Universitat Autònoma de Barcelona

ISGlobal Institut de Salut Global Barcelona

TRANSCRIPTOME PROFILING OF PATIENTS WHO DEVELOPED A SECOND TUMOR AFTER RADIOTHERAPY

Èlia Palma Rojo

MSc Bioinformatics, 2018-2019

Supervisor: Juan Ramón González

Co-supervisors: Gemma Armengol and Joan Francesc Barquinero

Academic tutor: Raquel Egea

July 2019

APPROVAL AND SIGNATURE

TRANSCRIPTOME PROFILING OF PATIENTS WHO DEVELOPED A SECOND TUMOR AFTER RADIOTHERAPY

Juan Ramón González	*Raquel Egea*	*Èlia Palma*
Supervisor	**Academic tutor**	**Student**
(Signature)	*(Signature)*	*(Signature)*
		

Universitat Autònoma de Barcelona

ISGlobal Institut de Salut Global Barcelona

TABLE OF CONTENTS

All the documents mentioned below can be access from https://github.com/isglobal-brge/master_thesis/tree/master/radiation_cancer

ABBREVIATIONS

Conditional maximum likelihood= CML
Cytotoxic T cells= CTLs
Differentially expressed/Differentially expressed genes= DE/DEGs
Differential expression analysis= DEA
DNA damage response= DDR
Expression heterogeneity= EH
False discovery rate= FDR
Fold Change= FC
Gene Ontology= GO
Generalized linear model= GLM
Gene set enrichment analysis= GSEA
Guanine-cytosine content= GC-content
Ionizing radiation= IR
Kyoto Encyclopedia of Genes and Genomes= KEGG
Molecular Signatures Database= MSigDB
Over representation analysis= ORA
RangedSummarizeExperiment= RSE
Reads per kilobase per million reads= RPKM
Ribonucleic acid sequencing= RNA-seq
Surrogate variable analysis= SVA
Surveillance, Epidemiology, and End Results= SEER
The Cancer Genome Atlas= TCGA
Trimmed Mean of M-values= TMM

ABSTRACT

One of the most used and effective treatments for cancer is radiotherapy. However, patients undergoing this kind of treatment may suffer from short-term and long-term adverse effects, even leading to a second tumor development. It is worth noticing the interpatient variation in radiosensitivity, which is influenced by variations in the DNA damage response (DDR) network and has an important genetic background. In second tumor development after radiotherapy and/or chemotherapy, the genetic pathways involved in radiosensitivity play an important role, but also those related to drug metabolism and response to treatment. The gene variants and gene expression variants involved in second cancer development after radiotherapy could be used as a biomarker to adjust the therapy to the individual needs, reducing this way the incidence of secondary malignancies. In this study, a script to identify these gene expression variants from ribonucleic acid sequencing (RNA-seq) data is performed by means of the softwares DESeq2 and limma+voom. After that, an enrichment analysis with the resulting differentially expressed genes (DEGs) is carried out using different over representation analysis (ORA) methods. According to the results obtained, DESeq2 is the most suitable software for this particular study, matching some of the DEGs found those already published. The enrichment analysis evidences an upregulation of cell motility and migration, NOD-like receptor, MAPK signaling, and cell proliferation pathways. On the other hand, a downregulation of the immune and stress responses is produced. Further studies should be performed to validate these results and assess their specific role in second tumor development.

INTRODUCTION

Cancer is one of the main afflictions of the population, being 439.2 per 100,000 the number of new cases per year in the United States (1). For this reason, preventing cancer, controlling its growth and eradicating it are currently some of the main goals of many cancer-related research teams (2).

One of the most effective cancer treatments is radiotherapy, which is used in 50% of cancer patients at some point in the course of the disease (3). Tumor exposure to ionizing radiation (IR) has several effects, among them DNA damage, either through direct interaction with the DNA or through the ionization of neighboring molecules, such as water and other organic compounds of the cell (Figure 1) (4). Cells can respond to this damage using the DDR pathways. A recruitment of factors is produced in order to activate different cellular responses according to the severity of the damage, cell type, and other factors. If the cells are not able to repair or a misrepair takes place, a genomic instability may occur leading to cell death or to cellular senescence (Figure 2)(5). Apoptosis is the cell death mechanism by excellence after an exposure to radiation, being the cells with high proliferation rate and high expression of pro-apoptosis genes the most sensitive to apoptosis (2). This way, the removal of cancerous cells is boosted, and tumor progression can be stopped.

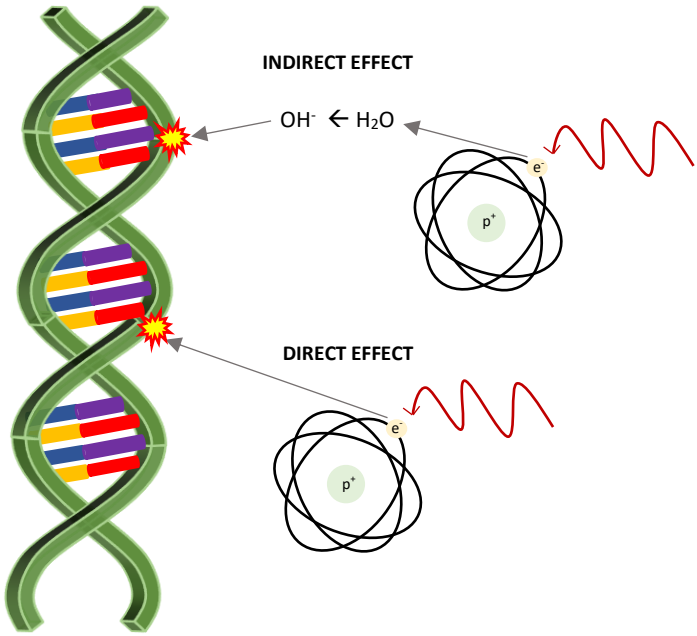

***Figure 1 Direct and indirect actions of IR**. In the indirect effect, free radicals such as hydroxyl are produced. These radicals produce molecular structural damages when interacting with the DNA. On the other hand, in the direct effect the radiation interacts directly with the DNA molecule disrupting its structure.*

However, radiotherapy may produce short-term toxicity and long-term consequences (Figure 2). Patients tend to recover well from the short-term or acute adverse effects.

On the contrary, the long-term adverse effects have a major impact on quality of life and survival (5).

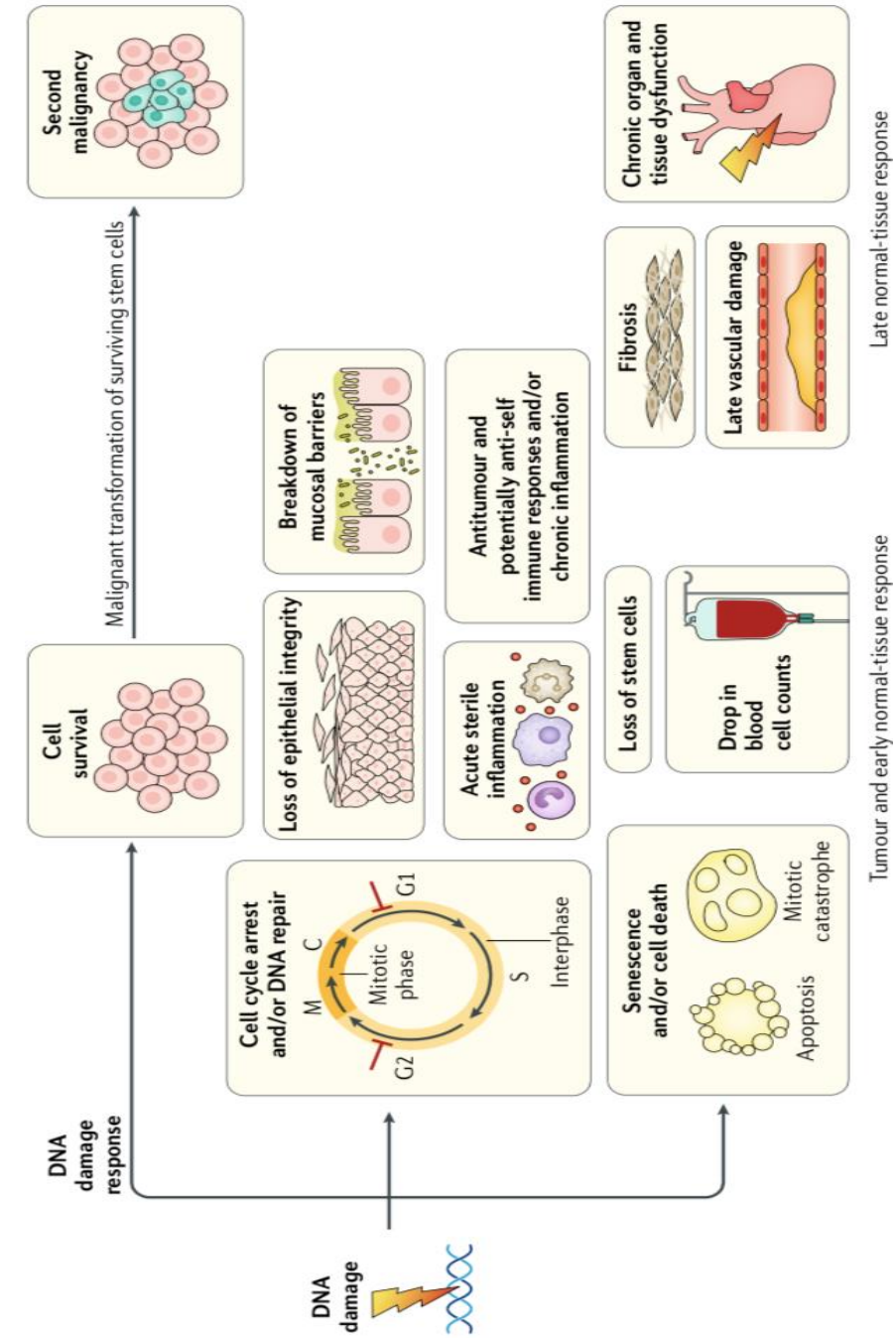

***Figure 2 Cellular response after radiotherapy**. Both, normal and tumor tissue suffer DNA damage after the exposure to therapeutic IR. After that, the irradiated cells stop the cell cycle to repair the damage. If the cell can repair the DNA, it survives without impairment. When the cell is not capable to completely repair the damage, the cell can enter into a senescence process or cell death (such as apoptosis or mitotic catastrophe). In the early response, a drastic drop in blood cell counts can occur. Due to IR an inflammation response is triggered, which contributes to tissue damage, but it is also very important for the induction of beneficial antitumor response. The surrounding normal tissue can suffer from late adverse effects, like transient mucositis and lymphopenia, and vascular damage, tissue fibrosis and organ dysfunction. Stem cells with residual DNA damage can suffer a malignant transformation leading to second cancers. Modified from ref. 5*

These side effects are due not only to the response of tumoral tissue to the exposure to radiotherapy, but especially to the interaction of IR with normal tissue. This damage to normal tissue is due to the loss of cells, bystander effects on unirradiated cells and inflammation. Tissues with high stem cell activity and regenerative capacity experience an acute response which leads to cell death, producing short-term toxicity. On the other hand, those with a low stem cell activity or structurally important cells respond to radiotherapy by permanent cell cycle arrest. This tends to produce long-term side effects which are usually irreversible (5).

Radiotherapy-related long-term toxicity depends on treatment regimen, the irradiated volume, location of the tumor, the age and sex of the patient. It is also worth noticing the interpatient variations in tumor and normal-tissue radiosensitivity influenced by variations in the DDR network. In this sense, the genetic component of radiosensitivity can reach 80% (5). Up to the present, the main candidate genes studied are implicated in DNA damage recognition and repair, free radical scavenging and anti-inflammatory response (3), among others (Table 1).

Table 1 Candidate genes for radiosensitivity and second tumor development after treatment.

Pathways	Candidate Genes (references)
DNA damage recognition and repair	*ATM* (3,6,7) *BRCA1* (3,6,7) *BRCA2* (3,6,7) *TP53 (3,6)* *TANC1* (8) *MSH2* (6,9) *MLH1* (6,9) *XRCC* (6,9) *RAD51* (6,9) *MDM2 (6)* *XRCC3 (6)*
Free radical scavenging	*SOD2 (3)* *TXNRD2* (8)
Anti-inflammatory/inflammatory response	*TGFB1* (3,8) *TNF* (8)
Apoptosis	*HSPB1* (8)
Drug metabolism/Response to treatment	*NQO1* (6,9) *GSTP1* (6,9) *CYP* (9) *GSTM1 (6)* *PRDM1 (6)*

Moreover, in some cases, normal cells with residual damage after radio/chemotherapy can become tumoral cells and cause second cancers, the leading cause of morbidity and mortality among cancer survivors. The occurrence of second cancers has increased considerably as a consequence of the enlargement of the population of cancer survivors

and its aging, reaching up to 19% of all cancer diagnoses in the United States between the years 2005 and 2009 according to the Surveillance, Epidemiology, and End Results (SEER) (6). It is worth noticing that one of the main risk factors for second cancer development is the primary cancer treatment. Both, radiotherapy and chemotherapy have been associated with an increased risk for second cancer, being in the case of radiotherapy the tissues in or near the radiotherapy treatment fields the ones with higher risk (6).

In second tumor development, other genetic components besides the ones above-mentioned for radiosensitivity, such as the drug metabolism pathways also play a key role, which suggests a larger effect of genetic factors than in sporadic cancers (Table 1). It is worth mentioning the importance of the interaction between genes in second cancer development (6). For example, the gene *XRCC* is involved in base excision repair and DNA single-strand breaks (10) and *RAD51* is one of the most important proteins for homologous recombination and plays a key role in DNA double-strand breaks repair (11) individuals with two variants of these genes that decreased the repair capacity (*XRCC Thr241Met* and *RAD51 G135C)* presented higher risk of therapy-related acute myeloid leukemia risk increases (6). One example of the relationship between radiotherapy and genetic susceptibility in second cancer development would be the case of women who carry *BRCA1*/*BRCA2* mutations, they have a 4-fold higher risk of contralateral breast cancer development after radiation exposure (7).

Gene variants and gene expression variants involved in second cancer development after radiotherapy could be used as a biomarker, adjusting this way the therapy given to the primary tumor in high risk patients and mitigating the incidence of second tumor development.

Thanks to the availability of RNA-seq data, a differential expression analysis (DEA) can be performed (12) to see which genes change their expression in patients with secondary cancer. The online resource recount2 contains RNA-seq gene counts in a RangedSummarizeExperiment (RSE) format and a phenotype information table. Specifically, information about the The Cancer Genome Atlas (TCGA) project which contains data of 33 different types of cancer, can be accessed (13). These RNA-seq profiles consist of integer counts of the number of reads mapping into a gene or exon (14). With this data, the DEGs can be identified.

The work performed by Costa-Silva, Domingues and Martins Lopes shows that limma+voom and DESeq2 are two of the most balanced softwares when considering the precision, accuracy and sensitivity for DEA (12). The common steps for both are:

1. **Normalization**. Required step to compare gene expression. It is used to take into account different factors that could affect the number of reads mapping to a gene, such as the sequencing and transcript length, and the guanine-cytosine content (GC-content). The length and GC-content are within sample effects, so they affect the comparison of

read counts between different genes in the same sample. As for the sequencing depth, the effect is produced between samples. When performing DEA, it is important to normalize between samples. Experimental variability, as the total number of molecules sequenced, can influence the total number of read counts in different samples. If a sample has more reads than another, genes that are not DE will tend to have higher read counts (15). Some of the normalization methods used are listed below (16):

- Reads per kilobase per million reads (RPKM)
- Trimmed Mean of M-values (TMM)
- DESeq
- Quantile

In the limma+voom pipeline, different normalization methods can be used. With the voom-conversion, a transformation of the normalized counts to logarithmic base 2 and estimation of the mean-variance relationship to determine a weigh to each observation, is performed (17).
In DESeq2 the non-DEGs get similar level counts. This model assumes that most of the genes do not change, and those that change are divided equally between up and down regulated genes. It calculates the normalization factors within statistical testing, but it does not transform the data. It takes the geometric mean of gene counts across all samples and divides the gene counts in a sample by the geometric mean. Then, it takes a median of these ratios, which will be the sample normalization factor.
In the presence of high-count genes, only DESeq and TMM (edgeR) are able to maintain a reasonable false positive rate without any loss of power (16).

2. **Dispersion estimation**. When comparing gene expression levels between groups, it is important to know also its within-group variability. The variability seen in counts is a sum of two things, the sample-to-sample variation (dispersion) and uncertainty in measuring expression by counting reads. A lot of replicates would be needed to perform the analysis, but for RNA-seq there are usually few replicates, so it is difficult to estimate within-group variability. One possible solution consists in pooling information across genes, look at other genes expressed at a similar level and assume that if they are similarly expressed, their variance would also be approximately the same. In low-count genes there is a lot of noise, so they are not reliable (18).
DESeq uses maximum likelihood to estimate gene-wise dispersion for each gene. As specified in the Figure 3, the gene wise dispersion estimates are shrinked towards the values predicted by the curve (in red). An empirical Bayes approach is used to shrink these values, being the amount of shrinkage dependent on several factors (19).

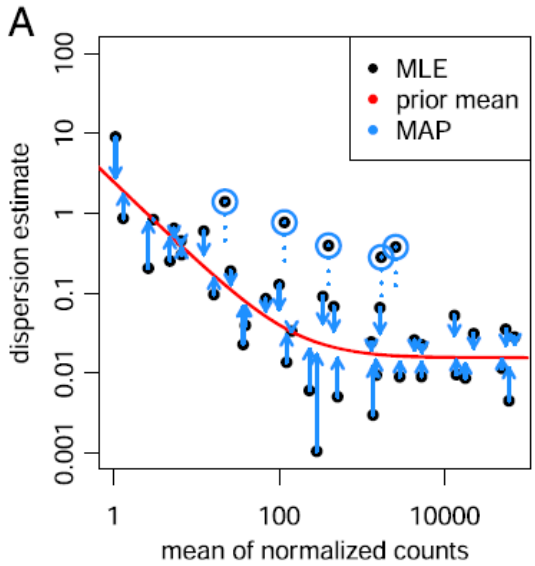

Figure 3 Dispersion estimation by DESeq2. *Gene-wise MLE are first obtained (black dots), then, a curve (red) is fitted to the MLEs to capture the trend of dispersion-mean dependence. This fit will be used as a mean for a second estimation, which results in the final MAP estimates of dispersion (arrows). The dispersion outliers (circled in blue) are not shrunk. MLE:maximum-likelihood estimate; MAP: maximum a posteriori. Modified from ref. 19.*

3. **Statistical modeling and testing**.

Parametric. the main parametric methods used in RNA-seq data are the Poisson, the negative binomial and the beta binomial.

Based on the **Poisson** distribution, a log-linear model can be used to model the mean difference between two samples adopting the likelihood ratio test. The main drawback of this model is that in a group there is usually a between-sample variation of sequencing reads for a gene which exceeds the mean (overdispersion), and that can not be explained using the Poisson model (16).

As for the **negative binomial**, it is used to accommodate the overdispersion among biological replicate data. Is the most common assumption in a parametric framework. The Conditional maximum likelihood (CML) is used to estimate the dispersion. For the statistical testing, the exact test can be used (16).

The **beta binomial** is another option to accommodate the overdispersion. It takes into account the between-library and within-library variations (16).

For all parametric models a classic log-likelihood ratio test can be employed (16).

Bayesian and empirical Bayesian. Bayesian false discovery rate (FDR) or local FDR can be used (16).

Nonparametric. Nonparametric methods can capture the distribution of the data in a more detailed way, a rigid model is not used in this case (12). For the statistical testing, the Wilcoxon statistic is a good option (16).

The limma+voom pipeline includes a linear modeling in order to analyze experiments with multiple treatment factors, quantitative weights to account for variation in precision between different observations, and empirical Bayes statistical methods to borrow strength between genes. The normal-based empirical Bayes procedures can

adapt to different datasets and provide exact type I error rate control even for experiments with small number of replicate samples (14).
On the other hand, DESeq2 is based on the negative binomial model and Generalized linear model (GLM). It shrinks log fold changes (FC) to zero when the counts are low using an empirical Bayes method (shrinkage is stronger when counts are low, dispersion is high, or there are only a few samples). It uses the Wald test for significance.

4. **Multiple testing correction.** There is a broad number of genes being tested, so it is possible that some genes get good p-values just by chance. To control this problem, p-values need to be corrected for multiple testing.

5. **Filtering**. The filtering reduces the severity of multiple testing correction by removing some genes (makes n smaller). The genes filtered out are those which have little chance of showing evidence for significant DE; these are genes not expressed, or expressed at very low level (low counts are unreliable) (16).

It is very important to keep in mind that there are several unmeasured factors that can influence the expression of a given gene, generating this way new sources of DE. The primary variables are the ones modeled in the analysis; the others can decrease the power to detect associations. Expression heterogeneity (EH) describes patterns of variation produced by these factors, which are usually technical, environmental, demographic or genetic. This EH is very high in human data, particularly in complex systems such as cancer. Normalization techniques are useful to adjust for variation due to laboratory and technical conditions, but not for all sources of systemic expression variation.
The Surrogate variable analysis (SVA) is used to identify, estimate and utilize the components of EH. After the detection of surrogate variables, they are incorporated to the analysis as covariates (20).
Once the analysis is performed, to understand the function, the place where it is performed, and the biological pathways enriched for these DEGs, an enrichment analysis must be performed. What is more, to study a collection of genes involved in the same process is more biologically intuitive and it is easier to understand. Terms statistically over or under represented within a list of interest can be identified (21).

The two most used methods to perform the enrichment are:

1. **ORA.** It assesses if a particular gene set contains many genes differentially expressed. This is the approach that will be used in this study.

2. **Gene set enrichment analysis (GSEA).** It looks if genes of a particular gene set gather at the top or bottom of a list with all the genes ordered by direction and magnitude of expression change.

There are different databases in order to perform the enrichment analysis, such as Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), WikiPathways and Molecular Signatures Database (MSigDB).

1. **GO**. It is the largest source of information about the functions of genes. It ranges from the molecular to the organism level (22,23).

2. **KEGG**. It contains information from a molecular level (24,25).

3. **WikiPathways**. It is an unceasingly updated pathway database. It is a more focused approach than GO (26).

4. **MSigDB**. It is a collection of gene sets including 8 different categories: hallmark gene sets (H), positional gene sets (C1), curated gene sets (C2), motif gene sets (C3), computational gene sets (C4), GO gene sets (C5), oncogenic signatures (C6) and immunogenic signatures (C7) (27).

OBJECTIVES AND HYPOTHESIS

Hypothesis

Pathways related mainly with DNA damage recognition and repair, but also with the anti-inflammatory response, free radical scavenging and apoptosis, will be enriched in genes found to be DE in patients who develop a second tumor after radiotherapy.

Objectives

The main objective of this project is to elaborate a script to find genes differentially expressed in patients who develop a second tumor after radiation therapy and to see which pathways are enriched in those genes using several databases.

There are also several sub-objectives intended to be accomplished:

- To determine whether the limma+voom or DEseq2 strategy leads to a likelier result considering the biological background.
- To look for the specific function of the most significative DEGs and assess their role in tumor development and progression.
- To compare the genes obtained with the DEA to the already published list of genes related with second tumor development

METODOLOGY

The package *tidyverse* (28) was used at several points through the analysis.

Obtaining the information

The data used to perform the analysis was downloaded from the recount2 website, specifically from the subsection TCGA (13). The file downloaded was in a RSE format with gene counts information for all kind of cancers. The information was loaded into R studio where only those samples belonging to normal tissue (at a distance greater than 2cm from tumor margin) and with information about second tumor development and radiotherapy administration were selected. To access to the phenotypic and counts data the package *SummarizedExperiment* (29) was needed.

Exploratory analysis

For the exploratory analysis, a contingency table was constructed in order to determine the number of individuals belonging to the subgroup of patients who had received radiotherapy and developed a second tumor.

The gender, the ethnicity, the primary cancer status, the age at diagnosis, and the first tumor tissue site were graphically represented separately for the mentioned group and the rest of the patients. For that purpose, the package *compareGroups* (30) was used.

Limma+voom pipeline

Normalization and filtering

Using the *edgeR* package (31), a DGEList object was generated by means of the read counts matrix. After that, a filtering of the data with zero or low counts was performed. The following step consisted in normalizing the data contained in the DGEList with the TMM method. The common dispersion (squared coefficient of variation) was also estimated.

Differential expression: limma + voom + SVA

A full model matrix with the adjustment variables and the variables of interest and a null model with only the adjustment variables were created. After that, the voom transformation was applied to the normalized DGEList using as a design the full model matrix and a plot with the mean-variance trend was generated.
In the next step, the number of latent factors was estimated and the *sva* function was used to estimate the surrogate variables. The SVA step was fundamental to correct for the different sample tissue origin. The full and null matrices were then modified including the unknown batch found with SVA.
The FC and standard errors were estimated by fitting a linear model for each gene using as a design the new generated matrices. The empirical Bayes smoothing was then applied to the standard errors.

For the SVA estimation the package *sva* (32) was used, and for the limma+voom pipeline the package *limma* (33).

DESeq2 pipeline

For the DESeq2 pipeline the package *DEseq2* (19) was used. The package *BiocParallel* (34) was also employed to parallelize the work and speed up the process.

Preparation of the data

A DESeqDataSet object was constructed with the count matrix, the phenotypic information and using as a design the interaction between the variables "*radiation therapy*" and *"new tumor event after initial treatment"*.
To reduce the memory size of the object and increase the speed of the transformation and testing functions, a pre-filtering was performed keeping only those rows with at least 10 reads.

Differential expression analysis and SVA

The DE analysis was performed using the function *DESeq*. The resulting object was transformed in order to include the SVA information and the function *DESeq* was used again with this new object.
The result table was generated using the function *results* and plotted using *plotMA*.
The same procedure but shrinking the results with *lfcShrink* was performed in parallel. Given the worse performance compared to the same protocol without shrinking, this step was finally omitted.

Annotation and enrichment analysis

For this section, the package *org.Hs.eg.db* (35) and *clusterProfiler*(26) were used.

First, a gene universe (all the genes) and a list with the DEGs were created with the results obtained from the limma+voom and the DESeq2 pipelines. These lists were then annotated in Entrez IDs. The gene symbol for the gene sets were also obtained.

For DESeq2, the DEGs were also plotted using the package *ggpubr* (36) showing in two different colors the genes up and down regulated. Two lists with the ensembl, entrez and gene symbol of the up and down regulated genes, were then created and mapped using *org.Hs.eg.db* (35).

GO

For the GO enrichment analysis, the package *GOstats* (37) was used.
This is the only enrichment analysis performed for both limma+voom and DESeq2 pipelines. After that, one or the other were chosen based on the results obtained to perform the analysis.
With the results obtained from the enrichment using the function *hyperGTest*, a html report was generated. In DESeq2, two more html were generated with the up and down regulated genes separately.

Using the function *groupGO* from *clusterProfiler*(26) genes were classified based on GO distribution and were graphically represented.
With *enrichGO* the enriched cellular components categories were found. To extract the associations between different biological categories a *cnetplot* and *emapplot* were performed.

KEGG

In this case, the enrichment was performed using the library *KEGG.db* (38) with all the genes, the downregulated genes, and upregulated genes separately. The *enrichKEGG* function from *clusterProfiler* (26) was used, and the results were represented in a bar plot.

WikiPathways

To simplify the script, the package *magrittr* (39) was used in this step.
First, the *Homo sapiens* gmt file containing all the genes in each of the human pathways was downloaded from the website http://data.wikipathways.org/current/gmt/, and opened into R. Then, the analysis was performed for all the genes and the results were then plotted.
To obtain the gene symbols, the package *DOSE* (40) was used.

MSigDB

For this section the package *msigdbr* (27) was used. The enrichment was performed with two different collections, the oncogenic signatures and the immunologic signatures, and with all the genes, the upregulated ones and the downregulated separately.

RESULTS

Cohort characteristics

Among the 104 patients from which there was information about normal tissue read counts, only 4 patients had received radiotherapy and afterwards developed a new tumor (Table 2). The cohort characteristics are specified in the Table 3.

Table 2 Cohort dimensions according to the therapy and second neoplasm presence.

	radiation therapy	
new tumor event	NO	YES
NO	52	35
YES	13	4

***Table 3 Cohort characteristics**. The "YES" column includes the patients that had received radiotherapy and developed a second tumor afterwards. The rest of the patients are represented in the "NO" column. The variables studied are the following ones: the gender of the patients, the ethnicity, the stage of the first tumor, the age at diagnosis of the first tumor, and the first tumor tissue site.*

	NO **N=100**	**YES** **N=4**
Gender		
Female	49 (49.0%)	3 (75.0%)
Male	51 (51.0%)	1 (25.0%)
Ethnicity		
Hispanic or Latino	6 (7.69%)	1 (25.0%)
Not Hispanic or Latino	72 (92.3%)	3 (75.0%)
Group		
Early	70 (78.7%)	0 (0.0%)
Late	19 (21.3%)	4 (100%)
Age at diagnosis	54.9	52.0
Tumor primary site		
Bile Duct	6 (6.0%)	0 (0.0%)
Bladder	3 (3.0%)	0 (0.0%)
Head and Neck	2 (2.0%)	1 (25.0%)
Kidney	28 (28.0%)	0 (0.0%)
Liver	2 (2.0%)	0 (0.0%)
Lung	2 (2.0%)	0 (0.0%)
Nervous System	2 (2.0%)	0 (0.0%)
Pancreas	3 (3.0%)	0 (0.0%)
Prostate	3 (3.0%)	0 (0.0%)
Skin	1 (1.0%)	0 (0.0%)
Stomach	11 (11.0%)	0 (0.0%)
Thymus	2 (2.0%)	0 (0.0%)
Thyroid	34 (34.0%)	3 (75.0%)
Uterus	1 (1.0%)	0 (0.0%)

Limma+voom DE and enrichment analysis

As it can be seen in Table 4, a small number of genes were found to be DE when performing the limma+voom analysis. The pathways enriched for the obtained list of genes are enumerated in the following link:

https://htmlpreview.github.io/?https://github.com/isglobal-brge/master_thesis/blob/master/radiation_cancer/goBPlimma.html

Table 4 Limma+voom DEGs. *ENSEMBL ID, ENTREZ ID and GENE SYMBOL of the DEGs found following the limma+voom pipeline.*

ENSEMBL	ENTREZID	SYMBOL
ENSG00000105048	7138	*TNNT1*
ENSG00000261175	102724344	*LINC02188*
ENSG00000181626	342850	*ANKRD62*
ENSG00000130957	8789	*FBP2*
ENSG00000166869	63928	*CHP2*
ENSG00000114638	7348	*UPK1B*
ENSG00000265933	400643	*LINC00668*
ENSG00000148513	91074	*ANKRD30A*
ENSG00000230438	221756	*SERPINB9P1*
ENSG00000227857	101929626	*LOC101929626*
ENSG00000115112	29842	*TFCP2L1*
ENSG00000249201	101928857	*CTD-3080P12.3*
ENSG00000042832	7038	*TG*
ENSG00000170807	442721	*LMOD2*
ENSG00000185038	339766	*MROH2A*
ENSG00000167080	124872	*B4GALNT2*
ENSG00000236611	102724679	*LINC02556*
ENSG00000231439	100132169	*WASIR2*
ENSG00000115541	3336	*HSPE1*
ENSG00000113070	1839	*HBEGF*
ENSG00000187416	375612	*LHFPL3*
ENSG00000007038	10942	*PTDD21*
ENSG00000229391	3128	*HLA-DRB6*
ENSG00000232629	3120	*HLA-DQB2*

These genes are mainly expressed in clatrin-coated endocytic vesicles and the striated muscle (Figure 4).

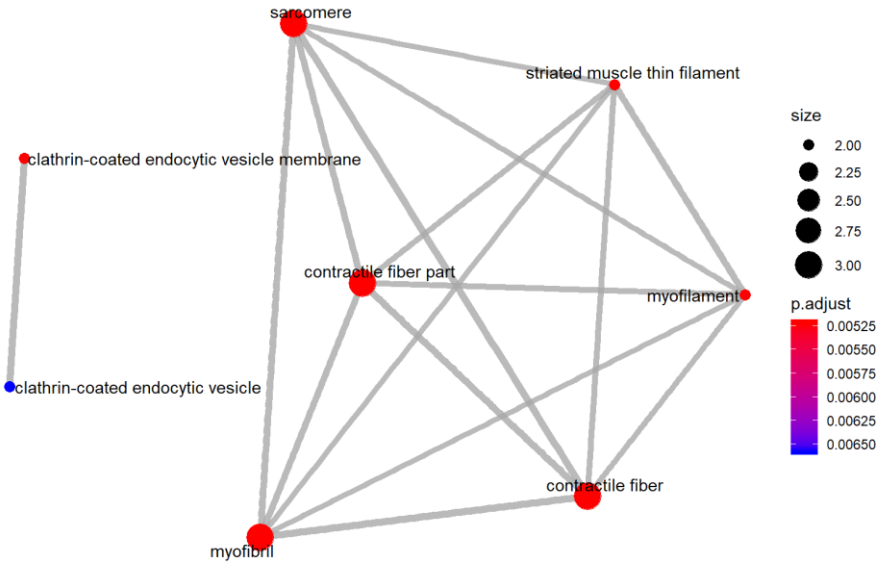

Figure 4 Main pathways enriched in DEGs. *Interaction between pathways, the number of DEGs in each pathway, and the obtained p value for the over-representation test.*

DESeq2 DE and enrichment analysis

After the DEA performed with DESeq2, a big number of genes were found to be DE. These genes were divided into two groups, the upregulated (122 genes) and the downregulated genes (182 genes) (Figure 5).

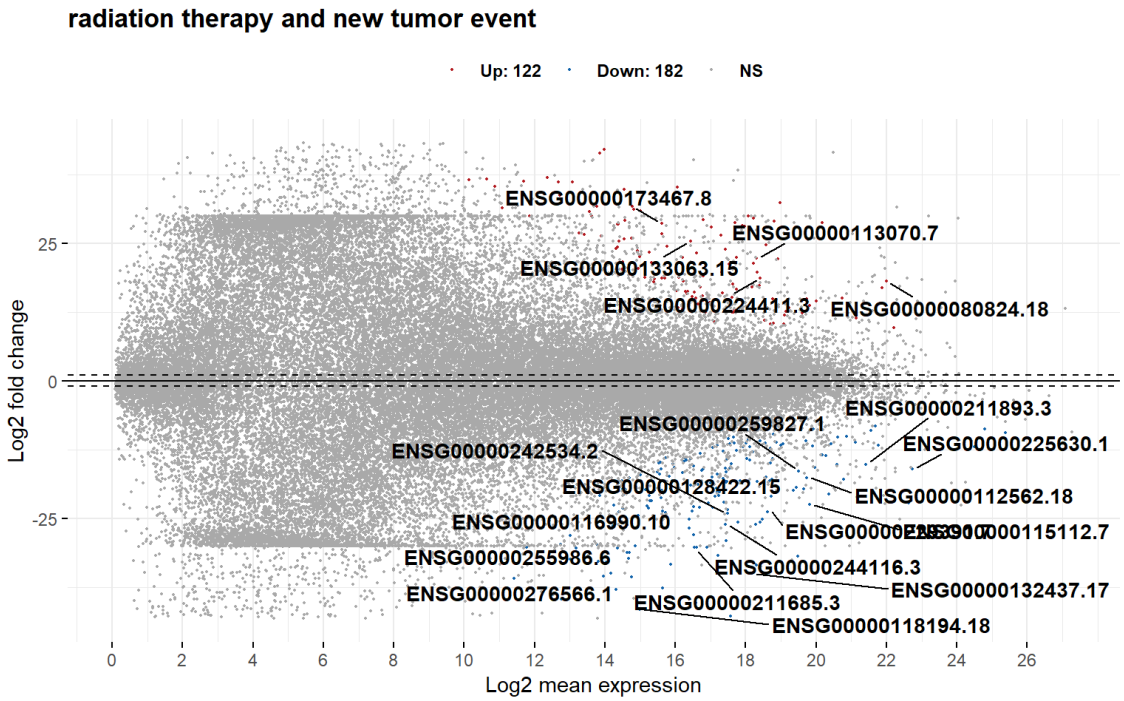

***Figure 5 Upregulated (red) and downregulated (blue) DEGs**. In the x axis the logarithm in base two of the mean expression of each gene is represented, and the y axis shows the log2 FC of the gene.*

Of the top ten genes according to the adjusted p-value, only three were found to be upregulated (red, Table 5), the rest showed a logFC<-1 (blue, Table 5).

***Table 5 Top ten DESeq2 DEGs**. Over-expressed genes are shown in red and downregulated genes in blue. The adjusted p-value and the bibliographic information of each DEG is also specified.*

Gene	Adjusted P value	Bibliographic information
HBEGF	4,07E-13	Enhances cell growth, widely expressed in tumors (41)
TFCP2L1	5,24E-12	Protective role against clear cell renal cell carcinoma (42)
HLA-DRB6	1,78E-11	Antigen presentation, pseudogene (43)
MYCL	1,81E-10	Cell proliferation, differentiation, survival and immune surveillance. Deregulated in a wide number of human cancers (44)
KRT17	2,75E-09	Tumor growth, motility, and invasion, upregulated in cancer (45)
DDC	1,90E-08	Enzymatic synthesis of dopamine, downregulated in laryngeal cancer (46)
TNNT2	3,76E-08	Cell growth, differentiation and proliferation (47)
SMOC2	6,81E-08	Cell-cycle progression, regulates effects of GF, mediates cell growth, proliferation and migration. Downregulated in various tumors (48)
HSP90AA1	7,39E-08	Involved in tumor progression and cancer cell invasion and chemotherapy resistance (49)
AGR3	7,66E-08	Protein folding, over-expressed in several cancers (43)

GO

The pathways enriched in upregulated DEGs are listed in https://htmlpreview.github.io/?https://github.com/isglobal-brge/master_thesis/blob/master/radiation_cancer/go_deseq_ur.html, and the downregulated DEGs in https://htmlpreview.github.io/?https://github.com/isglobal-brge/master_thesis/blob/master/radiation_cancer/go_deseq_dr.html.

According to GO, some of the main biological processes enriched in DEGs are the negative regulation of growth, cell chemotaxis, response to IFN-gamma, detoxification of copper ion and stress response to copper ion (Figure 6).

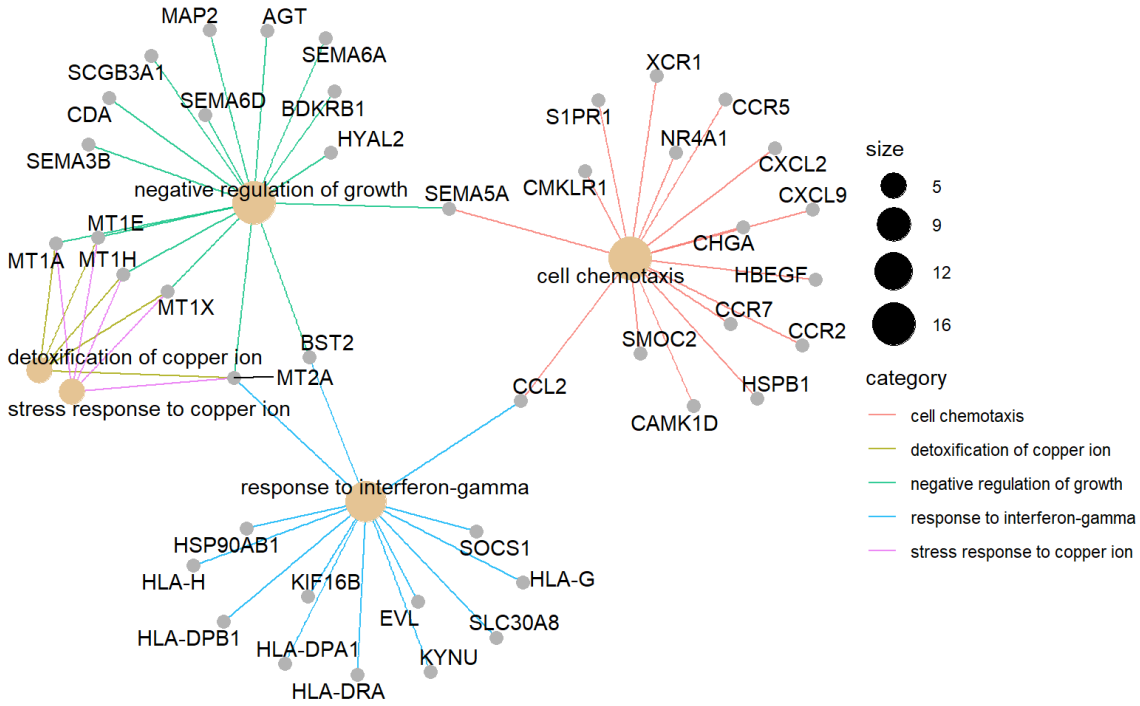

***Figure 6 Main pathways enriched in DEGs**. The main annotation categories are shown in pale brown and the size of the circle depends on the number of genes found in each of them. Each line color represents one of these categories and are used as a link between the DEGs and the category they belong to.*

KEGG

The main pathways enriched in DEGs found by KEGG are listed in https://htmlpreview.github.io/?https://github.com/isglobal-brge/master_thesis/blob/master/radiation_cancer/kegg_deseq_ur.html for the upregulated genes and in https://htmlpreview.github.io/?https://github.com/isglobal-brge/master_thesis/blob/master/radiation_cancer/kegg_deseq_dr.html for the downregulated genes. In this case, some medical conditions that are enriched in those upregulated or downregulated genes are listed, but also some of the categories mentioned in the enrichment analysis performed by GO.

Wikipathways

The categories found with Wikipathways were already found by GO or KEGG, being zinc homeostasis, copper homeostasis, lung fibrosis and allograft rejections (Figure 7).

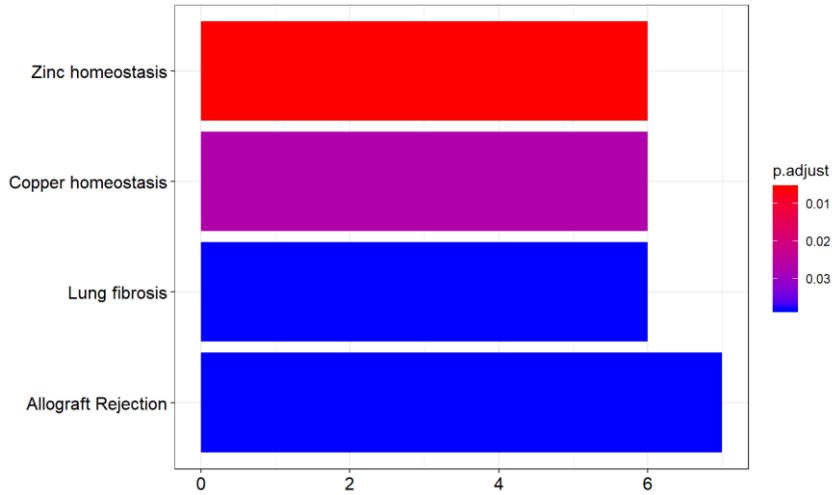

Figure 7 Categories enriched in DEGs according to Wikipathways.

MSigDB

Some of the enriched gene sets for the oncogenic category and the immunologic category are listed in the Tables 6 and 7 respectively. For a list of all the enriched gene sets, the final script can be consulted (https://github.com/isglobal-brge/master_thesis/blob/master/radiation_cancer/Script.md).

***Table 6 Enrichment analysis in the oncogenic category**. Main gene sets enriched in the upregulated and downregulated DEGs. Description extracted from http://software.broadinstitute.org/gsea/index.jsp*

ONCOGENIC SIGNATURES			
State	**Gene set**	**Description**	**P adjusted**
Upregulated	RPS14_DN.V1_UP	Genes up-regulated in CD34+ hematopoietic progenitor cells after knockdown of RPS14	0.01171826
	KRAS.600_UP.V1_DN	Genes down-regulated in four lineages of epithelial cell lines over-expressing an oncogenic form of KRAS gene	0.02819405
Downregulated	LEF1_UP.V1_DN	Genes down-regulated in DLD1 cells (colon carcinoma) over-expressing LEF1	0.006924768
	HINATA_NFKB_IMMU_INF	Immune or inflammatory genes induced by NF-kappaB in primary keratinocytes and fibroblasts.	0.006924768

Table 7 Enrichment analysis in the immunologic category. *Main gene sets enriched in the upregulated and downregulated DEGs.*

IMMUNOLOGIC SIGNATURES			
State	**Gene set**	**Description**	**P adjusted**
Upregulated	GSE10325_BCELL_VS_MYELOID_DN	Genes up-regulated in CD34+ hematopoietic progenitor cells after knockdown of RPS14	3.865116E-07
	GSE19888_ADENOSINE_A3R_INH_PRETREAT_AND_ACT_BY_A3r_VS_TCELL_MEMBRANES_ACT_MAST_CELL_UP	Genes up-regulated in HMC-1 (mast leukemia) cells: incubated with the peptide ALL1 and then treated with Cl-IB-MECA versus stimulation by T cell membranes.	2.541558E-05
Downregulated	GSE20715_0H_VS_48H_OZONE_LUNG_DN	Genes down-regulated in comparison of lung tissue from wild type mice subjected to ozone for 0 h versus that from wild type mice subjected to ozone for 48 h.	0.006608480
	GSE42021_CD24HI_TREG_VS_CD24HI_TCONV_THYMUS_DN	Genes down-regulated in CD42 high cells from thymus: T reg versus T conv.	0.009970054

R script

The script developed to find the DEGs and the enriched pathways and categories can be consulted in the following link:
https://github.com/isglobal-brge/master_thesis/blob/master/radiation_cancer/Script.md

DISCUSSION

In this study, two methods were applied to find DEGs. As mentioned above, the number of DEGs found with the limma+voom method (24 genes) is smaller than the one found with DESeq2 (304 genes). According to the literature, the DESeq2 is a more conservative method (17), so a bigger number of DEGs were expected to be found with the limma+voom method. A more detailed analysis of each step of both methods pipelines should be performed with the same and other datasets to assess their performance.

The enrichment analysis of the DEGs with GO shows that the main categories enriched in the limma+voom DEGs are striated muscle processes and clathrin-coated endocytic vesicles. On the other hand, the enrichment with the DESeq2 subset shows that the main pathways are related to negative regulation of growth, cell chemotaxis, response to interferon-gamma, detoxification of copper ion and stress response to copper ion. Given the considerable bigger number of DEGs found by DESeq2 and the similarity between the enriched pathways found with this subset of genes and the ones biologically expected, the rest of the analysis and the afterwards interpretation was performed only with the subset obtained by DESeq2.

Only five genes found in the DEA matched those found in the literature (Table 1). It is worth noticing that the in-common genes belong to the inflammatory, apoptosis and drug metabolism processes. The patients selected for this study had received radiotherapy or radiotherapy and chemotherapy, which explains the three genes found to be involved in drug metabolism. From these five genes, three were found to be upregulated in the subgroup of patients who had received radiotherapy and developed a second tumor, and two downregulated ("https://github.com/isglobal-brge/master_thesis/blob/master/radiation_cancer/upregulated_genes.csv", "https://github.com/isglobal-brge/master_thesis/blob/master/radiation_cancer/downregulated_genes.csv"). The first of the three upregulated genes is *HSPB1,* which expression is correlated with a poor clinical outcome in cancer given its possible role in cancer cell proliferation and metastasis, and in protection of tumoral cells from apoptosis (43). The second one is *GSTM3*, a glutathione S-transferase involved in detoxification of carcinogens, therapeutic drugs and products of oxidative stress. Variations on this gene can affect the susceptibility to carcinogens and toxins as well as change the efficacy of certain drugs and their toxicity (43). And the third one is *NQO1,* a member of the NADPH dehydrogenase family which prevents the one electron reduction of quinones that results in the production of radical species. Altered expression of this protein has been seen in many tumors (43).

As for the downregulated genes, *TGFB1* regulates cell proliferation, differentiation and growth, and modulates the expression and activation of IFN gamma and TNF alpha. It is frequently upregulated in tumor cells (43) but, contradictory, in this particular case it is

found downregulated. Due to its role in the immune response, the downregulation of this gene may reduce the attack of the immune system to the cancerous cells and promote this way the survival of malignant cells. The other downregulated gene is *CYP*, specifically the variants *CYP2S1*, *CYP11A1* and *CYP21A1P*. *CYP2S1* is involved in the metabolism of toxic and carcinogenic compounds as well as drugs (50), *CYP11A1* seems to be downregulated in many cancer types (51), and *CYP21A1P* is a pseudogene whose role in cancer is still unknown.

As far as the top ten genes are concerned and according to the literature, they all have a potential role in cancer development and progression, or in the immune response. Despite that, some genes that were found to be downregulated in this study are usually over-expressed in cancer (Bibliographic information, Table 5), so a broader research of each gene is needed to understand the effect of the change in the gene expression.

As for the enrichment analysis of the upregulated DEGs, according to GO there is a clear enrichment in pathways related to cell motility and migration (https://htmlpreview.github.io/?https://github.com/isglobal-brge/master_thesis/blob/master/radiation_cancer/go_deseq_ur.html). The enrichment on these pathways may indicate a higher susceptibility of the patients with a genetic predisposition to develop a second tumor after radiotherapy to experience tumoral cell invasion of neighboring tissues (52).

The KEGG analysis revealed also an enrichment in the NOD-like receptor signaling pathway. Alterations in *NOD2* and *NOD1* have been linked to cancer, although the mechanism by which they contribute to the initiation and progression of cancer is not well defined. Specifically, increased expression of these receptors has been reported in head and neck cancer (53). With KEGG, an over-expression of genes related to the MAPK signaling pathway was also detected. It has been widely studied the relationship of activating mutations of *RAS* (component of the MAPK pathway) and cancer; moreover, other components of the pathway, such as RAF, MEK and ERK, have also been linked to cancer (54). What is more, the genes activated by this signaling pathway have an important role in the stimulation of cell proliferation and differentiation, as well as in apoptosis regulation (55), but also in the response to extracellular stresses and inflammation-associated cancer development (56).

On the other hand, pathways related with the immune system, response to stress, negative regulation of growth and detoxification of copper ion were enriched in downregulated DEGs. According to GO and KEGG, the affected immune processes in patients developing a second tumor after radiotherapy are response to IFN-gamma and other cytokines, chemotaxis, antigen processing and presentation, and in general, dysregulation of all immune processes (https://htmlpreview.github.io/?https://github.com/isglobal-brge/master_thesis/blob/master/radiation_cancer/go_deseq_dr.html). The adaptive immune response mediated by cytotoxic T cells (CTLs) plays a very important role in

cancer secreting IFN gamma, a cytokine that generates an acute inflammation enabling the expansion of CTLS, the destruction of the tissue and a potential control or elimination of cancer. If there is a dysregulation in the processing and presentation of antigens, T cells are not activated, and this acute inflammation process is not stimulated, leading to a worst cancer outcome (57). As for the response to stress, it plays a crucial role in cancer development through several mechanisms. One of them, is the autophagy process triggered in response to stress stimuli (58). A lower expression of genes involved in this process, may be an indicator of a bad stress response, that could lead to an enhancement of tumor development. In relation to the genes involved in detoxification or stress response to copper ion, they are also involved in negative regulation of cell growth and response to INF-gamma (Figure 6), so presumably, these pathways are enriched in DEGs only by chance, but a deeper research should be performed.

As for the enrichment analysis performed with the downregulated DEGs to the oncogenic signatures category of MSigDB, the gene subset LEF1_UP.V1.DN corresponds to a set of downregulated genes in a colorectal adenocarcinoma cell line that overexpresses *LEF1* (Table 6). It is known that this gene is not expressed in normal adult colon tissue, but it is activated in colon carcinogenesis. What is more, the aberrant expression of *LEF1* is implicated in tumorogensis and cancer cell proliferation, migration and invasion (59). So, this set of downregulated genes may be a biomarker of *LEF1* expression.

The other relevant gene set found in this category with the downregulated DEGs, is the HINATA_NFKB_IMMU_INF (Table 6). This set is composed of immune or inflammatory genes induced by NF-kappa B in primary keratinocytes and fibroblasts. NF-kappa B induce proinflammatory and antiapoptotic genes, so the decrease of genes induced from NF-kappa B could be an indicator of a worst immune performance against cancerous cells.

However, given that only four individuals belonged to the cohort of patients who developed a second tumor after radiotherapy, further studies should be performed to validate the results.

CONCLUSIONS

Some of the genes and pathways found in this analysis have a clear role in second tumor development and progression. There is a clear enrichment in pathways related to cell motility and migration, very important processes when it comes to tissue invasion. The NOD-like receptor signaling pathway and MAPK pathways are also enriched in DE upregulated genes, both linked to cancer. Processes such as regulation of cell proliferation were also found to be dysregulated according to MSigDB.

As for the downregulated DEGs, there is an enrichment in pathways related to the immune response. Processes such as the response to cytokines, chemotaxis and antigen processing and presentation are enriched in these genes, which leads to a general disruption of the immune system and particularly, in the inflammatory response. But these are not the only pathways enriched in downregulated genes, the response to stress, which plays an important role in cancer development, was also found to be dysregulated.

All the mentioned pathways have a clear role in tumoral processes, but their specific function in second tumor development after radiotherapy is unknown. Further studies should be performed to assess their role and validate the results. What is more, given the small size of the cohort of this study, the obtained results must be also confirmed using a bigger number of individuals.

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin. 2019;69(1):7–34.
2. Mortezaee K, Najafi M, Farhood B, Ahmadi A, Potes Y, Shabeeb D, et al. Modulation of apoptosis by melatonin for improving cancer treatment efficiency: An updated review. Life Sci. 2019;228(May):228–41.
3. Barnett GC, Kerns SL, Noble DJ, Dunning AM, West CML, Burnet NG. Incorporating Genetic Biomarkers into Predictive Models of Normal Tissue Toxicity. Clin Oncol. 2015;27(10):579–87.
4. Desouky O, Ding N, Zhou G. Targeted and non-targeted effects of ionizing radiation. J Radiat Res Appl Sci. 2015;8(2):247–54.
5. De Ruysscher D, Niedermann G, Burnet NG, Siva S, Lee AWM, Hegi-Johnson F. Radiotherapy toxicity. Nat Rev Dis Prim. 2019;5(1).
6. Morton LM, Onel K, Curtis RE, Hungate EA, Armstrong GT. The Rising Incidence of Second Cancers: Patterns of Occurrence and Identification of Risk Factors for Children and Adults. Am Soc Clin Oncol Educ B. 2014;34:e57–67.
7. Bernstein JL, Thomas DC, Shore RE, Robson M, Jr JDB, Stovall M, et al. Public Access NIH Public Access. Eur J Cancer. 2014;49(14):2979–2985.
8. Kerns SL, Dorling L, Fachal L, Bentzen S, Pharoah PDP, Barnes DR, et al. Meta-analysis of Genome Wide Association Studies Identifies Genetic Markers of Late Toxicity Following Radiotherapy for Prostate Cancer. EBioMedicine. 2016;10:150–63.
9. Bhatia S. Genetic variation as a modifier of association between therapeutic exposure and subsequent malignant neoplasms in cancer survivors. Cancer. 2015;121(5):648–63.
10. Devi KR, Ahmed J, Narain K, Mukherjee K, Majumdar G, Chenkual S, et al. DNA Repair Mechanism Gene, XRCC1A (Arg194Trp) but not XRCC3 (Thr241Met) Polymorphism Increased the Risk of Breast Cancer in Premenopausal Females: A Case–Control Study in Northeastern Region of India. Technol Cancer Res Treat. 2017;16(6):1150–9.
11. Zhao M, Chen P, Dong Y, Zhu X, Zhang X. Relationship between Rad51 G135C and G172T variants and the susceptibility to cancer: A meta-analysis involving 54 case-control studies. PLoS One. 2014;9(1).
12. Costa-Silva J, Domingues D, Lopes FM. RNA-Seq differential expression analysis: An extended review and a software tool. PLoS One. 2017;12(12):1–18.
13. Collado-Torres L, Nellore A, Kammers K, Ellis SE, Taub MA, Hansen KD, et al. recount2: A multi-experiment resource of analysis-ready RNA-seq gene and exon count datasets [Internet]. Reproducible RNA-seq analysis using recount2. 2017 [cited 2019 Mar 20]. Available from: https://jhubiostatistics.shinyapps.io/recount/
14. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model

analysis tools for RNA-seq read counts. Genome Biol. 2014;15(2):R29.
15. Evans C, Hardin J, Stoebel DM. Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. Brief Bioinform. 2018;19(5):776–92.
16. Huang H-C, Niu Y, Qin L-X. Differential Expression Analysis for RNA-Seq: An Overview of Statistical Methods and Computational Software. Cancer Inform. 2015;14(Suppl 1):57.
17. Seyednasrollah F, Laiho A, Elo LL. Comparison of software packages for detecting differential expression in RNA-seq studies. Brief Bioinform. 2013;16(1):59–70.
18. Yoon S, Nam D. Gene dispersion is the key determinant of the read count bias in differential expression analysis of RNA-seq data. BMC Genomics. 2017;18(1):1–11.
19. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
20. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genet. 2007;3(9):1724–35.
21. Tipney H, Hunter L. An introduction to effective use of enrichment analysis software. Hum Genomics. 2010;4(3):202–6.
22. The Gene Ontology Consortium. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. 2019;47(D1):D330–8.
23. The Gene Ontology Consortium, Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. Gene Ontology : tool for the unification of biology. Nat Genet. 2011;25(1):25–9.
24. Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. New approach for understanding genome variations in KEGG. Nucleic Acids Res. 2019;47:D590–5.
25. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: New perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017;45(D1):D353–61.
26. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. Omi A J Integr Biol. 2012;16(5):284–7.
27. Dolgalev I. msigdbr: MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format. R Packag version 621. 2018;
28. Wickham H. tidyverse: Easily Install and Load the “Tidyverse.” R Packag version 121. 2017;
29. Morgan M, Obenchain V, Hester J, Pagès H. SummarizedExperiment: SummarizedExperiment container. R Packag version 1120. 2018;
30. Subirana I, Sanz H, Vila J. Building Bivariate Tables: The compareGroups Package for R. J Stat Softw. 2014;57(12):1–16.
31. McCarthy D, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res.

2012;40:4288–97.
32. T. Leek J, Evan Johnson W, Parker S. H, J. Fertig E, E. Jaffe A, D. Storey J, et al. sva: Surrogate Variable Analysis. R Packag version 3301.
33. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7).
34. Morgan M, Obenchain V, Lang M, Thompson R, Turaga N. BiocParallel: Bioconductor facilities for parallel evaluation. R Packag version 1166. 2019;
35. Carlson M. org.Hs.eg.db: Genome wide annotation for Human. R Packag version 370. 2018;
36. Kassambara A. ggpubr: "ggplot2" Based Publication Ready Plots. R Packag version 02. 2018;
37. Falcon S, Gentleman R. Using GOstats to test gene lists for GO term association. Bioinformatics. 2007;23(2):257–8.
38. Carlson M. KEGG.db: A set of annotation maps for KEGG. R Packag version 323.
39. Bache SM, Wickham H. magrittr: A Forward-Pipe Operator for R. R Packag version 15. 2014;
40. Guangchuang Y, Li-Gen W, Guang-Rong Y, Qing-Yu H. DOSE: an R/Bioconductor package for Disease Ontology Semantic and Enrichment analysis. Bioinformatics. 2015;31(4):608–9.
41. Ongusaha P, Kwak J, Zwible A, Macip S, Higashiyama S, Taniguchi N, et al. HB-EGF is a potent inducer of tumor growth and angiogenesis. Cancer Res. 2004;64(15):5283–90.
42. Kotarba G, Krzywinska E, Grabowska AI, Taracha A, Wilanowski T. TFCP2/TFCP2L1/UBP1 transcription factors in cancer. Cancer Lett. 2018;420:72–9.
43. Weizmann institute of science, LifeMap sciences. GeneCards: The Human Gene Database [Internet]. [cited 2019 Jun 10]. Available from: https://www.genecards.org/
44. Chen H, Liu H, Qing G. Targeting oncogenic Myc as a strategy for cancer treatment. Signal Transduct Target Ther. 2018;3(1):1–7.
45. Chivu-Economescu M, Dragu DL, Necula LG, Matei L, Enciu AM, Bleotu C, et al. Knockdown of KRT17 by siRNA induces antitumoral effects on gastric cancer cells. Gastric Cancer. 2017;20(6):948–59.
46. Patsis C, Glyka V, Yiotakis I, Fragoulis EG, Scorilas A. l-DOPA Decarboxylase (DDC) Expression Status as a Novel Molecular Tumor Marker for Diagnostic and Prognostic Purposes in Laryngeal Cancer. Transl Oncol. 2014;5(4):288–96.
47. Johnston JR, Chase PB, Pinto JR. Troponin through the looking-glass: emerging roles beyond regulation of striated muscle contraction. Oncotarget. 2018;9(1):1461–82.
48. Huang XQ, Zhou ZQ, Zhang XF, Chen CL, Tang Y, Zhu Q, et al. Overexpression of

SMOC2 attenuates the tumorigenicity of hepatocellular carcinoma cells and is associated with a positive postoperative prognosis in human hepatocellular carcinoma. J Cancer. 2017;8(18):3812–27.

49. Xiao X, Wang W, Li Y, Yang D, Li X, Shen C, et al. HSP90AA1-mediated autophagy promotes drug resistance in osteosarcoma. J Exp Clin Cancer Res. 2018;37:201.
50. Saarikoski ST, Rivera SP, Hankinson O, Husgafvel-Pursiainen K. CYP2S1: A short review. Toxicol Appl Pharmacol. 2005;207(2 SUPPL.):62–9.
51. Fan Z, Wang Z, Chen W, Cao Z, Li Y. Association between the CYP11 family and six cancer types. Oncol Lett. 2016;12(1):35–40.
52. Paul CD, Mistriotis P, Konstantopoulos K. Cancer cell motility: Lessons from migration in confined spaces. Nat Rev Cancer. 2017;17(2):131–40.
53. Saxena M, Yeretssian G. NOD-like receptors : master regulators of inflammation and cancer. 2014;5(July):1–16.
54. Germann UA, Furey BF, Markland W, Hoover RR, Aronov AM, Roix JJ, et al. Targeting the MAPK Signaling Pathway in Cancer : Promising Preclinical Activity with the Novel Selective ERK1 / 2 Inhibitor BVD-523 (Ulixertinib). 2017;2351–64.
55. ZHANG W, LIU HT. MAPK signal pathways in the regulation of cell proliferation in mammalian cells. Cell Res. 2006;12(1):9–18.
56. Huang P, Han J, Hui L. MAPK signaling in inflammation-associated cancer development. Protein Cell. 2010;1(3):218–26.
57. Disis ML. Immune Regulation of Cancer. Clin Oncol. 2010;28(29):4531–8.
58. Zelenka J, Konco M, Ruml T. Targeting of stress response pathways in the prevention and treatment of cancer. 2018;36(August 2017):583–602.
59. Santiago L, Daniels G, Wang D, Deng F, Lee P. Wnt signaling pathway protein LEF1 in cancer , as a biomarker for prognosis and a target for treatment. 2017;7(6):1389–406.