REPUBLIC OF TURKEY
ADANA ALPARSLAN TÜRKEŞ SCIENCE AND TECHNOLOGY UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
BIOENGINEERING DEPARTMENT

IDENTIFICATION OF KEY BIOMOLECULES IN ADRENOCORTICAL CANCER PROGRESSION VIA BIOINFORMATICS AND MACHINE LEARNING APPROACHES

AYŞE SAVAŞ ÖZMENOĞLU

M.Sc.

ADANA 2023

REPUBLIC OF TURKEY
ADANA ALPARSLAN TÜRKEŞ SCIENCE AND TECHNOLOGY UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
BIOENGINEERING DEPARTMENT

IDENTIFICATION OF KEY BIOMOLECULES IN ADRENOCORTICAL CANCER PROGRESSION VIA BIOINFORMATICS AND MACHINE LEARNING APPROACHES

AYŞE SAVAŞ ÖZMENOĞLU

M.Sc.

SUPERVISOR
Assoc. Prof. ESRA GÖV

ADANA, 2023

DECLARATION OF CONFORMITY

In this thesis study, which was prepared following the thesis writing rules of Adana Alparslan Türkeş Science and Technology University Institute of Graduate School, I declare that I provide all the information, documents, evaluations and results in accordance with scientific ethics and moral codes without resorting to any means or assistance that would be contrary to scientific ethics and traditions. I also declare that I refer to all of the articles I used in this study with appropriate references and accept all moral and legal consequences if a situation is found contrary to my statement regarding my work.

..../..../20..

[Signature]

Ayşe SAVAŞ ÖZMENOĞLU

ÖZET

ADRENOKORTİKAL KANSER İLERLEMESİNDE ANAHTAR BİYOMOLEKÜLLERİN BİYOİNFORMATİK VE MAKİNE ÖĞRENİMİ YAKLAŞIMLARI İLE TANIMLANMASI

Ayşe SAVAŞ ÖZMENOĞLU

Yüksek Lisans, Biyomühendislik Anabilim Dalı

Danışman: Doç. Dr. Esra GÖV

Ağustos 2023, 80 sayfa

Adrenal Adenomlar (ACA) adrenal bezde oluşan benign kitlelerdir, adrenokortikal karsinom (ACC) ise malign agresif bir tümördür. Bu çalışmada amacımız, adrenokortikal adenomun başlangıcından karsinomun ilerlemesine kadar kritik biyomoleküllerin biyoinformatik ve makine öğrenimi yaklaşımları ile araştırılmasıdır. Hem vaka durumunda pozitif korelasyona hem de kontrol durumunda negatif korelasyona sahip (veya tam tersi) istatistiksel olarak anlamlıSpearman'ın gen korelasyon katsayısını belirlemek için yeni bir analiz yöntemi kullanıldı.. ACA-Normal doku (ACA-N), ACC-Normal doku (ACC-N) ve ACC-ACA için üç RGEC ağı oluşturuldu ve RGEC ağlarının merkez genleri belirlendi. ACA-N grubunda PLA2G4A geni, ACC-N grubunda 14 gen (SAE1, ATRX, PNMT, RYBP, NCAPG2, PDK2, C7, RNF114, PTPRB, NPY1R, PARM1, PGK1, HMMR ve ETC2), ACC-ACA grubunda FMO2 ve UBE2S genleri merkez genler olarak tespit edildi. Elde edilen genlerin tanısal ve prognostik performansı, K Komşu Sınıflandırıcısı, Lojistik Regresyon, MLP Sınıflandırıcısı, Karar Ağacı Sınıflandırıcısı, Rastgele Orman Sınıflandırıcısı, Gradyan Geliştirme Sınıflandırıcısı, Cat Geliştirme Sınıflandırıcısı, LGBM Sınıflandırıcısı ve XGB Sınıflandırıcısını içeren makine öğrenimi (ML) sınıflandırma algoritmaları aracılığıyla test edildi. ACC-N grubu RCEG ağı için 14 merkezi genin, yaşayan ve vefat eden hastaların örneklerinde ayrım yapma performansına kıyasla ACC dokularını normal dokulardan daha iyi bir şekilde ayırdığı bulunmuştur. Bu varsayılan biyomoleküllerin keşfi, iyi huylu bir tümörün kötü huylu bir tümöre dönüşmesini önlemede de önemli bir adım olabilir.

Anahtar Kelimeler: adrenal kanser, biyobelirteçler, makine öğrenimi, gen korelasyonu

ABSTRACT

IDENTIFICATION OF KEY BIOMOLECULES IN ADRENOCORTICAL CANCER PROGRESSION VIA BIOINFORMATICS AND MACHINE LEARNING APPROACHES

Ayşe SAVAŞ ÖZMENOĞLU

M.Sc., Department of Bioengineering

Supervisor: Assoc. Prof. Dr. Esra GÖV

August 2023, 80 pages

Adrenal Adenomas (ACA) are benign masses that form in the adrenal gland, while adrenocortical carcinoma (ACC) is a malignant aggressive tumor. In the present study, our aim is to investigate critical biomolecules from the onset of adrenocortical adenoma to the progression of carcinoma with bioinformatics and machine learning approaches. We will use a new analysis method to identify a statistically significant Spearman's gene correlation coefficient that have both a positive correlation in the case state and a negative correlation in the control state (or vice versa). Three RGEC networks were recontructed for ACA-Normal tissue (ACA-N), ACC-Normal tissue (ACC-N), and ACC-ACA and hub genes of RGEC networks were identified. PLA2G4A gene in ACA-N group, 14 genes (i.e: SAE1, ATRX, PNMT, RYBP, NCAPG2, PDK2, C7, RNF114, PTPRB, NPY1R, PARM1, PGK1, HMMR and ETC2) in ACC-N group, the FMO2 and UBE2S genes in the ACC-ACA group were determined. Diagnostic and prognostic performance of the resultant genes were performed through machine learning (ML) classification algorithms including K Neighbor Classifier, Logistic Regression, MLP Classifier, Decision Tree Classifier, Random Forest Classifier, Gradient Enhancement Classifier, Cat Enhancement Classifier, LGBM Classifier, and XGB Classifier. It was found that 14 central genes of RCEG network for ACC-N group efficiently discriminate the ACC tissues from the normal tissues compared to the performance of discriminating in the live and dead specimens. The discovery of these putative biomolecules may also be an important step in preventing a benign tumor from turning into a malignant one.

Keywords: adrenal cancer, biomarkers, machine learning, gene correlation

Dedication

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CRH	: Corticotropin-releasing hormone
ACTH	: Adrenocorticotropic hormone
CT	: Computed Tomography
MRI	: Magnetic Resonance
ACA	: Adrenocortical Adenoma
ACC	: Adrenocortical Carcinoma
CS	: Cushing Syndrome
MEN	: Multiple Endocrine Neoplasia
MLT	: Machine Learning Techniques
DEG	: Differentially Expressed Gene
GEO	: Gene Expression Omnibus
RMA	: Robust Multi-Array Average
FC	: Fold Change
DAVID	: Database for Annotation, Visualization, and Integrated Discovery
RGEC	: Reverse Gene Expression Correlation
KEGG	: Kyoto Encyclopedia of Genes and Genomes
PCC	: Pearson Correlation Coefficient
SCC	: Spearman Correlation Coefficient
TCGA	: The Cancer Genome Atlas
KNN	: K-nearest Neighbor
MLP	: Multilayer Perceptrons
LGBM	: Light Gradient Boosting Machine

1. INTRODUCTION

The adrenal glands are important endocrine glands responsible for the production of steroid hormones, which are of vital importance in the body (Özgül et al., 2021). These glands, located on the upper part of both kidneys, consist of two parts called the cortex and medulla, which are quite different from each other (REMINISCENT, 2016). The adrenal glands, which are part of the endocrine system, work in an organized manner with the hypothalamus and pituitary gland. The main task of these glands is to produce various hormones in order to protect the body from dangers and control the physiological behaviors necessary for survival in cases where homeostasis cannot be maintained in mammals (Heck & Handa, 2019).

The hypothalamus-pituitary-adrenal axis (HPA axis) signaling pathway is responsible for the hormones produced in the cortex layer (Minami et al., 2019). Corticotropin-releasing hormone (CRH), which is activated in the hypothalamus, causes the production of adrenocorticotropic hormone (ACTH) from the pituitary gland. As a result, glucocorticoids, mineralocorticoids, and androgen hormones are secreted in the cortex. On the other hand, preganglionic sympathetic nerves play an active role in the production of hormones produced in the medulla, and catecholamines are secreted here (REMINISCENT, 2016).

Adrenal tumors occur as a result of functional disorders that occur during the secretion of these hormones (Miller and White, 2022). The symptoms that occur as a result of these disorders; diabetes mellitus, and high blood pressure, such as non-hormone-induced symptoms, the diagnosis of adrenal cancer is usually determined by computed tomography (CT) and magnetic resonance (MRI) taken during routine health checks in the interval 3%-8% of patients (Bourdeau et al., 2018; Hindman and Megibow, 2020).

To classify tumors, these imaging methods are used. Adrenal lesions can be determined to be benign or malignant by using factors such as tumor size, where the tumor is located, mitosis rate, tumor cell proliferation ability, and necrosis (Giordano et al., 2009). However, the incidence of these physiopathological developing cancers is too high to be underestimated by 3% -10% (Zennaro et al., 2017; Chehade et al., 2020).

There are three main reasons why ACC patients are admitted to the hospital the most. The first of these reasons is the symptoms that occur due to an increase in hormones in the body, another reason is symptoms such as abdominal pain, and bloating caused by the formation of a local tumor; finally, tumor development in ACC patients occurs accidentally during routine health checks (Else et al., 2014a).

In hormone-producing ACCs, symptoms may be due to the overproduction of cortisol and androgen hormones. Excessive cortisol production, that is, hypercortisolism, causes diabetes mellitus, osteoporosis, muscle weakness, and muscle atrophy (Thampi et al., 2016). In addition, an excess of the hormone cortisol causes the HSD11B12 system, a

glucocorticoid-activating enzyme, to be saturated, which leads to complaints of hypokalemia and hypertension as a result of glucocorticoid-mediated mineralocorticoid activation (Else et al., 2014b). Hypercortisolism is a clinical finding usually encountered in ACC patients (Berruti et al., 2014). This condition manifests itself with weakness in the muscles in patients.

As a result of overproduction of the androgen hormone, another adrenal gland hormone, symptoms such as menstrual irregularities, thickening of the voice, male pattern baldness, and hirsutism occur in women (Wang & Rainey, 2012). In men, hyperandrogenism is usually not observed, in addition, IGF-2-mediated hypoglycemia is commonly observed (Karimi et al., 2019).

The main aim of this thesis is to investigate critical biomolecules from the onset of ACA to the progression of ACC with bioinformatics and ML approaches. We will use a new analysis pipeline to determine a statistically significant inverse correlation of differentially expressed genes (DEGs) in two conditions including carcinoma, and adenoma. Validation of the resultant putative genes identified with opposite correlations will be performed through ML classification algorithms including K Neighbor Classifier, Logistic Regression, MLP Classifier, Decision Tree Classifier, Random Forest Classifier, Gradient Enhancement Classifier, Cat Enhancement Classifier, LGBM Classifier, and XGB Classifier. Putative key genes can provide important information about the transition from benign to malignant adrenocortical tumors. The discovery of these putative biomolecules may also be an important step in preventing a benign tumor from turning into a malignant one.

2. THEORETICAL FOUNDATIONS AND LITERATURE REWIEV

2.1. BENIGN TUMORS OF THE ADRENAL GLAND

Masses of 1 cm and above occurring in the adrenal gland are considered tumors (Belmihoub et al., 2017). The most common of these adrenal tumors are adrenocortical adenomas (ACA), most of which are found as benign masses (Zennaro et al., 2017). These ACAs, which are neoplasms of the adrenal cortex, are divided into functional and non-functional (functional-nonfunctional) according to their hormone secretion abilities (Wang et al., 2018; Mahmood and Anastasopoulou, 2021). It is very difficult to diagnose because no hormones are secreted when non-functional ACAs occur.

Adrenocortical Adenomas

2.1.1. Adrenal Incidentalomas

Adrenal tumors that occur randomly as a result of the increase in imaging techniques such as CT, MRI are called incidentalomas (Sherlock et al., 2020). Although the incidence was less in previous years, current studies show an increase in the incidence by 7.3% (Hanna et al., 2022). It has been found that the prevalence increases with age, and in addition, there is no significant difference between the incidence in men and women (Bancos, 2022). Adrenal incidentalomas occur during imaging during routine checkups because they do not show any specific symptoms. With the examination conducted after the discovery of the tumor, the findings of the disease are confirmed.

2.1.2. Cushing Syndrome (CS)

Cushing's syndrome occurs as a result of the hormone cortisol produced in excess amounts from the pituitary gland (Lacroix et al., 2015). In addition, it has been found that long-term exposure to glucocorticoids also triggers this disease (Pivonello et al., 2016). Approximately 15-20% of cases are caused by the adrenal gland (Lonser et al., 2017). Of these, 75% are ACAs and 25% are ACCs (Feelders et al., 2019). CS is divided into two types according to the symptoms it shows as obvious CS or subclinical CS.

2.1.2.1. Obvious cushing's syndrome

In obvious Cushing's syndrome, patients are usually admitted to the hospital with non-specific complaints such as weakness, weight gain, and depression (Eker, 2019). In addition, symptoms such as obesity, buffalo hump, excessive body hair, hypertension, and muscle weakening are often observed (Nieman, 2015). CS can be ACTH-dependent or independent of ACTH regulation. ACTH-dependent CS occurs with excessive ACTH release from the pituitary gland, malignant pancreatic tumor, small cell lung cancer, or excessive ACTH release from a tumor outside the pituitary (Reincke et al., 2015).

2.1.2.2. Subclinical Cushing's Syndrome

Although obvious CS is more common in adrenal incidentalomas, subclinical CS is also observed. In contrast to obvious CS, despite the increased amount of cortisol in the blood, there are no obvious symptoms such as buffalo hump, or obesity in subclinical CS (Iacobone et al., 2015). According to the hormonal tests performed, results such as low ACTH value in patients, lack of suppression of cortisol secretion, and upper limit of cortisol excreted in the urine in the last 24 hours indicate subclinical CS (Iwamoto et al., 2022).

2.1.3. Conn Syndrome (Hyperaldosteronism)

It occurs as a result of excessive aldosterone release from the adrenal cortex (Vaidya & Carey, 2020). Hypertension is accompanied by hypokalemia and hyporeninemia (Nagaraja et al., 2015). 50-70% of Conn Syndrome is caused by ACA and ACC (Araujo-Castro & Parra-Ramirez, 2022). The remaining part is formed by hyperplasia of the adrenal gland.

With the imaging methods such as CT and MRI used with the developing technology, ACA and hyperplasia can be distinguished (Dominguez et al., 2018). These aldosterone-producing masses are usually smaller than 2 cm and have a lipid load, low density, and their shape is oval (Papadopoulou-Marketou et al., 2015).

2.1.4. Pheochromocytoma

It is a tumor that accounts for 6.5% of adrenal incidences (Nguyen-Martin & Hammer, 2006). Chromaffin comes from cells in the sympathetic nervous system and produces catecholamine harmony (Eisenhofer et al., 2001). Although most adrenal tumors are produced in the adrenal cortex, a pheochromocytoma is a tumor of the adrenal medulla (Lenders & Eisenhofer, 2011). It has been found to be associated with diseases such as multiple endocrine neoplasia (MEN), Carney syndrome, von Hippel-Lindau, paraganglioma, and hypertension (Manger & Gifford Jr, 1993). It is known that less than 10% of tumors show malignancy (Harari & Inabnet III, 2011). In addition, the importance of imaging methods in accurately understanding malignancy is quite great. Tumors larger than 6 cm that invade or metastasize to nearby tissues or organs are tumors with malignant properties (Toniato et al., 2007).

2.1.5. Adrenocortical Carcinoma (ACC)

Another adrenal cancer is adrenocortical carcinoma (ACC), which is much more aggressive and fatal unlike ACA, although it is very rare with an incidence of 0.7-2.0 per million, the majority of which show malignancy (Zheng et al., 2016;Crona and Beuschlein, 2019). The 5-year survival rate after being diagnosed with ACC is less than 15% (Subramanian & Cohen, 2021). Although the age for the occurrence of the disease is not specified, it has been found that it is more common in early childhood and middle age groups (Datta and Roses, 2016). Considering the concept of gender, it has been found that it is more common in women than in men (Altieri et al., 2020).

2.2. FOLLOW-UP AND TREATMENT OF ADRENAL TUMORS

Although research is still ongoing in the treatment of ACA for patients, methods involving surgical and non-surgical supportive therapies are preferred by the majority (Lee and Jin, 2019). Although conventional surgery is not preferred in ACAS, minimally invasive surgery is the most commonly used method (Aydin et al., 2019).

In the treatment of ACC, treatment options are also limited because the disease is very rare (Altieri et al., 2020). In addition to the most preferred treatment methods such as surgery, chemotherapy, and radiation therapy, mitotane therapy is also used in metastatic tumors (Catalano et al., 2021). Since a powerful chemotherapeutic agent specific to the disease has not yet been found, chemotherapy treatment is usually used in combination with mitotane therapy in cases where surgical treatment cannot be applied (Board, 2019). Considering the rate of spread of recurrence and the fear of operative intervention caused by surgery in the

patient, it seems that this treatment method is not strong enough and its effect on survival is uncertain (Gaujoux et al., 2017).

The difficulty of diagnosis in ACA and ACC and the fact that the relationship of the findings related to the course of the disease to the disease is unknown are the biggest reasons why the disease cannot be prevented. This situation leads researchers to diagnostic studies. Although the imaging methods used at the diagnostic stage provide information about the benign-malignant nature of the tumor, determining the degree of the tumor, or the progression of cancer, they are quite inadequate considering the difficulty of application and the time to get results (D'amuri et al., 2019). Artificial intelligence and MLT, which have been frequently announced recently, also reveal quite practical and meaningful characteristic results in this regard.

2.3. MACHINE LEARNING TECHNIQUES

Machine learning is the programming of computers using the data taught to the computer and past knowledge and experiences in order to bring a certain criterion to the best level (Larranaga et al., 2006). This programming is based on the theory of making inferences from existing data with many algorithms and providing information about new data (Sevli and Sunday, 2020). Bioinformatics is a multi-faceted discipline that analyzes omic data using (Abdurakhmonov, 2016). The increasing use of omic data day by day and the formation of a big data mess once again demonstrates the need for bioinformatics methods. Although bioinformatics methods provide fast and accurate results through calculation and analysis, it needs faster and more efficient methods for storing big data that arise with developing technology and extracting the necessary information from big data (Shastry and Sanjay, 2020). Therefore, the integration of bioinformatics techniques with MLT will benefit from many aspects.

Because ACC is a malignant tumor that is difficult to diagnose, has a poor prognosis, and has a low survival rate, it has been frequently included in the literature. To determine new treatment methods for the disease, it is necessary to have a good understanding of the pathogenesis of the disease. Xing et al reported that fourteen candidate hub genes involved in the formation and development of the ACC by bioinformatics and MLT (Xing et al. 2019). In another study, a model for the prognosis of the disease was performed with MLT, which can efficiently process complex data, and the model successfully predicted the survival status of patients with ACC disease (Tang et al. 2023). Limited studies also reported that MLT-based analyses were performed to identify prognostic genes of ACC (Yi et al. 2022; Yan et al 2021). In addition to the studies in which biomarkers predicted for high survival rates of patients have been identified, there are also studies aimed at the identification of the main key genes in the tumor microenvironment (Guo et al. 2020; Li et al. 2020). Although several biomarkers in ACC have been identified, there are no studies that identify biomarkers based on gene correlation in cancer progression.

2.3.1. K-nearest Neighbor (KNN) Classifier

The KNN classifier is one of the classifiers that are common in ML algorithms. The basic principle of the classification tool is to measure the similarity or distance between the tested data and the training data (Abu Alfeilat et al., 2019). Thanks to the ease and efficiency of applying the classifier, it has been considered one of the top 10 algorithms used in data mining (Gou et al., 2019). Despite this, there are also negative aspects of the KNN classifier. The model makes a binary calculation including similarity or distance parameters by selecting the closest k sample for all training data (Maillo et al., 2017). Unlike other classifiers, KNN contains only two parameters; the distance parameter to measure the distance between the training data and the test data range, and the number of nearest neighbors parameter K (Zhang et al., 2019).

According to literature, it has been seen that the classifier greatly benefits researchers in the field of health. For example, in a study involving the investigation of many types of cancer (i.e: pancreatic cancer, ovarian cancer, and pancreatic cancer) through the nearest neighbor classifier and classification tree, the expression levels of five protein biomarkers were compared (Banaei et al., 2019).

In another study, nine central genes associated with the prognosis of ACC patients were obtained, and MLT algorithms, including KNN, were used to determine biomarkers with the highest survival accuracy (Yi et al., 2022).

Marquardt et al.was aimed to classify ACC disease according to the expression model.The data obtained from the TCGA data bank and used KNN and Random Forest algorithms presented a successful classification (Marquardt et al., 2021).

2.3.2. Logistic Regression Classification

Logistic regression classification is an algorithm that helps to understand the relationship between the probability of one of two different variables and the intrinsic properties of the other variable (Bisong & Bisong, 2019). The difference between logistic regression and linear regression is that a linear function is an infinite data pool, while logistic regression extracts the parameters of the data with the maximum probability of occurrence from the pool (Das, 2021).

Logistic regression has many advantages as well as disadvantages and is an algorithm under development. For example, training big data in big data analytics takes time, and the error rate increases as the data rate increases (Zou et al., 2019). Therefore, logistic regression is not preferred in such applications. In addition, it provides strong prediction performance about the data distribution they are trained in the analyses performed using the appropriate data set (Khosravi et al., 2019). The algorithm comes across in the field of health, as in many areas. For example, a classification can be made to predict the prognosis of a type of cancer. Thus,

the course of the disease and the chances of survival can be predicted (Campbell & Donner, 1989).

A study conducted in 2021 suggests that logistic regression algorithms are used to classify the expression levels of genes in cancer, but they will negatively affect the classification performance of genes unrelated to the target disease. (Zhou et al., 2021).

In a study conducted on adrenal incidentalomas, a hormonal hypersecretion prediction model was put forward based on the components of high metabolic syndrome of 125 patients. The results of multivariate discriminant analysis were compared with the results of traditional logistic regression analysis (Tsentidis et al., 2019).

Studies have also been conducted for ACC, where the frequency of research is higher compared to ACAS. Based on the multivariate logistic regression results of 751 patients who were examined taking into account parameters such as gender, race, and age, important inferences were made in predicting overall survival and disease survival (Li et al., 2018).

2.3.3. Multilayer Perceptrons (MLP) Classifier

MLP is a multi-layered supervised learning technic that is most commonly used in classification algorithms (Ramchoun et al., 2016). There are various activation functions in the input layer that allow the model to learn the data. The hidden layer is not exposed to direct input (Alnuaim et al., 2022). As the data in this layer increases, the depth of the layer also increases.The last layer is named as the output layer. From here, the necessary vector for the problem is obtained (Lin et al., 2022).

As in many fields, the algorithm plays an important role in cancer research. In order for personalized treatment to be applied, cancer subtypes must be well known. Although studies have been conducted in the area, it has been aimed to classify new cancer subtypes by using high-efficiency MLT methods (Luo et al., 2023).

Studies have recently focused on breast cancer, which is considered to contain a lot of data. In one study, MLP and Support Vector Machine (SVM) algorithms were used to diagnose breast cancer and then evaluate its performance (Ghosh et al., 2014).

2.3.4. Decision Tree Classifier

A decision tree is a flowchart-like classification algorithm used to determine the highest probability required to achieve a goal (Charbuty & Abdulazeez, 2021). This flow, starting from the root and extending to the leaves, is considered the rule in classification (Tanha et al., 2017). It is mostly used in the analysis of very complex events.

Although there are many methods for the diagnosis of adrenal tumors, they are not sufficient. For this purpose, in a study, adrenal steroids were examined in the blood sample taken from patients, and classification algorithm was made, obtaining useful results both in terms of diagnosis and accurate determination of cancer subtypes (Ku et al., 2021).

Similarly, in a study, the 24-hour urinary steroid profiles of patients were examined, and it was aimed to distinguish between ACA and ACC diseases using the decision tree model (Vogg et al., 2023).

2.3.5. Random Forest (RF) Classifier

RF is an algorithm that can perform analyses such as classification and regression, create a decision tree during learning, and then make numerical and class predictions (Qi, 2012).

In a study conducted on ACA and ACC data based on the algorithm, ACA and ACC samples obtained as a result of CT scanning were compared with a series of MLTs, including the RF algorithm, and as a result, it was found that the discrimination of benign and malignant tumors obtained by MLT results in 82% accuracy (Elmohr et al., 2019).

2.3.6. Light Gradient Boosting Machine (LGBM) Classifier

LGBM is an algorithm based on a tree-leaf mechanism, just like the Random Forest algorithm. The leaf with the most important feature is selected for the purpose of tree growth (Taha & Malebary, 2020). Thus, the dimensionality of the training data is reduced. LGBM is designed to be able to process big data quickly (Alzamzami et al., 2020).

It uses a sampling algorithm (Gradient-based One-Side Sampling (GOSS)) to determine the importance of training data and improve training. GOSS stores data with a large gradient, while random samples data with a small gradient (F. Li et al., 2018).

2.3.7. XGB Classifier

XGB is a histogram-based algorithm used in classification and regression problems (Alzamzami et al., 2020). Compared with RF, the ranking is different in terms of combining growth and decoupling data (AlThuwaynee et al., 2021).

In a study conducted, the cost and inefficiency of detecting DNA binding proteins experimentally were mentioned, and it was suggested that predicting these proteins with the help of classifier algorithms will give much more effective and accurate results (Q. Zhang et al., 2021).

3. MATERIALS AND METHODS

3.1 Research of High-throughput gene expression datasets

Microarray datasets related to ACC and ACA were comprehensively scanned in order to determine gene expression profiles. In this search, we focused on datasets containing healthy adenoma or carcinoma taken from the adrenal cortex or gland tissues. Comparative gene expression analyses were performed between three groups: ACA-Normal tissue (ACA-N), ACC-Normal tissue (ACC-N), and ACC-ACA. In the Gene Expression Omnibus (GEO) database (Barrett et al. 2012) as a result of the research, a high-efficiency data GSE10927 (Giordano et al. 2009), GSE12368 (Soon et al. 2009), GSE19750 (Demeure et al. 2013), and GSE75415 (West et al. 2007) datasets were obtained (Table 1). Datasets including Affymetrix Human Genome U133+ 2.0 Sequence and Affymetrix Human Genome U133A Sequence platforms were selected. Different microarray platforms have not been used since the common differentially expressed genes (DEGs) identified from the datasets will be used.

Table 3.1: Datasets used in the study and analysis design for differentially expressed genes.

ACCESSION ID OF DATASETS	SAMPLE DESCRIPTION	SAMPLE NUMBER	PLATFORM	REFERENCES
GSE10927-ACA-N	22 Adrenocortical Adenoma and 10 Normal Adrenal Cortex	22 Tumor-10 Normal tissue	Affymetrix Human Genome U133 Plus 2.0 Array	Giordano et al., 2009
GSE12368-ACA-N	16 Adrenocortical Adenoma and 6 Normal Adrenal Cortex	16 Tumor-6 Normal tissue	Affymetrix Human Genome U133 Plus 2.0 Array	Soon et al., 2009
GSE75415-ACA-N	5 Adrenocortical Adenoma and 7 Normal Adrenal Cortex	5 Tumor-7 Normal tissue	Affymetrix Human Genome U133A Array	West et al., 2007
GSE10927-ACC-N	33 Adrenocortical Carcinoma and 10 Normal Adrenal Cortex	33Tumor-10 Normal tissue	Affymetrix Human Genome U133 Plus 2.0 Array	Giordano et al., 2009
GSE12368-ACC-N	12 Adrenocortical Carcinoma and 6 Normal Adrenal Cortex	12 Tumor-6 Normal tissue	Affymetrix Human Genome U133 Plus 2.0 Array	Soon et al., 2009
GSE19750-ACC-N	44 Adrenocortical Carcinoma and 4 Normal Adrenal Glands	44 Tumor-4 Normal tissue	Affymetrix Human Genome U133 Plus 2.0 Array	Demeure et al., 2013
GSE75415-ACC-N	18 Adrenocortical Carcinoma and 7 Normal Adrenal Cortex	18 Tumor-7 Normal tissue	Affymetrix Human Genome U133A Array	West et al., 2007
GSE10927-ACC-ACA	33 Adrenocortical carcinomas and 22 Adrenocortical Adenomas	33 Tumor-22 Normal tissue	Affymetrix Human Genome U133 Plus 2.0 Array	Giordano et al., 2009
GSE12368-ACC-ACA	12 Adrenocortical carcinomas and 16 Adrenocortical Adenomas	12 Tumor-16 Normal tissue	Affymetrix Human Genome U133 Plus 2.0 Array	Soon et al., 2009
GSE75415-ACC-ACA	18 Adrenocortical Carcinoma and 5 Adrenocortical Adenoma	18 Tumor-5 Normal tissue	Affymetrix Human Genome U133A Array	West et al., 2007

3.2 Determination of Differentially Expressed genes (DEGs)

Each dataset was analyzed for three groups ACA and ACC comparing with normal tissues and ACC comparing with ACA to determine the genes which are differentially expressed (DEG). Raw data were normalized through Robust Multi-Array Average (RMA) (Bolstad et al. 2003) in the "affy" package of the R/Bioconductor studio software (R-Studio 64-Bit version) (Gautier et al. 2004). The LIMMA method (Smyth et al. 2005) was used to identify DEGs. In order to obtain statistically significant DEGs, genes with a p-value < 0.05 and a fold change (FC) > 1.5 and <0.67 were considered. The Jvenn web tool (Bardou et al. 2014) was used to determine the common DEGs for each three groups.

3.3 Biological insights of determined differentially expressed genes

Gene enrichment analysis was performed to determine the biological significance of the identified DEGs. For the analyses, "Database for Annotation, Visualization, and Integrated Discovery" (DAVID) (Sherman and Lempicki, 2009) and Metascape bioinformatics platforms (Zhou et al. 2019) were used.

As a result of the analyses, enrichment analysis was performed separately for the dataset in each group. In addition, biological processes, molecular functions, cellular components, signaling pathways, and protein-protein interactions of common DEGs were identified for all three groups. When the P value was <0.01, the results of the gene enrichment analysis were considered statistically significant.

3.4 Establishment of reverse gene expression correlation (RGEC) networks

For the reverse correlation network analysis, the correlation score critic (c-critic) values were calculated using the Pearson correlation coefficient/spearman correlation coefficient (PCC/SCC) values of the genes involved in the correlation network. The calculation is expressed as:

$$c_{critic} = mean\ SCC + 1.96 * std\ of\ SCC$$

SCC refers to the correlation coefficient of gene pairs in case samples and control samples. Each c-critic was calculated separately for case and control samples.

Binary genes that have both a positive correlation in the case state and a negative correlation in the control state (or vice versa) were selected. The selected genes were named reversely correlated genes. Afterward, reverse gene expression correlation (RGEC) networks were reconstructed for three groups by using reversely correlated genes. The RGEC networks which were visualized using Cytoscape (ver. 3.6) (Smoot et al., 2011), was represented as an undirected graph, where nodes represented reversely correlated genes and edges represented reverse correlation in the control vs normal state among gene expression. Hub genes which are central genes were identified according to degree and betweennes centrality score (Erceylan et al., 2021).

3.5 Performance evaluation of the RGEC genes with classification algorithms

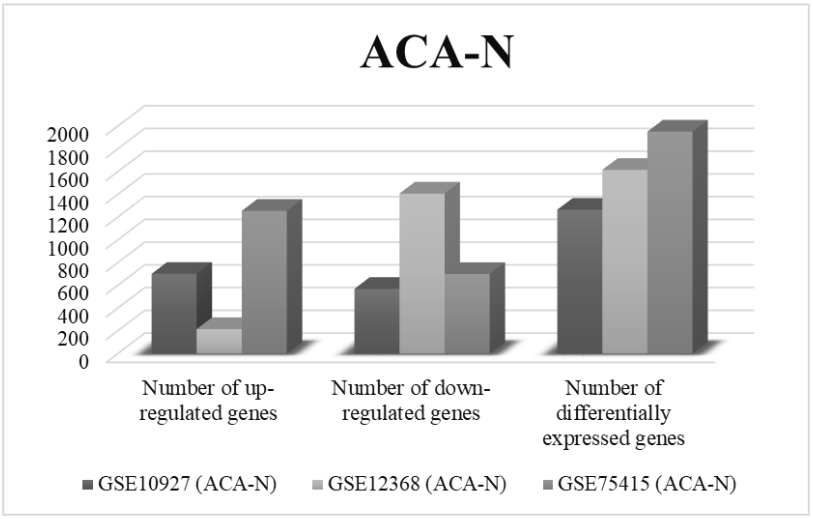
To evaluate the prognostic and diagnostic performance of the RGEC, we obtained gene expressions from ACC samples with clinical information in the TCGA (https://portal. gdc.cancer.gov/) and GEO. To determine the prognostic performance, z-normalization was applied to the RNA-Seq datasets of ACC (n = 78) including 50 alive and 28 dead samples obtained from TCGA. To evaluate the diagnostic performance, z-normalization was applied to the data obtained from RNA-Seq datasets of ACC (n = 78) in TCGA for ACC samples and 27 adrenal cortex samples obtained from GSE90713 (Weiss et al. 2017) ,GSE143383 (Fojo et al. 2020), GSE33371 (Heaton et al. 2012) and GSE75415 (West et al. 2007) in the GEO database. To better improve the interpretation of significant candidates, various classification techniques were used to identify potential new biomarker candidates. We implemented different classification algorithms, including K-Neighbors, MLP, Decision Tree, Random Forest, Gradient Boosting, CatBoost, LGBM, and XGB using the Python programming language (Schapire, 2015). The performance of these techniques was estimated based on the predictive F1, recall, precision and accuracy score of the classifiers.

4. RESULTS

4.1 Gene expression profiles of adrenocortical cancer progression

In the study, we compared the gene expression levels of ACA, ACC and normal tissues obtained from the adrenal gland and cortex layers. We obtained 701 up-regulated , 569

down-regulated 1265 DEGs, respectively, from the GSE10927 dataset analyzed in the ACA-N group; 215 up-regulated, 1407 down-regulated, 1615 DEGs, respectively, from the GSE12368 dataset; and 1255 up-regulated, 699 down-regulated, 1952 DEGs, respectively, from the GSE75415 dataset (Figure 4.1). From these three data sets, we determined the common 110 DEG (Fig.4.2).



ID of Datasets	Number of up-regulated genes	Number of down-regulated genes	Number of differentially expressed genes
GSE10927 (ACA-N)	701	569	1265
GSE12368 (ACA-N)	215	1407	1615
GSE75415 (ACA-N)	1255	699	1952

Figure 4.1. Column graphs showing the numbers of up-down regulated DEGs in the ACA-N.

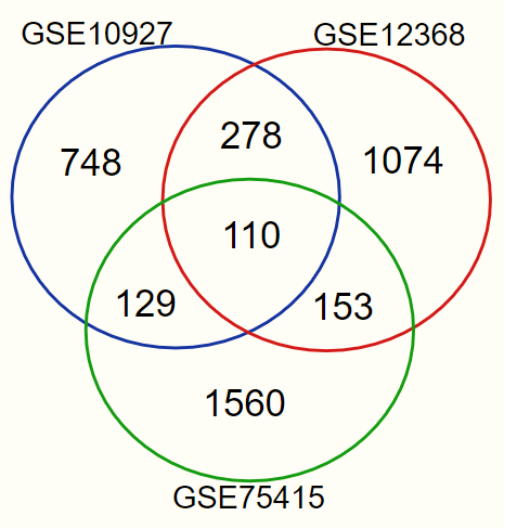
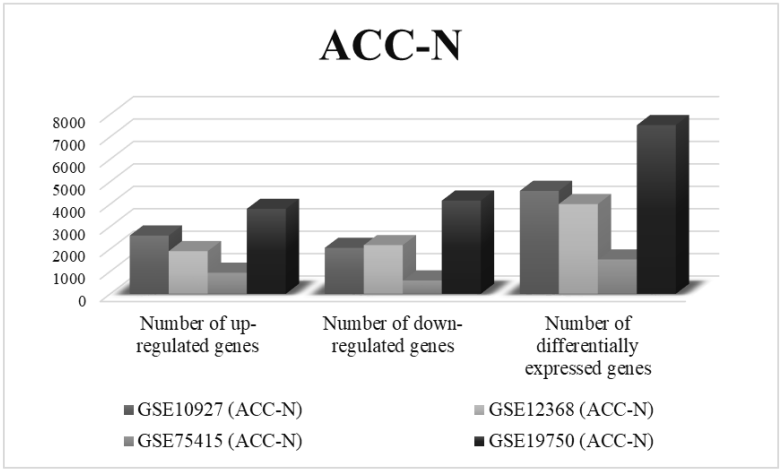


Figure 4.2. Venn Diagram showing common DEGs in ACA-N group.

Our analysis of the ACC-N group consisting of GSE10927, GSE12368 and GSE75415, GSE19750 datasets showed 2598 up-regulated, 2048 down-regulated, 4589 DEGs in GSE10927, respectively; GSE12368 showed 1902 up-regulated, 2170 down-regulated, 4000 DEGs; GSE75415 showed 939 up-regulated, 593 down-regulated, 1531 DEGs; and GSE19750 showed 3785 up-regulated, 4154 down-regulated, 7520 DEGs (Figure 4.3). From these four data sets, we have identified 454 DEGs that are common (Figure 4.4).



ID of Datasets	Number of up-regulated genes	Number of down-regulated genes	Number of differentially expressed genes
GSE10927 (ACC-N)	2598	2048	4589
GSE12368 (ACC-N)	1902	2170	4000
GSE75415 (ACC-N)	939	593	1531
GSE19750 (ACC-N)	3785	4154	7520

Figure 4.3 Column graphs showing the numbers of up-down regulated ACC-N genes and DEGs .

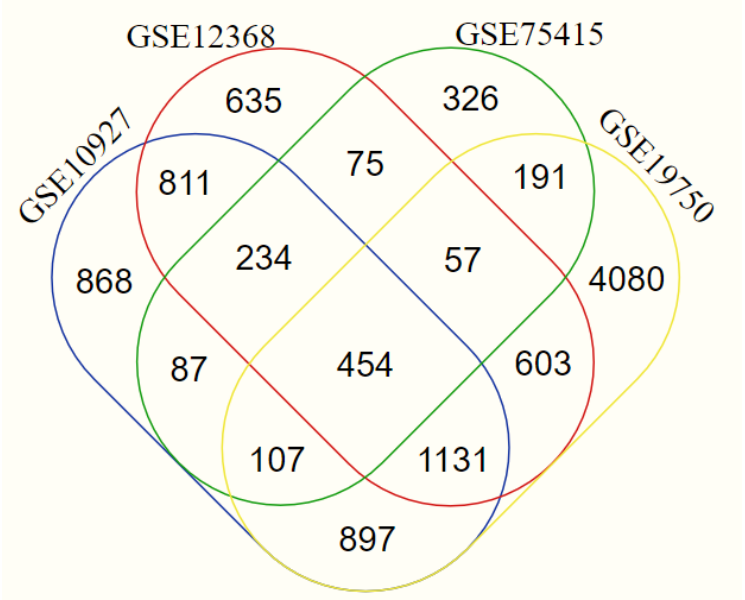
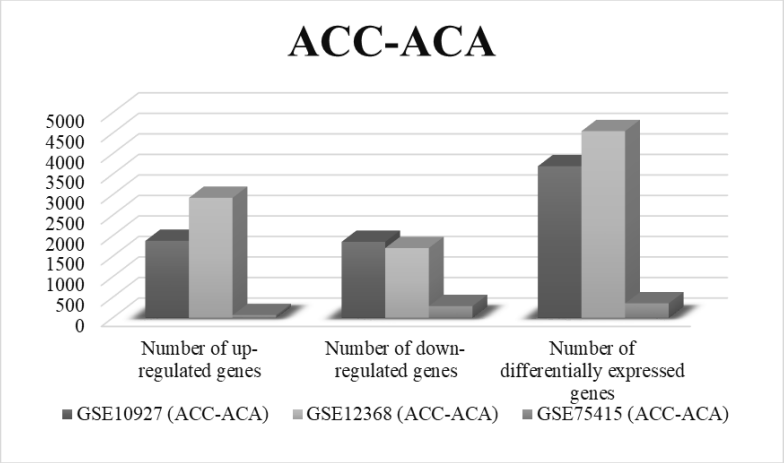


Figure 4.4 Venn diagram represents the number of common DEGs in the ACC-N.

Finally, we identified 1877 up-regulated, 1853 down-regulated, 3695 DEGs, respectively, from the GSE10927 dataset analyzed in the ACC-ACA group; we identified 2925 up-regulated, 1700 down-regulated, 4550 DEGs, respectively, from the GSE12368 dataset;

and we identified 67 up-regulated, 287 down-regulated, 354 DEGs, respectively, from the GSE75415 dataset (Figure 4.5). We have identified 128 DEGs in three datasets (Figure 4.6).



ID of Datasets	Number of up-regulated genes	Number of down-regulated genes	Number of differentially expressed genes
GSE10927 (ACC-ACA)	1877	1853	3695
GSE12368 (ACC-ACA)	2925	1700	4550
GSE75415 (ACC-ACA)	67	287	354

Figure 4.5 Column graphs showing the numbers of up-down regulated DEGs in ACC-ACA group.

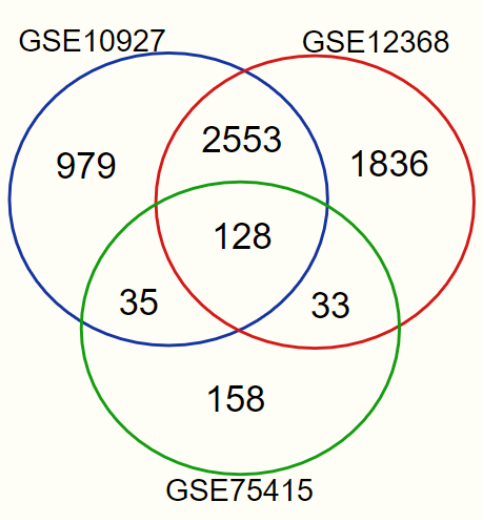


Figure 4.6 Venn diagram represents the number of common DEGs in ACC-ACA groups.

The data obtained from the enrichment analyses performed for each dataset have been tabulated. The genes in the GSE10927 dataset belonging to the ACA-N group have been found to be enriched with diseases such as Rheumatoid arthritis, Intestinal immune network

for IgA production, Type I diabetes mellitus, and biological pathways (Table A1). In addition, these genes are enriched in processes such as the NABA core matrix, vascular development, regulation of cell adhesion (Figure 4.7).

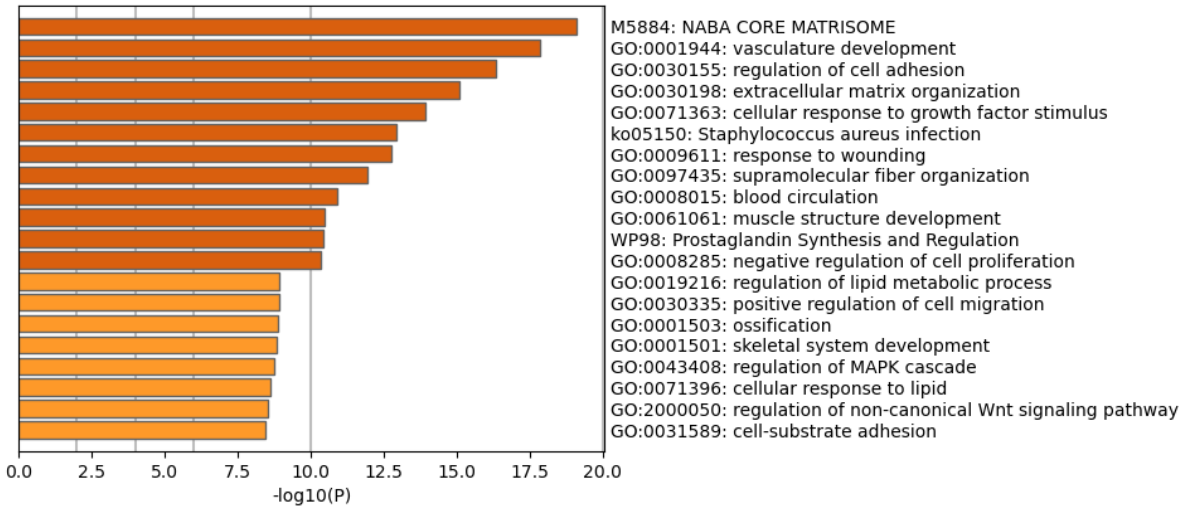


Figure 4.7 The results of the enrichment analysis of the GSE10927 dataset in the ACA-N group.

In the GSE12368 dataset, another ACA-N group data set, genes were associated with signal transmission mechanisms such as Neuronal System, Signal Transduction, and Transmission across Chemical Synapses (Table A2). Moreover, these genes are enriched with signaling pathways such as trans-synaptic signaling, regulation of cell projection organization and behavior (Figure 4.8).

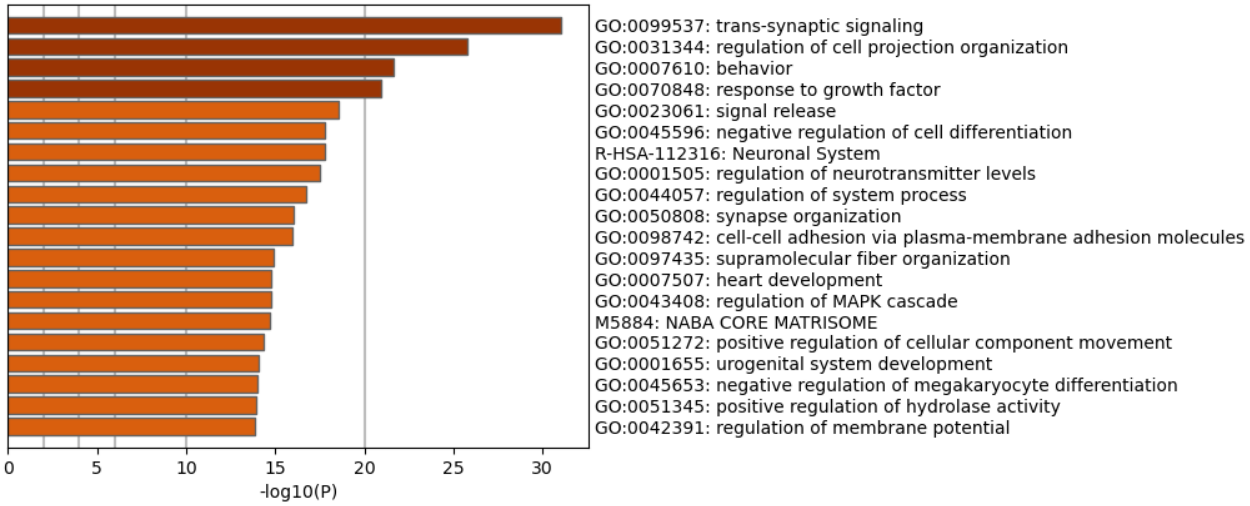


Figure 4.8 The results of the enrichment analysis of the GSE12368 dataset in the ACA-N group.

GSE75415, the last data set of the ACA-N group, showed that genes are enriched mostly in molecular mechanisms, for example, Diseases of signal transduction by growth factor receptors and second messengers, Downstream TCR signaling, Degradation of beta-catenin by the destruction complex (Table A3). Moreover, these genes are enriched with cellular responses to external stimuli, diseases of signal transduction by growth factor receptors and second messengers, diseases and receptors such as vesicle mediated transport (Figure 4.9).

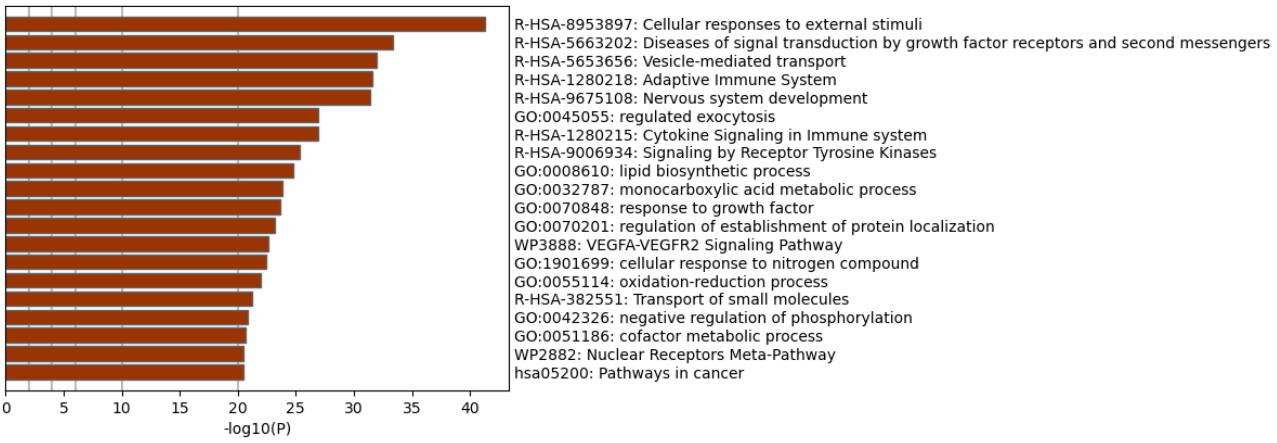


Figure 4.9 The results of the enrichment analysis of the GSE75415 dataset in the ACA-N group.

Because ACC tumors are more aggressive and malignant than ACA, it has been observed that the genes in the ACC-N group are more enriched with biological pathways associated with cancer. For example, the genes in the GSE10927 dataset in the ACC-N group are enriched with cancer-related processes such as cell cycle, mitotic cell cycle, Mitotic Prometaphase, Cell Division, and Regulation of cell cycle processes. (Table A4), (Figure 4.10).

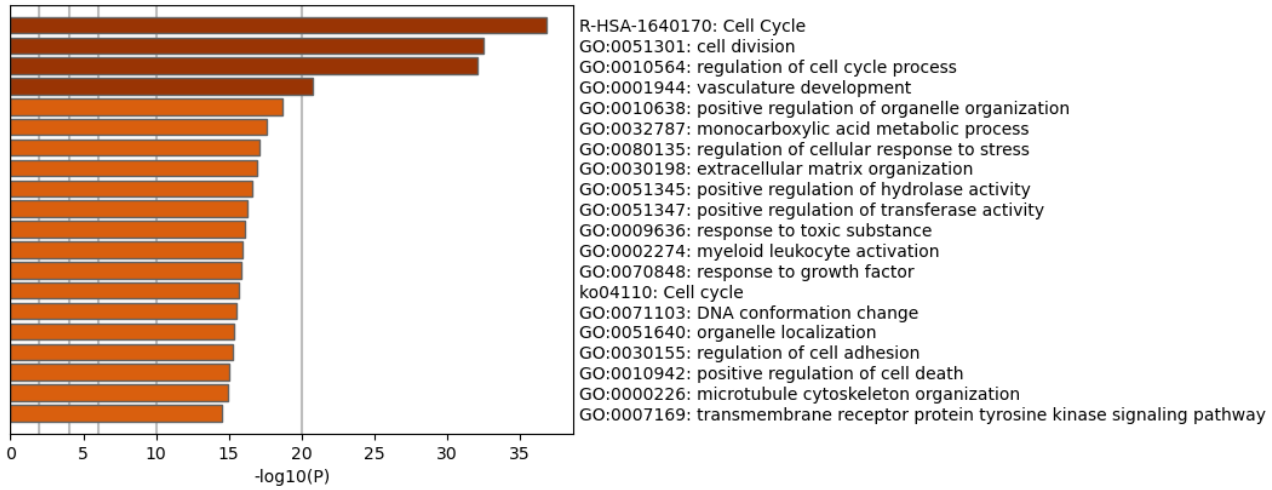


Figure 4.10 The results of the enrichment analysis of the GSE10927 dataset in the ACC-N group.

Similarly, as a result of the enrichment analysis with the genes in the GSE12368 dataset in the ACC-N group, these genes were associated with processes such as Activation of the pre-replicative complex, cell cycle, cell division, cell cycle phase transition and DNA strand elongation (Table A5), (Figure 4.11).

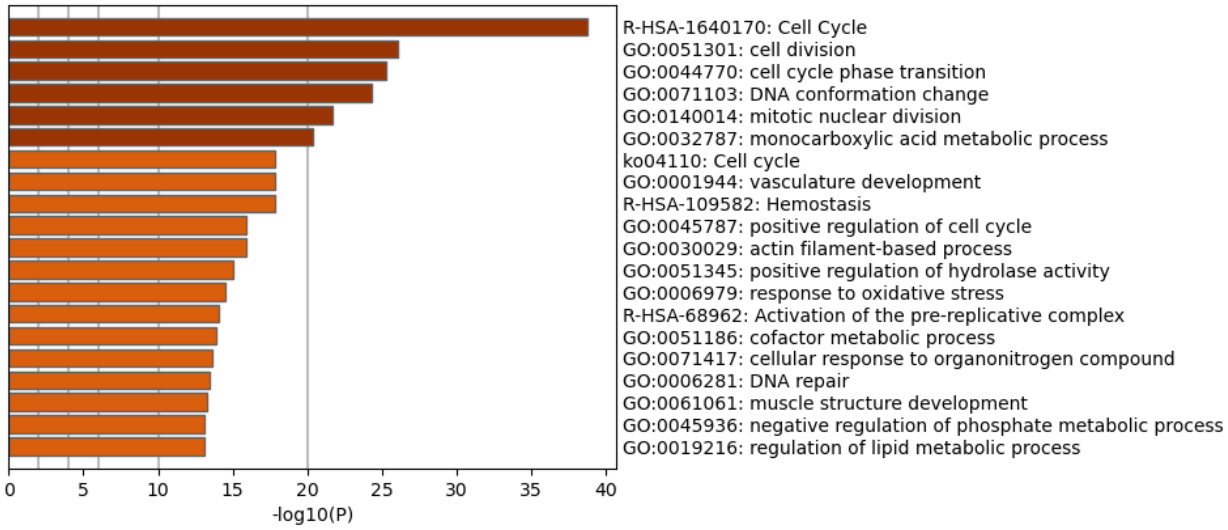


Figure 4.11 The results of the enrichment analysis of the GSE12368 dataset in the ACC-N group.

Another dataset of the ACC-N group, the GSE19750 genes, are mostly enriched in mechanisms that play an effective role in the formation of cancer. For example, gene expression (transcription), RNA polymerase II transcription, and generic transcription pathway, cell cycle phase transition, Alcholism, and positive regulation of cellular component biogenesis (Table A6), (Figure 4.12).

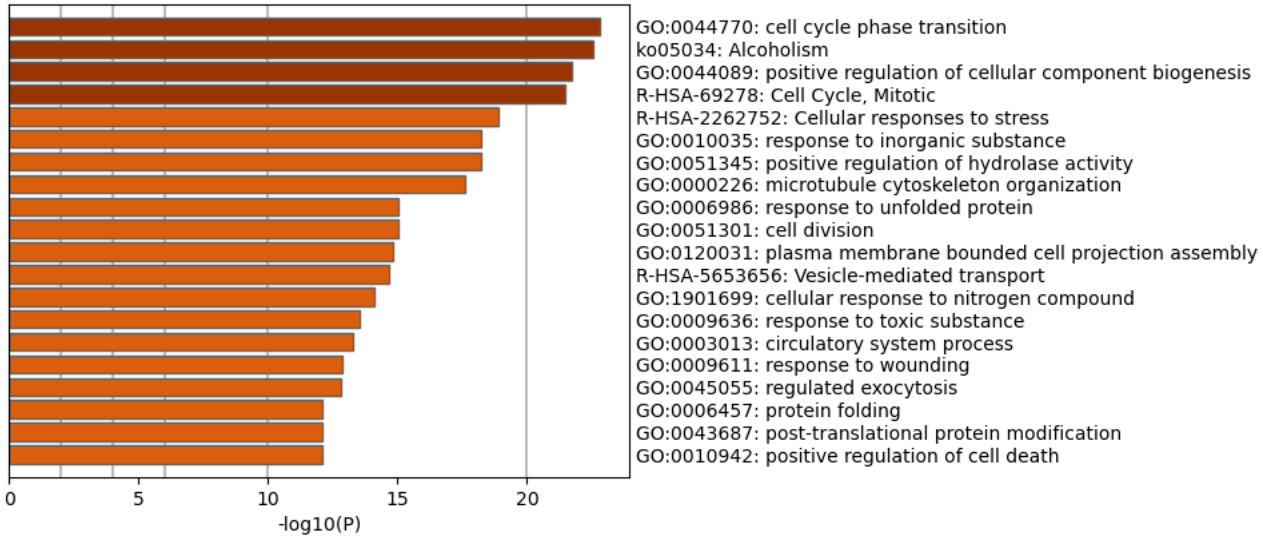


Figure 4.12 The results of the enrichment analysis of the GSE19750 dataset in the ACC-N group.

The GSE75415 genes, the last dataset of the ACC-N group, have been associated with cancer-related processes, just like the GSE10927 dataset genes, these are; mitotic cell cycle, cell cycle, cell cycle checkpoints (Table A7), (Figure 4.13).

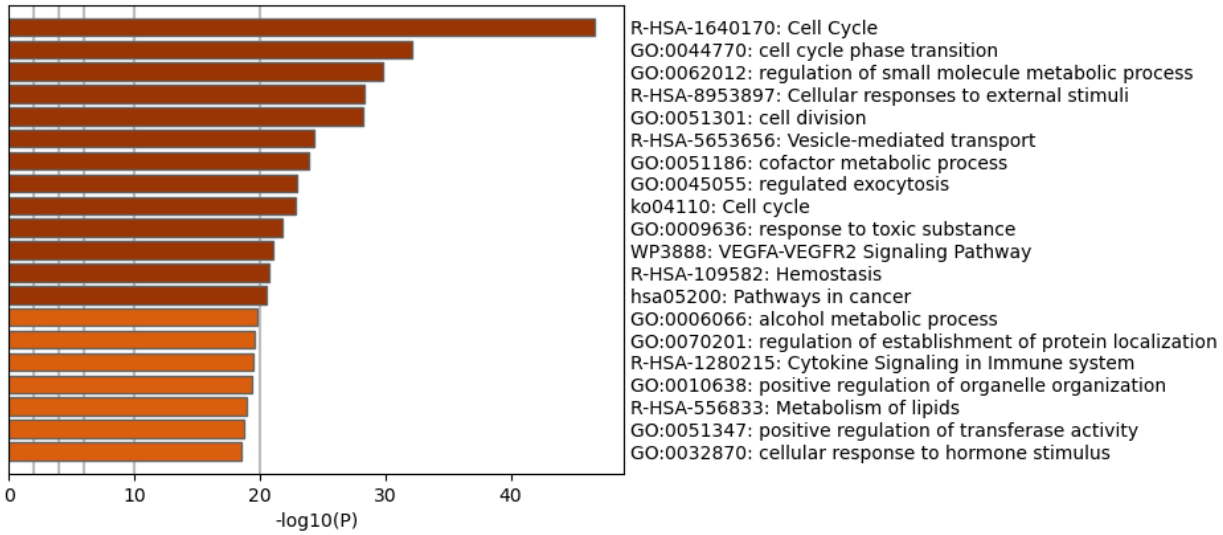


Figure 4.13 The results of the enrichment analysis of the GSE75415 dataset in the ACC-N group.

Another analysis in this study was based on the mutual comparison of ACC and ACA tumors. Enrichment analyses were also performed between ACC-ACA genes in order to better understand the difference between these two groups. As a result, interestingly, the genes in the GSE10927 and GSE12368 datasets are enriched with cancer-related processes (e.g. cell cycle, mitotic cell cycle, resolution of sister chromatid cohesion, cell division, DNA conformation change), just like in the ACC-N group (Tables A8, A9), (Figure 4.14 and 4.15).

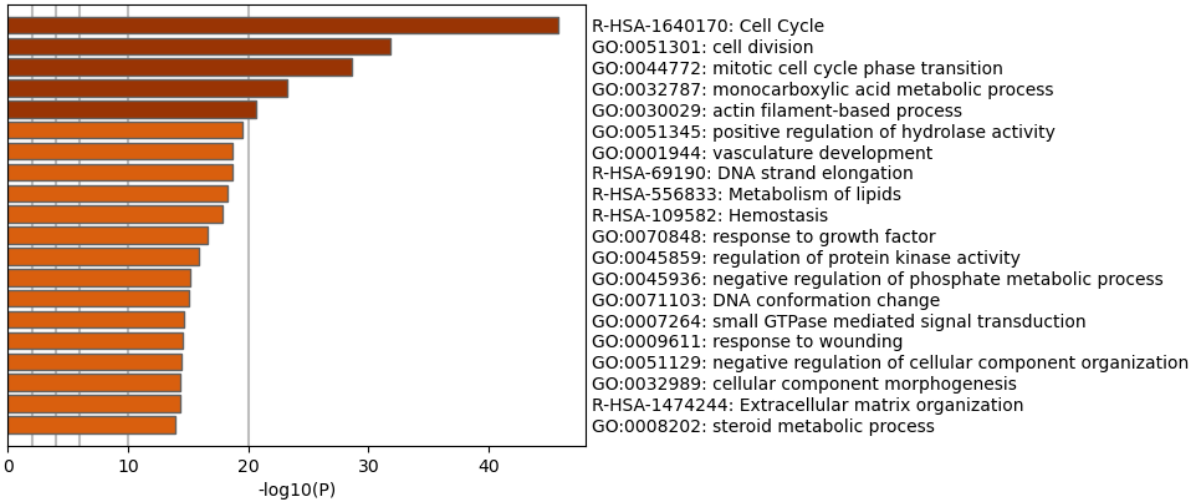


Figure 4.14 The results of the enrichment analysis of the GSE10927 dataset in the ACC-ACA group.

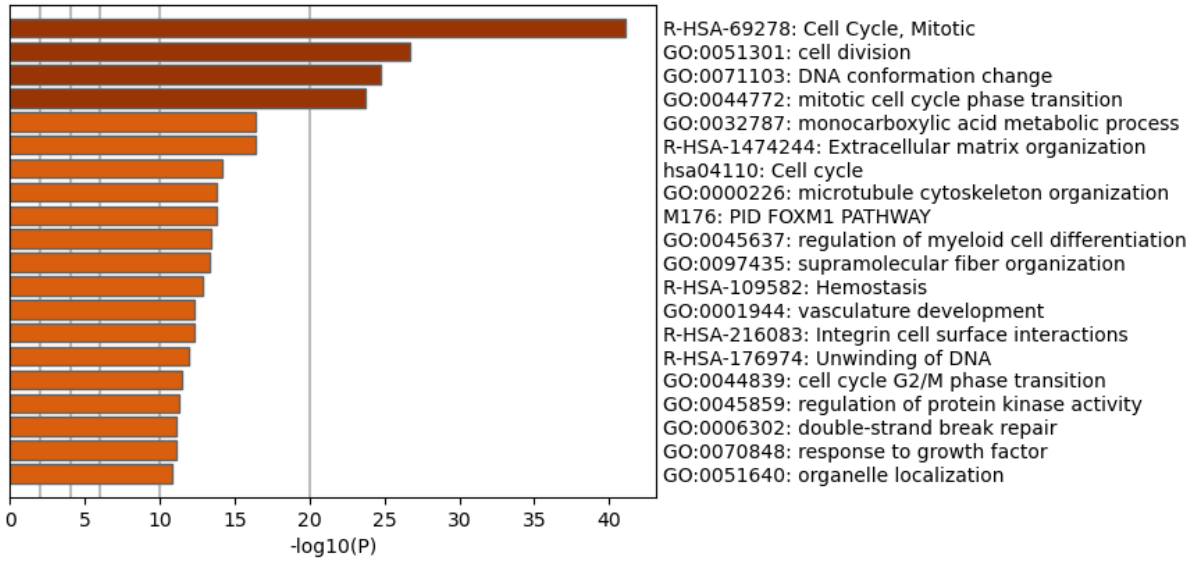


Figure 4.15 The results of the enrichment analysis of the GSE12368 dataset in the ACC-ACA group.

The genes of the GSE75415 dataset, which is the other dataset of the ACC-ACA group, are enriched quite differently from the other datasets. In this dataset, more specifically, antigen processing and presentation, phosphorylation of CD3 and TCR zeta chains, lymphocyte

mediated immunity, myeloid leukocyte activation, diseases, and processes such as Asthma is observed (Table A10).

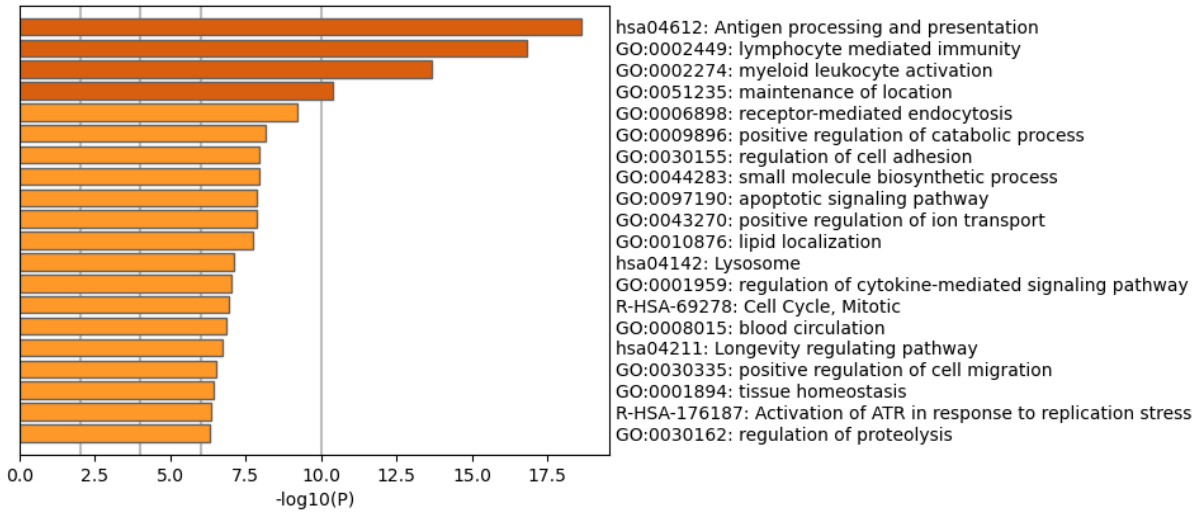


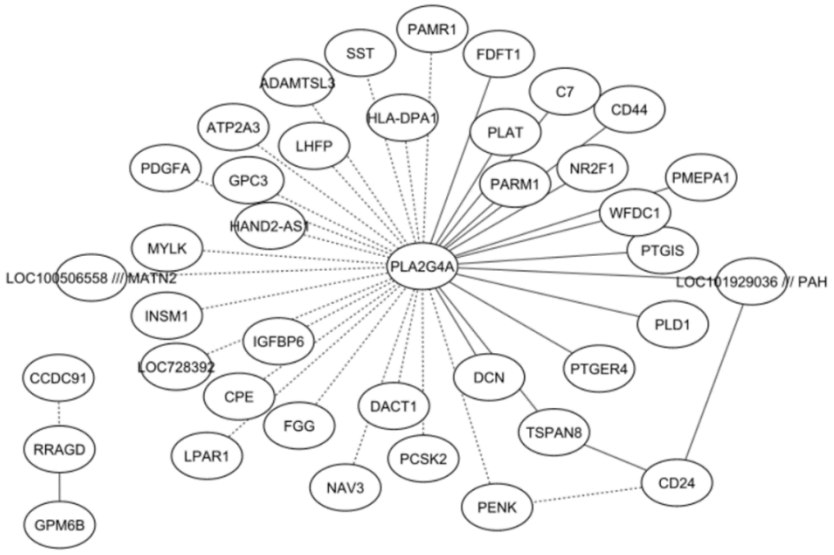
Figure 4.16 The results of the enrichment analysis of the GSE75415 dataset in the ACC-ACA group.

Comparative gene enrichment analysis was performed and 25 common significant KEGG pathways enriched by using the DEGs of each dataset (Figure 4.17). Pathways in cancer, proteoglycans in cancer, viral myocarditis, cAMP signaling pathway were identified as enriched terms in ACA-N group. Cell cycle, nucleocytoplasmic transport, p53 signaling pathway, cellular senescence, DNA replication, small cell lung cancer, oocyte meiosis, excision repair, progesterone-mediated-oocyte maturation were identified as enriched terms in ACC-N group. Interestingly, enriched terms in ACA-N group were almost down-regulated whereas enriched terms in ACC-N group were up-regulated pathways. More down-regulated pathways (80%) were observed in the ACC-ACA group compared to the ACC-N group (30.7%). Cell cycle and DNA replication were identified as up-regulated enriched terms in the ACC-ACA group.

Figure 4.17. The gene enrichment analysis of the DEGs belonging to each dataset through comparative and integrative perspective.

4.2 Reverse gene expression correlation (RGEC) networks in adrenocortical cancer progression

Identification of reversely correlated genes that have both a positive correlation in the case state and a negative correlation in the control state (or vice versa) from the common DEGs provides an in-depth understanding of the ACC progression. Thus, we aimed to determine reversely correlated genes, using Spearman's gene correlation coefficient. Three RGEC networks were recontructed by using binary genes that have both a positive correlation in the case state and a negative correlation in the control state (or vice versa). The RGEC network for ACA-N that was reconstructed included 40 nodes and 40 edges by using significant reversely pairwise gene correlations (Figure 4.18). RGEC network for ACC-N that was reconstructed included 288 nodes and 1460 edges by using significant reversely pairwise gene correlations (Figure 4.19), while there are 31 interactions among 30 RGEC genes belonging to the ACC-ACA group (Figure 4.20).



LINE TYPE	DESCRIPTION
----------	Negative correlation in score in case state and positive correlation score in control state
————	Positive correlation in score in case state and negative correlation score in control state

Figure 4.18 The RGEC networks for ACA-N groups. The correlation between genes with a negative score in case state and positive score in control state was shown by a dashed line, while the correlation between genes with positive score in case state and negative score in control state was shown by straight line.

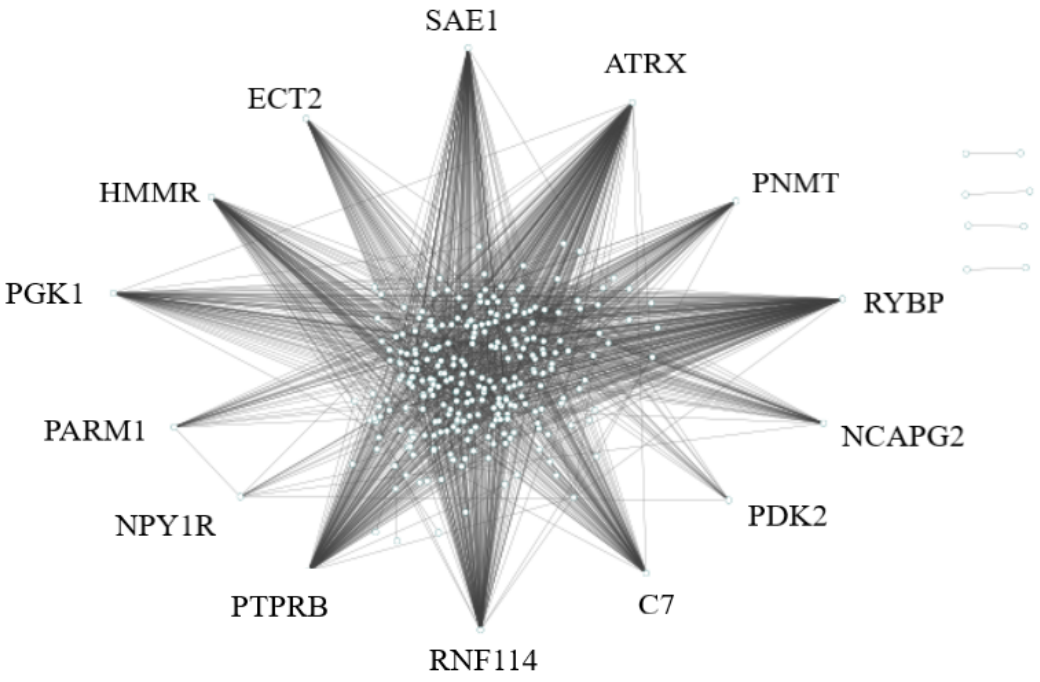
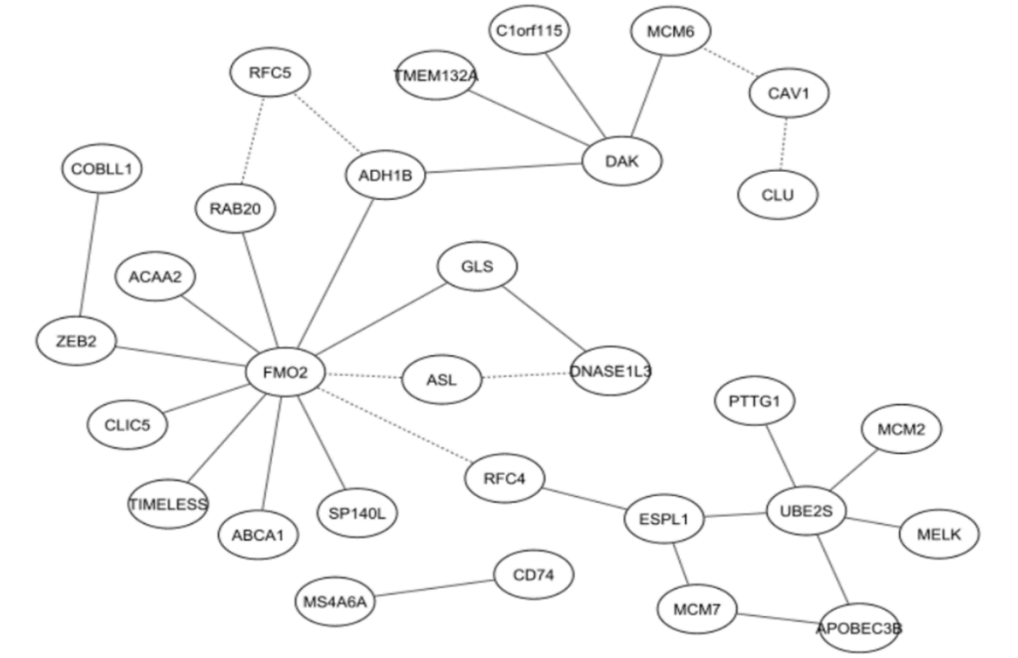


Figure 4.19 The RGEC networks for ACC-N group. The correlation between genes was shown by a straight line.



LINE TYPE	DESCRIPTION
----------	Negative correlation in score in case state and positive correlation score in control state
————	Positive correlation in score in case state and negative correlation score in control state

Figure 4.20. The RGEC networks for ACC-ACA group. The correlation between genes with a negative score in case state and positive score in control state was shown by a dashed line, while the correlation between genes with positive score in case state and negative score in control state was shown by straight line.

4.3 Central genes in RGEC networks

Hub genes which are named as central genes were identified according to degree and betweenness scores of the each RCEG network. It was determined PLA2G4A gene is the central gene in the RCEG network of ACA-N group (Figure 4.18). Since the number of genes is higher than in the ACC-N group compared the other nwtwork genes, many central genes have been identified icluding RYBP, RNF114, PTPRB, SAE1, PNMT, PGK1, HMMR, SSPN, NCAPG2, PARM1, ATRX, ECT2, PDK2, NPY1R and USP48 which sorted by degree score (Figure 4.19). In the network of ACC-ACA group, the FMO2 and UBE2S genes were defined

as central genes (Figure 4.20). The biological significance of these hub genes has been determined using the GeneCards database (Table 4.1).

Table 4.1 Biological importance of central genes in ACA-N, ACC-N and ACC-ACA group.

Group	**Hub Gene**	**Biological Importance**
ACA-N	PLA2G4A	It is involved in the remodeling of membrane lipids and the synthesis of lipid mediators in the inflammatory response.
ACC-N	SAE1	It regulates protein structure and intracellular localization.
ACC-N	ATRX	It is involved in transcriptional regulation and chromatin restructuring.
ACC-N	PNMT	It is involved in the regulation of epinephrine production.
ACC-N	RYBP	It is thought to be involved in activating DNA binding activity and transcription core regulatory activity.
ACC-N	NCAPG2	Together with the condensin I complex, it plays a role in chromosome assembly and segregation during mitosis.
ACC-N	PDK2	Diseases associated with PDK2 are Autosomal Dominant Polycystic Kidney Disease and Polycystic Kidney Disease.
ACC-N	C7	C7 gene acts as a membrane stabilizer.
ACC-N	RNF114	It participates in the regulation of many biological processes, such as the cell cycle, apoptosis, and osteoclastogenesis.
ACC-N	PTPRB	The PTBR gene regulates them by interacting with the neuronal receptor and cell-building molecules such as contact and tenasin C.
ACC-N	NPY1R	It plays a role in the mobilization of intracellular calcium and the inhibition of adenylate cyclase activity.
ACC-N	PARM1	It is known that this gene regulates telomerase activity, thereby enabling some prostate cells to resist apoptosis.
ACC-N	PGK1	This protein is secreted by tumor cells, where it participates in angiogenesis by acting as a reducing disulfide bond in the serine protease plasmin.
ACC-N	HMMR	This gene plays a role in cell mobility.
ACC-ACA	FMO2	This gene is associated with the disease Trimethylaminuria.
ACC-ACA	UBE2S	It acts as an important factor of the anaphase promoting complex/cyclosome (APC/C), a ubiquitin ligase regulated by the cell cycle that controls progression through mitosis.

We performed an enrichment analysis to determine the biological significance of RGEC genes showing a reverse correlation in the ACA-N group.

First, we performed an enrichment analysis of the genes that showed a negative correlation in the ACA samples and a positive correlation in the control samples. These genes are enriched Regulation of Insülin-like Growth factor (IGF) transport and uptake by Insulin-like Growth factor binding proteins, regulation of wound healing, G alpha signaling events.

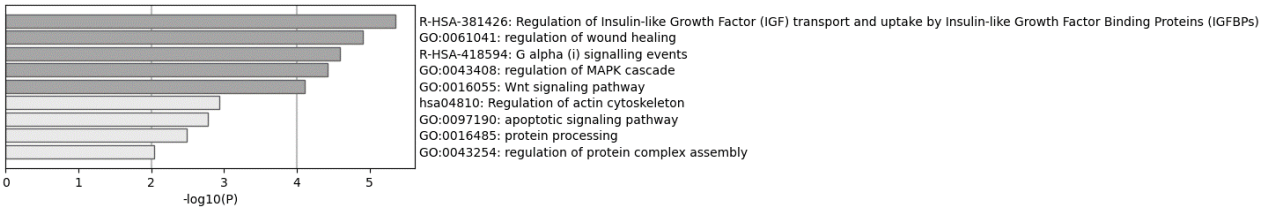


Figure 4.21 The results of the enrichment analysis of the negative correlation in the ACA samples and positive correlation in the control samples of the ACA-RGEC genes.

We performed an enrichment analysis of the genes that showed a positive correlation in the ACA samples and a negative correlation in the control samples. These genes are enriched wound healing, Prostaglandin synthesis and regulation, negative regulation of response to external stimulus.

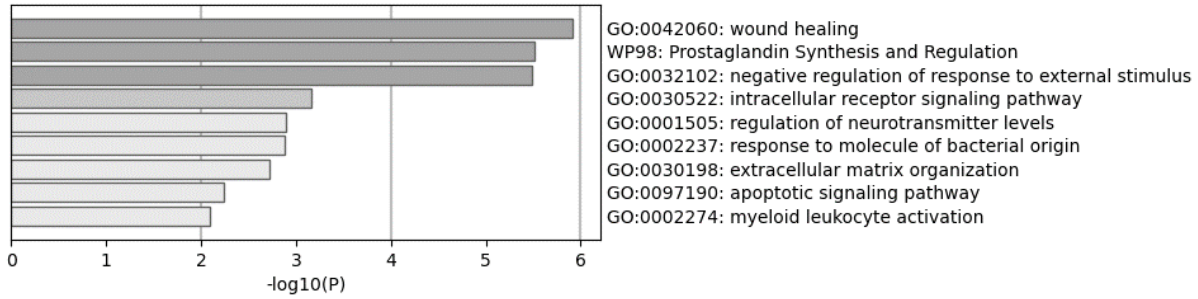


Figure 4.22 The results of the enrichment analysis of the positive correlation in the ACA samples and negative correlation in the control samples of the ACA-RGEC genes.

We performed gene enrichment analysis of 40 RGEC genes that showed a positive correlation in ACA samples and negative correlation in control samples and a positive correlation in negative control samples in ACA samples (Table 4.1).

Table 4.2 The results of the enrichment analysis of the 40 RGEC genes of ACA-N group.

Category	Term	P Value
REACTOME_PATHWAY	R-HSA-381426~Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs)	0.003856
KEGG_PATHWAY	hsa04072:Phospholipase D signaling pathway	0.005529
REACTOME_PATHWAY	R-HSA-162582~Signal Transduction	0.006808

We performed enrichment analysis on the RCEG network genes of ACC-N group via the METASCAPE web portal to identify biological process and pathways. Enrichment analysis was performed by dividing the RGEC genes of ACC-N group into two groups. The first group is 140 genes with a negative correlation in ACC samples and positive correlation in control samples, second group is 272 genes with positive correlation in ACC samples and negative correlation in control samples. The first group 140 genes were found to be associated with nuclear chromosome segregation, regulation of the cell cycle process, and regulation of chromosome segregation which are the top three enrcihed terms. Moreover, it has been observed small three modules of protein protein interaction (PPI) network of 140 genes. Gene enrichment analysis of 272 genes was performed. It was found that these genes are associated with Cell Cycle, PID Aurora B Pathway, and PID FOXM1 Pathway which are the top three enrcihed terms. Furhermore, it has been observed four modules of protein protein interaction (PPI) network of 272 genes.

We performed an enrichment analysis for our second group, ACC RGEC genes. We determined the biological pathways of genes that showed a negative correlation in ACC

samples and a positive correlation in control samples. These are nuclear chromosome segregation, regulation of cell cycle processes, regulation of chromosome segregation.

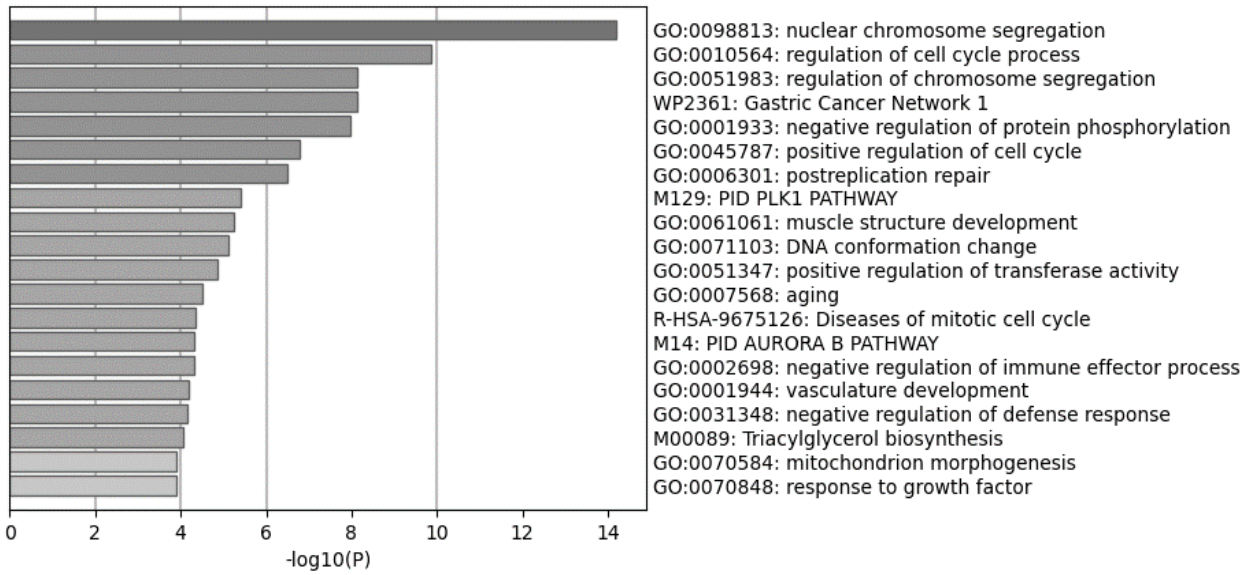
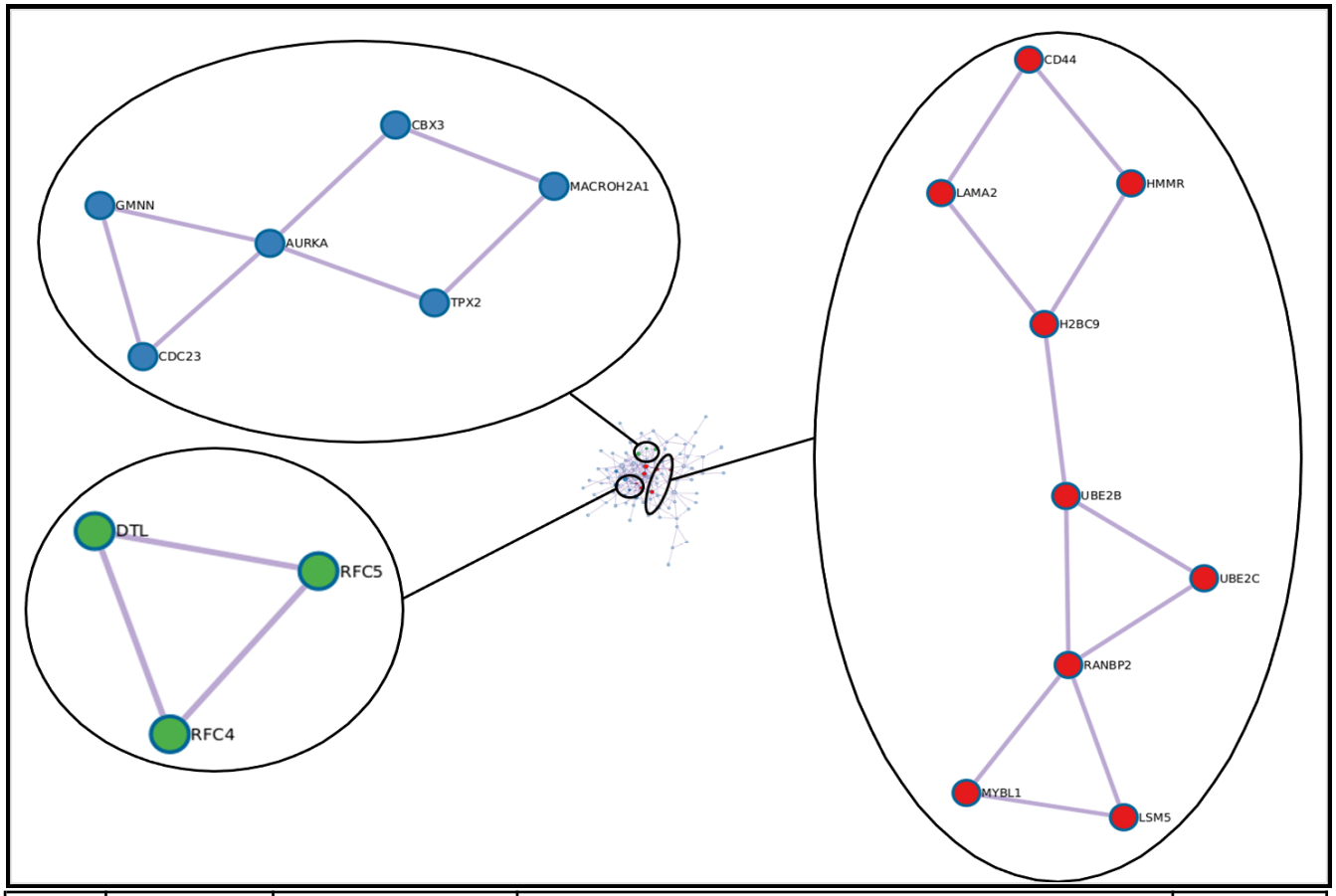


Figure 4.23 The results of the enrichment analysis of the negative correlation in the ACC samples and positive correlation in the control samples of the ACC-RGEC genes.



Color	MCODE	GO	Description	Log10(P)
	MCODE_1	R-HSA-8852135	Protein ubiquitination	-5.8
	MCODE_1	ko04512	ECM-receptor interaction	-5.7
	MCODE_1	hsa04512	ECM-receptor interaction	-5.6
	MCODE_2	GO:0044770	cell cycle phase transition	-7.4
	MCODE_2	GO:0010564	regulation of cell cycle process	-7
	MCODE_2	GO:0140014	mitotic nuclear division	-6.8
	MCODE_3	R-HSA-110314	Recognition of DNA damage by PCNA-containing replication complex	-9
	MCODE_3	GO:0042769	DNA damage response, detection of DNA damage	-8.6
	MCODE_3	GO:0019985	translation synthesis	-8.5

Figure 4.24 Protein protein interaction (PPI) network of reverse correlated genes with a negative correlation in ACC samples and positive correlation in control samples and it modules.

We determined the biological pathways of genes that showed a positive correlation in ACC samples and a negative correlation in control samples. These are cell cycle, cell cycle mitotic, PID AURORA B Pathway.

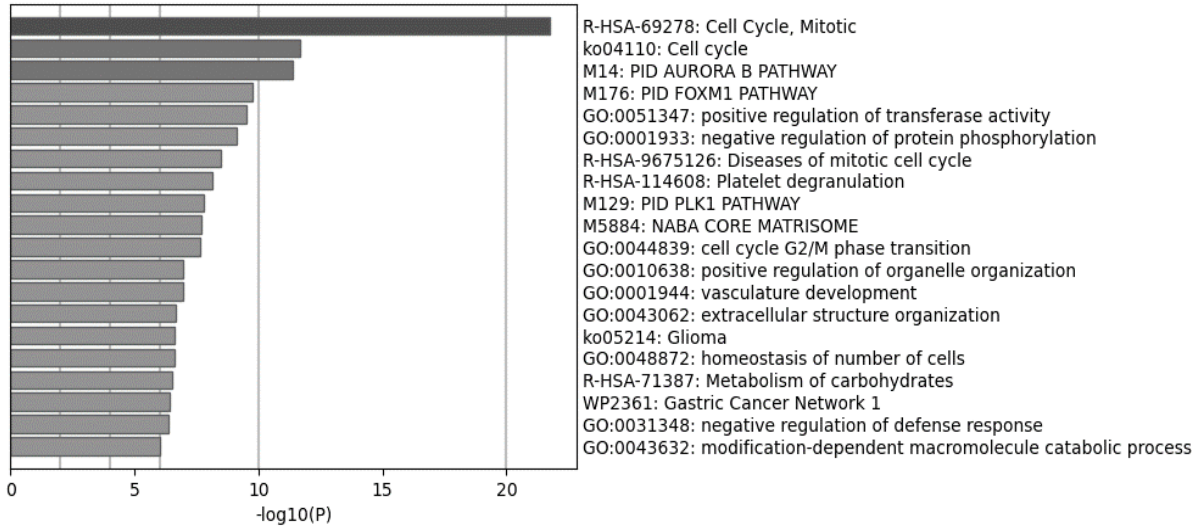
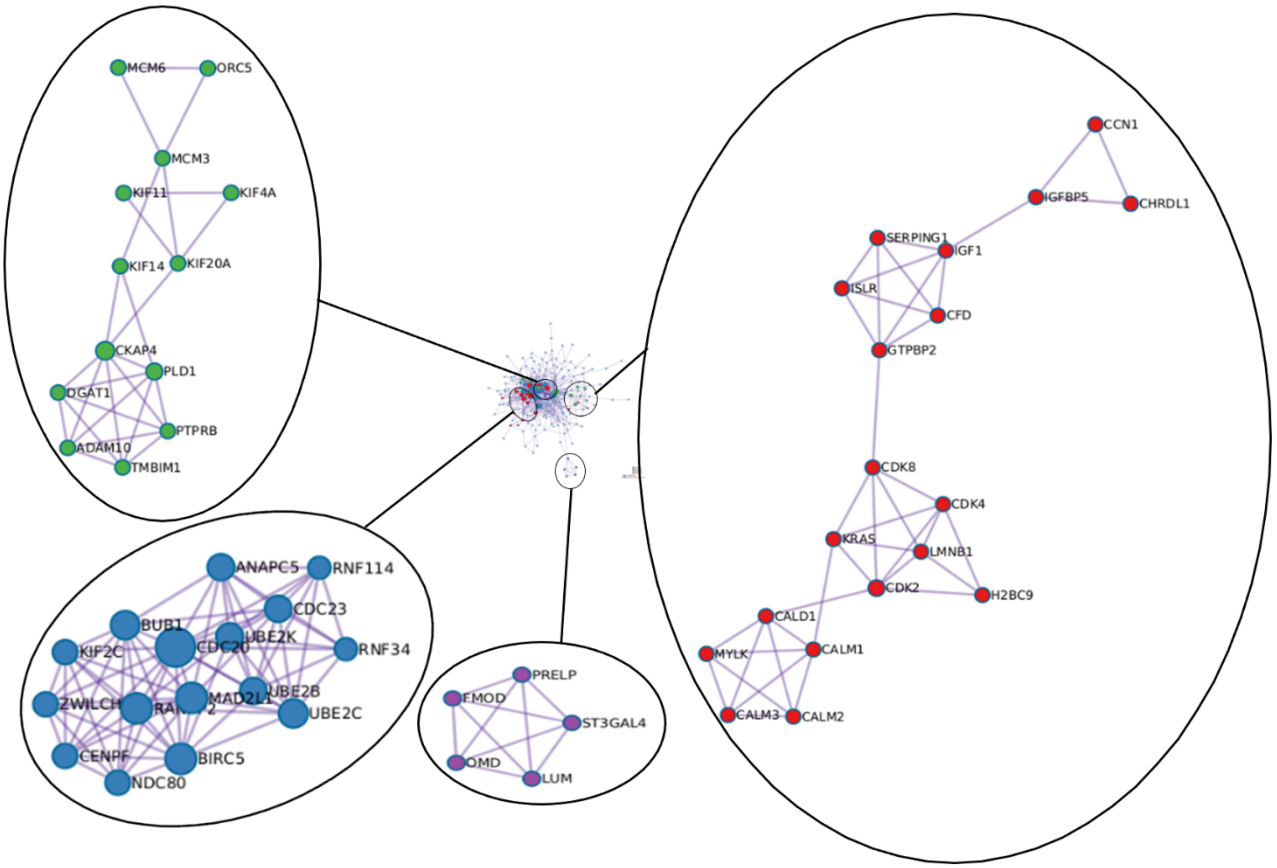


Figure 4.25 The results of gene enrichment analysis of reverse correlated genes with positive correlation in ACC samples and negative correlation in control samples of ACC-RGEC genes.



Color	MCODE	GO	Description	Log10(P)
	MCODE_1	R-HSA-114608	Platelet degranulation	-13.9
	MCODE_1	R-HSA-76005	Response to elevated platelet cytosolic Ca2+	-13.8
	MCODE_1	R-HSA-109582	Hemostasis	-11.7
	MCODE_2	R-HSA-69618	Mitotic Spindle Checkpoint	-25.8
	MCODE_2	R-HSA-2467813	Separation of Sister Chromatids	-22.9
	MCODE_2	R-HSA-68882	Mitotic Anaphase	-21.8
	MCODE_3	R-HSA-6798695	Neutrophil degranulation	-7.4
	MCODE_3	GO:0043312	neutrophil degranulation	-7.4
	MCODE_3	GO:0002283	neutrophil activation involved in immune response	-7.4
	MCODE_4	R-HSA-2022854	Keratan sulfate biosynthesis	-15.2

	MCODE_4	GO:0018146	keratan sulfate biosynthetic process	-15.2
	MCODE_4	GO:0042339	keratan sulfate metabolic process	-14.8

Figure 4.26 Protein protein interaction (PPI) network of reverse correlated genes with a positive correlation in ACC samples and negative correlation in control samples and it modules.

We performed gene enrichment analysis of 288 RGEC genes that showed a positive correlation in ACC samples and negative correlation in control samples and a positive correlation in negative control samples in ACC samples.

Table 4.3 The results of the enrichment analysis of the 288 RGEC genes of ACC-N group.

Category	Term	P Value
REACTOME_PATHWAY	R-HSA-69278~Cell Cycle, Mitotic	9.66E-14
REACTOME_PATHWAY	R-HSA-1640170~Cell Cycle	1.66E-11
KEGG_PATHWAY	hsa04110:Cell cycle	1.47E-08
REACTOME_PATHWAY	R-HSA-69620~Cell Cycle Checkpoints	4.99E-08
REACTOME_PATHWAY	R-HSA-69239~Synthesis of DNA	2.07E-07
REACTOME_PATHWAY	R-HSA-114608~Platelet degranulation	4.39E-07
REACTOME_PATHWAY	R-HSA-76005~Response to elevated platelet cytosolic Ca2+	6.82E-07
REACTOME_PATHWAY	R-HSA-69242~S Phase	1.12E-06
REACTOME_PATHWAY	R-HSA-9675126~Diseases of mitotic cell cycle	3.88E-06
REACTOME_PATHWAY	R-HSA-69618~Mitotic Spindle Checkpoint	4.87E-06

We determined the biological pathways of genes that showed a negative correlation in ACC samples and a positive correlation in ACA (control) samples. These genes are enriched DNA strand elongation, nucleic acid phosphodiester bond hydrolysis, and positive regulation of transferase activity.

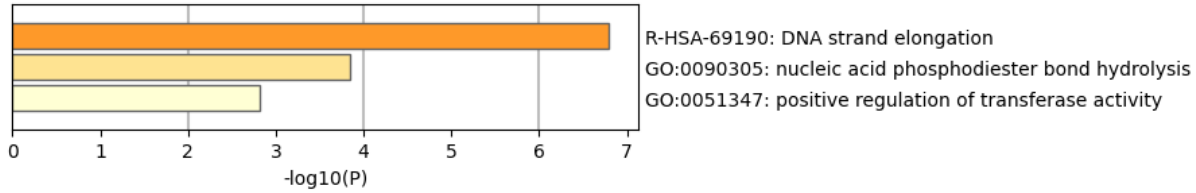


Figure 4.27 The results of the enrichment analysis of the negative correlation in the ACC samples and positive correlation in the ACA samples of the ACC-ACA-RGEC genes.

We determined the biological pathways of genes that showed a positive correlation in ACC samples and a negative correlation in ACA (control) samples. These genes are enriched MCM2-MCMMCM7 complex, organic hydroxy compound metabolic process, and negative regulation of chromosome organization.

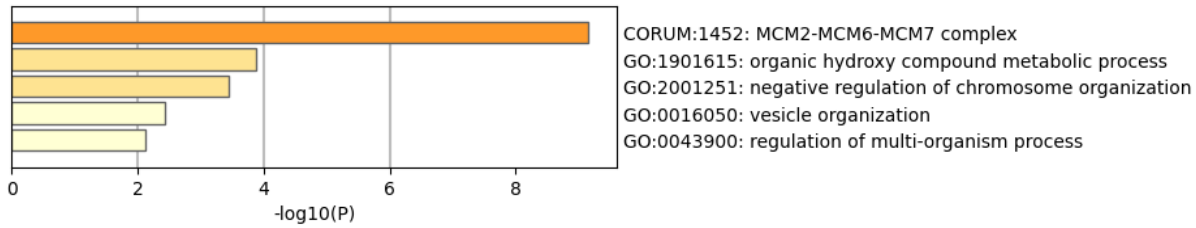
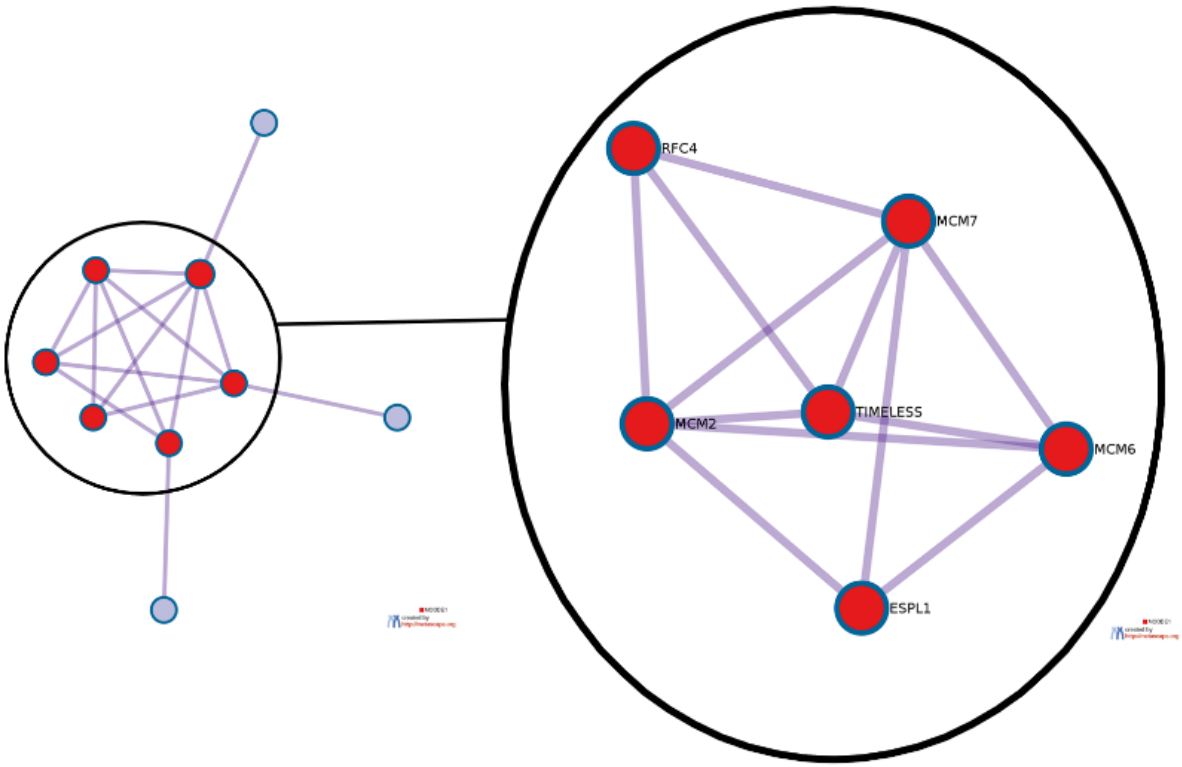


Figure 4.28 The results of the enrichment analysis of the positive correlation in the ACC samples and negative correlation in the ACA samples of the ACC-ACA-RGEC genes.



MCODE	GO	Description	Log10(P)
MCODE_1	CORUM:1452	MCM2-MCM6-MCM7 complex	-11.3
MCODE_1	R-HSA-69190	DNA strand elongation	-10.7
MCODE_1	CORUM:2792	MCM2-MCM4-MCM6-MCM7 complex	-10.7

Figure 4.29 Protein protein interaction (PPI) network of reverse correlated genes with a positive correlation in ACC samples and negative correlation in ACA samples and it modules.

We performed gene enrichment analysis of 30 RGEC genes that showed a sum of positive correlation in ACC samples and negative correlation in ACA samples and a negative correlation in ACC samples and positive corelation in ACA samples.

Table 4.4 The results of the enrichment analysis of the 30 RGEC genes of ACC-ACA group.

Category	Term	P Value
REACTOME_PATHWAY	R-HSA-69190~DNA strand elongation	3.46E-07
REACTOME_PATHWAY	R-HSA-176187~Activation of ATR in response to replication stress	6.32E-07
KEGG_PATHWAY	hsa03030:DNA replication	8.08E-07
REACTOME_PATHWAY	R-HSA-69239~Synthesis of DNA	2.67E-06
REACTOME_PATHWAY	R-HSA-69242~S Phase	1.15E-05
REACTOME_PATHWAY	R-HSA-69306~DNA Replication	2.24E-05
REACTOME_PATHWAY	R-HSA-69278~Cell Cycle, Mitotic	5.44E-05
REACTOME_PATHWAY	R-HSA-69620~Cell Cycle Checkpoints	0.000189
REACTOME_PATHWAY	R-HSA-1640170~Cell Cycle	0.000205
REACTOME_PATHWAY	R-HSA-176974~Unwinding of DNA	0.000228

4.4 Performance evaluation of reversely correlated genes via machine learning classification algorithms

Gene expression values of RGEC network for ACC-N were obtained from TCGA to elavulate the prognostic and diagnostic performance of RGEC genes. RNA-Seq datasets of ACC (n = 78) were downloaded and applied z normalization. Classification, a machine learning technique, was used to evaluate putative biomarker candidates identified by bioinformatic analysis. Since an effective potential biomarker needs to be able to distinguish the diseased cohort from the controls and dead patients from the alive patients we used various classification algorithms to determine the potential of key biomolecules. The performance of the key biomolecules was evaluated according to the predicted F1, recall, precision and accuracy score performance of the classification algorithms including CatBoost, LGBM, XGB, Gradient Boosting, Random Forest, Decision Tree, MLP, logistic regression and K-Neighbors. We hypothesized that the novel RGEC genes we provided here could efficiently distinguish diseased samples from controls and dead samples from the alive. We evaluated the key biomolecules based on the predictive accuracy score of the classifier for diagnostic and prognostic analysis and determined that the accuracy of the nine different classification algorithm ranged from 80% to 100% and 46% to 67%, respectively (Figure 4.30).

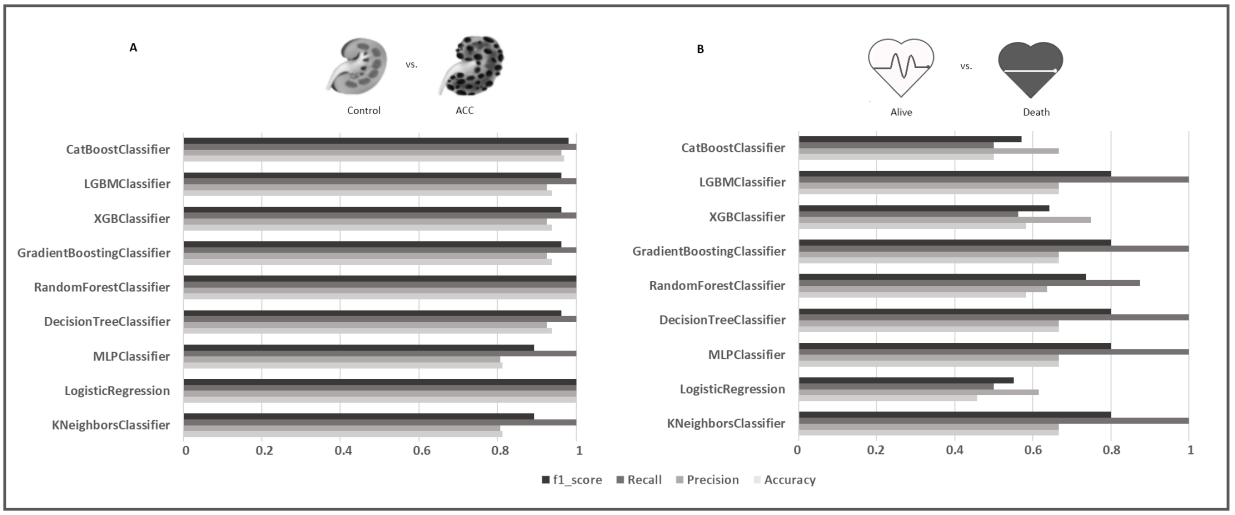


Figure 4.30 Machine learning analysis for reverse correlated genes. (A) The F1, recall, precision and accuracy score plot of nine different classification algorithms for discriminating between ACC tissues and control tissues. (B) The F1, recall, precision and accuracy score plot of nine different classification algorithms for discriminating between alive and dead samples with ACC.

5. DISCUSSION

Adrenal cancer is a type of cancer that occurs in the adrenal gland, whose benign masses are quite common, and the prognosis worsens significantly as the malignancy increases. The fact that the symptoms of the disease are not specific makes it difficult to diagnose and screen the prognosis (Terzolo et al. 2014). For this reason, the identification of biomolecules that play an active role in the formation and development of the disease progression is very important for the diagnosis of the disease and the transition to malignancy.

Bioinformatics techniques allow accurate analysis of omics data on the disease. Recently, artificial intelligence and MLT, which appeared with the development of technology, came into play. Much more rational results are obtained after integrating MLT and bioinformatics methods. In this study, it was used a new method to determine the statistically significant reverse correlated genes which are both a positive correlation in the case state and a negative correlation in the control state (or vice versa) and prognostic and diagnostic performance were evaluated via nine classification MLT algorithms. Bioinformatic analyzes were performed between ACA, ACC, and healthy samples to identify key biomolecules involved in adrenal cancer progression. Considering the batch effect of gene expression datasets, comparative

analyzes of tumor tissues (ACA and ACC) and normal tissues obtained from the adrenal cortex, were performed on the same microarray platform.

Each gene expression dataset was analyzed seperately and common DEGs were identified for ACA-N, ACC-N and ACC-ACA groups. Obtaining higher DEGs (454 genes) in the ACC-N group than in other groups (110 and 128 genes) has been associated with the malignancy of the disease. The up-regulated genes in the ACC-N and ACC-ACA groups are significantly enriched in the cell cycle and DNA replication. It can be said that tumor formation occurs due to alterations occurring in the cell cycle and changes in checkpoints. In addition, it can be suggested that DNA damage, and mutations occurring during apoptosis and replication are also effective in the formation and development of ACC. The fact that these biological pathways are widespread in both groups, ACC-N and ACC-ACA, can be interpreted as having an effect on the transformation of the disease from benign to malignant tumors.

Similarly, down-regulated genes obtained from ACC-N and ACC-ACA groups have been enriched in tuberculosis disease. According to information obtained from the literature, it has been found that people with tuberculosis disease have a risk of developing cancer (Anastasopoulou et al. 2019). The enrichment of proteoglycans in both the ACA-N and ACC-N groups confirms the effect of proteoglycans on circulating tumor cells that cause metastasis (Ahrens et al. 2020).

Common DEGs expression values were applied z normalization and gene pair correlations were calculated. Binary genes that have both a positive correlation in the case state and a negative correlation in the control state (or vice versa) were selected and named as reversely correlated genes. Afterward, reverse gene expression correlation (RGEC) networks were reconstructed. The central genes that have higher degree score than other genes, were identified. PLA2G4A gene in ACA-N group, 14 genes (i.e: SAE1, ATRX, PNMT, RYBP, NCAPG2, PDK2, C7, RNF114, PTPRB, NPY1R, PARM1, PGK1, HMMR and ETC2) in ACCN group, the FMO2 and UBE2S genes in the ACC-ACA group were determined. It was found that 14 central genes for RCEG network for ACC-N group efficiently discriminate the ACC tissues from the controls compared to the performance of discriminating in the live and dead samples. Since, it may suggest that key biomolecule candidates have the diagnostic capability of ACC.

A literature review was conducted to determine the role of these central genes in the progression of disease. Although the role of the ATRX gene in the formation and mechanism of cancer is not fully known, it is known that it is one of the most frequently mutated tumor suppressor gene in malignancies (Pang et al. 2023). Furthermore, the fact that it plays an important role in DNA damage repair and replication indicates that a disorder in this gene is effective in cancer formation and disease resistance (Pang et al. 2023). It has been determined that the ATRX gene is also a prognostic biomarker in ACC (Brondani et al. 2021). Although it is reported that the PNMT gene is generally associated with neurodegenerative diseases, the

expression of the PNMT in Pheochromocytoma, a subtype of ACA, has been studied and found to play a role in the molecular mechanism underlying the disease (Saxena et al. 2022). It has been reported that overexpression of the NCAPG2 gene increases malignancy in many types of cancer, including ACC, and causes alterations in the signaling pathways (Wang et al. 2023). The PDK2 gene is one of the inhibitors that facilitate tumor cell proliferation and induce apoptosis, an increase in activation of PDK2 has been found to be associated with ACC (Cui et al. 2020). It was also reported that NPY1R, FMO2 and UBE2S gene expression are associated with ACC via bioinformatics based analysis (Yao et al. 2020; Bao et al. 2022;)

It was reported that PTPRB gene is the prognostic genes in the ACC tumor microenvironment (Li et al. 2020). According to literature review, it was suggest that PLA2G4A, SAE1, RYBP, C7, RNF114, PARM1, PGK1, HMMR and ETC2 genes are novel genes for ACC. These genes provide very valuable information about the pathogenesis of the disease and prospective treatment methods. Although our study provides key genes and pathways for ACC progression, several limitations of the study should be noted as we could not validate the performance of discriminating in the ACA and ACC tissues because of insufficient datasets; thus, MLT analyses were applied ACC and control samples.

6. CONCLUSION

We present here for the first time the identification of key biomolecules of ACC progression gene correlation based new method approach and ML algorithms. The bioinformatics and MTL approach determined key genes including PLA2G4A gene in ACA-N group, 14 genes (i.e: SAE1, ATRX, PNMT, RYBP, NCAPG2, PDK2, C7, RNF114, PTPRB, NPY1R, PARM1, PGK1, HMMR and ETC2) in ACC-N group, the FMO2 and UBE2S genes in the ACC-ACA group. Moreover, it was found that 14 central genes for RCEG network for ACC-N group efficiently discriminate the ACC tissues from the controls compared to the performance of discriminating in the live and dead samples. Since, it may suggest that key biomolecule candidates have the diagnostic capability of ACC. It was reported that several biological processes and pathways are involved in the progression of ACC. It has been proven how important the MLT used in obtaining this result is for validation. While our study showed the benefit and usefulness of bioinformatics and MLT approaches in bringing out the key biomolecules that can be potential diagnostic, prognostic and treatment targets of ACC, the need for improved analysis and prospective clinical studies is still imperative.

7. REFERENCE

Abu Alfeilat, H. A., Hassanat, A. B., Lasassmeh, O., Tarawneh, A. S., Alhasanat, M. B., Eyal Salman, H. S., & Prasath, V. S. (2019). Effects of distance measure choice on k-nearest neighbor classifier performance: A review. *Big Data*, *7*(4), 221–248.

Ahrens, T.D., Bang-Christensen, S.R., Jørgensen, A.M. (2020). The role of proteoglycans in cancer metastasis and circulating tumor cell analysis. Front Cell Dev Biology 8, 749.

Alnuaim, A. A., Zakariah, M., Shukla, P. K., Alhadlaq, A., Hatamleh, W. A., Tarazi, H., Sureshbabu, R., & Ratna, R. (2022). Human-computer interaction for recognizing speech emotions using multilayer perceptron classifier. *Journal of Healthcare Engineering*, *2022*.

AlThuwaynee, O. F., Kim, S.-W., Najemaden, M. A., Aydda, A., Balogun, A.-L., Fayyadh, M. M., & Park, H.-J. (2021). Demystifying uncertainty in PM10 susceptibility mapping using variable drop-off in extreme-gradient boosting (XGB) and random forest (RF) algorithms. *Environmental Science and Pollution Research*, *28*(32), 43544–43566. https://doi.org/10.1007/s11356-021-13255-4

Altieri, B., Ronchi, C. L., Kroiss, M., & Fassnacht, M. (2020). Next-generation therapies for adrenocortical carcinoma. *Best Practice & Research Clinical Endocrinology & Metabolism*, *34*(3), 101434.

Alzamzami, F., Hoda, M., & Saddik, A. E. (2020). Light Gradient Boosting Machine for General Sentiment Classification on Short Texts: A Comparative Evaluation. *IEEE Access*, *8*, 101840–101858. https://doi.org/10.1109/ACCESS.2020.2997330

Anastasopoulou, A., Ziogas, D.C., Samarkos, M. (2019). Reactivation of tuberculosis in cancer patients following administration of immune checkpoint inhibitors: Current evidence and clinical practice recommendations. J Immunother Cancer, 7(1), 1–13.

Araujo-Castro, M., & Parra-Ramirez, P. (2022). Diagnosis of primary hyperaldosteronism. *Medicina Clínica (English Edition)*, *158*(9), 424–430.

Auwul, R., Rahman, R., Gov, E. (2021). Bioinformatics and machine learning approach identifies potential drug targets and pathways in COVID-19. Brief Bioinformatics, 22(5), bbab120.

Banaei, N., Moshfegh, J., Mohseni-Kabir, A., Houghton, J. M., Sun, Y., & Kim, B. (2019). Machine learning algorithms enhance the specificity of cancer biomarker detection using SERS-based immunoassays in microfluidic chips. *RSC Advances*, *9*(4), 1859–1868.

Bao, H., Luo, Y., Ding, G. (2022). A pan-cancer analysis of Ube2s in tumorigenesis, prognosis, pathway, immune infiltration and evasion, and therapy response from an immune-oncology perspective. J Oncol.

Bardou, P., Mariette, J., Escudié, F. (2014). jvenn: An interactive Venn diagram viewer. BMC Bioinformatics, 15(1), 1–7.

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., & Holko, M. (2012). NCBI GEO: Archive for functional genomics data sets—update. *Nucleic Acids Research*, *41*(D1), D991–D995.

Berruti, A., Fassnacht, M., Haak, H., Else, T., Baudin, E., Sperone, P., Kroiss, M., Kerkhofs, T., Williams, A. R., & Ardito, A. (2014). Prognostic role of overt hypercortisolism in completely operated patients with adrenocortical cancer. *European Urology*, *65*(4), 832–838.

Bisong, E., & Bisong, E. (2019). Logistic regression. *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*, 243–250.

Borges, K. S., Pignatti, E., Leng, S., Kariyawasam, D., Ruiz-Babot, G., Ramalho, F. S., Taketo, M. M., Carlone, D. L., & Breault, D. T. (2020). Wnt/β-catenin activation cooperates with loss of p53 to cause adrenocortical carcinoma in mice. *Oncogene*, *39*(30), 5282–5291.

Bolstad, B.M., Irizarry, R.A., Astrand, M. (2013). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics, 19, 185–193.

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, 144–152.

Bourdeau, I., El Ghorayeb, N., Gagnon, N., & Lacroix, A. (2018). Management of endocrine disease: Differential diagnosis, investigation and therapy of bilateral adrenal incidentalomas. *European Journal of Endocrinology*, *179*(2), R57–R67.

Brondani, V.B., Lacombe, A.M.F., Mariani, B.M.D.P. (2021). Low protein expression of both ATRX and ZNRF3 as novel negative prognostic markers of adult adrenocortical carcinoma. Int J Mol Sci, 22(3), 1238.

Campbell, M. K., & Donner, A. (1989). Classification efficiency of multinomial logistic regression relative to ordinal logistic regression. *Journal of the American Statistical Association*, *84*(406), 587–591.

Catalano, R., Giardino, E., Treppiedi, D., Mangili, F., Morelli, V., Elli, F. M., Serban, A. L., Luconi, M., Mannelli, M., & Spada, A. (2021). The cytoskeleton actin binding protein filamin A impairs both IGF2 mitogenic effects and the efficacy of IGF1R inhibitors in adrenocortical cancer cells. *Cancer Letters*, *497*, 77–88.

Catalano, R., Giardino, E., Treppiedi, D. (2021). The cytoskeleton actin binding protein filamin A impairs both IGF2 mitogenic effects and the efficacy of IGF1R inhibitors in adrenocortical cancer cells. Cancer Lett, 497, 77–88.

Charbuty, B., & Abdulazeez, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, *2*(01), Article 01. https://doi.org/10.38094/jastt20165

Chehade, M., Bullock, M., Glover, A., Hutvagner, G., & Sidhu, S. (2020). Key MicroRNA's and Their Targetome in Adrenocortical Cancer. *Cancers*, *12*(8), 2198.

Comertpay, B., & Gov, E. (2020). Identification of key biomolecules in rheumatoid arthritis through the reconstruction of comprehensive disease-specific biological networks. *Autoimmunity*, *53*(3), 156–166.

Crona, J., & Beuschlein, F. (2019). Adrenocortical carcinoma—Towards genomics guided clinical care. *Nature Reviews Endocrinology*, *15*(9), 548–560.

Cui, L., Cheng, Z., Liu, Y. (2020). Overexpression of PDK2 and PDK3 reflects poor prognosis in acute myeloid leukemia. Cancer Gene Ther, 27(1–2), 15–21.

Das, A. (2021). Logistic regression. In *Encyclopedia of Quality of Life and Well-Being Research* (pp. 1–2). Springer.

Datta, J., & Roses, R. E. (2016). Surgical management of adrenocortical carcinoma: An evidence-based approach. *Surgical Oncology Clinics*, *25*(1), 153–170.

Demeure, M.J., Coan, K.E., Grant, C.S. (2013). PTTG1 overexpression in adrenocortical cancer is associated with poor survival and represents a potential therapeutic target. Surgery, 154(6), 1405–1416.

Dominguez, A., Muppidi, V., & Gupta, S. (2018). *Hyperaldosteronism*.

Dong, K., Zhao, H., Tong, T., & Wan, X. (2016). NBLDA: Negative binomial linear discriminant analysis for RNA-Seq data. *BMC Bioinformatics*, *17*(1), 1–10.

Dudoit, S., Fridlyand, J., & Speed, T. P. (2002). Comparison of discrimination methods for the classification of tumors using gene expression data. *Journal of the American Statistical Association*, *97*(457), 77–87.

Eker, F. (2019). *Nonfonksiyonel adrenal insidentalomaların bisfenol A ile ilişkisi*.

Eisenhofer, G., Huynh, T.-T., Hiroi, M., & Pacak, K. (2001). Understanding catecholamine metabolism as a guide to the biochemical diagnosis of pheochromocytoma. *Reviews in Endocrine and Metabolic Disorders*, *2*, 297–311.

Elmohr, M. M., Fuentes, D., Habra, M. A., Bhosale, P. R., Qayyum, A. A., Gates, E., Morshid, A. I., Hazle, J. D., Elsayes, K. M. 2019. "Machine learning-based texture analysis

for differentiation of large adrenal cortical tumours on CT", Clinical radiology, 74(10), 818-e1.

Erceylan, Ö. F., Savaş, A., & Göv, E. (2021). Targeting the tumor stroma: Integrative analysis reveal GATA2 and TORYAIP1 as novel prognostic targets in breast and ovarian cancer. *Turkish Journal of Biology*, *45*(2), 127–137.

Feelders, R. A., Newell-Price, J., Pivonello, R., Nieman, L. K., Hofland, L. J., & Lacroix, A. (2019). Advances in the medical treatment of Cushing's syndrome. *The Lancet Diabetes & Endocrinology*, *7*(4), 300–312.

Fojo, T., Huff, L., Litman, T. (2020). Metastatic and recurrent adrenocortical cancer is not defined by its genomic landscape. BMC Med Genomics, 13(1):165.

Gaujoux, S., Weinandt, M., Bonnet, S. (2017). Surgical treatment of adrenal carcinoma. J Visc Surg, 154(5), 335–343.

Gautier, L., Cope, L., Bolstad, B.M. (2004). affy—Analysis of Affymetrix GeneChip data at the probe level. Bioinformatics, 20(3), 307–315.

Ghosh, S., Mondal, S., & Ghosh, B. (2014). A comparative study of breast cancer detection based on SVM and MLP BPN classifier. *2014 First International Conference on Automation, Control, Energy and Systems (ACES)*, 1–4.

Goksuluk, D., Zararsiz, G., Korkmaz, S., Eldem, V., Zararsiz, G. E., Ozcetin, E., Ozturk, A., & Karaagaoglu, A. E. (2019). MLSeq: Machine learning interface for RNA-sequencing data. *Computer Methods and Programs in Biomedicine*, *175*, 223–231.

Gou, J., Ma, H., Ou, W., Zeng, S., Rao, Y., & Yang, H. (2019). A generalized mean distance-based k-nearest neighbor classifier. *Expert Systems with Applications*, *115*, 356–372.

Gov, E., & Arga, K. Y. (2017). Differential co-expression analysis reveals a novel prognostic gene module in ovarian cancer. *Scientific Reports*, *7*(1), 1–10.

Gov, E., Kori, M., & Arga, K. Y. (2017). Multiomics analysis of tumor microenvironment reveals Gata2 and miRNA-124-3p as potential novel biomarkers in ovarian cancer. *Omics: A Journal of Integrative Biology*, *21*(10), 603–615.

Gov, E. (2019). İnsan Protein Etkileşim Ağı Kullanarak Tiroid Karsinomu İle İlgili Moleküler Hedef ve Biyoişaretçi Adayların Belirlenmesi. *International Journal of Advances in Engineering and Pure Sciences*, *31*(3), 245–254.

Han, H., Cho, J.-W., Lee, S., Yun, A., Kim, H., Bae, D., Yang, S., Kim, C. Y., Lee, M., & Kim, E. (2018). TRRUST v2: An expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Research*, *46*(D1), D380–D386.

Harari, A., & Inabnet III, W. B. (2011). Malignant pheochromocytoma: A review. *The American Journal of Surgery*, *201*(5), 700–708.

Heaton, J.H., Wood, M.A., Kim, A.C. (2012). Progression to adrenocortical tumorigenesis in mice and humans through insulin-like growth factor 2 and β-catenin. Am J Pathol, 181(3):1017-33.

Hsu, S.-D., Lin, F.-M., Wu, W.-Y., Liang, C., Huang, W.-C., Chan, W.-L., Tsai, W.-T., Chen, G.-Z., Lee, C.-J., & Chiu, C.-M. (2011). miRTarBase: A database curates experimentally validated microRNA–target interactions. *Nucleic Acids Research*, *39*(suppl_1), D163–D169.

Iacobone, M., Citton, M., Scarpa, M., Viel, G., Boscaro, M., & Nitti, D. (2015). Systematic review of surgical treatment of subclinical Cushing's syndrome. *Journal of British Surgery*, *102*(4), 318–330.

Iwamoto, Y., Tatsumi, F., Itoh, T., Sasaki, T., Mori, S., Sugisaki, T., Nakao, E., Ohnishi, M., Kusano, T., & Takenouchi, H. (2022). Usefulness of cortisol/ACTH ratio (CAR) for diagnosis of cushing's syndrome: Comparison of CAR with findings in dexamethasone suppression test. *Scientific Reports*, *12*(1), 17680.

Karimi, F., Dehghanian, A., Fallahi, M., & Dalfardi, B. (2019). Pure Androgen-Secreting Adrenocortical Carcinoma Presenting with Hypoglycemia. *Archives of Iranian Medicine (AIM)*, *22*(9).

Khosravi, P., Liang, Y., Choi, Y., & Broeck, G. V. den. (2019). What to expect of classifiers? Reasoning about logistic regression with missing features. *ArXiv Preprint ArXiv:1903.01620*.

Kori, M., Gov, E., & Arga, K. Y. (2016). Molecular signatures of ovarian diseases: Insights from network medicine perspective. *Systems Biology in Reproductive Medicine*, *62*(4), 266–282.

Kori, M., Gov, E., & Arga, K. Y. (2019). Novel genomic biomarker candidates for cervical cancer as identified by differential co-expression network analysis. *Omics: A Journal of Integrative Biology*, *23*(5), 261–273.

Ku, E. J., Lee, C., Shim, J., Lee, S., Kim, K.-A., Kim, S. W., Rhee, Y., Kim, H.-J., Lim, J. S., Chung, C. H., Chun, S. W., Yoo, S.-J., Ryu, O.-H., Cho, H. C., Hong, A. R., Ahn, C. H., Kim, J. H., & Choi, M. H. (2021). Metabolic Subtyping of Adrenal Tumors: Prospective Multi-Center Cohort Study in Korea. *Endocrinology and Metabolism*, *36*(5), 1131–1141. https://doi.org/10.3803/EnM.2021.1149

Lacroix, A., Feelders, R. A., Stratakis, C. A., & Nieman, L. K. (2015). Cushing's syndrome. *The Lancet*, *386*(9996), 913–927.

Larranaga, P., Calvo, B., Santana, R. (2006). Machine learning in bioinformatics. Brief Bioinformatics, 7(1), 86-112.

Lee, K. A., & Jin, H. Y. (2019). Non-surgically treated case of nonfunctioning ruptured adrenal adenoma in a patient on hemodialysis. *Endocrine*, *65*(1), 213–216.

Lenders, J. W., & Eisenhofer, G. (2011). Pathophysiology and diagnosis of disorders of the adrenal medulla: Focus on pheochromocytoma. *Comprehensive Physiology*, *4*(2), 691–713.

Li, F., Zhang, L., Chen, B., Gao, D., Cheng, Y., Zhang, X., Yang, Y., Gao, K., Huang, Z., & Peng, J. (2018). A Light Gradient Boosting Machine for Remainning Useful Life Estimation of Aircraft Engines. *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 3562–3567. https://doi.org/10.1109/ITSC.2018.8569801

Li, X., Gao, Y., Xu, Z. (2020). Identification of prognostic genes in adrenocortical carcinoma microenvironment based on bioinformatic methods. Cancer Med, 9(3), 1161–1172.

Li, Y., Bian, X., Ouyang, J., Wei, S., He, M., & Luo, Z. (2018). Nomograms to predict overall survival and cancer-specific survival in patients with adrenocortical carcinoma. *Cancer Management and Research*, 6949–6959.

Lin, R., Zhou, Z., You, S., Rao, R., & Kuo, C.-C. J. (2022). Geometrical interpretation and design of multilayer perceptrons. *IEEE Transactions on Neural Networks and Learning Systems*.

Lonser, R. R., Nieman, L., & Oldfield, E. H. (2017). Cushing's disease: Pathobiology, diagnosis, and management. *Journal of Neurosurgery*, *126*(2), 404–417.

Loscalzo, J. (2017). *Network medicine*. Harvard University Press.

Luo, J., Feng, Y., Wu, X., Li, R., Shi, J., Chang, W., & Wang, J. (2023). ForestSubtype: A cancer subtype identifying approach based on high-dimensional genomic data and a parallel random forest. *BMC Bioinformatics*, *24*(1), 1–19.

Maillo, J., Ramírez, S., Triguero, I., & Herrera, F. (2017). kNN-IS: An Iterative Spark-based design of the k-Nearest Neighbors classifier for big data. *Knowledge-Based Systems*, *117*, 3–15.

Manger, W. M., & Gifford Jr, R. W. (1993). Pheochromocytoma: Current diagnosis and management. *Cleveland Clinic Journal of Medicine*, *60*(5), 365–378.

Marquardt, A., Landwehr, L.-S., Ronchi, C. L., Di Dalmazi, G., Riester, A., Kollmannsberger, P., Altieri, B., Fassnacht, M., & Sbiera, S. (2021). Identifying new potential biomarkers in adrenocortical tumors based on mRNA expression data using machine learning. *Cancers*, *13*(18), 4671.

Mohan, D. R., Lerario, A. M., & Hammer, G. D. (2018). Therapeutic targets for adrenocortical carcinoma in the genomics era. *Journal of the Endocrine Society*, *2*(11), 1259–1274.

Nagaraja, V., Eslick, G. D., & Edirimanne, S. (2015). Recurrence and functional outcomes of partial adrenalectomy: A systematic review and meta-analysis. *International Journal of Surgery*, *16*, 7–13.

Nguyen-Martin, M. A., & Hammer, G. D. (2006). Pheochromocytoma: An update on risk groups, diagnosis, and management. *Hospital Physician*, *42*(2), 17.

Nieman, L. K. (2015). Cushing's syndrome: Update on signs, symptoms and biochemical screening. *European Journal of Endocrinology*, *173*(4), M33–M38.

Pang, Y., Chen, X., Ji, T. (2023).The Chromatin Remodeler ATRX: Role and Mechanism in Biology and Cancer. Cancers, 15(8), 2228.

Papadopoulou-Marketou, N., Vaidya, A., Dluhy, R., & Chrousos, G. P. (2015). *Hyperaldosteronism*.

Pignatti, E., & Flück, C. E. (2021). Adrenal cortex development and related disorders leading to adrenal insufficiency. *Molecular and Cellular Endocrinology*, *527*, 111206.

Pivonello, R., Isidori, A. M., De Martino, M. C., Newell-Price, J., Biller, B. M., & Colao, A. (2016). Complications of Cushing's syndrome: State of the art. *The Lancet Diabetes & Endocrinology*, *4*(7), 611–629.

Qi, Y. 2012. "Random forest for bioinformatics", Ensemble machine learning, 307-323.

Ramchoun, H., Ghanou, Y., Ettaouil, M., & Janati Idrissi, M. A. (2016). *Multilayer perceptron: Architecture optimization and training*.

Reincke, M., Ritzel, K., Oßwald, A., Berr, C., Stalla, G., Hallfeldt, K., Reisch, N., Schopohl, J., & Beuschlein, F. (2015). A critical reappraisal of bilateral adrenalectomy for ACTH-dependent Cushing's syndrome. *European Journal of Endocrinology*, *173*(4), M23–M32.

Saxena, S., Murthy, T.K., Chandramohan, V. (2022). In-silico analysis of deleterious single nucleotide polymorphisms of PNMT gene. Mol Simul, 48(16), 1411–1425.

Schapire, R. Machine Learning Algorithms for Classification; Princeton University: Princeton, NJ, USA, 2015; p. 10.

Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, *4*(1), 44–57.

Smyth, G.K. (2005). LIMMA: Linear models for microarray data. In: Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Gentleman R, Carey VJ, Huber W, Irizarry RA, and Dudoit S, eds. Springer, New York, New York, USA. pp. 397–420.

Soon, P.S.H., Gill, A.J., Benn, D.E. (2009). Microarray gene expression and immunohistochemistry analyses of adrenocortical tumors identify IGF2 and Ki-67 as useful in differentiating carcinomas from adenomas. Endocr Relat Cancer, 16(2), 573–583.

Subramanian, C., Cohen, M.S. (2021). Identification of novel lipid metabolic biomarkers associated with poor adrenocortical carcinoma prognosis using integrated bioinformatics. Surgery.

Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P.-L., & Ideker, T. (2011). Cytoscape 2.8: New features for data integration and network visualization. *Bioinformatics*, *27*(3), 431–432.

Taha, A. A., & Malebary, S. J. (2020). An Intelligent Approach to Credit Card Fraud Detection Using an Optimized Light Gradient Boosting Machine. *IEEE Access*, *8*, 25579–25587. https://doi.org/10.1109/ACCESS.2020.2971354

Tang, J., Fang, Y., Xu, Z. (2023). Establishment of prognostic models of adrenocortical carcinoma using machine learning and big data. Front Surg, 9, 966307.

Tanha, J., van Someren, M., & Afsarmanesh, H. (2017). Semi-supervised self-training for decision tree classifiers. *International Journal of Machine Learning and Cybernetics*, *8*(1), 355–370. https://doi.org/10.1007/s13042-015-0328-7

Terzolo, M., Daffara, F., Ardito, A. (2014). Management of adrenal cancer: a 2013 update. J Endocrinol Invest. 37, 207-217.

Toniato, A., Boschin, I. M., Opocher, G., Guolo, A., Pelizzo, M., & Mantero, F. (2007). Is the laparoscopic adrenalectomy for pheochromocytoma the best treatment? *Surgery*, *141*(6), 723–727.

Tsentidis, C., Bampilis, A., Ntova, V., Fragkos, D., Panos, C., Limniati, C., & Kassi, G. (2019). Metabolic syndrome as a predictor of adrenal functional status: A discriminant multivariate analysis versus logistic regression analysis. *Hormone and Metabolic Research*, *51*(01), 47–53.

Vaidya, A., & Carey, R. M. (2020). Evolution of the primary aldosteronism syndrome: Updating the approach. *The Journal of Clinical Endocrinology & Metabolism*, *105*(12), 3771–3783.

Vogg, N., Müller, T., Floren, A., Dandekar, T., Riester, A., Dischinger, U., Kurlbaum, M., Kroiss, M., & Fassnacht, M. (2023). Simplified urinary steroid profiling by LC-MS as diagnostic tool for malignancy in adrenocortical tumors. *Clinica Chimica Acta*, *543*, 117301. https://doi.org/10.1016/j.cca.2023.117301

Wang, F., Liu, J., Zhang, R., Bai, Y., Li, C., Li, B., Liu, H., & Zhang, T. (2018). CT and MRI of adrenal gland pathologies. *Quantitative Imaging in Medicine and Surgery*, *8*(8), 853.

Wang, J.-J., Peng, K.-Y., Wu, V.-C., Tseng, F.-Y., & Wu, K.-D. (2017). CTNNB1 mutation in aldosterone producing adenoma. *Endocrinology and Metabolism*, *32*(3), 332–338.

Wang, T., & Rainey, W. E. (2012). Human adrenocortical carcinoma cell lines. *Molecular and Cellular Endocrinology*, *351*(1), 58–65.

Witten, D. M. (2011). Classification and clustering of sequencing data using a Poisson model. *The Annals of Applied Statistics*, *5*(4), 2493–2518.

Weiss, I.D., Huff, L.M., Evbuomwan, M.O. (2017). Screening of cancer tissue arrays identifies CXCR4 on adrenocortical carcinoma: correlates with expression and quantification on metastases using ⁶⁴Cu-plerixafor for PET. Oncotarget, 8(43):73387-73406

West, A.N., Neale, G.A., Pounds, S. (2007). Gene expression profiling of childhood adrenocortical tumors. Cancer Res, 67(2):600-8.

Yi, X., Wan, Y., Cao, W., Peng, K., Li, X., & Liao, W. (2022). Identification of Four Novel Prognostic Biomarkers and Construction of Two Nomograms in Adrenocortical Carcinoma: A Multi-Omics Data Study via Bioinformatics and Machine Learning Methods. *Frontiers in Molecular Biosciences*, *9*, 878073.

Yuan, L., Qian, G., Chen, L., Wu, C.-L., Dan, H. C., Xiao, Y., & Wang, X. (2018). Co-expression network analysis of biomarkers for adrenocortical carcinoma. *Frontiers in Genetics*, *9*, 328.

Zennaro, M.-C., Boulkroun, S., & Fernandes-Rosa, F. (2017). Genetic causes of functional adrenocortical adenomas. *Endocrine Reviews*, *38*(6), 516–537.

Zhang, Q., Liu, P., Wang, X., Zhang, Y., Han, Y., & Yu, B. (2021). StackPDB: Predicting DNA-binding proteins based on XGB-RFE feature optimization and stacked ensemble classifier. *Applied Soft Computing*, *99*, 106921. https://doi.org/10.1016/j.asoc.2020.106921

Zhang, Y., Cao, G., Wang, B., & Li, X. (2019). A novel ensemble method for k-nearest neighbor. *Pattern Recognition*, *85*, 13–25.

Zheng, S., Cherniack, A. D., Dewal, N., Moffitt, R. A., Danilova, L., Murray, B. A., Lerario, A. M., Else, T., Knijnenburg, T. A., & Ciriello, G. (2016). Comprehensive pan-genomic characterization of adrenocortical carcinoma. *Cancer Cell*, *29*(5), 723–736.

Zhou, Z., Huang, H., & Liang, Y. (2021). Cancer classification and biomarker selection via a penalized logsum network-based logistic regression model. *Technology and Health Care*, *29*(S1), 287–295.

Zou, X., Hu, Y., Tian, Z., & Shen, K. (2019). Logistic regression model optimization and case analysis. *2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT)*, 135–139.

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., Benner, C., & Chanda, S. K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications*, *10*(1), 1–10.

APPENDIX

Table A1: The results of enrichment analysis of the GSE10927 dataset in the ACA-N group.

Category	Term	P Value
KEGG_PATHWAY	hsa05323:Rheumatoid arthritis	3.03E-08
KEGG_PATHWAY	hsa04672:Intestinal immune network for IgA production	3.814E-08
KEGG_PATHWAY	hsa04940:Type I diabetes mellitus	4.529E-08
KEGG_PATHWAY	hsa05310:Asthma	4.61E-08
KEGG_PATHWAY	hsa05416:Viral myocarditis	1.165E-07
KEGG_PATHWAY	hsa05332:Graft-versus-host disease	2.709E-07
REACTOME_PATHWAY	R-HSA-202430~Translocation of ZAP-70 to Immunological synapse	5.535E-07
KEGG_PATHWAY	hsa05330:Allograft rejection	6.159E-07
KEGG_PATHWAY	hsa04145:Phagosome	1.046E-06
KEGG_PATHWAY	hsa05150:Staphylococcus aureus infection	1.221E-06
REACTOME_PATHWAY	R-HSA-162582~Signal Transduction	1.24E-06
KEGG_PATHWAY	hsa04640:Hematopoietic cell lineage	1.992E-06
KEGG_PATHWAY	hsa05321:Inflammatory bowel disease	2.178E-06
REACTOME_PATHWAY	R-HSA-202427~Phosphorylation of CD3 and TCR zeta chains	2.527E-06
REACTOME_PATHWAY	R-HSA-389948~PD-1 signaling	3.93E-06
KEGG_PATHWAY	hsa05140:Leishmaniasis	4.377E-06
KEGG_PATHWAY	hsa05200:Pathways in cancer	1.208E-05
REACTOME_PATHWAY	R-HSA-202433~Generation of second messenger molecules	2.152E-05
KEGG_PATHWAY	hsa05320:Autoimmune thyroid disease	2.801E-05
KEGG_PATHWAY	hsa05164:Influenza A	2.815E-05

Table A2: The results of the enrichment analysis of the GSE12368 dataset in the ACA-N group.

Category	Term	P Value
REACTOME_PATHWAY	R-HSA-112316~Neuronal System	6.46E-12
REACTOME_PATHWAY	R-HSA-162582~Signal Transduction	2.01E-11
REACTOME_PATHWAY	R-HSA-112315~Transmission across Chemical Synapses	4.66E-11
KEGG_PATHWAY	hsa04713:Circadian entrainment	1.24E-09
KEGG_PATHWAY	hsa04024:cAMP signaling pathway	4.49E-09
KEGG_PATHWAY	hsa04728:Dopaminergic synapse	2.75E-08
REACTOME_PATHWAY	R-HSA-112314~Neurotransmitter receptors and postsynaptic signal transmission	2.91E-07
KEGG_PATHWAY	hsa05031:Amphetamine addiction	4.36E-07
KEGG_PATHWAY	hsa04925:Aldosterone synthesis and secretion	2.48E-06
KEGG_PATHWAY	hsa04725:Cholinergic synapse	2.49E-06

REACTOME_PATHWAY	R-HSA-936837~Ion transport by P-type ATPases	3.8E-06
REACTOME_PATHWAY	R-HSA-397014~Muscle contraction	6.4E-06
KEGG_PATHWAY	hsa04020:Calcium signaling pathway	7.82E-06
REACTOME_PATHWAY	R-HSA-8986944~Transcriptional Regulation by MECP2	1.25E-05
REACTOME_PATHWAY	R-HSA-9675108~Nervous system development	1.35E-05
KEGG_PATHWAY	hsa04261:Adrenergic signaling in cardiomyocytes	1.61E-05
REACTOME_PATHWAY	R-HSA-9006934~Signaling by Receptor Tyrosine Kinases	1.84E-05
KEGG_PATHWAY	hsa05200:Pathways in cancer	2.47E-05
REACTOME_PATHWAY	R-HSA-5576892~Phase 0 - rapid depolarisation	2.91E-05
KEGG_PATHWAY	hsa04724:Glutamatergic synapse	3.22E-05

Table A3: The results of the enrichment analysis of the GSE75415 dataset in the ACA-N group.

Category	Term	P Value
REACTOME_PATHWAY	R-HSA-5663202~Diseases of signal transduction by growth factor receptors and second messengers	1.89E-14
REACTOME_PATHWAY	R-HSA-202424~Downstream TCR signaling	7.55E-14
REACTOME_PATHWAY	R-HSA-195253~Degradation of beta-catenin by the destruction complex	5.3E-13
REACTOME_PATHWAY	R-HSA-8852276~The role of GTSE1 in G2/M progression after G2 checkpoint	6.49E-12
REACTOME_PATHWAY	R-HSA-168256~Immune System	7.83E-12
REACTOME_PATHWAY	R-HSA-8953897~Cellular responses to stimuli	1.02E-11
REACTOME_PATHWAY	R-HSA-187577~SCF(Skp2)-mediated degradation of p27/p21	1.47E-11
REACTOME_PATHWAY	R-HSA-1430728~Metabolism	1.52E-11
REACTOME_PATHWAY	R-HSA-9759194~Nuclear events mediated by NFE2L2	1.57E-11
REACTOME_PATHWAY	R-HSA-9675108~Nervous system development	3.81E-11
REACTOME_PATHWAY	R-HSA-9755511~KEAP1-NFE2L2 pathway	4.18E-11
REACTOME_PATHWAY	R-HSA-69563~p53-Dependent G1 DNA Damage Response	4.63E-11
REACTOME_PATHWAY	R-HSA-69580~p53-Dependent G1/S DNA damage checkpoint	4.63E-11
REACTOME_PATHWAY	R-HSA-202403~TCR signaling	5.11E-11
REACTOME_PATHWAY	R-HSA-382556~ABC-family proteins mediated transport	5.92E-11
REACTOME_PATHWAY	R-HSA-5653656~Vesicle-mediated transport	6.75E-11
REACTOME_PATHWAY	R-HSA-453279~Mitotic G1 phase and G1/S transition	8.09E-11
REACTOME_PATHWAY	R-HSA-191273~Cholesterol biosynthesis	9.12E-11
REACTOME_PATHWAY	R-HSA-69615~G1/S DNA Damage Checkpoints	1.16E-10
REACTOME_PATHWAY	R-HSA-5678895~Defective CFTR causes cystic fibrosis	1.54E-10

Table A4: The results of the enrichment analysis of the GSE10927 dataset in the ACC-N group.

Category	Term	P Value

REACTOME_PATHWAY	R-HSA-1640170~Cell Cycle	2.32E-16
REACTOME_PATHWAY	R-HSA-69278~Cell Cycle, Mitotic	7.27E-15
REACTOME_PATHWAY	R-HSA-68877~Mitotic Prometaphase	2.54E-12
KEGG_PATHWAY	hsa04110:Cell cycle	2.35E-11
REACTOME_PATHWAY	R-HSA-2500257~Resolution of Sister Chromatid Cohesion	3.53E-11
REACTOME_PATHWAY	R-HSA-141424~Amplification of signal from the kinetochores	1.57E-10
REACTOME_PATHWAY	R-HSA-141444~Amplification of signal from unattached kinetochores via a MAD2 inhibitory signal	1.57E-10
REACTOME_PATHWAY	R-HSA-69618~Mitotic Spindle Checkpoint	1.95E-10
REACTOME_PATHWAY	R-HSA-69620~Cell Cycle Checkpoints	2.93E-10
REACTOME_PATHWAY	R-HSA-69190~DNA strand elongation	7.27E-10
REACTOME_PATHWAY	R-HSA-176187~Activation of ATR in response to replication stress	1.47E-09
REACTOME_PATHWAY	R-HSA-9648025~EML4 and NUDC in mitotic spindle formation	3.45E-09
REACTOME_PATHWAY	R-HSA-68962~Activation of the pre-replicative complex	1.7E-08
REACTOME_PATHWAY	R-HSA-5663220~RHO GTPases Activate Formins	5.95E-08
KEGG_PATHWAY	hsa04115:p53 signaling pathway	9.58E-08
REACTOME_PATHWAY	R-HSA-212436~Generic Transcription Pathway	1.5E-07
REACTOME_PATHWAY	R-HSA-6804756~Regulation of TP53 Activity through Phosphorylation	1.67E-07
REACTOME_PATHWAY	R-HSA-69242~S Phase	2.06E-07
REACTOME_PATHWAY	R-HSA-2555396~Mitotic Metaphase and Anaphase	3.22E-07
REACTOME_PATHWAY	R-HSA-74160~Gene expression (Transcription)	4.15E-07

Table A5: The results of the enrichment analysis of the GSE12368 dataset in the ACC-N group.

Category	**Term**	**P Value**
REACTOME_PATHWAY	R-HSA-68962~Activation of the pre-replicative complex	5.07E-11
REACTOME_PATHWAY	R-HSA-1640170~Cell Cycle	8.82E-11
REACTOME_PATHWAY	R-HSA-69190~DNA strand elongation	1.91E-10
REACTOME_PATHWAY	R-HSA-69278~Cell Cycle, Mitotic	2.24E-10
KEGG_PATHWAY	hsa04110:Cell cycle	7.41E-10
REACTOME_PATHWAY	R-HSA-141424~Amplification of signal from the kinetochores	1.07E-09
REACTOME_PATHWAY	R-HSA-141444~Amplification of signal from unattached kinetochores via a MAD2 inhibitory signal	1.07E-09
REACTOME_PATHWAY	R-HSA-176187~Activation of ATR in response to replication stress	2.07E-09
REACTOME_PATHWAY	R-HSA-2500257~Resolution of Sister Chromatid Cohesion	4.6E-09
REACTOME_PATHWAY	R-HSA-69618~Mitotic Spindle Checkpoint	5.21E-09
REACTOME_PATHWAY	R-HSA-69620~Cell Cycle Checkpoints	8.14E-09
REACTOME_PATHWAY	R-HSA-68877~Mitotic Prometaphase	1.25E-08

REACTOME_PATHWAY	R-HSA-9648025~EML4 and NUDC in mitotic spindle formation	2.07E-08
KEGG_PATHWAY	hsa03030:DNA replication	3.54E-08
REACTOME_PATHWAY	R-HSA-176974~Unwinding of DNA	1.51E-07
REACTOME_PATHWAY	R-HSA-162582~Signal Transduction	2.02E-07
REACTOME_PATHWAY	R-HSA-453279~Mitotic G1 phase and G1/S transition	4.78E-07
REACTOME_PATHWAY	R-HSA-5663220~RHO GTPases Activate Formins	1.95E-06
REACTOME_PATHWAY	R-HSA-69206~G1/S Transition	3.2E-06
KEGG_PATHWAY	hsa00280:Valine, leucine and isoleucine degradation	6.08E-06

Table A6: The results of the enrichment analysis of the GSE19750 dataset in the ACC-N group.

Category	**Term**	**P Value**
REACTOME_PATHWAY	R-HSA-74160~Gene expression (Transcription)	9.16E-46
REACTOME_PATHWAY	R-HSA-73857~RNA Polymerase II Transcription	1.77E-36
REACTOME_PATHWAY	R-HSA-212436~Generic Transcription Pathway	2.53E-30
REACTOME_PATHWAY	R-HSA-1640170~Cell Cycle	1.57E-29
REACTOME_PATHWAY	R-HSA-69278~Cell Cycle, Mitotic	4.54E-25
REACTOME_PATHWAY	R-HSA-68877~Mitotic Prometaphase	3.26E-16
KEGG_PATHWAY	hsa05168:Herpes simplex virus 1 infection	2.96E-15
KEGG_PATHWAY	hsa04110:Cell cycle	4.1E-15
REACTOME_PATHWAY	R-HSA-3700989~Transcriptional Regulation by TP53	9.88E-14
REACTOME_PATHWAY	R-HSA-5633007~Regulation of TP53 Activity	2.28E-12
REACTOME_PATHWAY	R-HSA-68886~M Phase	5.87E-12
REACTOME_PATHWAY	R-HSA-69620~Cell Cycle Checkpoints	3.32E-11
REACTOME_PATHWAY	R-HSA-72203~Processing of Capped Intron-Containing Pre-mRNA	6.3E-11
REACTOME_PATHWAY	R-HSA-2500257~Resolution of Sister Chromatid Cohesion	3.71E-10
REACTOME_PATHWAY	R-HSA-69618~Mitotic Spindle Checkpoint	4.82E-10
KEGG_PATHWAY	hsa03013:Nucleocytoplasmic transport	1.18E-09
REACTOME_PATHWAY	R-HSA-6804756~Regulation of TP53 Activity through Phosphorylation	1.24E-09
REACTOME_PATHWAY	R-HSA-141424~Amplification of signal from the kinetochores	1.64E-09
REACTOME_PATHWAY	R-HSA-141444~Amplification of signal from unattached kinetochores via a MAD2 inhibitory signal	1.64E-09
REACTOME_PATHWAY	R-HSA-9716542~Signaling by Rho GTPases, Miro GTPases and RHOBTB3	3.67E-09

Table A7: The results of the enrichment analysis of the GSE75415 dataset in the ACC-N group.

Category	**Term**	**P Value**

REACTOME_PATHWAY	R-HSA-69278~Cell Cycle, Mitotic	8.51E-19
REACTOME_PATHWAY	R-HSA-1640170~Cell Cycle	2.03E-17
REACTOME_PATHWAY	R-HSA-69620~Cell Cycle Checkpoints	5.01E-14
REACTOME_PATHWAY	R-HSA-453279~Mitotic G1 phase and G1/S transition	1.06E-13
REACTOME_PATHWAY	R-HSA-69239~Synthesis of DNA	2.17E-13
REACTOME_PATHWAY	R-HSA-69242~S Phase	2.76E-13
REACTOME_PATHWAY	R-HSA-68882~Mitotic Anaphase	2.78E-13
REACTOME_PATHWAY	R-HSA-2555396~Mitotic Metaphase and Anaphase	3.42E-13
KEGG_PATHWAY	hsa04110:Cell cycle	6.87E-13
REACTOME_PATHWAY	R-HSA-69206~G1/S Transition	1.01E-12
REACTOME_PATHWAY	R-HSA-2467813~Separation of Sister Chromatids	1.22E-12
REACTOME_PATHWAY	R-HSA-453276~Regulation of mitotic cell cycle	7.25E-12
REACTOME_PATHWAY	R-HSA-174143~APC/C-mediated degradation of cell cycle proteins	7.25E-12
REACTOME_PATHWAY	R-HSA-69052~Switching of origins to a post-replicative state	2.98E-11
REACTOME_PATHWAY	R-HSA-176409~APC/C:Cdc20 mediated degradation of mitotic proteins	8.31E-11
REACTOME_PATHWAY	R-HSA-176814~Activation of APC/C and APC/C:Cdc20 mediated degradation of mitotic proteins	1.21E-10
REACTOME_PATHWAY	R-HSA-8852276~The role of GTSE1 in G2/M progression after G2 checkpoint	1.21E-10
REACTOME_PATHWAY	R-HSA-68949~Orc1 removal from chromatin	4.08E-10
REACTOME_PATHWAY	R-HSA-69275~G2/M Transition	4.75E-10
REACTOME_PATHWAY	R-HSA-176408~Regulation of APC/C activators between G1/S and early anaphase	4.96E-10

Table A8: The results of the enrichment analysis of the GSE10927 dataset in the ACC-ACA group.

Category	**Term**	**P Value**
REACTOME_PATHWAY	R-HSA-1640170~Cell Cycle	1.75E-19
REACTOME_PATHWAY	R-HSA-69278~Cell Cycle, Mitotic	8.56E-18
REACTOME_PATHWAY	R-HSA-2500257~Resolution of Sister Chromatid Cohesion	9.85E-16
REACTOME_PATHWAY	R-HSA-68877~Mitotic Prometaphase	3.09E-15
REACTOME_PATHWAY	R-HSA-9648025~EML4 and NUDC in mitotic spindle formation	2.2E-14
REACTOME_PATHWAY	R-HSA-141444~Amplification of signal from unattached kinetochores via a MAD2 inhibitory signal	1.16E-13
REACTOME_PATHWAY	R-HSA-141424~Amplification of signal from the kinetochores	1.16E-13
REACTOME_PATHWAY	R-HSA-69620~Cell Cycle Checkpoints	1.58E-13
REACTOME_PATHWAY	R-HSA-69618~Mitotic Spindle Checkpoint	2.75E-13
REACTOME_PATHWAY	R-HSA-69190~DNA strand elongation	5.58E-12
REACTOME_PATHWAY	R-HSA-5663220~RHO GTPases Activate Formins	6.29E-12
KEGG_PATHWAY	hsa04110:Cell cycle	1.15E-11
REACTOME_PATHWAY	R-HSA-2555396~Mitotic Metaphase and Anaphase	3.83E-10
REACTOME_PATHWAY	R-HSA-68882~Mitotic Anaphase	7.52E-10

REACTOME_PATHWAY	R-HSA-2467813~Separation of Sister Chromatids	8.3E-10
REACTOME_PATHWAY	R-HSA-176187~Activation of ATR in response to replication stress	5.55E-09
REACTOME_PATHWAY	R-HSA-453279~Mitotic G1 phase and G1/S transition	1.16E-08
REACTOME_PATHWAY	R-HSA-68962~Activation of the pre-replicative complex	1.41E-08
REACTOME_PATHWAY	R-HSA-69273~Cyclin A/B1/B2 associated events during G2/M transition	8.2E-08
REACTOME_PATHWAY	R-HSA-176974~Unwinding of DNA	8.79E-08

Table A9: The results of the enrichment analysis of the GSE12368 dataset in the ACC-ACA group.

Category	**Term**	**P Value**
REACTOME_PATHWAY	R-HSA-69278~Cell Cycle, Mitotic	7.19E-15
REACTOME_PATHWAY	R-HSA-1640170~Cell Cycle	3.74E-14
REACTOME_PATHWAY	R-HSA-2500257~Resolution of Sister Chromatid Cohesion	1.1E-12
REACTOME_PATHWAY	R-HSA-212436~Generic Transcription Pathway	1.23E-12
REACTOME_PATHWAY	R-HSA-68877~Mitotic Prometaphase	3.47E-12
REACTOME_PATHWAY	R-HSA-74160~Gene expression (Transcription)	1.34E-11
REACTOME_PATHWAY	R-HSA-141424~Amplification of signal from the kinetochores	2.52E-11
REACTOME_PATHWAY	R-HSA-141444~Amplification of signal from unattached kinetochores via a MAD2 inhibitory signal	2.52E-11
REACTOME_PATHWAY	R-HSA-73857~RNA Polymerase II Transcription	8.29E-11
REACTOME_PATHWAY	R-HSA-9648025~EML4 and NUDC in mitotic spindle formation	1.59E-10
REACTOME_PATHWAY	R-HSA-69190~DNA strand elongation	2.59E-10
REACTOME_PATHWAY	R-HSA-69618~Mitotic Spindle Checkpoint	3.47E-10
REACTOME_PATHWAY	R-HSA-5663220~RHO GTPases Activate Formins	1.1E-09
REACTOME_PATHWAY	R-HSA-176187~Activation of ATR in response to replication stress	3.88E-09
REACTOME_PATHWAY	R-HSA-68962~Activation of the pre-replicative complex	6.52E-09
KEGG_PATHWAY	hsa05168:Herpes simplex virus 1 infection	7.15E-09
REACTOME_PATHWAY	R-HSA-69620~Cell Cycle Checkpoints	8.14E-09
REACTOME_PATHWAY	R-HSA-2555396~Mitotic Metaphase and Anaphase	1.55E-07
KEGG_PATHWAY	hsa04110:Cell cycle	1.58E-07
REACTOME_PATHWAY	R-HSA-68882~Mitotic Anaphase	2.65E-07

Table A10: The results of the enrichment analysis of the GSE75415 dataset in the ACC-ACA group.

Category	**Term**	**P Value**
KEGG_PATHWAY	hsa04612:Antigen processing and presentation	2.64E-13
REACTOME_PATHWAY	R-HSA-202427~Phosphorylation of CD3 and TCR zeta chains	1.48E-11
KEGG_PATHWAY	hsa05310:Asthma	7.8E-11
REACTOME_PATHWAY	R-HSA-202430~Translocation of ZAP-70 to Immunological synapse	1.01E-10
REACTOME_PATHWAY	R-HSA-2132295~MHC class II antigen presentation	2.71E-10

REACTOME_PATHWAY	R-HSA-389948~PD-1 signaling	8.26E-10
KEGG_PATHWAY	hsa05330:Allograft rejection	9.54E-10
REACTOME_PATHWAY	R-HSA-877300~Interferon gamma signaling	1.6E-09
REACTOME_PATHWAY	R-HSA-202433~Generation of second messenger molecules	2.36E-09
KEGG_PATHWAY	hsa05332:Graft-versus-host disease	3.11E-09
KEGG_PATHWAY	hsa04940:Type I diabetes mellitus	4.1E-09
KEGG_PATHWAY	hsa04145:Phagosome	4.21E-09
KEGG_PATHWAY	hsa05145:Toxoplasmosis	1.07E-08
KEGG_PATHWAY	hsa05416:Viral myocarditis	1.63E-08
KEGG_PATHWAY	hsa04672:Intestinal immune network for IgA production	1.82E-08
KEGG_PATHWAY	hsa05320:Autoimmune thyroid disease	4.38E-08
KEGG_PATHWAY	hsa05140:Leishmaniasis	2.98E-07
KEGG_PATHWAY	hsa04659:Th17 cell differentiation	3.19E-07
KEGG_PATHWAY	hsa05323:Rheumatoid arthritis	3.5E-07
REACTOME_PATHWAY	R-HSA-168256~Immune System	3.53E-07