RNA-seq analysis of glioma combination drug therapies

Marina Vallejo[1*]***, Maria Artigues***[1**]***, Pau Torren***[1***] ***and Sara Vega***[1****]

[1]Universitat Pompeu Fabra

[*]marina.vallejo01@estudiant.upf.edu (mailto:marina.vallejo01@estudiant.upf.edu)
[**]maria.artigues01@estudiant.upf.edu (mailto:maria.artigues01@estudiant.upf.edu)
[***]pau.torren01@estudiant.upf.edu (mailto:pau.torren01@estudiant.upf.edu)
[****]sara.vega02@estudiant.upf.edu (mailto:sara.vega02@estudiant.upf.edu)

2022-06-15

1 Introduction

Diffuse Midline Gliomas (DMGs) (https://www.cancer.gov/rare-brain-spine-tumor/tumors/diffuse-midline-gliomas) are a type of lethal tumours that affect the glial cells. In this study, they focus on a subtype of DMGs, **Diffuse Intrinsic Pontine Gliomas (DIPG)**. They tend to have quick growth so children affected are usually diagnosed very soon. Despite the early diagnosis, the median overall survival is only 9-10 months, as nowadays treatment is limited to radiotherapy.

In the recent years, **Panobinostat** has been reported as a promising treatment drug for this disease. Nevertheless, in DIPG preclinical studies with this drug, resistances have arised. The aim of this study is to characterize combinational drug therapy and find new vulnerabilities in these kind of cancers. For this, the authors perform a **High-throughput drug screening** and, later on, a **RNA-Seq experiment** with the most promising drug combinations (Lin et al. 2019). In this context, the **transcriptional data** allows to characterize the metabolic state of the DIPG models under different drugs and identify the underlying metabolic effects.

2 Quality assessment

Before starting with the analysis, it is important to perform a **Quality assessment**. This collection of steps will allow us to understand our data and get rid of those samples that do not meet the quality standards.

2.1 Data import and cleaning

First of all, we will import the raw table of counts available in **.rds** (https://functionalgenomics.upf.edu/courses/IEO/projects/datasets/GSE123278.rds) format.

```
se <- readRDS(file.path(system.file("extdata",
                                    package="IEOprojectGlioma"),
                        "GSE123278.rds"))
se
class: RangedSummarizedExperiment
dim: 25120 17
metadata(4): experimentData annotation ensemblVersion urlProcessedData
assays(1): counts
rownames(25120): 1 10 ... 9994 9997
rowData names(5): gene_id gene_biotype description gene_id_version
  symbol
colnames(17): SRR8272120 SRR9652348 ... SRR9652362 SRR9652363
colData names(37): title geo_accession ... agent:ch1 cell type:ch1
```

Data is stored in a **RangedSummarizedExperiment** class. It is a sort of matrix container where the columns are samples, and rows are features of interest. We have a total of **25120 genes** and **17 samples**.

Now we are going to explore **rowData**:

```
cat("Dimensions: ", dim(rowData(se)),"\n","\n") # check dimensions
Dimensions:  25120 5


head(rowData(se)) # get first rows
DataFrame with 6 rows and 5 columns
                  gene_id   gene_biotype            description
              <character>    <character>            <character>
1         ENSG00000121410 protein_coding alpha-1-B glycoprote..
10        ENSG00000156006 protein_coding N-acetyltransferase ..
100       ENSG00000196839 protein_coding adenosine deaminase ..
1000      ENSG00000170558 protein_coding cadherin 2 [Source:H..
10000     ENSG00000117020 protein_coding AKT serine/threonine..
100008586 ENSG00000236362 protein_coding G antigen 12F [Sourc..
             gene_id_version      symbol
                 <character> <character>
1         ENSG00000121410.11        A1BG
10         ENSG00000156006.4        NAT2
100       ENSG00000196839.12         ADA
1000       ENSG00000170558.8        CDH2
10000     ENSG00000117020.16        AKT3
100008586  ENSG00000236362.8     GAGE12F
```

Note that **gene id** is encoded using Stable IDs, which follow the pattern: **ENS/species prefix/feature type prefix/a unique eleven digit number**. They also provide a short gene description, among others.

Now we are going to explore **colData**:

```
cat("Dimensions: ", dim(colData(se)),"\n","\n")
Dimensions:  17 37

head(colData(se), n=5)
DataFrame with 5 rows and 37 columns
                   title geo_accession               status submission_date
             <character>   <character>          <character>     <character>
SRR8272120    DIPG13-Mar2    GSM3498943 Public on Nov 29 2019     Dec 03 2018
SRR9652348   DIPG6-Combo1    GSM3930449 Public on Nov 29 2019     Jul 08 2019
SRR9652349   DIPG6-Combo2    GSM3930450 Public on Nov 29 2019     Jul 08 2019
SRR9652350 DIPG6-Control1    GSM3930451 Public on Nov 29 2019     Jul 08 2019
SRR9652351 DIPG6-Control2    GSM3930452 Public on Nov 29 2019     Jul 08 2019
           last_update_date        type channel_count        source_name_ch1
                <character> <character>   <character>           <character>
SRR8272120      Nov 29 2019         SRA             1 patient-derived cell..
SRR9652348      Nov 29 2019         SRA             1 patient-derived cell..
SRR9652349      Nov 29 2019         SRA             1 patient-derived cell..
SRR9652350      Nov 29 2019         SRA             1 patient-derived cell..
SRR9652351      Nov 29 2019         SRA             1 patient-derived cell..
           organism_ch1  characteristics_ch1  characteristics_ch1.1
            <character>          <character>           <character>
SRR8272120 Homo sapiens    cell type: DIPG13  agent: 20nM Marizomib
SRR9652348 Homo sapiens cell type: SU-DIPG-6 agent: Panobinostat ..
SRR9652349 Homo sapiens cell type: SU-DIPG-6 agent: Panobinostat ..
SRR9652350 Homo sapiens cell type: SU-DIPG-6            agent: DMSO
SRR9652351 Homo sapiens cell type: SU-DIPG-6            agent: DMSO
           treatment_protocol_ch1    growth_protocol_ch1 molecule_ch1
                      <character>           <character>  <character>
SRR8272120 Cells were plated at.. Cells are grown in T..    polyA RNA
SRR9652348 Cells were plated at.. Cells are grown in T..    polyA RNA
SRR9652349 Cells were plated at.. Cells are grown in T..    polyA RNA
SRR9652350 Cells were plated at.. Cells are grown in T..    polyA RNA
SRR9652351 Cells were plated at.. Cells are grown in T..    polyA RNA
             extract_protocol_ch1 extract_protocol_ch1.1 extract_protocol_ch1.2
                      <character>           <character>           <character>
SRR8272120 RNA-seq samples were.. After extraction fro.. Libraries were seque..
SRR9652348 RNA-seq samples were.. After extraction fro.. Libraries were seque..
SRR9652349 RNA-seq samples were.. After extraction fro.. Libraries were seque..
SRR9652350 RNA-seq samples were.. After extraction fro.. Libraries were seque..
SRR9652351 RNA-seq samples were.. After extraction fro.. Libraries were seque..
             taxid_ch1      description description.1        data_processing
           <character>      <character>   <character>           <character>
SRR8272120        9606 poly(A)+ RNA-seq   DIPG13-Mar2 Mapping: RNA-seq wit..
SRR9652348        9606                               Mapping: RNA-seq wit..
SRR9652349        9606                               Mapping: RNA-seq wit..
SRR9652350        9606                               Mapping: RNA-seq wit..
SRR9652351        9606                               Mapping: RNA-seq wit..
                data_processing.1      data_processing.2      data_processing.3
                      <character>           <character>           <character>
SRR8272120 Bed files generated .. bedGraph files were .. Gene expression: For..
SRR9652348 Bed files generated .. bedGraph files were .. Gene expression: For..
SRR9652349 Bed files generated .. bedGraph files were .. Gene expression: For..
SRR9652350 Bed files generated .. bedGraph files were .. Gene expression: For..
SRR9652351 Bed files generated .. bedGraph files were .. Gene expression: For..
            data_processing.4      data_processing.5 platform_id data_row_count
                  <character>           <character> <character>    <character>
SRR8272120 Genome_build: hg19 Supplementary_files_..    GPL18573              0
SRR9652348 Genome_build: hg19                           GPL18573              0
SRR9652349 Genome_build: hg19                           GPL18573              0
SRR9652350 Genome_build: hg19                           GPL18573              0
SRR9652351 Genome_build: hg19                           GPL18573              0
               instrument_model library_selection library_source
                     <character>       <character>    <character>
SRR8272120 Illumina NextSeq 500              cDNA transcriptomic
SRR9652348 Illumina NextSeq 500              cDNA transcriptomic
SRR9652349 Illumina NextSeq 500              cDNA transcriptomic
SRR9652350 Illumina NextSeq 500              cDNA transcriptomic
SRR9652351 Illumina NextSeq 500              cDNA transcriptomic
           library_strategy              relation             relation.1
                <character>           <character>           <character>
SRR8272120          RNA-Seq BioSample: https://w.. SRA: https://www.ncb..
SRR9652348          RNA-Seq BioSample: https://w.. SRA: https://www.ncb..
SRR9652349          RNA-Seq BioSample: https://w.. SRA: https://www.ncb..
SRR9652350          RNA-Seq BioSample: https://w.. SRA: https://www.ncb..
SRR9652351          RNA-Seq BioSample: https://w.. SRA: https://www.ncb..
           supplementary_file_1             agent:ch1 cell type:ch1
                     <character>           <character>   <character>
SRR8272120                 NONE         20nM Marizomib        DIPG13
SRR9652348                 NONE Panobinostat and Mar..     SU-DIPG-6
SRR9652349                 NONE Panobinostat and Mar..     SU-DIPG-6
SRR9652350                 NONE                   DMSO     SU-DIPG-6
SRR9652351                 NONE                   DMSO     SU-DIPG-6
```

colData contains phenotypic data. At the beginning we can check that our data set has **37 phenotypic variables** (columns).

Note that there's the column **geo_accession**, which contains **GSM** identifiers. These identifiers can be used in order to check if there are any technical replicates, as they define individual samples.

Now we will check if there are **technical replicates**:

```
length(unique(se$geo_accession))
[1] 17
table(lengths(split(colnames(se), se$geo_accession)))

 1
17
```

Our data set doesn't have any technical replicates, there are no repetitions in the **geo_accession** variable.

Convert to DGEList and check data dimensions:

```
dge <- DGEList(counts=assays(se)$counts, genes=rowData(se))

cat("Dimensions: ",dim(dge))
Dimensions:  25120 17
```

As we are working with **RNA expression** data, we can calculate the expression units. They can be seen as a digital measure to quantify the abundance of transcripts. **log2 counts per million reads mapped (CPM)**, consists in counting the sequenced fragments scaled by the total number of reads and multiplied by a million.

Calculate **log2 CPM**:

```
assays(se)$logCPM <- cpm(dge, log=TRUE) # logical, if TRUE then log2 values are returned.
assays(se)$logCPM[1:5, 1:5] # check the 5 top rows
      SRR8272120 SRR9652348 SRR9652349 SRR9652350 SRR9652351
1       1.683627  0.1645094  0.1403699  0.6753762  0.5097146
10     -3.337912 -3.3379119 -3.3379119 -3.3379119 -3.3379119
100     4.257046  3.3465005  3.8285967  2.3306139  2.4325264
1000    9.179276  7.2764631  7.2811980  7.9228441  7.8941422
10000   6.765192  6.0183313  5.9079829  6.6407650  6.5136130
```

Check other categorical variables that contain further information about the experiment:

- `cell type:ch1` → cell type
- `agent:ch1` → treatment received

Create a table containing these 2 variables:

Table 1: **Number of samples for each cell type and treatment.**Columns show different cell lines, rows show different treatments.

	DIPG13	QCTB-R059	SU-DIPG-6
20nM Marizomib	1	0	0
DMSO	0	2	2
Marizomib	0	2	2
Panobinostat	0	2	2
Panobinostat and Marizomib	0	2	2

In the Table 1 we have **treatment** as rows and **cell line** as columns, with 5 and 3 levels each one.

Further information about the **cell lines** (sex/diagnosis age/survival/tumor type/tissue obtention/prior therapy):

- **DIPG13**: 6 years / F / 4 months / DIPG, WHO grade IV / postmortem autopsy / XRT
- **QCTB-R059**: 10 years / F / 1 month / Pediatric GBM, WHO grade IV / surgical resection / None
- **SU-DIPG-6**: 7 years / F / 6 months / DIPG, WHO grade III / postmortem autopsy / XRT, vorinostat

Further information about the **treatments**:

- **DMSO (dimethyl sulfoxide)**: cells treated with this compound are considered the control of the experiment.
- **Panobinostat**: the plated cells were treated with 50nM of panobinostat. This drug is a Histone deacetylase (HDAC) inhibitor.
- **Marizomib**: the plated cells were treated with 20nM of marizomib. This drug is a proteasome inhibitor.
- **Panobinostat and Marizomib**: also called “combo”, the plated cells were treated with both drugs (50nM of panobinostat and 20nM of marizomib).

If we take a look at the table, we can appreciate that not all the cell lines studied have the same number of replicates nor treatments, as the cell type **DIPG13** only has one sample treated with **Marizomib 20nM**. We should consider whether it is appropriate to remove this sample, as we won’t be able to identify changes in the expression due to the lack of information.

If we want to know about the experimental protocols used to treat the different cell lines, we can access the variables associated with technical factors:

```
se$treatment_protocol_ch1[1]
[1] "Cells were plated at 200k cells/mL and treated with DMSO control, 50nM Panobinostat, 20nM Marizomib, 50nM Panobin
se$growth_protocol_ch1[1]
[1] "Cells are grown in Tumor Stem Media (TSM), consisting of Neurobasal (-A), DMEM F12, B27 (-A), bFGF, EGF, PDGF-AB,
se$extract_protocol_ch1[1]
[1] "RNA-seq samples were pelleted and lysed in Trizol reagent then subsequently precipitated and processed through a
```

Table 2: **Phenotypic variables.** Each row shows a sample.

Identifier	Cell line	Treatment	Group
SRR8272120	DIPG13	20nM Marizomib	DIPG13-Mar2
SRR9652348	SU-DIPG-6	Panobinostat and Marizomib	DIPG6-Combo1
SRR9652349	SU-DIPG-6	Panobinostat and Marizomib	DIPG6-Combo2
SRR9652350	SU-DIPG-6	DMSO	DIPG6-Control1
SRR9652351	SU-DIPG-6	DMSO	DIPG6-Control2
SRR9652352	SU-DIPG-6	Marizomib	DIPG6-Mar1
SRR9652353	SU-DIPG-6	Marizomib	DIPG6-Mar2
SRR9652354	SU-DIPG-6	Panobinostat	DIPG6-Pano1
SRR9652355	SU-DIPG-6	Panobinostat	DIPG6-Pano2
SRR9652356	QCTB-R059	Panobinostat and Marizomib	R059-Combo1
SRR9652357	QCTB-R059	Panobinostat and Marizomib	R059-Combo2
SRR9652358	QCTB-R059	DMSO	R059-Control1
SRR9652359	QCTB-R059	DMSO	R059-Control2
SRR9652360	QCTB-R059	Marizomib	R059-Mar1
SRR9652361	QCTB-R059	Marizomib	R059-Mar2
SRR9652362	QCTB-R059	Panobinostat	R059-Pano1
SRR9652363	QCTB-R059	Panobinostat	R059-Pano2

Later on, we will use the group identifier (last column of table 2) to group together the samples with same cell line and treatment by deleting the number at the end.

2.2 Sequencing depth

Figure 1 below shows the library size per sample in increasing order. They are coloured by cell type, as shown in the legend.

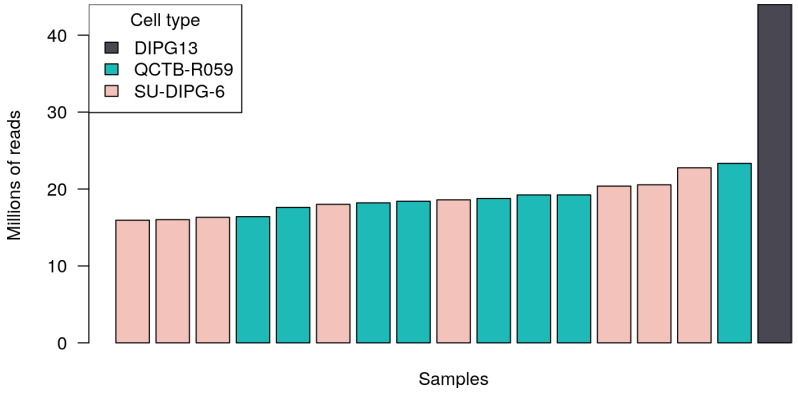


Figure 1: **Library sizes**
Samples are shown in increasing library size order and colored by cell type.

We can see a big difference in the sequencing depth of the cell line **DIPG-13** in comparison with the other samples, having almost twice as many reads as the rest.

For this reason, and since, as mentioned before, we only have one sample for this cell line with only one treatment, we decide to remove the **DIPG-13** sample from the analysis.

```
## remove DIPG13 sample
mask <- se$title != "DIPG13-Mar2"
se_wo13 <- se[, mask]
dge_wo13 <- dge[, mask]
print(unique(se_wo13$`cell type:ch1`)) ## we can see that DIPG13 is not in the new dataset
[1] "SU-DIPG-6" "QCTB-R059"
print(dim(assay(se_wo13))) ## we have one less column
[1] 25120    16
print(dim(dge_wo13))
[1] 25120    16
```

2.3 Distribution of expression levels among samples

The Figure 2 below shows the distribution of expression levels per sample, as logarithmic CPM.

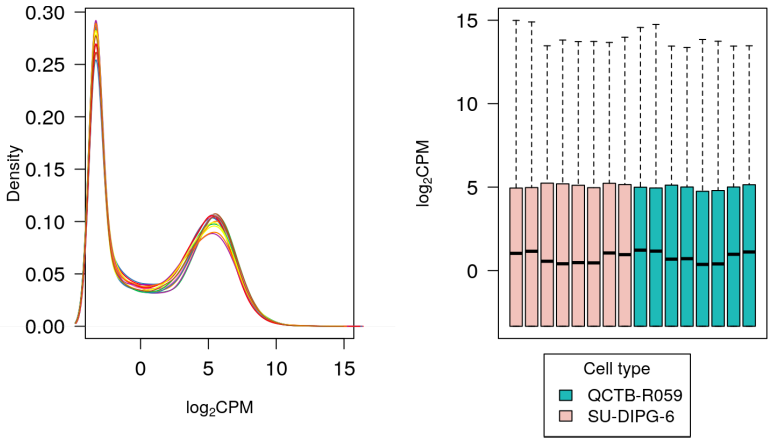


Figure 2: **Distribution of expression levels among samples**

There are no major differences in the expression distribution across the samples.

2.4 Distribution of expression levels among genes

Figure 3 below shows the distribution of average expression among genes.

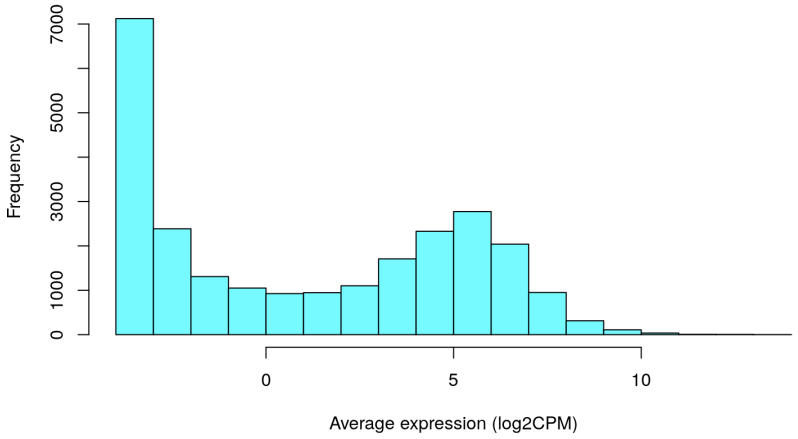


Figure 3: **Distribution of average expression levels among genes**

We see many lowly-expressed genes that need to be filtered in the next step.

2.5 Filtering of lowly-expressed genes

We will filter the lowly-expressed genes to avoid expression-dependent biases in the samples. To achieve this, we have decided to apply a cut-off value of 1 in all samples.

```
mask <- rowMeans(assays(se_wo13)$logCPM) > 1
se.filtered <- se_wo13[mask, ]
dge.filtered <- dge_wo13[mask, ]
dim(se.filtered)
[1] 12326    16
```

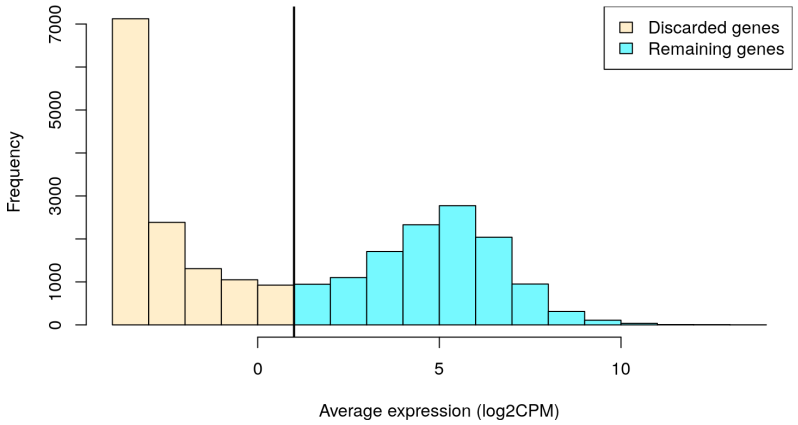


Figure 4: **Distribution of average expression of genes after applying a cut-off**
The cut-off is represented by the vertical line.

After the filtering we are left with 12326 genes.

2.6 Normalization

Here we are going to take the data previously filtered and calculate the **normalization factors** in order to scale the raw library sizes.

It can be done using the function `calcNormFactors` from the `edgeR` package. The default method uses the **Trimmed Mean of M-values (TMM)** between all the sample pairs.

```
dge.filtered <- calcNormFactors(dge.filtered)
```

Now we are going to replace the previously calculated **log2 CPM** for the normalized values:

```
assays(se.filtered)$logCPM <- cpm(dge.filtered, log=TRUE,
                                  normalized.lib.sizes=TRUE)
```

2.7 MA-plots

MA plots are a very useful tool to assess the normalization results. They can be used to find differences between measurements taken in two different samples.

The figure 5 below shows the MA-plots for each sample, where **M** is the difference between the log intensity and the average and **A** is the average of the log intensity.

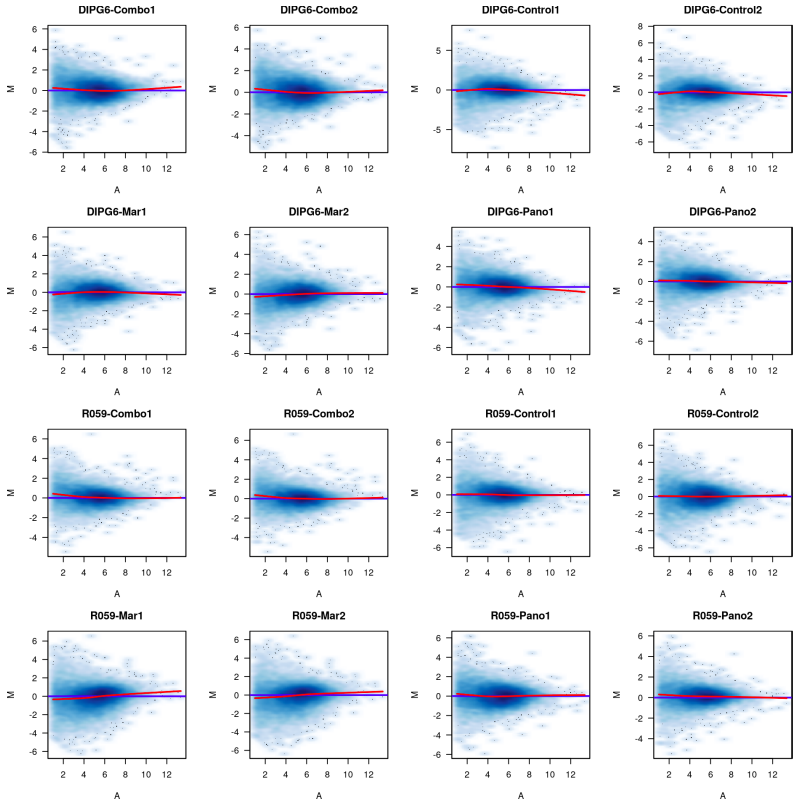


Figure 5: **MA plots of filtered and normalized expression values for each sample**

Here we can see that by filtering has been well performed as we don't appreciate substantial bias in the expression values.

2.8 Experimental design and batch identification

To assess whether there is a possible **batch effect** that should be addressed, we will take a look into the details of the experimental design:

Table 3: **Number of samples for each cell type and treatment.** Columns show sample cell line and rows show sample treatment.

	QCTB-R059	SU-DIPG-6
DMSO	2	2
Marizomib	2	2
Panobinostat	2	2
Panobinostat and Marizomib	2	2

As we have commented before, the number of samples studied is the same for each condition.

Table 4: **Number of samples for each cell type and experimental protocol.** Columns show sample cell line and rows show experimental protocol

	QCTB-R059	SU-DIPG-6
RNA-seq samples were pelleted and lysed in Trizol reagent then subsequently precipitated and processed through a Zymo RNA Clean and Concentrator-5 column	8	8

If we retrieve the experimental protocols by which the data was obtained, we also establish that all the cells underwent the same conditions and processes.

In order to assess in a visual way if there is **batch effect** we can compute:

- Hierarchical clustering
- MDS Plot

Hierarchical clustering of samples:

We will need to group the samples by their treatment and cell type so we have decided to create a new variable called groupname that contains this information:

```
## group by cell type and treatment
se.filtered$groupname <- factor(unname(sapply(se.filtered$title, function(x) gsub("-", ".", substring(x, 1, nchar(x)-1
se.filtered$groupname
 [1] DIPG6.Combo   DIPG6.Combo   DIPG6.Control DIPG6.Control DIPG6.Mar
 [6] DIPG6.Mar     DIPG6.Pano    DIPG6.Pano    R059.Combo    R059.Combo
[11] R059.Control  R059.Control  R059.Mar      R059.Mar      R059.Pano
[16] R059.Pano
8 Levels: DIPG6.Combo DIPG6.Control DIPG6.Mar DIPG6.Pano ... R059.Pano
table(se.filtered$groupname)

  DIPG6.Combo DIPG6.Control     DIPG6.Mar    DIPG6.Pano    R059.Combo
            2             2             2             2             2
 R059.Control      R059.Mar     R059.Pano
            2             2             2
```

After this, we can use this new label to identify the samples in further analysis, such as the hierarchical clustering:

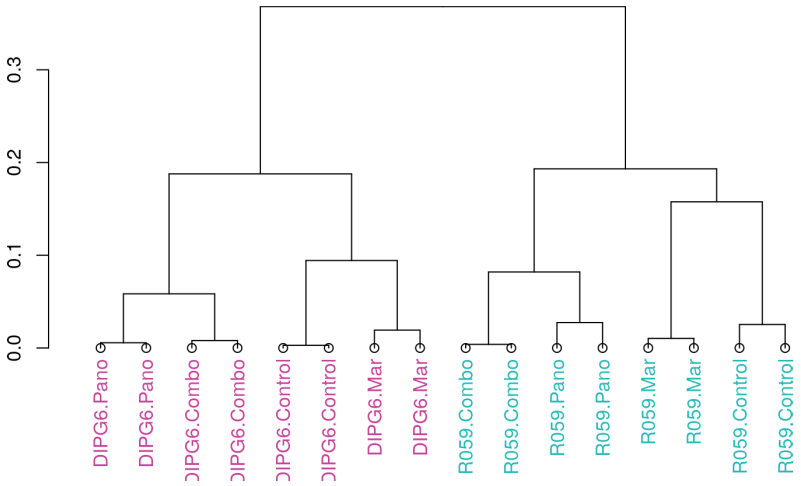


Figure 6: **Hierarchical clustering of the samples**
Labels correspond to sample group, while colors indicate sample cell type.

MDS plot:

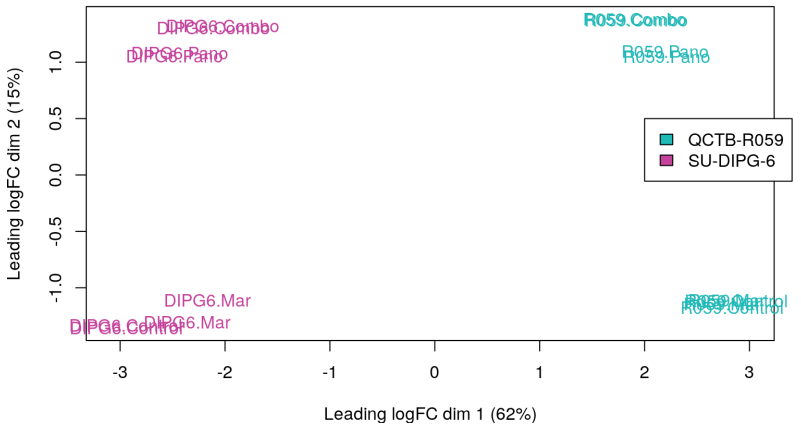


Figure 7: **MDS plot of the samples**
Labels indicate sample group, while colors indicate sample cell type.

As we can see in the figure 6 and 7, the cell lines (**QCTB-R059** and **SU-DIPG-6**) are clearly differentiated in two clusters according to the 1st dimension. Moreover, we can see how in both cell lines (**QCTB-R059** and **SU-DIPG-6**) there is a clear separation between treatments, where **Control** and **Marizomib** treatments group together forming a cluster and **Panobinostat** and **Panobinostat & Marizomib (Combo)** form another cluster.

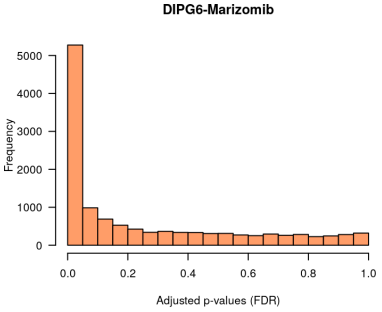
3 Differential expression

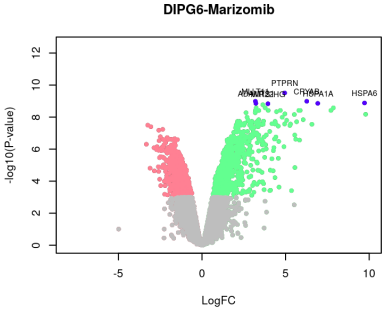
To further explore the effect of the different treatments in the cell lines, we are going to identify the **genes differential expressed** between the treatment cells in respect to the control samples. To perform this study we are applying a **factorial design approach**, as we want to make multiple pairwise comparisons between the same cell type.

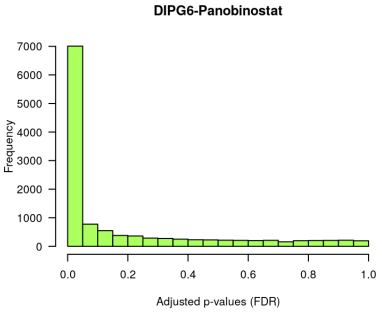
First of all, we need to create a model matrix taking into account both cell type and treatment as covariates, already encoded in the previously created variable 'groupname'. Then, a model is fitted with the function 'glmQLFit()', from 'EdgeR' to prepare the model to conduct genewise statistical tests for a given coefficient or contrast. The last step is to create a contrast matrix, where we specify all the comparisons between groups we want to perform.

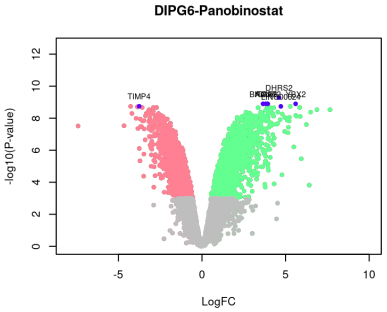
```
co <- se.filtered$groupname
mod <- model.matrix(~ 0 + co, colData(se.filtered))
dge.filtered <- estimateDisp(dge.filtered, mod)
fit <- glmQLFit(dge.filtered, mod)

cont.matrix <- makeContrasts(DIPG6.Mar=coDIPG6.Mar-coDIPG6.Control,
                             DIPG6.Pano=coDIPG6.Pano-coDIPG6.Control,
                             DIPG6.Combo=coDIPG6.Combo-coDIPG6.Control,
                             R059.Mar=coR059.Mar-coR059.Control,
                             R059.Pano=coR059.Pano-coR059.Control,
                             R059.Combo=coR059.Combo-coR059.Control,
                             levels=mod)
```

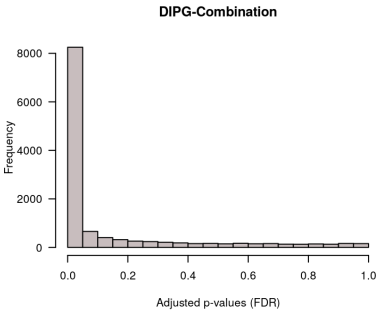
Once we have built our model, we need to extract the relevant information for our analysis. Below can be found different plots that help us understand the differences in expression between the compared samples. We considered significant differential expressed genes those with a p-value lower than 0.001. The p-value distributions show how many differential expressed genes are found in the treatment samples when compared against the control, as the distributions are generally skewed to the left, we can establish that the proportion of differential expressed genes is very high in the treatment samples. Moreover, the volcano plots show the proportion of how many of those genes are up-regulated or down-regulated gene. Furthermore, we have labeled the top 7 more significant genes in every sample, being these the ones with lower p-value.

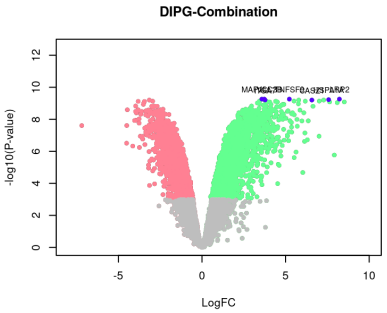


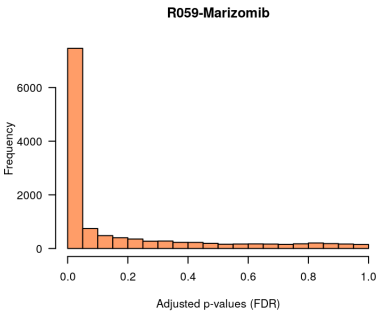


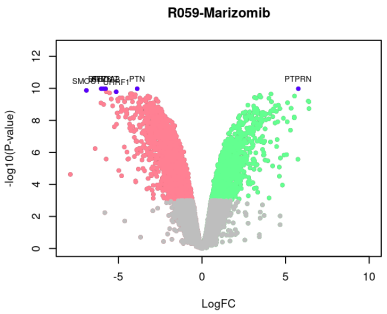












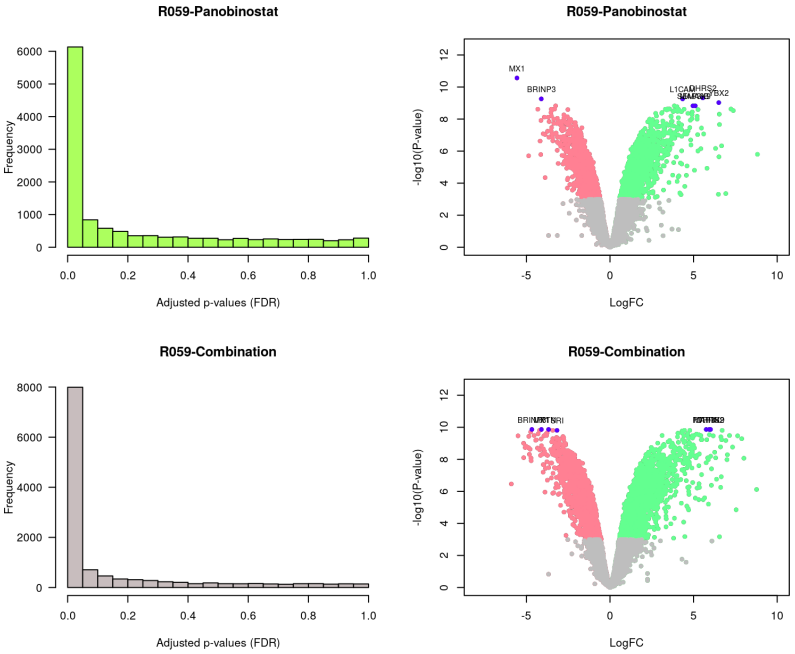


Figure 8: **P-value distributions and volcano plots for each sample group against control**
P-value distributions are colored by treatment. Volcano plots are colored as follows: significant down-regulated genes in red, significant up-regulated genes in green, non-significant genes in grey and top 7 genes with lower p-value in blue and labeled.

A trend can be appreciated in all 6 comparisons, with similar p-value distributions and volcano plot shapes. In the last ones we can see separated by colors the non-significant genes (gray points), the up-regulated ones (green points) and the down-regulated (red points). We find more divergence regarding the top 7 differential expressed genes, as they are up-regulated or down-regulated depending on the sample. At first glance, we see a more clear difference between cell types than between treatments, regarding the over or under expression of those top differential expressed genes.

Table 5: **Amount of differential expressed genes sorted by quantity of expression**

	DIPG6.Mar	DIPG6.Pano	DIPG6.Combo	R059.Mar	R059.Pano	R059.Combo
Down-regulated	949	1814	2550	2341	1317	2467
Not Significant	10355	8568	7137	8262	9426	7423
Up-regulated	1022	1944	2639	1723	1583	2436

In table 5 we can see the results per group of the **Differential Expression** analysis. For each group we can see the number of genes **up-regulated**, **down-regulated** and **non-significant**.

As we want to focus on the genes **up-regulated** and **down-regulated** we will create a second table with only them (getting rid of the **non-significant**) and showing for each case the top 7 DE genes (the ones with lower p-value).

Table 6: **Results of the differential expression analysis.** The table shows the results of the differential expression analysis of each sample group compared to the corresponding control group. The amount of DE genes are computed as the ones with FDR < 0.05. Top 7 DE genes are the ones with lower p-value.

Sample group	Amount of DE genes	Top 7 DE genes
DIPG6.Mar	1971	PTPRN, CRYAB, MLLT11, HSPA6, HSPA1A, ADAMTS1, MIR22HG
DIPG6.Pano	3758	DHRS2, ITGA7, YBX2, BAIAP3, KANK2, TIMP4, LINC00624
DIPG6.Combo	5189	TNFSF9, MAP1LC3B, VCL, LRP2, HSPA1A, ITGA7, CASZ1
R059.Mar	4064	ATP1A2, PTPRN, PTN, SEZ6, BTBD17, SMOC1, UHRF1
R059.Pano	2900	MX1, DHRS2, L1CAM, BRINP3, YBX2, MAP3K9, SEMA6B
R059.Combo	4903	DHRS2, PTPRN, BRINP3, MX1, MAP3K9, PTN, SRI

In table 6 can be found the **top 7 differential expressed genes** for each comparison between the treatments and the controls. We can appreciate that for each group the amount of differential expressed (DE) genes is in the order of the thousands.

From this table we see different interesting things:

- The highest and lowest number of DE genes are found on the cell line SU-DIPG-6. The **Combination** treatment for this certain cell line is the top scorer with **5189 DE genes** and **Marizomib** treatment the lowest scorer with **1971 DE genes**. It is interesting to see that variation within treatments in this cell line is greater than in the other cell line, QCTB-R059.
- The 2 higher number of DE genes are the ones corresponding to the **Combination** treatment (**5189 DE genes for SU-DIPG-6** and **4903 DE genes for QCTB-R059**). When applying this treatment the number of DE genes is greater than with other treatments, no matter the cell type.

After carefully checking the previous table and searching for common genes for each pairwise combination of treatments (only taking into account the top 7 genes present in the previous table), we can see the following common genes:

- Common DE genes in DIPG6-Marizomib and in R059-Marizomib: PTPRN
- Common DE genes present in DIPG6-Panobinostat and in R059-Panobinostat: DHRS2, YBX2
- No common DE genes in DIPG6-Combination and R059-Combination.

Here we can find those genes that are repeated between cell lines when applying the same treatment.

For the **Marizomib** treatment we only have one common DE gene:

- **PTPRN**: this gene encodes for the protein Tyrosine phosphatase receptor type N. This protein has an important role in the **regulation of the secretion pathways** of various neuroendocrine cells. It has been found in literature (Wang et al. 2021) that the **down-regulation** of it **reduced the proliferation and migration of glioma cells**, while its **up-regulation** produced the reversed effect, **inducing the proliferation of the glioma cells**. The authors finally state that reducing the expression of PTPRN could be used as a therapeutic strategy in glioma cells.

For the **Panobinostat** treatment we have two common DE genes:

- **DHRS2** : It may be considered an interesting case as these gene is found to be described by previous authors (Zhou et al. 2017) as a gene with a **tumor suppressing role**.
- **YBX2**: it encodes for the protein Y-box binding protein 2. It has been previously associated with properties of germ and cancer cells. Some authors (Suzuki et al. 2021) hypothesize that this gene may contribute to the characteristics of cancer stem cells. Whereas we couldn't find specific bibliography relating YBX2 with glioma, we found an interesting article (Gong et al. 2020) relating YBX1 with glioma. **YBX1** encodes for Y-box binding protein 1 (a protein from the same family) and its **overexpression is associated with the progression of glioma** with an influence in patient survival.

4 Functional analysis

In this part of our analysis we are going to identify the **enriched genes** in the samples. In order to do it we are going to follow a **Gene Set Enrichment Analysis** methodology.

In order to retrieve a data object containing the gene sets and their member genes, we are going to use the library msigdbr (https://cran.r-project.org/web/packages/msigdbr/vignettes/msigdbr-intro.html). We are going to use the hallmark gene sets, to retrieve the principal pathways in which our DE genes are involved and see what processes are more affected by the different treatments in the samples.

```
h_gene_sets = msigdbr(species = "human", category = "H")
head(h_gene_sets)
# A tibble: 6 × 15
  gs_cat gs_subcat gs_name gene_symbol entrez_gene ensembl_gene human_gene_symb…
  <chr>  <chr>     <chr>   <chr>             <int> <chr>        <chr>
1 H      ""        HALLMA… ABCA1                19 ENSG0000016… ABCA1
2 H      ""        HALLMA… ABCB8             11194 ENSG0000019… ABCB8
3 H      ""        HALLMA… ACAA2             10449 ENSG0000016… ACAA2
4 H      ""        HALLMA… ACADL                33 ENSG0000011… ACADL
5 H      ""        HALLMA… ACADM                34 ENSG0000011… ACADM
6 H      ""        HALLMA… ACADS                35 ENSG0000012… ACADS
# … with 8 more variables: human_entrez_gene <int>, human_ensembl_gene <chr>,
#   gs_id <chr>, gs_pmid <chr>, gs_geoid <chr>, gs_exact_source <chr>,
#   gs_url <chr>, gs_description <chr>
```

Now we are going to use the library fgsea (https://bioconductor.org/packages/release/bioc/vignettes/fgsea/inst/doc/fgsea-tutorial.html) to generate the **GSEA** plots. This will allow us to check the genes ranked per pathway, for each of the 6 samples that we have got. In order to get just a representative selection, we only select the **20 values** that have the lowest p-values and then we order it by the **Normalized Enrichment Score (NES)**.

The **Enrichment Score (ES)** is the degree to which a certain gene set is over-represented at the top or bottom of the ranked list of genes in the expression dataset.

The **NES** is an statistic for checking gene set enrichment results. It is basically a normalization of the **ES**, and with that it takes into account differences in gene set size and in-correlations between gene sets and expression data set.

Pathway	Gene ranks	NES	pval	padj
HALLMARK_E2F_TARGETS		-2.61	1.3e-19	6.7e-18
HALLMARK_G2M_CHECKPOINT		-2.26	4.1e-12	6.8e-11
HALLMARK_MITOTIC_SPINDLE		-1.71	2.2e-05	9.3e-05
HALLMARK_UV_RESPONSE_DN		1.60	1.6e-03	4.1e-03
HALLMARK_XENOBIOTIC_METABOLISM		1.71	9.9e-04	2.7e-03
HALLMARK_COAGULATION		1.73	1.8e-03	4.6e-03
MARK_INTERFERON_GAMMA_RESPONSE		1.79	2.1e-04	6.4e-04
HALLMARK_MTORC1_SIGNALING		1.81	2.5e-05	9.5e-05
LMARK_INTERFERON_ALPHA_RESPONSE		1.85	3.7e-04	1.1e-03
HALLMARK_COMPLEMENT		1.86	3.4e-05	1.1e-04
HALLMARK_IL2_STAT5_SIGNALING		1.86	2.9e-05	1.0e-04
HALLMARK_MYOGENESIS		1.90	2.0e-05	9.0e-05
HALLMARK_UV_RESPONSE_UP		1.96	3.7e-06	2.1e-05
HALLMARK_P53_PATHWAY		1.99	3.5e-07	2.2e-06
EPITHELIAL_MESENCHYMAL_TRANSITION		2.08	7.8e-08	5.6e-07
HALLMARK_HYPOXIA		2.10	2.3e-08	1.9e-07
_REACTIVE_OXYGEN_SPECIES_PATHWAY		2.18	5.2e-06	2.6e-05
HALLMARK_INFLAMMATORY_RESPONSE		2.30	1.1e-08	1.1e-07
HALLMARK_APOPTOSIS		2.36	1.2e-10	1.5e-09
HALLMARK_TNFA_SIGNALING_VIA_NFKB		2.65	3.8e-18	9.4e-17

Figure 9: **GSEA results for DIPG6-Marizomib**
Figure shows Top 20 gene sets with lowest p-value, ordered by NES.

Pathway	Gene ranks	NES	pval	padj
HALLMARK_E2F_TARGETS		-2.46	2.4e-15	1.2e-13
HALLMARK_G2M_CHECKPOINT		-2.45	9.2e-15	1.5e-13
HALLMARK_MYC_TARGETS_V1		-2.44	5.9e-15	1.5e-13
HALLMARK_MYC_TARGETS_V2		-2.36	5.0e-08	6.0e-07
HALLMARK_ANGIOGENESIS		-1.49	3.3e-02	8.3e-02
HALLMARK_APICAL_JUNCTION		1.37	1.8e-02	4.8e-02
HALLMARK_KRAS_SIGNALING_DN		1.44	1.8e-02	4.8e-02
HALLMARK_UV_RESPONSE_UP		1.44	1.1e-02	3.5e-02
HALLMARK_COAGULATION		1.46	1.4e-02	4.0e-02
EPITHELIAL_MESENCHYMAL_TRANSITION		1.47	3.1e-03	1.1e-02
HALLMARK_INFLAMMATORY_RESPONSE		1.53	7.9e-03	2.6e-02
HALLMARK_MYOGENESIS		1.58	1.4e-03	5.6e-03
HALLMARK_APOPTOSIS		1.59	1.3e-03	5.5e-03
MARK_INTERFERON_GAMMA_RESPONSE		1.65	4.6e-04	2.3e-03
HALLMARK_COMPLEMENT		1.66	8.0e-04	3.6e-03
HALLMARK_HEME_METABOLISM		1.67	2.8e-04	1.6e-03
LMARK_INTERFERON_ALPHA_RESPONSE		1.77	2.2e-04	1.4e-03
HALLMARK_P53_PATHWAY		1.85	2.2e-06	1.5e-05
HALLMARK_TNFA_SIGNALING_VIA_NFKB		1.96	4.6e-07	3.8e-06
HALLMARK_HYPOXIA		2.02	6.0e-08	6.0e-07

Figure 10: **GSEA results for DIPG6-Panobinostat**
Figure shows Top 20 gene sets with lowest p-value, ordered by NES.

Pathway	Gene ranks	NES	pval	padj
HALLMARK_E2F_TARGETS		-2.66	1.0e-19	5.2e-18
HALLMARK_G2M_CHECKPOINT		-2.32	3.5e-13	5.9e-12
HALLMARK_MYC_TARGETS_V1		-2.15	7.2e-11	6.0e-10
HALLMARK_MYC_TARGETS_V2		-2.07	3.1e-05	9.8e-05
ALLMARK_OXIDATIVE_PHOSPHORYLATION		-1.59	1.3e-04	3.3e-04
HALLMARK_XENOBIOTIC_METABOLISM		1.72	3.2e-04	8.0e-04
HALLMARK_HEME_METABOLISM		1.74	8.5e-05	2.3e-04
HALLMARK_UV_RESPONSE_UP		1.85	1.6e-05	5.8e-05
HALLMARK_COAGULATION		1.90	8.2e-05	2.3e-04
LMARK_INTERFERON_ALPHA_RESPONSE		1.96	2.2e-05	7.3e-05
HALLMARK_MYOGENESIS		1.99	5.2e-07	2.2e-06
HALLMARK_IL2_STAT5_SIGNALING		2.01	3.0e-07	1.4e-06
HALLMARK_INFLAMMATORY_RESPONSE		2.02	3.6e-06	1.4e-05
EPITHELIAL_MESENCHYMAL_TRANSITION		2.03	1.2e-07	6.5e-07
HALLMARK_COMPLEMENT		2.04	1.4e-07	7.1e-07
MARK_INTERFERON_GAMMA_RESPONSE		2.07	8.1e-08	5.1e-07
HALLMARK_P53_PATHWAY		2.25	4.9e-12	4.9e-11
HALLMARK_APOPTOSIS		2.25	2.7e-10	1.9e-09
HALLMARK_HYPOXIA		2.31	2.5e-12	3.1e-11
HALLMARK_TNFA_SIGNALING_VIA_NFKB		2.52	2.1e-16	5.3e-15

Figure 11: **GSEA results for DIPG6-Combination**
Figure shows Top 20 gene sets with lowest p-value, ordered by NES.

Pathway	Gene ranks	NES	pval	padj
HALLMARK_E2F_TARGETS		-2.47	2.4e-20	1.2e-18
HALLMARK_G2M_CHECKPOINT		-1.96	1.7e-08	2.9e-07
HALLMARK_ESTROGEN_RESPONSE_LATE		-1.57	2.8e-03	7.2e-03
HALLMARK_MITOTIC_SPINDLE		-1.36	1.6e-02	4.0e-02
HALLMARK_KRAS_SIGNALING_UP		1.61	1.3e-03	4.0e-03
HALLMARK_PROTEIN_SECRETION		1.65	2.3e-03	6.3e-03
HALLMARK_IL2_STAT5_SIGNALING		1.70	2.1e-04	7.5e-04
HALLMARK_MTORC1_SIGNALING		1.71	1.7e-05	1.0e-04
EPITHELIAL_MESENCHYMAL_TRANSITION		1.75	5.1e-05	2.1e-04
HALLMARK_IL6_JAK_STAT3_SIGNALING		1.77	1.8e-03	5.2e-03
HALLMARK_UV_RESPONSE_UP		1.78	5.5e-05	2.1e-04
LMARK_INTERFERON_ALPHA_RESPONSE		1.79	3.1e-04	1.0e-03
HALLMARK_APOPTOSIS		1.84	3.3e-05	1.8e-04
HALLMARK_HYPOXIA		1.91	3.8e-06	2.7e-05
HALLMARK_INFLAMMATORY_RESPONSE		1.92	5.5e-05	2.1e-04
HALLMARK_P53_PATHWAY		1.94	7.8e-07	7.8e-06
MARK_INTERFERON_GAMMA_RESPONSE		1.98	2.9e-06	2.4e-05
_REACTIVE_OXYGEN_SPECIES_PATHWAY		2.04	4.0e-05	2.0e-04
LMARK_UNFOLDED_PROTEIN_RESPONSE		2.16	1.3e-07	1.7e-06
HALLMARK_TNFA_SIGNALING_VIA_NFKB		2.57	4.0e-14	9.9e-13

Figure 12: **GSEA results R059-Marizomib**
Figure shows Top 20 gene sets with lowest p-value, ordered by NES.

Pathway	Gene ranks	NES	pval	padj
LMARK_INTERFERON_ALPHA_RESPONSE		-2.65	9.2e-13	2.3e-11
HALLMARK_MYC_TARGETS_V2		-2.37	1.3e-08	1.6e-07
HALLMARK_MYC_TARGETS_V1		-2.34	4.5e-14	2.3e-12
HALLMARK_E2F_TARGETS		-2.23	4.8e-12	8.1e-11
MARK_INTERFERON_GAMMA_RESPONSE		-2.11	7.2e-08	6.0e-07
HALLMARK_G2M_CHECKPOINT		-1.99	2.2e-08	2.2e-07
HALLMARK_GLYCOLYSIS		1.26	8.8e-02	2.3e-01
HALLMARK_XENOBIOTIC_METABOLISM		1.26	9.9e-02	2.3e-01
HALLMARK_APOPTOSIS		1.28	8.1e-02	2.3e-01
HALLMARK_KRAS_SIGNALING_DN		1.30	8.6e-02	2.3e-01
HALLMARK_KRAS_SIGNALING_UP		1.30	7.1e-02	2.3e-01
HALLMARK_APICAL_SURFACE		1.36	8.2e-02	2.3e-01
HALLMARK_PANCREAS_BETA_CELLS		1.50	4.4e-02	1.6e-01
HALLMARK_SPERMATOGENESIS		1.54	1.4e-02	5.4e-02
HALLMARK_TNFA_SIGNALING_VIA_NFKB		1.56	2.4e-03	1.0e-02
HALLMARK_HEME_METABOLISM		1.61	7.0e-04	3.4e-03
HALLMARK_APICAL_JUNCTION		1.62	7.0e-04	3.4e-03
HALLMARK_P53_PATHWAY		1.64	7.4e-04	3.4e-03
HALLMARK_UV_RESPONSE_UP		1.79	3.5e-05	2.2e-04
HALLMARK_HYPOXIA		1.88	6.8e-06	4.8e-05

Figure 13: **GSEA results R059-Panobinostat**
Figure shows Top 20 gene sets with lowest p-value, ordered by NES.

Pathway	Gene ranks	NES	pval	padj
HALLMARK_E2F_TARGETS		-2.60	6.3e-19	3.1e-17
HALLMARK_MYC_TARGETS_V1		-2.44	9.4e-16	2.3e-14
HALLMARK_G2M_CHECKPOINT		-2.02	1.2e-08	1.2e-07
HALLMARK_MYC_TARGETS_V2		-2.01	1.9e-05	1.0e-04
LMARK_INTERFERON_ALPHA_RESPONSE		-1.96	5.1e-05	2.5e-04
ALLMARK_OXIDATIVE_PHOSPHORYLATION		-1.93	1.1e-07	7.7e-07
MARK_INTERFERON_GAMMA_RESPONSE		-1.41	7.8e-03	2.3e-02
EPITHELIAL_MESENCHYMAL_TRANSITION		1.40	1.3e-02	3.5e-02
HALLMARK_IL2_STAT5_SIGNALING		1.40	1.5e-02	3.6e-02
HALLMARK_INFLAMMATORY_RESPONSE		1.48	1.1e-02	3.0e-02
ALLMARK_ESTROGEN_RESPONSE_EARLY		1.51	4.1e-03	1.3e-02
HALLMARK_MTORC1_SIGNALING		1.54	1.4e-03	4.9e-03
HALLMARK_KRAS_SIGNALING_UP		1.59	3.2e-03	1.1e-02
HALLMARK_HEME_METABOLISM		1.62	4.4e-04	1.7e-03
LMARK_UNFOLDED_PROTEIN_RESPONSE		1.68	3.8e-04	1.6e-03
HALLMARK_APOPTOSIS		1.70	3.3e-04	1.5e-03
HALLMARK_UV_RESPONSE_UP		1.89	4.8e-06	3.0e-05
HALLMARK_HYPOXIA		2.07	1.5e-08	1.3e-07

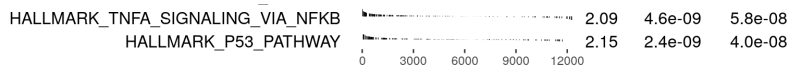


Figure 14: **GSEA results R059-Combination**
Figure shows Top 20 gene sets with lowest p-value, ordered by NES.

In the graphs above (figures 9, 10, 11, 12, 13, 14) we find the 20 pathways with the lowest p-values in which our differential expressed genes are involved, ordered by their Normalized Enrichment Score (NES). There is a graph for each sample that we are working with, and the sign of the NES score indicates the direction of the expression: if it is negative the pathway is down-regulated and if it is positive, the pathway is up-regulated.

Between all samples we can find common hallmark pathways down-regulated, such as E2F transcription factors and G2M checkpoint components. On the contrary, we find up-regulated pathways such as p53, TNFA via NFkB, Apoptosis and Hypoxia. These pathways are involved in the maintenance of cell viability and its proliferation, the down-regulated ones are typically involved in cell proliferation, while the up-regulated ones adopt more of a tumor suppressing role. These pathways are very extensive and are involved also in many different cellular processes, it is kind of expected that they are affected by all treatments.

Regarding the differences between treatments, we have noted that we find down-regulated the MYC targets pathway only in the cells treated with Panobinostat and Combination. This pathway is, again, involved in many cellular processes such as cell proliferation (c-Myc is a known oncogene), maturation and death and its differential expression is maintained between both cell lines with the same treatments.

Also it is worth noting the presence of the Unfolded Protein Response (UPR) as up-regulated in QCTB-R059 cells treated with Marizomib and Combination. This pathway is conformed of genes that are typically up-regulated during the response of unfolded proteins, a cellular stress response related to the endoplasmic reticulum. Interestingly, this is not found differentially expressed in any SU-DIPG-6 cell samples.

5 Discussion

Taking into account the results obtained by our analysis, we can highlight some important points to understand the expression landscape of the studied cancer samples. In relation to the amount of differential expressed genes sorted by quantity of expression (table 5), we can see that between cell lines (comparing the same treatment between cell lines), there are not a lot of differences in the amount of differential expressed genes, with the exception of marizomib. The majority of differences in the amount of differential expressed genes are seen when comparing treatments within the same cell line. When comparing the amount of DE genes within the same cell line (comparing the treatments), we can see how combination treatment had the highest amount of significant differentially expressed genes (down-regulated and up-regulated). In SU-DIPG-6, the treatment with less significant DE genes was with only marizomib, but in QCTB-R059, the treatment with less significant DE genes was panobinostat.

Having said that, we can see how the marizomib treatment in QCTB-R059 has much more down-regulated genes compared to the marizomib treatment in SU-DIPG-6 (2341 and 949 respectively). This difference may happen because of the difference in cell type. SU-DIPG-6 cells were obtained in early postmortem autopsy from a DIPG grade III tumor in the pons, and had the TP53 and H3.3K27M mutated. However, QCTB-R059 cells were obtained in a surgical resection from a pediatric glioblastoma in the thalamus and only had H3.3K27M mutated. Moreover, the patient from which the SU-DIPG-6 cells where obtained had been previously treated with selective adjuvant radiotherapy and vorinostat (inhibitor of histone deacetylase). Knowing that the nature of the patient-derived cell lines could explain the differences found between treatments, it would be interesting to obtain more samples of different glioma patients and include them in our analysis.

Regarding our results and the ones in the original paper, we have found some common findings in the **differential expressed pathways** in both cell lines. As we have mentioned earlier in the functional analysis part, we see many pathways related to **cell proliferation** affected by the treatments. This effect was kind of expected, as we are working with patient-derived cancer cell lines, the treatments will increase the expression of tumor suppressing pathways and down-regulate the oncogenic ones.

In the original publication, Lin et al. find a consistent up-regulation of the UPR gene set in the samples treated with Marizomib and Combination across patient-derived cultures. Our results are partially consistent with this finding, as we also found an up-regulation of this gene set in the Marizomib and Combination QCTB-R059 samples (figures 12, 14), but we didn't find it between the top 20 differentially expressed gene sets in the SU-DIPG-6 samples (figures 9, 11). Other proteasome inhibitors such as Bortezomib, have been shown to promote the up-regulation of components in the unfolded protein response (UPR) due to the induction of stress in the endoplasmic reticulum (Mujtaba and Dou 2011). Marizomib, also a proteasome inhibitor, could act in a similar way, as we have reported an upregulation of this gene set in half of the samples treated with Marizomib.

Along with the previous finding, we found a down-regulation of the oxidative phosphorylation gene sets in the combination samples of both cell lines (figures 11, 14), which is also a highlighted finding in the original paper. This down-regulation is only present in the combination samples and it is not differentially expressed in samples treated with only Panobinostat or Marizomib. This fact could pinpoint the down-regulation of the oxidative phosphorylation gene set to a **synergic effect** of the combination of both treatments, rather than their separate effects.

Interestingly, we also found a cell proliferation pathway (hallmark Myc targets) that appeared differentially expressed (down-regulated) in only Panobinostat and Combination samples in both cell lines (figures 10, 11, 13, 14). Although the effect of histone deacetylases inhibitors like Panobinostat in c-Myc expression has not yet been fully understood, there have been reports where the addition of Panobinostat down-regulated the Myc expression in cancer cells (Nebbioso et al. 2017). These reports are consistent with our results, as in all samples treated with Panobinostat or in combination with Marizomib show down-regulation of hallmark Myc targets.

6 Conclusions

Diffuse midline gliomas and diffuse intrinsic pontine gliomas are known for being lethal childhood cancers without an effective treatment. Here, we studied the expression profiles of some patient-derived cell lines treated with experimental drugs and identified the principal gene sets affected by them. Our observations direct us to believe that the combination of Marizomib and Panobinostat lead SU-DIPG-6 and QTCB-R059 cells to metabolic collapse. Our GSEA enrichment results show a down-regulation of known oncogens (E2F, MYC) and oxidative phosporylation gene sets in the samples treated with the drug combination, as well as changes in expression of cell cycle and apoptosis related gene sets which is consistent with the results reported in the paper. The original study also includes extra experimental assays to verify the cytotoxic effects of this treatment on Diffuse Midline Glioma cells, which we have not reproduced. These assays demonstrate that glioma cells are sensitive to metabolic dysregulation and that metabolic collapse is one of the main causes of cytotoxicity of this treatment on glioma cells.

To further assess the benefits of combining both drugs to treat diffuse midline gliomas, another control of healthy cells could be added to the analysis. The comparison of treated tumor cells with treated healthy cells would determine whether the treatment effect is specific for tumor cells or affects all cells equally.Moreover, additional research on this topic may also include extending the analysis to more patients with both similar and different tumor features (for instance, different tumor grade or different genetic background). That would be useful to validate the current results and to check any differences in the treatment effect due to different kinds of tumor.

As we have remarked before, there is not a specific nor effective treatment to deal with this kind of cancers, so it is very important to develop combinational drug strategies to treat them. Here we pinpoint the combination of Marizomib and Panobinostat as a promising new therapy in hopes to accelerate the process of finding an effective treatment for this disease.

7 Session information

```
sessionInfo()
R version 4.1.3 (2022-03-10)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Ubuntu 20.04.4 LTS

Matrix products: default
BLAS/LAPACK: /usr/lib/x86_64-linux-gnu/openblas-pthread/libopenblasp-r0.3.8.so

locale:
 [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
 [3] LC_TIME=en_US.UTF-8        LC_COLLATE=en_US.UTF-8
 [5] LC_MONETARY=en_US.UTF-8    LC_MESSAGES=en_US.UTF-8
 [7] LC_PAPER=en_US.UTF-8       LC_NAME=C
 [9] LC_ADDRESS=C               LC_TELEPHONE=C
[11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C

attached base packages:
[1] stats4    stats     graphics  grDevices utils     datasets  methods
[8] base

other attached packages:
 [1] tidyr_1.2.0                 msigdbr_7.5.1
 [3] RColorBrewer_1.1-3          fgsea_1.20.0
 [5] sva_3.42.0                  BiocParallel_1.28.3
 [7] genefilter_1.76.0           mgcv_1.8-40
 [9] nlme_3.1-157                geneplotter_1.72.0
[11] annotate_1.72.0             XML_3.99-0.9
[13] AnnotationDbi_1.56.2        lattice_0.20-45
[15] edgeR_3.36.0                limma_3.50.3
[17] IEOprojectGlioma_1.2.0      SummarizedExperiment_1.24.0
[19] Biobase_2.54.0              GenomicRanges_1.46.1
[21] GenomeInfoDb_1.30.1         IRanges_2.28.0
[23] S4Vectors_0.32.4            BiocGenerics_0.40.0
[25] MatrixGenerics_1.6.0        matrixStats_0.61.0
[27] usethis_2.1.5               kableExtra_1.3.4
[29] knitr_1.38                  BiocStyle_2.22.0

loaded via a namespace (and not attached):
 [1] colorspace_2.0-3       ellipsis_0.3.2         rprojroot_2.0.3
 [4] XVector_0.34.0         fs_1.5.2               rstudioapi_0.13
 [7] farver_2.1.0           roxygen2_7.1.2         remotes_2.4.2
[10] bit64_4.0.5            fansi_1.0.3            xml2_1.3.3
[13] splines_4.1.3          cachem_1.0.6           pkgload_1.2.4
[16] jsonlite_1.8.0         png_0.1-7              BiocManager_1.30.16
[19] compiler_4.1.3         httr_1.4.2             assertthat_0.2.1
[22] Matrix_1.4-1           fastmap_1.1.0          cli_3.2.0
[25] htmltools_0.5.2        prettyunits_1.1.1      tools_4.1.3
[28] gtable_0.3.0           glue_1.6.2             GenomeInfoDbData_1.2.7
[31] dplyr_1.0.8            fastmatch_1.1-3        Rcpp_1.0.8.3
[34] jquerylib_0.1.4        vctrs_0.4.1            Biostrings_2.62.0
[37] babelgene_22.3         svglite_2.1.0          xfun_0.30
[40] stringr_1.4.0          ps_1.6.0               brio_1.1.3
[43] testthat_3.1.3         rvest_1.0.2            lifecycle_1.0.1
[46] devtools_2.4.3         zlibbioc_1.40.0        scales_1.2.0
[49] parallel_4.1.3         yaml_2.3.5             gridExtra_2.3
[52] memoise_2.0.1          ggplot2_3.3.5          sass_0.4.1
[55] stringi_1.7.6          RSQLite_2.2.12         highr_0.9
[58] desc_1.4.1             pkgbuild_1.3.1         rlang_1.0.2
[61] pkgconfig_2.0.3        systemfonts_1.0.4      bitops_1.0-7
[64] evaluate_0.15          purrr_0.3.4            labeling_0.4.2
[67] tidyselect_1.1.2       bit_4.0.4              processx_3.5.3
[70] magrittr_2.0.3         bookdown_0.26          R6_2.5.1
[73] magick_2.7.3           generics_0.1.2         DelayedArray_0.20.0
[76] DBI_1.1.2              pillar_1.7.0           withr_2.5.0
[79] survival_3.3-1         KEGGREST_1.34.0        RCurl_1.98-1.6
[82] tibble_3.1.6           crayon_1.5.1           KernSmooth_2.23-20
[85] utf8_1.2.2             rmarkdown_2.13         locfit_1.5-9.5
[88] grid_4.1.3             data.table_1.14.2      blob_1.2.3
[91] callr_3.7.0            digest_0.6.29          webshot_0.5.3
[94] xtable_1.8-4           munsell_0.5.0          viridisLite_0.4.0
[97] bslib_0.3.1            sessioninfo_1.2.2
```

References

Gong, Huilin, Shan Gao, Chenghuan Yu, Meihe Li, Ping Liu, Guanjun Zhang, Jinning Song, and Jin Zheng. 2020. “Effect and Mechanism of YB-1 Knockdown on Glioma Cell Growth, Migration, and Apoptosis.” *Acta Biochimica Et Biophysica Sinica* 52 (2): 168–79. https://doi.org/10.1093/abbs/gmz161 (https://doi.org/10.1093/abbs/gmz161).

Lin, Grant L., Kelli M. Wilson, Michele Ceribelli, Benjamin Z. Stanton, Pamelyn J. Woo, Sara Kreimer, Elizabeth Y. Qin, et al. 2019. “Therapeutic Strategies for Diffuse Midline Glioma from High-Throughput Combination Drug Screening.” *Science Translational Medicine* 11 (519). https://doi.org/10.1126/scitranslmed.aaw0064 (https://doi.org/10.1126/scitranslmed.aaw0064).

Mujtaba, Taskeen, and Q Ping Dou. 2011. “Advances in the Understanding of Mechanisms and Therapeutic Use of Bortezomib.” *Discovery Medicine* 12 (67): 471.

Nebbioso, Angela, Vincenzo Carafa, Mariarosaria Conte, Francesco Paolo Tambaro, Ciro Abbondanza, Joost Martens, Matthias Nees, et al. 2017. “C-Myc Modulation and Acetylation Is a Key HDAC Inhibitor Target in CancerMyc Modulation by HDACi in Cancer.” *Clinical Cancer Research* 23 (10): 2542–55.

Suzuki, Izumi, Sachiko Yoshida, Kouichi Tabu, Soshi Kusunoki, Yumiko Matsumura, Hiroto Izumi, Kazuo Asanoma, et al. 2021. “Ybx2 and Cancer Testis Antigen 45 Contribute to Stemness, Chemoresistance and a High Degree of Malignancy in Human Endometrial Cancer.” *Scientific Reports* 11 (1). https://doi.org/10.1038/s41598-021-83200-5 (https://doi.org/10.1038/s41598-021-83200-5).

Wang, Dong, Fan Tang, Xi Liu, Yueshan Fan, Yu Zheng, Hao Zhuang, Budong Chen, Jie Zhuo, and Bo Wang. 2021. “Expression and Tumor-Promoting Effect of Tyrosine Phosphatase Receptor Type n (PTPRN) in Human Glioma.” *Frontiers in Oncology* 11. https://doi.org/10.3389/fonc.2021.676287 (https://doi.org/10.3389/fonc.2021.676287).

Zhou, Y, L Wang, X Ban, T Zeng, Y Zhu, M Li, X-Y Guan, and Y Li. 2017. “Dhrs2 Inhibits Cell Growth and Motility in Esophageal Squamous Cell Carcinoma.” *Oncogene* 37 (8): 1086–94. https://doi.org/10.1038/onc.2017.383 (https://doi.org/10.1038/onc.2017.383).