatic and non-metastatic patients



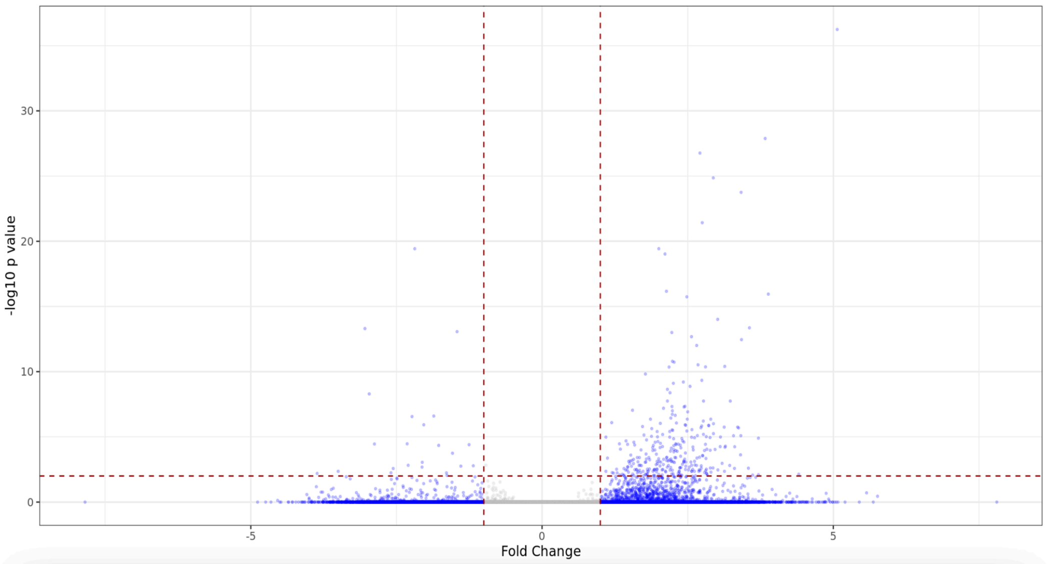
The KM plot shows potential differences in survival between the metastatic group and non-metastatic (Primary solid Tumor) group for TCGA SKCM patients. Since the p-value is less than 0.0001, there is potentially enough evidence to conclude that there are differences in the survival rates between the metastatic and non-metastatic groups for TCGA SKCM patients. However, there were lots of censored points along both strata and the Primary solid Tumor group seems to drop off early, which could affect the strength of the evidence for this difference conclusion. Therefore, it would be hard to generalize this conclusion to an entire population rather than the sample analyzed.

2. **Expression differences between metastatic and non-metastatic patients**



The EnhancedVolcano plot displays the differentially expressed genes (DEGs) between metastatic and non-metastatic (primary solid tumor) TCGA SKCM patient groups, controlling for covariates of treatment effects (radiation, chemotherapy, immunotherapy, molecular therapy, and vaccine), race, gender, and vital status using DESeq2. We see DEGs greater than the log2FC threshold (+1) and less than the log2FC threshold (-1), indicating that these genes were upregulated and downregulated respectively between metastatic and nonmetastatic groups. Furthermore, we see only 1 DEG greater than the -log10(padj) threshold of 0.05, showing a statistically significant difference in expression for such genes. We can conclude that there are some DEGs that are upregulated and some are downregulated and that there seems to be only one DEG that is significant yet not differentially expressed; therefore, there seems to be no statistically significant difference in expression in genes between metastatic and non-metastatic TCGA SKCM patient groups.

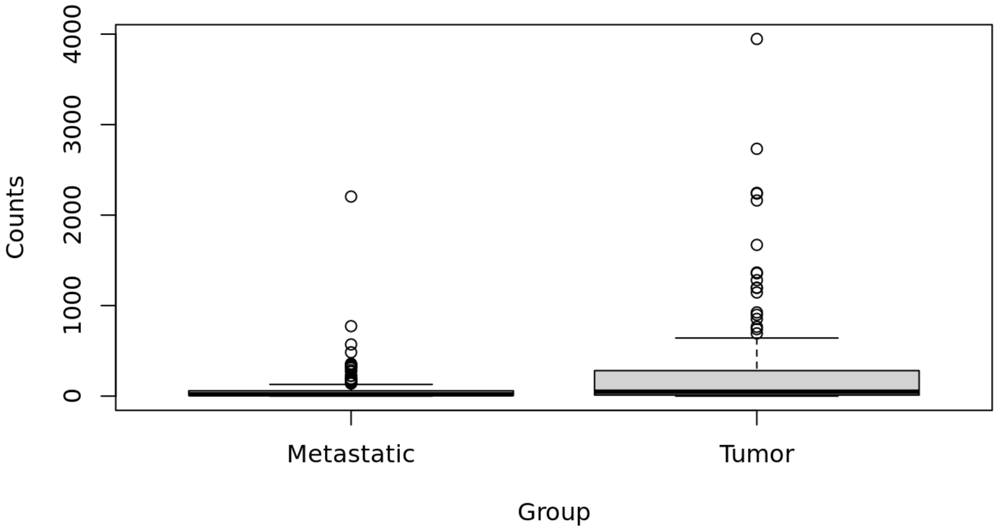
3. **Methylation differences between metastatic and non-metastatic patients**

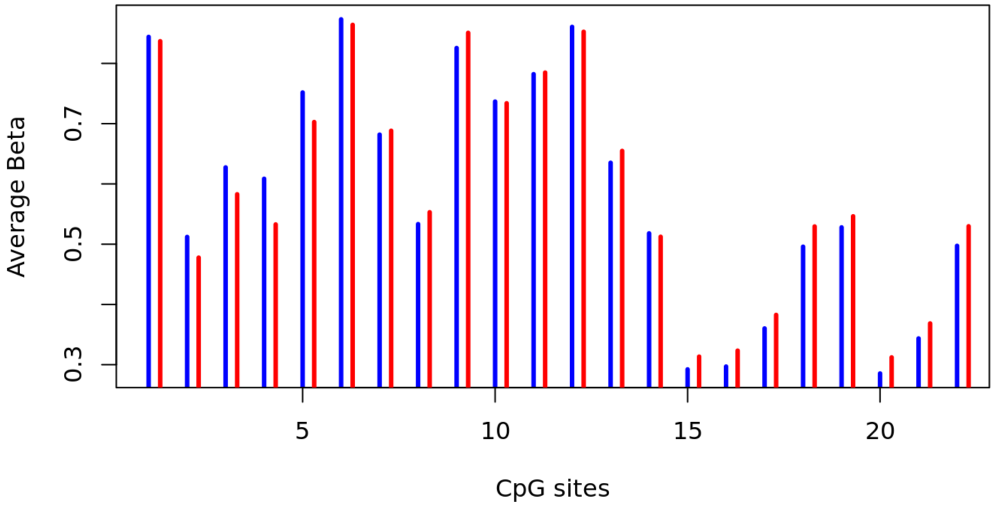


The volcano plot displays naive differential methylation between non-metastatic (Primary solid Tumor) and metastatic TCGA patient groups. The blue dots represent changes in methylation, where the blue dots on the right, which have an FC greater than the threshold of +1, indicate hypermethylation in the tumor, and the blue dots on the left, which have an FC less than the threshold of -1, indicate hypomethylation in the tumor. However, there are a lot of hypermethylated or hypomethylated points that are not significant as they lie under the threshold for -log10p-value. Since there are more statistically significant points with significant differences in methylation for those that are hypermethylated than those that are hypomethylated, we can conclude that there is evidence that there seems to be more patients with tumors that are hypermethylated than hypomethylated between the TCGA SKCM metastatic and non-metastatic patient groups.

4. **Direct comparison of transcriptional activity to methylation status for 10 genes**

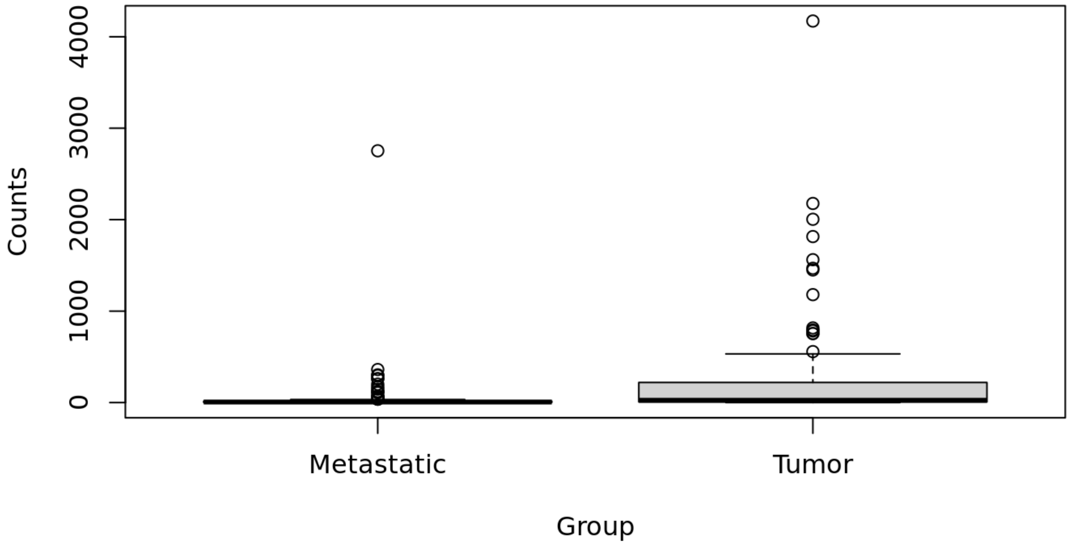
   a. TTC22

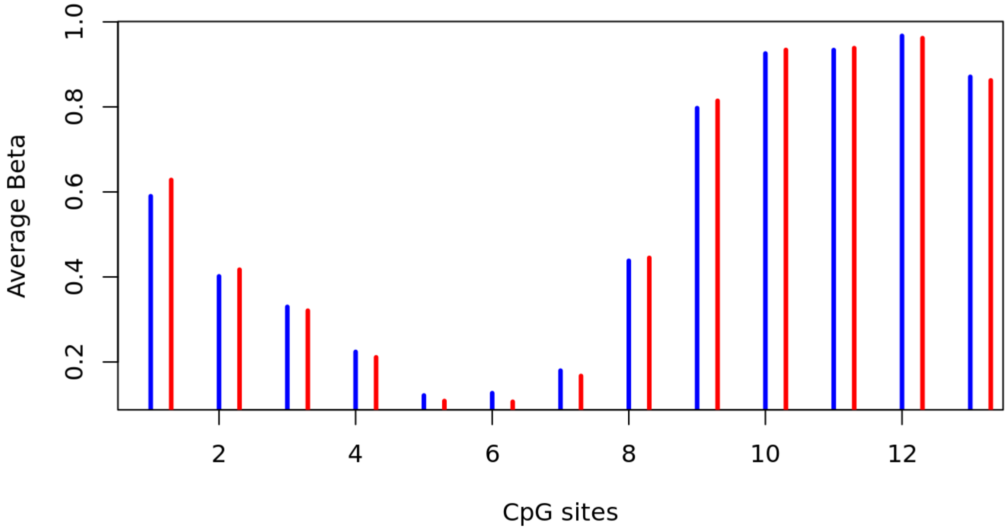




The boxplot shows a higher count of transcriptomic expression for the TTC22 gene for the Primary solid tumor group than in metastatic groups. The horizontal line plot shows differences in average methylation beta values where red is the tumor group and blue is the metastatic group at 23 different CpG sites for the TTC22 gene. Overall, there seem to be some differences between the metastatic and non-metastatic groups, where metastatic patients have higher average beta methylation values in earlier CpG sites and non-metastatic/tumor patients have higher average beta methylation values in later CpG sites. There seems to be a significant average difference in transcriptomic activity for the metastatic and non-metastatic TCGA SKCM patients.

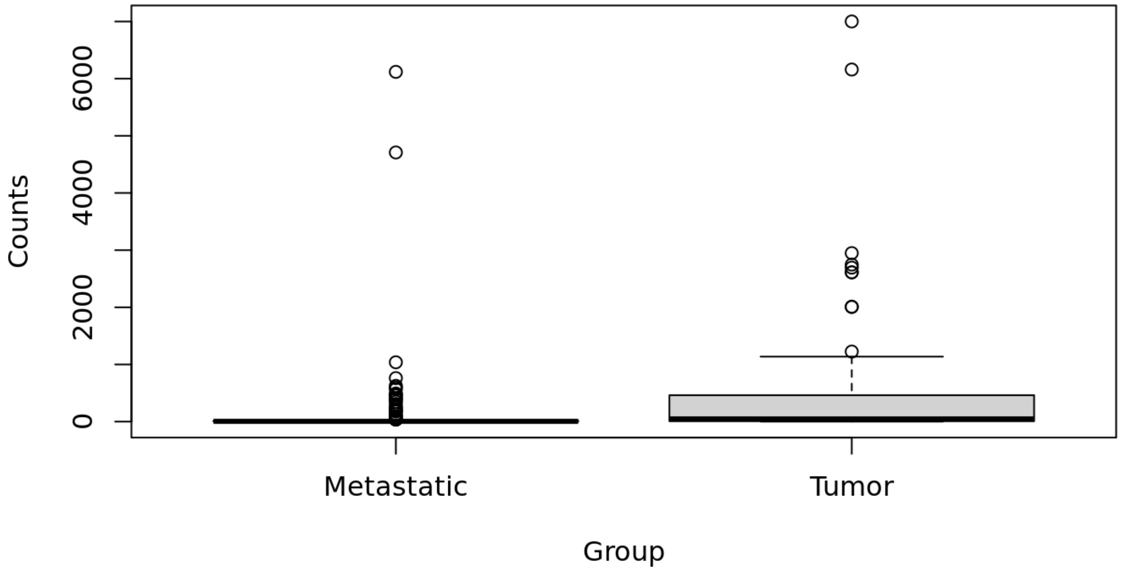
b. EPN3

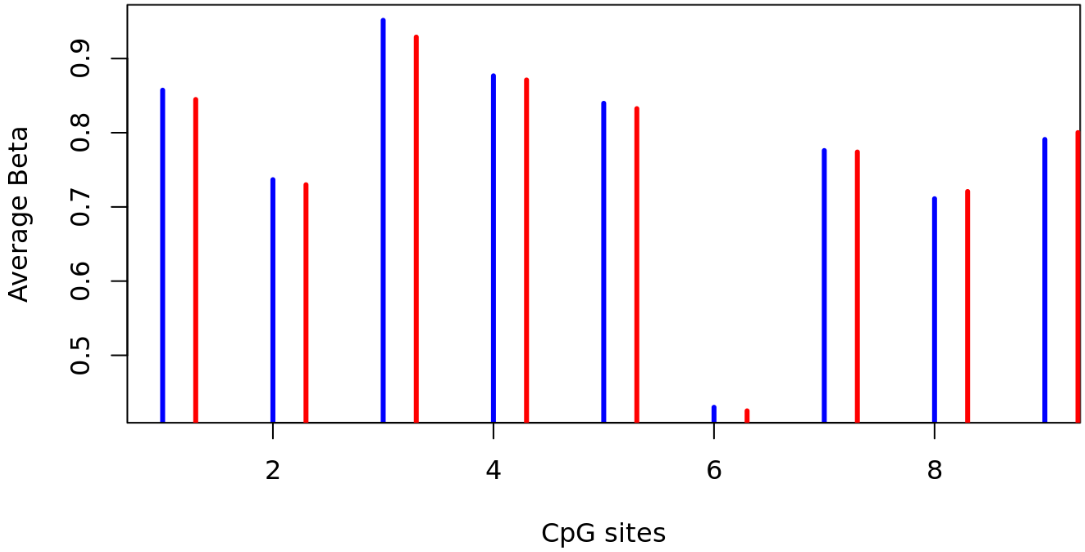




The boxplot shows a significantly higher average count of transcriptomic expression for the EPN3 gene for the Primary solid tumor (non-metastatic) group than in metastatic groups. The horizontal line plot shows differences in average methylation beta values where red is the tumor group and blue is the metastatic group at 13 different CpG sites for the EPN3 gene. Overall, there are not too many differences in methylation values between the metastatic and non-metastatic groups with non-metastatic patients having a relatively higher average beta methylation value at a couple of CpG sites, but there seems to be a significant average difference in transcriptomic activity for the metastatic and non-metastatic TCGA SKCM patients.

c. PRSS8

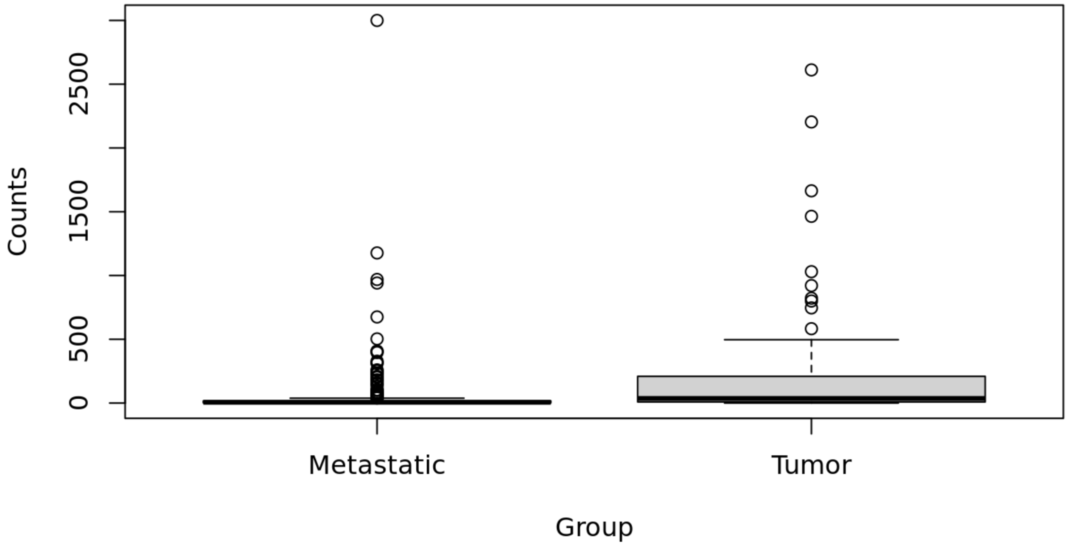


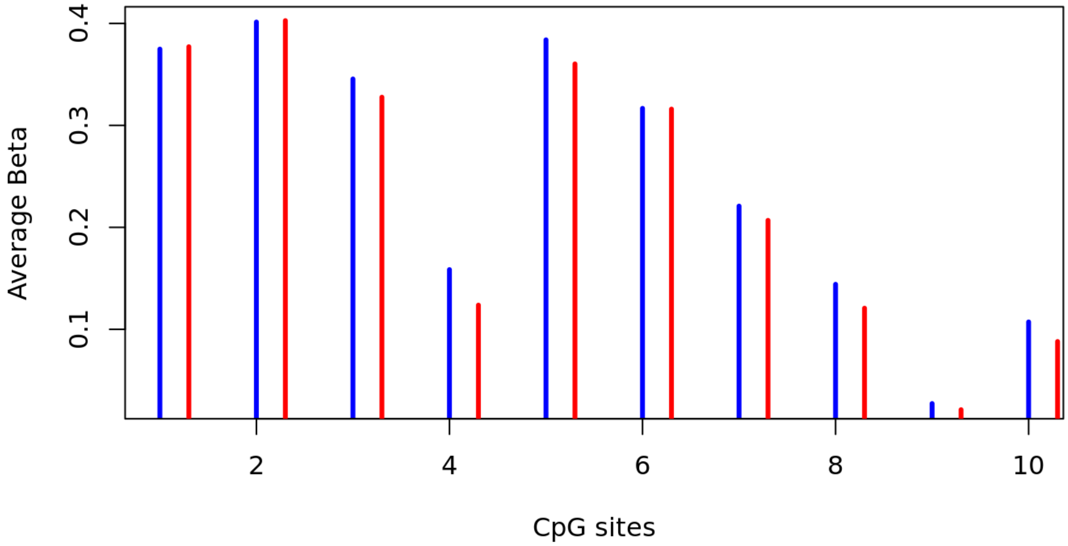


The boxplot shows a significantly higher average count of transcriptomic expression for the PRSS8 gene for the primary solid tumor (non-metastatic) group than for the metastatic group.

The horizontal line plot shows differences in average methylation beta values where red is the tumor group and blue is the metastatic group at 9 different CpG sites for the PRSS8 gene. Overall, there are not too many differences in beta methylation values between the metastatic and non-metastatic groups, but there seems to be a significant average difference in transcriptomic activity for the metastatic and non-metastatic TCGA SKCM patients.

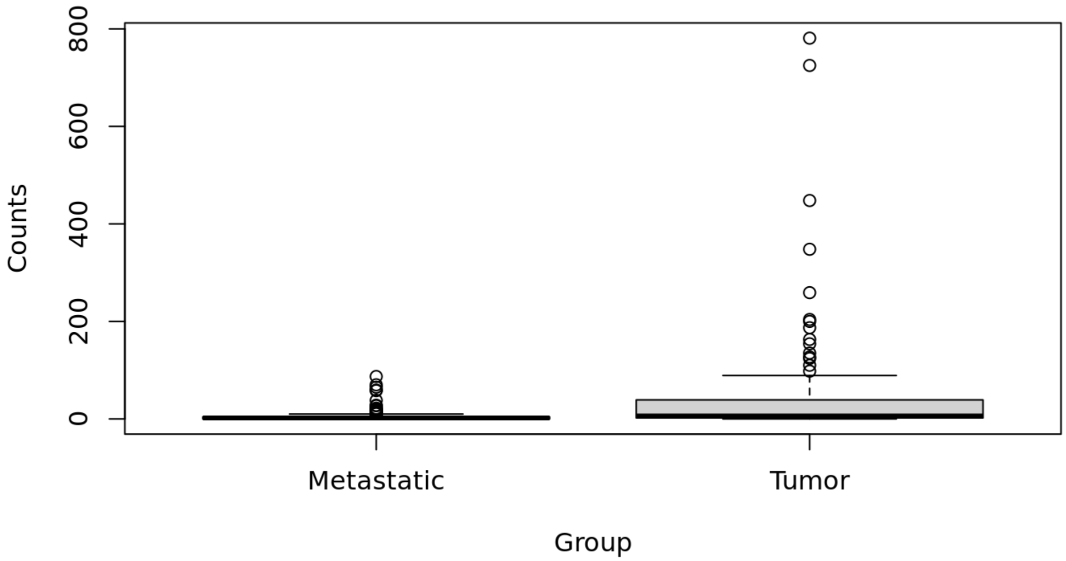
d. DLX3

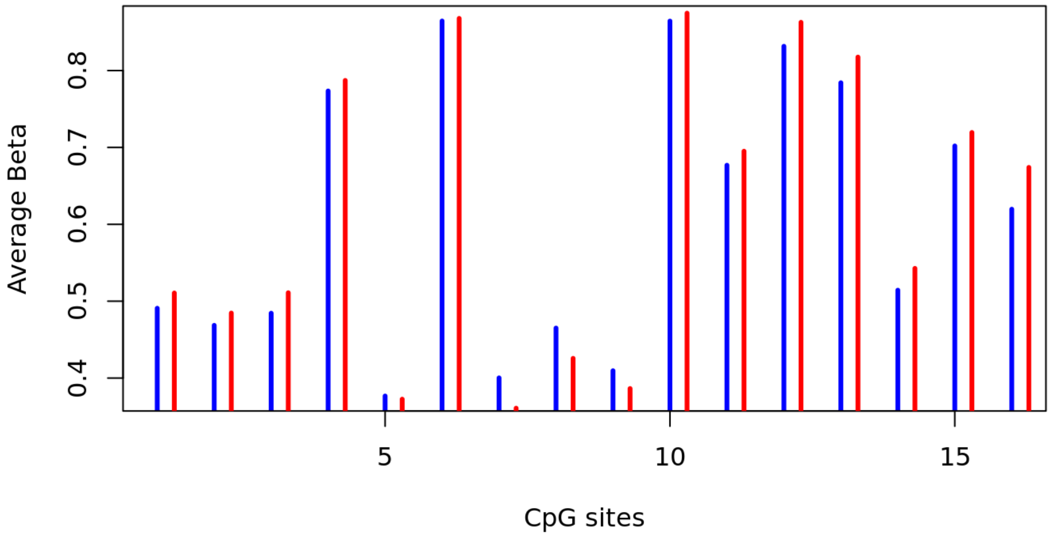




The boxplot shows a higher average count of transcriptomic expression for the DLX3 gene for the primary solid tumor (non-metastatic) group than for the metastatic group. The horizontal line plot shows differences in average methylation beta values where red is the tumor group and blue is the metastatic group at 10 different CpG sites for the DLX3 gene. There are some differences in beta methylation values between the metastatic and non-metastatic groups, where the metastatic group has a higher average beta methylation value than the non-metastatic (tumor) group, and there seems to be a significant average difference in transcriptomic activity for the metastatic and non-metastatic TCGA SKCM patients.

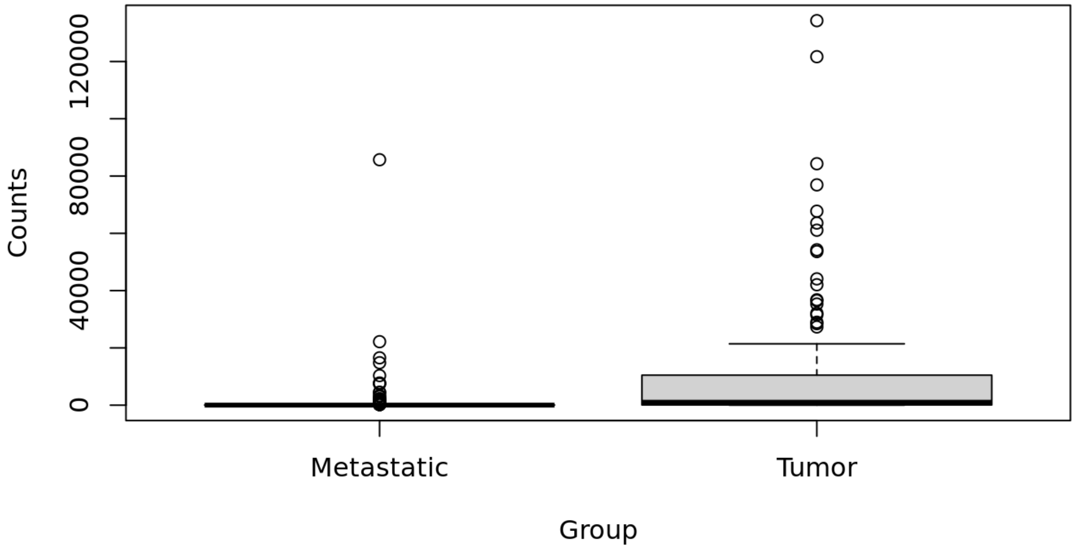
e. CNGB1

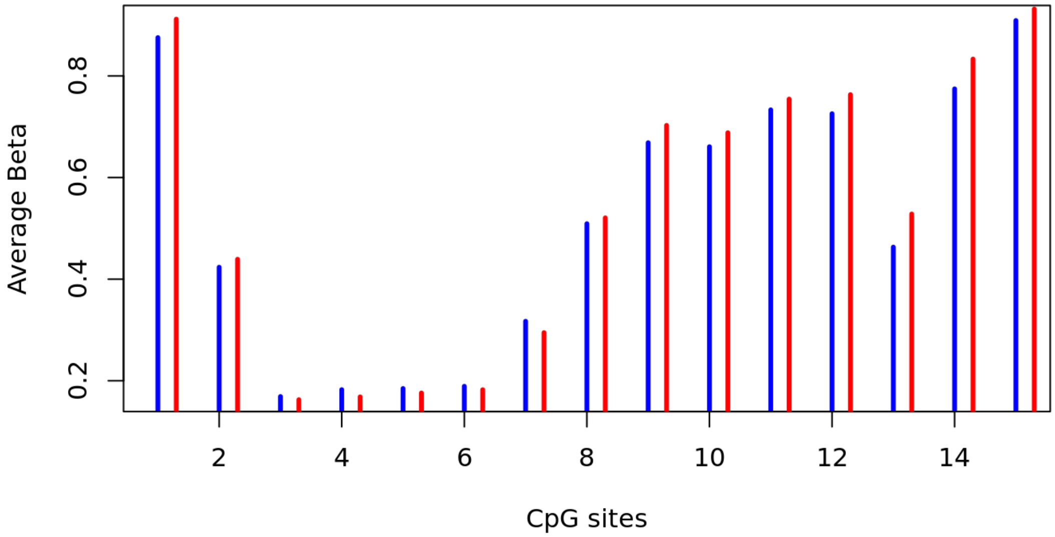




The boxplot shows a higher average count of transcriptomic expression for the CNGB1 gene for the primary solid tumor (non-metastatic) group than for the metastatic group. The horizontal line plot shows differences in average methylation beta values where red is the tumor group and blue is the metastatic group at 16 different CpG sites for the CNGB1 gene. There are some differences in beta methylation values between the metastatic and non-metastatic groups, where the tumor/non-metastatic group has mostly higher average beta methylation values than the metastatic group. There also seems to be a significant average difference in transcriptomic activity for the metastatic and non-metastatic TCGA SKCM patients.

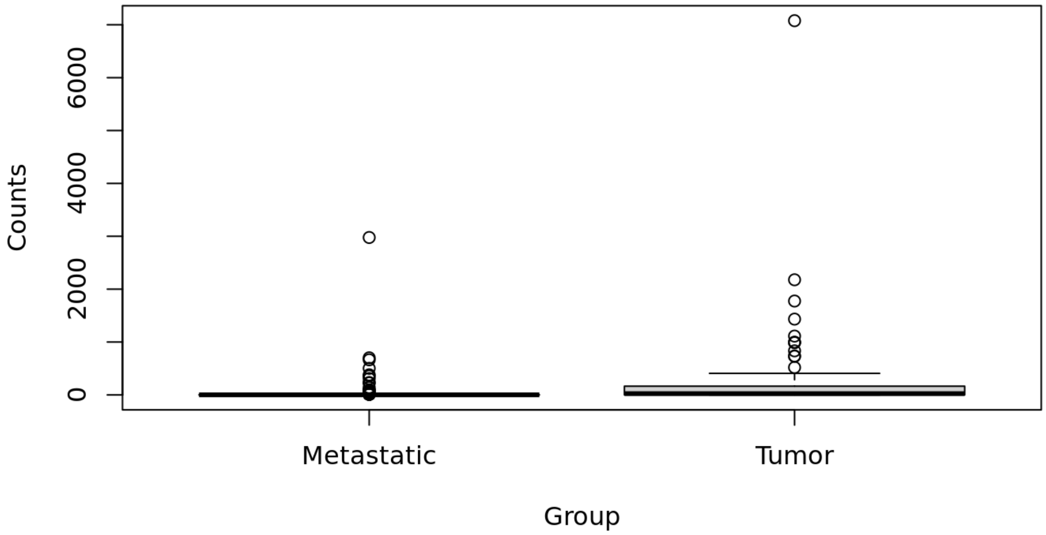
f. PKP1

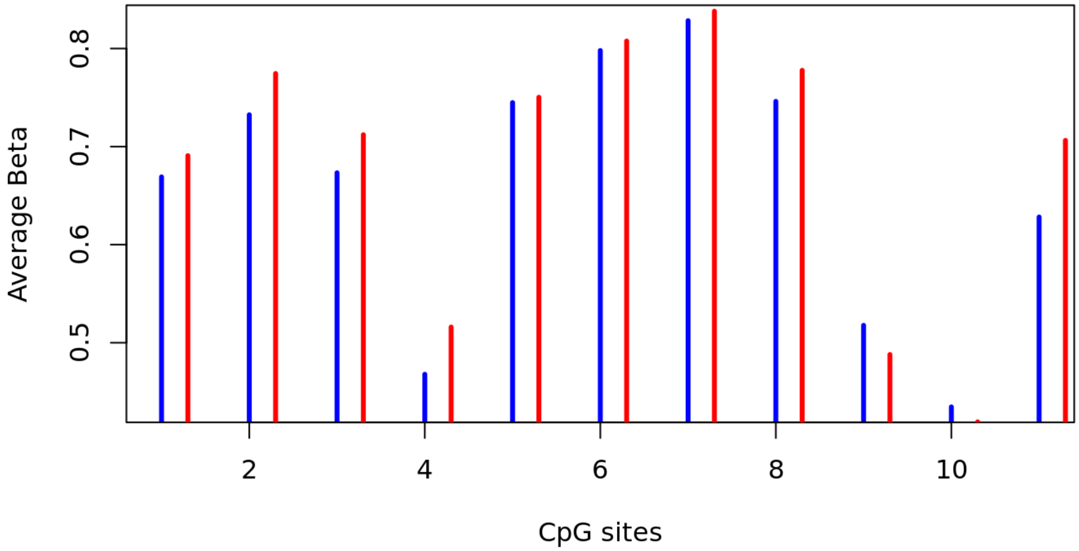




The boxplot shows a higher average count of transcriptomic expression for the PKP1 gene for the primary solid tumor (non-metastatic) group than for the metastatic group. The horizontal line plot shows differences in average methylation beta values where red is the tumor group and blue is the metastatic group at 15 different CpG sites for the PKP1 gene. There are some differences in beta methylation values between the metastatic and non-metastatic groups, where the tumor/non-metastatic group has mostly higher average beta methylation values than the metastatic group; there also seems to be a significant average difference in transcriptomic activity for the metastatic and non-metastatic TCGA SKCM patients.

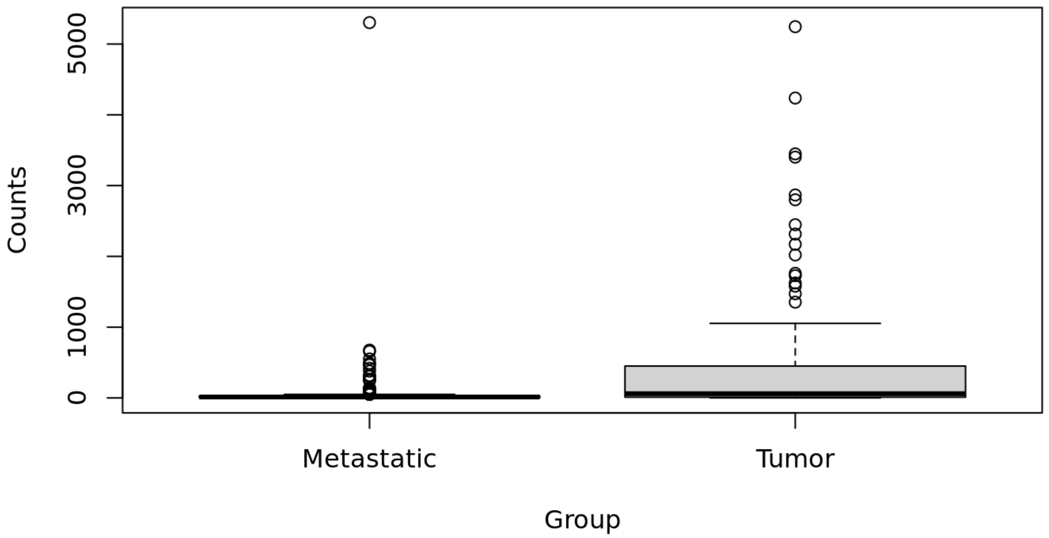
g. GRHL2

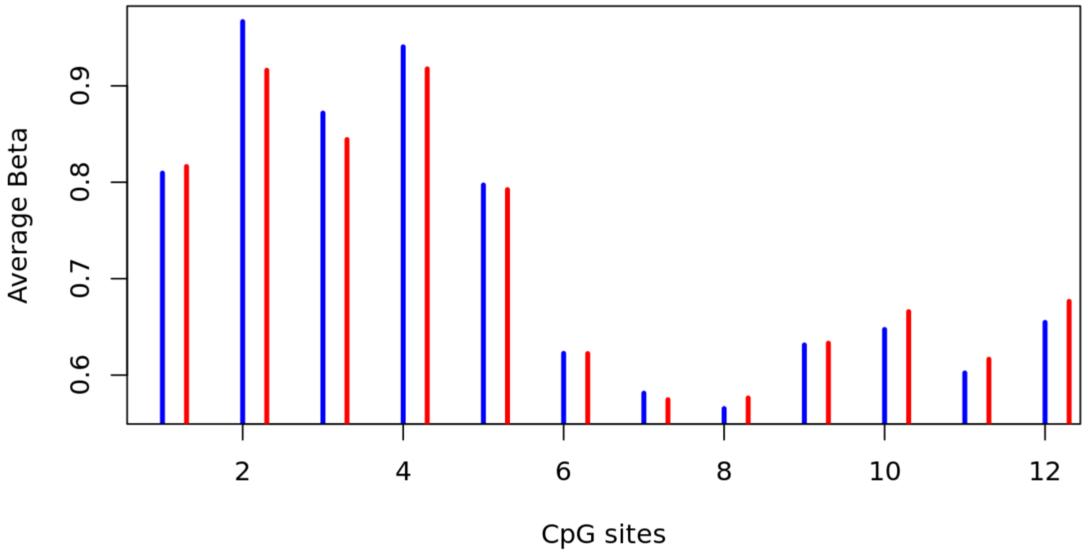




The boxplot shows a somewhat higher average count of transcriptomic expression for the GRHL2 gene for the primary solid tumor (non-metastatic) group than for the metastatic group. The horizontal line plot shows differences in average methylation beta values where red is the tumor group and blue is the metastatic group at 11 different CpG sites for the GRHL2 gene. There are some differences in beta methylation values between the metastatic and non-metastatic groups, where the tumor/non-metastatic group has mostly higher average beta methylation values than the metastatic group but the metastatic group has higher average beta methylation values for the 9 and 10 CpG sites; there also seems to be a significant average difference in transcriptomic activity for the metastatic and non-metastatic TCGA SKCM patients.

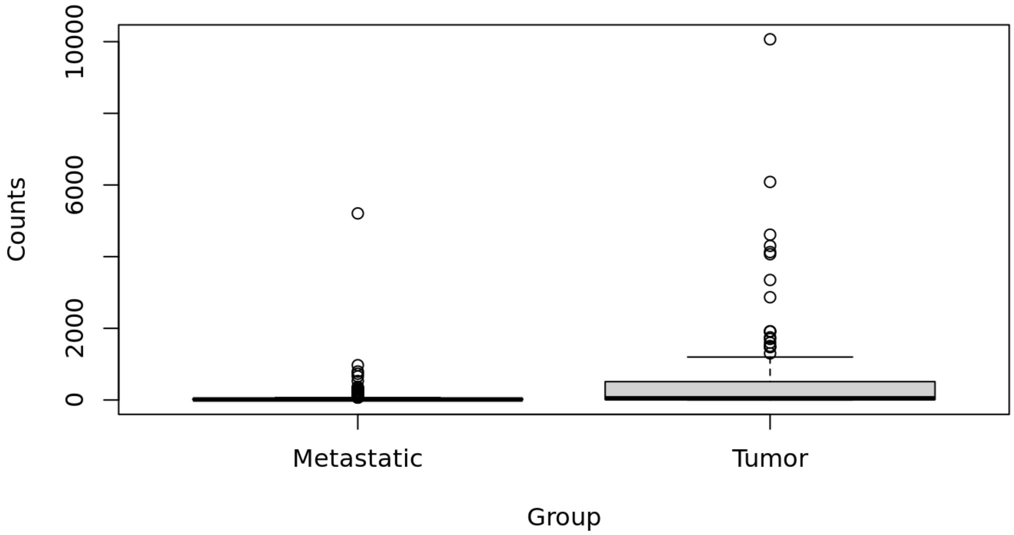
h. ESRP2

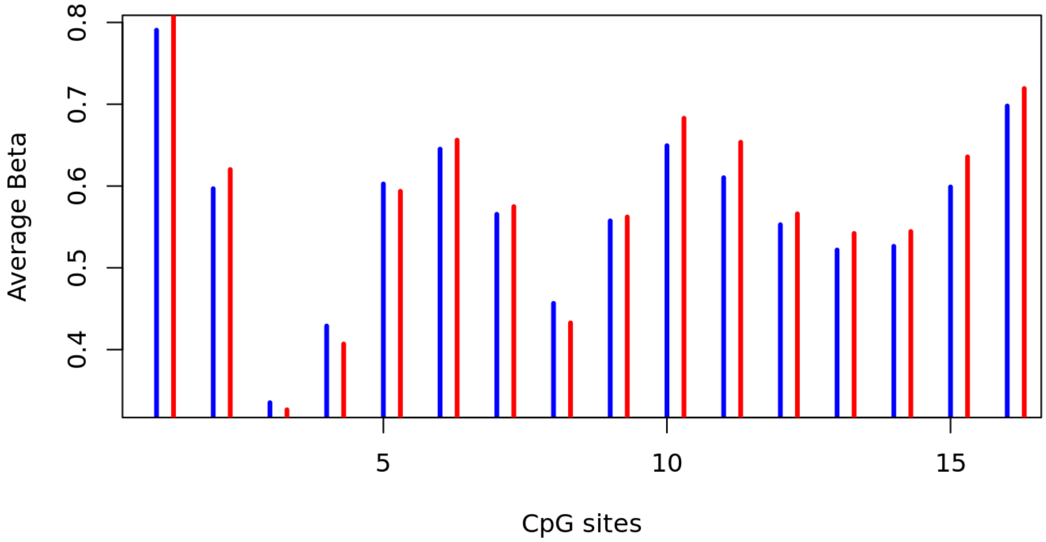




The boxplot shows a higher average count of transcriptomic expression for the ESRP2 gene for the primary solid tumor (non-metastatic) group than for the metastatic group. The horizontal line plot shows differences in average methylation beta values where red is the tumor group and blue is the metastatic group at 12 different CpG sites for the ESRP2 gene. There are some differences in beta methylation values between the metastatic and non-metastatic groups, where the tumor/non-metastatic group has lower average beta methylation values at earlier CpG sites then higher average beta methylation values at later CpG sites; there also seems to be a significant average difference in transcriptomic activity for the metastatic and non-metastatic TCGA SKCM patients.

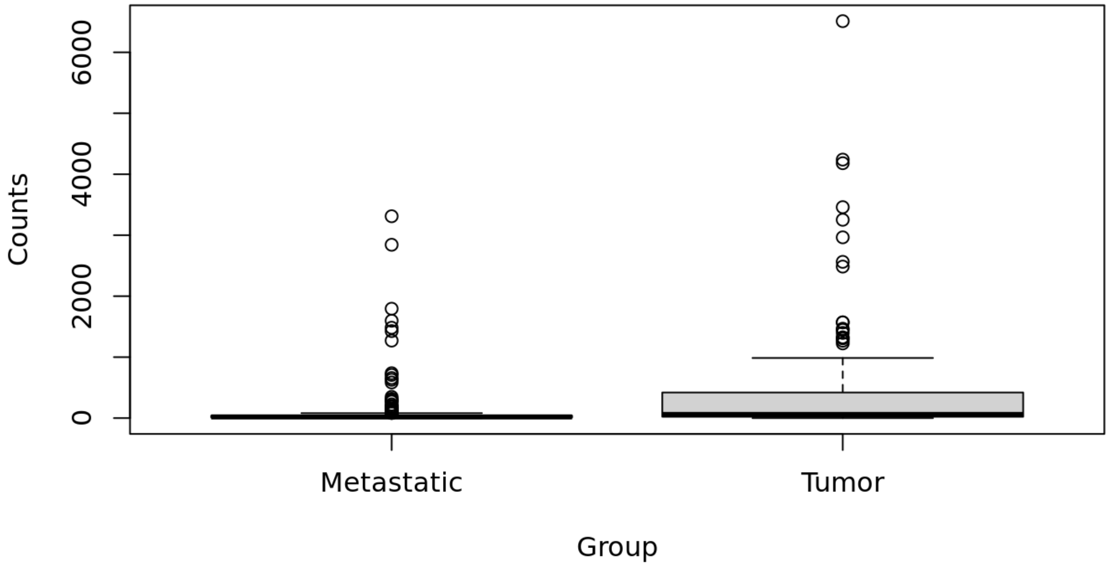
i. EPHX3

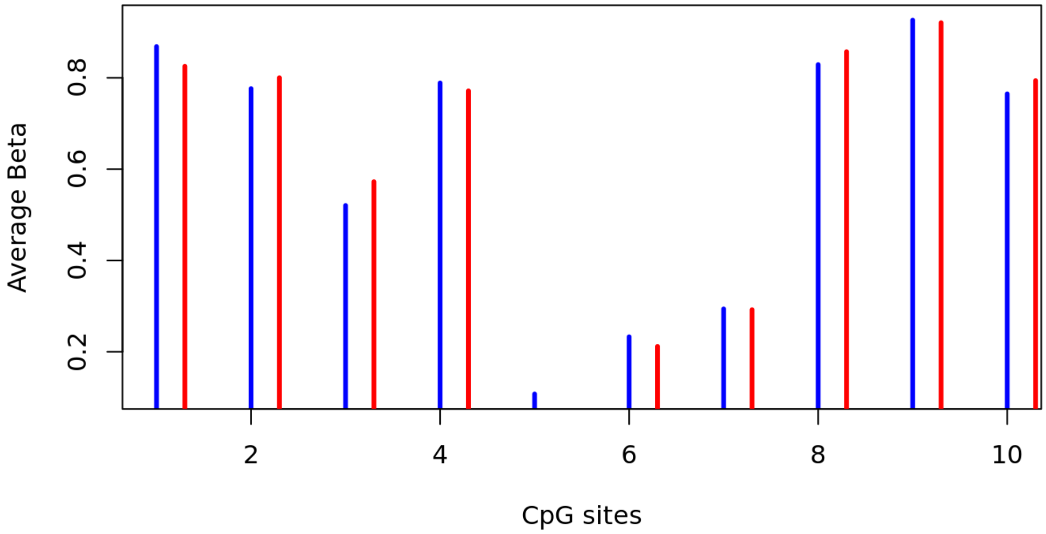




The boxplot shows a higher average count of transcriptomic expression for the EPHX3 gene for the primary solid tumor (non-metastatic) group than for the metastatic group. The horizontal line plot shows differences in average methylation beta values where red is the tumor group and blue is the metastatic group at 16 different CpG sites for the EPHX3 gene. There are differences in beta methylation values between the metastatic and non-metastatic groups, where the tumor/non-metastatic group has mostly higher average beta methylation values at CpG sites but some sites have lower average beta methylation values; there also seems to be a significant average difference in transcriptomic activity for the metastatic and non-metastatic TCGA SKCM patients.

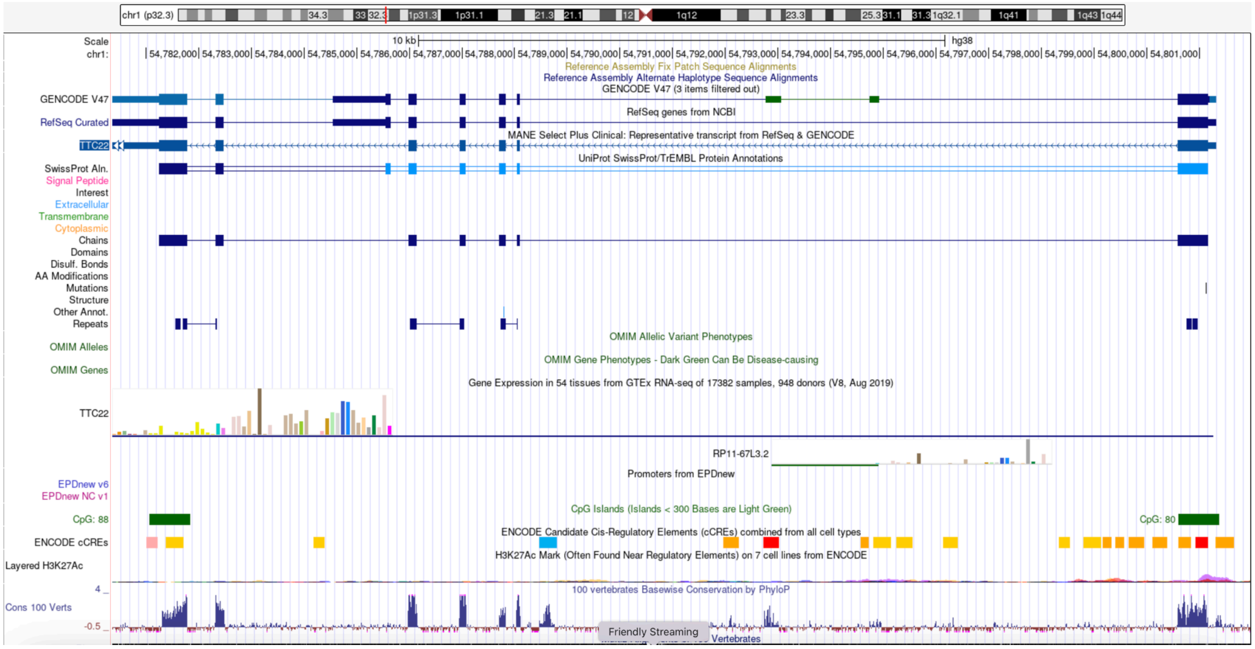
j. SCNN1A





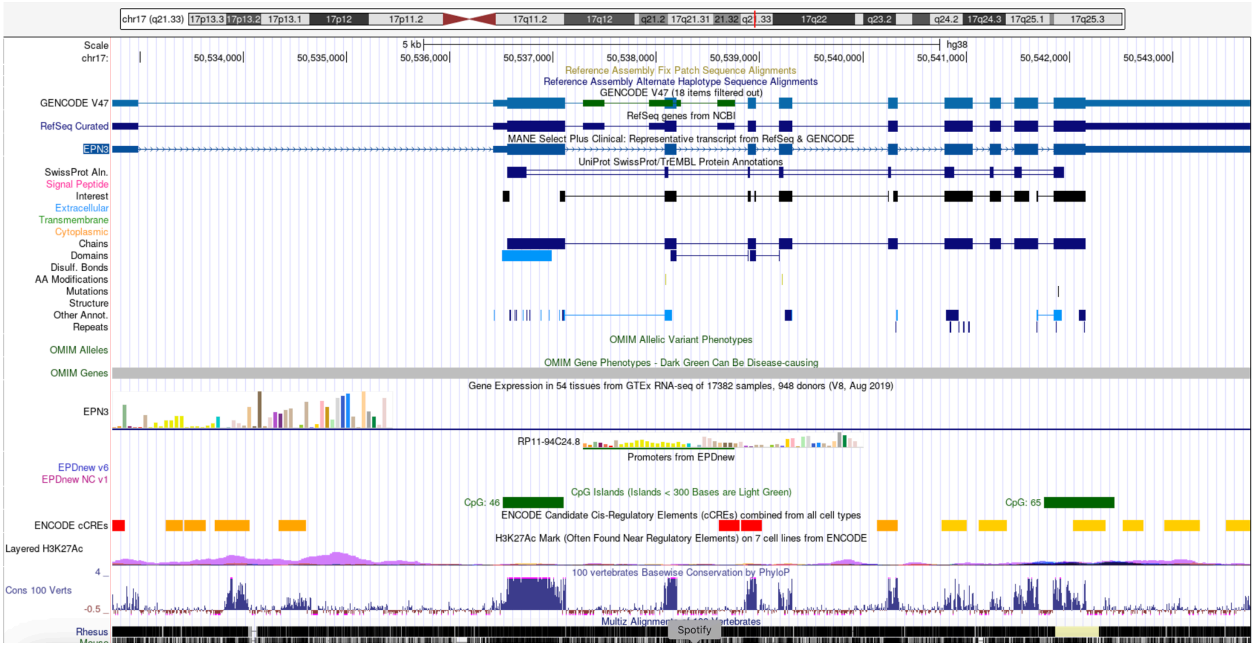
The boxplot shows a higher average count of transcriptomic expression for the SCNN1A gene for the primary solid tumor (non-metastatic) group than for the metastatic group. The horizontal line plot shows differences in average methylation beta values where red is the tumor group and blue is the metastatic group at 10 different CpG sites for the SCNN1A gene. There are differences in beta methylation values between the metastatic and non-metastatic groups, where the tumor/non-metastatic group and metastatic group fluctuate in having higher average beta methylation values between CpG sites; there also seems to be a significant average difference in transcriptomic activity for the metastatic and non-metastatic TCGA SKCM patients.

5. **Visualization of CpG sites and protein domains for 3 genes (UCSC genome browser)**
    a. TTC22


chr1 (p32.3)
Scale
chr1:
10 kb
hg38
54,782,000 | 54,783,000 | 54,784,000 | 54,785,000 | 54,786,000 | 54,787,000 | 54,788,000 | 54,789,000 | 54,790,000 | 54,791,000 | 54,792,000 | 54,793,000 | 54,794,000 | 54,795,000 | 54,796,000 | 54,797,000 | 54,798,000 | 54,799,000 | 54,800,000 | 54,801,000
Reference Assembly Fix Patch Sequence Alignments
Reference Assembly Alternate Haplotype Sequence Alignments
GENCODE V47 (3 items filtered out)
GENCODE V47
RefSeq genes from NCBI
RefSeq Curated
MANE Select Plus Clinical: Representative transcript from RefSeq & GENCODE
TTC22
UniProt SwissProt/TrEMBL Protein Annotations
SwissProt Aln.
Signal Peptide
Interest
Extracellular
Transmembrane
Cytoplasmic
Chains
Domains
Disulf. Bonds
AA Modifications
Mutations
Structure
Other Annot.
Repeats
OMIM Allelic Variant Phenotypes
OMIM Alleles
OMIM Gene Phenotypes - Dark Green Can Be Disease-causing
OMIM Genes
Gene Expression in 54 tissues from GTEx RNA-seq of 17382 samples, 948 donors (V8, Aug 2019)
TTC22
RP11-67L3.2
Promoters from EPDnew
EPDnew v6
EPDnew NC v1
CpG Islands (Islands &lt; 300 Bases are Light Green)
CpG: 88
CpG: 80
ENCODE Candidate Cis-Regulatory Elements (cCREs) combined from all cell types
ENCODE cCREs
H3K27Ac Mark (Often Found Near Regulatory Elements) on 7 cell lines from ENCODE
Layered H3K27Ac
4
100 vertebrates Basewise Conservation by PhyloP
Cons 100 Verts
-0.5
Friendly Streaming
Vertebrates

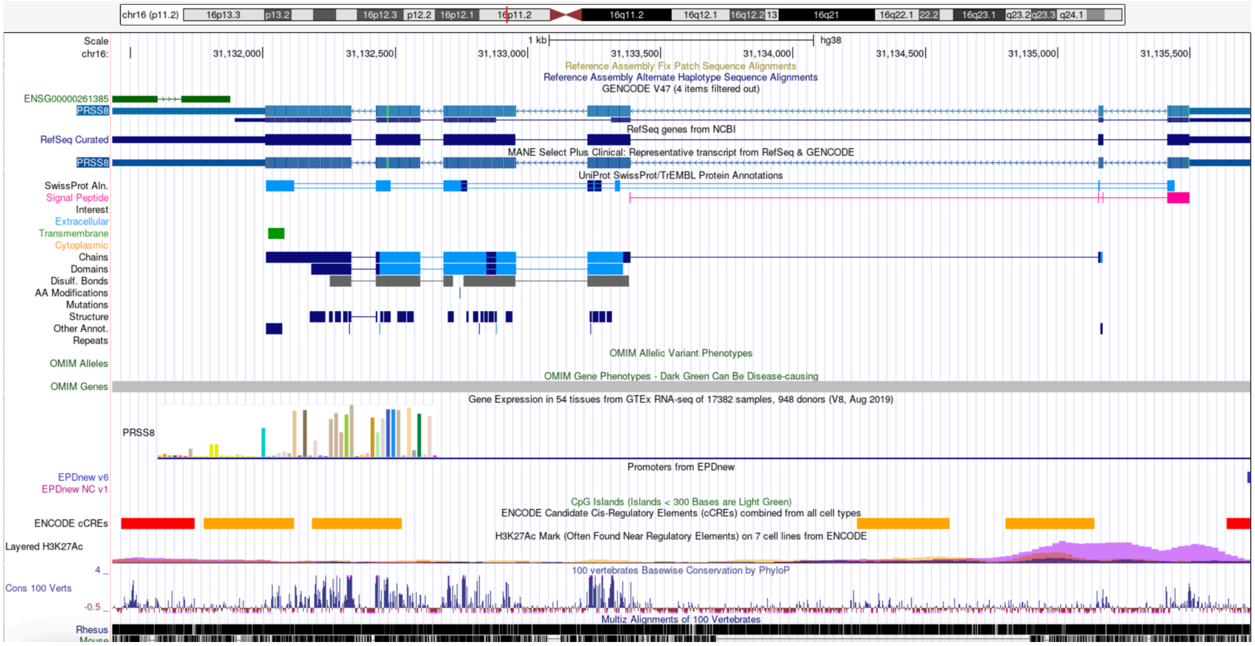
For the TTC22 gene, there seems to be increased expression at CpG islands 88 and 80 for protein chains and repeats. In correlation with the previous TTC22 plot, there seems to be greater methylation activity at earlier CpG sites for the metastatic group and greater methylation activity at later CpG sites for the non-metastatic (tumor) group. We can potentially conclude that, on average, earlier CpG sites will have higher protein chain and repeats expression with higher methylation expression for metastatic patients and likewise for later CpG sites for tumor group patients for the TTC22 gene. According to Fujiwara et al. (2019), we see TTC22 was hypermethylated for melanoma samples using melanocytes as a comparison, which somewhat supports the findings on hypermethylation of TTC22.

    b. EPN3


chr17 (q21.33)
Scale
chr17:
5 kb
hg38
50,534,000
50,535,000
50,536,000
50,537,000
50,538,000
50,539,000
50,540,000
50,541,000
50,542,000
50,543,000
Reference Assembly Fix Patch Sequence Alignments
Reference Assembly Alternate Haplotype Sequence Alignments
GENCODE V47 (18 items filtered out)
GENCODE V47
RefSeq genes from NCBI
RefSeq Curated
MANE Select Plus Clinical: Representative transcript from RefSeq & GENCODE
EPN3
UniProt SwissProt/TrEMBL Protein Annotations
SwissProt Aln.
Signal Peptide
Interest
Extracellular
Transmembrane
Cytoplasmic
Chains
Domains
Disulf. Bonds
AA Modifications
Mutations
Structure
Other Annot.
Repeats
OMIM Allelic Variant Phenotypes
OMIM Alleles
OMIM Gene Phenotypes - Dark Green Can Be Disease-causing
OMIM Genes
Gene Expression in 54 tissues from GTEx RNA-seq of 17382 samples, 948 donors (V8, Aug 2019)
EPN3
RP11-94C24.8
Promoters from EPDnew
EPDnew v6
EPDnew NC v1
CpG Islands (Islands &lt; 300 Bases are Light Green)
CpG: 46
CpG: 65
ENCODE Candidate Cis-Regulatory Elements (cCREs) combined from all cell types
ENCODE cCREs
H3K27Ac Mark (Often Found Near Regulatory Elements) on 7 cell lines from ENCODE
Layered H3K27Ac
100 vertebrates Basewise Conservation by PhyloP
4
Cons 100 Verts
-0.5
Multiz Alignments of 100 Vertebrates
Rhesus
Spotify

For the EPN3 gene, there seems to be increased expression at CpG islands 46 and 65 for protein chains, interest, and domains. In correlation with the previous EPN3 plot, there seems to be greater methylation activity at earlier CpG sites for the tumor/non-metastatic group and greater methylation activity at later CpG sites for the metastatic group. We can potentially conclude that, on average, earlier CpG sites will have higher protein chain, interest, and domains expression with higher methylation expression for non-metastatic patients and likewise for later CpG sites for metastatic group patients for the EPN3 gene. According to Tong et al. (2021), overexpression of 40% of breast cancers can be used to predict "distant metastasis," which somewhat supports the findings of later CpG sites with such higher protein expression for metastatic TCGA SKCM patients for the EPN3 gene.

c. PRSS8


chr16 (p11.2)
16p13.3
p13.2
16p12.3
p12.2
16p12.1
16p11.2
16q11.2
16q12.1
16q12.2
13
16q21
16q22.1
22.2
16q23.1
q23.2
q23.3
q24.1
Scale
chr16:
1 kb
hg38
31,132,000|
31,132,500|
31,133,000|
31,133,500|
31,134,000|
31,134,500|
31,135,000|
31,135,500|
Reference Assembly Fix Patch Sequence Alignments
Reference Assembly Alternate Haplotype Sequence Alignments
GENCODE V47 (4 items filtered out)
ENSG00000261385
PRSS8
RefSeq genes from NCBI
RefSeq Curated
MANE Select Plus Clinical: Representative transcript from RefSeq & GENCODE
PRSS8
UniProt SwissProt/TrEMBL Protein Annotations
SwissProt Aln.
Signal Peptide
Interest
Extracellular
Transmembrane
Cytoplasmic
Chains
Domains
Disulf. Bonds
AA Modifications
Mutations
Structure
Other Annot.
Repeats
OMIM Allelic Variant Phenotypes
OMIM Alleles
OMIM Gene Phenotypes - Dark Green Can Be Disease-causing
OMIM Genes
Gene Expression in 54 tissues from GTEx RNA-seq of 17382 samples, 948 donors (V8, Aug 2019)
PRSS8
Promoters from EPDnew
EPDnew v6
EPDnew NC v1
CpG Islands (Islands < 300 Bases are Light Green)
ENCODE Candidate Cis-Regulatory Elements (cCREs) combined from all cell types
ENCODE cCREs
H3K27Ac Mark (Often Found Near Regulatory Elements) on 7 cell lines from ENCODE
Layered H3K27Ac
4 _
100 vertebrates Basewise Conservation by PhyloP
Cons 100 Verts
-0.5 _
Multiz Alignments of 100 Vertebrates
Rhesus
Mouse

For the PRSS8 gene, there seems to be increased expression at no specific CpG islands for transmembrane, chains, domains, disulfide bonds, and structure proteins. In correlation with the previous PRSS8 plot, there seems to be greater methylation activity at earlier CpG sites for the metastatic group and greater methylation activity at later CpG sites for the non-metastatic (tumor) group. We can potentially conclude that, on average, earlier CpG sites will have higher transmembrane and chain protein expression with higher methylation expression for metastatic patients and higher chains, domains, disulfide bonds, and structure protein expression for later CpG sites for tumor group patients for the PRSS8 gene. According to Henehan and Benedetto (2019), PRSS8 acts as one of the "most relevant activators" for PAR2, which is associated with its upregulation in malignant melanoma. This research somewhat supports the findings as it is found to have higher transmembrane and chain protein expression in correlation with higher methylation expression in malignant TCGA SKCM patients for the PRSS8 gene.

References

Fujiwara, S., Nagai, H., Jimbo, H., Jimbo, N., Tanaka, T., Inoie, M., & Nishigori, C. (2019). Gene Expression and Methylation Analysis in Melanomas and Melanocytes From the Same Patient: Loss of NPM2 Expression Is a Potential Immunohistochemical Marker for Melanoma. *Frontiers in Oncology*, *8*. https://doi.org/10.3389/fonc.2018.00675

Henehan, M., & Benedetto, A. D. (2019). Update on protease‐activated receptor 2 in cutaneous barrier, differentiation, tumorigenesis and pigmentation, and its role in related dermatologic diseases. *Experimental Dermatology*, *28*(8), 877–885. https://doi.org/10.1111/exd.13936

Tong, X., Qu, X., & Wang, M. (2021). A Four-Gene-Based Prognostic Model Predicts Overall Survival in Patients With Cutaneous Melanoma. *Frontiers in Oncology*, *11*. https://doi.org/10.3389/fonc.2021.639874