IDENTIFICATION OF CRITICAL PATHWAYS ALTERED BY RADIATION EXPOSURE AND DRUG TARGET ANALYSIS

By
Christopher Betzing

A THESIS

Presented to the Department of Bioinformatics & Computational Biology
and the Oregon Health & Science University School of Medicine
in partial fulfillment of the requirements for the degree of
Master of Science
August 2013

School of Medicine
Oregon Health & Science University

Certificate of Approval

This is to certify that the Master's thesis of

Christopher R. Betzing

"Identification of Critical Pathways Altered by Radiation Exposure and Drug Target Analysis"

Has been approved

Dr. Eilis Boudreau, Thesis Advisor

Dr. Shannon McWeeney, Committee Member

Dr. Charles Thomas, Committee Member

Table of Contents

List of Tables

Acknowledgements

My thesis chair, Dr. Eilis Boudreau, provided me with the background and training necessary to acquire a writing style that enabled me to concisely develop my thesis work. Dr. Boudreau made herself available to answer any questions that arose and helped me to guide my specific aims and thesis write up.

I would like to express my deep appreciation for the support and mentorship of my thesis committee. Dr. Shannon McWeeney shared her expertise with me on all the technical aspects of my thesis. Dr. McWeeney took time out of her extremely busy schedule to provide me with exceptional guidance and encouragement. Dr. Mcweeney's expertise and oversight helped me to develop the technical skills necessary to complete this study.

In addition, Dr. Charles Thomas has been a major guiding force in my academic career for several years and this thesis would not have been possible without his support. He has provided me with in depth experiences in the field of radiation medicine and access to materials that strengthened my background and enabled me to perform this study.

Dr. Wayne Zundel's help in providing with the original concept of my thesis and introducing me to the various database repositories is greatly appreciated.

I would also to thank the staff and students at the OHSU Bioinformatics Department for providing me with an excellent educational environment.

Abstract

Ionizing radiation (IR) is commonly used in the treatment of cancer through radiation therapy. In addition, exposure to IR is a vital safety risk that nuclear workers constantly face. The effect of IR on human cell, both normal and malignant, is biologically significant. The purpose of this study is three fold: 1) to determine which genes are differentially expressed in cells exposed to radiation, 2) to discover critical pathways affected by these genes, 3) and identify drugs that significantly target the differentially expressed pathways. The raw data for this study consists of 5 Gene Expression Omnibus (GEO) data sets that contain microarray expression data for normal or cancerous cell lines exposed to IR. A secondary analysis on these data sets utilizes various Bioconductor R packages in the analysis of genes that are differentially expressed and the identification of critical pathways. Numerous genes and pathways were found to be differentially expressed in this analysis. In addition, six pathways were differentially expressed across all cancer cell data sets. No pathways that were differentially expressed across all normal cell data sets. The critical Reactome pathways that were significant across all cancer cell data sets are the following: 1) "Cell Cycle, Mitotic", 2) "DNA Replication", 3) "Mitotic M-M/G1 phases", 4) "AKT phosphorylates targets in the cytosol", 5) "M Phase", 6) "Mitotic Prometaphase". A drug pathway analysis was performed on these pathways and the genes that were differentially expressed and members of these pathways. The drug pathway analysis found that six significant gene targets. These six gene targets are: POLA1, POLA2, PLK1, CENPE, AURKB, CDKN1B, and RB1. The results of our study are significant because they allow

for the identification of potential genes and pathways that could be used as genetic markers in radiation therapy. In addition, the identification of drug targets will allow for testing to determine if drugs that target these genes affect the radioresistance or radiosensitivity of cancerous cells upon exposure to IR.

Introduction

1.1.1 Ionizing Radiation

Ionizing radiation is a form of radiation that ionizes atoms along its trajectory and can deposit energy on the medium it passes through [1]. Exposure to ionizing radiation (IR) is known to increase an individual's chance of developing cancer and can lead to severe medical complications and death [2]. Mitigating the risk for exposure of nuclear workers or the general public to IR is extremely important. Previous studies have shown that individuals who work with radioactive material are exposed to higher levels of radiation and have a higher chance to develop medical complications [3]. Ionizing radiation can affect human cells through direct or indirect effects [1]. Direct effects occur when the radiation directly reacts with a cell [1]. The most damaging example of a direct effect would be radiation interacting with the DNA of a cell [1]. The other method of interaction is called indirect effects and consists of the creation of free radicals through the interaction of the radiation with water or other compounds in the cell [1]. These free radicals can damage the cell and cellular DNA [1]. The direct or indirect interaction of radiation with cellular DNA can cause double or single stranded breaks in the DNA [1]. The damages caused by IR through these effects, particularly double stranded breaks as they are more difficult to repair, can lead to mutations or cell death [1]. Therefore, it is important to note the potential of IR exposure to cause the development of cancer [1,2,3]. Finally, there are many different medical applications of IR that exist, including the exploitation of IR characteristics to treat cancer and other medical conditions [1,4,5].

1.1.2 Medical Applications of Ionizing Radiation

A common practice in cancer treatment is radiation therapy, which utilizes different forms of IR to treat patients [4]. A large percentage of cancer patients receive radiation therapy at some point during their treatment [1]. Radiation therapy is widely practiced because it has been shown to increase the survival rate of some cancer patients [5]. There are two different techniques for delivering radiation therapy. The first technique involves the delivery of an external beam of radiation to the cancerous tumor [4]. The second technique is brachytherapy, which is the placement of a radioactive source next to the tumor (intracavitary) or directly inside the gross tumor (interstitial). Furthermore, brachytherapy can be delivered via low-dose rate (permanent or temporary) or high -dose rate (usually temporary) techniques. Radiation therapy involves the targeting of cancerous tissue with IR in an effort to kill the cancerous cells [4]. It is well known that most types of cancer are significantly more sensitive to radiation damage than normal cells [4]. However, the use of radiation therapy has some serious side effects. The main obstacle in the use of radiation therapy is the damage to normal tissue that occurs. Inevitably, some healthy cells located near the targeted cancerous tumor are exposed to ionizing radiation [4]. This can lead to the destruction of healthy tissue and the increased risk of the patient developing a secondary cancer in the future [6]. Therefore, due to the serious medical conditions that can arise from exposure to IR and the medical applications of IR, it is imperative that novel therapeutic techniques are developed in order to mitigate the negative effects of IR exposure to healthy tissue and increase the effect of IR on cancerous tissue in order to increase the effectiveness of radiation therapy.

1.2.1 Data Utilized

The raw data utilized in this study consisted of gene expression data from GEO datasets that contain Affymetrix microarray analysis data. GEO is a public collection of microarray data that is compiled from past experiments. Previous studies have been performed utilizing gene microarray gene expression data in the field of radiation medicine. These studies have found a number of genes that are differentially expressed in human tissue exposed to radiation [2,7,8,10,11,12,13]. For example, the original studies that utilized the GEO data sets that were selected for this study found that genes that were critical to DNA damage and replication were differentially expressed [14,15,16,17,18]. In addition, each of these studies found that p53 and genes associated with p53 expression were differentially expressed. This result was also observed in this study and is expected given the significant role that p53 expression plays in cell apoptosis [11,17,18].

The analyses performed in the referenced studies provide evidence that shows gene and pathway expression analysis can be performed to determine a list of candidate genes and pathways that could potentially contribute to variable radiosensitivity or radioresistance in human cancer and normal cell lines. However, few studies have performed a comprehensive meta-analysis utilizing multiple microarray datasets. In addition, most of the previous studies have focused primarily on the expression analysis of known metabolic pathways and genes that are widely known to be associated with cancer, such as p53 [11]. The few studies that have performed a comprehensive meta-analysis to identify differential pathway expression in human tissue exposed to radiation have focused on individual cell types [19]. The analysis performed in this thesis is novel because of its meta-analysis of multiple microarray gene expression data sets from different cell types. In addition, the experimental analysis that was performed allowed for the identification of biological pathways that have not been previously associated with changes in radioresistance or radiosensitivity.

1.2.2 Gene Expression Analysis

The analysis of gene expression in Affymetrix microarrays is a well-known process [20]. The basic premise for performing a gene expression analysis with microarray technology is that the level of a specific RNA in a cell is positively correlated with the expression of the gene coding for that RNA [20,21,22,23]. The Affymetrix microarrays consist of oligonucleotide probes that are complimentary to a specific RNA sequence [20,23]. The three most common types of Affymetrix microarray platforms are the HGU 133 Plus 2.0, Human Exon 1.0 ST, and Human gene 1.0 ST arrays. The probes in the HGU 133 Plus 2.0 array are 25 nucleotides in length and contain a complimentary sequence that perfectly matches the target RNA [15,23,24,25]. However, to account for noise and random binding, an oligonucleotide probe that contains 1 altered base pair, usually at the 13th nucleotide, is created [20,24,25]. In contrast, the Human Exon 1.0 ST and Human Gene 1.0 ST arrays use probes that do not match with any human RNA as negative control probes [24,25]. Therefore, the intensity of the negative control probes for each array can be utilized to filter the microarray data in order to remove non-binding probes.

The use of gene expression data is typically analyzed by using the probe intensity values to compare the expression of a specific gene between a treatment and control group [2, 26]. If the difference of the intensity value between the two groups is significant, then it can be concluded that the gene is differentially expressed in the treatment group [2,26]. If the intensity of a probe is statistically higher in one group, then its corresponding gene is concluded to be expressed at a higher level than the other group. Likewise, if the intensity of the probe is lower than the other group then the gene is concluded to be expressed at a lower level than the other group. If the difference between the intensity of a specific probe is not significantly different between the two groups then it is concluded that the probe's corresponding gene is not differentially expressed between the two groups [26]. One issue with this form of testing with microarray data is correcting for multiple testing. Multiple testing is an issue because there are thousands of probes corresponding to a specific gene on a microarray. Therefore, it is necessary to perform a correction for multiple testing to lower the occurrences of false positives.

The probe IDs and their corresponding intensity values are reported in the GEO datasets, which can be utilized for secondary analysis. GEO data sets typically contain replicates for their control and treatment groups. This is important because performing a replicate microarray analysis with the replicate data sets for each group helps to account for any sources of error that may affect the data. The secondary meta-analysis of gene expression data from multiple GEO datasets can introduce a possible bias from batch effect, which should be addressed. [21,27] The batch effect bias may occur due to the use of data from different experiments. Normalization is one common method that can be utilized to remove batch effect and other sources of bias. The process of normalization will allow for the removal or mitigation of various sources of experimental bias [21]. The Robust Multi-Array Analysis (RMA) is an appropriate method for normalizing Affymetrix microarray data and correcting for batch effect and other sources of bias [24,25]. The use of quality control, background correction, and normalization methods are critical for the mitigation of potential biases that may occur during a gene expression analysis and are vital steps in any analysis [22,24,25,28].

1.2.3 Annotation

The use of different Affymetrix microarray platforms in a meta-analysis can cause some issues regarding gene annotation. Therefore, identifying which gene is associated with a specific piece RNA, and the RNA that is complementary to a probe, is very important. The process for identify the genes that are represented by specific probe IDs is called annotation [25,29]. This process is extremely important because faulty annotation can lead to the misidentification of differentially expressed genes and faulty data [25,29]. Therefore, for a gene expression analysis to ensure that each probeset is annotated to the correct gene, an efficient annotation method is needed. One effective tool for properly annotating gene expression data from various Affymetrix platforms is the Bioconductor package Annmap. This package matches the various probeset IDs from the microarray to specific genes by comparing the probe sequences with a reference genome [1]. Any unrealizable probesets, such as probe sets that contain probes that do not map to the genome or have multiple targets, are removed [1]. This is an effective means of annotation because it ensures that the probeset IDs that are determined to be differentially expressed are annotated to the correct gene. Without proper annotation, a gene expression analysis could misidentify genes that are differentially expressed through the incorrect mapping of a probeset to its corresponding gene.

1.2.4 Pathway Analysis

The data obtained from the identification of genes that are differentially expressed has previously been successfully utilized to determine which pathways are differentially expressed between a treatment and control group [30]. For the pathway expression analysis of microarray data, the Bioconductor package Graphite and the Reactome pathway database are useful tools. The Reactome database consists of pathways for various biologically processes [31]. The base unit of the pathways in the Reactome database is a connection of various reactions and a pathway comprised of the interactions of the base units [31]. The Reactome database is often used because the development of pathways with a base unit of biological reactions allows for a comprehensive pathway analysis that accounts for the interactions of the various gene products [31].

For a pathway analysis to be performed, a gene expression analysis needs to be performed in order to determine which genes to test for pathway membership. The biological products of these genes are utilized to determine their membership in the Reactome pathways and to perform the pathway analysis [29,31]. The Bioconductor package GRAPHITE offers the use of the method of Signaling Pathway Impact Analysis (SPIA) for pathway analysis [29]. The SPIA pathway analysis utilizes the fold change of the differentially expressed genes and the interactions between various gene products to determine the amount of perturbation that occurs in a specific pathway [29,32]. The results of this pathway analysis are of great significance because the statistical analysis performed is influenced by knowledge of the biological processes and interactions of the gene products for the differentially expressed genes that are being analyzed [29,32]. However, one common issue in pathway analysis is the concern of multiple testing[32]. A pathway analysis performs many statistical tests for each pathway, which can lead to a high number of false positives. Therefore, a form of correction for multiple testing is an important aspect of pathway analysis. The use of the analysis of the perturbation of the biological pathways along with the pathway enrichment analysis and multiple testing is a vital aspect in ensuring that a pathway analysis has a high significance.

1.2.5 Drug Target Analysis

MetaDrug is a comprehensive knowledgebase that contains information regarding protein-protein interaction and the reactions of various gene productions to create an interaction network in the context of drug toxicology [33]. This database can utilize input data in the form of genes that are differentially expressed and pathways that are significantly enriched with these gens [33]. One goal of this study is to identify pharmaceutical compounds that target genes and pathways that are differentially expressed in human cells exposed to radiation. Therefore, the MetaDrug database was utilized in order to identify these pharmaceutical compounds through an analysis of the MetaDrug interaction network using the microarray data consisting of differentially expressed genes and pathways that were obtained in our analysis. The use of MetaDrug is vital to the identification of novel therapeutic targets in the field of radiation medicine as the identification of significant pharmaceutical compounds that target pathways and genes that are differentially expressed upon exposure to radiation can allow for the identification of currently existing drugs that could be utilized in the field of radiation oncology or radiation protection.

1.3.1 Previous Studies

Microarray array technology has been used to analyze the biological response of human cells to exposure to ionizing radiation. These past studies have found that a large fraction of the genes that are differentially expressed upon exposure to IR are involved with cell cycle regulation, apoptosis, and DNA repair pathways [34]. However, one major concern in performing a gene expression analysis is the issue of false positives and multiple testing [34]. However, this study will perform statistical corrections for the issue of multiple testing. In addition to studies utilizing one dataset, some meta-analyses studying the effects of IR on human cells have been performed using microarray technology [19]. One of the most meta-analyses most relevant to this study by Kim et al. involved the exposure of the NCI-60 cancer cell lines to 2 Gy of IR [19]. This study identified genes that were differentially expressed between each of the cancer cell lines upon exposure to IR and in agreement with past studies found that most of the differentially expressed genes were members of cell cycle regulation and DNA replication pathways [19]. In addition to the field of radiation medicine and the treatment of cancer with IR, studies have also been conducted in the field of radiation protection using microarray technology. One study by Fachin et al. performed a gene expression analysis on nuclear workers that were exposed to IR during their normal work routine [35]. This study also found that the genes that were differentially expressed were members of pathways that centered on cell cycle regulation and stress response/DNA repair [35].

These previous studies have shown the feasibility of using microarray technology to analyze the biological response to IR in human cells. In addition, the various fields covered in these previous studies ranging from radiation treatment to radiation protection have shown that our current study has a great significant in the field of radiation medicine as a whole. In contrast to the past research that has been performed in this area, our study seeks to not only perform a gene expression analysis, but to perform a pathway analysis and a drug pathway target analysis. This will allow for the determination of which genes are differentially expressed upon exposure to IR, which pathways these genes are significantly perturbing, and which drugs significantly target the differentially expressed pathways.

1.3.2 Rationale and Objective

The long term goal of this study is to identify critical pathways with altered gene expression in human cells exposed to radiation and to identify pharmaceutical compounds that target these pathways. This goal is based on our central hypothesis that changes in gene expression in cells exposed to radiation is a common response in all cells and the identification of genes and pathways that have altered expression when exposed to radiation will allow for the identification of candidate drugs that can be used for novel approaches in radiation medicine to treat, utilize, or moderate the effects of exposure to IR. The specific aims of this study are the following:

1. Determine which genes are differentially expressed in human cells after exposure to radiation through the secondary analysis of GEO datasets.
2. Identify which biological pathways are significantly enriched with the genes that are differentially expressed after exposure to ionizing radiation.
3. Identify pharmaceutical drugs that target significantly target the biological pathways that are differentially expressed in human cells exposed to ionizing radiation.

Methods

2.1.1 Data

The data for this project consists of GEO datasets containing Affymetrix microarray expression data. These datasets contain microarray data for a treatment group and a control group. The treatment group consists of the gene expression data for the human cell line exposed to radiation. The control group consists of the gene expression data for the same cell line without exposure to radiation. In addition, there was a set criterion that was used to select the data sets for this study, which is outlined below.

The GEO datasets used in this study were selected based on the following criteria:

1. The microarray expression data was collected from human cells that received a dose of 2Gy, 5Gy, or 10Gy.
2. The RNA post radiation extraction time must range from 2 to 8 hours.
3. The control group for each data set must contain the same cell lines as the treatment group and be collected using the same laboratory techniques.
4. The treatment and control groups in each dataset must contain at least 1 replicate (2 different microarray expression data for each of the treatment and control group).

These dose values for the criteria were chosen due to their clinical significance as the use of 2Gy, 5Gy, or 10Gy dose fractions is common in radiation therapy [36]. In addition, the RNA extraction time post radiation values for the datasets ranging from 2 hours to 8 hours was chosen because these times are relatively close. The use of expression data with small differences in DNA extraction time has been found to allow for an accurate gene expression analysis and should have little negative effect on the data quality [37]. A control group was required for each analysis because an analysis of the same cell line without exposure to radiation is needed to determine which genes are differentially expressed when the same cell line is exposed to radiation. Data sets were required to contain at least 1 replicate because this would help to identify any bias or sources of error in the data.

One of the 5 GEO data sets contains microarray data for normal cell lines and cancerous cell lines. Therefore, this data set was split and the normal and cancerous sub-datasets were treated as individual data sets. The 5 GEO datasets utilized in this project are also divided into two groups. These two groups are labeled Control and Treatment. The control group consists of four datasets that contain gene expression data for normal cells exposed to radiation. The second group will consist of the two data sets that contain data for cancerous tissue exposed to radiation.

A list of the GEO datasets, cell types, radiation exposure in Gray (Gy), post radiation RNA extraction time, and group assignment can be seen in table 1.

GEO Dataset	Platform	Cell Type (All Human)	Dose Received (Gray)	Post Radiation RNA Extraction Time	Group
GSE26841	GPL5175 [HuEx-1_0-st] Affymetrix Human Exon 1.0 ST Array [transcript (gene) version]	Primary Fibroblast	10Gy	4 hours	Normal (1)
GSE37668	GPL6244 [HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]	Embryonic Human Fibroblasts	2Gy	2,4,8 hours	Normal (1)
GSE41840	GPL5175 [HuEx-1_0-st] Affymetrix Human Exon 1.0 ST Array [transcript (gene) version]	Primary Fibroblast	10Gy	4 hours	Normal (1)
GSE30240	GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	G361, HepG2, TK6, U2OS, BJ cell lines	5Gy	3,6 hours	Normal(1) and Cancer(2)
GSE20549	GPL6244 [HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array[transcript (gene) version]	Lung Cancer Cell Lines H460 and H1299	2Gy	2,4,8 hours	Cancer (2)

Table 1: List of GEO datasets and their corresponding attributes.

2.1.2 Data Import

Various R Bioconductor packages were used to perform the gene expression and pathway analysis on the data from the GEO datasets. The specific R code that was utilized in this study can be seen in Appendix 1.The CEL files for the samples for each of the GEO data sets were downloaded from the GEO website and unzipped to the local drive. The Bioconducter package Oligo was then utilized to import the raw intensity data for each dataset into R. Once the raw intensity data for each data set was imported, quality control was performed. First, a histograph and boxplot of each of the samples within a data set was performed. This allowed for the identification of any obviously corrupted data. To analyze the quality of the data, a proble-level model was fit to the data using the Bioconductor package Olgi. Next, the Normalized Unscaled Standard Error (NUSE) was calculated and graphed for each of the samples in the data set. The NUSE plots are boxplots of the normalized standard error from the probe-level model fit for each sample [25]. These NUSE plots were analyzed in order to identify any abnormal arrays. A Relative Log Expression (RLE) plot was also performed from the probe-level model. The RLE plot was also useful in analyzing the quality of the arrays and ensuring that none of the arrays contain abnormal values with respect to the other samples in the dataset [25].

2.2.1 Normalization and Probeset Filtering

After quality control was performed on the arrays in the data set, the expression data was normalized. The normalization and summarization o the expression data was performed using the Robust Multi-Array Analysis (RMA) technique. The RMA method is extremely popular and creates a linear model of the log-scale expression data in order to perform a background correction and normalization of the data [25,29].

After the microarray expression data was background corrected and normalized, the probesets were filtered based on their intensity. For the gene filtering step, the packages limma, affyPLM, and genefilter were used. The gene filtering technique used in this study consisted of obtaining the intensity of all of the negative control probes in a dataset and creating a normal distribution of those intensities. Next, the 85th percentile value of that distribution was calculated. This value was used as the filtering value cutoff so that at least 1/3 of the probeset intensities for a specific probeset ID had to be above this value. This was performed in an effort to remove irrelevant or non-binding probesets and probesets for genes that are not expressed [25,29]. For the datasets with the HuEx 1.0 ST and Gene 1.0 ST platforms, the anti-genomic negative control probesets were obtained using the Oligo packages. These anti-genomic probesets consist of probesets that correspond to genes that are not present in humans and are therefore not expect to be expressed in the array [25,29]. Therefore, the intensity value for these probesets is likely to be caused by non-specific binding and other sources of noise [25,29].

For the datasets using the HGU 133 Plus 2.0 platform, the negative strand matching probesets were used as negative control. These are probesets that are not expected to hybridize to any RNA that is present in human cells [25,29]. To obtain the probeset IDs for the negative strand matching probesets, the NetAffx annotation files (release 33) for the HGU 133 Plus 2.0 platform were downloaded from the Affymetrix website. This file contains the probeset IDs for all of the probesets in the array. This information was used to obtain all negative strand matching probes. The intensity of all of the negative strand matching probesets was collected and a normal distribution was created as well. The value of the 85th percentile was also used as the filtering cutoff value.

2.2.2 Gene Expression Analysis

The gene expression analysis in this study was performed by fitting a linear model to the normalized and filtered expression data utilizing the Bioconductor package limma. This analysis separated the expression samples in a dataset into two groups, control or treatment. The control group consisted of the intensity values for each probeset for the samples that weren't exposed to radiation and the treatment group consisted of the intensity values for each probeset for the samples that were exposed to radiation. For this gene expression analysis, correction for multiple testing must be performed. In the gene expression analysis there were thousands of probesets that were tested for differential expression. This means that a large number of the probes that are not differentially expressed would be expected to be significant at a p-value cutoff of 0.05 because of random chance [25,29]. Therefore, to correct for the issue of multiple testing, The Benjamini and Hochberg's method to control the false discovery rate was used [25,29]. The cutoff value for the adjusted p-value given through this method used in this study was 0.05. Therefore, after the differentially gene expression analysis and correction for multiple testing, any probesets that were found to have an adjusted p-value of less than 0.05 were identified as being differentially expressed upon exposure to radiation.

2.2.3 Annotation

The output from the gene expression analysis consisted of the probeset IDs along with the log intensity and log fold change of the corresponding probeset IDs. There are various annotation techniques that can be used to identify the genes that correspond to each probeset [25,29]. However, many annotation techniques do not map probeset IDs to the same genes and can give differing results [25,29,38]. In addition, the use of HuEx 1.0 ST and HuGene 1.0 ST arrays presents further annotation issues. It is possible for some probesets to contain sequences that would map to many different genes [38,39]. This would make it difficult to determine which gene is being expressed and the true intensity value for that gene [38,39]. In addition, many genes undergo alternate splicing, which can add further complications when attempting to annotate the HuEx 1.0 ST array [38,39]. To achieve accurate annotation for the gene expression analysis performed on each of the different Affymetrix platforms, the Bioconductor package Annmap was used. The Annmap package utilizes the unique exon structure of a specific gene and the specific transcripts that can be mapped to a gene to provide accurate gene annotation [38,39].

Annmap aligns each of the probeset sequences with a reference human genome in order to determine the specific genes that correspond to that probeset [38,39]. This allows for the identification of probesets that are considered to be unreliable. Unreliable probesets are defined as probesets that either map to the genome multiple times or do not map to the genome [38,39]. The annotation performed in this study removed all unreliable probesets in order to increase the accuracy of the gene annotation. This annotation approach with Annmap was used for the differential gene expression analysis for the data sets that contained microarray data on the HGU 133 Plus 2.0, HuGene 1.0 ST, and HuEx 1.0 ST Affymetrix platforms.

The output from Annmap is the Ensemble gene ID for each of the differentially expressed genes. However, the pathway analysis performed in this study requires Entrez Gene ID. Therefore, to annotation the Ensemble gene IDs to Entrez IDs the Bioconductor package BioMart was used. BioMart is a commonly used annotation tool that can effectively annotate between Ensemble IDs and Entrez IDs [33]. The human genome is well annotated and the annotation approach utilizing BioMart and Annmap is an efficient method for ensuring that the differentially expressed probesets are mapped correctly to their corresponding genes and that the output data from the gene expression analysis is formatted correctly for the pathway analysis.

2.2.4 Pathway Analysis

The pathway analysis for this study was performed with the Bioconductor package Graphite, which utilizes the package SPIA. The input for these packages consisted of a named vector that contained the log fold change and Entrez gene ID for each gene that was found to be differentially expressed. These values were used to determine which

Reactome biological pathways were significantly inhibited or activated by the differential expression of the significant genes.

The packages Graphite and SPIA use an effective and unique approach to pathway analysis. The pathway analysis performed by these packages uses a interaction network that allows for the creation of a perturbation p-value, which is calculated based on the effect that the differentially expressed genes have on the pathway [29]. This factor is calculated by determining the relationship of each of the genes in the pathway and using the log fold change values for the differentially expressed genes to calculate how much the pathway was perturbed [29]. These pathway analysis packages also take the topology of the pathway into account [29] For example, a gene with many interactions at the beginning of a pathway that is differentially expressed would perturb the pathway more than a gene that is near the end of the pathway with few gene product interactions [29]. In addition, a p-value is also calculated based on how many genes would be expected to be differentially expressed within the pathway based on random chance [29]. These two attributes are used to calculate an overall p-value for the significance of the pathway [29]. However, like in the gene expression analysis, the pathway analysis must include a correction for multiple testing since there are hundreds of pathways that are being tested. For this study, a FDR adjusted p-value cutoff of 0.05 was used. The output of this pathway analysis was the identification of pathways that are significantly inhibited or activated by the differential expression of the significant genes that were determined in our gene expression analysis.

This pathway analysis used the Reactome database for biological pathways in order to ensure that the pathway analysis was significant. The Reactome database consists of pathways that are built with a base unit of a reaction between the gene products [31]. This is very important to the significance of the pathway analysis since the Reactome pathway accounts for the numerous interactions and reactions between gene products [31]. In addition, the Reactome database is a free open-source pathway database that is constantly updated [31].

In summary, the pathway analysis in this section was performed using the Bioconductor packages Graphite and SPIA. The input for this analysis consisted of the gene IDs for the differentially expressed genes obtained in the gene expression analysis and their respective log fold change values. The output from this analysis was the identification of the pathways that are significantly inhibited or activated by exposure to ionizing radiation.

2.2.5 Drug Target Analysis

The drug target analysis for this study was performed using MetaDrug, a curated drug knowledgebase. The drug target analysis was performed for the six pathways that were found to be differentially expressed among all data sets in the cancer group. These six pathways were: Cell Cycle Mitotic, DNA Replication, Mitotic M-M/G1 phases, AKT phosphorylates targets in the cytosol, M Phase, and Mitotic Prometaphase. The input for this analysis consisted of the name of each pathway that was differentially expressed and the genes that were differentially expressed in each pathway. After the analysis was performed for each pathway, the names of the drugs that significantly targeted each pathway and the gene targets of the drug were recorded. The drug output from this analysis was compared with the results from the gene expression analysis to ensure that the drugs found in the analysis targeted genes that were found to be differentially expressed in our gene expression analysis. Finally, the drug targets that were found in this analysis were compared across each significant pathway. The results of this analysis can be seen in Table 3.

Results

3.1.1 Gene Expression and Pathway Analysis

The complete list of pathways that were obtained from the pathway analysis and were found to be either significantly inhibited or activated by exposure to ionizing radiation in each data set can be seen in table 2.

GEO Dataset (Group)	Differentially Expressed Reactome Pathways
GSE 26841 (Normal)	No Significant Pathways
GSE 41840 (Normal)	No Significant Pathways
GSE 37668 (Normal)	"Mitotic G1-G1/S phases"
GSE 30240 (Normal)	"APC/C:Cdc20 mediated degradation of Cyclin B", "M Phase", "Mitotic Prometaphase",

	"Cell Cycle, Mitotic", "DNA Replication", "Mitotic M-M/G1 phases"
GSE 20549 (Cancer)	"AKT phosphorylates targets in the cytosol", "Meiotic Recombination", "RNA, Polymerase I Transcription", "RNA Polymerase I Promoter Clearance", "Activation of the pre-replicative complex", "RNA Polymerase I Promoter Opening", "Telomere Maintenance", "RNA Polymerase I, RNA Polymerase III, and Mitochondrial Transcription", "Chromosome, Maintenance", "G/M, Checkpoints", "Activation of ATR in response to replication stress", "Mitotic, G-G/S, phases", "G/S, Transition", "Cell Cycle, Mitotic", "Metabolism of carbohydrates", "Metabolism of RNA", "SLC-mediated transmembrane transport", "Processing of Capped Intron-Containing Pre-mRNA", "mRNA Processing", "Influenza Life Cycle", "Influenza Infection", "Transcription", "SemaD induced cell migration and growth-cone collapse", "SemaD in semaphorin signaling", "Export of Viral Ribonucleoproteins from Nucleus", "Eukaryotic Translation Elongation", "Translation", "Peptide chain elongation", "Glucose transport", "Hexose transport", "Transport of Mature mRNA derived from an Intron-Containing Transcript", "Transport of Ribonucleoproteins into the Host Nucleus", "Transport of Mature Transcript to Cytoplasm", "NEP/NSInteracts with the Cellular Export Machinery", "Chromosome Maintenance", "DNA Replication", "Mitotic M-M/G phases", "Apoptosis"
GSE 30240 (Cancer)	"M Phase", "Mitotic Prometaphase", "Cell Cycle, Mitotic", "DNA Replication", "Mitotic M-M/G1 phases", "APC/C-mediated degradation of cell cycle proteins"

Table 2: Differentially expressed pathways by GEO dataset.

It should be noted that many of the pathways are known to have a high level of biological significance, such as "apoptosis" and "cell cycle, mitotic" [19,34,5]. A large proportion of the pathways that were found to be significant are related to cell cycle regulation. In addition, there were some pathways that were novel and unexpected, namely "Export of Viral Ribonucleoproteins from Nucleus", "Influenza Life Cycle", "Influenza Infection". However, the following pathways were found in every data set using cancerous cell lines: Cell Cycle Mitotic, DNA Replication, Mitotic M-M/G1 phases, AKT phosphorylates targets in the cytosol, M Phase, and Mitotic Prometaphase. There were no pathways that were found to be differentially expressed in each of the normal cell type groups. This provides further evidence for the significance of cell cycle regulation in cancerous cells that are exposed to ionizing radiation. In addition, these pathways may provide possible genetic markers for use in radiation therapy for cancer.

3.1.2 Drug Pathway Analysis

The drug pathway analysis that was performed on the six pathways that were differentially expressed in each of the cancer data sets was performed separately for each pathway. After this analysis, the drugs that significantly targeted each of the pathways and their respective gene targets were determined. The six drug targets that were targeted in at least 1 of the six pathways are: POLA1, POLA2, PLK1, CENPE, AURKB, CDKN1B, and RB1. The drug gene targets, the effect of the drug on its respective pathway, and the number of pathways that are affected by the gene target are outlined in table 3.

Drug Target	Effect	Number of Significant Pathways Targeted
PLK1	Inhibited	5/6
CENPE	Inhibited	5/6
POLA1	Inhibited	3/6
POLA2	Inhibited	3/6
CDKN1B	Inhibited	3/6
AURKB	Inhibited	2/6
RB1	Inhibited	2/6

Table 3: List of drug targets found in drug target analysis

Discussion

4.1.1 Gene and Pathway analysis in Perspective

A large portion of the pathways that were found to be significantly inhibited or activated after the gene expression analysis and pathway analysis was performed were involved with cell cycle regulation or DNA repair. For example, some of the pathways that were found to be significant were "Apoptosis", "Cell Cycle, Mitotic", "DNA replication", and other pathways that were involved in various steps of the cell cycle. These pathways are extremely similar to the pathways found in previous studies. Previous studies that performed gene expression analyses have found that the genes that were differentially expressed were members of pathways that are involved in cell cycle regulation, apoptosis, and DNA replication [19,34,35]. However, the analysis performed in this study utilized thorough statistical procedures through a comprehensive pathway analysis that included the known biological reactions and interactions of the gene products of the genes that were found to be differentially expressed. This is significant because it provides further evidence that the pathways that were found to be significant are critical to the radiation response of normal and cancerous human cells.

4.1.2 Drug Pathway Analysis

The identification of the drugs that have a significant impact on the pathways that were differentially expressed in each cancerous data set is very significant. Our findings suggest that the drugs that were identified in our drug pathway analysis could have a potential use in the field of radiation medicine. In order to provide further evidence of the impact of these drugs on cancerous tissue during radiation therapy, further study with experimental validation is needed. One possible future experimental validation is a survival analysis between cancerous tissue that is exposed to IR and cancerous tissue that is treated with each of the drugs that were found to be significant and exposed to IR. This would allow for the determination of the effects of the drugs on cancerous tissue exposed to IR. If the validation shows that these drugs have a negative impact on the survivability of cancerous tissue upon exposure to radiation, they could potential be used during radiation therapy to lower the possibility of recurrence occurring post treatment.

4.1.3 Impact of the Results and Perspective

This study successfully identified the genes that are differentially expressed upon exposure to IR, the pathways that are significantly inhibited or activated during exposure to IR, and the pharmaceutical compounds that significantly target these pathways. These results contain data that is clinically significant and contains numerous medical applications with high impact. The identification of pathways that are significantly inhibited or activated by exposure to IR will allow for the identification of novel therapeutic targets that can be utilized to develop new drugs in the field of radiation therapy or radiation protection.

The six significant pathways identified in this study were critical to the cell cycle or DNA repair. This is consistent with the findings of other previous studies [19,34,35]. In addition, the pathways that were significant in only one data set could be implicated in radiation response and involved DNA repair pathways. DNA repair pathways have also been found to be differentially expressed in human cells exposed to IR [14,15,16,17,18,35]. The identification of critical pathways and genes that are differentially expressed allows for the future identification of genetic markers that can be utilized to predict patient outcome to radiation treatment. In addition, the identification of currently developed drugs that significantly target the pathways that are inhibited or activated by exposure to IR and their respective drug targets allows for the identification of drugs that can be used for a novel purpose in the field of radiation medicine. This has a high degree of clinical significance because novel drugs or drug applications that either increases the radioresistance of normal cells or radiosensitivity of cancerous tissue will have a positive impact on patient treatment. The identification of drugs that increase the radioresistance of normal cells would have applications in the field of radiation protection as these compounds could be utilized to protect nuclear workers from IR exposure. Drugs that significantly increase the radiosensitivity of cancerous cells would be extremely useful in radiation therapy as they would make the cancerous tissue more responsive to treatment and reduce the risk of cancer recurrence. This study lays the foundation for futures studies to determine the effect of drugs that significantly target these differentially expressed pathways on the radiosensitivity and radioresistance of human cells and the identification of genetic markers that can be used in radiation therapy.

The targets of the drugs that were identified are also of significant interest. The targets of these drugs are: POLA1, POLA2, PLK1, CENPE, AURKB, CDKN1B, and RB1. There have been many studies that have identified PLK1 as a potential therapeutic target in radiation medicine [40,41,42,43]. Many studies involving RNAi knockdowns of PLK1 in cancerous tissue have also been performed. These studies found that inhibition of PLK1 stunted cell growth and increased cell death in cancerous tissues [40,41,42,43]. In addition, Harris et al found that the use of BI 2536, a drug that inhibits PLK1 and was found in our drug analysis, increases the radiosensitivity of cancerous cells to radiation [42]. Therefore, the numerous studies that show the effect of PLK1 inhibition on cancerous cell growth and exposure to radiation provide further evidence for our pathway and drug target analysis results in the identification of PLK1 as a possible genetic hotspot effecting cancerous cell radiosensitivity.

In addition to PLK1, previous studies have found that CENPE expression was linked to radiation induced fibrosis [44]. This suggests that CENPE could potentially be used as a genetic marker to determine patient outcome, in particular the development of radiation fibrosis, after radiation therapy. However, there are no comprehensive studies to our knowledge that discussed the use of CENPE as potential therapeutic targets in radiation therapy.

The significance of the identified gene targets of POLA1 is also supported by previous studies. A study by Toukoki et al found that POLA1 has a significant role in the response of streptococcus to peroxide stress [45]. This is significant to radiation response because the indirect effect of radiation on human cells damages DNA in a similar mechanism to peroxide [1]. The mechanism for DNA damage from the indirect effects of radiation involves the formation of free radicals from the interaction of radiation with water molecules, which damages DNA using the same mechanism involved in peroxide stress [1]. Therefore, this provides significant evidence that POLA1 may have a significant role in the radiosensitivity of cancerous cells.

The significant gene target of POLA2 has also been previously studied. Previous studies have found that POLA2 is differentially expressed in cancerous cells that were exposed to chemotherapy and ionizing radiation [46,47]. This is significant because these studies provide further evidence of the potential role of POLA2 as a genetic marker for radiation response or treatment. In addition, it provides further support for our identification of this gene as a possible therapeutic target to increase cancer cell radiosensitivity.

The gene target of AURKB that our study identified has been found to be a significant gene of interest in radiation response. Many papers have previously performed knockdowns of AURKB and found that AURKB inhibition is associated with increased cancer cell death [48,49]. In addition, an increase in the radiosensitivity of cancer cells was observed upon the inhibition of AURKB [48,49]. Therefore, the significance of this identified gene target has been found to be very high in previous studies and our identification of this gene provides further evidence for its role in the radiation sensitivity of cancerous cells.

The gene target of CDKN1B that was identified by our analysis has also been associated with changes in radiosensitivity in previous studies. It has been shown that knocking down CDKN1B increases the radioresistance of cancerous cells and that higher expression of CDKN1B is correlated with increased radiosensitivity [50,51]. In addition, the last gene target, RB1, was found to be associated with increased radioresistance [52]. RB1 is known to play a role in homologous DNA repair, which is critical to repairing double strand breaks that occur upon exposure to ionizing radiation [52].

The association of these gene targets with changes in the radiosensitivity or radioresistance of cancerous cells is significant as it allows for future research to be performed determining if these genes could be utilized as therapeutic targets for new drugs. In addition, some of these gene targets could potentially serve as genetic markers to predict patient outcome or response to different treatments.

Conclusion

The analysis performed in this study is novel in its approach as a meta-analysis of microarray data on human normal and cancerous cells exposure to IR. In addition, the thorough use of novel statistical approaches that incorporate the biological interactions and reactions between gene products allowed for a robust and high impact pathway analysis. This study has identified critical pathways that are significant to the biological response of human cells to IR and has laid the foundation for further pharmaceutical research of drugs that target these specific pathways in order to alter the radiosensitivity or radioresistance of normal and cancerous tissue. In addition, the identification of current drugs that target the pathways that are significantly affected by exposure to IR is significant and allows for the identification of new potential drugs that can be utilized in the field of radiation medicine or radiation protection. Given the results of this study in relationship to previous studies, it is evident that biological pathways involved with cell cycle regulation, apoptosis, and DNA repair are critical to the response of human cells to IR. However, the novel approach in this study has allowed for the identification of specific pathways and gene product interactions that are effect by IR exposure. In addition, this paper has outlined the high potential for the use of gene expression and pathway analyses in the identification of pharmaceutical compounds that can target the specific biological pathways that are critical to IR exposure. This paper also outlined the power of meta-analysis of microarray data and the potential use of online repositories to perform future studies on any biological question of interest. Finally, this paper has shown that this approach can also be used to analyze the biological pathways critical to any potential question of interest and the identification pharmaceutical compounds and gene hotspots that can potentially be used in a novel manner.

References

1. Baskar R, Lee KA, Yeo R, Yeoh KW. Cancer and Radiation Therapy: Current Advances and Future Directions. Int J Med Sci 2012; 9(3):193-199. doi:10.7150/ijms.3635. Available from http://www.medsci.org/v09p0193.htm
2. Meadows SK, Dressman HK, Muramoto GG, Himburg H, Salter A, Wei Z, et al. Gene Expression Signatures of Radiation Response are Specific, Durable, and Accurate in Mice and Humans. PLoS One. 2008 Apr 2; 3(4):e1912.
3. Ritz B. Radiation Exposure and Cancer Mortality in Uranium Processing Workers. Epidemiology. 1999 Sep;10(5):531-8.
4. Baskar R, Lee KA, Yeo R, Yeoh KW. Cancer and Radiation Therapy: Current Advances and Future Directions. Int J Med Sci. 2012 Feb 27; 9(3):193-199.
5. Trock B, Han M, Freedland S, Humphreys EB, DeWeese, TL, Partin W, et al. Prostate Cancer-Specific Survival Following Salvage Radiotherapy vs Observation in Men With Biochemical Recurrence After Radical Prostatectomy. JAMA. 2008 Jun 18; 299(23):2760-9.
6. Kleinerman RA, Tucker MA, Tarone RE, Abramson DH, Seddon JM, Stovall M, et al. Risk of New Cancers After Radiotherapy in Long-Term Survivors of Retinoblastoma: An Extended Follow-up. J Clin Oncol. 2005 Apr 1; 23(10):2272-9.
7. Alsner J, Rødningen OK, Overgaard J. Differential gene expression before and after ionizing radiation of subcutaneous fibroblasts identifies breast cancer patients resistant to radiation-induced fibrosis. Radiother Oncol. 2007 June; 83(3):261-6.

8. Amundson SA, Grace MB, McLeland CB, Epperly MW, Yeager A, Zhan Q. Human in Vivo Radiation-Induced Biomarkers: Gene Expression Changes in Radiotherapy patients. Cancer Res. 2004 Sep 15; 64(18):6368-71.

9. Du XL, Jiang T, Wen ZQ, Li QS, Gao R, Wang F. Differential expression profiling of gene response to ionizing radiation in two endometrial cancer cell lines with distinct radiosensitivities. Oncol Reps. 2009 Mar; 21(3):625-34.

10. Guo WF, Lin RX, Huang J, Zhou Z, Yang J, Guo GZ, et al. Identification of differentially expressed genes contributing to radioresistance in lung cancer cells using microarray analysis. Radiat Res. 2005 Jul; 164(1):27-35.

11. Lee YS, Oh JH, Yoon S, Kwon MS, Song CW, Kim KH, et al. Differential Gene Expression Profiles of Radioresistant Non-Small-Cell Lung Cancer Cell Lines Established by Fractionated Irradiation: Tumor Protein p53-Inductible Protein 3 Confers Sensitivity to Ionizing Radiation. Int J Radiat Oncol Biol Phys. 2010 Jul 1; 77(3):858-66.

12. Stassen T, Port M, Nuyken I, Abend M. Radiation-induced gene expression in MCF-7 cells. Int J Radiat Biol. 2003 May; 79(5):319-31.

13. Xu QY, Gao Y, Liu Y, Yang WZ, Xu XY. Identification of differential gene expression profiles of radioresistant lung cancer cell line established by fractionated ionizing radiation in vitro. Chin Med J. 2008 Sep 20; 121(18):1830-7.

14. Forrester HB, Li J, Hovan D, Ivashkevich AN et al. DNA repair genes: alternative transcription and gene expression at the exon level in response to the DNA damaging agent, ionizing radiation. PLoS One 2012;7(12):e53358. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23285288

15. Kim K, Yoo H, Joo K, Jung Y, Jin J, Kim Y, Jin S, et al. Time-course analysis of DNA damage response-related genes after in vitro radiation in H460 and H1299 lung cancer cell lines. Exp Mol Med. 2011 July 31; 43(7);419-426. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3158501/

16. Parvin B, Groesser T, Jakkula L, Han J. Time-varying expression data from WI38 cell lines that were exposed to a challenge dose with or without a priming dose. NCBI. 2012 May 1. Available from: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE37668

17. Rashi-Elkeles S, Elkon R, Shavit S, Lerenthal Y et al. Transcriptional modulation induced by ionizing radiation: p53 remains a central player. Mol Oncol 2011 Aug;5(4):336-48. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21795128

18. Sung HY, Wu HG, Ahn JH, Park WY. Dcr3 inhibit p53-dependent apoptosis in y-irradiated lung cancer cells. Int J Radiat Biol. 2010 Sep; 86(9):780-90. Available from: http://informahealthcare.com/doi/abs/10.3109/09553002.2010.484481

19. Kim HS, Kim SC, Kim SJ, Park CH, Jeung HC, Kim YB. Identification of a Radiosensitivity Signature Using Integrative Metaanalysis of Published Microarray Data for NCI-60 Cancer Cells. BMC Genomics. 2012 Jul 30;13:348.

20. Butte A. The Use and Analysis of Microarray Data. Nat Rev Drug Discov. 2002 Dec 1; 1(12): 951-60.

21. Smyth GK, Speed T. Normalization of cDNA Microarray Data. Methods. 2003 Dec; 31(4):265-73.

22. Zakharkin SO, Kim K, Mehta T, Chen L, Barnes S, Scheirer KE. Sources of Variation in Affymetrix Microarray Experiments. BMC Bioinformatics. 2005 Aug 29; 6:214.

23. Zhang Y, Szustakowsky J, Schinke M. Bioinformatics Analysis of Microarray Data. Methods Mold Biol. 2009; 573:259-84.

24. Gohlmann H, Talloen W, Etheridge AM, Gross LJ, Lenhart S, Maini P, et al. Gene Expression Studies Using Affymetrix Microarrays. Boca Raton: Chapmann & Hall; 2009.

25. Draghici S, Briton NF, Lin X, Safer H, Schneider MV, Singh M, et al. Statistics and Data Analysis for Microarrays Using R and Bioconductor. 2nd ed. Boca Raton: Taylor & Francis Group; 2012.

26. Lovén J, Orlando DA, Sigova AA, Lin CY, Rahl PB, Burge CB, et al. Revisiting Global Gene Expression Analysis. Cell. 2012 Oct 26;151(3):476-82.

27. Wang D, Cheng L, Wang M, Wu R, Li P, Li B, et al. Extensive Increase of Microarray Signals in Cancer Calls for novel Normalization Assumptions. Comput Biol Chem. 2011 Jun; 35(3):126-30.

28. Wu Z, Irizarry R, Gentleman R, Murillo FM, Spencer F. A Model Based Background Adjustment for Oligonucleotide Expression Arrays. JHU Biostat. 2004 May. Working Paper 1.

29. Sales G, Calura E, Cavalieri D, Romualdi C. graphite - A Bioconductor package to convert pathway topology to gene network. BMC Bioinformatics. 2012 Jan 31;13:20.

30. Tian L, Greenberg SA, Kong SW, Altschuler J, Kohane IS, Park PJ. Discovering Statistically Significant Pathways in Expression Profiling Studies. 2005 Sep 20; 102(38):13544-9.

31. Joshi-Tope G, Gillespie M, Vastrik I, D'Eustachio P, Schmidt E, de Bono B, et al. Reactome: a knowledgebase of biological pathways. Nucleic Acids Res. 2005 Jan 1; 1:33(Database issue):D428-32.

32. Tarca AL, Draghici S, Khatri P, Hassan SS, Mittal P, Kim JS, et al. A novel signaling pathway impact analysis. Bioinformatics. 2009 Jan 1; 25(1):75-82.

33. Ekins S, Kirillov E, Rakhmatulin EA, Nikolskaya T. A novel method for visualizing nuclear hormone receptor networks relevant to drug metabolish. Drug Metab Dispos. 2005 Mar; 33(3):474-81.

34. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A. 2001 Apr 24; 98(9):5116-5121. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC33173/

35. Fachin AL, Mello SS, Sandrin-Garcia P, Junta CM, Ghilardi-Netto T, Donadi EA, et al. Gene expression profiles in radiation workers occupationally exposed to ionizing radiation. J Radiat Res. 2009 Jan; 501(1):61-71.

36. Glimelius B, Pahlman L. Preoperative Radiotherapy for Rectal Cancer: Hypofractionation with Multiple Fractions (25-25Gy). Ann Ital Chir. 2001 Sep-Oct72(5):539-47.

37. Toonen EJM, Gilissen C, Franke B, Kievit W, Eijsbouts A, den Broder AA, et al. Validation Study of Existing Gene Expression Signatures for Anti-TNF Treatment in patients with Rheumatoid Arthritis. PLoS One. 2012;7(3):e33199.

38. Yates T, Okoniewski MJ, Miller CJ. X:Map: annotation and visualization of genome structure for Affymetrix exon array analysis. Nucleic Acids Res. 2008 Jan;36(Database issue):D780-6. Epub 2007 Oct 1. Available from: http://nar.oxfordjournals.org/content/36/suppl_1/D780.full

39. Allen J, Wang S, Chen M, Girard L, Minna J, Xie Y, et al. Probe Mapping Across Multiple Microarray Platforms. Brief Bioinform. 2012 Sep;13(5):547-54.

40. Tandle AT, Kramp T, Kil WJ, Halthore A, Gehlhaus K, Shankavaram U, et al. Inhibition of polo-like kinase 1 in glioblastoma multiforme induces mitotic catastrophe and enhances radiosenstitisation. Eur J Cancer. 2013 Jun 18. pii: S0959-8049(13)00410-3. Available from: http://www.ejcancer.com/article/S0959-8049%2813%2900410-3/abstract

41. Gerster K, Shi W, Ng B, Yue S, Ito E, Waldron J, et al. Targeting polo-like kinase 1 enhances radiation efficacy for head-and-neck squamous cell carcinoma. Int J Radiat Oncol Biol Phys. 2010 May 1; 77(1):253-60. Available from: http://www.redjournal.org/article/S0360-3016%2809%2903568-8/abstract

42. Harris PS, Venkataraman S, Alimova I, Birks DK, DonsonAM, Knipstein J, et al. Polo-like kinase 1 (PLK1) inhibition suppresses cell growth and enhances radiation sensitivity in medulloblastoma cells. BMC Cancer. 2012 Mar 5; 12:80. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22390279

43. Luo J, Emanuele MJ, Li D, Creighton CJ, Schlabach MR, Westbrooke TF, et al. A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. 2008 May 29; 137(5)835-48. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19490893

44. Edvardsen H, Landmark-Hoyvik H, Reinertsen KV, Zhao X, Grenaker-Alnaes GI, Nebdal D, et al. SNP in *TXNRD2* Associated With Radiation-Induced Fibrosis: A Study of Genetic Variation in Reactive Oxygen Species Metabolism and Signaling. Int J Radiat Oncol Biol Phys. 2013 Jul 15;86(3):791-9. Available from: http://www.redjournal.org/article/S0360-3016%2813%2900217-4/abstract

45. Toukoki C, Gryllos I. PolA1, a putative DNA polyermase I, is coexpressed with PerR and contributes to peroxide stress defenses of group A streptococcus. 2013 Feb; 195(4):717-25. Available from:http://www.ncbi.nlm.nih.gov/pubmed/22390279

46. Zhou T, Chou J, Mullen T, Elkon R, Zhou Y, Simpson D, et al. Identification of Primary Transcriptional Regulation of Cell Cycle-Regulated Genes upon DNA Damage. Cell Cycle. 2007 April 15; 6(8):972-81. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2117899/

47. Roe OD, Szulkin A, Anderssen E, Flatberg A, Sandeck H, Amundsen T, et al. Molecular resistance fingerprint of pemetrexed and platinum on a long-term survivor of mesothelioma. PLoS One. 2012; 7(8):e40521. Available from: http://www.ncbi.nlm.nih.gov/pubmed/?term=POLA2+roe

48. Niermann KJ, Moretti L, Giacalone NJ, Sun Y, Schleicher SM, Kopsombut P, et al. Enhanced radiosensitivity of androgen-resistance prostate cancer: AZF1152-mediated Aurora Kinase B inhibition. Radiat Res. 2011 Apr; 175(4):444-51. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21222513

49. Tao Y, Zhang P, Girdler F, Frascogna V, Castedo M, Bourhis J, et al. Enhancement of radiation response in p53-deficient cancer cells by the Aurora-B kinase inhibitor AZD1152. Oncogene. 2008 May 22; 27(23):3244-55. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18084327

50. Zhang C, Wang G, Kang C, Du Y, Pu P. [Up-regulation of p27(kip1) by miR-221/222 antisense oligonucleotides enhances the radiosensitivity of U251 glioblastoma]. 2009 Dec; 26(6):634-8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19953484

51. Tong Q, Zhang W, Jin S, Li S, Chen Z. The relationship between p27(kip1) expression and the change of radiosensitivity of esophageal carcinoma cells. 2011 Feb; 46(2):173-6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20923380

52. Yang Y, Tian S, Brown B, Chen P, Hu H, Xia L, et al. The Rb1 gene inhibits the viability of retinoblastoma cells by regulating homologous recombination. 2013 Jul; 32(1): 137-43. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23670186

53. Gautier L, Cope L, Bolstad BM, Irizarry RA. Affy-Analysis of Affymetrix GeneChip Data at the Probe Level. Bioinformatics. 2004 Feb 12; 20(3):307-15.

54. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for Integration and Interpretation of Large-Scale Molecular Data Sets. Nucleic Acid Res. 2012 Januar; 40(D1):D109-D114.

55. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M, Hirakawa M. KEGG for Representation and Analysis of Molecular Networks Involving Diseases and Drugs. 2010 January; 38:D355-D360.

56. Kis E, Szatmári T, Keszei M, Farkas R, Esik O, Lumniczky K, et al. Microarray Analysis of Radiation Response Genes in Primary Human Fibroblasts. Int J Radiat Oncol Biol Phys. 2006 Dec 1; 66(5):1506-14.

57. Kodama K, Horikoshi M, Toda K, Yamada S, Hara K, Irie J, et al. Expression-Based Genome-Wide Association Study Links the Receptor CD44 in Adipose Tissue with Type 2 Diabetes. Proc Natl Acad Sci USA. 2012 May 1; 109(18):7049-54.

58. Miecznikowski JC, Wang D, Liu S, Sucheston L, Gold D. Comparative Survival Analysis of Breast Cancer Microarray Studies Identifies Important Prognostic Genetic Pathways. BMC Cancer. 2010 Oct 21;10:573.

59. Ponette V, Le Péchoux C, Deniaud-Alexandre E, Fernet M, Giocanti N, Tourbez H, et al. Hyperfast, early cell response to ionizing radiation. Int J Radiat Biol. 200 Sep;76(9):1233-43.

60. Samet J. Radiation and Cancer Risk: a Continuing Challenge for Epidemiologists. Environ Health. 2001 Apr 5; 10(Suppl 1):S4.

61. Servant N, Gravier E, Gestraud P, Laurent C, Paccard C, Biton A, et al. EMA - A R Package for Easy Microarray Data Analysis. BMC Res Notes. 2010 Nov 3;3:277.

62. Smedley D, Halder S, Ballester B, Holland R, London D, Thorisson G, et al. BioMart - biological queries made easy. BMC Genomics. 2009 Jan 14; 10:22. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2649164/

63. Smyth G. Limma: Linear models for microarray data. In: Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W, editors. Bioinformatics and Computational Biology Solutions using R and Bioconductor. New York: Springer; 2005. p. 397-420.

64. Soh D, Dong D, Guo Y, Wong L. Consistency, Comprehensiveness, and Compatibility of Pathway Databases. BMC Bioinformatics. 2010; 11:449.

65. Stevens JR, Doerge RW. Combining Affymetrix Microarray Results. BMC Bioinformatics. 2005 Mar 17; 6:57.

66. Stuart D Pepper, Emma K Saunders, Laura E Edwards, Claire L Wilson, Crispin J Miller. The Utility of MAS5 Expression Summary and Detection Call Algorithms. BMC Bioinformatics. 2007; 8:273.

67. Therneau TM, Ballman KV. What Does PLIER Really Do? Cancer Inform. 2008 Aug 27; 6:423-431.

68. Thomas R, de la Torre L, Chang X, Mehrotra S. Validation and Characterization of DNA Microarray Gene Expression Data Distribution and Associated Moments. BMC Bioinformatics. 2010 Nov 24;11:576.

69. Tsai MH, Cook JA, Chandramouli GV, DeGraff W, Yan H, Zhao S. Gene expression profiling of breast, prostate, and glioma cells following single versus fractionated doses of radiation. 2007 Apr 15; 67(8):3845-52.

70. Verducci JS, Melfi VF, Lin S, Wang Z, Roy S, Sen CK. Microarray Analysis of Gene Expression: Considerations in Data Mining and Statistical Treatment. Physiol Genomics. 2006 May 16;25(3):355-63.

71. Wong WC, Loh M, Eisenhaber F. On the Necessity of Different Statistical Treatment for Illumina Beadchip and Affymetrix GeneChip Data and its Significance for Biological Interpretation. Biol Direct. 2008 Jun 3; 3:23.

Appendix

1.1 R code for differential expression and pathway analysis for HGU 133 Plus 1.0 Platform.

#Quality Control

```
library(oligo)
setwd("C:/Users/Chris/Documents/Thesis/R/GSE30240_C1_Cell/U2OS_6hr")
dat=read.celfiles(list.celfiles("C:/Users/Chris/Documents/Thesis/R/GSE30240_C1_Cell/U2OS_6
hr"))
#Adjust the path to the CEL files
library("RColorBrewer")
usr.col=brewer.pal(9, "Set1")
mycols=rep(c("blue ", "red"), c(3,3))
hist(dat, lty=rep(1,6), col=mycols)
Legend("topright", rownames(pData(dat)), lty=rep(1,6), col=mycols, cex=0.6)
boxplot(dat,col=mycols,las=3,cex.axis=0.5,names=sampleNames(dat))
plm <- fitProbeLevelModel(dat)
NUSE(plm)
RLE(plm)
eset <- rma(dat)

bgp=read.csv(file.choose(), header=FALSE)
#Choose the csv file containing the negative control probeset IDs.
sql <- paste("select * from pmfeature where fsetid in ('",
paste(bgp[,1], collapse = "','"), "');", sep = "")
```

# Gene Filtering

```
library(limma)
library(affyPLM)
TS <- gl(2,3,length=6, labels=c("control", "treatment") )
#Adjust TS for each data set
design<-model.matrix(~0+TS)
colnames(design)<-levels(TS)
fit1<-lmFit(eset,design)
cont.matrix<-makeContrasts(contrast=treatment-control, levels=design)
fit1<-contrasts.fit(fit1, cont.matrix)
fit1<-eBayes(fit1)
topall2<-topTable(fit1, coef="contrast", number=nrow(eset), adjust="BH", p.value=1,)
limmaDEgene<-topTable(fit1, coef="contrast", number=nrow(eset), adjust="BH", p.value=1)
#p-value cutoff of 1. This step just gives all probe IDs and intensity.
```

```
bgp=as.matrix(bgp)
holder=c()
x=1
while (x< dim(limmaDEgene[1])+1){
if (limmaDEgene[x,][[1]][1]%in%bgp==TRUE){
holder=c(holder,limmaDEgene[x,][[3]][1])}
x=x+1}

Cutoff=quantile(holder, 0.85)
library(genefilter)
f1<-pOverA(1/3,Cutoff)
ff<-filterfun(f1)
index<-genefilter(eset,ff)
eset1<-eset[index,]
#Filters genes based on the negative control distribution.
```

# DE TEST

```
library(limma)
library(affyPLM)
TS <- gl(2,3,length=6, labels=c("control", "treatment") )
#Adjust TS for each data set
design<-model.matrix(~0+TS)
colnames(design)<-levels(TS)
fit1<-lmFit(eset1,design)
cont.matrix<-makeContrasts(contrast=treatment-control, levels=design)
fit1<-contrasts.fit(fit1, cont.matrix)
fit1<-eBayes(fit1)
topall2<-topTable(fit1, coef="contrast", number=nrow(eset1), adjust="BH", p.value=1,)
limmaDEgene<-topTable(fit1, coef="contrast", number=nrow(eset1), adjust="BH",
p.value=0.05)
dim(limmaDEgene)
testinglist=limmaDEgene[,1]
```

# Annotation

```
library(annmap)
annmapConnect('human70')
arrayType(name='HG-U133Plus2')
geneListExp=probesetToGene( testinglist, as.vector=TRUE, rm.unreliable=TRUE )
```

# PATHWAY ANALYSIS

```
library(annmap)
annmapConnect('human70')
```

```
arrayType(name='HG-U133Plus2')
inten=limmaDEgene[,1:2]
L=dim(limmaDEgene)[1]
EnsIntensity=c()
EnsID=c()
z=1
while (z<L+1){
geneList=probesetToGene(limmaDEgene[z,][[1]][1], as.vector=TRUE, rm.unreliable=TRUE )
if (is.null(geneList)){}
else {
if (length(geneList) ==1){
EnsIntensity=c(limmaDEgene[z,][[2]][1], EnsIntensity)
EnsID=c(geneList, EnsID)}
else{
for (x in geneList){
EnsID=c(x, EnsID)}
t=0
T=length(geneList)
while (t<T){
EnsIntensity=c(limmaDEgene[z,][[2]][1], EnsIntensity)
t=t+1}}}
z=z+1}

Together=data.frame(EnsID=EnsID, EnsIntensity=EnsIntensity)
averageInt=ave(Together[,2], Together[,1])
Together2=data.frame(EnsID=EnsID, EnsIntensity=EnsIntensity, averageInt = averageInt)
FinalTable= Together2[!duplicated(Together2["EnsID"]),]
FinalTable=FinalTable[,-2]
EnsList=as.vector(FinalTable[,1])
EnsListInt=as.vector(FinalTable[,2])

library(biomaRt)
mart <- useMart(biomart = "ensembl", dataset = "hsapiens_gene_ensembl")
GOterms = getBM(attributes=c('ensembl_gene_id','entrezgene'),filters='ensembl_gene_id',
values=EnsList, mart=mart)
GOterms=na.omit(GOterms)

OutputTable= data.frame(EnsID=GOterms$ensembl_gene_id, Entrez=GOterms$entrezgene,
Intensity=FinalTable[match(GOterms$ensembl_gene_id, FinalTable$EnsID ), 2])
GeneNames=OutputTable[,2]
DE_genes=OutputTable[,3]

RepeatGeneID = data.frame(GeneNames=GeneNames, DE_genes=DE_genes)
Average2=ave(RepeatGeneID[,2], RepeatGeneID[,1])
RepeatGeneID2=data.frame(GeneNames=GeneNames, DE_genes=DE_genes,
Average2=Average2)
```

```
RepeatGeneID2=RepeatGeneID2[!duplicated(RepeatGeneID2["GeneNames"]),]
GeneNames=RepeatGeneID2[,1]
DE_genes= RepeatGeneID2[,3]
names(DE_genes)=GeneNames

#Get Gene Universe
TS <- gl(2,3,length=6, labels=c("control", "treatment") )
design<-model.matrix(~0+TS)
colnames(design)<-levels(TS)
fit1<-lmFit(eset1,design)
cont.matrix<-makeContrasts(contrast=treatment-control, levels=design)
fit1<-contrasts.fit(fit1, cont.matrix)
fit1<-eBayes(fit1)
topall2<-topTable(fit1, coef="contrast", number=nrow(eset1), adjust="BH", p.value=1,)
limmaDEgene<-topTable(fit1, coef="contrast", number=nrow(eset1), adjust="BH", p.value=1)
dim(limmaDEgene)
testinglist=limmaDEgene[,1]

library(annmap)
annmapConnect('human70')
arrayType(name='HG-U133Plus2')
geneList=probesetToGene( testinglist, as.vector=TRUE, rm.unreliable=TRUE )
universe=geneList

GOtermsUniverse =
getBM(attributes=c('ensembl_gene_id','entrezgene'),filters='ensembl_gene_id', values=universe,
mart=mart)
GOtermsUniverse=GOtermsUniverse[,2]
GOtermsUniverse =unique(GOtermsUniverse)
GOtermsUniverse[GOtermsUniverse!=""]
GOtermsUniverse <- GOtermsUniverse [!is.na(GOtermsUniverse)]
ALL_genes=GOtermsUniverse

library(SPIA)
res=spia(de=DE_genes,all=ALL_genes,organism="hsa",nB=2000,plots=FALSE,beta=NULL,combine="fisher",verbose=FALSE)
res$Name=substr(res$Name,1,10)
res[1:20,-12]

library(graphite)
prepareSPIA(reactome, "prepareSPIApathway", print.names = FALSE)
a=runSPIA(de=DE_genes, all=ALL_genes, "prepareSPIApathway")

TextOutput=c()
x=1
while (x<length(a[,1])+1){
```

```
if (a[x,][8][1]<=0.05){
TextOutput=c(TextOutput, a[,1][x])}
x=x+1}

x=1
FinalText=c()
while (x<length(TextOutput)+1){
FinalText=c(FinalText, TextOutput[x])
p=reactome[[TextOutput[x]]]
pEntrez <- convertIdentifiers(p, "entrez")
pathwayGenes=nodes(pEntrez)
for (z in pathwayGenes){
if (z%in%GeneNames){
FinalText=c(FinalText, z)}}
x=x+1}

write(FinalText, file = "AllDEtesting.txt",
ncolumns = if(is.character(x)) 1 else 5,
append = FALSE, sep = ",")
```

1.2 R code for differential expression and pathway analysis HuGene 1.0 ST and HuEx 1.0 ST Platforms.

#Quality Control

```
library(oligo)
library(annmap)
annmapConnect('human70')
setwd("C:/Users/Chris/Documents/Thesis/R/GSE26841_N1_Cell/Fib_4hr_sample2")
dat=read.celfiles(list.celfiles("C:/Users/Chris/Documents/Thesis/R/GSE26841_N1_Cell/Fib_4hr
_sample2"))
#Adjust the path to the CEL files
library("RColorBrewer")
usr.col=brewer.pal(9, "Set1")
mycols=rep(usr.col,  each=2)
hist(dat, lty=rep(1,length(rownames(pData(dat)))), col=mycols)
legend("topright", rownames(pData(dat)), lty=rep(1,length(rownames(pData(dat)))), col=mycols,
cex=0.6)
boxplot(dat,col=mycols,las=3,cex.axis=0.5,names=sampleNames(dat))
plm <- fitProbeLevelModel(dat, target= "probeset")
NUSE(plm)
RLE(plm)
eset <- rma(dat, target= "probeset")
```

# Gene Filter

```
controlProbesets <- getProbeInfo(eset, field=c('fid', 'fsetid', 'type'), target='probeset', subset= type == 'control->bgp->antigenomic')
controlProbesets = controlProbesets[,2]
controlProbesets=unique(controlProbesets)
eset2 = eset[featureNames(eset) %in% controlProbesets,]
#Create new eset2 with just control probesets

library(limma)
library(affyPLM)
TS <- gl(2,2,length=4, labels=c("control", "treatment") )
design<-model.matrix(~0+TS)
colnames(design)<-levels(TS)
fit1<-lmFit(eset2,design)
cont.matrix<-makeContrasts(contrast=treatment-control, levels=design)
fit1<-contrasts.fit(fit1, cont.matrix)
fit1<-eBayes(fit1)
topall2<-topTable(fit1, coef="contrast", number=nrow(eset2), adjust="BH", p.value=1,)
limmaDEgene<-topTable(fit1, coef="contrast", number=nrow(eset2), adjust="BH", p.value=1)
#p-value cutoff of 1. This step just gives all probe IDs and intensity.

holder=c()
x=1
while (x< dim(limmaDEgene[1])+1){
if (limmaDEgene[x,][[1]][1]%in%controlProbesets==TRUE){
holder=c(holder,limmaDEgene[x,][[3]][1])}
x=x+1}

Cutoff=quantile(holder, 0.85)
library(genefilter)
f1<-pOverA(1/3,Cutoff)
ff<-filterfun(f1)
index<-genefilter(eset,ff)
eset1<-eset[index,]
#Filters genes based on the negative control distribution.
```

# TEST

```
TS <- gl(2,2,length=4, labels=c("control", "treatment") )
#Adjust TS for each data set
design<-model.matrix(~0+TS)
colnames(design)<-levels(TS)
fit1<-lmFit(eset,design)
cont.matrix<-makeContrasts(contrast=treatment-control, levels=design)
```

```
fit1<-contrasts.fit(fit1, cont.matrix)
fit1<-eBayes(fit1)
topall2<-topTable(fit1, coef="contrast", number=nrow(eset), adjust="BH", p.value=1,)
limmaDEgene<-topTable(fit1, coef="contrast", number=nrow(eset), adjust="BH", p.value=0.05)
dim(limmaDEgene)
testinglist=limmaDEgene[,1]

# Annotation
library(annmap)
annmapConnect('human70')
geneListExp=probesetToGene( testinglist, as.vector=TRUE, rm.unreliable=TRUE )

#Pathway Analysis

library(annmap)
annmapConnect('human70')
inten=limmaDEgene[,1:2]
L=dim(limmaDEgene)[1]
EnsIntensity=c()
EnsID=c()
z=1
while (z<L+1){
geneList=probesetToGene(limmaDEgene[z,][[1]][1], as.vector=TRUE, rm.unreliable=TRUE )
if (is.null(geneList)){}
else {
if (length(geneList) ==1){
EnsIntensity=c(limmaDEgene[z,][[2]][1], EnsIntensity)
EnsID=c(geneList, EnsID)}
else{
for (x in geneList){
EnsID=c(x, EnsID)}
t=0
T=length(geneList)
while (t<T){
EnsIntensity=c(limmaDEgene[z,][[2]][1], EnsIntensity)
t=t+1}}}
z=z+1}

Together=data.frame(EnsID=EnsID, EnsIntensity=EnsIntensity)
averageInt=ave(Together[,2], Together[,1])
Together2=data.frame(EnsID=EnsID, EnsIntensity=EnsIntensity, averageInt = averageInt)
FinalTable= Together2[!duplicated(Together2["EnsID"]),]
FinalTable=FinalTable[,-2]
EnsList=as.vector(FinalTable[,1])
EnsListInt=as.vector(FinalTable[,2])
```

```
library(biomaRt)
mart <- useMart(biomart = "ensembl", dataset = "hsapiens_gene_ensembl")
GOterms = getBM(attributes=c('ensembl_gene_id','entrezgene'),filters='ensembl_gene_id',
values=EnsList, mart=mart)
GOterms=na.omit(GOterms)

OutputTable= data.frame(EnsID=GOterms$ensembl_gene_id, Entrez=GOterms$entrezgene,
Intensity=FinalTable[match(GOterms$ensembl_gene_id, FinalTable$EnsID ), 2])
GeneNames=OutputTable[,2]
DE_genes=OutputTable[,3]

RepeatGeneID = data.frame(GeneNames=GeneNames, DE_genes=DE_genes)
Average2=ave(RepeatGeneID[,2], RepeatGeneID[,1])
RepeatGeneID2=data.frame(GeneNames=GeneNames, DE_genes=DE_genes,
Average2=Average2)
RepeatGeneID2=RepeatGeneID2[!duplicated(RepeatGeneID2["GeneNames"]),]
GeneNames=RepeatGeneID2[,1]
DE_genes= RepeatGeneID2[,3]
names(DE_genes)=GeneNames
#This brings the intensity values over with the gene IDs.

#Get Gene Universe
TS <- gl(2,2,length=4, labels=c("control", "treatment") )
#Adjust TS for each data set
design<-model.matrix(~0+TS)
colnames(design)<-levels(TS)
fit1<-lmFit(eset1,design)
cont.matrix<-makeContrasts(contrast=treatment-control, levels=design)
fit1<-contrasts.fit(fit1, cont.matrix)
fit1<-eBayes(fit1)
topall2<-topTable(fit1, coef="contrast", number=nrow(eset1), adjust="BH", p.value=1,)
limmaDEgene<-topTable(fit1, coef="contrast", number=nrow(eset1), adjust="BH", p.value=1)
dim(limmaDEgene)
testinglist=limmaDEgene[,1]

library(annmap)
annmapConnect('human70')
geneList=probesetToGene( testinglist, as.vector=TRUE, rm.unreliable=TRUE )
universe=geneList

GOtermsUniverse =
getBM(attributes=c('ensembl_gene_id','entrezgene'),filters='ensembl_gene_id', values=universe,
mart=mart)
GOtermsUniverse=GOtermsUniverse[,2]
GOtermsUniverse =unique(GOtermsUniverse)
```

```
GOtermsUniverse[GOtermsUniverse!=""]
GOtermsUniverse <- GOtermsUniverse [!is.na(GOtermsUniverse)]
ALL_genes=GOtermsUniverse

library(SPIA)
res=spia(de=DE_genes,all=ALL_genes,organism="hsa",nB=2000,plots=FALSE,beta=NULL,combine="fisher",verbose=FALSE)
res$Name=substr(res$Name,1,10)
res[1:20,-12]

library(graphite)
prepareSPIA(reactome, "prepareSPIApathway", print.names = FALSE)
a=runSPIA(de=DE_genes, all=ALL_genes, "prepareSPIApathway")

TextOutput=c()
x=1
while (x<length(a[,1])+1){
if (a[x,][8][1]<=0.05){
TextOutput=c(TextOutput, a[,1][x])}
x=x+1}


x=1
FinalText=c()
while (x<length(TextOutput)+1){
FinalText=c(FinalText, TextOutput[x])
p=reactome[[TextOutput[x]]]
pEntrez <- convertIdentifiers(p, "entrez")
pathwayGenes=nodes(pEntrez)
for (z in pathwayGenes){
if (z%in%GeneNames){
FinalText=c(FinalText, z)}}
x=x+1}

write(FinalText, file = "AllDEtesting.txt",
ncolumns = if(is.character(x)) 1 else 5,
append = FALSE, sep = ",")
```