Identifying genes and pathways as an underlying cause of inverse comorbidity between cancer and Alzheimer's disease

By

Tayyaba Alvi

00000273864

MS-BI-3

Supervised by:

Dr Mehak Rafiq

School of Interdisciplinary Engineering and Sciences (SINES)

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

DEDICATION

Dedicated to my beloved parents, who gave me the little they had to ensure I have the opportunity of an education. Without their efforts, struggles, patience, encouragement and love, nothing of this would have been possible.

DECLARATION

I Tayyaba Alvi, hereby declare that work presented in this thesis is result of my own work except specific reference is made to the work of others wherever due. I also declare that the content presented in this thesis is original and have not been submitted in whole or in part to this university or to any other university for any other degree or qualification.

_____

Name: Tayyaba Alvi

Batch reg no: (NUST00000273864-MSBI-Fall18

ACKNOWLEDGEMENT

All the praises be to Almighty Allah, the most compassionate and the most merciful, who has bestowed upon me the power and ability to think and grow, empowering me to play my role in conveying a little share of my knowledge. I would like to express my deepest gratitude to my supervisor, Dr Mehak Rafiq whose sincerity and encouragement I will never forget. I am thankful for the extraordinary experiences she arranged for me and for providing opportunities for me to grow professionally. I am much obliged to my guidance committee members Dr Rehan Zafar Paracha, Dr Maria Shabbir and Dr Zamir Hussain has always been available for their humble assistance at various stages of my study and provided me with their valuable feedback and opinion.

I am immensely grateful for my parents whose constant love and support keep me motivated and confident. My accomplishments and success are because they believed in me. I am thankful to them for always teaching me great virtues. Deepest thanks to my siblings, who keep me grounded, remind me of what is important in life, and are always supportive of my adventures.

I would highly appreciate all the faculty members at SINES, NUST who have helped me directly or indirectly in the successful completion of my thesis and study. I am thankful to my friends Maleeha Ahmad, Aqsa Khalid, Farhana Riaz, Mehar Masood, Shayan Danish and Abeera Fatima for their moral support and inspiration. I highly appreciate the support, advice and continuous motivation of my friend Noor Us Subah. I am thankful to my research group at Data Analytics Lab for always encouraging and motivating me.

Table of Contents

List of Abbreviations

AD	Alzheimer's Disease
BAM	Binary Alignment Format
CDK	Cyclin Dependent Kinase
CI	Confidence Interval
CNS	Central Nervous System Disorder
DE	Differential Expression
DEG	Differentially Expressed Genes
EBI	European Bioinformatics Institute
EC	Oesophageal Cancer
ENA	European Nucleotide Archive
ERK	Extracellular signal-regulated kinase
ENCODE	The Encyclopaedia of DNA Elements
FOXO	Forkhead Box O
FPKM	Fragments per kilobase of exon per million mapped fragments
GBM	Glioblastoma
GO	Gene Ontology
GSEA	Gene Set Enrichment Analysis
HCC	Hepatocellular Carcinoma
HTS	High Throughput Sequencing
IC	Inverse Comorbidity
KEGG	Kyoto Encyclopaedia of Genes and Genomes

LC	Liver Cancer
MAPK	Mitogen activated protein kinase
MEK	Mitogen activated kinase
MTOR	Mammalian target of rapamycin
PCA	Principal Component Analysis
QC	Quality Control
SAM	Sequence Alignment Map

List of Figures

List of Tables

ABSTRACT

Epidemiological studies and clinical evidence suggest an inverse comorbidity between cancer and Alzheimer's Disease (AD). Several epidemiological studies and biological evidence suggested liver cancer (HCC) and Oesophageal cancer (EC) as most related to AD. Gene expression studies conducted on microarray platforms have also reported genes and biological pathways as inversely regulated between both diseases. However, to best of our knowledge no study has reported molecular level association between AD and EC. Therefore, this study conducts meta-differential expression analysis to get DEGs and pathways inversely regulated between cancer (Oesophageal cancer, Liver cancer) and Alzheimer disease. Two RNA-seq datasets from AD patients and controls from similar brain regions and four datasets from two different cancer types were subjected to differential expression analysis. Meta-analysis was performed using p-value combination method implicated in metaRNASeq. Intersection for inverse genes in each pair was calculated and significance of overlap was tested through Fisher exact test. Inversely regulated genes were then subjected to pathway analysis. Both cancer types (EC, LC) showed a significant overlap in gene expression deregulation in the opposite direction from AD. Functional enrichment comparison revealed several biological pathways which were affected jointly in both diseases, including PI3K/AKT/MTOR, GABAergic, Central Carbon Metabolism in Cancer and Metabolic pathways. Sixteen pathways were found to be deregulated in opposite direction. These results reinforce the previously proposed gene candidates and pathways like PI3K/AKT/MTOR and GABAergic synapse as contributing factor towards true inverse association between cancers and AD. It also reveals new potential candidates and pathways that can be involved.

Chapter 1

INTRODUCTION

In this chapter, details about terms and scope, background, general topic, and evaluation of current findings will be discussed. This will help in understanding the background, important terms, and value of this study.

1.1 Background

Inverse comorbidity (IC) occurs when the presence of one disease can prevent the risk of being susceptible to another. One such association exists between cancer and Alzheimer's disease (AD). Cancer and AD are among the most age-associated diseases and the leading cause of human death worldwide. The very first evidence for this association was surfaced from autopsy studies. Many patients with confirmed AD were observed to show a lesser probability of developing incidental cancer than non-AD patients. However, with these findings from epidemiological studies, it was not obvious if the lower rate of developing the disease than controls is due to some biological link between diseases or other factors such as mortality, age, diagnosis, and treatment are involved. Strikingly, findings from multiple epidemiological studies were highly consistent. Although there are certain treatment and confounding factors that might play a role in these observations. The amount of epidemiological evidence was enough to conclude true inverse comorbidity between AD and cancer. This finding provided the basis for researchers to analyse both AD and cancer from a new perspective. The main cause of AD is neurodegeneration while cancer is caused by uncontrolled cell proliferation. In this sense, both diseases are at opposite ends of two extremes of cell cycle, so they are considered to share many common genes and biological pathways. Although there have not been enough molecular level studies conducted on the subject, some studies have successfully identified genes and pathways significantly explaining the association. While reviewing cancer and AD association, several epidemiological studies and biological evidence suggested liver cancer (HCC) and Oesophageal cancer (EC) as most related to AD. According to a review, HCC and EC are associated with 51% and 33% lower risk of AD, respectively. Considering this evidence both cancer types were selected for this study.

1.2 Alzheimer's Disease

Alzheimer's disease (AD) is a neurological disorder that destroys neuronal cells. The early symptoms of AD include memory loss and delayed thinking skills and eventually damages the brain's ability to perform the simplest tasks. In most people, these late-onset symptoms appear in mid 60's. Late-onset AD is the most common type of dementia among the elderly. Another type of AD, early-onset AD is very rare and occurs between a person's 30's or mid 60's. The requisite disease has been named after Dr Alois Alzheimer. The changes in brain tissues were noticed by Dt Alzheimer in 1906 along with the formation of abnormal clumps. The clumps were actual extracellular beta amyloid plaques and intracellular neurofibrillary tangles (Tiwari *et al.*, 2019). Henceforth, till date extracellular beta amyloid plaques and intracellular neurofibrillary tangles are known as primary pathological hallmark of Alzheimer disease. Additionally, the loss of connections between neutrons in brain has been found another feature of Alzheimer disease. Neuron cells have been functioned for the transformation of message between parts of the brain and from brain to the muscles followed by different organs in the body. All above, a lot of other complex brain changes have had been found playing role in Alzheimer's disease. Therefore, the part of brain damage involved in memory initially takes place that includes hippocampus and entorhinal cortex. Furthermore, in the long term it effects areas of cerebral cortex responsible for reasoning, language and social behaviour eventually damaging other areas of the brain.

1.3 Cancers

Cancer is known as a disease in which body cells starts an uncontrollable growth which spreads to other parts of the body. Cancerous cell might affect any part of the body tissue with trillions of abnormal growths of the cells. The normal cell growth of human body is followed by cell division by multiplication for the formation of new cells in the body when needed. Therefore, the process involves cell growth followed by decay of old cells when become damaged and new cells are generated to take place when needed by the body. But every so often, the orderly process fails to happen, and damaged cell start growing abnormally, divide and spread abnormally without self-destructing as programmed to do so. Further, the cells might form tumours called lumps of tissue. These tumours can be cancerous and might not, normally known as benign.

Furthermore, prevalence of cancer is genetic caused by changes in the gene that are responsible for the cell function specifically their growth and division. Moreover, the genetic cause of cancer is the error that occurs as cell divide and cause damage to DNA due to harmful environmental substances. These include toxic compounds and chemicals found in tobacco smoke and UV rays from sun and mutations inherited in the genetic pathway from family. Furthermore, one of the reasons of later age cancer has been found as the body eliminates damaged DNA before it turns cancerous, but this ability goes down with aging. For this reason, there is a high chance of cancer in later life.

1.4 Two opposite extremes of the same paradigm

Because of their opposing nature AD and cancer has attracted the imagination of many researchers worldwide. The main cause of AD is neurodegeneration while cancer is caused by uncontrolled cell proliferation. In this sense, both diseases are at opposite ends of two extremes of cell cycle. In the process of cell survival and cell death, there are certain mechanics involved. These can be mostly associated with cell cycle regulation machinery, mutations, DNA damage and immune system responses. In this regard, they are considered to share many common genes and biological pathways. In an attempt to analyse the role of these factors in the inverse comorbidity of cancer and AD, many of these genes and pathways were seen to be regulated in opposite directions. While there is a plethora of research available for cancer, this is not true in case of AD. However, the corelation of both these diseases with each other and involvement of a common primary culprit can provide a basis to explore AD from new perspectives.

Although the molecular patterns involved in these diseases have been explored individually, there has not been enough research carried out in exploring their association. This information makes it crucial to explore biological factors responsible for the association.

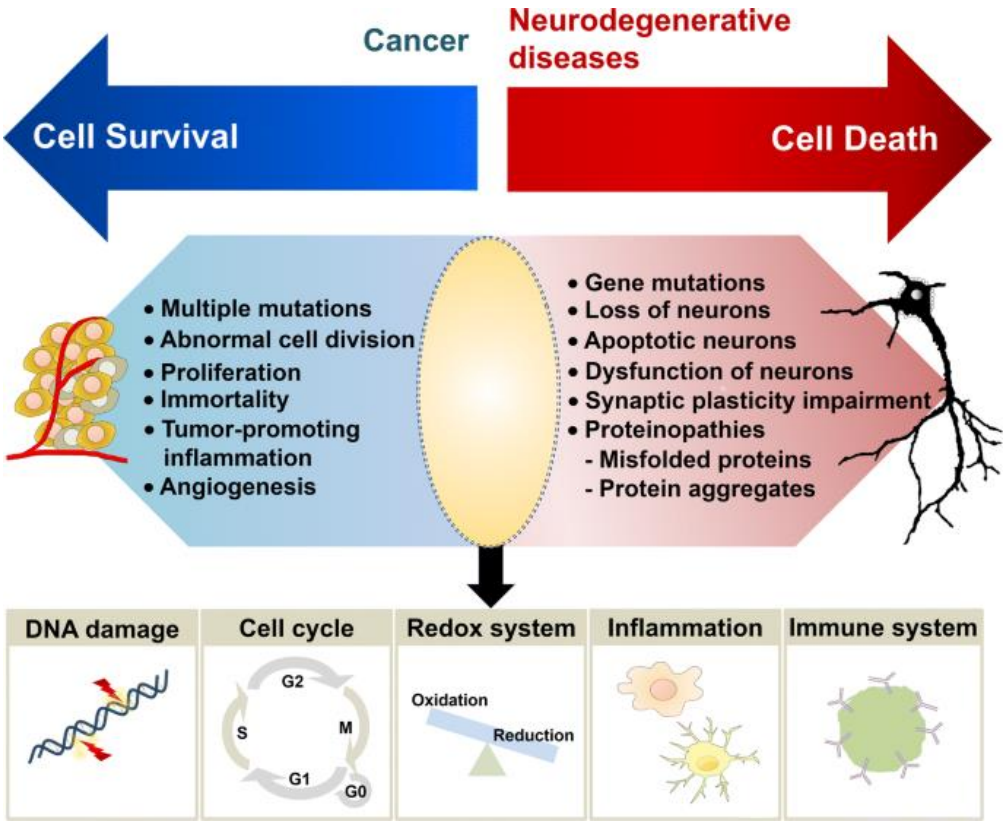


Figure 1-1: Featured hallmarks of cancer and neurodegenerative diseases. Pathophysiological features of two representative age-related diseases, cancer and neurodegenerative diseases are shown. Mechanisms are inversely undergone in these two diseases to lead to cell survival and cell death in cancer and neurodegenerative diseases, respectively. DNA damage, cell cycle aberrations, redox imbalance, inflammation, and immunity are closely associatedging shared characteristics between cancer and neurodegenerative diseases [2].

1.5 Cell cycle in AD

In general, every cell in the body goes through mitosis at some point. The cell cycle is being regulated constantly in a controlled environment. There are four phases of the cell cycle namely G1 which involve cell growth, S phase involves DNA synthesis, G2 involves cell growth and eventually M phase also known as the mitotic phase where the cell divides. However, once the neuronal cells mature, they do not divide and remain in a dormant state for the rest of their lives. The reason behind this is still not fully understood but it has been suggested that a non-canonical pathway, mediated by DNA polymerase beta is involved. It was observed that degenerative neurons in affected areas of the brain have replicated their DNA, but no similar pattern was observed in other healthy parts of the brain. According to a hypothesis, an abnormal re-entry into cell cycle is the cause of neurodegeneration.

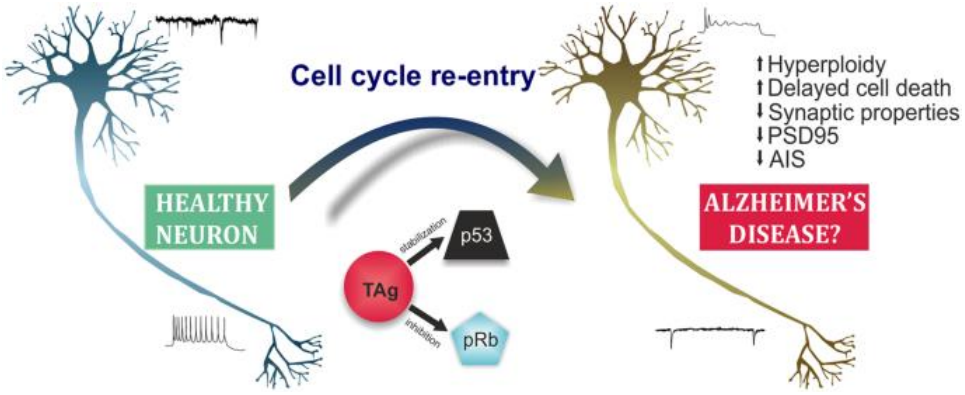


Figure 1-2: Scheme summarizing induced cell cycle re-entry in healthy neurons. The capacity of TAg to inhibit p53 might help to prevent apoptosis in neurons that undergo cell cycle re-entry, which nevertheless undergoes delayed non-apoptotic cell death. Cell cycle re-entry elicits functional changes in neurons, which might contribute to cognitive impairment and neuronal death susceptibility as observed in AD. [3]

1.6 Cell cycle in cancer

Cell cycle division involves a series of coordinated events, illustrated in **Figure** 1-3: Molecular Aspects of the Mammalian Cell Cycle and Cancer [4]. There are two types of cell cycle control mechanisms. One mechanism involves a cascade of protein phosphorylations that programmes cells from one stage to another and set checkpoints that regulate the completion of crucial events. The proteins involved in regulating the cell cycle mostly belong to the kinase family. These kinase proteins usually require association with another protein subunit called cyclin. The cyclin-dependent kinase (CDK) combines with its cyclin partner and creates an active complex with substrate specificity. To ensure well regulation and transition of cell cycle stages, the CDK-cyclin complex's activity is monitored through phosphorylation and dephosphorylation.

The second type of cell cycle regulation and control is through signals, called cell cycle checkpoints. Cell cycle checkpoints detect abnormalities in crucial events as chromosomal segregation and DNA replication. When an abnormality is detected through these signals, the progression to the next stage is delayed until the effect of mutation or abnormality is averted. As compared to CDK's, the extent of checkpoints signal is not as obvious because they are not involved in every cell cycle.

Since cell cycle machinery controls cell division and proliferation, the relation between cancer and the cell cycle is obvious. Interestingly, all cancers exhibit abnormal cell cycle behaviours caused

by insensitivity to signals and mutation in regulatory proteins. Almost in all types of cancer, there is a combination of two or more faulty regulatory mechanisms. Thus, there can be several genes and pathways involved in regulation and signal transduction that leads to cancer.

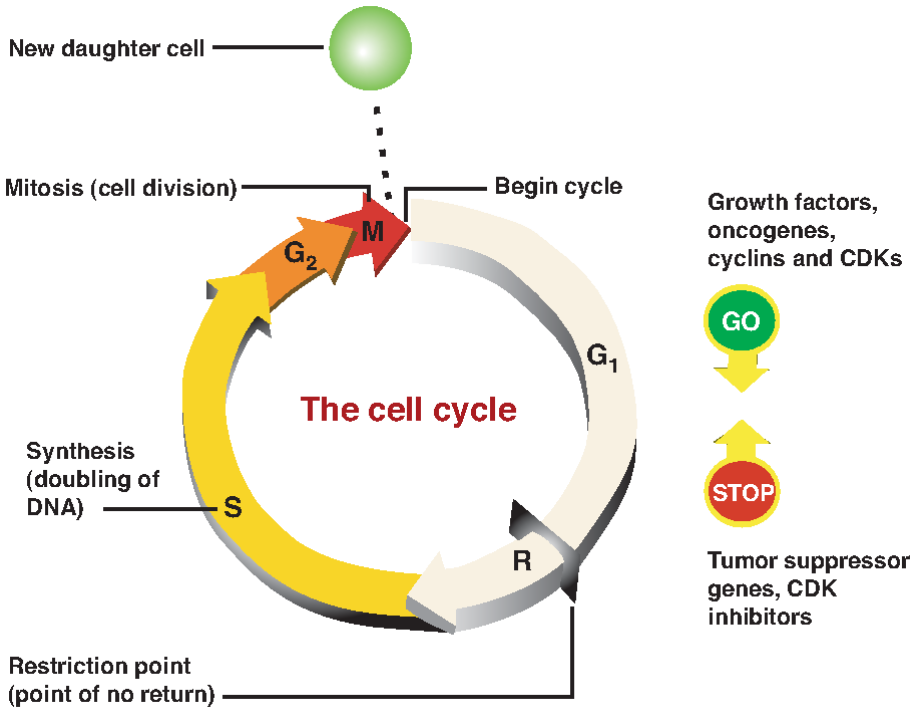


Figure 1-3: Molecular Aspects of the Mammalian Cell Cycle and Cancer [4].

1.7 Gene expression, a fundamental intermediate phenotype

Gene expression is the fundamental link between DNA sequence and individual-level phenotypes such as disease susceptibility. Gene expression quantification techniques are used to measure gene expression from mRNA transcripts abundances. There is a plethora of research available to infer certain phenotypes by measuring gene expression.

In the human body, there are approximately 20,000 protein-coding genes as annotated by the GENCODE project [5]. A gene is comprised of certain regions such as coding and non-coding sequences, regulatory regions such as promoters and other distal regulatory elements such as silencers and enhancers[6], [7]. The protein-coding sequences of a gene are responsible for two later parts of central dogma, transcription, and translation. In the process of transcription, RNA polymerase transcribes the coding or sense strand of the DNA in 5’ to 3’ direction to produce a

single strand RNA complement. The sequence of a eukaryotic gene typically consists of exons and introns. During post transcription processing introns are spliced out by spliceosome and exons are retained, which are the protein-coding parts of a gene. Although a great proportion of the human genome does not code for proteins, more than 80% of the genome have assigned biochemical functions. These processes are mostly believed to be of regulatory importance. Gene expression is a tightly regulated process, it makes certain that genes are expressed in the correct amount in certain tissues or cell types at the appropriate developmental stage. Therefore, hidden patterns of gene regulation can be of importance when explaining the behaviours of genes.

1.7.1 Regulation of gene expression

In the human genome, there are around 20,000 genes. The genetic material of different cell types is fundamentally similar, but the expression of genes varies from cell to cell. Moreover, the same cell can have a different expression over different developmental stages or different conditions. This is because genes are being regulated differently across cells and tissue types. For instance, some genes called housekeeping genes that are involved in fundamental cell processes such as DNA production enzymes and glycolysis etc, are expressed in almost every cell. On the other hand, tissue-specific genes are expressed in certain types of cells and not in others like neurons. To understand tissue functions, it is crucial to understand the pattern of gene expression. In addition, the role of this spatiotemporal pattern of gene expression is crucial in our understanding of cell and tissue behaviour under pathogenesis [8]. Gene regulation is a tightly controlled process that determines which gene to turn on and off. There are many cellular and genetic factors involved in the regulation process from early developmental stages to the death of an organism. Several approaches are being used to understand different gene regulatory patterns.

Recent advancement in high-throughput gene expression profiling technologies has enabled us to understand and study gene expression regulation patterns. The Genotype-Tissue Expression (GTEx) project is to date the largest resource to characterise tissue-specific regulation of transcriptome (‘Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multi-tissue gene regulation in humans.’, 2015; Aguet *et al.*, 2017). Gene expression regulation is crucial for cellular activities. It prepares cells to respond to external signals as well internal stimuli. It is a collection of comprehensive processes at various stages of transcription and translation. For example, the production rate of mRNA and proteins can be influenced by chromatin accessibility

which can be affected by methylation, histone modification, splicing events, mRNA stability and at last post-translational modifications [11].

Regulatory components can be classified into two types: cis- and trans-regulatory elements, with the former being adjacent DNA sequences and the latter being DNA-binding proteins including transcription factors (TFs). Typical cis-regulatory elements include promoters and enhancers, and their epigenetic modifications (e.g., histone modifications) and binding proteins that affect transcription initiation and regulation [12]. TFs can increase or suppress the transcriptional activity by binding to cis-regulatory elements of target genes. TFs contain DNA-binding domains that target specific DNA sequences on the genome [13]. In addition, genes in the same biochemical pathway interact with each other and are often co-regulated, and a network framework can provide improved modelling of the relationships among transcripts. Thus, a better understanding of regulatory patterns of genes can provide a better insight to study their expression in different tissue types.

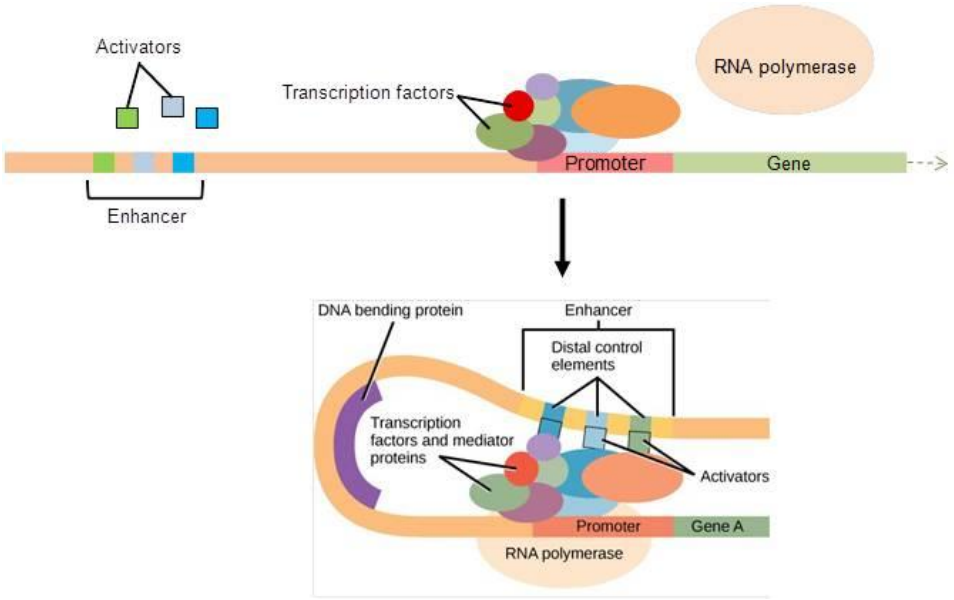


Figure 1-4: Each gene has a promoter upstream of the coding sequence. The promoter binds to transcription factors and helps RNA polymerase to bind and start transcription. Bottom. Many genes also have upstream enhancers. Enhancers bind activators, bend around, and help RNA polymerase start transcription. (Matt. i, no date)

1.7.2 High-throughput transcriptome data

A transcriptome is defined as the complete set of RNA transcripts in one or a group of cells, including mRNAs and non-coding RNAs. Unlike the genome, which is fixed in an organism, the

transcriptome varies, sometimes dramatically, across cell types, developmental stages, and external environmental stimuli. Thus, characterising transcriptomes provides insights into the function of genes, gene expression variation, and gene regulation under different conditions. Two main high-throughput technologies are used to quantify gene expression profiles: hybridisation-based approaches such as oligonucleotide microarrays and sequence-based approaches such as RNA sequencing (RNA-seq) [15], [16]. Microarray gene expressing profiling measures gene expression levels by hybridisation between probes attached to specific spots on the array and complementary DNA (cDNA) synthesised from mRNA molecules in given samples. Like other types of DNA microarrays, probes on the array are designed based on sequences of identified and annotated genes. The probe sequences can uniquely match certain gene transcripts. Meanwhile, RNA-seq, 10 based on recent deep-sequencing technologies, does not rely on prior knowledge of gene sequences. It has advantages in the aspect of, for example, identifying novel gene isoforms derived from alternative splicing, as well as allowing for measuring abundances of different gene isoforms. RNA-seq can also provide information on allele-specific expression at heterozygous sites [17]. Unlike microarrays that are prone to probe saturation, RNA-seq can detect a broader range of gene expression levels. Microarrays have been more widely adopted due to the lower cost and relatively advanced methods. While both RNA-seq and microarrays are affected by technical variations such as batch effects, those affecting microarrays tend to be better understood and can be corrected statistically. RNA-seq used to be less affordable compared with microarray platforms; however, the cost of RNA-seq is decreasing and it is becoming more widely used in this field.

1.8 Problem Statement

Epidemiological studies have suggested an inverse association between Oesophageal cancer and Alzheimer's disease. However, no study has reported this association with respect to inverse gene expression. Therefore, this study aims to identify genes and pathways involved in inverse comorbidity between AD and cancer using the RNA-seq approach.

1.9 Aims and objectives.

1. Identifying differentially expressed genes in AD, EC, and HCC through differential expression meta-analysis.

2. Identifying inversely expressed genes between AD & cancers.
3. Testing for significance of true inverse association by fisher exact test.
4. Elucidating the role/effect of genes and pathways in both phenotypes through gene set enrichment analysis.

Chapter 2

LITERATURE REVIEW

In this chapter, the current knowledge, findings, as well as theoretical and methodological contributions which helped in formulating and conducting this study will be discussed.

2.1 Disease Comorbidity

Disease Comorbidity is defined as the presence of more than one disease in an individual, it is also known as multimorbidity [18]. The comorbid diseases can exist together, or one disease can follow the other after a period. In most conditions, these diseases can seem not to be linked but rather independent of each other, but comorbid diseases influence each other [19].

Disease comorbidity has emerged as a major problem in treatment, as treating patients with multiple diseases is complicated because it requires certainty in diagnosis and treatment [20], [21]. Moreover, patients with comorbid diseases have a higher rate of longer hospitalisation and mortality which is time and cost consuming [22]–[24]. Furthermore, if the patient receives multiple drug treatments for multiple conditions at the same time, it might cause serious side effects due to drug interactions [25], [26]. However, the underlying mechanisms and patterns of disease comorbidities are not well studied [27]. Therefore, in recent years, it has emerged as a hot topic in scientific research both from the molecular level and clinical observation perspectives. Comorbidity can be further divided into two types named as inverse and direct comorbidity.

2.1.1 Direct Comorbidity

Direct comorbidity refers to the type of comorbidity in which diseases co-occur in a way that the presence of one disease can increase the chance of acquiring the other. For example, an increased risk of diabetes mellitus type 2 is associated with an increased risk of cardiac arrest [28]. Direct comorbidity was also observed between Alzheimer's disease (AD) and glioblastoma [29]. Although there can be certain biological and non-biological factors contributing to this association, direct comorbidities are often observed in age-associated diseases.

2.1.2 Inverse Comorbidity

Inverse comorbidity (IC) refers to the type of comorbidity in which diseases co-occur in a way that the presence of one disease can reduce the chances of having the other [30]. This type of relation among diseases protects individuals from a condition when they are exposed to the other. The paradoxical nature of IC violates common sense thus attracting the attention of many scientists and researchers. The concept of IC is based on these two premises: firstly, certain non-biological and environmental factors make certain conditions protective of another. Secondly, a type of disease such as Alzheimer's disease (AD) has certain biological aspects (genes, pathways) involved, which can protect against another disease such as cancers [30]. Therefore, these factors can influence an individual with disease A meanwhile protecting the same individual against disease B.

Several genes have been studied for their inverse role between different diseases. One of such examples is DSCR1 (Down's syndrome candidate region 1) which is involved in Down's syndrome but is protective against solid tumour mechanism. The encoded protein by DSCR1 suppresses angiogenesis signalling. As angiogenesis is an important mechanism for tumour growth, overexpression of DSCR1 in Down's syndrome can cause suppression of angiogenesis, as a result protecting the individual from cancers [31].

2.2 Inverse Comorbidity between cancer and Alzheimer's disease

2.2.1 Epidemiological evidence

Cancer and AD are the two most common age-associated diseases. However, little was known about their association until recently. The very first evidence for this association was surfaced from autopsy studies. Many patients with confirmed AD were observed to show a lesser probability of developing incidental cancer than non-AD patients [13],[14]. This was intriguing since there might be a higher probability to suffer from an undiagnosed disease for patients who are unable to report symptoms, based on the previous disease record. In a study on Japanese atomic bomb survivors who are at increased risk of cancer, it was observed that patients who were clinically diagnosed with AD had a 70% less prior cancer history than age-matched controls [34]. Several studies were conducted carefully following these observations which also confirmed an inverse association between AD and cancers. In 2005, a longitudinal study conducted on 6,000 participants over a

period of 10 years also confirmed that a history of cancer reduced the risk of AD. Moreover, the prevalence of AD was also observed to be associated with a reduced risk of cancer [35]. Interestingly, both studies did not find out any such association between AD and vascular dementia (VaD). A similar epidemiological study in the United States over a period of 15 years further validated this association [36].

However, with these findings from epidemiological studies, it was not obvious if the lower rate of developing the disease than controls is due to some biological link between diseases or other factors such as mortality, age, diagnosis, and treatment are involved. In 2011, the Framingham heart study with clearly defined cohorts of cancer and AD patients concluded an inverse comorbidity rate of 0.67 with a 95% CI (confidence interval) This observation was compared with the cohort excluding patients who died before the age of 80 and concluded similar statistics (IC= 0.67; CI= 95%) [37]. Another study from Italy further validated that the inverse association was present both before and after diagnosis as well it is independent of survival and mortality, demonstrating that incidence of AD is associated with a 50% lower risk of cancer and a history of cancer is associated with 35% lower risk of AD [38]. The inverse association with cancer was also reported in other neurodegenerative diseases such as Parkinson's disease [39], [40]. However, all these studies reported no such association between cancers and VaD. A study considered the possibility of both direct and inverse comorbidity between cancer and AD. Metanalysis on two distinct types of cancers was carried out which concluded findings aligning with previous studies that AD is associated with a lower risk of lung cancer but higher risk of glioblastoma [41]. It should be noted that glioblastoma was included in that study as a type of brain cancer which in term can disrupt the homeostasis of the brain by affecting it directly. Hence, resulting in a direct correlation with neurodegenerative diseases such as AD. Several other studies on different ethnic groups also produced similar statistics.

Strikingly, findings from multiple epidemiological studies were highly consistent. Although there are certain treatment and confounding factors that might play a role in these observations. The amount of epidemiological evidence was enough to conclude true inverse comorbidity between AD and cancer. This finding provided the basis for researchers to analyse both AD and cancer from a new perspective. Although there have not been enough molecular level studies conducted on the subject, some studies have successfully identified genes and pathways significantly

explaining the association. A better understanding of the underlying mechanism for inverse comorbidity between cancer and AD could help to develop improved treatments.

2.2.2 Gene Expression and Meta-Analysis studies

Because of their opposing nature AD and cancer has attracted the imagination of many researchers from around the globe. Several hypotheses have surfaced explaining the biological mechanism behind this association. Some hypotheses have been mentioned earlier in chapter 1. Moreover, some studies have explored genes and pathways deregulated in opposite directions. Also, many transcriptomic level studies have explored most types of cancers and AD individually. However, there are only a few studies conducted on gene expression analysis to validate inverse comorbidity between these two at the molecular level.

The earliest study in this regard was carried out on gene expression data from these 3 different cancer types, Lung (LC), Colorectal (CRC) and Prostate (PC) with 3 different CNS (Central Nervous System) disorders AD, PD and SCZ (Schizophrenia) [42]. Microarray expression data for all disease types were collected from publicly available datasets and a meta-analyses approach was used to measure gene expression across multiple samples. Most of the common genes found between CNS disorders and cancers were found to be deregulated in opposite directions. Previously reported genes for their role in inverse comorbidity such as PIN1 and P53 expression was analysed and confirmed that it was downregulated in AD and PD and upregulated in CRC. Another gene that stood out as an indicator of strong inverse association was ATP13A2 gene. Previously, loss of functional mutation in ATP13A2 have been linked to early-onset PD and some somatic mutations have been seen in the case of cancer [43]. It was also identified to be upregulated in all three types of cancer and downregulated in AD and PD. Overall, the study reported 74 genes were reported to be downregulated in all CNS disorders and simultaneously upregulated in all cancer types. On the other hand, 19 genes were reported as upregulated in CNS disorders but downregulated in cancer types. **Table 2-1**

Table 2-1 DEGs are significantly deregulated in opposite directions in the three CNS disorders and three types of Cancer [42].

DEGs significantly upregulated in the three CNS disorders and downregulated in the three Cancer types (q-value<0.05).
"MT2A, MT1X, NFKBIA, AC009469.1, DHRS3, CDKN1A, TNFRSF1A, CRYBG3, IL4R, MT1M, FAM107A, ITPKC, MID1, IL11RA, AHNAK, KAT2B, BCL2, PTH1R, NFASC"
DEGs significantly downregulated in the three CNS disorders and upregulated in the three Cancer types (q-value<0.05).
"(PPIAP11, IARS, GGCT, NME2, GAPDHP1, CDC123, PSMD8, MRPS33, FIBP, OAZ2, IARS2, SLC35B1, APOO, TMEM189-UBE2V1, VDAC1, TMED3, SMS, DNM1L, PRPS1, SRSF2, TMEM14D, TOMM70A, ATP6V1C1, NUP93, MRPL15, UBA5, PPIH, SMYD3, NIT2, SRD5A1, NUDT21, MRPL12, EEF1E1, MRPS7, TTPAL, BZW1P2, RP11-552M11.4, TSN, MECR, ZWINT, RPRD1A, UCHL5, NHP2P2, TFB2M, FEN1, CGREF1, IMPAD1, ARL1, ACLY, MRPL42, LSM4, KPNA1, TIMM23B, RP11-164O23.5, RP11-762H8.2, FARSA, MRPL4, API5, RP3-425P12.4, RFC3, RANBP9, TFCP2, GMDS, CCNB1, TMEM177, GUF1, HSPA13, NMD3, GCFC2, TUBGCP5, TBCE, YKT6, PHF14, BRCC3..."

In order to elucidate the role of these genes on molecular bases pathway enrichment analyses was performed. To account for an inverse association, pathways that were being deregulated in opposite directions in at least one disease as compared to the 3 in the other group were considered. Interestingly, out of 30 pathways, 24 were reported to be deregulated in opposite directions, including the p53 signalling pathway [42]. A detailed insight into these pathways resulted in most pathways involved in cellular processing, genetic information processing, environmental information processing, meta metabolism, organismal systems. Although this study provided significant insight and a molecular basis to further explore the behaviour of candidate genes and pathways, it is debatable whether the expression changes observed are due to treatment or some other environmental factor instead of a true inverse comorbidity role.

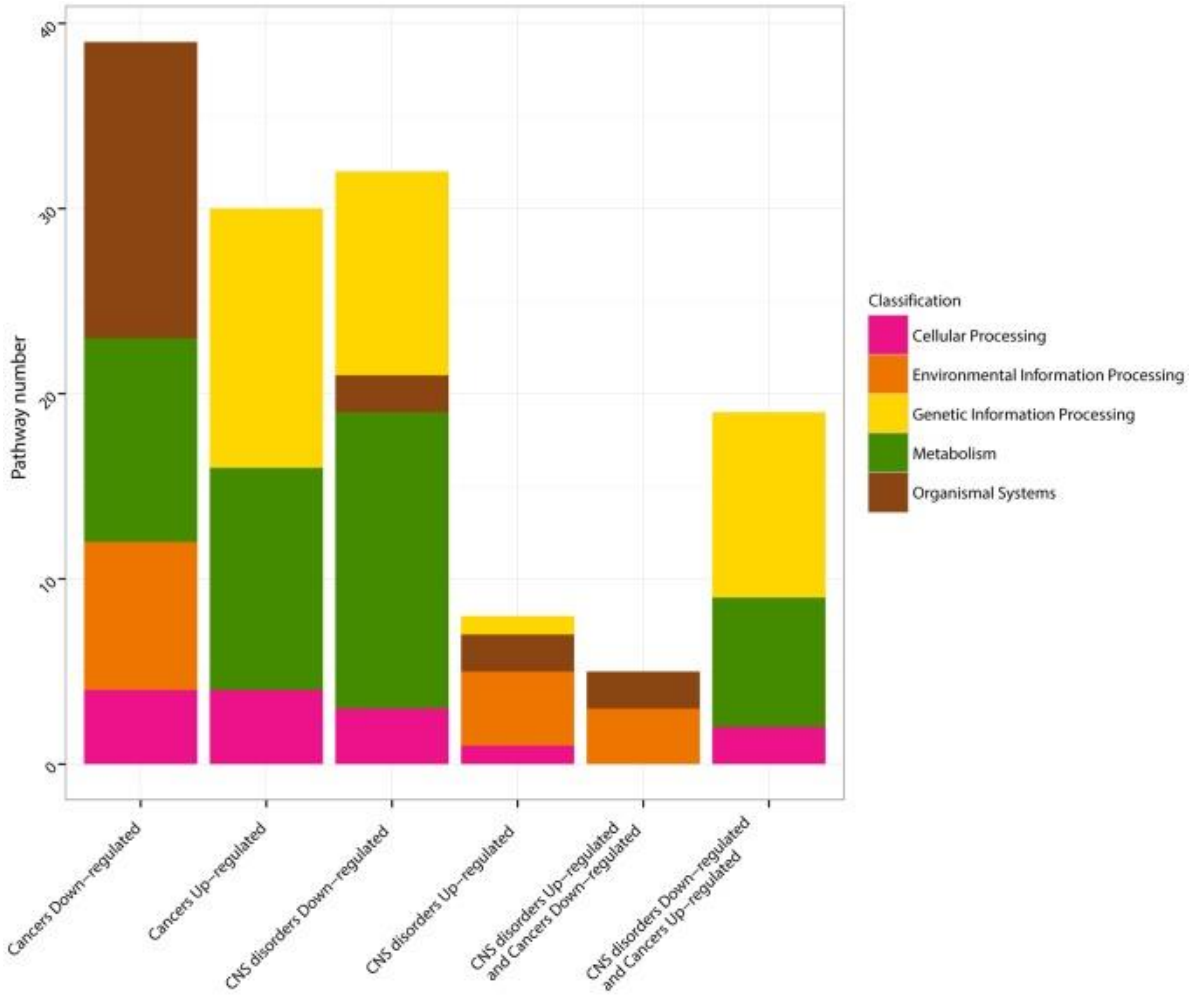


Figure 2-1 KEGG (Kyoto Encyclopaedia of Genes and Genomes) pathways identified by GSEA (Genome Set Enrichment Analysis) as significantly deregulated in cancers and CNS disorders [42].

A study conducted on data published by Ibáñez et al [42] to investigate the role of differential gene deregulation with respect to protein formation. The idea was to analyse the physicochemical properties of deregulated genes and investigate the possibility to distinguish CNS disorders and cancer based on structural disorder. The study concluded that structural disorder is a significant factor to differentiate cancer and CNS disorders [44].

The most recent gene expression study to investigate inverse comorbidity between AD and cancers revealed a considerable number of genes deregulated in opposite directions in AD and lung cancer. Transcriptome meta-analyses study investigated AD association with Lung Cancer (LC) and Glioblastoma (GBM) [29]. Glioblastomas are the most common type of brain tumours in adults [41]. Previously epidemiological studies have reported a direct association between AD and GBM

[37], [45]. Similar meta-analyses differential expression techniques were applied to check comorbidities between AD & GBM and AD & LC. The results were consistent with epidemiological findings as more genes were observed being deregulated in opposite direction in the case of AD and LC, whereas most genes were observed being deregulated in the same direction in the case of AD & GBM.

Figure 2-2 Significantly differentially expressed genes (sDEG) overlap between AD, GBM and LC. The overlap is shown here on deregulation in opposite and similar directions between all three diseases in separate groups [29].

The molecular significance of inverse comorbidity between AD and cancer is evident from these transcriptome expression analysis studies. To further investigate the interesting behaviour of involved genes, a more in-depth analysis is required.

2.3 Motivation

Advancement in bioinformatics research methods has made it possible to explore gene expression data. There are multiple tools and pipelines available for gene expression estimation chapter 1. This approach can be implemented to elucidate the role of individual genes and pathways related

to a certain phenotype which can also be comparable to gene expression changes directly. Thus, an extensive literature review has helped to formulate the following findings which are the core motivation of this study.

- Epidemiological studies were directed towards finding a correlation between AD and Cancers.

- Several studies concluded the existence of inverse comorbidity between two diseases.

- Gene Expression studies targeted to identify inversely expressed genes that might explain the biological plausibility of this relationship.

- There are 3 transcriptome level studies, that were able to identify molecular cause of inverse association. However, none of these studies included Oesophageal cancer that is a highly correlated cancer with AD. Therefore, this study aims to find molecular cause of this association.

Chapter 3

METHODOLOGY

This chapter describes in detail the materials and methods used in this study to achieve previously formulated objectives. The datasets used in this study are all secondary datasets that are publicly available on GEO NCBI and ENA (European Nucleotide Archive). The whole analysis was carried out on a Linux environment using Ubuntu (18.08). All the methods, tools and pipelines were optimized before use.

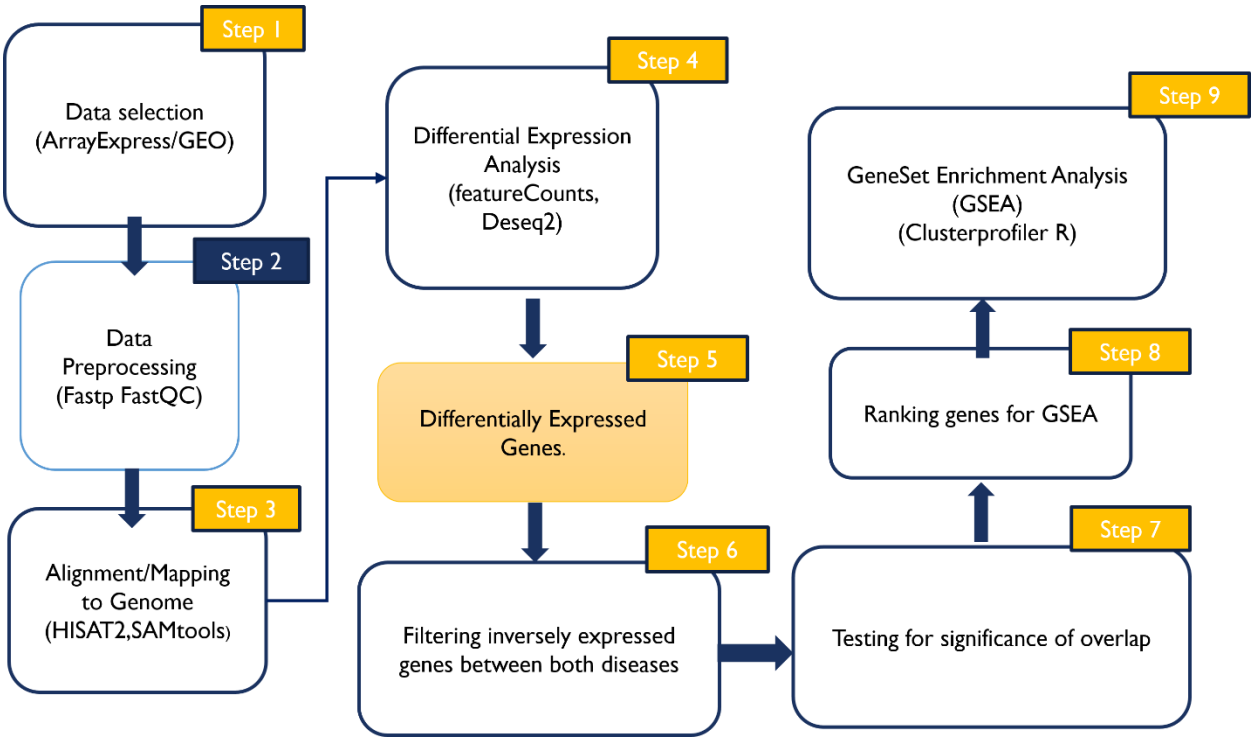


Figure 3-1 Overall workflow of methodology

The core objective of this study is to find out SNPs (Single Nucleotide Polymorphism) present in inversely regulated genes in cancers and Alzheimer's disease which might explain the opposite regulation of the cell cycle. To achieve this, the first step is to find deferentially expressed genes that are being regulated in opposite directions and then perform variant analysis to find and compare SNPs present in those genes. Following methods were applied to achieve this objective.

3.1 Data Acquisition

All the data used in this study is secondary and downloaded from different online repositories such as GEO NCBI, SRA, and ENA. RNA-seq datasets were the major requirement of this study which were collected for three different diseases. Apart from RNA-seq datasets, some other necessary data files were downloaded. The types of data required, and their properties are described in detail in this section.

3.1.1 RNA-seq Data

The major purpose of transcriptome level sequencing is to study transcript level changes. However, several bioinformatics pipelines have made it possible to analyse variants through RNA-seq data. This study was conducted to analyse and identify variants present in inversely expressed genes in two different types of disease. RNA-seq data sets were downloaded from online repositories GEO and ENA in FastQ format. To limit the false positives and for more accurate variant detection only paired-end data was used. Data selection criteria are as follows.

- Homo Sapiens & RNA seq
- No cell lines
- Normal and Disease
- Not treated and knockdown
- Age-matched controls
- Paired end
- Preferably adjacent normal

According to this criterion, the following datasets were selected and downloaded.

	Accession	No. of samples	Platform
Alzheimer's Disease	GSE53697	17 (8 N, 9 AD)	Illumina HiSeq 2500
	GSE95587	30 (11 N, 19 AD)	Illumina HiSeq 2500
Oesophageal Cancer	GSE130078	46 (23 N, 23 T)	Illumina HiSeq 2000
	GSE111011	14 (7 N, 7 T)	Illumina HiSeq 2500
Liver Cancer	GSE197214	18 (9 N, 9 T)	Illumina HiSeq 2000
	GSE105130	50 (25 N, 25 T)	Illumina HiSeq 2500

Table 3-1: RNA-seq Datasets for Alzheimer's Disease, Esophageal Cancer, Liver cancer. Accession number, sample number, type and sequencing platform are given.

3.1.2 Reference Genome and Annotation Files

The human reference genome (also known as reference assembly) is representative of all the genes present in a human species. It is a digital nucleic acid sequence database, assembled by researchers as a representative example of a human set of genes. Reference genomes of species are freely accessible online at various databases, using specified browsers, for example, Ensembl or UCSC Genome Browser [46]. The Human Reference genome has different versions but the most recent one is hg38 [47]. A genomic feature file used for annotation purposes is a tab-delimited file that contains specific information about genome coordinates. Both genome reference file (fasta format) and annotation file (GTF format) were downloaded from the ensemble.

3.1.3 Genomic Variation Files for Variant Analysis

In response to the general need of cataloguing human genome variation data, several databases have been designed that stores variation data from a large set of samples. These databases are then used for association studies. One of such projects is the 1000 genome project where they have

reconstructed the genomes of 2504 individuals from 26 different populations [48]. Another such project specifically was designed to catalogue data for human single nucleotide polymorphism by NCBI [49]. The dbsnp database is being widely used for genome variation data specific to SNPs. In this study, both were used to access genomic variation data needed for variant analysis.

3.2 Differential gene expression analysis

The selected datasets for this study were subjected to differential gene expression analysis, which involves performing statistical analysis on read count data to discover quantitative changes in two or more different experimental groups. The workflow for the analysis is given below followed by the steps performed during this analysis described in detail.

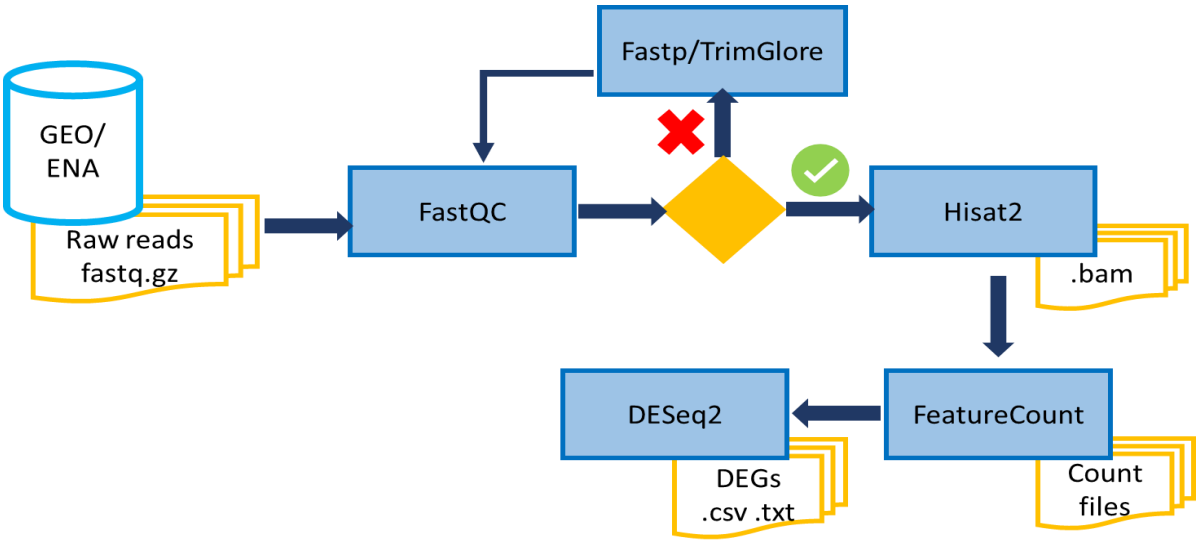


Figure 3-2 Differential Expression Analysis workflow

3.2.1 Quality Control and Pre-Processing

Data quality assessment is a crucial step when working with any genomic data. Before going into further downstream analysis, data quality is evaluated based on certain parameters. Many tools are available to check and correct data quality. For high throughput sequencing data, FastQC is one of the best and most used quality assessment tools [50].

3.2.1.1 FastQC

FastQC is designed to detect and highlight problems in a wide variety of high throughput sequencing data. It provides a set of analyses that can be used to give a quick assessment of data that if it has any problems before going into downstream analysis. FastQC takes a FASTQ, BAM or SAM file as input and provides summaries about different quality check matrices. It also provides an HTML report to view certain quality check graphs. Among the analysis modules provided by FastQC, "per base sequence quality" and "Overrepresented sequences" are the most important checks to minimize error rates in downstream analysis. "Per base sequence quality" gives insight into the quality score distribution of all bases on each position of a read. Whereas "Overrepresented sequence analysis" helps in identifying contaminations.

In this study, all the raw FASTQ files were subjected to FastQC by running the following script which allows analysis of multiple files.

Sample command:

for f in /RawData/.fastq.gz; do fastqc $f --outdir /fastqc -p 32 done*

3.2.1.2 Fastp

Once the quality control step has identified problems with data, the next step is to perform pre-processing to correct for technical biases detected by FastQC. There can be certain quality check fails for which pre-processing can be applied such as adapter sequences, low base quality, over-represented sequences, duplication level, N bases and poor-quality score distributions.

Fastp is an ultra-fast all in one pre-processor for problems in FASTQ data. It is a c++ developed tool and caters to many FASTQ quality problems in a single scan. It can also use up to 16 multiple threads to allow fast processing. Despite it is multiple functionalities, Fastp has proven to be 2-3 times faster than other pre-processing tools such as Cutadapt and Trimmomatric [51]. Fastp has many options to perform analysis, however, it has most of the options enabled by default. To disable or enable options depending on the data quality reports user can use certain flags e.g. Over-representation analysis is disabled by default, in case of the presence of over-represented

sequences in data flag -p can be used. In this study. Fastp was installed using GitHub source, following command was used to pre-process files individually.

Once the pre-processing was done, quality control was again applied to the preprocessed file. In case of successful pre-processing operation, data were subjected to further analysis. Any failure in the quality control step demands more pre-processing. The following command was used to process fastq files.

Sample command:

fastp -i SRR10066648/SRR10066648_1.fastq.gz -o LCSample2.fastq.gz -I SRR10066648/SRR10066648_2.fastq.gz -O LC2sample2.fastq.gz --correction -p -w 16 -z 5

3.2.2 Mapping

After quality checking and trimming of sequences, the following step is the sequence alignment to the reference genome. HISAT2 was used for RNA-seq read alignment to the reference genome. The alignment scheme of HISAT2 allows gaps to give speed, sensitivity, and accuracy across the alignment. HISAT2 is a universally used tool for aligning reads to a reference genome at a faster rate than STAR with similar accuracy (Kim, Langmead and Salzberg, 2015). HISAT2 is the core of the next version of Tophat. This tool also detects novel splice variants. The latest version of HISAT2 aligns with genotype variants, achieving higher accuracy. Reduces the cost than STAR (<8 GB for mapping to the human genome using default settings). The output files obtained from the tool for each dataset were aligned files in BAM format (Binary version of SAM format), and text files having the alignment summary. After alignment, the next step is to get a count matrix containing the count of each gene. For this reason, the featureCounts tool is used (Stephani, 2011). The following command was used to get alignment files in bam format.

Sample command:

hisat2 -x alignment/Index/grch38/genome -p 32 --dta --summary-file ADF81SRR545.txt -1 SRR5305545_1.fastq.gz -2 SRR5305545_2.fastq.gz | samtools view -bs -> ADF81SRR545.bam

3.2.3 Quantification

Feature Counts tool is the same as HTseq but much faster. The output is a little bit changed due to varying expression assignment approaches (Kim, Langmead and Salzberg, 2015). This tool takes SAM or BAM files as input and counts reads for each genomic feature. For this purpose, it requires a human reference genome annotation file along with SAM files. The output is a simple tab-delimited text file containing counts of all genes and featureCounts summary files (Stephani, 2011). The following command was used to get count files for one of the AD datasets.

Sample command:

featureCounts -T 32 -p -t exon -g gene_name -s 1 -a Homo_sapiens.GRCh38.84.gtf -o coungenets.txt ADF81SRR545.sorted.bam ADF85SRR515.sorted.bam ADF85SRR527.sorted.bam ADF86SRR531.sorted.bam ADF88SRR512.sorted.bam ADF89SRR541.sorted.bam ADF91SRR537.sorted.bam ADF91SRR538.sorted.bam ADF91SRR566.sorted.bam ADF92SRR517.sorted.bam ADM82SRR567.sorted.bam ADM85SRR562.sorted.bam ADM87SRR581.sorted.bam ADM89SRR568.sorted.bam ADF91SRR540.sorted.bam ADM77SRR580.sorted.bam ADM82SRR557.sorted.bam ADF82SRR586.sorted.bam ConF91SRR513.sorted.bam ConF78SRR542.sorted.bam ConF87SRR525.sorted.bam ConF90SRR575.sorted.bam ConF97SRR576.sorted.bam ConM80SRR556.sorted.bam ConM81SRR582.sorted.bam ConM81SRR593.sorted.bam ConM86SRR530.sorted.bam ConM87SRR554.sorted.bam ConM93SRR543.sorted.bam

3.2.4 Differential Expression Analysis

FeatureCounts outputs a tabular file containing gene counts per sample. This file can be used to check differential expression between two conditions. Differential expression analysis was performed on all three cancer types and AD. DESeq2 is an R Bioconductor implemented method used to detect differentially expressed genes. It uses the Wald test method to test for differential expression.

Null and alternative hypotheses are:

Ho: There is no differential expression of mRNAs

Ha: There is a differential expression of mRNAs

3.2.4.1 DESeq2

Deseq2 is one of the most used R Bioconductor packages [52]. It takes un-normalized count matrices as input to find differentially expressed genes under different conditions [53]. Like edgeR, it is also based on the negative binomial method and applies Fisher exact test to find DE genes. The normalization method used in DESeq2 is different from edgeR. It normalizes the data using the median of ratios of all gene counts in each sample over the geometric mean. Differential expression analysis by DESeq2 is done in these steps.

- Modelling raw counts for each gene:
    - Estimate size factors (accounts for differences in library size).
    - Estimate dispersions.
    - GLM (Generalized Linear Model) fit for each gene.
- Testing for differential expression (Wald test).

Typically, the output of DESeq2 contains 7 columns and rows corresponding to genes in the experiment. The columns which are used to extract differentially expressed genes are usually the pvalue, padj and log2FoldChange. The DESeq2 R code used in this study comes with other functionalities such as annotating gene information, normalization, transformation, and visualizing data in different plots. Biomart was used to convert gene IDs into gene symbols. Transcripts and genes were selected with the significance level <0.05 and a fold change value of 0.5 for both up and down-regulation. DESeq2 R code outputs a differential expression file, a normalized count file and plots for graphical visualization of data.

3.2.5 Meta Differential Expression Analysis

Meta-analysis for studying gene expression is a common technique, previously used for microarray studies. However, in most cases, the methods used for microarray expression analysis are not applicable for RNA-seq. Although, the importance of integrating multiple RNA-seq studies is much more crucial given the lower number of replicates in HTS experiments. In this study, to mitigate the effect of cross-sample variation in similar experiments, a meta differential expression analysis was performed. To perform this analysis an R package metaRNASeq available at CRAN (http://cran.r-project.org/web/packages/metaRNASeq), was used [54].

MetaRNASeq implies two p-value combination methods, fisher and inverse normal. However, for biological relevance, the p-value combination methods are built on the fact that both experimental results are obtained by using the same test-statistic method. In this case, DESeq2 was applied which implements Welch's test and was true for all studies. The p-values obtained from the DESeq2 method were subjected to metaRNASeq for meta differential expression analysis.

The code used for metaRNASeq is attached in supplementary. The final output of the tool is up-regulated and down-regulated genes in a disease that passed the p-value criteria for meta-analysis tests.

3.2.6 Genes with Inverse Expression

After differential expression analysis has been performed, the next step is to find out genes that are being regulated in opposite direction. For this purpose, FC (fold change) was used as criteria to define up and down-regulation. All the common differentially expressed genes were collected by using the Excel **VLOOKUP** function, then the same function was applied by using FC>0.5 as upregulated and FC<-0.5 as downregulated genes to get inversely expressed genes between both phenotypes.

3.2.7 Fisher-exact test for significance of the overlap

To account for true biological overlap between cancer and AD, the fisher-exact test was applied to calculate the significance of the overlap between AD and cancer. The significance was compared between cancers, AD and cancers, and self-overlap for each disease. R package GeneOverlap was used [55]. It tests two gene sets and calculates the significance of their intersection given a random number of background genes. The code is attached in supplementary.

3.2.8 Ranking based on disease association

The inverse regulated genes list was then subject to disease association analysis. GeneCard suit is an online web application platform that provides several tools to interpret genetic information without the need of installing and using many other expert bioinformatics tools. To extract a ranked list of genes for further downstream analysis, results obtained from metaRNASeq was submitted to the GeneCards VarElect tool [56]. The VarElect tool gives insight into extensive disease-gene relationship data. It gives both direct and indirect relationships of a gene to a given phenotype. In

our case, cancer and AD were selected as phenotypes of interest and inverse analysis gene list was given as input.

3.2.9 Gene Set Enrichment Analysis

After getting inversely expressed genes, the next step is the gene set enrichment analysis, abbreviated as GSEA. GSEA is a statistical method that describes the statistical significance of the genes to specific Gene Ontology (GO), KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways Reactome, Biocarta and many other pathway analyses. GSEA identifies the set of enriched genes in a particular dataset compared to control. Enrichment analysis is performed using the clusterprofiler R package.

GSEA computes enrichment analysis based on Enrichment score (ES) and Nominal P-value. ES determines which gene set is over-represented in the top and bottom of the ranked correlated gene list. Positive enrichment score (ES) shows gene set enriched at the top of the list, whereas negative ES shows gene set enriched at the bottom. The nominal P-value evaluates the statistical significance of ES, showing the likelihood that this gene set is enriched in a pathway.

A list of ranked genes based on disease association obtained from GeneCards was subjected to GSEA. Gene ontologies (GO) and KEGG pathways databases were used to obtain functional categories of genes. R package clusterprofiler and fgsea were used to get GSEA results.

Chapter 4

RESULTS AND DISCUSSION

In this chapter, results from all major steps of this study will be reported. A thorough discussion on these results will be conducted to answer the research questions formulated during this study.

4.1 Differential Expression

The first objective of this study was to identify differentially expressed genes in individual diseases. To perform this analysis a detailed methodology has been described in Chapter 3. Before performing the differential expression analysis, pre-processing, and mapping of individual samples were performed. Mapping results are attached in the Appendix.

4.1.1 Alzheimer's Disease

Dataset search was a crucial step for this study. After extensive literature and repositories (ArrayExpress and GEO) search, only two datasets of AD matched our pre-defined criteria. Results of RNA-Seq analysis performed on Alzheimer's datasets are described below.

- **GSE53697**

This dataset downloaded from ENA (PRJNA232669) is a part of the super series GSE53699. Tissue samples from the frozen human brain were obtained from Mount Sinai Brain Bank. According to data description from the primary publisher, brains was dissected along with Brodmann areas. A total of 9 samples were extracted from the dorsolateral prefrontal cortex from AD affected advanced stage brains (between 4 and 5). Eight age and gender-matched controls were selected. RNA was extracted from all samples and prepared for sequencing, following the Illumina high-throughput sequencing guidelines. A detailed description of samples is attached in Appendix.

The first step was to perform a quality check using FastQC. All samples passed the quality filters and were subjected to mapping using HISAT2. The alignment rates of each sample are attached in Appendix. Next, featureCount was used to quantify expression using the FPKM approach. The count files for all samples were next used to perform differential expression analysis.

R package DESeq2 was used to perform differential expression analysis. ("GEO Accession viewer") Different plots were generated to visualize data and results. **Figure** 4-1 shows the PCA plot which is a dimension reduction technique used to cluster data based on phenotypes. Normal and diseased samples formed distinct clusters, explaining variation in data based on condition. This can also be observed in the case of the heatmap shown in **Figure**. The heatmap represents sample to sample distance. The darker colour represents a smaller distance which means that the samples are closely related. The lighter the colour the more distant are the samples. As it can be seen in **Figure** 4-1, diseased samples are clustered together and are distant from the control group cluster.

Figure 4-1 shows a volcano plot representing differentially expressed genes. The threshold applied for significance can also be seen. The x-axis represents log2FoldChange which is a measure for the difference of expression between two phenotypes for a given gene.

$$\text{FoldChange} = \text{Counts(Treatment)} / \text{Counts(Control)}$$

Here, a FoldChange equal to 1 means no difference, and a FoldChange value greater or lesser than 1 means that gene is up or downregulated, respectively. The minimum threshold for a gene to be considered differentially expressed is if it has at least 50% change. Therefore, FoldChange value 1.5 is considered as an appropriate threshold and the log2FoldChange value corresponding to it is 0.5. The Y-axis represents the statistical significance of calculated FoldChange values across samples. The hypothesis to be tested in this case is.

H_0 Genes are not differentially expressed > G1(Treatment/ G1 (control) = 1

H_A Genes are differentially expressed > G1(Treatment/ G1 (control) ≠ 1

Deseq2 uses a Wald-test to calculate test statistics and p-value. P-value threshold 0.05 is used which represents a false positive rate of 5%. All the genes which passed both thresholds are represented in coloured dots.

A total of 1032 genes passed both criteria and were found to be differentially expressed in **GSE53697.**

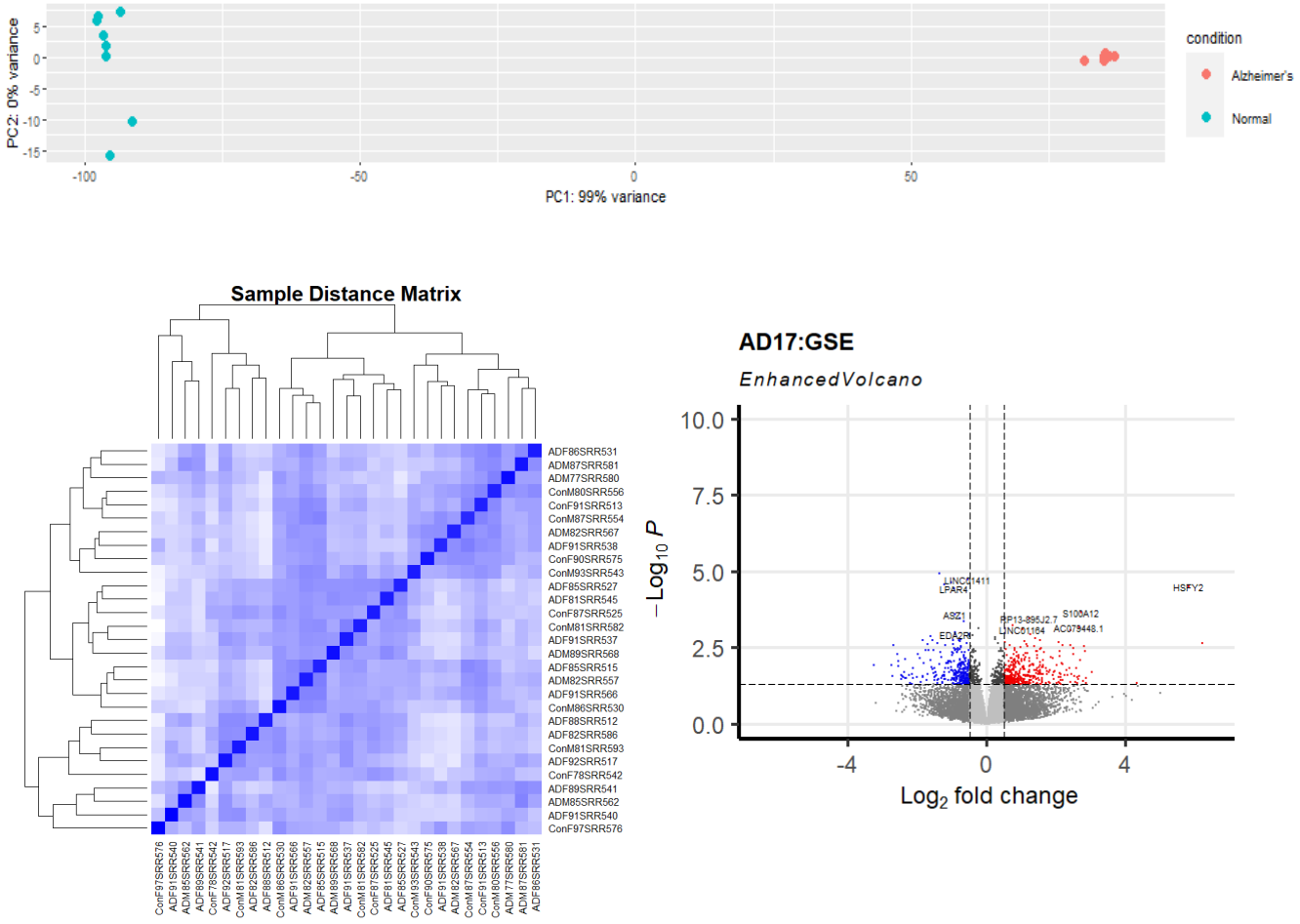


Figure 4-1 PCA plot representing sample in clusters. B) Heatmap showing samples correlation and distances. c) Volcano plot showing differentially expressed genes. x-axis on volcano plot represents Log2FoldChange, y-axis represents p values. The dots in volcano plots represents genes, dotted line represents cut-offs.

- **GSE95587**

Another RNA-seq dataset was downloaded from ENA. This dataset contains a total of 117 samples (33 control, 84 AD) extracted from fusiform gyrus tissue. The AD progression stages in this dataset are ranged between 2 to 6. For minimizing confounding factors between 2 datasets, only samples belonging to stage 4 were selected. This filter returned a total of 30 samples (11 Normal, 19 AD).

All samples were pre-processed, and quality checked. One sample (accession) failed quality criteria and was excluded from this study. Next, alignment to human reference was performed. The alignment rates are provided in supplementary.

Figure 4-2 shows the PCA plot for this dataset. PCA explains percentage variability in the data based on selected phenotype. Samples were not clustered distinctively based on phenotype, which means there is not much variability based on disease and control conditions. However, PC1 was able to explain 30% of the variability. The volcano plot shows differentially expressed genes. A total of 5647 genes were found to be significantly differentially expressed.

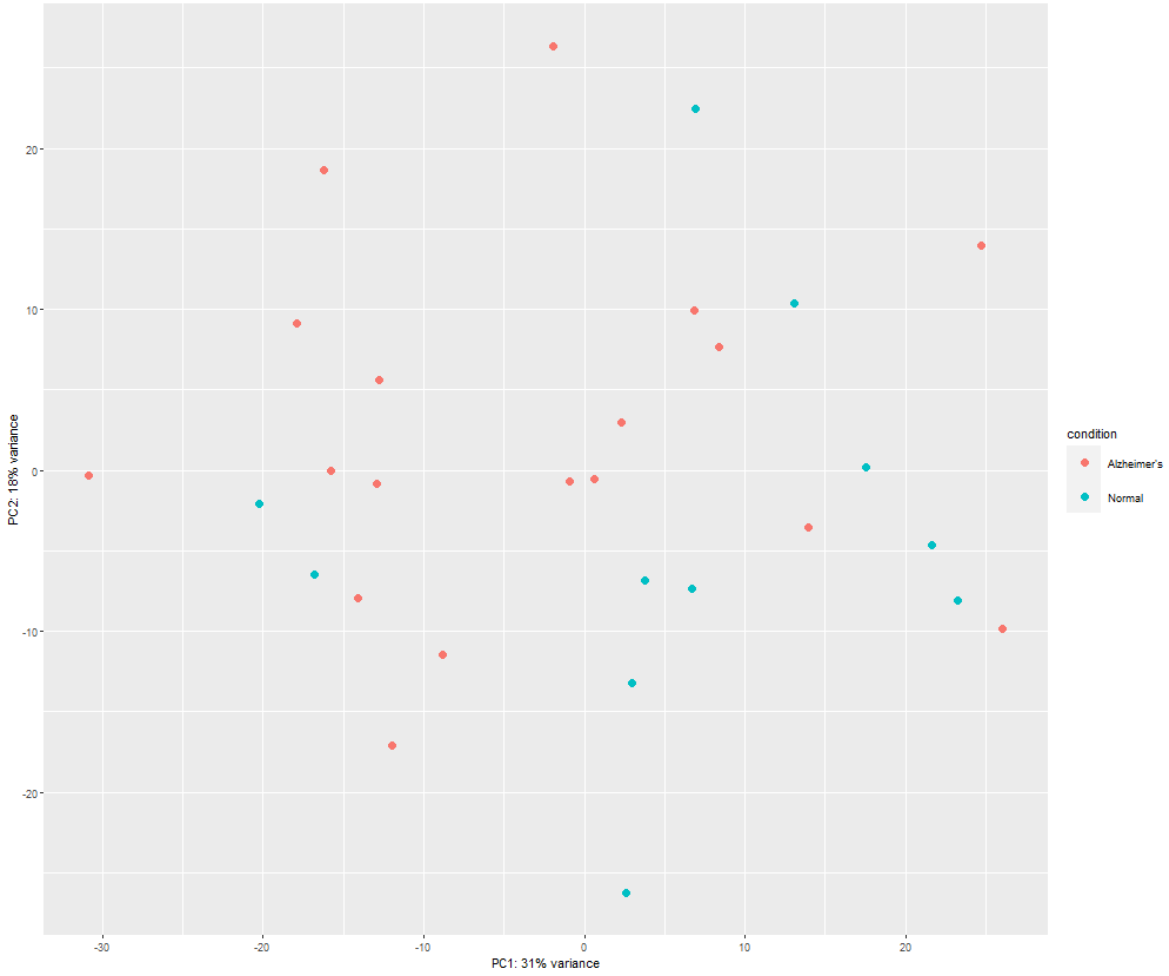


Figure 4-2 PCA plot showing distinct clusters based on disease and healthy phenotypes. The blue colour represents a normal sample, and the red colour represents cancer samples. PC1 and PC2 are plotted to give a two-dimensional representation of multi-dimensional data.

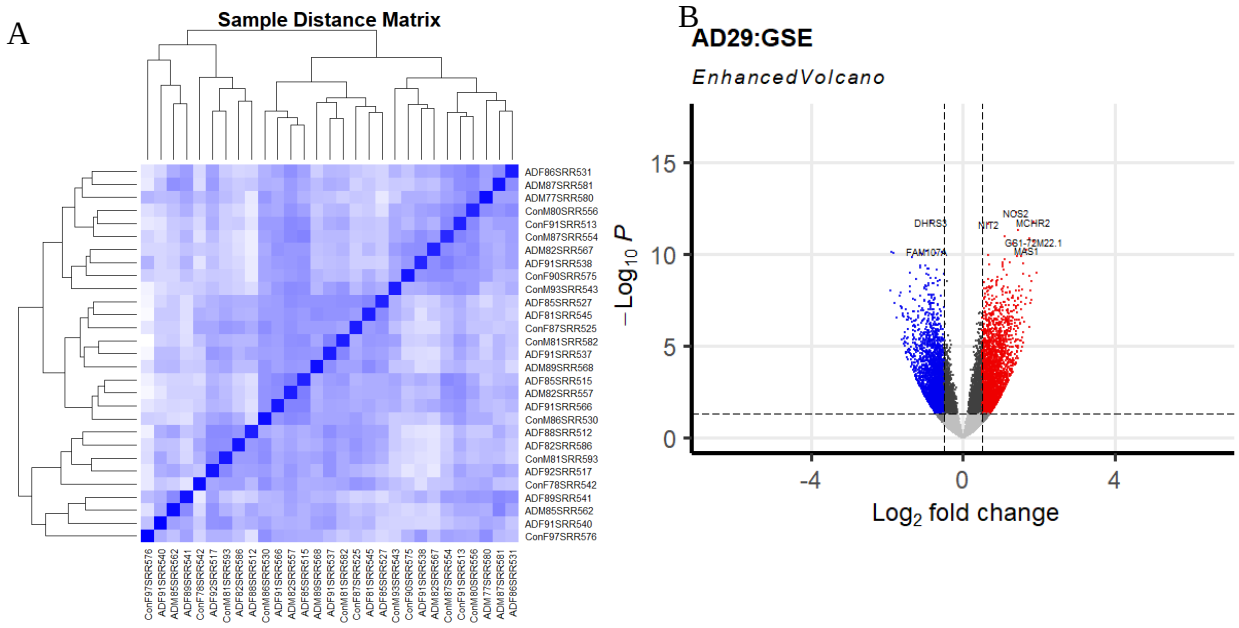


Figure 4-3 A) Heatmap representing sample distances. Most samples were clustered based on normal and cancer phenotypes. In sample names C represents cancer and N represents Normal samples. B) Volcano plot showing differentially expression. Red dots represent up-regulated genes, blue dots represent down-regulated genes.

4.1.2 Liver Cancer (HCC)

While reviewing cancer and AD association, several epidemiological studies and biological evidence suggested liver cancer or hepatocellular carcinoma (HCC) as most related to AD. According to a review, liver cancer is associated with a 51% lower risk of AD. Considering this evidence liver cancer was selected for this study [57].

- **GSE105103**

This dataset was downloaded from ENA. Total RNA-Seq was performed on tissue samples extracted from 25 HCC tumour tissues and 25 adjacent normal tissues. Most of these patients were HCC stages 1, 2 and 3. Clinical data can be viewed here.

Quality check and alignment was performed. All samples produced >90% alignment rate and were further subjected to quantification. Quantification produced a total of 9163 genes as expressed, out of which 1398 genes were found to be significantly differentially expressed.

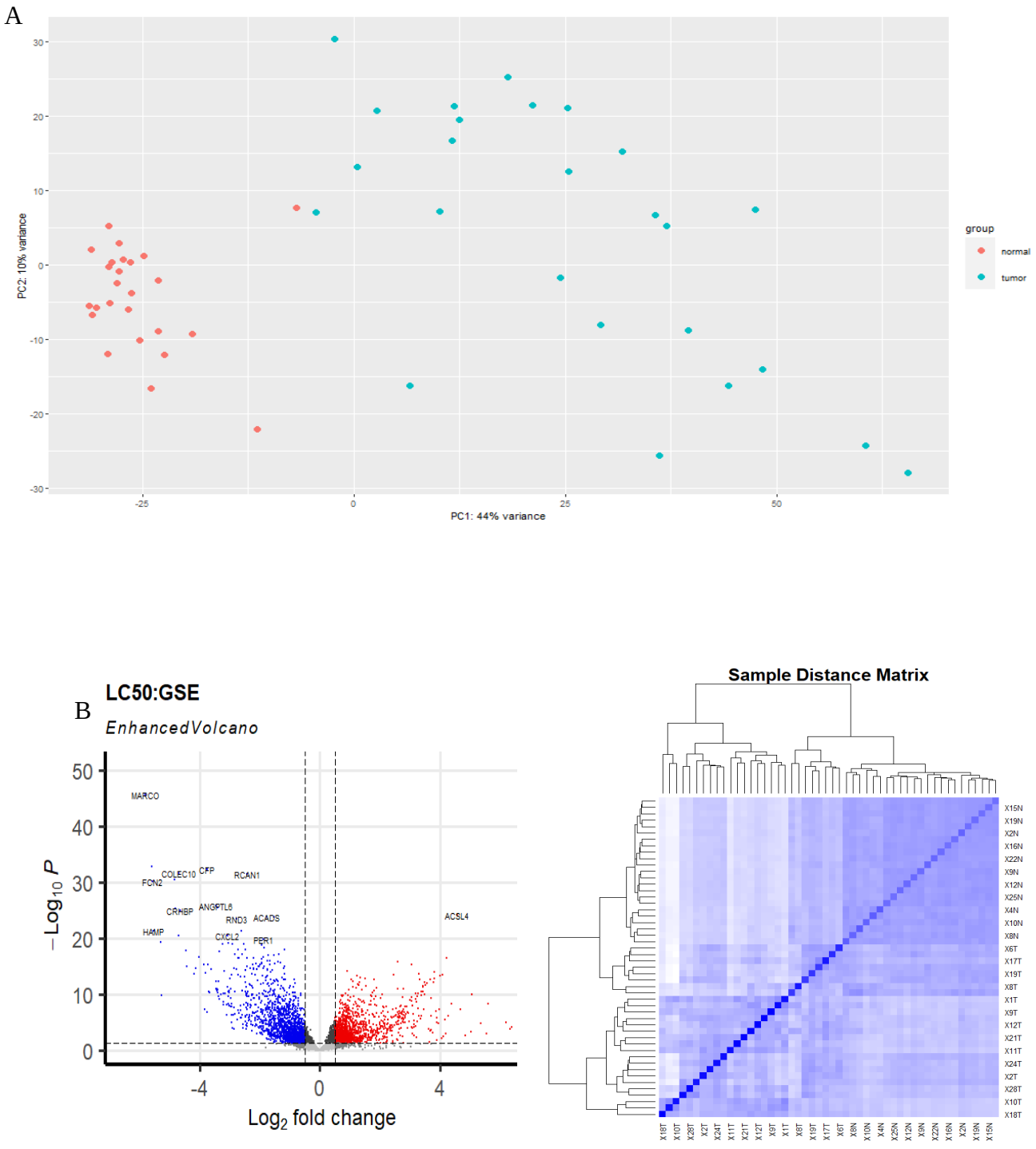


Figure 4-4 PCA plot showing distinct clusters based on disease and healthy phenotypes. The blue colour represents a normal sample, and the red colour represents cancer samples. PC1 and PC2 are plotted to give a two-dimensional representation of multi-dimensional data. B) Heatmap representing sample distances. Most samples were clustered based on normal and cancer phenotypes. In sample names even numbers represent cancer and odd numbers represent Normal samples. C) Volcano plot showing differential expression. Red dots represent up-regulated genes, blue dots represent down-regulated genes.

- **GSE97214**

Another dataset of HCC was analysed in this study. Total RNA-seq was performed on 9 tumour tissues and 9 adjacent normals. Sequencing was performed on Illumina HiSeq 2500 sequencer with 125 bp paired end reads. The tumour samples were extracted based on differentiation stages as high, moderate, and low.

Raw sequencing read was quality filtered and adapters were removed. Alignment was performed according to the methods described in chapter 3. All samples produced more than a 90% alignment rate. PCA plot in **Figure** 4-5 shows poorly formed clusters based on phenotypes. Although PC1 was able to explain 38% variability. Heatmap in **Figure** 4-6 shows most diseased samplers clustered together. A total of 963 genes were found to be differentially expressed. The volcano plot in **Figure** 4-6 represents DEGs that passed both p-value and log2FoldChange cut-offs.

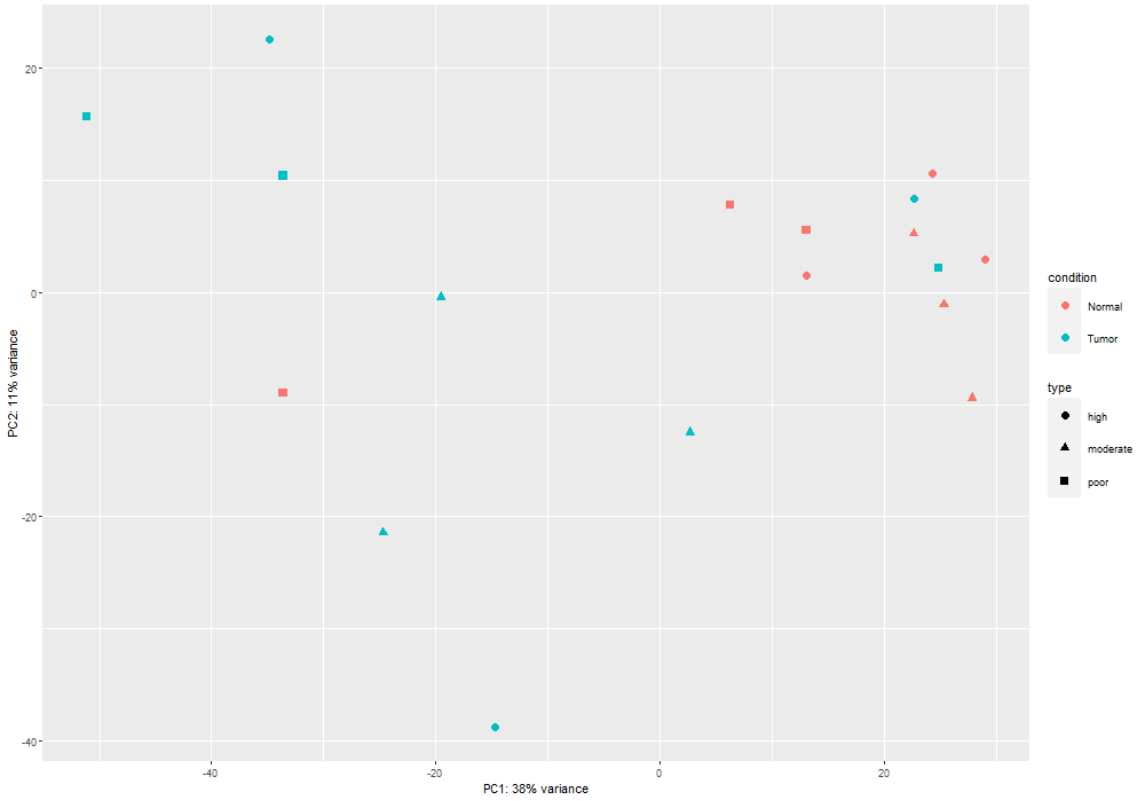


Figure 4-5 PCA plot showing distinct clusters based on disease and healthy phenotypes. The blue colour represents a normal sample, and the red colour represents cancer samples. PC1 and PC2 are plotted to give a two-dimensional representation of multi-dimensional data.

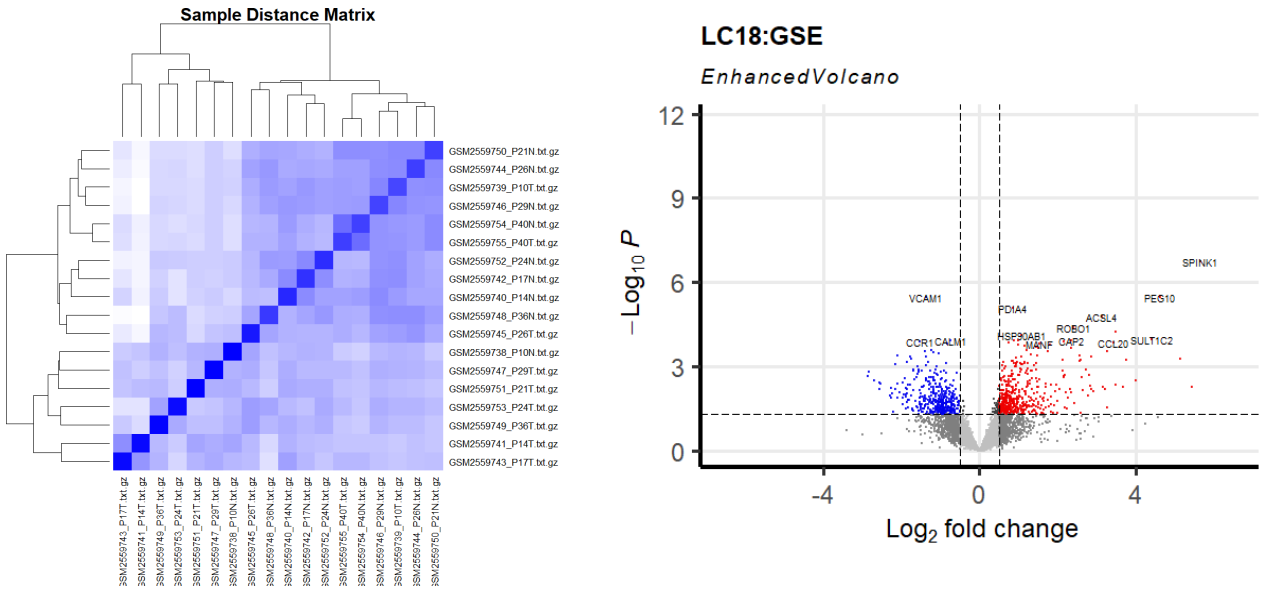


Figure 4-6 A) Heatmap representing sample distances. Most samples were clustered based on normal and cancer phenotypes. In sample names C represents cancer and N represents Normal samples. B) Volcano plot showing differentially expression. Red dots represent up-regulated genes, blue dots represent down-regulated genes.

4.1.3 Oesophageal Cancer (EC)

While reviewing the cancer and AD association, several epidemiological studies and biological evidence suggested Oesophageal cancer (EC) as one of the most related cancers to AD. According to a review, EC is associated with a 33% lower risk of AD. Considering this evidence EC was selected for this study [58].

- **GSE130078**

This dataset is part of an experimental series “HERES, a lncRNA that regulates canonical and noncanonical Wnt signalling pathways via interaction with EZH2” [59]. High throughput RNA sequencing (RNA-seq) was performed on 23 Korean EC patients with matched normals.

The raw RNA-seq data in fastq format were subjected to quality filtration, adapters were removed and all samples passing the quality control parameters were subjected to alignment. All 46 samples were mapped to GRCH38 and resulted in a more than 90% alignment rate. After quantifying expression from mapped files with featureCounts, differential expression was performed. **Figure**

4-7 shows the PCA plot. Based on disease and healthy phenotypes samples were able to form distinct clusters with PC1 explaining 54% variability in data. A heatmap is also shown in **Figure** 4-8 representing samples forming clusters with a similar phenotype.

A total of 6839 genes were found to be differentially expressed, out of those 3992 were down-regulated and 2847 were up-regulated. **Figure** 4-8 shows a volcano plot of differentially expressed genes. P-value cut-off of 0.05 and log2foldchange cut-off of 0.5 was applied. All genes passing these criteria were labelled as differentially expressed.

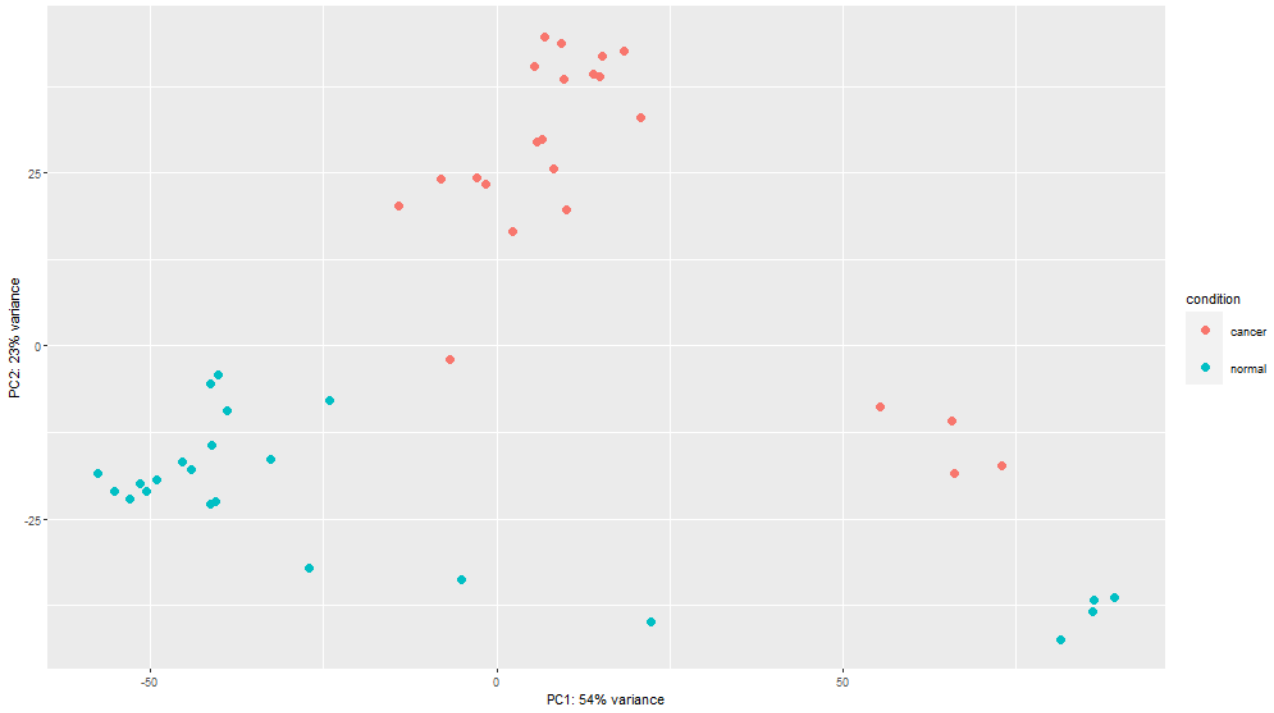


Figure 4-7 PCA plot of GSE130078 showing distinct clusters based on disease and healthy phenotypes. The blue colour represents a normal sample, and the red colour represents cancer samples. PC1 and PC2 are plotted to give a two-dimensional representation of multi-dimensional data.

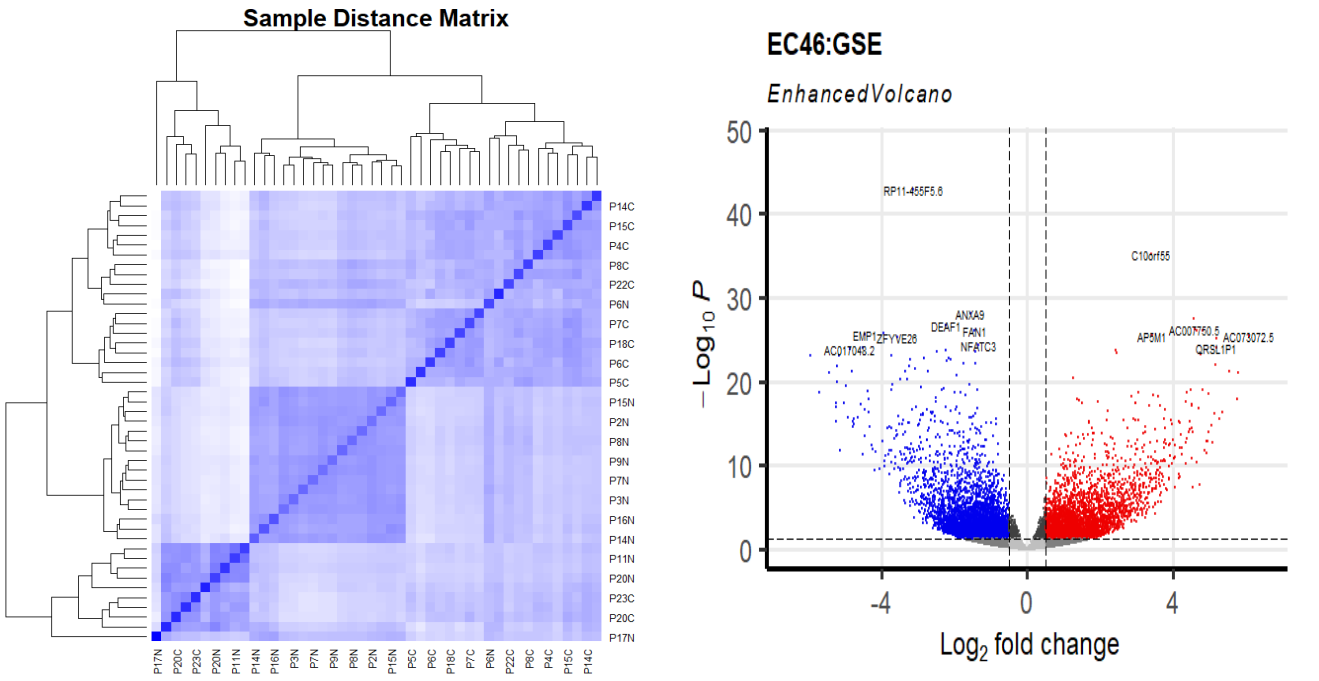


Figure 4-8 A) Heatmap representing sample distances. Most samples were clustered based on normal and cancer phenotypes. In sample names C represents cancer and N represents Normal samples. B) Volcano plot showing differentially expression. Red dots represent up-regulated genes, blue dots represent down-regulated genes.

- **GSE111011**

Another dataset was obtained from NCBI GEO containing 14 samples (7 tumours and 7 matched normal). The RNA-Seq dataset was obtained with Illumina HiSeq 2500. Quality control and filtration were applied, adapters, low-quality reads and ambiguous reads were removed. Then the reads were mapped and aligned to the human reference genome (GRCH38).

After quantification with featureCounts, 58828 features were extracted which were then subjected to differential expression using DESEeq2. A total of 7160 genes were found to be differentially expressed, out of these 3998 were down-regulated and 3162 were up-regulated. **Figure** 4-9 shows a PCA plot representing samples forming distinct clusters based on cancer and normal phenotypes. Two EC samples formed a separate cluster distant from both healthy and cancer clusters. Nonetheless, PC1 was able to explain 94% variability within these clusters. There is also a heatmap shown in **Figure** 4-9, odd numbers in sample names represents normal and even numbers represent cancer samples. It is evident from the heatmap that samples were clustered together based on phenotypes and we expect many genes affected. The volcano plot in **Figure** 4-9 shows differential

expression. P-value cut-off of 0.05 and log2foldchange cut-off of 0.5 was applied. All genes passing these criteria were labelled as differentially expressed.

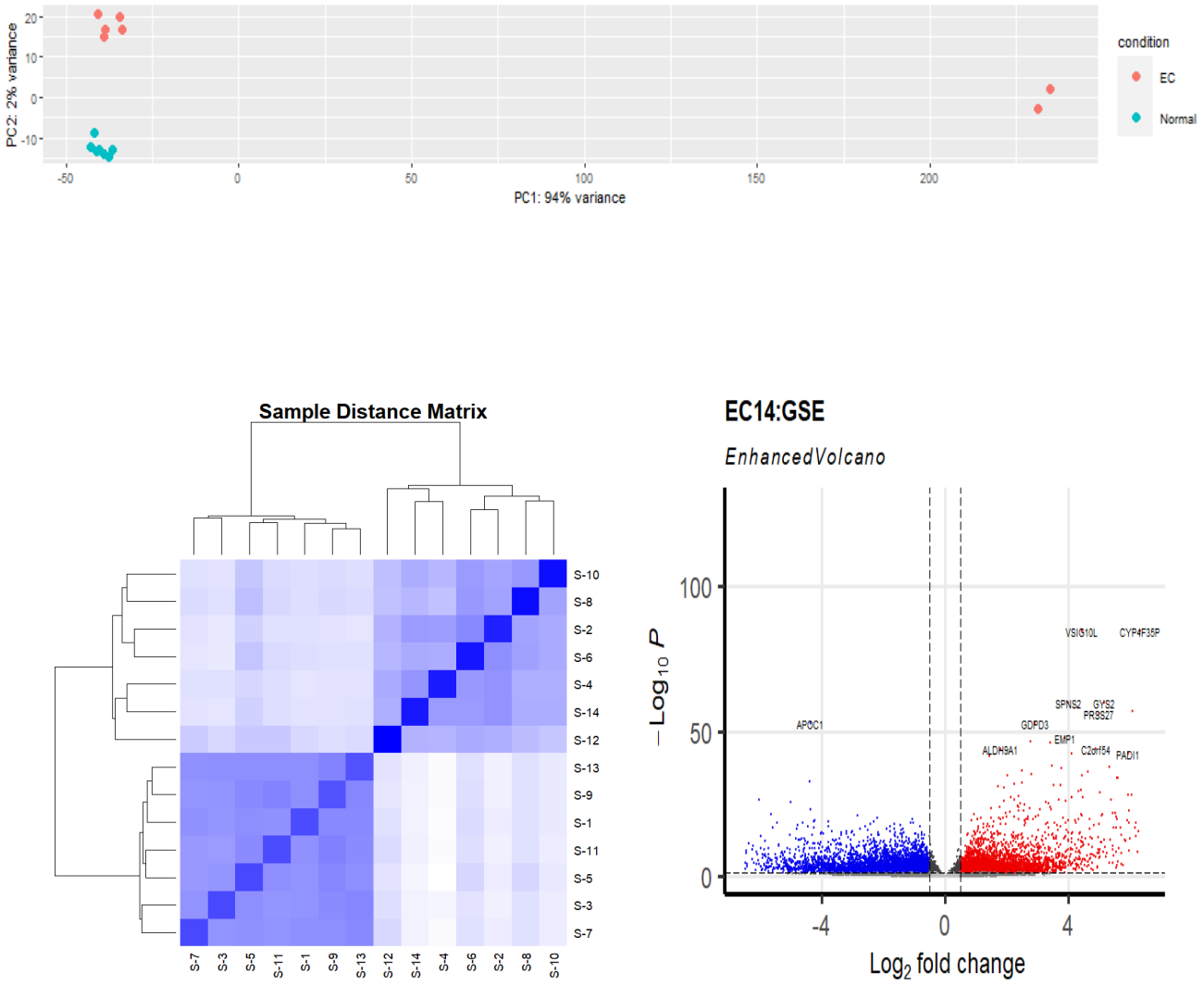


Figure 4-9 A) PCA plot showing distinct clusters based on disease and healthy phenotypes. Blue colour represents normal sample and red colour represents cancer samples. PC1 and PC2 are plotted to give a two-dimensional representation of multi-dimensional data. B) Heatmap representing sample distances. Most samples were clustered based on normal and cancer phenotypes. In sample names even numbers represent cancer and odd numbers represent Normal samples. C) Volcano plot showing differentially expression. Red dots represent up-regulated genes; blue dots represent down-regulated genes.

4.2 Meta-Analysis

To compare results generated from differential expression analysis, results for each disease were subjected to meta-analysis. Different p-value combination methods were applied for meta differential expression analysis. Both p-value combinations were able to detect a greater number of genes than the conservative approach of the global comparison method. In addition, many genes were found to be common between Fisher p-value and invnorm methods. The Venn diagrams presented in **Figure** 4-10 compare the list of differentially expressed genes resulting from the different methods for all 3 diseases. More results for the individual disease are described below.

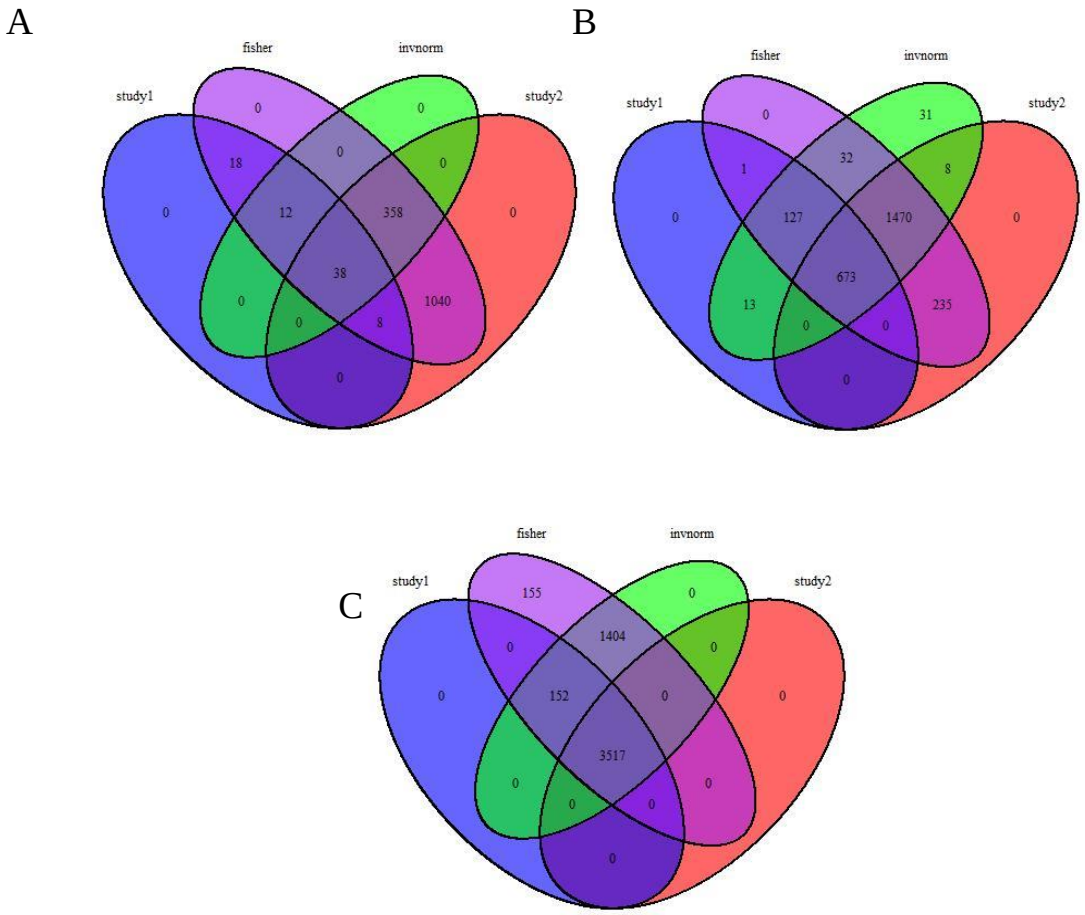


Figure 4-10 Comparison of results from differential analyses of AD(A), HCC(B) and EC(C). Venn diagram presents the results of the differential analysis for the two meta-analysis methods (Fisher and inverse normal), the global analysis (DESeq (study)), and the intersection of individual per-study analyses (Individual). **Figure** made using the VennDiagram package.

4.2.1 AD differential gene expression meta-analysis

A total of 1032 and 5647 genes were found to be differentially expressed in GSE53697 and GSE95587, respectively. These DEGs were subjected to metanalysis which resulted in 1474 DEGs. Out of these, 1039 genes were up-regulated and 435 were down-regulated. These observations were made on genes passing both p-value combination methods to avoid any errors. Full AD differential expression and meta-analysis results are available in supplementary.

4.2.2 Cancer differential gene expression meta-analysis

A high number of genes were found to be differentially expressed in cancer datasets. A total of 1398 and 963 genes were found to be differentially expressed in HCC datasets (GSE105103, GSE97214 respectively). Meta-analysis resulted in a total of 2490 DEGs in HCC. Out of these, 1263 genes were up-regulated and 1327 were down-regulated. On the other hand, 7160 and 6839 genes were found to be differentially expressed in EC datasets (GSE111011, GSE130078). A total of 7676 genes were extracted as differentially expressed as the result of the meta-analysis. A substantial proportion of genes 6006 was up-regulated, whereas 1670 genes were found to be down-regulated. These observations were made on genes passing both p-value combination methods to avoid any errors. Full HCC and EC differential expression and meta-analysis results are available in supplementary.

4.3 Genes regulated in opposite directions between cancer and AD

To investigate whether differentially expressed genes in AD are deregulated in opposite directions in cancer, all possible AD and cancer pairs were subjected to the intersection as follow.

1: Genes upregulated in AD but downregulated in at least one cancer type

2: Gene downregulated in AD but upregulated in at least one cancer type

3: Genes deregulated between AD and all cancer types

Following results were observed.

1: 154 DEGs that were up-regulated in AD were found down-regulated in EC, and 23 DEGs that were down-regulated in AD were found to be up-regulated in EC.

2: In the case of HCC, 72 genes down-regulated in HCC were up-regulated in AD, and 29 genes up-regulated in HCC were down-regulated in AD.

3: Collectively, 242 genes were found deregulated in opposite direction between AD and both types of cancers.

Figure 4-11 shows the overall intersection between all possible AD and cancer pairs. There was a significant inverse overlap observed between AD and cancers. The overlap was more between AD-up and cancer-down. However, a significance test will be more appropriate to explain this overlap.

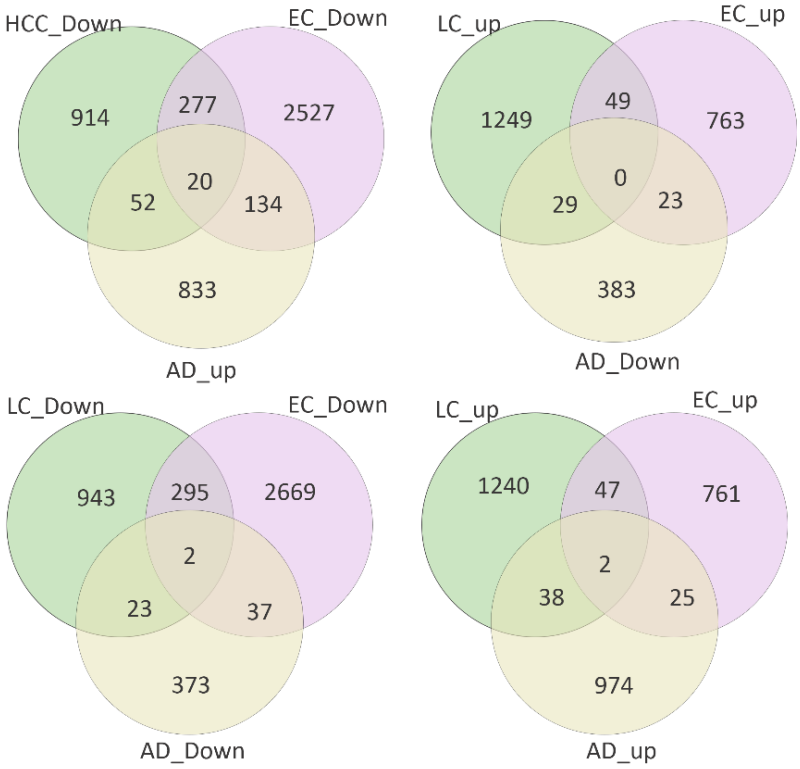


Figure 4-11 Venn Diagrams representing the number of genes common between AD, HCC, and EC up and down-regulation profiles. (a) and (b) represents the inverse overlap between cancers and AD. (c) and (d) represents the direct overlap between AD and cancers.

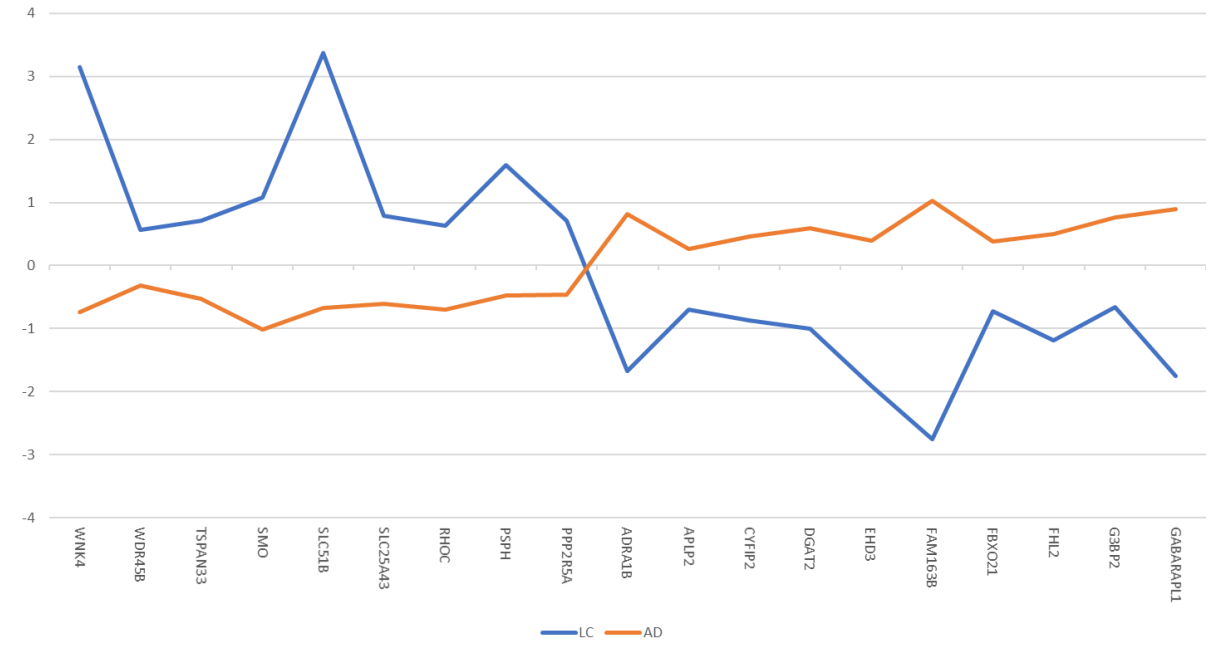


Figure 4-12 Line plot showing genes deregulated in opposite directions between HCC and AD.

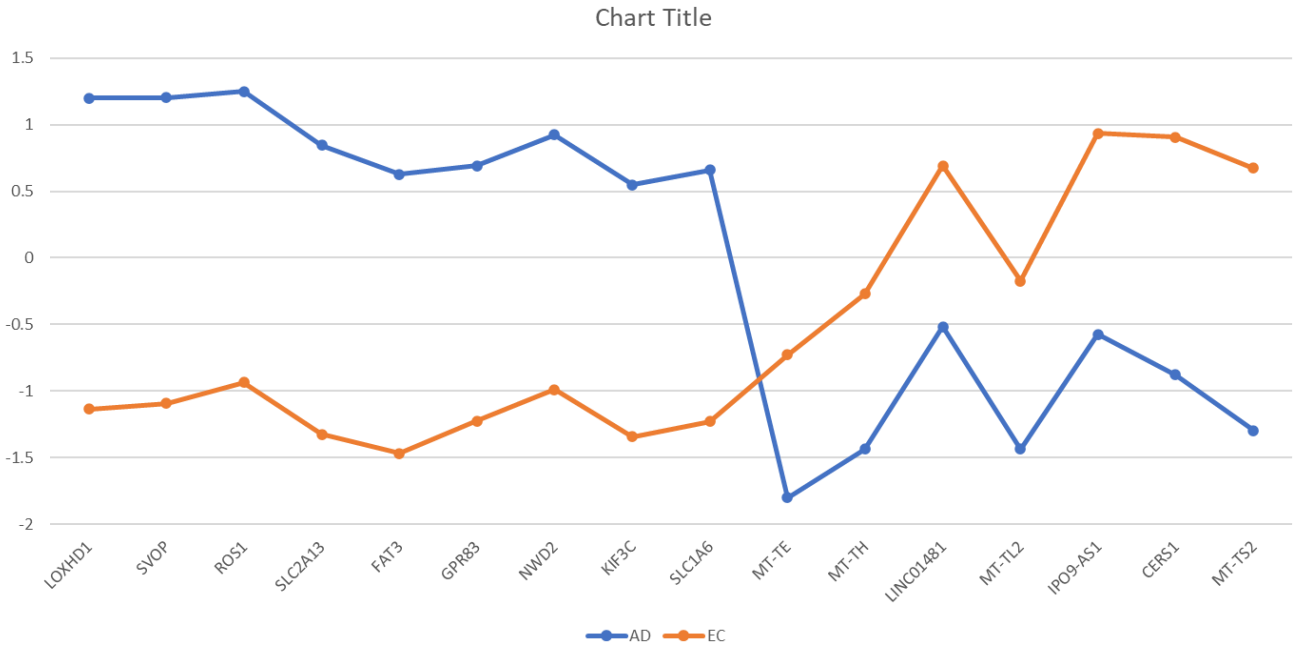


Figure 4-13 Line plot showing genes deregulated in opposite directions between EC and AD.

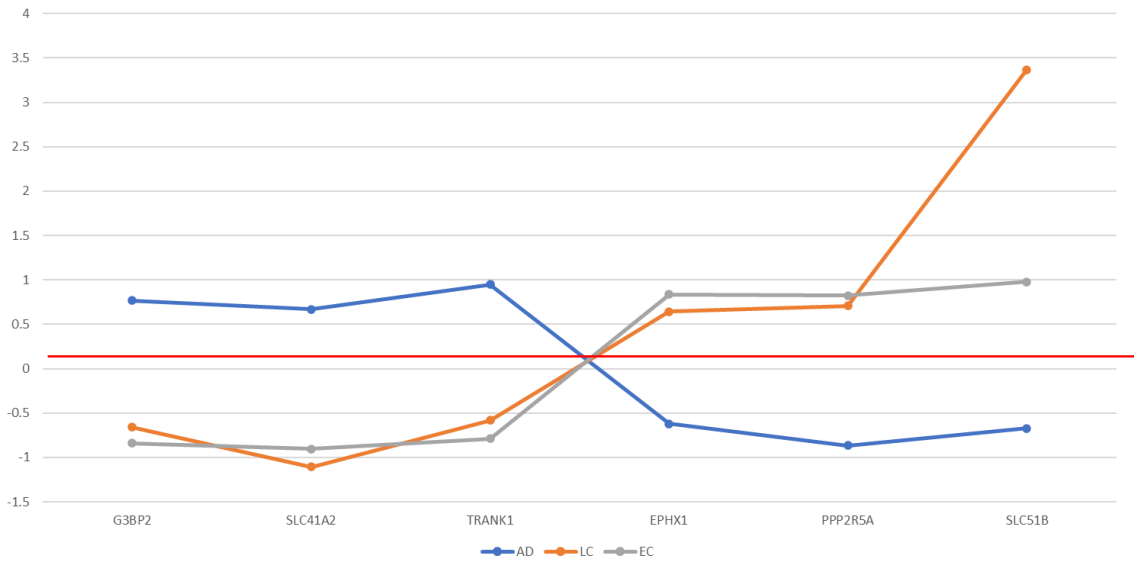


Figure 4-14 Line plot showing genes deregulated in opposite directions between cancers and AD.

4.3.1 Significance of Overlap

To determine the statistical significance of this overlap, a fisher t-test was applied using the "GeneOverlap" R package. The overlap was tested for each cancer type individually. P-value cut-off 0f 0.05 was considered for significance. Any association below the threshold was considered non-significant (N.S). There was a significant association between AD-UP & EC-DOWN, AD-UP & HCC-Down, AD-Down & EC-up, and AD-Down & HCC-up. No other significant association was observed. To further investigate the occurrence of a true inverse association, cancer types were compared to each other on the same parameters. There was a significant association between HCC-up and EC-up, EC-down, and HCC-down also showed significant association. However, the p-value was not significant enough to account for an inverse association between cancers.

Figure 4-15 represents the heatmap for the significance of overlap. The darker colour represents a more significant association, while the lighter colour represents lesser significance.

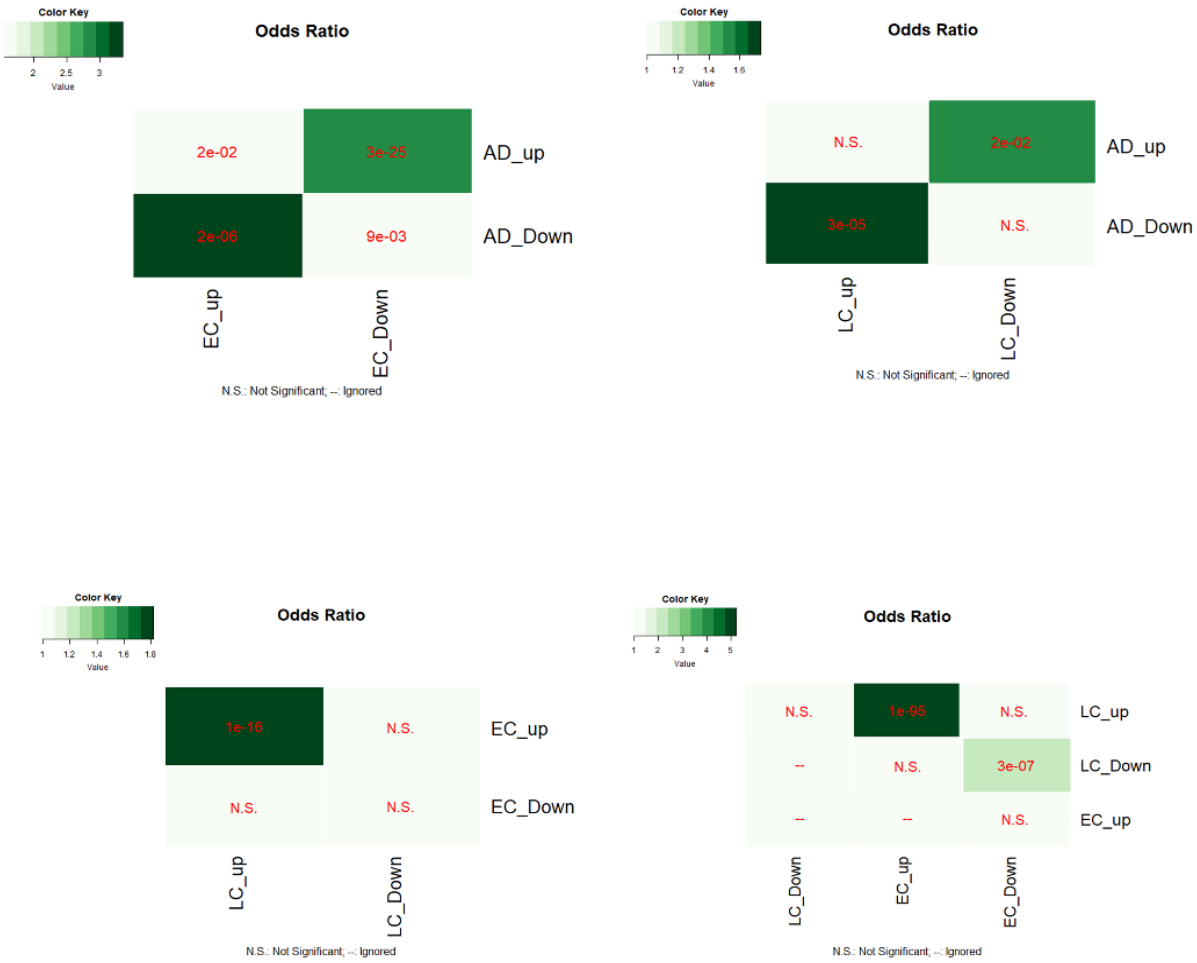


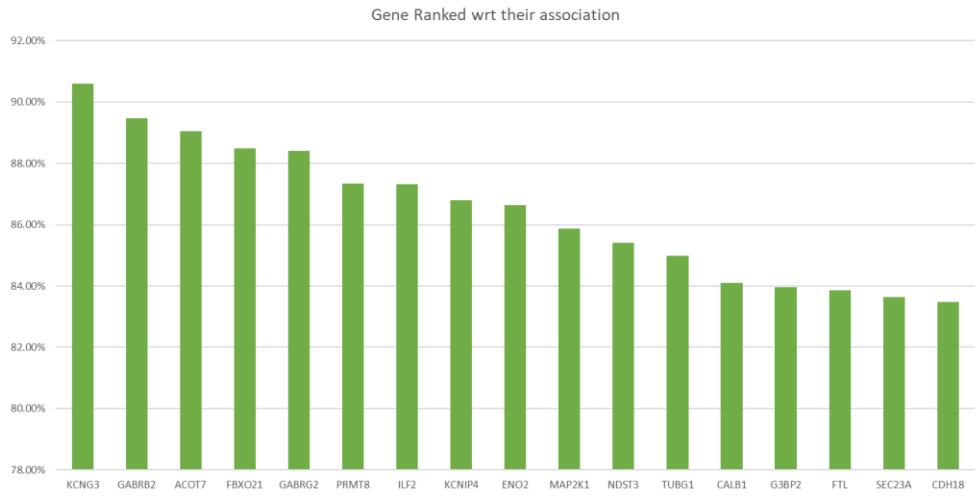
Figure 4-15 Heatmaps showing the significance of the overlap between all possiblepairspackage "GeneOverlap" was used for testing significance. A P-value of greater than 0.05 is labelled as N.S(Non-Significant). (A) Gene overlap test between AD and EC (B) Gene Overlap test between AD and HCC. (C) Gene Overlap test between HCC and ECC. (D) Cancers between and self overlap.

4.3.2 Ranking genes based on their association with disease

Inversely expressed genes were next subjected to functional analysis. First, genes were listed according to their disease association with the disease as reported by GeneCard suit. Two hundred and twenty-four genes were found to be directly related to either cancer or AD or both. Six genes were reported as indirectly associated with AD and cancer through 19 different genes. Genes were ranked according to their association score. The top 10 inversely regulated genes are given in the table.

Table 4-1 List of inversely expressed genes was given to GeneCards suit and disease association scores were obtained. The table represents genes associated with both phenotypes and their scores.

Symbol	Description	Matched Phenotypes	Matched Phenotypes count	Global Rank (Total Genes 26634)	-LOG10(P)	Score	Average Disease-Causing Likelihood
MSH2	MutS Homolog 2	Alzheimer, Cancer	2	9	3.45	118.04	65.16%
RAD51C	RAD51 Paralog C	Alzheimer, Cancer	2	34	2.88	75.66	64.31%
RET	Ret Proto-Oncogene	Alzheimer, Cancer	2	39	2.82	71.70	55.50%
MAP2K1	Mitogen-Activated Protein Kinase Kinase 1	Alzheimer, Cancer	2	78	2.51	51.68	85.87%
FH	Fumarate Hydratase	Alzheimer, Cancer	2	125	2.31	42.50	78.01%
SDHA	Succinate Dehydrogenase Complex Flavoprotein Subunit A	Alzheimer, Cancer	2	192	2.12	34.78	52.33%
ATR	ATR Serine/Threonine Kinase	Alzheimer, Cancer	2	362	1.85	25.21	62.28%
NOS2	Nitric Oxide Synthase 2	Alzheimer, Cancer	2	423	1.78	22.98	46.59%
ROS1	ROS Proto-Oncogene 1, Receptor Tyrosine Kinase	Cancer	1	228	2.05	22.66	22.31%
ENO2	Enolase 2	Alzheimer, Cancer	2	599	1.63	17.99	86.64%



4.4 Gene-Set Enrichment Analysis (GSEA)

4.4.1 Alzheimer's Disease

Gene set enrichment analysis suggested that genes upregulated in AD are mostly associated with Metabolic pathways, GABAergic synapse, Glutamatergic synapse, Neuroactive ligand-receptor interaction, synaptic signaling, cell regulation and Neuroactive ligand-receptor interaction. Nicotine addiction and Morphine addiction were also among the top up-regulated pathways. Genes downregulated in AD samples were majorly associated with negative regulation of cellular processes and epithelial migration, NFKB signaling and PI3K-Akt signaling pathway. Full AD GSEA enrichment results can be found in supplementary.

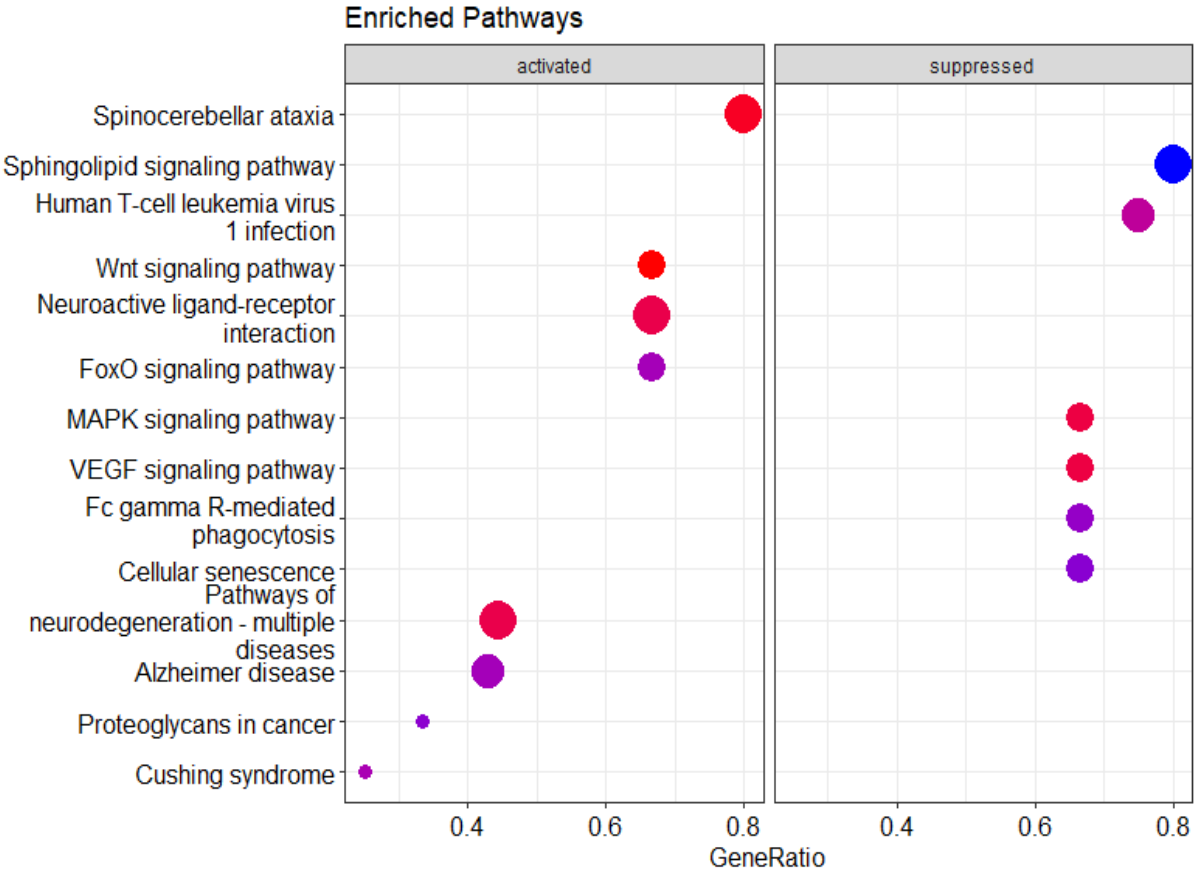


Figure 4-16 Top KEGG pathways found to be regulated in AD. A ranked gene list was given to GSEA. The size of dot is proportional to gene set size. Pathways with a positive NES score are activated and with a negative NES score are supressed.

4.4.2 Cancers

Enrichment analysis showed that processes related to DNA replication, cell cycle, and mitotic division were up regulated. Also, pathways like cytokine-cytokine receptor ligation, PI3K-Akt signalling pathway and TNF signalling pathway were up-regulated in both cancer types. On the other hand, KEGG pathways like Neuroactive ligand-receptor interaction, Bile secretion, Nucleocytoplasmic transport, Spinocerebellar ataxia were commonly down-regulated in cancers. Pathways downregulated in EC were mostly related to neurodegeneration. **Figure** 4-17 Top KEGG pathways found to be regulated in cancers. A ranked gene list was given to GSEA. The size of dot is proportional to gene set size. Pathways with a positive NES score are activated and with a negative NES score are supressed. represents dot plot of top pathways found to be regulated in both cancers.

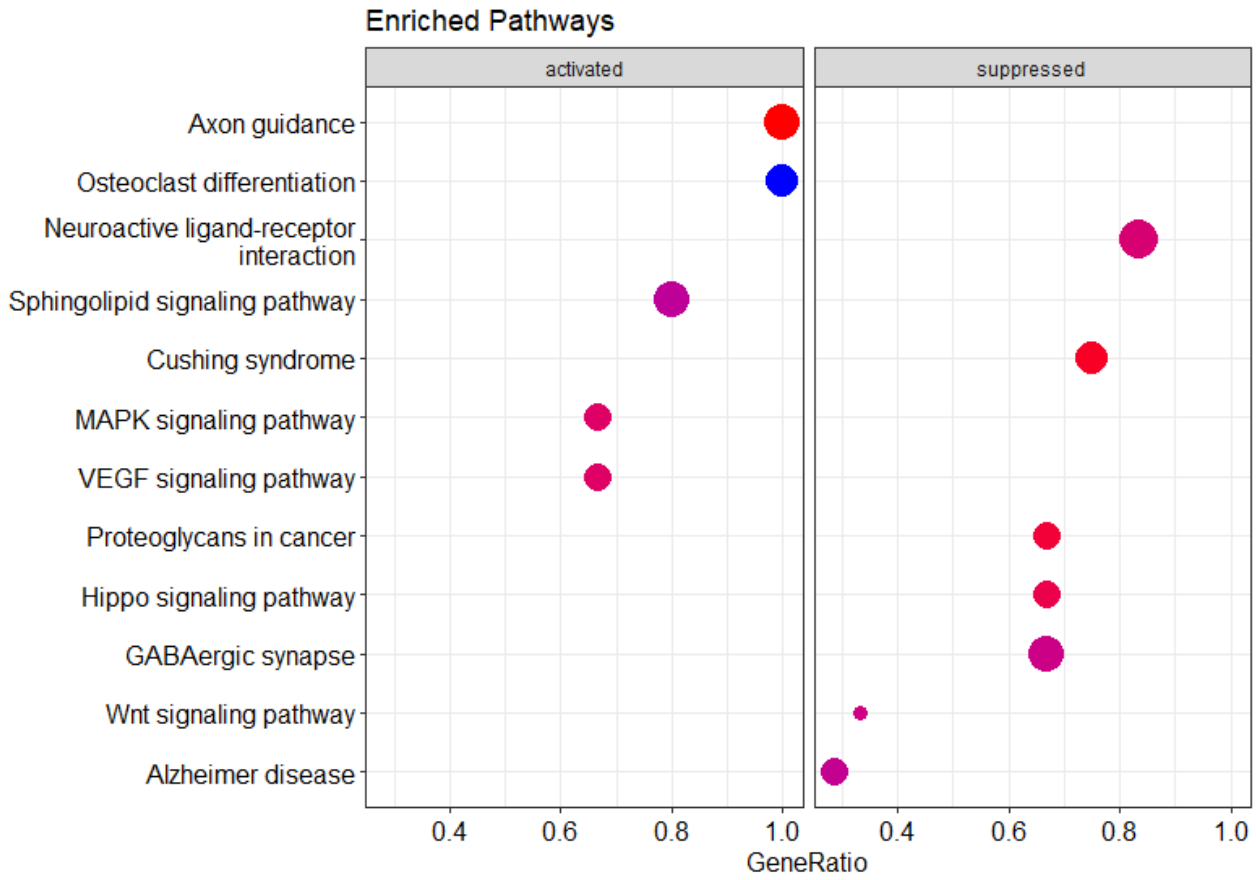


Figure 4-17 Top KEGG pathways found to be regulated in cancers. A ranked gene list was given to GSEA. The size of dot is proportional to gene set size. Pathways with a positive NES score are activated and with a negative NES score are supressed.

4.4.3 AD and Cancer Pathways comparison

To sketch the landscape of inversely deregulated pathways between AD and cancer, GSEA KEGG results from individual studies were compared. Metabolic pathways, Glycolysis and oxidative phosphorylation were found to be up regulated in both diseases. Protein production pathways such as ribosomes were found to be deregulated in opposite directions in both diseases. Most of the pathways upregulated in cancer were found to be downregulated in AD. Pathways associated with cancer like the TGF-beta signalling pathway, Hepatocellular carcinoma and PI3K-Akt signalling pathway were upregulated in cancer but downregulated in AD. On the other hand, Neuronal pathways like GABAergic synapse, Neuroactive ligand-receptor interaction, Retrograde endocannabinoid signalling, Glutamatergic synapse, Synaptic vesicle cycle and Spinocerebellar ataxia were up-regulated in AD but downregulated in cancers. A total of sixteen pathways were found to be deregulated in opposite directions between AD and cancer.

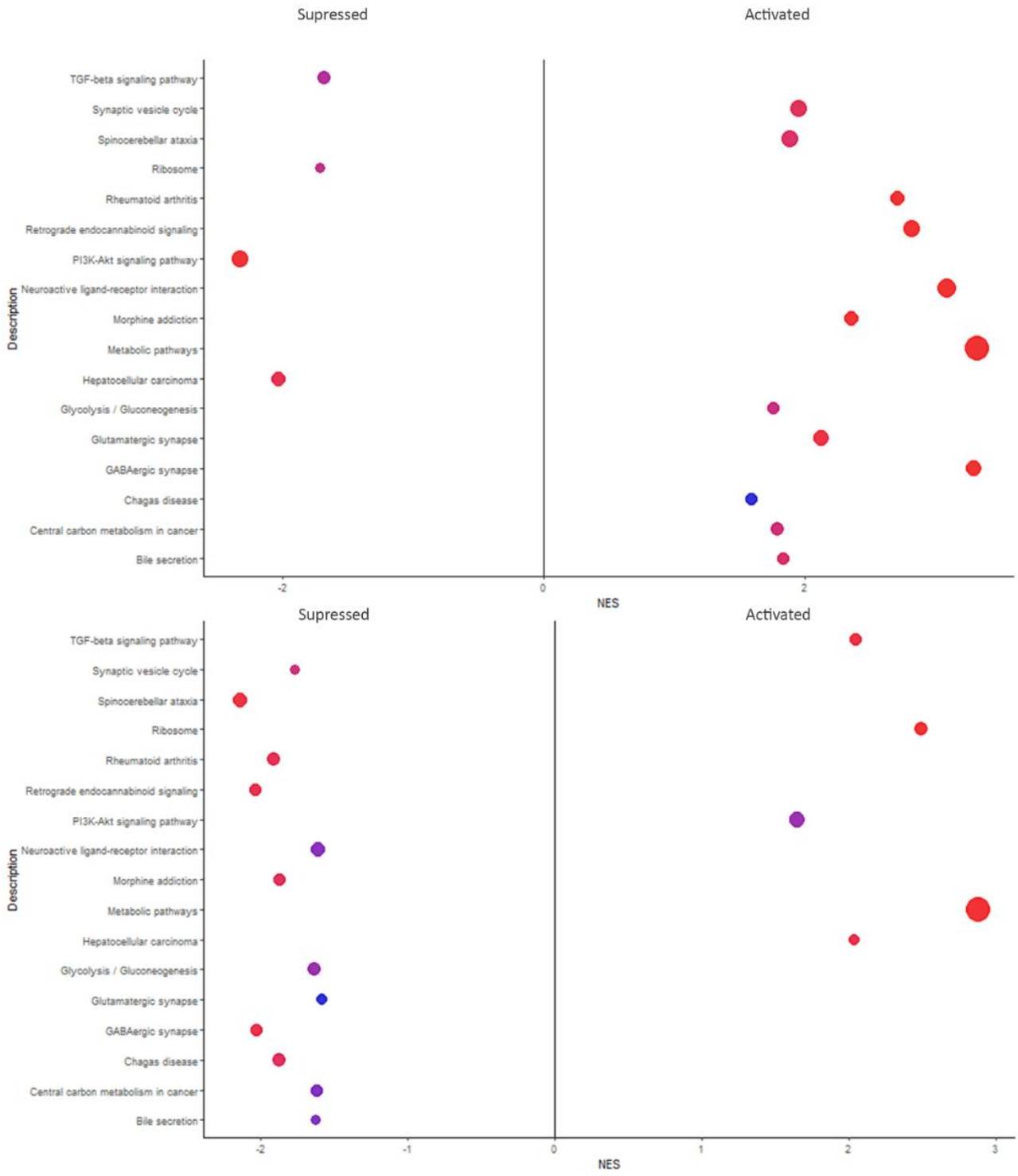


Figure 4-18 KEGG pathways inversely regulated between cancer and AD. A) shows regulation of pathways in AD. B) shows regulation of pathways in cancers. The size of circle is proportional to number of genes and NES represents enrichment score.

4.5 Discussion

To the best of our knowledge, this is the first study to include EC for studying the molecular association between AD and cancer. Among other cancers, EC has been previously reported as one of the most comorbid cancers with AD. We found a positive pattern of association between both types of cancers in this study and a negative pattern of association between AD and both types of cancers. This observation is in an agreement with previous expression studies and epidemiological studies reporting decreased risk of AD in cancer patients and vice versa.

The differential expression results reported genes in AD that have been previously linked to cancer. For example, CDKN2C (Cyclin-dependent kinase inhibitor 2C) is a gene that encodes for a protein involved in cell cycle regulation, was found to be down-regulated in AD. Furthermore, MAP2K1, the gene coding for primary protein in RAS/MAPK pathways which is involved in the growth and division of cells, was also found up-regulated in AD. Genes like CCNYL2 and NWD2 which are directly linked to cancer genes TP53 and MTOR respectively were also found inversely expressed. Mutations in the MSH2 (MutS homolog2) gene have been associated with several types of cancers. It is involved in DNA replication and required for DNA mismatch repair recognition. This kind of mismatch repair is also associated with the risks of AD. Many other cancers associated genes were found deregulated in opposite direction in AD. The p-value to test for the deregulation of these genes was also found to be significant. Thus, a molecular interpretation for inverse comorbidity that downregulation of certain genes reducing the risk of cancer would increase the risk of developing AD, while the upregulation of other genes which is increasing the risk of cancer would be decreasing the risk of developing AD.

Most commonalities between cancer and AD surrounds cell cycle related mechanisms. While AD pathophysiology is mostly described by beta amyloid deposition and neurofibrillary tangles, it is often explained by two hit cell cycle theory [60]. The two-hit theory demonstrates cell cycle re-entry as an important hallmark for AD and a commonality with cancer. According to this theory, for a normal cell to become AD it must first go through cell cycle re-entry process. Normally a mitogenic stimuli can cause neuronal cell to re-enter G2 phase of cell cycle, however, this effect is mitigated through apoptosis process. However, due to oxidative stress in aged brain tissues as a

second hit, the apoptosis process is avoided resulting in proliferation of neuronal cells and AB deposition.

The initiation of abnormal cell response correlates with oxidative stress that is usually triggered by the overexpression of ROS gene. Our result demonstrates ROS1 upregulation in AD. Several cell signalling associated pathways that malfunction in oxidative stress were also found to be deregulated in AD and cancer.

4.5.1 MAPK signalling pathway

The MAPKs are serine /threonine kinases that are involved in many cellular processes in diverse cell types. ERK1/2 are members a subfamily of MAPKs that are particular to extracellular signal regulated kinases [61]. ERK1/2 are well studied in mice for their protective role in stress conditions. Under stress conditions, MEK (mitogen activated kinase) phosphorylates ERK1/2 for their activation. Once activated ERK1/2 can bind with many substrates. Interaction with different substrates leads to dual nature of their action [61]. For example, MEK1/2 and ERK are both associated with upregulation of matrix metalloproteinase (MMP) and protecting cancer cells. In addition, ERK correlates with regulating pro-apoptic genes.

FoxO (Forkeahead box O) is a subfamily of transcription factors involved in cell fate decision and their function is also reported as tumor suppressors in a wide range of cancers. FoxOs are involved in a variety of cellular function such as apoptosis, differentiation, proliferation, and oxidative stress [62]. Activation of FoxO is tightly regulated by phosphorylation, acetylation and ubiquitination. FoxOs are activated by phosphorylation by several kinases such as PI3k (Phosphoinositide 3-kinases) and ERK1/2 (extracellular signal-regulated kinase 1 and 2) [63]. The role of Raf/MEK/ERK pathway in cell proliferation, apoptosis, and survival is documented.

Our results demonstrate over-expression of MEK1/2 gene in cancers which directly phosphorylates ERK1/2 which in turn hyper phosphorylate FoxO proteins. Hyper phosphorylation of FoxO protein through ERK1/2 leads to activation MDM2 gene and downregulation of pro-apoptic gene BIM. The GSEA analysis also resulted in activation of FoxO signalling pathway through MEK/ERK cascade. The initiation of apoptosis through FoxO proteins correlates with

oxidative stress. Paradoxically, MAPK signalling pathway is supressed in case of AD which would lead to activation of pro-apoptic genes causing cell death [64].

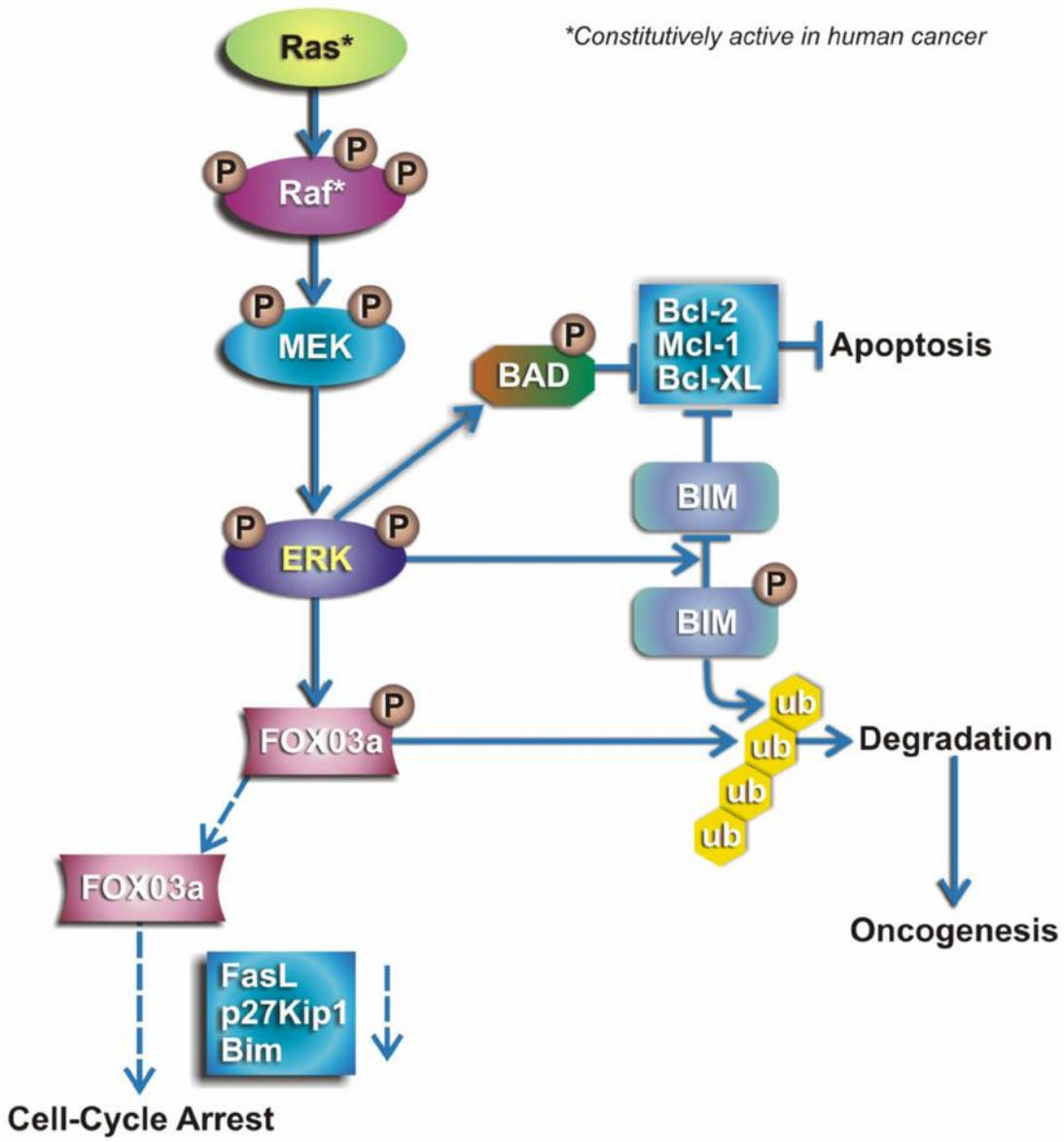


Figure 4-19 ERK1/2 promoting tumorigenesis by phosphorylating BIM, and thereby inhibiting apoptosis. Moreover, activation of FOXO3a facilitates its interaction with MDM2 that enhances cell survival [1].

4.5.2 PI3K/AKT/MTOR signalling Pathway

The PI3K/AKT/mTOR signalling pathway is an important target for molecular abnormalities between AD and cancer, which makes it a good comorbid candidate associating both diseases. Previously, it has been discussed as a potential candidate in terms of explaining cancer and AD

comorbidity. It is one of the most frequently affected pathways in human cancers and is involved in several key regulatory events. Our results showed that PI3K/AKT/MTOR is being downregulated in AD but upregulated in cancer.

4.5.3 GABAergic synapse

GABA (γ-aminobutyric acid) is the principal inhibitory neurotransmitter in the mammalian nervous system. It has been demonstrated that neurons with GABA receptors are resistant to AD pathology. Furthermore, some studies demonstrated that upregulation of the GABAA receptor increases the intracellular calcium levels and is involved in the activation of mitogen-activated protein kinase/extracellular signal-regulated kinase (MAPK/ERK) cascade which is another important hallmark of cancer. Our results showed downregulation of GABAergic synapse in AD and up-regulation in cancer.

Some other pathways like TGF-beta signalling, Central Carbon Metabolism in Cancer also showed a similar pattern. PI3K/AKT/mTOR has been previously studied regarding AD and cancer comorbidity and our findings align with previous transcriptome level studies. However, GABAergic synapse plays a vital role in AD and cancer comorbidity but have not been explored in this regard.

4.6 Limitations

Although this study presents novel insights into AD and cancer comorbidity, it has its limitations. Firstly, this study relied on publicly available transcriptomic data and the number of available datasets for AD is scarce. This fact undermines the statistical significance of meta-analysis. Furthermore, it lacks the necessary experimental information to adjust for batch effects. Also, this study does not include any counter diseases to account for a different comorbid scenario. Finally, based on previous studies on the subject, we conclude that including more datasets and diseases would not only increase the statistical significance of the findings but also strengthen the biological interpretations.

Chapter 5

CONCLUSION

A substantial amount of epidemiologic and scientific evidence suggests inverse comorbidity between cancer and Alzheimer's disease. Implication of both direct and indirect association depending on types of cancer has been explored. The relationship between AD and cancer is complex and involvement of many pathways and genes was observed. Among other, deregulation of the cell cycle can be considered as one of the primary and common culprits between two diseases. In this study, patterns of deregulation of common pathways were observed. There was a significant overlap between genes deregulated in the opposite direction between cancer and AD. In conclusion, deregulation of cell cycle as a result of activated PI3K/AKT/mTOR pathway can be considered as a trigger for neurodegeneration in AD and can be considered as an overlapping link between AD and cancer. GABAergic synapse is being studied for its involvement in cancers, our results demonstrate its involvement in inverse comorbidity between AD and cancer. Although many other common pathologies are yet to be explored, cell cycle related mechanisms may provide new insights in understanding either diseases or both. Additionally, future studies are required to investigate epigenetic factors involved in gene regulation that can serve as biomarkers of risk for both diseases, as well as repurposing use of existing cancer biomarkers towards AD treatment and diagnosis.

REFERENCES

[1] Y. Mebratu and Y. Tesfaigzi, “How ERK1/2 activation controls cell proliferation and cell death is subcellular localization the answer?,” *Cell Cycle*, vol. 8, no. 8, pp. 1168–1175, 2009, doi: 10.4161/cc.8.8.8147.

[2] J. Seo and M. Park, “Molecular crosstalk between cancer and neurodegenerative diseases,” *Cell. Mol. Life Sci.*, vol. 77, no. 14, pp. 2659–2680, 2020, doi: 10.1007/s00018-019-03428-3.

[3] E. Barrio-Alonso, A. Hernández-Vivanco, C. C. Walton, G. Perea, and J. M. Frade, “Cell cycle reentry triggers hyperploidization and synaptic dysfunction followed by delayed cell death in differentiated cortical neurons,” *Sci. Rep.*, vol. 8, no. 1, pp. 1–14, 2018, doi: 10.1038/s41598-018-32708-4.

[4] T. Sandal, “Molecular Aspects of the Mammalian Cell Cycle and Cancer,” *Oncologist*, vol. 7, no. 1, pp. 73–81, 2002, doi: 10.1634/theoncologist.7-1-73.

[5] J. Harrow *et al.*, “GENCODE: the reference human genome annotation for The ENCODE Project,” *Genome Res.*, vol. 22, no. 9, pp. 1760–1774, Sep. 2012, doi: 10.1101/gr.135350.111.

[6] G. A. Maston, S. K. Evans, and M. R. Green, “Transcriptional regulatory elements in the human genome.,” *Annu. Rev. Genomics Hum. Genet.*, vol. 7, pp. 29–59, 2006, doi: 10.1146/annurev.genom.7.080505.115623.

[7] L. A. Pennacchio, W. Bickmore, A. Dean, M. A. Nobrega, and G. Bejerano, “Enhancers: five essential questions.,” *Nature reviews. Genetics*, vol. 14, no. 4. pp. 288–295, Apr. 2013, doi: 10.1038/nrg3458.

[8] F. Aguet and K. G. Ardlie, “Tissue Specificity of Gene Expression,” *Curr. Genet. Med. Rep.*, vol. 4, no. 4, pp. 163–169, 2016, doi: 10.1007/s40142-016-0105-2.

[9] "Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans.," *Science*, vol. 348, no. 6235, pp. 648–660, May 2015, doi: 10.1126/science.1262110.

[10] F. Aguet *et al.*, "Genetic effects on gene expression across human tissues," *Nature*, vol. 550, no. 7675, pp. 204–213, 2017, doi: 10.1038/nature24277.

[11] M. Melé *et al.*, "Human genomics. The human transcriptome across tissues and individuals.," *Science*, vol. 348, no. 6235, pp. 660–665, May 2015, doi: 10.1126/science.aaa0355.

[12] T.-K. Kim and R. Shiekhattar, "Architectural and Functional Commonalities between Enhancers and Promoters.," *Cell*, vol. 162, no. 5, pp. 948–959, Aug. 2015, doi: 10.1016/j.cell.2015.08.008.

[13] P. J. Mitchell and R. Tjian, "Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins.," *Science*, vol. 245, no. 4916, pp. 371–378, Jul. 1989, doi: 10.1126/science.2667136.

[14] K. R. Mattaini, "chapter 17. regulation of gene expression," in *introduction to molecular and cell biology*, .

[15] D. J. Lockhart and E. A. Winzeler, "Genomics, gene expression and DNA arrays.," *Nature*, vol. 405, no. 6788, pp. 827–836, Jun. 2000, doi: 10.1038/35015701.

[16] Z. Wang, M. Gerstein, and M. Snyder, "RNA-Seq: a revolutionary tool for transcriptomics.," *Nat. Rev. Genet.*, vol. 10, no. 1, pp. 57–63, Jan. 2009, doi: 10.1038/nrg2484.

[17] A. Oshlack, M. D. Robinson, and M. D. Young, "From RNA-seq reads to differential expression results," *Genome Biol.*, vol. 11, no. 12, p. 220, 2010, doi: 10.1186/gb-2010-11-12-220.

[18] R. Tabarés-Seisdedos and A. Baudot, "Editorial: Direct and Inverse Comorbidities Between

Complex Disorders," *Front. Physiol.*, vol. 7, p. 117, Mar. 2016, doi: 10.3389/fphys.2016.00117.

[19] C. Rubio-Perez *et al.*, "Genetic and functional characterization of disease associations explains comorbidity," *Sci. Rep.*, vol. 7, no. 1, pp. 1–14, 2017, doi: 10.1038/s41598-017-04939-4.

[20] B. E Yu *et al.*, "The analysis of associations between cytokine network genes and inverse co-morbidity of bronchial asthma and tuberculosis," *Biomed. Genet. Genomics*, vol. 1, no. 5, 2016, doi: 10.15761/bgg.1000122.

[21] J. X. Hu, C. E. Thomas, and S. Brunak, "Network biology concepts in complex disease comorbidities.," *Nat. Rev. Genet.*, vol. 17, no. 10, pp. 615–629, Oct. 2016, doi: 10.1038/nrg.2016.87.

[22] D. E. Weiner *et al.*, "Kidney disease as a risk factor for recurrent cardiovascular disease and mortality1 1The Atherosclerosis Risk in Communities Study, Cardiovascular Health Study, and the Framingham Heart and Framingham Offspring studies are conducted and supported by the N," *Am. J. Kidney Dis.*, vol. 44, no. 2, pp. 198–206, 2004, doi: https://doi.org/10.1053/j.ajkd.2004.04.024.

[23] B. Starfield, K. W. Lemke, T. Bernhardt, S. S. Foldes, C. B. Forrest, and J. P. Weiner, "Comorbidity: implications for the importance of primary care in 'case' management.," *Ann. Fam. Med.*, vol. 1, no. 1, pp. 8–14, 2003, doi: 10.1370/afm.1.

[24] S. Teven *et al.*, "mor tality from coronary hear t disease in sub jec ts with and without type 2 diabetes mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects with and without prior myocardial infarction a bstract Background Typ," *N. Engl. J. Med.*, vol. 339, pp. 229–234, 1998

[25] T. G. von Lueder and D. Atar, "Comorbidities and Polypharmacy," *Heart Fail. Clin.*, vol. 10, no. 2, pp. 367–372, 2014, doi: https://doi.org/10.1016/j.hfc.2013.12.001.

[26] A. Levin *et al.*, "Cardiovascular disease in patients with chronic kidney disease: Getting to

the heart of the matter," *Am. J. Kidney Dis.*, vol. 38, no. 6, pp. 1398–1407, 2001, doi: https://doi.org/10.1053/ajkd.2001.29275.

[27] F. He, G. Zhu, Y. Y. Wang, X. M. Zhao, and D. S. Huang, "PCID: A novel approach for predicting disease comorbidity by integrating multi-scale data," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 14, no. 3, pp. 678–686, 2017, doi: 10.1109/TCBB.2016.2550443.

[28] S. Akın and C. Bölük, "Prevalence of comorbidities in patients with type–2 diabetes mellitus," *Prim. Care Diabetes*, vol. 14, no. 5, pp. 431–434, 2020, doi:10.1016/j.pcd.2019.12.006.

[29] J. Sánchez-Valle *et al.*, "A molecular hypothesis to explain direct and inverse co-morbidities between Alzheimer's Disease, Glioblastoma and Lung cancer," *Sci. Rep.*, vol. 7, no. 1, pp. 1–12, 2017, doi: 10.1038/s41598-017-04400-6.

[30] R. Tabarés-Seisdedos and J. L. Rubenstein, "Inverse cancer comorbidity: a serendipitous opportunity to gain insight into CNS disorders," *Nat. Rev. Neurosci.*, vol. 14, no. 4, pp. 293–304, 2013, doi: 10.1038/nrn3464.

[31] K. Baek *et al.*, "HHS Public Access," vol. 459, no. 7250, pp. 1126–1130, 2009, doi: 10.1038/nature08062.Down.

[32] J. A. N. CORSELLIS, *MENTAL ILLNESS AND THE AGEING BRAIN The Distribution of Pathological Change in a Mental Hospital Population*. Oxford University Press, London (1962), 1962.

[33] D. R. Royall and R. K. Mahurin, "ECF Deficits and Anorectic Behavior," *J. Am. Geriatr. Soc.*, vol. 39, no. 8, pp. 840–841, 1991, doi: 10.1111/j.1532-5415.1991.tb02714.x.

[34] M. Yamada *et al.*, "Prevalence and risks of dementia in the Japanese population: RERF's Adult Health Study Hiroshima subjects," *J. Am. Geriatr. Soc.*, vol. 47, no. 2, pp. 189–195, 1999, doi: 10.1111/j.1532-5415.1999.tb04577.x.

[35] C. M. Roe, M. I. Behrens, C. Xiong, J. P. Miller, and J. C. Morris, "Alzheimer disease and

cancer," *Neurology*, vol. 64, no. 5, pp. 895 LP – 898, Mar. 2005, doi: 10.1212/01.WNL.0000152889.94785.51.

[36] M. Ganguli, H. H. Dodge, C. Shen, R. S. Pandav, and S. T. DeKosky, "Alzheimer Disease and Mortality: A 15-Year Epidemiological Study," *Arch. Neurol.*, vol. 62, no. 5, pp. 779–784, May 2005, doi: 10.1001/archneur.62.5.779.

[37] J. A. Driver *et al.*, "Inverse association between cancer and Alzheimer{\textquoteright}s disease: results from the Framingham Heart Study," *BMJ*, vol. 344, 2012, doi: 10.1136/bmj.e1442.

[38] M. Musicco *et al.*, "Inverse occurrence of cancer and Alzheimer disease," *Neurology*, vol. 81, no. 4, pp. 322 LP – 328, Jul. 2013, doi: 10.1212/WNL.0b013e31829c5ec1.

[39] N. Vanacore, S. Spila-Alegiani, R. Raschetti, and G. Meco, "Mortality cancer risk in parkinsonian patients: a population-based study.," *Neurology*, vol. 52, no. 2, pp. 395–398, Jan. 1999, doi: 10.1212/wnl.52.2.395.

[40] F. Catalá-López *et al.*, "Inverse and direct cancer comorbidity in people with central nervous system disorders: A meta-analysis of cancer incidence in 577,013 participants of 50 observational studies," *Psychother. Psychosom.*, vol. 83, no. 2, pp. 89–105, 2014, doi: 10.1159/000356498.

[41] S. Lehrer, "Glioblastoma and dementia may share a common cause.," *Med. Hypotheses*, vol. 75, no. 1, pp. 67–68, Jul. 2010, doi: 10.1016/j.mehy.2010.01.031.

[42] K. Ibáñez, C. Boullosa, R. Tabarés-Seisdedos, A. Baudot, and A. Valencia, "Molecular Evidence for the Inverse Comorbidity between Central Nervous System Disorders and Cancers Detected by Transcriptomic Meta-analyses," *PLoS Genet.*, vol. 10, no. 2, pp. 1–7, 2014, doi: 10.1371/journal.pgen.1004173.

[43] M. J. Devine, H. Plun-Favreau, and N. W. Wood, "Parkinson's disease and cancer: two wars, one front.," *Nat. Rev. Cancer*, vol. 11, no. 11, pp. 812–823, Oct. 2011, doi: 10.1038/nrc3150.

[44] P. Klus, D. Cirillo, T. Botta Orfila, and G. Gaetano Tartaglia, “Neurodegeneration and cancer: where the disorder prevails,” *Sci. Rep.*, vol. 5, pp. 1–7, 2015, doi: 10.1038/srep15390.

[45] M. Musicco *et al.*, “Inverse occurrence of cancer and Alzheimer disease: a population-based incidence study.,” *Neurology*, vol. 81, no. 4, pp. 322–328, Jul. 2013, doi: 10.1212/WNL.0b013e31829c5ec1.

[46] T. J. P. Hubbard *et al.*, “Ensembl 2007,” *Nucleic Acids Res.*, vol. 35, no. SUPPL. 1, pp. 610–617, 2007, doi: 10.1093/nar/gkl996.

[47] J. Navarro Gonzalez *et al.*, “The UCSC genome browser database: 2021 update,” *Nucleic Acids Res.*, vol. 49, no. D1, pp. D1046–D1057, 2021, doi: 10.1093/nar/gkaa1070.

[48] A. Auton *et al.*, “A global reference for human genetic variation,” *Nature*, vol. 526, no. 7571, pp. 68–74, 2015, doi: 10.1038/nature15393.

[49] S. T. Sherry *et al.*, “DbSNP: The NCBI database of genetic variation,” *Nucleic Acids Res.*, vol. 29, no. 1, pp. 308–311, 2001, doi: 10.1093/nar/29.1.308.

[50] Andrews, “No Title,” *FastQC: a quality control tool for high throughput sequence data.*, 2010.

[51] S. Chen, Y. Zhou, Y. Chen, and J. Gu, “Fastp: An ultra-fast all-in-one FASTQ preprocessor,” *Bioinformatics*, vol. 34, no. 17, pp. i884–i890, 2018, doi: 10.1093/bioinformatics/bty560.

[52] K. Froussios *et al.*, “How well do RNA-Seq differential gene expression tools perform in a complex eukaryote? A case study in Arabidopsis thaliana,” *Bioinformatics*, vol. 35, no. 18, pp. 3372–3377, 2019, doi: 10.1093/bioinformatics/btz089.

[53] M. I. Love, W. Huber, and S. Anders, “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2,” *Genome Biol.*, vol. 15, no. 12, p. 550, Dec. 2014, doi: 10.1186/s13059-014-0550-8.

[54] A. Rau, G. Marot, and F. Jaffrézic, “Differential meta-analysis of RNA-seq data from multiple studies.,” *BMC Bioinformatics*, vol. 15, p. 91, Mar. 2014, doi: 10.1186/1471-2105-15-91.

[55] S. Is. Shen L, “GeneOverlap: Test and visualize gene overlaps. R package version 1.30.0.” 2021.

[56] G. Stelzer *et al.*, “VarElect: the phenotype-based variation prioritizer of the GeneCards Suite,” *BMC Genomics*, vol. 17, no. 2, p. 444, 2016, doi: 10.1186/s12864-016-2722-2.

[57] S. Colby, “Cancer and Chemotherapy Are Associated With a Reduced Alzheimer’s Risk,” *Neurol. Rev.*, 2013.

[58] S. Colby, “Cancer and Chemotherapy Are Associated With a Reduced Alzheimer’s Risk,” *Neurol. Rev.*, 2013.

[59] B. H. You, J. H. Yoon, H. Kang, E. K. Lee, S. K. Lee, and J. W. Nam, “HERES, a lncRNA that regulates canonical and noncanonical Wnt signaling pathways via interaction with EZH2,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 116, no. 49, pp. 24620–24629, 2019, doi: 10.1073/pnas.1912126116.

[60] A. G. Clark and E. Paluch, “Cell Cycle in Development,” *Cell Cycle*, vol. 53, pp. 31–73, 2011, doi: 10.1007/978-3-642-19065-0.

[61] E. K. Kim and E.-J. Choi, “Pathological roles of MAPK signaling pathways in human diseases,” *Biochim. Biophys. Acta - Mol. Basis Dis.*, vol. 1802, no. 4, pp. 396–405, 2010, doi:10.1016/j.bbadis.2009.12.009.

[62] M. Farhan, H. Wang, U. Gaur, P. J. Little, J. Xu, and W. Zheng, “FOXO Signaling Pathways as Therapeutic Targets in Cancer,” *Int. J. Biol. Sci.*, vol. 13, no. 7, pp. 815–827, 2017, doi: 10.7150/ijbs.20052.

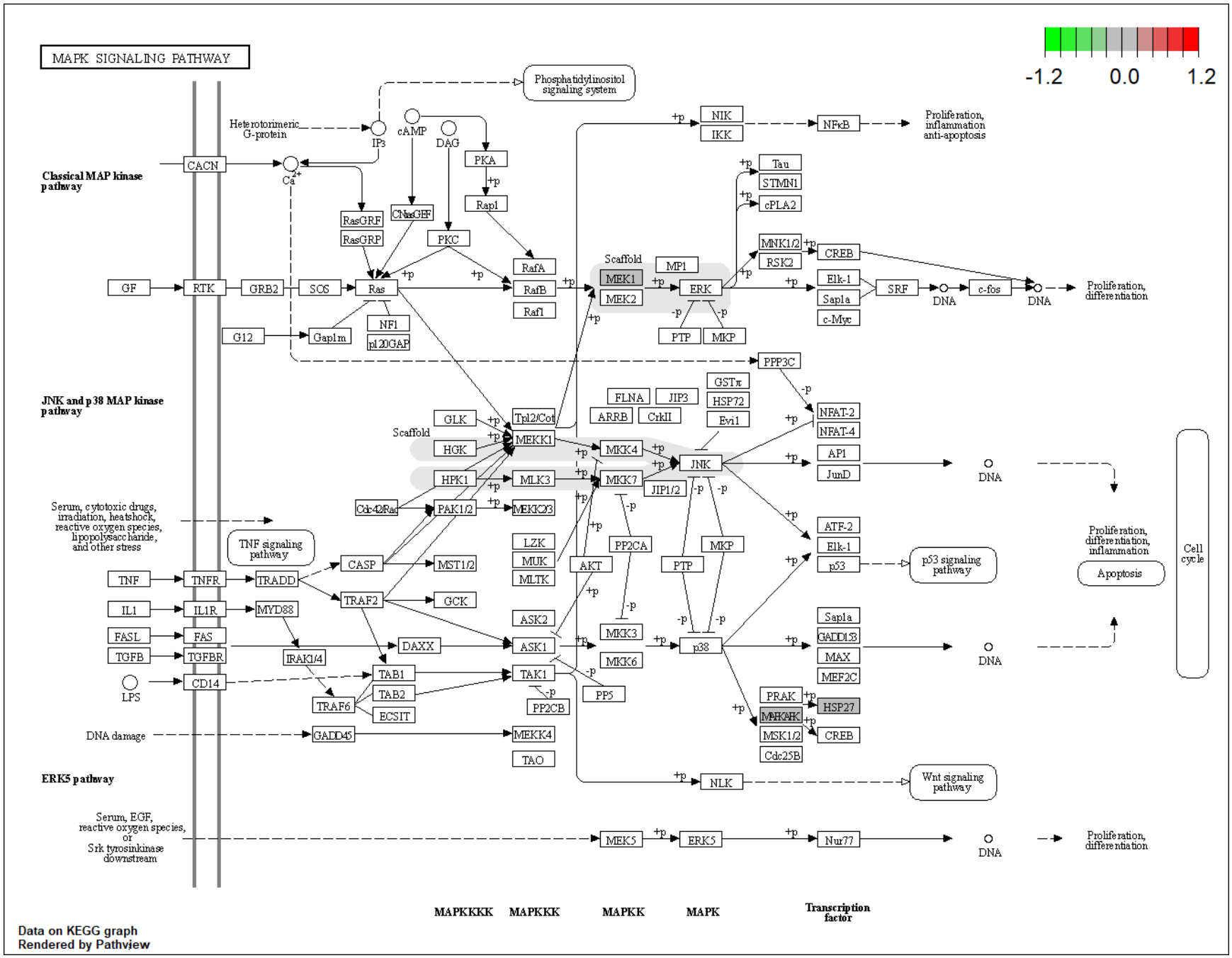
[63] Y. Jiramongkol and E. W.-F. Lam, “FOXO transcription factor family in cancer and metastasis,” *Cancer Metastasis Rev.*, vol. 39, no. 3, pp. 681–709, 2020, doi:

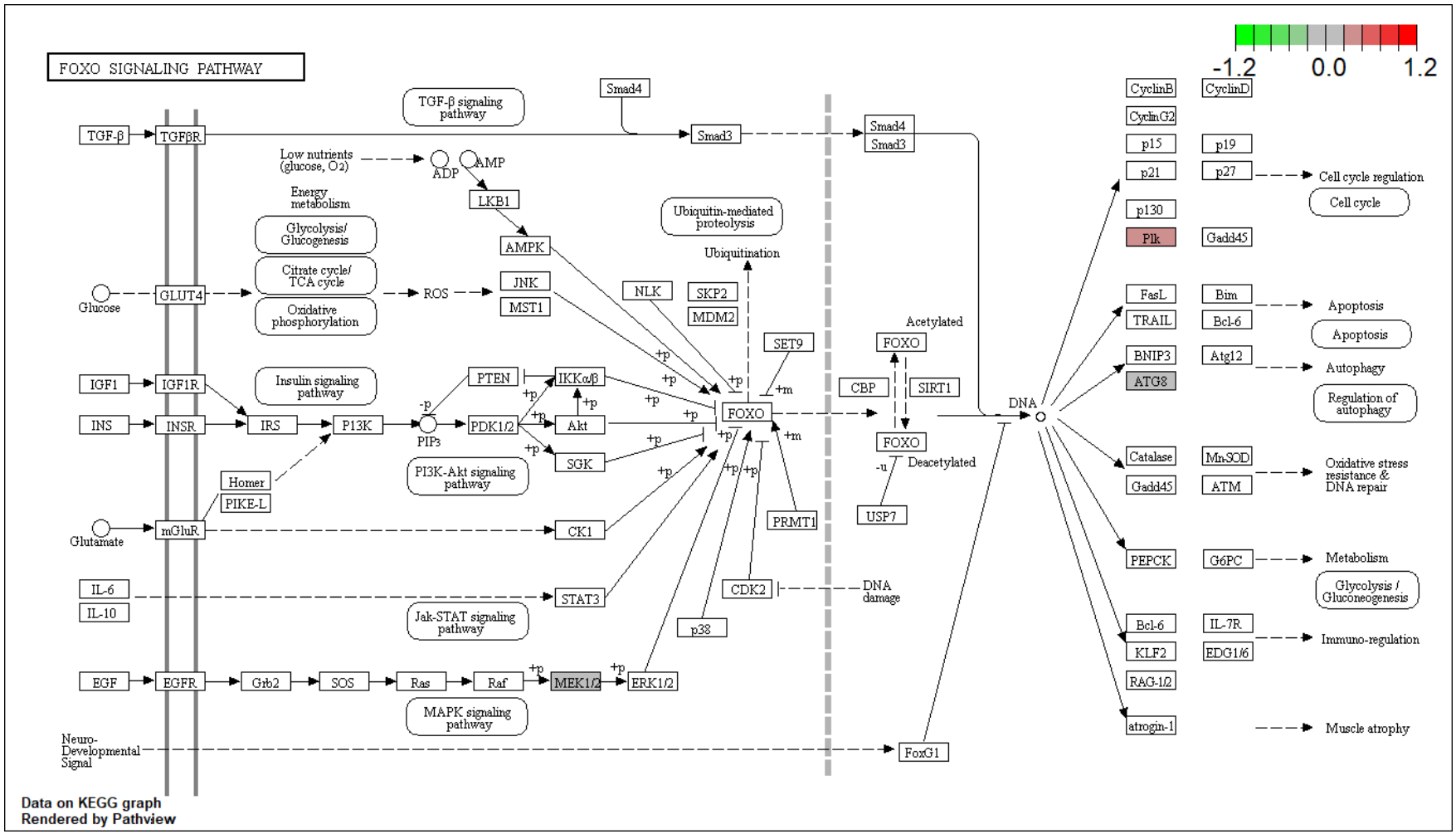
10.1007/s10555-020-09883-w.

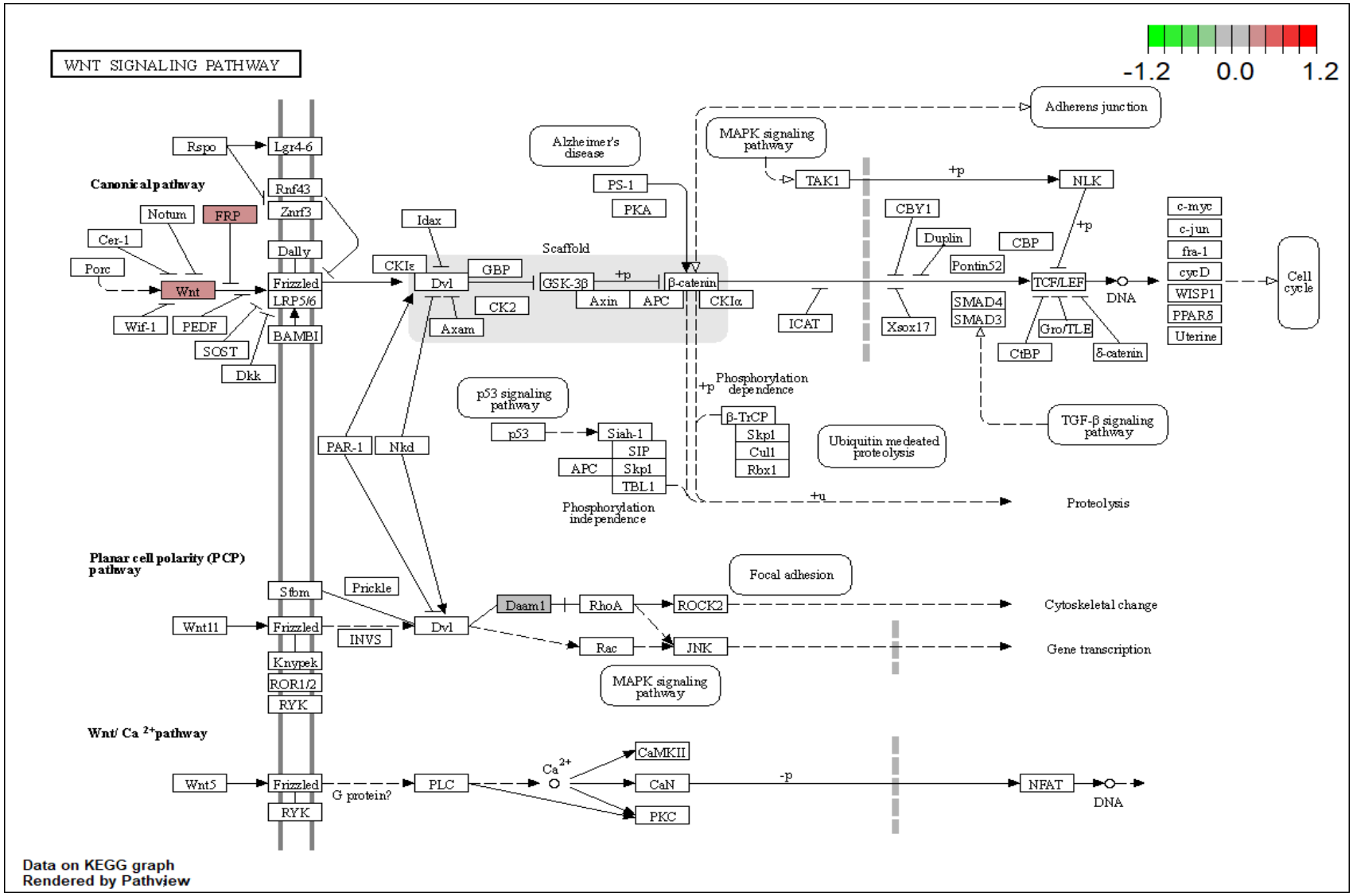
[64] Z. Lu and S. Xu, “ERK1/2 MAP kinases in cell survival and apoptosis,” *IUBMB Life*, vol. 58, no. 11, pp. 621–631, 2006, doi: 10.1080/15216540600957438.

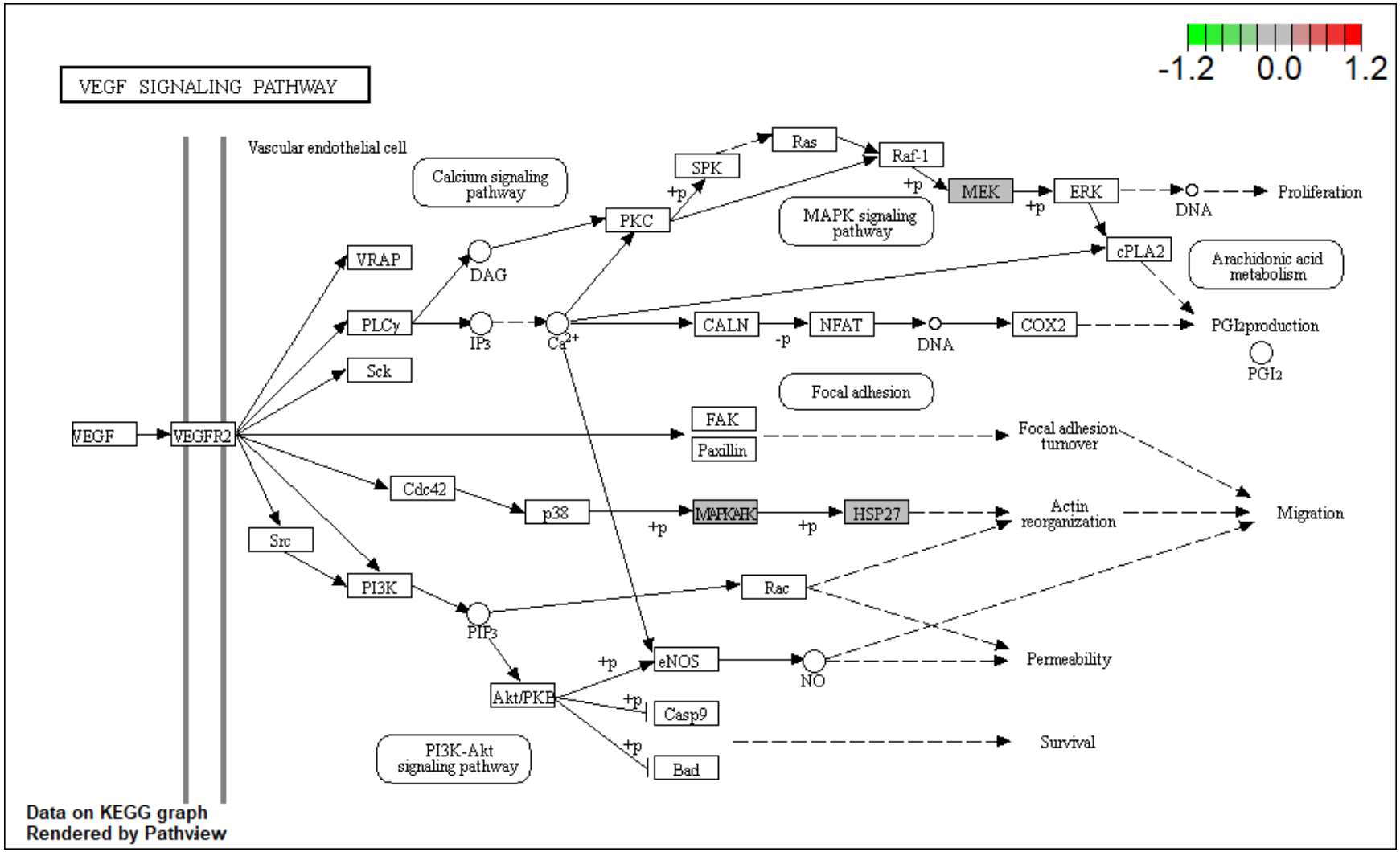
Pathways in AD

ID	Description	setSize	enrichmer	NES	pvalue
hsa04310	Wnt signaling pathway	3	0.848507	1.408121	0.046078
hsa05017	Spinocerebellar ataxia	5	0.734132	1.389109	0.075004
hsa05022	Pathways of neurodegeneration - multiple diseases	9	0.607389	1.333081	0.112974
hsa04010	MAPK signaling pathway	3	-0.7919	-1.41051	0.115909
hsa04370	VEGF signaling pathway	3	-0.7919	-1.41051	0.115909
hsa04080	Neuroactive ligand-receptor interaction	6	0.665448	1.319028	0.125125
hsa05166	Human T-cell leukemia virus 1 infection	4	-0.61966	-1.20767	0.250535
hsa04934	Cushing syndrome	4	0.640673	1.144486	0.284887
hsa05010	Alzheimer disease	7	0.558045	1.152025	0.28754
hsa04068	FoxO signaling pathway	3	0.684053	1.135206	0.297118
hsa04666	Fc gamma R-mediated phagocytosis	3	-0.62553	-1.11417	0.32983
hsa05205	Proteoglycans in cancer	3	0.669111	1.110409	0.330714
hsa05014	Amyotrophic lateral sclerosis	6	0.560804	1.111607	0.337437
hsa04218	Cellular senescence	3	-0.61702	-1.09902	0.342045
hsa05033	Nicotine addiction	3	0.659894	1.095113	0.352597
hsa05200	Pathways in cancer	8	0.510273	1.089601	0.370304
hsa04270	Vascular smooth muscle contraction	3	0.652253	1.082432	0.372785
hsa04071	Sphingolipid signaling pathway	5	-0.48574	-1.01396	0.410158
hsa01100	Metabolic pathways	21	0.395725	1.051812	0.415373
hsa04020	Calcium signaling pathway	3	0.617484	1.024733	0.458622


MAPK SIGNALING PATHWAY
-1.2
0.0
1.2
Phosphatidylinositol signaling system
Heterotorimeric G-protein
IP3
cAMP
DAG
PKA
Rap1
RasGRF
RasGRP
CNksGEF
PKC
CACN
Ca2+
Classical MAP kinase pathway
GF
RTK
GRB2
SOS
Ras
RafA
RafB
Raf1
MEK1
MEK2
Scaffold
MP1
ERK
PTP
MKP
NIK
IKK
NFκB
Proliferation, inflammation anti-apoptosis
Tau
STMN1
cPLA2
MNK1/2
RSK2
CREB
Elk-1
Sap1a
c-Myc
SRF
DNA
c-fos
DNA
Proliferation, differentiation
G12
Gap1m
NF1
p120GAP
PPP3C
JNK and p38 MAP kinase pathway
GLK
HGK
HPK1
Cdc42/Rac
PAK1/2
Tpl2/Cot
MEKK1
MLK3
MEKK2/3
Scaffold
FLNA
ARRB
JIP3
CrkII
GSTπ
HSP72
Evi1
MKK4
MKK7
JIP1/2
JNK
NFAT-2
NFAT-4
AP1
JunD
DNA
ATF-2
Elk-1
p53
p53 signaling pathway
Proliferation, differentiation, inflammation
Apoptosis
Cell cycle
Serum, cytotoxic drugs, irradiation, heatshock, reactive oxygen species, lipopolysaccharide, and other stress
TNF signaling pathway
CASP
MST1/2
LZK
MUK
MLTK
AKT
PP2CA
PTP
MKP
TNF
TNFR
TRADD
TRAF2
GCK
IL1
IL1R
MYD88
FASL
FAS
TGFB
TGFBR
IRAK1/4
DAXX
ASK2
ASK1
MKK3
MKK6
p38
Sap1a
GADD153
MAX
MEF2C
DNA
LPS
CD14
TAB1
TAB2
ECSIT
TAK1
PP2CB
PP5
TRAF6
PRAK
MAPKAPK
HSP27
MSK1/2
CREB
Cdc25B
DNA damage
GADD45
MEKK4
TAO
ERK5 pathway
NLK
Wnt signaling pathway
Serum, EGF, reactive oxygen species, or Src tyrosinkinase downstream
MEK5
ERK5
Nur77
DNA
Proliferation, differentiation
MAPKKKK
MAPKKK
MAPKK
MAPK
Transcription factor
Data on KEGG graph
Rendered by Pathview

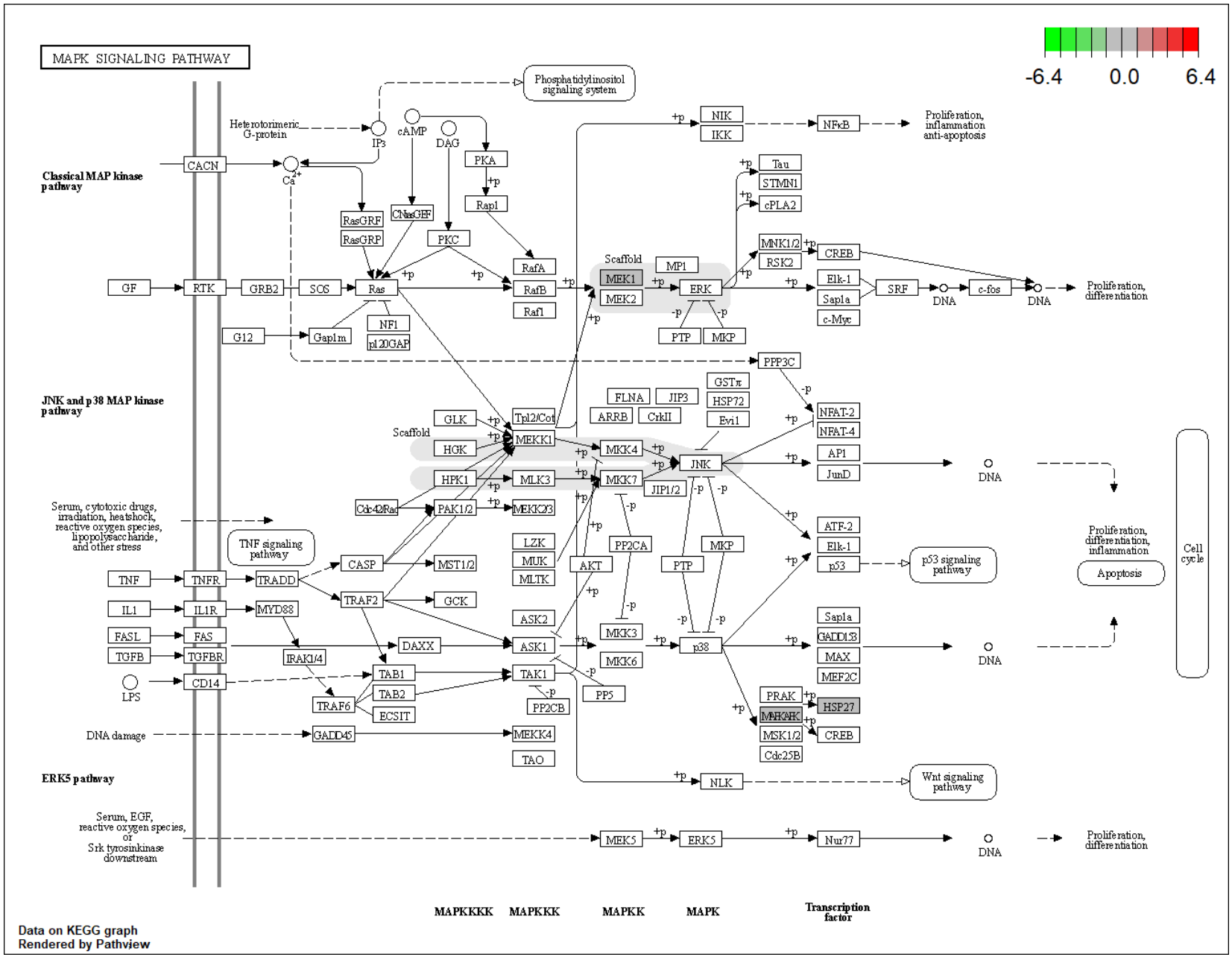

FOXO SIGNALING PATHWAY
-1.2
0.0
1.2
TGF-β
TGFβR
TGF-β signaling pathway
Smad4
Smad3
Smad4
Smad3
Low nutrients (glucose, O2)
ADP
AMP
LKB1
Energy metabolism
Glycolysis/ Glucogenesis
Citrate cycle/ TCA cycle
Oxidative phosphorylation
AMPK
Ubiquitin-mediated proteolysis
Ubiquitination
Glucose
GLUT4
ROS
JNK
MST1
NLK
SKP2
MDM2
SET9
IGF1
IGF1R
Insulin signaling pathway
INS
INSR
IRS
P13K
PIP3
PTEN
PDK1/2
IKKα/β
Akt
SGK
PI3K-Akt signaling pathway
FOXO
CBP
SIRT1
Acetylated
FOXO
Deacetylated
-u
USP7
PRMT1
DNA
Homer
PIKE-L
Glutamate
mGluR
CK1
IL-6
IL-10
STAT3
Jak-STAT signaling pathway
CDK2
DNA damage
p38
EGF
EGFR
Grb2
SOS
Ras
Raf
MEK1/2
ERK1/2
MAPK signaling pathway
Neuro-Developmental Signal
FoxG1
CyclinB
CyclinD
CyclinG2
p15
p19
p21
p27
Cell cycle regulation
Cell cycle
p130
Plk
Gadd45
FasL
Bim
Apoptosis
TRAIL
Bcl-6
Apoptosis
BNIP3
Atg12
Autophagy
ATG8
Regulation of autophagy
Catalase
MnSOD
Oxidative stress resistance & DNA repair
Gadd45
ATM
PEPCK
G6PC
Metabolism
Glycolysis / Gluconeogenesis
Bcl-6
IL-7R
Immuno-regulation
KLF2
EDG1/6
RAG-1/2
atrogin-1
Muscle atrophy
Data on KEGG graph
Rendered by Pathview

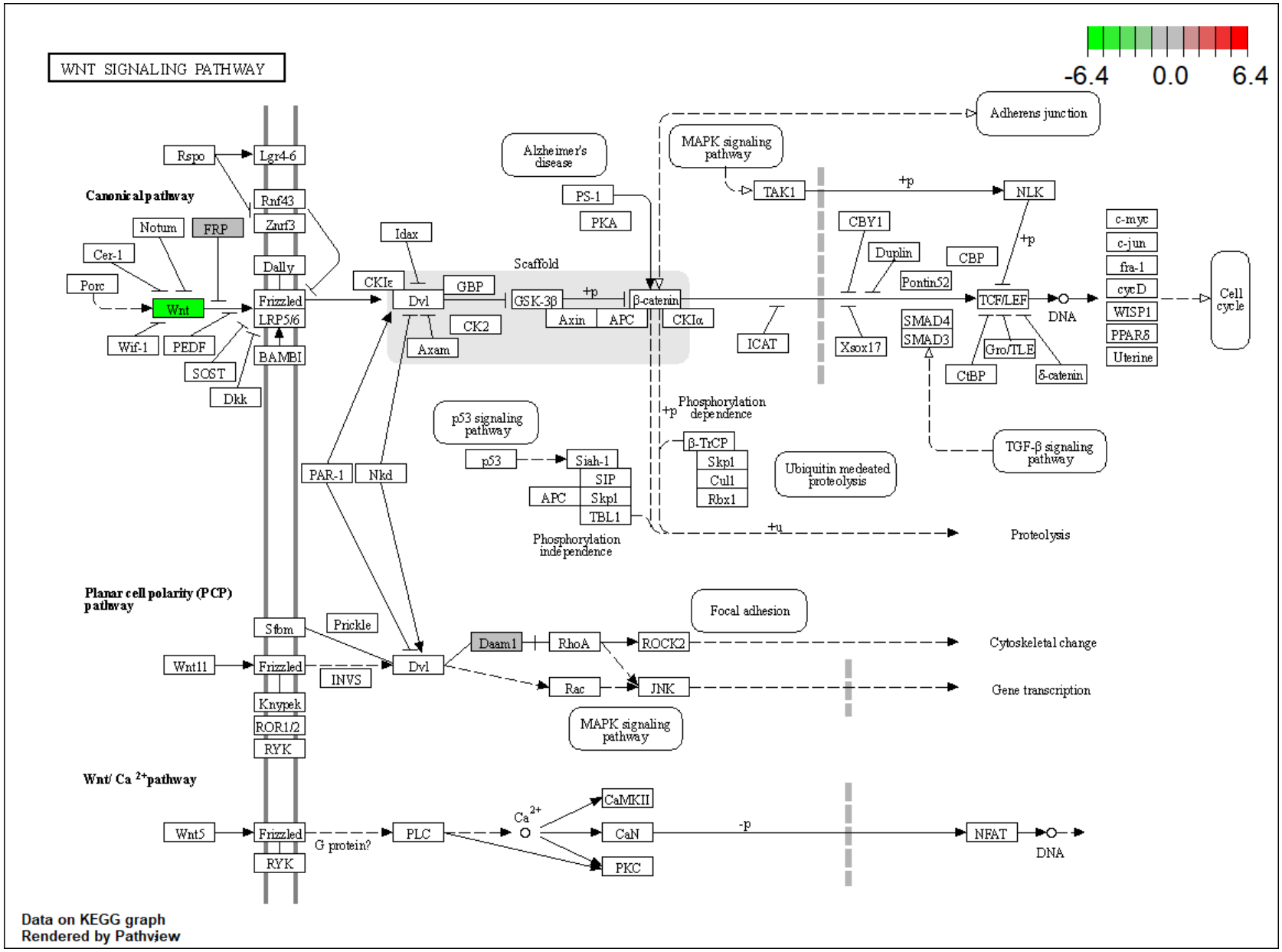

WNT SIGNALING PATHWAY
-1.2
0.0
1.2
Canonical pathway
Rspo
Lgr4-6
Rnf43
Znrf3
Notum
FRP
Cer-1
Dally
Porc
Wnt
Frizzled
LRP5/6
Wif-1
PEDF
SOST
Dkk
BAMBI
Idax
CKIε
GBP
Dvl
Axam
CK2
Scaffold
GSK-3β
+p
β-catenin
Axin
APC
CKIα
Alzheimer's disease
PS-1
PKA
MAPK signaling pathway
TAK1
+p
NLK
Adherens junction
CBY1
Duplin
CBP
Pontin52
+p
TCF/LEF
DNA
c-myc
c-jun
fra-1
cycD
WISP1
PPARδ
Uterine
Cell cycle
ICAT
Xsox17
SMAD4
SMAD3
Gro/TLE
CtBP
δ-catenin
TGF-β signaling pathway
p53 signaling pathway
p53
Siah-1
SIP
APC
Skp1
TBL1
+p
Phosphorylation dependence
β-TrCP
Skp1
Cul1
Rbx1
Ubiquitin mediated proteolysis
PAR-1
Nkd
Phosphorylation independence
+u
Proteolysis
Planar cell polarity (PCP) pathway
Stbm
Prickle
Wnt11
Frizzled
INVS
Dvl
Daam1
RhoA
ROCK2
Focal adhesion
Cytoskeletal change
Rac
JNK
Gene transcription
MAPK signaling pathway
Knypek
ROR1/2
RYK
Wnt/ Ca 2+pathway
Wnt5
Frizzled
RYK
G protein?
PLC
Ca2+
CaMKII
CaN
PKC
-p
NFAT
DNA
Data on KEGG graph
Rendered by Pathview

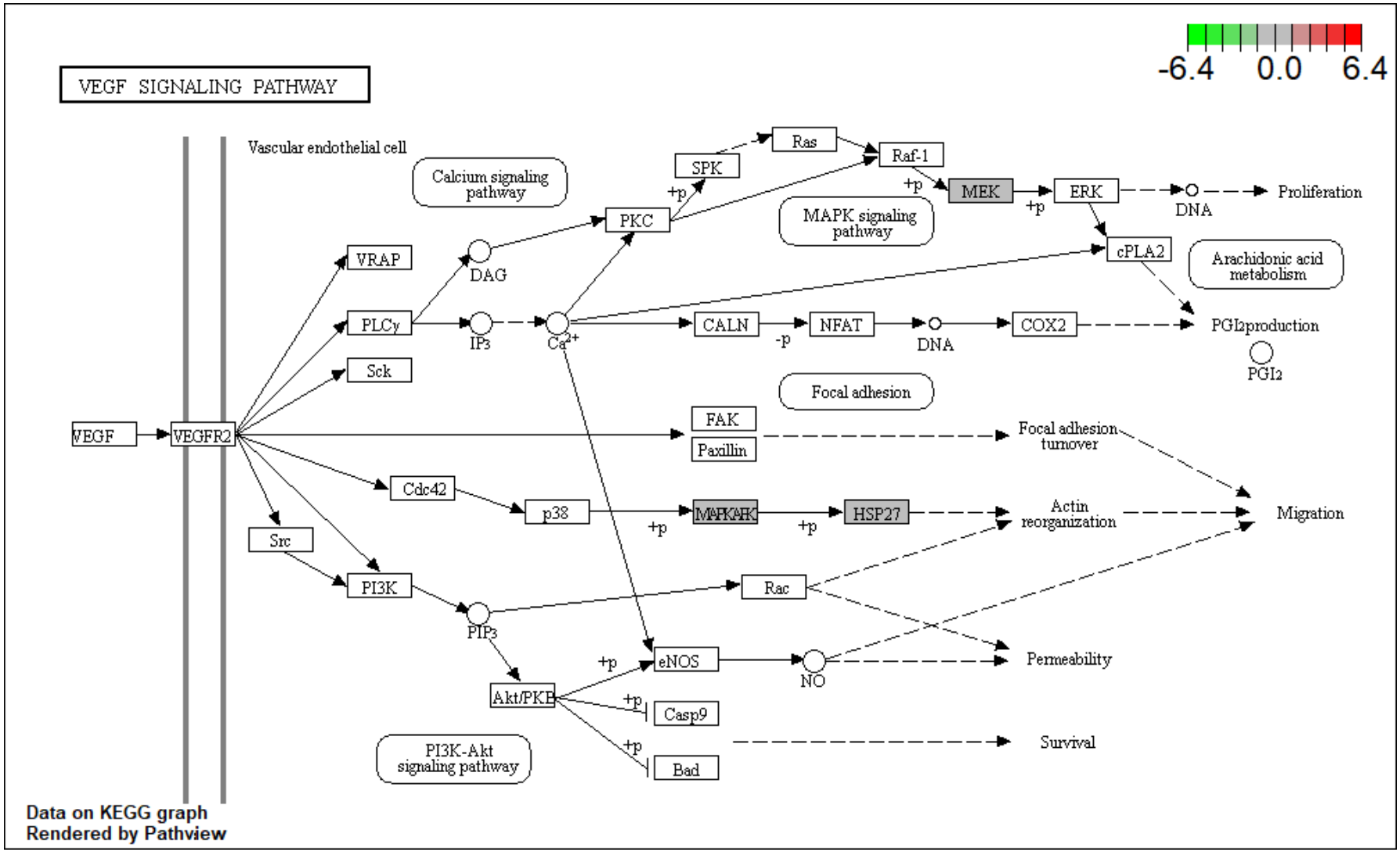

VEGF SIGNALING PATHWAY
-1.2
0.0
1.2
Vascular endothelial cell
Calcium signaling pathway
MAPK signaling pathway
Arachidonic acid metabolism
Focal adhesion
PI3K-Akt signaling pathway
VEGF
VEGFR2
VRAP
PLCγ
Sck
Cdc42
Src
PI3K
DAG
IP3
PIP3
PKC
SPK
Ras
Raf-1
MEK
ERK
DNA
Proliferation
cPLA2
CALN
NFAT
COX2
PGI2production
PGI2
Ca2+
FAK
Paxillin
Focal adhesion turnover
p38
MAPKAPK
HSP27
Actin reorganization
Migration
Rac
eNOS
NO
Permeability
Akt/PKB
Casp9
Bad
Survival
+p
-p
Data on KEGG graph
Rendered by Pathview

Pathways in Cancers

ID	Description	setSize	enrichmer	NES	pvalue
hsa04360	Axon guidance	4	0.863248	1.816262	0.010076
hsa04934	Cushing syndrome	4	-0.83116	-1.43924	0.032136
hsa05205	Proteoglycans in cancer	3	-0.85212	-1.38019	0.050443
hsa04390	Hippo signaling pathway	3	-0.83466	-1.35191	0.07116
hsa04010	MAPK signaling pathway	3	0.75415	1.419202	0.100448
hsa04370	VEGF signaling pathway	3	0.75415	1.419202	0.100448
hsa04080	Neuroactive ligand-receptor interaction	6	-0.69334	-1.31621	0.115946
hsa04310	Wnt signaling pathway	3	-0.79148	-1.28197	0.133313
hsa04727	GABAergic synapse	6	-0.67738	-1.28591	0.140349
hsa05010	Alzheimer disease	7	-0.65326	-1.28353	0.15195
hsa04071	Sphingolipid signaling pathway	5	0.566524	1.299016	0.161471
hsa05033	Nicotine addiction	3	-0.77021	-1.24752	0.173848
hsa05165	Human papillomavirus infection	5	-0.68167	-1.24591	0.183467
hsa04723	Retrograde endocannabinoid signaling	4	-0.70966	-1.22885	0.197727
hsa05032	Morphine addiction	4	-0.70966	-1.22885	0.197727
hsa04380	Osteoclast differentiation	3	0.617021	1.161146	0.274738
hsa05022	Pathways of neurodegeneration - multiple diseases	9	-0.55781	-1.1529	0.293045


MAPK SIGNALING PATHWAY
-6.4
0.0
6.4
Classical MAP kinase pathway
JNK and p38 MAP kinase pathway
ERK5 pathway
Phosphatidylinositol signaling system
Heterotorimeric G-protein
cAMP
DAG
IP3
CACN
Ca2+
PKA
Rap1
RasGRF
RasGRP
CNksGEF
PKC
GF
RTK
GRB2
SOS
Ras
RafA
RafB
Raf1
G12
Gap1m
NF1
p120GAP
Scaffold
MP1
MEK1
MEK2
ERK
PTP
MKP
NIK
IKK
NFκB
Proliferation, inflammation anti-apoptosis
Tau
STMN1
cPLA2
MNK1/2
RSK2
CREB
Elk-1
Sap1a
c-Myc
SRF
DNA
c-fos
Proliferation, differentiation
PPP3C
FLNA
JIP3
HSP72
ARRB
CrkII
Evi1
GSTπ
NFAT-2
NFAT-4
AP1
JunD
GLK
HGK
HPK1
Tpl2/Cot
MEKK1
MLK3
MKK4
MKK7
JNK
JIP1/2
Cdc42/Rac
PAK1/2
MEKK2/3
Serum, cytotoxic drugs, irradiation, heatshock, reactive oxygen species, lipopolysaccharide, and other stress
TNF signaling pathway
CASP
MST1/2
TRADD
TRAF2
GCK
LZK
MUK
MLTK
PP2CA
AKT
ATF-2
Elk-1
p53
p53 signaling pathway
Proliferation, differentiation, inflammation
Apoptosis
Cell cycle
TNF
TNFR
IL1
IL1R
MYD88
FASL
FAS
TGFB
TGFBR
IRAK1/4
DAXX
ASK2
ASK1
MKK3
MKK6
p38
Sap1a
GADD153
MAX
MEF2C
LPS
CD14
TAB1
TAB2
TAK1
TRAF6
ECSIT
PP5
PP2CB
PRAK
MAPKAPK
HSP27
MSK1/2
CREB
Cdc25B
DNA damage
GADD45
MEKK4
TAO
NLK
Wnt signaling pathway
Serum, EGF, reactive oxygen species, or Src tyrosinkinase downstream
MEK5
ERK5
Nur77
MAPKKKK
MAPKKK
MAPKK
MAPK
Transcription factor
Data on KEGG graph
Rendered by Pathview


WNT SIGNALING PATHWAY
-6.4
0.0
6.4
Canonical pathway
Rspo
Lgr4-6
Rnf43
Znrf3
Notum
FRP
Cer-1
Dally
Porc
Wnt
Frizzled
LRP5/6
Wif-1
PEDF
SOST
Dkk
BAMBI
Idax
CKIε
GBP
Dvl
CK2
Axam
Scaffold
GSK-3β
+p
β-catenin
Axin
APC
CKIα
Alzheimer's disease
PS-1
PKA
MAPK signaling pathway
TAK1
+p
NLK
Adherens junction
CBY1
Duplin
CBP
Pontin52
+p
TCF/LEF
DNA
c-myc
c-jun
fra-1
cycD
WISP1
PPARδ
Uterine
Cell cycle
SMAD4
SMAD3
Gro/TLE
CtBP
δ-catenin
ICAT
Xsox17
TGF-β signaling pathway
p53 signaling pathway
p53
Siah-1
SIP
Skp1
TBL1
+p
Phosphorylation dependence
β-TrCP
Skp1
Cul1
Rbx1
Ubiquitin mediated proteolysis
PAR-1
Nkd
Phosphorylation independence
+u
Proteolysis
Planar cell polarity (PCP) pathway
Sfrm
Prickle
Wnt11
Frizzled
INVS
Dvl
Daam1
RhoA
ROCK2
Focal adhesion
Cytoskeletal change
Rac
JNK
Gene transcription
MAPK signaling pathway
Knypek
ROR1/2
RYK
Wnt/ Ca 2+pathway
Wnt5
Frizzled
RYK
G protein?
PLC
Ca2+
CaMKII
CaN
PKC
-p
NFAT
DNA
Data on KEGG graph
Rendered by Pathview


-6.4
0.0
6.4
VEGF SIGNALING PATHWAY
Vascular endothelial cell
Calcium signaling pathway
MAPK signaling pathway
Arachidonic acid metabolism
Focal adhesion
PI3K-Akt signaling pathway
VEGF
VEGFR2
VRAP
PLCγ
Sck
DAG
IP3
Ca2+
PKC
SPK
Ras
Raf-1
MEK
ERK
DNA
Proliferation
cPLA2
CALN
NFAT
COX2
PGI2production
PGI2
FAK
Paxillin
Focal adhesion turnover
Cdc42
p38
MAPKAPK
HSP27
Actin reorganization
Migration
Src
PI3K
PIP3
Rac
eNOS
NO
Permeability
Akt/PKB
Casp9
Bad
Survival
+p
-p
Data on KEGG graph
Rendered by Pathview