# BINF6000 Group Project
# Group C

Rocco Ferguson (46992550), Blake Sanders (45915637), Bhoomika Shashidhara (47395578), Chloe Hu (47812851), Salsabila Luqyana (48492634), Peter Huang (45367115)

## Problem

- Effective treatment of breast cancer through anticancer drugs, such as antineoplastic drugs, requires identification of abnormal/dysregulated genes or mutations present in tumour cells.
    - Drugs target these sites.

- ***TP53*** and ***ERBB2*** - two somatic variants associated with breast cancer phenotypes.
    - Neither found in MCF7 breast cancer cell line; response to specific drugs unknown.

- Must identify molecular mechanisms associated with MCF7 phenotype and cell drug response.

## Aims

1. Identification of any genes that are differentially expressed in MCF7 cells relative to healthy tissue.
    - RNA-seq
2. Identification of any histone modifications that may be present in the genome of MCF7 cells.
    - ChIP-seq
3. In the identified genes, are there associations between these expression levels and histone modifications?
    - Modifications inside/outside promoter regions; are any multi-modified?

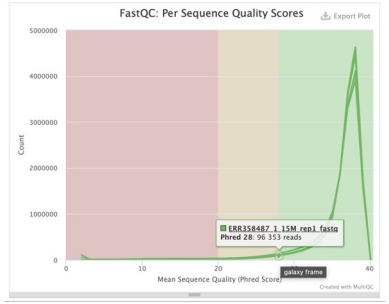
# Aim 1 (1,2,3)

## User Workflow

workflow: https://usegalaxy.org.au/u/rocco_ferguson/h/wk-10-13-group-project

**Quality Control**

FastQC on all sequences




```
=== Summary ===

Total read pairs processed:         15,000,000
Pairs written (passing filters):    15,000,000 (100.0%)

Total basepairs processed: 3,000,000,000 bp
  Read 1: 1,500,000,000 bp
  Read 2: 1,500,000,000 bp
Quality-trimmed:               88,282,455 bp (2.9%)
  Read 1:    30,370,510 bp
  Read 2:    57,911,945 bp
Total written (filtered):  2,911,717,545 bp (97.1%)
  Read 1: 1,469,629,490 bp
  Read 2: 1,442,088,055 bp
```

Filter to 28 because most sequence already above 28

MultiQC

Cutadapt

Kmer length of 7

Trimming and filtering

**Mapping**

HISAT2

Output: BAM file
Sequences and quality information

| QNAME | FLAG | RNAME |
|---|---|---|
| @HD VN:1.0 SO:coordinate | | |
| @SQ SN:chr10 LN:135534747 | | |
| @SQ SN:chr11 LN:135006516 | | |

```
HISAT2 summary stats:
    Total pairs: 15000000
        Aligned concordantly or discordantly 0 time: 557466 (3.72%)
        Aligned concordantly 1 time: 13682951 (91.22%)
        Aligned concordantly >1 times: 683939 (4.56%)
        Aligned discordantly 1 time: 75644 (0.50%)
    Total unpaired reads: 1114932
        Aligned 0 time: 557650 (50.02%)
        Aligned 1 time: 477171 (42.80%)
        Aligned >1 times: 80111 (7.19%)
    Overall alignment rate: 98.14%
```

**Counting**

featureCount

Output: Table with counts
Number of reads mapped to each gene with their ID

| GeneId | HISAT2 on data 60 and data 59: aligned reads (BAM) |
|---|---|
| GeneId | HISAT2 on data 60 and data 59: aligned reads (BAM) |
| ENSG00000223972.4 | 0 |
| ENSG00000227232.4 | 123 |
| ENSG00000243485.2 | 0 |
| ENSG00000237613.2 | 0 |

**Identification of differentially expressed features**

DESeq2

Pupa vs adult stage
Outcome: 13% upregulated, 14% downregulated

*log2FC| > 1*

Filter data

Dataset Peek

| | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| FBgn0000046 | 1196.84650159346 | 2.3726389331293 | 0.0613791898162397 | 38.6553946420245 | 0 | 0 | chr3R | 13425982 | 13428375 | + | protein_coding | Act87E |
| FBgn0036881 | 1018.69651584813 | 4.7663189355152 | 0.126388767984919 | 39.5909869636856 | 0 | 0 | chr3L | 19532742 | 19538289 | + | protein_coding | Cpr76Bd |
| FBgn0031940 | 2867.59813933142 | 3.72474415968364 | 0.0562619818264029 | 66.2035719197156 | 0 | 0 | chr2L | 7743681 | 7744839 | - | protein_coding | CG7214 |
| FBgn0031942 | 1603.80631480283 | -5.75217171347188 | 0.098154517677417 | -58.6032293732652 | 0 | 0 | chr2L | 7752169 | 7753186 | - | protein_coding | CG7203 |
| FBgn0032538 | 1308.42819964878 | 2.49914496193023 | 0.0660287089555218 | 41.6324956077593 | 0 | 0 | chr2L | 13932558 | 13934692 | + | protein_coding | CG16885 |

Annotation

Annotate DESeq2

# Aim 2 (1,2,3)

## Developer workflow

Run FastQC for quality control

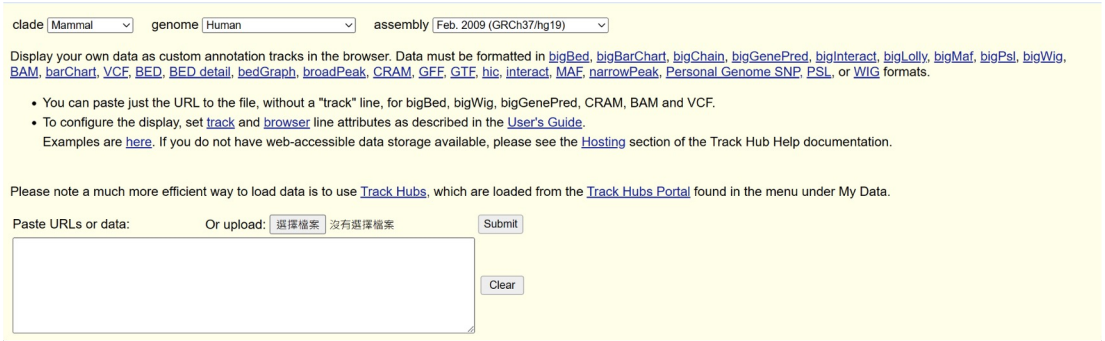
Align reads with Bowtie2

Convert SAM to BAM

Peak calling with MACS2 for the full dataset

Associate the priority genes and peaks with bedtools.

Visualize the peaks by genome browser

# Aim 2 (1,2) Developer workflow

```
#!/bin/bash

fastqc /home/binf6_03/B2.fastq.gz -o /home/binf6_03/ws3data
fastqc /home/binf6_03/control.fastq.gz -o /home/binf6_03/ws3data

bowtie2 -x /home/binf6_03/hg19 -U /home/binf6_03/control.fastq.gz > /home/binf6_03/ws3data/control.sam
bowtie2 -x /home/binf6_03/hg19 -U /home/binf6_03/B2.fastq.gz > /home/binf6_03/ws3data/B2.sam

samtools view -bS /home/binf6_03/ws3data/control.sam > /home/binf6_03/ws3data/control.bam
samtools view -bS /home/binf6_03/ws3data/B2.sam > /home/binf6_03/ws3data/B2.bam

samtools sort /home/binf6_03/ws3data/control.bam -o /home/binf6_03/ws3data/control_sorted.bam
samtools sort /home/binf6_03/ws3data/B2.bam -o /home/binf6_03/ws3data/B2_sorted.bam

macs2 callpeak -t /home/binf6_03/ws3data/B2_sorted.bam -c /home/binf6_03/ws3data/control_sorted.bam -f BAM -g hs -n B2 --outdir /home/binf6_03/ws3output --nomodel --extsize 147

bedtools intersect -a 20gene.bed -b ws3output/B2_summits.bed -wa -wb > genes_with_peaks.bed
```

```
Analysis complete for control.fastq.gz
4865409 reads; of these:
  4865409 (100.00%) were unpaired; of these:
    77692 (1.60%) aligned 0 times
    3134965 (64.43%) aligned exactly 1 time
    1652752 (33.97%) aligned >1 times
98.40% overall alignment rate
8735116 reads; of these:
  8735116 (100.00%) were unpaired; of these:
    204959 (2.35%) aligned 0 times
    6110274 (69.95%) aligned exactly 1 time
    2419883 (27.70%) aligned >1 times
97.65% overall alignment rate
```

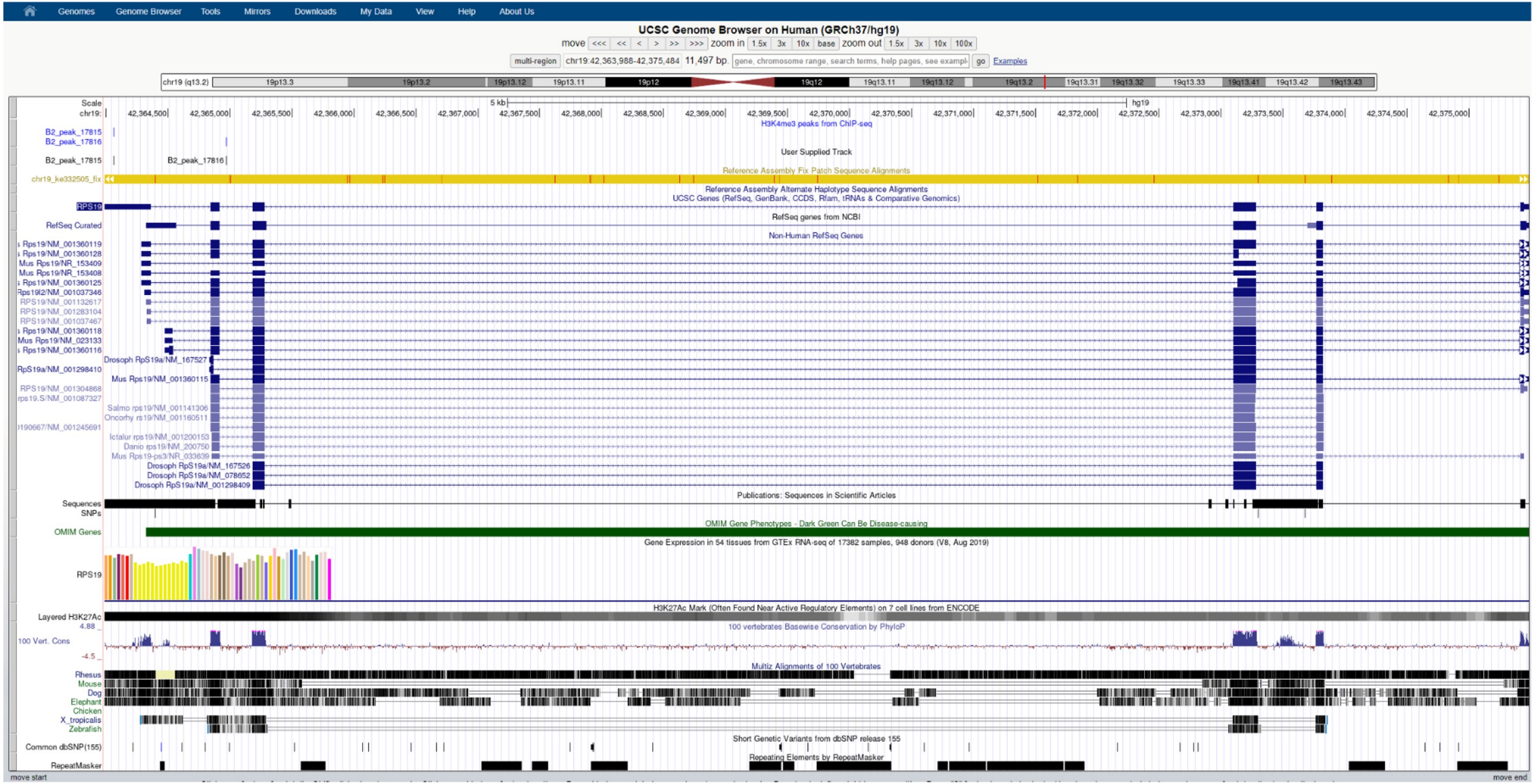
```
chr5    134363424    134369964    PITX1
chr5    176516551    176525126    FGFR4
chr19   42364325     42376993     RPS19
chr6    34505579     34524110     SPDEF
chr9    139956579    139965028    SAPCD2
chr1    214776532    214837914    CENPF
chr7    128097323    128097514    HILPDA
chr2    10262695     10271546     RRM2
chr5    68462837     68474070     CCNB1
chr4    175411328    175444049    HPGD
chr10   62538212     62554610     CDK1
chr7    158424003    158497520    NCAPG2
chr2    74425690     74442424     MTHFD2
chr5    126112315    126172712    LMNB1
chr8    124332091    124408705    ATAD2
chr1    204100190    204121307    ETNK2
chr6    27806440     27806820     HIST1H2BD
chr3    172468475    172539264    ECT2
chr4    120980579    120988013    MAD2L1
chr2    11674242     11782912     GREB1
```

20gene.bed file

```
track name="H3K4me3 Peaks" description="H3K4me3 peaks from ChIP-seq" visibility=full color=0,0,255
chr1    28308    28309    B2_peak_1    5.58909
chr1    28989    28990    B2_peak_2    52.5859
chr1    29544    29545    B2_peak_3    24.1675
chr1    459339   459340   B2_peak_4    5.25652
chr1    459744   459745   B2_peak_5    6.66766
```

B2_summits.bed file

# Aim 2 (4)


Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
View
Help
About Us
UCSC Genome Browser on Human (GRCh37/hg19)
move
zoom in
zoom out
multi-region
chr19:42,363,988-42,375,484
11,497 bp.
gene, chromosome range, search terms, help pages, see exampl
go
Examples
chr19 (q13.2)
Scale
chr19:
5 kb
hg19
H3K4me3 peaks from ChIP-seq
B2_peak_17815
B2_peak_17816
User Supplied Track
Reference Assembly Fix Patch Sequence Alignments
chr19_ke332505_fix
Reference Assembly Alternate Haplotype Sequence Alignments
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
RPS19
RefSeq Curated
RefSeq genes from NCBI
Non-Human RefSeq Genes
Publications: Sequences in Scientific Articles
Sequences
SNPs
OMIM Genes
OMIM Gene Phenotypes - Dark Green Can Be Disease-causing
Gene Expression in 54 tissues from GTEx RNA-seq of 17382 samples, 948 donors (V8, Aug 2019)
H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
Layered H3K27Ac
100 vertebrates Basewise Conservation by PhyloP
100 Vert. Cons
Multiz Alignments of 100 Vertebrates
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Common dbSNP(155)
Short Genetic Variants from dbSNP release 155
RepeatMasker
Repeating Elements by RepeatMasker
move start
move end

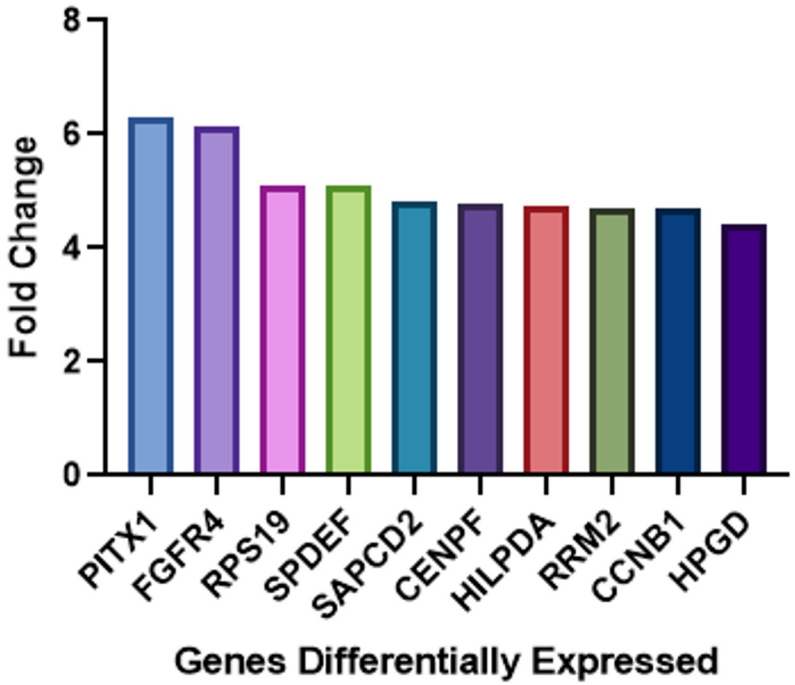
# Aim 1 (2)

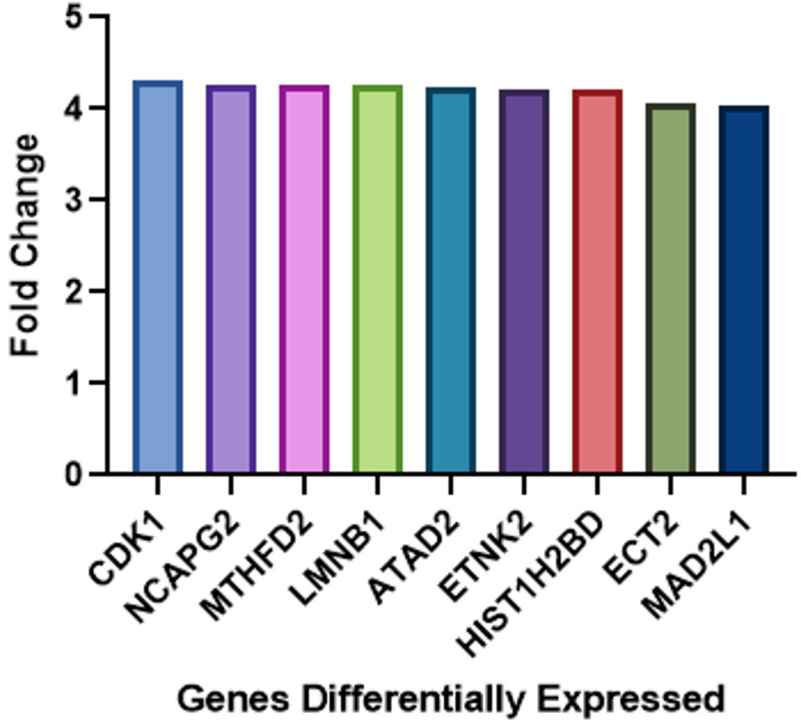
## (Q1) How many genes are significantly over expressed in MCF-7 breast cancer cells relative to normal breast tissue?

| Gene | Fold Change | q-value |
|---|---|---|
| PITX1 | 6.30 | 0 |
| FGFR4 | 6.12 | 0 |
| RPS19 | 5.11 | 0 |
| SPDEF | 5.10 | 0 |
| SAPCD2 | 4.84 | 0 |
| CENPF | 4.77 | 0 |
| HILPDA | 4.76 | 0 |
| RRM2 | 4.70 | 0 |
| CCNB1 | 4.70 | 0 |
| HPGD | 4.41 | 0 |

| Gene | Fold Change | q-value |
|---|---|---|
| CDK1 | 4.39 | 0 |
| NCAPG2 | 4.31 | 0 |
| MTHFD2 | 4.25 | 0 |
| LMNB1 | 4.25 | 0 |
| ATAD2 | 4.25 | 0 |
| ETNK2 | 4.24 | 0 |
| HIST1H2BD | 4.22 | 0 |
| ECT2 | 4.22 | 0 |
| MAD2L1 | 4.06 | 0 |
| GREB1 | 4.04 | 0 |

## Fold Change in Genes Between MCF-7 and Normal Breast Cell Lines



## q-value of Differentially Expressed Genes

- All differentially expressed genes had an adjusted p-value (q-value) of 0
- This suggests none of the genes are differentially expressed by chance
- In reality a q-value of 0 isn’t possible
- However, as the expression differs significantly, the q-value is so close to 0, it is just reported as 0

# Aim 1 (3)

## (Q2) Are Any of The Identified Over-Expressed Genes Implicated in Cancer

| Gene Name | Relation to Cancer |
|---|---|
| PITX1 | PITX1 dysfunction induces oncogenic pathways and cancer development. |
| FGFR4 | Genetic aberrations in FGFR4 are prevalent among breast cancer. |
| RPS19 | RPS19 is upregulated in human breast cancer cells. |
| SPDEF | SPDEF is upregulated in breast cancer and associated with tumor progression and poor prognosis. |
| SAPCD2 | SAPCD2 promotes the progression of breast cancer. |
| CENPF | Overexpression of CENPF predicts shorter survival and higher recurrence in breast cancer. |
| HILPDA | HILPDA promotes cancer progression via hypoxia-dependent and independent pathways. |
| RRM2 | RRM2 acts as a pro-metastatic factor to facilitate breast cancer metastasis. |
| CCNB1 | CCNB1 contributes to lymphovascular invasion in breast cancer. |
| HPGD | HPGD is highly expressed in metastatic and aggressive breast cancer and promotes migration. |

| Gene Name | Relation to Cancer |
|---|---|
| CDK1 | Deregulation of CDK1 has been closely associated with cancer. |
| NCAPG2 | NCAPG2 is involved in cancer, WNT signaling pathway, ubiquitin mediated proteolysis and focal adhesion. |
| MTHFD2 | MTHFD2 is upregulated in cancer, promotes growth and metastasis and correlates with poorer survival. |
| LMNB1 | LMNB1 is abnormally expressed in cancer. |
| ATAD2 | ATAD2 is highly expressed in breast cancer and is associated with poor prognosis. |
| ETNK2 | ETNK2 may promote the development of cancer through the HIPPO and EMT pathways. |
| HIST1H2BD | Mutations in H2B genes, especially HIST1H2BD, have been dominantly found in cancer cells. |
| ECT2 | ECT2 is overexpressed in various cancers and predicts poor prognosis. |
| MAD2L1 | MAD2L1 over-expression plays an important role in tumor proliferation and metastasis. |
| GREB1 | GREB1 is over expressed in malignant tumors and promotes cell proliferation. |

## Is it Likely The Genes Identified Are Linked to Cancer?

- All 20 identified differentially expressed genes had a reported implication in cancer development
- A majority of these had direct links to breast cancer development
- Most implicated genes were involved in cell cycle regulation and cellular growth control
- This suggests the genes identified likely are influential in the development of cancer in MCF7 cells
- From reported functions in literature, it can be determined the genes identified have a probable link to cancer progression

# Aim 2 (2,3)

## (Q3) Do the over expressed genes have a histone modification (H3K4me3) in their promoter region?

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| chr5 | 134363424 | 134369964 | PITX1 | chr5 | 134369143 | 134369144 | B2_peak_11046 | 11.4335 |
| chr19 | 42364325 | 42376993 | RPS19 | chr19 | 42364970 | 42364971 | B2_peak_7291 | 16.8846 |
| chr6 | 34505579 | 34524110 | SPDEF | chr6 | 34523222 | 34523223 | B2_peak_11755 | 33.1861 |
| chr9 | 139956579 | 139965028 | SAPCD2 | chr9 | 139963797 | 139963798 | B2_peak_13975 | 9.27428 |
| chr9 | 139956579 | 139965028 | SAPCD2 | chr9 | 139964109 | 139964110 | B2_peak_13976 | 8.60639 |
| chr7 | 128097323 | 128097514 | HILPDA | chr7 | 128097437 | 128097438 | B2_peak_12777 | 11.4335 |
| chr2 | 10262695 | 10271546 | RRM2 | chr2 | 10262976 | 10262977 | B2_peak_7613 | 10.7283 |
| chr5 | 68462837 | 68474070 | CCNB1 | chr5 | 68463189 | 68463190 | B2_peak_10764 | 16.8846 |
| chr4 | 175411328 | 175444049 | HPGD | chr4 | 175443408 | 175443409 | B2_peak_10578 | 10.7283 |
| chr10 | 62538212 | 62554610 | CDK1 | chr10 | 62538412 | 62538413 | B2_peak_1652 | 3.6006 |
| chr10 | 62538212 | 62554610 | CDK1 | chr10 | 62538756 | 62538757 | B2_peak_1653 | 5.36059 |
| chr7 | 158424003 | 158497520 | NCAPG2 | chr7 | 158497131 | 158497132 | B2_peak_12941 | 38.2787 |
| chr2 | 74425690 | 74442424 | MTHFD2 | chr2 | 74425932 | 74425933 | B2_peak_7887 | 14.4925 |
| chr2 | 74425690 | 74442424 | MTHFD2 | chr2 | 74426724 | 74426725 | B2_peak_7888 | 12.1921 |
| chr5 | 126112315 | 126172712 | LMNB1 | chr5 | 126112966 | 126112967 | B2_peak_10988 | 11.4335 |
| chr5 | 126112315 | 126172712 | LMNB1 | chr5 | 126113650 | 126113651 | B2_peak_10989 | 26.45 |
| chr5 | 126112315 | 126172712 | LMNB1 | chr5 | 126114388 | 126114389 | B2_peak_10990 | 12.1921 |
| chr8 | 124332091 | 124408705 | ATAD2 | chr8 | 124408124 | 124408125 | B2_peak_13291 | 20.1932 |
| chr8 | 124332091 | 124408705 | ATAD2 | chr8 | 124408543 | 124408544 | B2_peak_13292 | 16.0617 |
| chr1 | 204100190 | 204121307 | ETNK2 | chr1 | 204120674 | 204120675 | B2_peak_1249 | 16.0617 |
| chr6 | 27806440 | 27806820 | HIST1H2BD | chr6 | 27806503 | 27806504 | B2_peak_11525 | 12.9624 |
| chr3 | 172468475 | 172539264 | ECT2 | chr3 | 172468853 | 172468854 | B2_peak_9913 | 30.2278 |
| chr4 | 120980579 | 120988013 | MAD2L1 | chr4 | 120987804 | 120987805 | B2_peak_10431 | 8.60639 |

**RPS19**: Peak at 42364970 near the TSS at 42364325 (645 bp upstream).
**HILPDA**: Peak at 128097437 near the TSS at 128097323 (114 bp upstream).
**RRM2**: Peak at 10262976 near the TSS at 10262695 (281 bp upstream).
**CDK1**: Peaks at 62538412 and 62538756 near the TSS at 62538212 (200 bp and 544 bp upstream).
**MTHFD2**: Peak at 74425932 near the TSS at 74425690 (242 bp upstream).
**HIST1H2BD**: Peak at 27806503 near the TSS at 27806440 (63 bp upstream).
**ECT2**: Peak at 172468853 near the TSS at 172468475 (378 bp upstream).
**MAD2L1**: Peak at 120987804 near the TSS at 120988013 (209 bp downstream).

# Aim 3 (1)

## (Q4) Are the Histone Modifications Impacting Gene Expression?

Genes with H3K4me3 in their promoter region

| |
|---|
| RPS19 |
| HILPDA |
| CDK1 |
| MTHFD2 |
| HIST1H2B2 |
| ECT2 |
| MAD2L1 |
| RRM2 |

- The histone modification screened for was H3K4me3
- H3K4me3 is trimethylation of the histone tail at location H3K4 and is known to increase transcription by providing an adapter site for bringing transcription factors to bind with promoters
- H3K4me3 peaks were seen in promoter regions for 8 of the 20 identified genes
- This suggests H3K4me3 modifications may have driven gene expression in these genes, suggesting a mechanism causing their overexpression

## Why Aren’t All Genes Affected by This Modification?

- 12 of the 20 implicated genes did not have H3K4me3 peaks in their promoter regions
- This suggests their overexpression was likely not driven by this modification
- In reality there are many histone modifications that are possible to affect gene expression
- Screening for different histone modifications may give better insight into what epigenetic marks were driving the increased transcription of the othe 12 implicated genes

# Findings and Evidence

**(Q5) Justification of selected histone modification**

- In this analysis, the histone modification screened was H3K4me3 as it is known to be associated with active transcription.
- It acts like a necessary indicator of transcriptional activity due to their modification said to mark promoters of actively transcribed genes.

Further justification of the selected histone modification was considered based on:

- **Biological Relevance,**
- **Technical Feasibility and,**
- **Data Availability.**

**Would including multiple modifications be beneficial?**

- ➢ It is beneficial to include Multiple histone modifications in ChIP-seq sequencing in Breast Cancer.
- ➢ Including modifications like H3K36me3 and H3K27ac can provide information on the regulatory environment of a gene.
- ➢ Helps in distinguishing between different regulatory mechanisms and stages of transcription.
- ➢ Gives an insight in Epigenetic Interactions by highlighting regulatory elements and their role in disease/ cellular context.
- ➢ Can indicate dynamic regulatory mechanisms by identifying important epigenetic switches which affect gene expression.

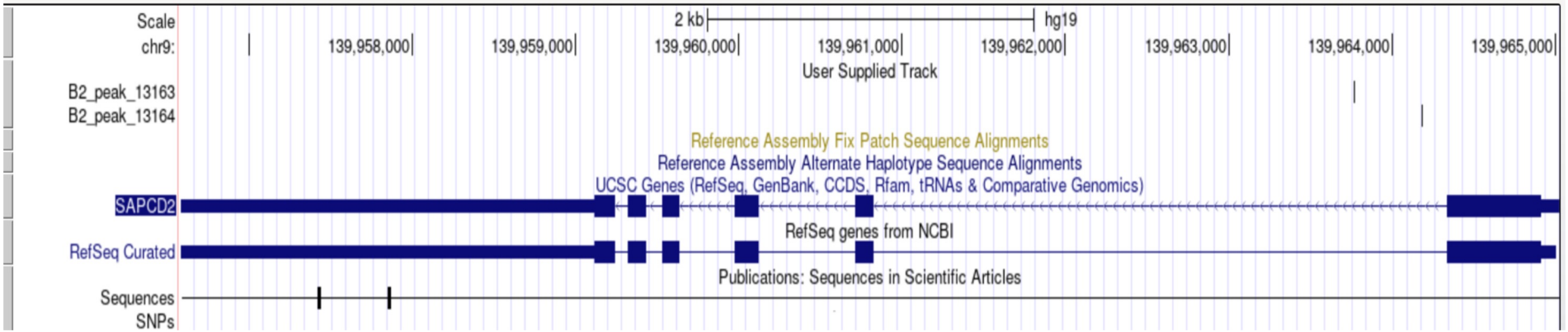
# Aim 3 (2)

## (Q6) Are there histone modifications outside of promoter regions? What role would these play?

There are 12 peaks outside of gene-promoter regions

Functions of these peaks could be for:

- Gene transcription
- DNA replication, recombination and damage repair.



## (Q7) Is using only over expressed genes better than including all differentially expressed genes?

- One limitation of using only the set of priority genes rather than all differentially expressed genes in the analysis is the **biases** incorporated during the computation of dispersion results where **high false positive and false negative rates** occur.

- The generally low sample sizes of RNA Seq datasets may result in the **unreliable gene dispersion estimation**.

- Other than that, the uncorrected P-values as a result of the biases, may overlap even between highly similar datasets which later will reduce the ability to compare across the RNA seq studies (Mukamel, 2021).

- Reducing the sequencing in RNA Seq analysis can be a problem since the gene expression trends identified with a nominal P-value will not necessarily hold with higher sample size and more robust statistic (Vannan et al, 2023)

- Another reason is that using only the set of priority genes does not measure all genes accurately especially when the genes of interest are unknown which results in not knowing which genes are affected. The set of priority genes may not be the cause of the disease of interest.

**Q8: Name one advantage of using biological replicates in ChIP-seq and RNA-seq analyses.**

- Allows researchers to measure and account for random/natural variation between different, distinct biological samples.

- Increases statistical power and reliability of results; differentiation between random noise and actual biological signals.

- Critical role in ensuring observed changes in protein-DNA interactions and/or gene expression are representative of biological conditions being studied, not phenomena such as sample idiosyncrasies or other artifacts.

- Helps ensure results are not specific to particular condition/sample, instead across differing populations/conditions; supports findings across a broader biological context.

# Reference List


Aljohani AI, Toss MS, Green AR, Rakha EA. The clinical significance of cyclin B1 (CCNB1) in invasive breast cancer with emphasis on its contribution to lymphovascular invasion development. Breast Cancer Research and Treatment. 2023;198(3):423-35.

Arimura Y, Ikura M, Fujita R, Noda M, Kobayashi W, Horikoshi N, et al. Cancer-associated mutations of histones H2B, H3.1 and H2A.Z.1 affect the structure and stability of the nucleosome. Nucleic Acids Res. 2018;46(19):10007-18.

Baker AL, Du L. The Function and Regulation of SAPCD2 in Physiological and Oncogenic Processes. J Cancer. 2022;13(7):2374-87.

Fu J, Zhang J, Chen X, Liu Z, Yang X, He Z, et al. ATPase family AAA domain-containing protein 2 (ATAD2): From an epigenetic modulator to cancer therapeutic target. Theranostics. 2023;13(2):787-809.

Lehtinen L, Vainio P, Wikman H, Reemts J, Hilvo M, Issa R, et al. 15-Hydroxyprostaglandin dehydrogenase associates with poor prognosis in breast cancer, induces epithelial–mesenchymal transition, and promotes cell migration in cultured breast cancer cells. The Journal of Pathology. 2012;226(4):674-86.

Li J, Wan X, Xie D, Yuan H, Pei Q, Luo Y, et al. SPDEF enhances cancer stem cell-like properties and tumorigenesis through directly promoting GALNT7 transcription in luminal breast cancer. Cell Death & Disease. 2023;14(8):569.

Li Q, Tong D, Jing X, Ma P, Li F, Jiang Q, et al. MAD2L1 is transcriptionally regulated by TEAD4 and promotes cell proliferation and migration in colorectal cancer. Cancer Gene Therapy. 2023;30(5):727-37.

Liu C, Zhou X, Zeng H, Wu D, Liu L. HILPDA Is a Prognostic Biomarker and Correlates With Macrophage Infiltration in Pan-Cancer. Front Oncol. 2021;11:597860.

Liu Y, Wang C, Li J, Zhu J, Zhao C, Xu H. Novel Regulatory Factors and Small-Molecule Inhibitors of FGFR4 in Cancer. Front Pharmacol. 2021;12:633453.

Markiewski MM, Vadrevu SK, Sharma SK, Chintala NK, Ghouse S, Cho J-H, et al. The Ribosomal Protein S19 Suppresses Antitumor Immune Responses via the Complement C5a Receptor 1. The Journal of Immunology. 2017;198(7):2989-99.

Massacci G, Perfetto L, Sacco F. The Cyclin-dependent kinase 1: more than a cell cycle regulator. British Journal of Cancer. 2023;129(11):1707-16.

Mukamel EA. Multiple Comparisons and Inappropriate Statistical Testing Lead to Spurious Sex Differences in Gene Expression. Biological Psychiatry. 2022;91(1):e1-e2.

Ohira T, Nakagawa S, Takeshita J, Aburatani H, Kugoh H. PITX1 inhibits the growth and proliferation of melanoma cells through regulation of SOX family genes. Scientific Reports. 2021;11(1):18405.

Qin H, Lu Y, Du L, Shi J, Yin H, Jiang B, et al. Pan-cancer analysis identifies LMNB1 as a target to redress Th1/Th2 imbalance and enhance PARP inhibitor response in human cancers. Cancer Cell International. 2022;22(1):101.

# Reference List


Ramos L, Henriksson M, Helleday T, Green AC. Targeting MTHFD2 to Exploit Cancer-Specific Metabolism and the DNA Damage Response. Cancer Research. 2024;84(1):9-16.

Ren W, Yang S, Chen X, Guo J, Zhao H, Yang R, et al. NCAPG2 Is a Novel Prognostic Biomarker and Promotes Cancer Stem Cell Maintenance in Low-Grade Glioma. Front Oncol. 2022;12:918606.

Shinzawa K, Matsumoto S, Sada R, Harada A, Saitoh K, Kato K, et al. GREB1 isoform 4 is specifically transcribed by MITF and required for melanoma proliferation. Oncogene. 2023;42(42):3142-56.

Vannan A, Dell'Orco M, Perrone-Bizzozero NI, Neisewander JL, Wilson MA. An approach for prioritizing candidate genes from RNA-seq using preclinical cocaine self-administration datasets as a test case. G3 Genes|Genomes|Genetics. 2023;13(10).

Yi M, Zhang D, Song B, Zhao B, Niu M, Wu Y, et al. Increased expression of ECT2 predicts the poor prognosis of breast cancer patients. Experimental Hematology & Oncology. 2022;11(1):107.

Zhang J, Wang Z, Liu Z, Chen Z, Jiang J, Ji Y, et al. CENPF promotes the proliferation of renal cell carcinoma in vitro. Translational Andrology and Urology. 2023;12(2):320-9.

Zheng D, Jin L, Chen B, Qi Y, Bhandari A, Wen J, et al. The ETNK2 gene promotes progression of papillary thyroid carcinoma through the HIPPO pathway. J Cancer. 2022;13(2):508-16.

Zhuang S, Li L, Zang Y, Li G, Wang F. RRM2 elicits the metastatic potential of breast cancer cells by regulating cell invasion, migration and VEGF expression via the PI3K/AKT signaling. Oncol Lett. 2020;19(4):3349-55.